Лекция 11

Вероятность: первые шаги

Когда мы рассматриваем какой-то сложный процесс, то во многих ситуациях результат процесса не предопределён, но возможны количественные утверждения о мере этой неопределённости. В этих ситуациях и возникает понятие «вероятности», точнее его математическая модель.

Простейшей является модель с равновозможными исходами. В этой модели есть несколько возможных результатов процесса (*исходов*) и все они одинаково возможны. Например, если мы подбрасываем монетку, то можно считать, что «орёл» и «решка» равновозможны. Аналогично, если мы бросаем шестигранный кубик, то любая из граней может оказаться верхней, как говорят, *с равной вероятностью*. *Событием* в этой модели называется некоторое подмножество множества возможных исходов. Мы считаем эти исходы благоприятными и интересуемся вероятностью этого события, то есть шансами того, что это событие произойдёт. Вероятностью события называется отношение числа благоприятных исходов к числу всех исходов или, другими словами, доля числа благоприятных среди всевозможных исходов.

Например, если мы подбрасываем монетку и интересуемся событием «выпадет орёл», то благоприятный исход один, а всего возможных исхода два. Таким образом вероятность выпадения орла равна $1/2$. Если же мы бросаем кубик, на гранях которого написаны числа от 1 до 6, и хотим посчитать вероятность того, что выпавшее число делится на 3, то благоприятными исходами будет выпадение 3 и 6, а всего исходов 6. Так что вероятность рассматриваемого события равна $1/3$.

Интуитивно равновозможность можно трактовать так: если повторять один и тот же эксперимент много раз, то все исходы будут встречаться примерно одинаковое число раз. И тогда доля благоприятных исходов среди всех будет приблизительно равна вероятности интересующего нас события.

Точный смысл, стоящий за этой интуицией, не так легко сформулировать. Теория вероятностей как математическая дисциплина и её продолжение — математическая статистика — предоставляют язык, на котором такая формулировка возможна. Мы здесь ограничимся *элементарной теорией вероятностей*, которую можно рассматривать как естественное обобщение перечислительной комбинаторики.

11.1 Элементарная теория вероятностей: определения

Теперь скажем всё то же самое более формально. *Вероятностным пространством* называется конечное множество U, его элементы называются *возможными исходами*. На вероятностном пространстве задана функция $\Pr\colon U \to [0,1]$, такая что $\sum_{x\in U} \Pr[x] = 1$. Функция Pr называется *вероятностным распределением*, а число $\Pr[x]$ называется *вероятностью исхода* $x \in U$. *Событием* называется произвольное подмножество $A \subseteq U$. Исходы, входящие в событие A, называются *благоприятными* (для события A). Вероятностью события A называется число $\Pr[A] = \sum_{x\in A} Pr[x]$.

В *модели с равновозможными исходами* функция p задается формулой $\Pr[x] = 1/|U|$ для всякого $x \in U$ (такое распределение называют также *равномерным*). Тогда вероятность события A равна $\Pr[A] = |A|/|U|$.

Приведём несколько примеров равномерных распределений на разных множествах и вычисления вероятностей событий. Фактически, это задачи перечислительной комбинаторики. Мы также указываем названия, которые обычно для них используются. Важно понимать, что все эти «монетки» и «кости» — условные названия, которые приняты для наглядности. Насколько подбрасывания реальной монеты соответствуют математической модели, — трудный вопрос и мы его здесь не обсуждаем.

Пример 11.1 («Подбрасывание монеты»)**.** Вероятностное пространство: числа 0 и 1. Все исходы равновозможны.

Пример 11.2 («Подбрасывание 6 монет»)**.** Вероятностное пространство: двоичные последовательности длины 6. Все исходы равновозможны.

Пример события, то есть множества в этом пространстве: ровно три элемента последовательности равны 1 (на неформальном жаргоне говорят «выпало три решки»). Чтобы найти вероятность этого события, нужно подсчитать количество исходов в нём и поделить на общее количество исходов.

Обе задачи легко решаются средствами перечислительной комбинаторики. Общее количество двоичных последовательностей длины 6 равно 2^6. Количество последовательностей, в которых ровно три единицы, равно $\binom{6}{3}$ (нужно выбрать из множества 6 позиций 3-элементное подмножество — те позиции, на которых стоят единицы).

Поэтому вероятность события равна

$$\frac{\binom{6}{3}}{2^6} = \frac{20}{64} = \frac{5}{16}.$$

Пример 11.3 («Подбрасывание n монет»)**.** Вероятностное пространство: двоичные слова длины n. Все слова равновозможны. Какова вероятность события «на i-й позиции в слове стоит 1»?

Всего исходов 2^n (количество двоичных слов длины n). Интересующее нас событие содержит 2^{n-1} исходов: каждый такой исход задаётся выбором 0 или 1 для всех позиций кроме i-й. Вероятность этого события равна по определению $2^{n-1}/2^n = 1/2$, как и вероятность дополнительного события «на i-й позиции в слове стоит 0».

Пример 11.4 («Подбрасывание двух игральных костей»)**.** Вероятностное пространство: последовательности (x_1, x_2) длины 2, состоящие из целых чисел в диапазоне от 1 до 6. Все исходы равновозможны. Нужно найти вероятность события «сумма выпавших чисел равна 7».

Общее количество исходов $6 \cdot 6 = 36$. Исходов, отвечающих указанному событию, ровно 6: если на первом месте в благоприятном исходе стоит число i, то на втором обязано стоять число $7 - i$. Поэтому вероятность равна $6/36 = 1/6$.

Пример 11.5 («Подбрасывание трёх игральных костей»)**.** Вероятностное пространство: последовательности (x_1, x_2, x_3) длины 3, состоящие из целых чисел в диапазоне от 1 до 6. Все исходы равновозможны.

Найдём вероятность события «сумма чисел в последовательности чётна».

Общее количество исходов $6 \cdot 6 \cdot 6 = 216$. Если на первых двух кубиках выпали числа i, j, то в благоприятном исходе есть три возможности для числа на третьем кубике (три чётных числа, если сумма $i + j$ чётна, три нечётных числа, если $i + j$ нечётна).

Общее количество благоприятных исходов $6 \cdot 6 \cdot 3$, поэтому вероятность равна $6 \cdot 6 \cdot 3/6 \cdot 6 \cdot 6 = 1/2$.

Пример 11.6 («Случайная перестановка»)**.** Вероятностное пространство: перестановки $(a_1, a_2, \ldots, a_n)$ чисел от 1 до n. Все исходы равновозможны.

Проверим, что вероятность события «$a_2 > a_1 > a_3$» равна $1/6$.

Общее количество исходов (т.е. перестановок n элементов) равно $n!$. Разобьём их на непересекающиеся группы так, что в каждой группе на местах с 4-го по n-е стоят одни и те же числа.

В каждой из построенных групп $3! = 6$ перестановок (столькими способами можно разместить 3 оставшихся числа на первых трёх местах). Из этих 6 перестановок ровно одна является благоприятным исходом. Поэтому вероятность события равна $1/6$.

В общем случае каждому исходу приписана некоторая вероятность, но теперь вероятности уже не равны: некоторые исходы вероятнее других. Можно понимать это так, что теперь у каждого исхода есть вес, и у более вероятных исходов вес больше. В этом случае мы тоже могли бы сказать, что для подсчёта вероятностей событий нужно сложить веса благоприятных исходов и поделить на сумму весов всех исходов. Но чтобы не приходилось каждый раз делить, удобно считать, что сумма весов всех исходов равна 1. В частности, рассматривая одноэлементные события, мы получаем, что вес каждого исхода — это и есть его вероятность.

Пример 11.7. Пусть вероятностное пространство состоит из двоичных слов длины n и известно, что для каждого i вероятности событий «на i-й позиции в слове стоит 0», «на i-й позиции в слове стоит 1» равны $1/2$. Из этого не следует, что все слова равновозможны.

Действительно, рассмотрим такое вероятностное распределение на этом пространстве: вероятности слов из одних нулей и из одних единиц равны $1/2$, а ве-

роятности остальных слов равны 0. Легко видеть, что вероятности указанных выше событий также равны 1/2.

Пример 11.8 («Подбрасывание 5 одинаковых костей»)**.** Вероятностное пространство: такие функции из $\{1,2,3,4,5,6\}$ в натуральные числа, что сумма всех значений функции равна 5. Вероятность исхода f равна

$$\frac{5!}{6^5 \prod_{i=1}^{6} f(i)!}$$

(договоримся, что $0! = 1$).

Пример выглядит замысловато. Неясно даже, почему сумма таких вероятностей по всем таким функциям равна 1. Проверим это.

Рассмотрим многочлен от 6 переменных $(x_1 + x_2 + x_3 + x_4 + x_5 + x_6)^5$. Если в нём раскрыть скобки и не приводить подобные, получится 6^5 слагаемых (из каждой скобки выбираем одно из 6 слагаемых.

А если привести подобные, то получатся слагаемые вида

$$C_f x_1^{f(1)} x_2^{f(2)} x_3^{f(3)} x_4^{f(4)} x_5^{f(5)} x_6^{f(6)},$$

причём сумма показателей равна 5. То есть, эти слагаемые находятся во взаимно однозначном соответствии с исходами нашего вероятностного пространства.

Коэффициенты C_f в сумме дают 6^5, так что

$$\sum_f \frac{C_f}{6^5} = 1,$$

суммирование по всему вероятностному пространству.

Осталось вычислить значения C_f. Это делается аналогично формуле для биномиальных коэффициентов: выберем множество скобок, из которых взяты $f(1)$ переменных x_1, потом из оставшихся скобок выберем те, из которых взяты $f(2)$ двоек и т.д.:

$$C_f = \binom{5}{f(1)} \cdot \binom{5-f(1)}{f(2)} \cdot \ldots = \frac{5!}{f(1)!(5-f(1))!} \cdot \frac{(5-f(1))!}{f(2)!(5-f(1)-f(2))!} \cdot \ldots$$

После сокращения в числителях и знаменателях останется

$$C_f = \frac{5!}{\prod_{i=1}^{6} f(i)!}.$$

Поясним неформальный смысл этого примера. Представьте, что подбрасываются 5 одинаковых игральных костей. «Одинаковых» означает, что когда кости лежат на столе после броска, вы можете увидеть, какие очки выпали на каждой, но не можете их различить между собой.

Исходы в такой модели неравновозможны. Вероятность появления пяти единиц в 120 раз меньше вероятности появления 1, 2, 3, 4, 5 (сформулируйте это утверждение точно в терминах исходов указанного выше вероятностного пространства и проверьте результат по формуле для вероятностей).

Вероятностные пространства можно задавать по-разному, это оказывается полезным при решении задач.

Приведём один из основных примеров такого рода. Пусть вероятностное пространство состоит из перестановок чисел от 1 до n и все перестановки равновозможны. Оказывается, эту вероятностную модель можно получить другим способом. Выберем случайно и равновозможно первое число в перестановке. Затем второе число выберем случайно и равновозможно среди оставшихся и т.д.

Как задать исходы в модели последовательного случайного выбора? Каждый исход задаётся последовательностью $i_1, \ldots, i_n$, где i_k указывает на порядковый номер числа среди не использованных на предыдущих шагах. Все исходы равновозможны.

На рисунке изображен пример для $n = 3$. Исходы в таком представлении — это пути из корня дерева в листья. На каждом ребре написан элемент соответствующей последовательности (каждое ребро входит ровно в один путь из корня в листья). По каждым листом написана перестановка, которая кодируется путём в этот лист.

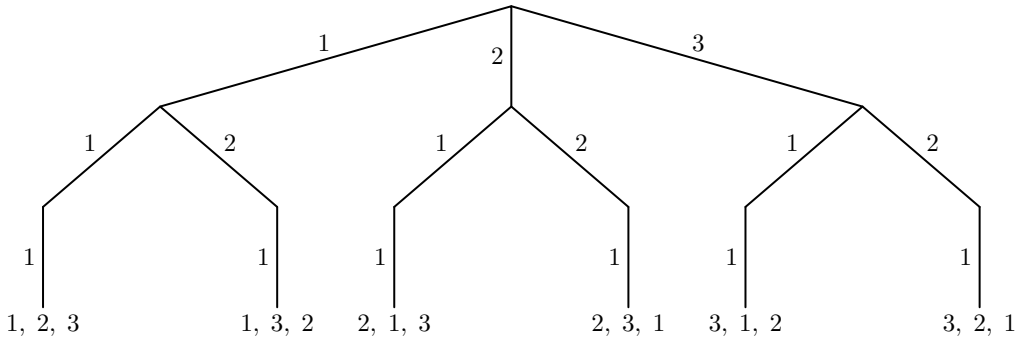

Рис. 11.1: Дерево порождения перестановок

Поскольку после $(k-1)$-го шага остаётся $n-k+1$ чисел, возможные последовательности номеров задаются условиями $1 \leqslant i_k \leqslant n-k+1$ для каждого k. Всего таких последовательностей $n \cdot (n-1) \cdot \ldots \cdot 1 = n!$, поэтому вероятность каждой перестановки в модели последовательного выбора равна $1/n!$, т.е. все перестановки равновозможны.

Аналогичным образом возможно задавать и другие вероятностные пространства. Например, пространство из примера 11.3 (двоичные строки длины n) задаётся полным бинарным деревом, вероятности на всех рёбрах равны $1/2$.

Заметим также, что вероятности выбора рёбер дерева, исходящих из данной вершины, вообще говоря, могут различаться. Чтобы получить распределение вероятностей на листьях дерева, достаточно удовлетворить двум условиям: (А) сумма вероятностей на всех рёбрах, исходящих из вершины, равна 1; (Б) вероятность листа равна произведению вероятностей на рёбрах пути, ведущего в этот лист.

На рис. 11.2 показан пример неравномерного распределения, удовлетворяющего этим условиям. Сумма вероятностей листьев (на рисунке написаны справа) равна 1, хотя это не так легко проверить сложением шести дробей. Удобнее провести про-

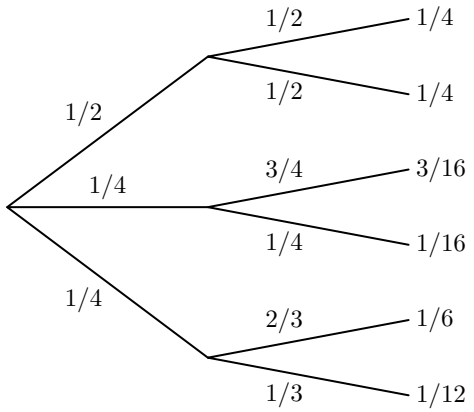

Рис. 11.2: Неравные вероятности выбора на каждом шаге

верку иначе, вынося общие множители:

$$\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{4}\cdot\frac{3}{4}+\frac{1}{4}\cdot\frac{1}{4}+\frac{1}{4}\cdot\frac{2}{3}+\frac{1}{4}\cdot\frac{1}{3}=$$
$$=\frac{1}{2}\cdot\left(\frac{1}{2}+\frac{1}{2}\right)+\frac{1}{4}\cdot\left(\frac{1}{4}+\frac{3}{4}\right)+\frac{1}{4}\cdot\left(\frac{2}{3}+\frac{1}{3}\right)=\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 1+\frac{1}{4}\cdot 1=1$$

Задача 11.9. Докажите, что и в общем случае при соблюдении условий (А) и (Б) сумма вероятностей в листьях равна 1.

Очень часто задачи про вероятности формулируются как «роман-задачи»: вместо явного указания вероятностного пространства и событий в нём, в задаче описывается некая условная ситуация, «как бы из жизни». В школьной математике такого рода задачи также активно используются: задачи про трубы и бассейны, про бригады полотёров или про велосипедистов, которые обгоняют идущие против течения катера, и т.п. Смысл в таких задачах простой: даже если не интересоваться приложениями (обычно такие задачи слишком условны, чтобы представлять интерес для приложений), решение таких задач помогает лучше представить стоящую за ними математику (в школьных примерах — уравнения и системы уравнений; в примерах про вероятности — понятия вероятностного пространства и события).

Решение таких «роман-задач» про вероятности следует всегда начинать с явной формулировки вероятностной модели и событий, о которых идёт речь в задаче. Приведём пару примеров.

Задача 11.10. Десять учеников сдают экзамен по десяти билетам. Ученики по очереди заходят в кабинет и вытягивают случайный билет из оставшихся (в частности, последний берет единственный оставшийся билет). Вася выучил только один билет.

Какова вероятность, что Васе достанется билет, который он знает, если **a)** Вася тянет билет первым? **b)** Вася тянет билет последним?

Решение. В пункте (a) ясно, что Вася вытягивает случайно и равновозможно один из 10 билетов (вероятностное пространство: множество билетов, все исходы равновозможны). Поэтому вероятность вытянуть благоприятный билет 1/10.

Оказывается, в пункте (b) вероятность такая же. Чтобы ввести вероятностное пространство, занумеруем студентов числами от 1 до 10 в порядке очереди. Билеты также занумеруем числами от 1 до 10, номер 10 присвоим тому билету, который Вася выучил.

Тогда исходами будут перестановки чисел от 1 до 10. Все исходы равновозможны: процесс последовательного выбора билетов как раз и представляется в виде дерева последовательного случайного выбора.

Событие, вероятность которого нас интересует, состоит в том, что на 10-м месте стоит билет номер 10 (Вася идёт последним и знает только билет №10.)

Всего исходов 10!, а благоприятных — 9!. Искомая вероятность равна $9!/10! = 1/10$. □

Наметим другое решение задачи: все события «билет №10 стоит на месте i» имеют равные вероятности из симметрии. В пункте (a) мы уже нашли вероятность события «билет №10 стоит на месте 1». Поэтому в пункте (b) ответ такой же.

Конечно, наблюдение о симметрии задачи нужно как-то обосновать и сделать это зачастую не проще, чем решить задачу другим способом. Приведём поучительный пример.

Задача 11.11. В самолёт по очереди заходят 100 пассажиров. Первый садится на случайное место. Каждый следующий садится на своё место, если оно свободно, и на случайное место, если его место занято. Какова вероятность того, что последний пассажир сядет на своё место?

Решение. Начнём с простого, короткого, но неформального решения задачи.

Какие места могут быть свободными перед посадкой последнего пассажира? Это либо его место, либо место первого пассажира. Ведь по условию каждый пассажир кроме первого занимает своё место, если оно свободно.

Получается, что все исходы принадлежат ровно одному из двух событий: «последний пассажир сел на своё место» и «последний пассажир сел на место первого пассажира». Значит, в сумме вероятности этих событий дают 1.

Кроме того, вероятности этих событий одинаковы: если первый и последний пассажиры обменяются билетами, это не изменит рассадку, так как действия первого и последнего пассажиров не зависят от номера билета. Но события при этом переставляются: исходы, которые были в первом событии, попадают теперь во второе, и наоборот.

Итак, вероятности двух указанных событий равны и в сумме дают 1. Поэтому каждая из них равна 1/2.

Обратите внимание, что это неформальное решение нарушает наше главное правило: начинать решение «роман-задачи» с явной формулировки вероятностной модели и событий, о которых идёт речь в задаче. К тому же, это решение подозрительно напоминает фольклорное рассуждение: «Какова вероятность встретить на улице

динозавра? Она равна $1/2$ — либо встретишь, либо нет». Ошибки в неформальных решениях вероятностных задач — обычное дело, в социальных сетях можно легко найти примеры бурных обсуждений таких неправильных решений.

Тем не менее, в отличие от динозавров, решение задачи про посадку в самолёт можно сделать корректным. Ниже приводится такое корректное решение. Заметим, что задачу можно решить и многими другими способами (придумайте парочку), аккуратная запись некоторых существенно короче. Смысл этого упражнения в том, чтобы показать, как можно превращать неформальные рассуждения в строгие.

Начнём с определения исходов. Занумеруем пассажиров в порядке очереди числами от 1 до 100. Исходом будет рассадка пассажиров по местам, то есть некоторая перестановка чисел от 1 до 100. Вероятности исходов неодинаковы. Их можно определить процессом последовательного случайного выбора: первый пассажир выбирает позицию в перестановке согласно равномерному распределению; i-й по очереди пассажир выбирает либо своё место, если оно свободно, с вероятностью 1 (вероятности остальных вариантов при этом равны 0), в противном случае он выбирает одно из свободных мест согласно равномерному распределению на них.

Чтобы определить вероятность исхода, нужно перемножить вероятности для всех выборов.

Полученное распределение не равномерное. Например, любая перестановка, в которой первый пассажир сидит на своём месте, а какой-то пассажир — не на своём, имеет вероятность 0.

Заметим также, что вероятности перестановок-исходов зависят от раздачи билетов, то есть соответствия между пассажирами и местами. Обозначим через π это соответствие: у первого пассажира билет на место $\pi(1)$, у второго — на место $\pi(2)$ и т.д.

Мы фактически определили не одно распределение, а 100! распределений на одном и том же вероятностном пространстве. Одно такое распределение получается из другого перенумерацией мест, то есть перестановкой исходов. Кажется расточительным использовать 100! в сущности одинаковых распределений, но это помогает в объяснении трюка с симметрией.

Обозначим через G_π событие «последний пассажир сел на своё место», а через B_π — его дополнение, то есть «последний пассажир сел на место первого». Как мы уже говорили, каждый исход попадает в одно из этих событий, поэтому

$$\Pr_\pi[G_\pi] + \Pr_\pi[B_\pi] = 1.$$

Индекс π указывает на распределение, задаваемое раздачей билетов π.

Нетрудно видеть, что вероятности интересующих нас событий не зависят от раздачи билетов, то есть $\Pr_\pi[G_\pi] = \Pr_\sigma[G_\sigma]$ для любых π, σ: при перенумерации мест перестановки, в которых последний сидит на своём месте, переходят в точности в перестановки, в которых последний сидит на своём месте.

Рассмотрим две раздачи билетов

$$\pi = (1,\ 2\ , \ldots, 99,\ 100), \quad \sigma = (100,\ 2\ , \ldots, 99,\ 1)$$

(первый и последний обменялись билетами).

Для любой рассадки α выполняется равенство

$$\Pr_{\pi}[\alpha] = \Pr_{\sigma}[\alpha].$$

Действительно, при вычислении вероятности исхода α все числа на пути из корня дерева случайного выбора в лист α одинаковы в обоих случаях. Первое равно $1/100$ (так как первый пассажир выбирает одно из 100 мест согласно равномерному распределению). Последнее равно 1 (так как у последнего нет выбора). Все промежуточные вероятности выборов равны, так как π и σ различаются только местами первого и последнего и это различие не меняет количества возможных выборов для i-го пассажира.

Однако интересующие нас события переставляются: $\alpha \in G_\pi$ (то есть $a_{100} = 100$) тогда и только тогда, когда $\alpha \in B_\sigma$ (ещё раз напомним, что последний пассажир садится либо на своё место, либо на место первого). Поэтому

$$\Pr_{\pi}[G_\pi] = \Pr_{\sigma}[B_\sigma].$$

Поскольку $\Pr_\sigma[G_\sigma] = \Pr_\pi[G_\pi]$, то получаем $\Pr_\sigma[G_\sigma] = \Pr_\sigma[B_\sigma]$, то есть $\Pr_\sigma[G_\sigma] = 1/2$. □

11.2 Вероятность объединения событий

События по определению являются множествами, поэтому для них определены все теоретико-множественные операции. В этом разделе мы рассмотрим объединения событий.

Лемма 11.1. *Если $A, B \subseteq U$, то $\Pr[A \cup B] \leqslant \Pr[A] + \Pr[B]$. Если кроме того $A \cap B = \varnothing$, то $Pr[A \cup B] = \Pr[A] + \Pr[B]$.*

Доказательство. Действительно,

$$\Pr[A \cup B] = \sum_{x \in A \cup B} \Pr[x] \leqslant \sum_{x \in A} \Pr[x] + \sum_{x \in B} \Pr[x] = \Pr[A] + \Pr[B].$$

Если при этом множества A и B не пересекаются, то неравенство в этой цепочке обращается в равенство. □

События A и B, которые не могут произойти одновременно, то есть для которых $A \cap B = \varnothing$, называются *несовместными*. Таким образом вероятность объединения несовместных событий равна сумме их вероятностей.

Заметим, что если речь идёт о модели с равновозможными исходами, то вычисления и преобразования вероятности по существу мало отличаются от обычной комбинаторики. Нужно точно так же подсчитать нужное количество исходов, только в конце ещё разделить его на количество всех исходов.

В частности, как следствие принципа включений и исключений для множеств мы можем сразу получить принцип включений и исключений для вероятностей в равновозможной модели.

Следствие 11.2. *В равновозможной модели для произвольных множеств $A_1, \ldots, A_n \subseteq U$ верно*

$$\Pr[A_1 \cup A_2 \cup \cdots \cup A_n] = \sum_i \Pr[A_i] - \sum_{i<j} \Pr[A_i \cap A_j] + \cdots = \sum_{\varnothing \neq I \subset \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left[\bigcap_{i \in I} A_i\right]. \tag{11.1}$$

Доказательство. Действительно, по принципу включений и исключений для мощностей, аналогичное равенство верно для мощностей множеств. Чтобы получить равенство (11.1) для вероятностей, достаточно поделить обе части равенства на $|U|$. □

Оказывается, что на самом деле принцип включений и исключений верен не только для равновозможной модели, но и для любой другой.

Лемма 11.3. *Для всякой вероятностной модели и для произвольных множеств $A_1, \ldots, A_n \subseteq U$ верно*

$$\Pr[A_1 \cup A_2 \cup \cdots \cup A_n] = \sum_i \Pr[A_i] - \sum_{i<j} \Pr[A_i \cap A_j] + \cdots = \sum_{\varnothing \neq I \subset \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left[\bigcap_{i \in I} A_i\right]. \tag{11.2}$$

Доказательство. Здесь мы уже не можем просто сослаться на принцип включений и исключений для множеств. Однако, мы можем по существу повторить старое доказательство.

Для начала разберёмся со случаем двух множеств:

$$\begin{aligned}
\Pr[A_1 \cup A_2] &= \Pr\left[(A_1 \setminus A_2) \cup (A_1 \cap A_2) \cup (A_2 \setminus A_1)\right] \\
&= \Pr[A_1 \setminus A_2] + \Pr[A_1 \cap A_2] + \Pr[A_2 \setminus A_1] = \\
&= \Pr[A_1 \setminus A_2] + 2\Pr[A_1 \cap A_2] + \Pr[A_2 \setminus A_1] - \Pr[A_1 \cap A_2] = \\
&= \Pr[A_1] + \Pr[A_2] - \Pr[A_1 \cap A_2],
\end{aligned}$$

где первое и последнее равенство верны поскольку мы имеем дело с объединениями несовместных событий.

Теперь мы можем повторить индуктивное доказательство принципа включений и исключений для множеств. База нами уже доказана.

Для доказательства шага индукции заметим, что

$$\begin{aligned}
\Pr\left[(A_1 \cup \cdots \cup A_{n-1}) \cup A_n\right] &= \Pr\left[A_1 \cup \cdots \cup A_{n-1}\right] + \Pr[A_n] - \Pr\left[A_n \cap (A_1 \cup \cdots \cup A_n)\right] = \\
&= \Pr\left[A_1 \cup \cdots \cup A_{n-1}\right] + \Pr[A_n] - \Pr\left[(A_n \cap A_1) \cup \cdots \cup (A_n \cap A_{n-1})\right].
\end{aligned}$$

Здесь мы воспользовались принципом включений и исключений для двух множеств. Осталось для каждого из объединений $A_1 \cup \cdots \cup A_{n-1}$, $(A_n \cap A_1) \cup \cdots \cup (A_n \cap A_{n-1})$ воспользоваться предположением индукции. □

В комбинаторике одним из стандартных приложений формулы включений и исключений является задача о количестве перестановок n различных объектов так, чтобы ни один не остался на своём месте. В терминах теории вероятности мы можем оценить долю таких перестановок.

Лемма 11.4 (Задача о беспорядках)**.** *Рассмотрим случайные перестановки n различных объектов, то есть рассмотрим вероятностное пространство всех перестановок n заданных объектов, причём все перестановки равновозможны. Пусть A_n — событие, означающее, что все объекты после перестановки оказались не на своих изначальных местах. Тогда $\lim_{n\to\infty} \Pr[A_n] = 1/e$.*

Доказательство. Зафиксируем n и обозначим через B_i, где $i = 1, \ldots, n$, событие «объект с номером i остался на месте». Тогда $B_1 \cup \ldots \cup B_n$ означает, что хотя бы один из элементов остался на месте. Дополнительное событие к этому – как раз событие A_n.

Применим формулу включений и исключений к событию $B_1 \cup \ldots \cup B_n$. Для этого нам нужно будет посчитать вероятности событий $\bigcap_{i\in I} B_i$ для всевозможных $I \subseteq [n]$. Но это сделать несложно. Подходящие перестановки – это в точности перестановки, оставляющие на месте элементы из I, и переставляющие остальные элементы произвольным образом. Таких перестановок $(n-|I|)!$. Таким образом, для всякого I

$$\Pr\left[\bigcap_{i\in I} B_i\right] = \frac{(n-|I|)!}{n!}.$$

Множеств I размера k всего $\binom{n}{k}$, так что по формуле включений и исключений мы получаем

$$\begin{aligned}\Pr\left[B_1 \cup \ldots \cup B_n\right] &= \binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} + \ldots + (-1)^{n+1}\binom{n}{n}\frac{1}{n!} \\ &= \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n+1}\frac{1}{n!}.\end{aligned}$$

Тогда для вероятности события A_n мы получаем формулу

$$\Pr[A_n] = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n\frac{1}{n!}.$$

Теперь мы сошлёмся на факт из математического анализа, а именно, что эта сумма совпадает с началом ряда Тейлора для функции e^x в точке $x = -1$. Более точно, воспользуемся тем, что $e^{-1} = \sum_{k=0}^{\infty}(-1)^k/k!$. Отсюда

$$\lim_{n\to\infty} \Pr[A_n] = 1/e.$$

□

11.3 Вероятностный метод

Мы уже готовы показать один из мощных методов, часто применяемых в комбинаторике и теоретической Computer Science, а именно, вероятностный метод. Этот метод позволяет доказывать существование объектов с заданными свойствами. Для этого выбирается случайный объект из некоторого семейства и показывается, что он удовлетворяет нужным свойствам с положительной вероятностью. Мы воспользуемся этим методом для нижней оценки чисел Рамсея. Напомним, что числом Рамсея $R(n,k)$ называется минимальное число вершин, что всякий граф на этих вершинах содержит клику размера n или независимое множество размера k. Мы видели, что числа Рамсея не слишком большие, а именно $R(n,k) \leqslant \binom{n+k-2}{k-1}$. Теперь мы покажем, что эти числа и не слишком маленькие.

Теорема 11.5. *Для всякого $k \geqslant 3$ верно $R(k,k) > 2^{(k-1)/2}$.*

Доказательство. Рассмотрим $n = 2^{(k-1)/2}$ вершин и рассмотрим на них случайный граф. То есть в качестве вероятностного пространства U мы рассматриваем все графы на этих вершинах, и мы приписываем каждому из них одинаковую вероятность. Можно сразу заметить, что всего таких графов $2^{\binom{n}{2}}$. Действительно, для каждой пары вершин есть два выбора: либо добавить эту пару в рёбра графа, либо нет. Всего пар вершин $\binom{n}{2}$, так что графов получается $2^{\binom{n}{2}}$.

Оценим вероятность того, что случайный граф содержит клику или независимое множество размера k. Обозначим это событие через A. Наша цель – показать, что эта вероятность меньше 1. Тогда мы получим, что существует граф без клики и независимого множества размера k. Чтобы оценить эту вероятность, разобьем событие в объединение нескольких событий. Для этого для всякого подмножества $W \subseteq V$ множества вершин, такого что $|W| = k$ рассмотрим событие A_W, состоящее в том, что в случайном графе W – клика или независимое множество. Нетрудно видеть, что

$$A = \bigcup_{W \subseteq V, |W|=k} A_W,$$

а значит

$$\Pr[A] \leqslant \sum_{W \subseteq V, |W|=k} \Pr[A_W].$$

Теперь оценим вероятность отдельного события A_W. Посчитаем количество графов, попадающих в это событие. Ребра между вершинами в W в таком графе должны либо все присутствовать, либо все отсутствовать. Ребра, хотя бы один конец которых лежит вне W, могут быть произвольными. Количество ребер, у которых хотя бы один конец лежит вне W, есть $\binom{n}{2} - \binom{k}{2}$ (все ребра минус ребра в W). Таким образом, количество таких графов есть $2 \cdot 2^{\binom{n}{2}-\binom{k}{2}}$, где первая двойка отвечает за выбор ребер внутри W, а второй множитель – за выбор остальных ребер. Тогда получается, что

$$\Pr[A_W] = \frac{2^{\binom{n}{2}-\binom{k}{2}+1}}{2^{\binom{n}{2}}} = 2^{-\binom{k}{2}+1}.$$

Таким образом, при $k \geqslant 3$

$$\Pr[A] \leqslant \sum_{W \subseteq V, |W|=k} 2^{-\binom{k}{2}+1} = \binom{n}{k} 2^{-\binom{k}{2}+1} \leqslant \frac{n^k}{2 \times 3} 2^{-\binom{k}{2}+1} =$$
$$= 2^{k(k-1)/2-\binom{k}{2}+1}/6 = 1/3.$$

Следовательно, вероятность дополнения события A положительна, а значит существует граф на N вершинах без клик и независимых множеств размера k. □

Заметим, что наше доказательство неконструктивно: мы не строим граф, в котором нет клики и независимого множества, мы только доказываем, что он существует.

11.4 Условные вероятности

Помимо вероятностей тех или иных событий бывает нужным говорить и о вероятностях одних событий при условии других. Неформально говоря, мы хотим определить вероятность выполнения события A в том случае, когда событие B выполняется. В терминах вероятностного пространства определение этого понятия довольно естественное: нужно сузить вероятностное пространство на множество B. Так, для равновозможной модели мы получаем, что вероятность A при условии B есть просто $|A \cap B|/|B|$, то есть число благоприятных исходов поделенное на число всех исходов (после сужения всего вероятностного пространства до B). В случае произвольного вероятностного пространства нужно учесть веса исходов, то есть нужно сложить вероятности исходов в $A \cap B$ и поделить на сумму вероятностей исходов в B.

Таким образом, мы приходим к формальному определению. *Условной вероятностью события A при условии B* называется число

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Заметим, что условная вероятность имеет смысл, только если $\Pr[B] > 0$. Иначе знаменатель обращается в ноль.

Определение условной вероятности можно переписать следующим образом:

$$\Pr[A \cap B] = \Pr[B] \cdot \Pr[A|B].$$

Другими словами, чтобы найти вероятность пересечения событий A и B достаточно найти вероятность события B и условную вероятность события A при условии события B.

Задача 11.12. В классе 50% мальчиков; среди мальчиков 60% любят мороженое. Какова доля мальчиков, любящих мороженое, среди учеников класса? Как переформулировать этот вопрос в терминах вероятностей?

Задача 11.13. Приведите примеры, в которых условная вероятность $\Pr[A \mid B]$ больше вероятности $\Pr[A]$, меньше её, а также равна ей.

Понятие условной вероятности не всегда легко соотнести с интуицией, основанной на обыденной жизни. Тут особенно важно следовать сформулированному нами правилу, явно обозначать вероятностное пространство и распределение на нём, после чего пользоваться формальным определением условной вероятности. Приведем несколько примеров.

Задача 11.14. Есть три внешне одинаковых мешочка. В одном лежит две золотых монеты, во втором — одна золотая и одна серебряная монета, в третьем — две серебряные. Случайно и равновозможно выбирается один из мешочков, затем из него случайно и равновозможно достают монету.

Какова вероятность того, что выбран мешок с золотыми монетами при условии, что выбранная монета золотая?

Поначалу может показаться, что вероятность очевидно равна $1/2$: золотая монета есть в двух мешках, так что мы выбрали один из них, и при этом мешки равноправны. Однако, это не так. Давайте аккуратно разберемся почему.

Решение. Исходами в данном случае являются монеты и все шесть исходов равновозможны. (В данном случае опять имеет место ситуация последовательного случайного выбора и на каждом шаге выбора вероятности одинаковые. Нарисуйте дерево для этого выбора.)

Событие $A =$«выбранная монета золотая» состоит из 3 исходов. Событие $B =$ «выбран мешок с золотыми монетами» состоит из двух исходов, причём оба они содержатся в A (напомним ещё раз, что события — это множества, в данном случае — подмножества монет). Значит, $A \cap B = B$ и поэтому

$$\Pr[B \mid A] = \frac{\Pr[A \cap B]}{\Pr[A]} = \frac{2}{3}\,.$$

□

Задача 11.15. Есть девять коробок и один шарик. Шарик помещается в одну из коробок по следующему правилу. Сначала случайно и равновозможно выбирается коробка. Затем с вероятностью $1/2$ в неё кладётся шарик, а с вероятностью $1/2$ — нет. Найдите вероятность того, что в последней коробке шарик есть при условии, что в остальных коробках его нет.

Решение. Перенумеруем коробки числами от 1 до 9 и будем считать, что последняя коробка имеет номер 9. Какое в этой задаче вероятностное пространство? Исход — это пара (i, j), где i — номер выбранной коробки, а j равно 1 или 0 (пусть 1 означает, что шарик положили в коробку). Все исходы равновозможны, так как опять имеем дело с ситуацией последовательного случайного выбора и на каждом шаге выбора вероятности одинаковые.

Событие-условие «в коробках с номерами от 1 до 8 шарика нет» состоит из 10 исходов:

$$(i, 0),\ 1 \leqslant i \leqslant 8, \quad (9, 0),\ (9, 1)$$

(первые 8 из этих исходов означают, что выбрана коробка i и в неё не положили шарик; последние два означают, что выбрана коробка 9: тогда шарика в остальных коробках заведомо нет). Интересующее нас событие состоит из одного исхода $(9, 1)$ (выбрана коробка 9 и в неё положили шарик).

Это событие содержится в событии-условии, поэтому искомая условная вероятность равна $1/10$. □

Из формального определения условной вероятности можно получить на первый взгляд неожиданные утверждения.

Лемма 11.6 (формула Байеса)**.** *Если вероятность событий A и B положительна, то*

$$\Pr[A|B] = \Pr[A] \cdot \frac{\Pr[B|A]}{\Pr[B]}.$$

У этой леммы есть вполне конкретный практический смысл, который мы проиллюстрируем на условном примере. Рассмотрим некоторую болезнь, и предположим, что для её обнаружения мы можем делать недорогой анализ, который при этом с заметной вероятностью выдаёт неправильный результат, а также есть дорогостоящее исследование, которое уже наверняка сообщает, болен ли человек. Мы хотим обнаруживать болезнь следующим образом: сначала человек сдаёт недорогой анализ, а если он дал положительный результат, то проверяем человека с помощью дорогостоящего исследования. При таком подходе возможно, что больной человек не будет обнаружен нашим анализом. Можем ли мы на основании статистических данных понять, как часто это происходит? На первый взгляд кажется, что этого сделать нельзя, ведь мы же не обнаруживаем этих больных. Но оказывается, что на самом деле это можно сделать при помощи формулы Байеса. Пусть B — событие «быть больным», а A — событие «получить положительный результат на анализе». Собрав статистику, мы можем узнать $\Pr[A]$ — вероятность для человека получить положительный анализ.[1] Можно также оценить $\Pr[B]$ — долю больных рассматриваемой болезнью. Наконец, можно оценить и $\Pr[B|A]$ — вероятность того, что человек болеет при условии, что он получил положительный результат анализа. (Опять же, мы можем собрать статистику.) Теперь по формуле Байеса мы можем посчитать $\Pr[A|B]$ — вероятность того, что больной человек будет обнаружен нашим анализом. А это как раз то, что мы хотели найти.

Доказательство леммы 11.6. Доказательство формулы Байеса почти очевидно. Достаточно просто записать вероятность события $A \cap B$ через условные вероятности двумя способами:

$$\Pr[A \cap B] = \Pr[B] \cdot \Pr[A|B] = \Pr[A] \cdot \Pr[B|A].$$

Теперь второе равенство сразу даёт формулу Байеса. □

[1] Строго говоря, из статистических данных можно лишь получать более-менее достоверные оценки вероятностей. В нашем условном примере мы пренебрегаем этой трудной проблемой.

Понятие условной вероятности позволяет нам также говорить о независимых событиях. Неформально, событие A не зависит от события B, если вероятность события A при условии события B такая же, как и вероятность A при условии не выполнения события B. Формально удобнее выбрать другое определение: событие A *не зависит от события* B, если

$$\Pr[A] = \Pr[A|B].$$

Чтобы не возникало никаких тонкостей с нулевыми вероятностями полезно условиться, что вероятности событий A и B ненулевые.

Из определения условной вероятности мы сразу получаем эквивалентное определение независимости событий. Событие A не зависит от события B, если

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].$$

Из этой формы определения видно замечательное свойство независимости событий: она симметрична. То есть, событие A не зависит от события B тогда и только тогда, когда событие B не зависит от события A.

Несложными алгебраическими преобразованиями проверяются и другие свойства независимых событий. Приведём два наиболее важных, оставляя доказательства в качестве упражнения для читателя.

Задача 11.16. Пусть события A и B независимы, и вероятность события $\bar{B}$ положительна. Докажите, что события A и $\bar{B}$ независимы. Здесь и далее $\bar{B} = U \setminus B$ обозначает событие, дополнительное к B.

Задача 11.17. Докажите, что события A и B, для которых $0 < \Pr[B] < 1$, независимы тогда и только тогда, когда $\Pr[A \mid B] = \Pr[A \mid \bar{B}]$.

Приведём несколько примеров проверки независимости событий. Первый очень простой, но важный.

Пример 11.18 (продолжение примера 11.3)**.** В примере 11.3 мы проверили, что для равномерного распределения на двоичных словах длины n вероятности событий «на i-й позиции стоит α» равны $1/2$ (здесь $1 \leqslant i \leqslant n$ и $\alpha \in \{0, 1\}$).

Теперь проверим, что эти события при $i \neq j$ независимы. Действительно, есть 2^{n-2} слов, принимающих заданные значения α и β в позициях i и j. Поэтому вероятность пересечения любых двух таких событий равна $2^{n-2}/2^n = 1/4 = 1/2 \cdot 1/2$.

Точно так же проверяется и независимость более сложных событий вида «на позициях $i_1, \ldots, i_s$ стоят символы $\alpha_1, \ldots, \alpha_s$», если множества позиций не пересекаются.

Менее очевидные примеры связаны с событиями на случайных перестановках.

Пример 11.19. Выбирается случайно и равновозможно перестановка $x_1, x_2, \ldots, x_{49}$ чисел от 1 до 49.

(а) Независимы ли события «x_{24} больше всех последующих» и «x_{25} больше всех последующих»?

Как мы уже видели выше в примере 11.10, другой способ получить случайно и равновозможно перестановку состоит в том, чтобы выбирать каждый x_i случайно и равновозможно из оставшихся к этому моменту вариантов. Поэтому вероятности событий «x_{24} больше всех последующих» и «x_{25} больше всех последующих» равны $1/26$ и $1/25$ соответственно (в первом случае нужно выбрать на 24-м шаге максимальное из оставшихся 26 чисел, во втором аналогично).

Вероятность пересечения этих событий совпадает с вероятностью выбрать на 24-м шаге максимальное из 26 имеющихся чисел, а на втором — максимальное из 25 чисел, оставшихся к этому шагу. Вероятность такого события равна

$$\frac{1}{26 \cdot 25}$$

(исходы — упорядоченные пары из 26 чисел, благоприятный исход один).

Поскольку вероятность пересечения событий равна произведению вероятностей событий, эти события независимы.

(б) Независимы ли события «$x_{24} > x_{25}$» и «$x_{25} > x_{26}$»?

Вероятности этих событий равны $1/2$. Перестановок, для которых $x_{24} > x_{25}$, ровно столько, сколько перестановок, для которых $x_{24} < x_{25}$: меняя местами 24-е и 25-е числа, мы из перестановки, для которой $x_{24} > x_{25}$, получаем перестановку, для которой $x_{24} < x_{25}$.

Аналогично рассуждаем и для второго события.

Теперь найдём вероятность пересечения рассматриваемых событий. Это событие «$x_{24} > x_{25} > x_{26}$». Рассуждаем аналогично примеру 11.6. Разобьём перестановки на группы из шести перестановок, в каждой группе одна и та же тройка чисел занимает места с 24-го по 26-е в каком-то порядке, а на остальных местах числа расставлены одинаково.

В каждой шестёрке есть ровно одна перестановка из события «$x_{24} > x_{25} > x_{26}$». Поэтому вероятность этого события $1/6$.

Так как

$$\frac{1}{2} \cdot \frac{1}{2} \neq \frac{1}{6}\,,$$

рассматриваемые события не являются независимыми.

Из этих вычислений заключаем, что вероятность события «$x_{25} > x_{26}$» при условии «$x_{24} > x_{25}$» равна $1/3 < 1/2$.

Это довольно удивительно с обыденной точки зрения. Представьте, что в лототроне изначально находятся 49 шаров, пронумерованных числами от 1 до 49. Шары выкатываются из лототрона по очереди. Предположим, что номер 24-го шара оказался больше номера 25-го. Каковы шансы, что номер 25-го шара будет больше номера 26-го? Мало кто из людей способен догадаться до точного значения $1/3$, большинство предположит, что эта вероятность больше.

Другим полезным утверждением об условных вероятностях является формула полной вероятности.

Лемма 11.7 (Формула полной вероятности)**.** *Пусть $B_1, \ldots, B_n$ — разбиение вероятностного пространства U, то есть $U = B_1 \cup \ldots \cup B_n$, где $B_i \cap B_j = \varnothing$ при $i \neq j$. Пусть также $\Pr[B_i] > 0$ для всякого i. Тогда для всякого события A*

$$\Pr[A] = \sum_{i=1}^{n} \Pr[A|B_i] \cdot \Pr[B_i].$$

Доказательство.

$$\Pr[A] = \sum_{i=1}^{n} \Pr[A \cap B_i] = \sum_{i=1}^{n} \Pr[A|B_i] \cdot \Pr[B_i],$$

где первое равенство получается по формуле сложения вероятностей непересекающихся событий, а второе равенство – по определению условной вероятности. □

Приведём примеры использования формулы полной вероятности.

Пример 11.20 (Продолжение примера 11.19)**.** Выбирается случайно и равновозможно перестановка $x_1, x_2, \ldots, x_{49}$ чисел от 1 до 49.

Какова вероятность события «$x_{25} > x_{26}$» при условии «$x_{24} < x_{25}$»?

Обозначим эту условную вероятность p. По формуле полной вероятности

$$\begin{aligned}\Pr[\text{«}x_{25} > x_{26}\text{»}] = {} & \Pr[\text{«}x_{25} > x_{26}\text{»} \mid \text{«}x_{24} < x_{25}\text{»}] \cdot \Pr[\text{«}x_{24} < x_{25}\text{»}] + \\ & + \Pr[\text{«}x_{25} > x_{26}\text{»} \mid \text{«}x_{24} > x_{25}\text{»}] \cdot \Pr[\text{«}x_{24} > x_{25}\text{»}].\end{aligned}$$

В примере 11.19 мы уже посчитали все вероятности, кроме искомой. Из формулы полной вероятности получаем

$$\frac{1}{2} = p \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2},$$

т.е. $p = 2/3$. Конечно, посчитать искомую вероятность можно и другим способом, аналогично примеру 11.19.

В приложениях очень часто формула условной вероятности применяется вместе с формулой Байеса. Приведём условный пример. Для краткости мы пропускаем точные определения вероятностного пространства и вероятностного распределения для этого примера. Читателю рекомендуется восстановить их самостоятельно в качестве упражнения.

Пример 11.21. Редакционную колонку в некотором издании каждый раз пишет один из журналистов X и Y. Журналист X пишет колонку в два раза чаще журналиста Y. Известно, что X допускает фактические ошибки в 25% статей, а Y — в 50% статей.

(a) С какой вероятностью в случайной редакционной колонке обнаружится фактическая ошибка? Предполагаем, что выбор колумниста и наличие фактических ошибок в статье происходят случайно.

По условию задачи события «колонку писал журналист X» и «колонку писал журналист Y» образуют разбиение всего множества исходов, причём вероятность первого события $2/3$, а второго — $1/3$. Вероятность фактической ошибки первого журналиста $1/4$, а второго — $1/2$. Получаем по формуле полной вероятности

$$\begin{aligned}\Pr[\text{ошибка}] = \\ = \Pr[\text{ошибка} \mid \text{писал } X] \cdot \Pr[\text{писал } X] + \Pr[\text{ошибка} \mid \text{писал } Y] \cdot \Pr[\text{писал } Y] = \\ = \frac{1}{4} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}\,.\end{aligned}$$

(б) В редакционной колонке обнаружена фактическая ошибка. С какой вероятностью её написал журналист X?

Формула Байеса даёт

$$\Pr[\text{писал } X \mid \text{ошибка}] = \Pr[\text{писал } X] \cdot \frac{\Pr[\text{ошибка} \mid \text{писал } X]}{\Pr[\text{ошибка}]} = \frac{2}{3} \cdot \frac{1/4}{1/3} = \frac{1}{2}\,.$$

Приведём ещё пример использования формулы полной вероятности в теории графов.

Пример 11.22. Пусть G — простой неориентированный граф с n вершинами $v_1, \ldots, v_n$, степени которых равны некоторому числу d (такой граф называется регулярным). Как мы уже доказывали, у него $nd/2$ рёбер. Рассмотрим два вероятностных распределения на его рёбрах. Первое — равномерное, то есть каждое из $nd/2$ рёбер выбирается с одинаковой вероятностью. А второе такое: сначала случайным образом выбирается вершина (каждая с одинаковой вероятностью), а затем случайно с равными вероятностями выбирается ребро, соседнее с этой вершиной. Если немного подумать, то интуитивно понятно, что эти два распределения одинаковые — каждое ребро выбирается в обоих распределениях с одной и той же вероятностью. Но как это аккуратно доказать?

Рассмотрим произвольное ребро e и событие A, означающее, что выбрано это ребро. Подсчитаем $\Pr[A]$ в каждом из распределений. В первом случае это несложно — у нас задано равновероятное распределение на $nd/2$ рёбрах, так что вероятность равна $2/nd$. Чтобы посчитать вероятность во втором случае для всякого $i \in \{1, \ldots, n\}$ рассмотрим событие B_i, состоящее в том, что была выбрана вершина v_i. Эти события образуют разбиение вероятностного пространства. Так что по формуле полной вероятности получаем

$$\Pr[A] = \sum_{i=1}^{n} \Pr[A|B_i] \cdot \Pr[B_i].$$

На вершинах у нас задано равновероятное распределение, так что для каждого i $\Pr[B_i] = 1/n$. Теперь подсчитаем условную вероятность $\Pr[A|B_i]$. Если вершина v_i не является концом ребра e, то ребро никак не может быть выбрано, так что условная вероятность в этом случае равна нулю. Если же v_i является концом ребра e, то вероятность, что мы выберем его равна $1/d$: мы равновероятно выбираем одно из d

рёбер с концом в v_i. У ребра два конца, так что все слагаемые кроме двух равны 0, а каждое из двух оставшихся равны $(1/d)\cdot(1/n)$. Таким образом, вероятность события A в случае второго распределения также равна $2/dn$, а значит оба распределения совпадают.

Следующий пример использования формулы полной вероятности показывает как упрощать вероятностное пространство, используя соображения симметрии.

Пример 11.23. Из n-элементного множества выбираются случайно, равновозможно и независимо два k-элементных множества X и Y. Какова вероятность события «$X \cap Y = \varnothing$»?

Условие означает, что вероятностное пространство — пары (X, Y) k-элементных подмножеств n-элементного множества, вероятности всех исходов одинаковы.

Событие «$X \cap Y = \varnothing$» не изменяется при переобозначении элементов множества. Поэтому вероятности условных событий «$Y \cap X = \varnothing \mid X = A$» и «$Y \cap X = \varnothing \mid X = B$» одинаковы для любых A и B. Из формулы полной вероятности получаем, что для любого множества A

$$\Pr[X \cap Y = \varnothing] = \Pr[Y \cap X = \varnothing \mid X = A].$$

Пусть $A = \{1, \ldots, k\}$. Вероятность выбрать k-элементное множество Y, которое не содержит ни одного элемента из A, равна

$$\frac{\binom{n-k}{k}}{\binom{n}{k}}$$

(в числителе стоит количество k-элементных подмножеств в дополнении к A, а в знаменателе — количество k-элементных подмножеств в n-элементном множестве).

Преобразуем это выражение:

$$\frac{\binom{n-k}{k}}{\binom{n}{k}} = \left(1 - \frac{k}{n}\right) \cdot \left(1 - \frac{k}{n-1}\right) \cdot \ldots \cdot \left(1 - \frac{k}{n-k+1}\right) \leqslant \left(1 - \frac{k}{n}\right)^k.$$

При больших n и $k \approx c\sqrt{n}$ последнее выражение оценивается как e^{-c^2} (здесь мы опускаем подробности).

Получаем неожиданный с точки зрения интуиции факт: весьма малые случайные подмножества большого конечного множества почти заведомо пересекаются. Более того, мы получили количественную оценку этой «малости»: $k/\sqrt{n} \to \infty$ при $n \to \infty$. Аналогичными рассуждениями можно также проверить, что случайные множества размера $o(\sqrt{n})$ почти заведомо не пересекаются.

11.5 Случайная величина, математическое ожидание

Случайная величина — это числовая функция на вероятностном пространстве, то есть функция вида $f\colon U \to \mathbb{R}$. То есть, по сути, случайная величина — это обычная

числовая функция, но теперь на её аргументах задано вероятностное распределение. Таким образом, например, мы можем говорить о вероятности того, что случайная величина f равна какому-то конкретному значению a: это есть просто вероятность события $\{u \in U \mid f(u) = a\}$. Случайные величины представляют собой числовые характеристики вероятностных экспериментов и, на самом деле, мы с ними уже неоднократно сталкивались, просто не говорили об этом. Например, если мы бросаем кубик, то исходом эксперимента является выпадение той или иной грани, а случайной величиной — число написанное на грани (каждой грани соответствует своё число — это функция).

Важным параметром случайной величины является её математическое ожидание. Неформально, это число, которое мы будем получать в среднем, если будем повторять эксперимент много раз и каждый раз смотреть на значение случайной величины.

Более конкретно, пусть вероятностное пространство состоит из k исходов, случайная величина $f\colon U \to \mathbb{R}$ принимает на них значения $a_1, \ldots, a_k$ соответственно и вероятности исходов равны $p_1, \ldots, p_k$ соответственно. В частности, $\sum_{i=1}^{k} p_i = 1$. Предположим, что выбор случайного элемента из U повторяется n раз. Если n достаточно большое, то случайная величина f примет значение a_1 примерно $p_1 n$ раз, значение a_2 — примерно $p_2 n$ раз, и так далее, значение a_k — примерно $p_k n$ раз. (Этому утверждению можно придать строгий смысл. Простейший случай разобран в разделе 11.7.) Подсчитаем теперь примерное среднее арифметическое значений случайной величины f в этих экспериментах:

$$\frac{a_1 p_1 n + a_2 p_2 n + \ldots + a_k p_k n}{n} = \sum_{i=1}^{k} a_i p_i.$$

Этот неформальный рассказ приводит нас к следующему строгому (и очень важному) определению.

Математическим ожиданием случайной величины f, принимающей значения $a_1, \ldots, a_k$ с вероятностями $p_1, \ldots, p_k$ соответственно, называется величина

$$E[f] = \sum_{i=1}^{k} a_i p_i.$$

Например, случайная величина, равная числу, выпадающему на грани кубика, принимает значения $1, 2, 3, 4, 5, 6$ с вероятностями $1/6$. Её математическое ожидание равно

$$1 \cdot (1/6) + 2 \cdot (1/6) + 3 \cdot (1/6) + 4 \cdot (1/6) + 5 \cdot (1/6) + 6 \cdot (1/6) = 21/6 = 3.5.$$

То есть, при бросании кубика мы будем в среднем получать число 3.5 .

Математическое ожидание с одной стороны является осмысленной характеристикой случайной величины, а с другой обладает свойствами, делающими работу с математическими ожиданиями удобной.

Лемма 11.8 (линейность математического ожидания)**.** *Пусть $f\colon U \to \mathbb{R}$ и $g\colon U \to \mathbb{R}$ — две случайные величины на одном и том же вероятностном пространстве. Тогда*

$$E[f+g] = E[f] + E[g].$$

Доказательство. Его несложно получить непосредственно из определения математического ожидания (собственно, из чего же ещё?).

Пусть вероятностное пространство U состоит из исходов $u_1, \ldots, u_k$ с вероятностями $p_1, \ldots, p_k$ соответственно. Тогда

$$E[f+g] = \sum_{i=1}^{k} (f(u_i) + g(u_i))p_i = \sum_{i=1}^{k} (f(u_i))p_i + \sum_{i=1}^{k} (g(u_i))p_i = E[f] + E[g].$$

□

Линейность математического ожидания во многих случаях заметно упрощает вычисления.

Пример 11.24 (Задача о днях рождения)**.** Рассмотрим n случайных людей и посмотрим на количество совпадений дней рождения у них, то есть на количество пар людей, имеющих день рождения в один день. Каким в среднем будет это число?

Сформулируем вопрос точно. Вероятностное пространство: всюду определённая функция из n-элементного множества людей $\{x_1, \ldots, x_n\}$ в 365-элементное множество дней в году. Все исходы равновозможные.[2]

Обозначим случайную величину, равную количеству пар людей с совпадающими днями рождения, через f. Нам требуется посчитать математическое ожидание случайной величины f. Но при этом случайная величина довольно сложная, и подсчитывать математическое ожидание непосредственно из определения трудно.

Идея состоит в следующем: давайте разобьём сложную случайную величину f в сумму нескольких простых случайных величин. Тогда мы сможем подсчитать отдельно математические ожидания всех простых величин, а затем, пользуясь линейностью математического ожидания, просто сложить результаты.

Обозначим через g_{ij} случайную величину, равную 1, если у людей x_i и x_j дни рождения совпадают, и равную 0 в противном случае. Тогда можно заметить, что

$$f = \sum_{i<j} g_{ij}.$$

Подсчитаем математическое ожидание случайной величины g_{ij}. Нетрудно увидеть, что вероятность того, что у двух случайных людей дни рождения совпадают, равна $1/365$, так что с вероятностью $1/365$ случайная величина равна 1, и с вероятностью

[2] Мы выбрали такой вариант для простоты — в реальной жизни нужно учесть существование високосных лет и неравномерное распределение дней рождения в году. Да и независимость дней рождения в реальной жизни неочевидна — вспомним о близнецах.

$1-1/365$ равна 0. Так что $E[g_{ij}] = 1/365$ (для всякой пары i, j). Для математического ожидания f из линейности получаем

$$E[f] = E[\sum_{i<j} g_{ij}] = \sum_{i<j} E[g_{ij}] = \sum_{i<j} 1/365 = \frac{n(n-1)}{2 \cdot 365}.$$

Например, если число людей n больше 27, то $E[f] > 1$, то есть естественно ожидать, что будет больше одного совпадения дней рождений, что может показаться противоречащим интуиции (поэтому эту задачу иногда называют «парадоксом дней рождения»).

На самом деле, зная немного анализа, можно убедиться, что если распределение дней рождения по дням года считать неравномерным (что ближе к реальности), то математическое ожидание только вырастет.

Пример 11.25. Выбирается случайное множество двоичных строк длины n. (Все подмножества множества $\{0,1\}^n$ равновероятны.)

(а) Чему равно математическое ожидание суммарного числа единиц в строках этого подмножества?

Обозначим суммарное количество единиц S. Для каждой двоичной строки w обозначим через S_w случайную величину, которая равна $|w|$, т.е. количеству 1 в строке w, если эта строка попала в случайное множество, и 0 в противном случае. Тогда

$$S = \sum_w S_w.$$

Поскольку все подмножества равновероятны, вероятность для каждого попасть в случайное множество равна $1/2$ (одинаково количество тех подмножеств, которые содержат w, и тех, которые не содержат).

Поэтому математическое ожидание S_w равно

$$E[S_w] = \frac{1}{2} \cdot |w| + \frac{1}{2} \cdot 0 = \frac{|w|}{2}.$$

Значит, $E[S]$ равно общему количеству единиц в двоичных строках длины n, делённому на 2. В каждом разряде во всех строках вместе ровно 2^{n-1} единиц (нулей и единиц в каждом разряде поровну). Поэтому общее количество единиц равно $n2^{n-1}$, откуда получаем $E[S] = n2^{n-2}$.

(б) Тот же вопрос, но выбирается случайное подмножество, в котором ровно k строк.

Рассматриваем аналогичные случайные величины S и S_w как в п. (а). Но теперь у нас пространство состоит из подмножеств мощности k.

Какова вероятность события «строка w попала в случайное k-элементное подмножество строк X»? Всего k-элементных подмножеств $\binom{2^n}{k}$, а тех, которые содержат данную строку w, — $\binom{2^n-1}{k-1}$. Вероятность равна отношению этих двух чисел.

Получаем

$$E[S_w] = |w| \frac{\binom{2^n-1}{k-1}}{\binom{2^n}{k}} = \frac{|w|k}{2^n}\ .$$

Отсюда, пользуясь вычислениями из п. (а), находим ответ

$$E[S] = \sum_w E[S_w] = \frac{k}{2^n} \sum_w |w| = \frac{k}{2^n} \cdot n2^{n-1} = \frac{kn}{2}.$$

С помощью математического ожидания можно обобщить вероятностный метод.

Лемма 11.9 («среднее не больше максимума и не меньше минимума»)**.** *Пусть для какой-то случайной величины $f\colon U \to \mathbb{R}$ верно $E[f] = C$. Тогда существует такой исход $u \in U$, что $f(u) \geqslant C$. Аналогично, существует и такой исход $u \in U$, что $f(u) \leqslant C$.*

Доказательство. Докажем первое утверждение леммы, второе доказывается аналогично.

На самом деле, доказательство довольно простое. Предположим, что утверждение неверно, а значит для всякого $u \in U$ верно $f(u) < C$.

Тогда

$$E[f] = \sum_{u \in U} \Pr[u] f(u) < \sum_{u \in U} \Pr[u] C = C,$$

противоречие. □

Задача 11.26. Объясните, почему старая формулировка вероятностного метода является частным случаем новой формулировки.

Новая формулировка удобна в некоторых случаях. Разберём один такой пример.

Рассмотрим простой неориентированный граф $G = (V, E)$. *Разрезом* графа называется разбиение множества его вершин на два непересекающихся подмножества: $V = V_1 \cup V_2$, $V_1 \cap V_2 = \varnothing$. Мы говорим, что ребро попадает в разрез, если один его конец лежит в V_1, а другой в V_2. Размером разреза называется число рёбер, попадающих в разрез. Нас будут интересовать большие разрезы графа.

Теорема 11.10. *Всякий граф $G = (V, E)$ имеет разрез размера не меньше $|E|/2$.*

Доказательство. Рассмотрим случайный разрез графа G. Более точно, мы берём равномерное распределение на множестве всех разрезов. Разрез задается подмножеством $S \subseteq V$: такому подмножеству ставится в соответствие разрез $(S, V \setminus S)$. Всего подмножеств (а значит и разрезов) 2^n, так что вероятность каждого разреза есть $1/2^n$. Аналогично примеру 11.18 можно проверить, что для каждой пары вершин $x \neq y$ все четыре события «$x \in S, y \in S$», «$x \notin S, y \in S$», «$x \in S, y \notin S$», «$x \notin S, y \notin S$» имеют вероятность $1/4$.

Итак, рассмотрим случайный разрез и рассмотрим случайную величину f, равную размеру разреза. Посчитаем её математическое ожидание. Для этого, как и раньше, стоит разбить случайную величину в сумму более простых случайных величин. Для всякого $e \in E$ рассмотрим случайную величину f_e, равную 1, если ребро e входит в разрез, и равную 0 в противном случае. Тогда нетрудно видеть, что $f = \sum_{e \in E} f_e$, а значит

$$E[f] = \sum_{e \in E} E[f_e].$$

Однако, для случайной величины f_e математическое ожидание уже нетрудно посчитать. Действительно, для всякого фиксированного ребра e вероятность, что оно попадёт в разрез равна $1/2$. А значит, $E[f_e] = 1/2$ для всякого $e \in E$, откуда

$$E[f] = \sum_{e \in E} 1/2 = |E|/2.$$

Из этого следует, что есть конкретный разрез, содержащий не меньше $|E|/2$ рёбер. □

На самом деле, этот результат не сильно удивителен, его можно доказать и «руками».

Задача 11.27. Постройте алгоритм, работающий за полиномиальное время и строящий разрез размера не меньше $|E|/2$.

Оказывается, что если проводить вероятностное рассуждение аккуратнее, то можно получить чуть более сильную оценку на размер разреза.

Теорема 11.11. *Рассмотрим граф $G = (V, E)$, в котором количество вершин $|V| = 2n$ — чётно. Тогда в G существует разрез размера не меньше $\frac{|E|n}{2n-1}$.*

Доказательство. Как и в прошлый раз, всякий разрез можно задать множеством $S \subseteq V$. Рассмотрим равномерное распределение на множествах $S \subseteq V$, таких что $|S| = n$ (в этом отличие от прошлого рассуждения).

Случайные величины f и f_e определим также как и в прошлом доказательстве. Оценим вероятность того, что $f_e = 1$. Число благоприятных исходов равно $2\binom{2n-2}{n-1}$, где двойка отвечает за выбор конца ребра e, лежащего в S, а биномиальный коэффициент отвечает за выбор остальных элементов S. Число всех исходов равно $\binom{2n}{n}$, так что

$$E[f_e] = \Pr[f_e = 1] = \frac{2\binom{2n-2}{n-1}}{\binom{2n}{n}} = \frac{2 \cdot n \cdot n}{2n \cdot (2n-1)} = \frac{n}{(2n-1)}.$$

Тогда аналогично доказательству предыдущей теоремы получаем

$$E[f] = \sum_{e \in E} E[f_e] = \frac{|E|n}{2n-1},$$

а значит существует такой разрез, в котором не меньше $\frac{|E|n}{2n-1}$. □

Такой разрез тоже можно построить напрямую, но это уже заметно сложнее. А наше доказательство было в сущности не очень сложным (во всяком случае, технически).

Задача 11.28. Докажите, что если в графе $G = (V, E)$ число вершин $|V| = 2n + 1$ — нечётно, то в нём есть разрез размера не меньше $\frac{|E|(n+1)}{2n+1}$.

Математическое ожидание позволяет давать оценки вероятностей некоторых событий.

Лемма 11.12 (неравенство Маркова)**.** *Пусть f — случайная величина, принимающая только неотрицательные значения. Тогда для всякого $\alpha > 0$ верно*

$$\Pr[f \geqslant \alpha] \leqslant \frac{E[f]}{\alpha}.$$

То есть, вероятность того, что случайная величина f сильно больше своего математического ожидания, не слишком велика (заметим, что лемма становится содержательной, когда $\alpha > E[f]$.

Доказательство. Взглянем на нужное нам неравенство с другой стороны. Нам нужно доказать, что

$$E[f] \geqslant \alpha \cdot \Pr[f \geqslant \alpha].$$

Пусть случайная величина f принимает значения $a_1, \ldots, a_k$ с вероятностями $p_1, \ldots, p_k$. Запишем, чему равно её математическое ожидание по определению:

$$E[f] = a_1 p_1 + a_2 p_2 + \ldots + a_k p_k.$$

Посмотрим отдельно на те a_i, которые меньше α, и отдельно на те a_i, которые не меньше α. Если первые заменить на ноль, то сумма может только уменьшиться. Если вторые заменить на α, то сумма также может только уменьшиться. После таких замен, у нас остаётся сумма нескольких слагаемых, каждое из которых есть αp_i, где p_i — вероятность некоторого значения случайной величины, не меньшего α. Нетрудно видеть, что такая сумма как раз равна $\alpha \cdot \Pr[f \geqslant \alpha]$, и лемма доказана. □

Задача 11.29. Где в нашем доказательстве мы использовали неотрицательность случайной величины? Остается ли лемма верной, если убрать условие неотрицательности случайной величины?

Задача 11.30. В лотерее на выигрыши уходит 40% от стоимости проданных билетов. Каждый билет стоит 100 рублей. Докажите, что вероятность выиграть 5000 рублей (или больше) меньше 1%.

Пример 11.31. Приведём алгоритмический пример применения неравенства Маркова.

Некоторые алгоритмы, использующие случайные числа, работают так, что всегда выдают верный ответ, но время работы может зависеть от значения случайных чисел и при некотором невезении алгоритм может работать долго. В таких ситуациях, чтобы тем не менее сказать что-то о времени работы алгоритма, говорят о среднем времени работы алгоритма. Действительно, время работы в данном случае — случайная величина (зависящая от случайных чисел, используемых алгоритмом), и среднее время работы — это просто математическое ожидание этой случайной величины.

Предположим, что у нас есть такой алгоритм A, работающий, скажем, за среднее время $O(n^2)$, где n — размер входных данных. Для наших практических целей хотелось бы, чтобы алгоритм всегда (то есть независимо от случайных чисел) заканчивал свою работу за время $O(n^2)$. Чтобы добиться этого, мы готовы даже смириться с тем, что в $0,01\%$ случаев алгоритм будет выдавать неправильный ответ. Можем ли мы получить такой алгоритм?

Оказывается, можем. Обозначим среднее время работы алгоритма A через T, и рассмотрим следующий алгоритм: запускаем алгоритм A и ждём пока он сделает $10000 \cdot T$ шагов. Если алгоритм успел выдать ответ, прекрасно. Если нет, выдаём произвольный ответ. Идея в том, что алгоритм A с очень большой вероятностью закончит свою работу за $10000 \cdot T$ шагов. Действительно, обозначим через f (неотрицательную) случайную величину, равную времени работы алгоритма A. Тогда $E[f] = T$. По неравенству Маркова получаем

$$\Pr[f > 10000 \cdot T] \leqslant \frac{T}{10000 \cdot T} = 1/10000.$$

Заметим, что время работы нового алгоритма действительно есть $O(n^2)$ (по сравнению со старым алгоритмом оно просто умножилось на константу), а ошибка может произойти только если старый алгоритм работал дольше $10000 \cdot T$ шагов. По нашей оценке это происходит с вероятностью не больше 0.01%.

11.6 Частота орлов при подбрасывании монеты и биномиальные коэффициенты

Если подбрасывать «честную» монету много раз, то разумно ожидать, что количество выпавших орлов будет примерно равно половине от числа подбрасываний. Именно это интуитивное наблюдение было положено в основу понятия «вероятность». Но что значит «примерно равно»? Оказывается, этому интуитивно ожидаемому результату можно придать точные количественные характеристики.

Во-первых, уточним, что разные подбрасывания «честной» монеты независимы. Если кто-то будет подбрасывать монету и после выпадения четвёртого орла подряд переворачивать монету, такие подбрасывания вряд ли стоит считать «честными».

Поэтому мы будем считать все возможные результаты n подбрасываний равновозможными. Будем записывать результаты, указывая 1, если выпал орёл, и 0, если выпала решка.

Как легко видеть, количество вариантов n подбрасываний, в которых выпало k орлов, равно количеству двоичных слов длины n, в которых ровно k единиц (и $n-k$ нулей), т.е. равно биномиальному коэффициенту

$$\binom{n}{k}.$$

Если же нас интересуют события вида «выпало не меньше k орлов» или «количество орлов не меньше k_1 и не больше k_2», то их вероятности — суммы биномиальных коэффициентов.

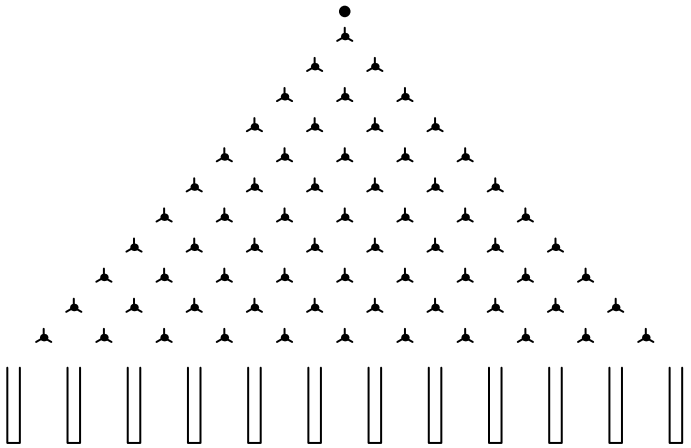

Рис. 11.3: Принципиальная схема доски Гальтона

Итак, нам нужно оценивать величины биномиальных коэффициентов. Это можно делать по-разному. Скажем, можно находить нужные величины экспериментально. Для этого можно использовать прибор, называемый доской Гальтона. Его схема изображена на рисунке 11.3.

Если набросать через такую решётку много шариков (в оригинальном исполнении это были бобы), то бункеры будут заполнены как раз пропорционально величине соответствующих биномиальных коэффициентов.

Конечно, в каждом конкретном эксперименте заполнение бункеров будет разным. Но форма этого распределения при достаточно большом количестве шариков всё больше будет напоминать кривую, изображённую на рисунке 11.4.

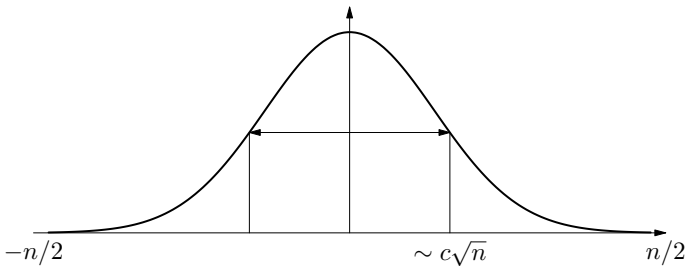

Рис. 11.4: Биномиальные коэффициенты: взгляд издалека

Задача 11.32. Рисунок 11.4 неточный: для наглядности масштаб по оси абсцисс выбран неравномерным. Попробуйте представить, как издалека выглядит график биномиальных коэффициентов при равномерном масштабе по оси абсцисс.

Другой подход к оценке сумм биномиальных коэффициентов состоит в использовании математики вместо бобов и гвоздиков.

Задача 11.33. Докажите, что биномиальные коэффициенты $\binom{n}{k}$ увеличиваются с ростом k вплоть до $n/2$, а затем убывают.

А насколько велик центральный коэффициент $\binom{2n}{n}$? Ясно, что совсем маленьким он быть не может: всего есть $2n+1$ коэффициент, их сумма равна 2^{2n}. Поэтому центральный коэффициент уж никак не меньше среднего значения:

$$\binom{2n}{n} \geqslant \frac{2^{2n}}{2n+1}\,.$$

Но интересно также получить верхнюю оценку для центрального биномиального коэффициента. В терминах вероятности это вероятность того, что в результате $2n$ подбрасываний монеты выпало ровно n орлов. Это то самое значение числа орлов, которое подсказывает нам интуиция. Мы пока лишь убедились, что эта вероятность не слишком мала, она не меньше $1/(2n+1)$.

Оказывается, она и не очень велика. Чтобы получить количественную оценку, можно воспользоваться асимптотической формулой Стирлинга для факториала:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{11.3}$$

(предел отношения этих выражений при $n \to \infty$ равен 1).

Подставляя в формулу для биномиального коэффициента, получаем

$$\binom{2n}{n} = \frac{(2n)!}{n!\cdot n!} \sim \frac{\sqrt{4\pi n}}{2\pi n}\left(\frac{2n}{e}\right)^{2n}\left(\frac{e}{n}\right)^{2n} = \frac{1}{\sqrt{\pi n}}2^{2n}. \tag{11.4}$$

Из этой оценки видим, что с ростом n вероятность получить ровно половину орлов стремится к нулю. Если вам показали результаты 10000 подбрасываний, в которых получилось ровно 5000 орлов, это повод задуматься о «честности» монеты.

Оценки биномиальных коэффициентов с помощью формулы Стирлинга довольно точные, но асимптотические и не очень простые из-за множителей вида $\sqrt{2\pi n}$. Очень часто оказываются удобными более грубые, но более простые оценки биномиальных коэффициентов. Приведём самую популярную пару.

Лемма 11.13. *Для всех n и k, где $k \leqslant n$, верно*

$$\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} < \left(\frac{en}{k}\right)^k. \tag{11.5}$$

Доказательство левого неравенства в (11.5). Запишем выражение для биномиального коэффициента:

$$\binom{n}{k} = \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} = \frac{n}{k}\cdot\frac{n-1}{k-1}\cdot\ldots\cdot\frac{n-(k-1)}{k-(k-1)}\,.$$

Дроби в это произведении увеличиваются слева направо, так как

$$\frac{a-1}{b-1} \geqslant \frac{a}{b} \quad \text{при } a \geqslant b > 1.$$

Заменяя каждую из этих k дробей на наименьшую среди них (это как раз n/k), получаем левую часть левого неравенства в (11.5). □

Доказательство правого неравенства в (11.5). Это неравенство уже требует хоть какого-нибудь анализа (откуда-то должно взяться число e).

Мы используем неравенство

$$e > \left(1 + \frac{1}{k}\right)^k, \tag{11.6}$$

которое сразу следует из самого элементарного определения числа e. Перемножим неравенства (11.6) для $k = 1, 2, \ldots, n-1$, получим

$$e^{n-1} > \left(\frac{2}{1}\right)^1 \cdot \left(\frac{3}{2}\right)^2 \cdot \ldots \cdot \left(\frac{n}{n-1}\right)^{n-1} = \frac{n^{n-1}}{(n-1)!} = \frac{n^n}{n!},$$

т.е. $n! > e(n/e)^n$. Отсюда

$$\binom{n}{k} < \frac{n^k}{k!} < \frac{1}{e} \cdot \left(\frac{en}{k}\right)^k,$$

что и требовалось, так как $e > 1$ (и даже $e > 2$, см. (11.6) при $k = 1$). □

Приведём ещё одно полезное неравенство для биномиальных коэффициентов.

Лемма 11.14. *Докажите, что для любых целых чисел k, t, удовлетворяющих условиям $0 < t \leqslant k \leqslant n/2$, выполняется неравенство*

$$\binom{n}{k-t} < \frac{k}{t^2}\binom{n}{k}.$$

Доказательство. По формуле для биномиальных коэффициентов получаем

$$\binom{n}{k} \Big/ \binom{n}{k-t} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} \cdot \frac{(k-t)!}{n \cdot (n-1) \cdot \ldots \cdot (n-k+t+1)} =$$
$$= \frac{(n-k+t) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot (k-t+1)} = \frac{n-k+t}{k} \cdot \frac{n-k+t-1}{k-1} \cdot \ldots \cdot \frac{n-k+1}{k-t+1}$$

Поскольку $k \leqslant n/2$, числители дробей больше знаменателей. Как и раньше, самая маленькая дробь в этом произведении — первая. Поэтому получаем оценку

$$\binom{n}{k} \Big/ \binom{n}{k-t} > \left(\frac{n-k+t}{k}\right)^t = \left(1 + \frac{n-2k+t}{k}\right)^t \geqslant 1 + \frac{t(n-2k+t)}{k} > \frac{t^2}{k},$$

что и даёт искомое неравенство. (Предпоследнее неравенство — это неравенство Бернулли.) □

Из этой леммы следует, что биномиальные коэффициенты довольно быстро убывают, начиная с расстояния $c\sqrt{n}$ от центрального. Более точные оценки приводятся в следующем разделе.

11.7 Большие уклонения: неравенство Чернова

Обозначим через X_n случайную величину, равную количеству выпавших орлов после n подбрасываний «честной» монеты, а через $\xi_n = X_n/n$ — частоту выпавших орлов. При больших n частота с очень большой вероятностью оказывается близкой к $1/2$. Имеет место следующая оценка, которая очень удобна в приложениях вероятностного метода в комбинаторике, а также во всевозможной теоретической информатике.

Теорема 11.15 (неравенство Чернова)**.** $\Pr\left[|X_n - \frac{n}{2}| > \varepsilon n\right] = \Pr\left[|\xi_n - \frac{1}{2}| > \varepsilon\right] < 2e^{-2\varepsilon^2 n}$.

Это и есть количественная формулировка того интуитивного представления, с которого мы начали обсуждение.

Заметим, что неравенство Чернова симметрично, оно оценивает вероятность уклонений частоты от $1/2$ в обе стороны. В силу симметрии биномиальных коэффициентов

$$\binom{n}{k} = \binom{n}{n-k}$$

достаточно оценивать вероятность превышения частоты над $1/2$ (это объясняет множитель 2 в правой части неравенства Чернова).

План доказательства неравенства Чернова. Изложим вначале общую схему доказательства, пропуская доказательства технических утверждений.

Оказывается, неравенство Чернова — это частный случай неравенства Маркова для подходящим образом подобранной функции от величины X_n.

Удобнее перейти к случайным величинам $Y_n = 2X_n - n$. Если X_n равна сумме n случайных величин, принимающих независимо и равновероятно значения 0 и 1, то Y_n равна сумме величин $y_{n,i}$, каждая из которых независимо принимает случайно и равновероятно значения -1 и $+1$.

Но это не всё: нужно взять экспоненту от Y_n с удачным основанием. Определим случайную величину

$$Z_n = e^{\lambda Y_n} = \prod_i e^{\lambda y_{n,i}}.$$

Оказывается, что в данном случае математическое ожидание произведения равно произведению математических ожиданий сомножителей (в отличие от линейности, мультипликативность не всегда выполняется для математических ожиданий):

$$E[Z_n] = \prod_i E[e^{\lambda y_{n,i}}] = \left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^n = (\operatorname{ch}\lambda)^n. \tag{11.7}$$

В последнем равенстве использовано определение функции гиперболического косинуса $\operatorname{ch} x$. Здесь мы его используем лишь для краткости.

Интересующее нас событие $X_n - n/2 > \varepsilon n$ записывается через случайную величину Y_n как $Y_n > 2\varepsilon n$, а через величину Z_n как $Z_n > e^{2\lambda\varepsilon n}$.

Применим неравенство Маркова к величине Z_n:

$$\Pr[Z_n > e^{2\lambda\varepsilon n}] \leqslant \frac{E[Z_n]}{e^{2\lambda\varepsilon n}} = \left(\frac{\operatorname{ch}\lambda}{e^{2\lambda\varepsilon}}\right)^n$$

Осталось выбрать λ, чтобы сделать дробь в основании степени поменьше. Для этого нужна ещё одна порция анализа, а именно, неравенство

$$\operatorname{ch} x \leqslant e^{x^2/2}. \tag{11.8}$$

Подставляя это неравенство, получаем при $\lambda = 2\varepsilon$

$$\Pr[Z_n > e^{2\lambda\varepsilon n}] \leqslant e^{(2\varepsilon^2 - 4\varepsilon^2)n},$$

что и даёт неравенство Чернова. □

Для завершения доказательства нам нужно проверить два технических утверждения.

Доказательство формулы (11.7). Запишем выражение для математического ожидания Z_n:

$$E[X_n] = \sum_{w\in\{0,1\}^n} 2^{-n} Z_n(w) = 2^{-n} \sum_{w\in\{0,1\}^n} \prod_{i=1}^{n} e^{\lambda(2w_i-1)}.$$

Каждое слагаемое является произведением, в котором стоят e^λ (если $w_i = 1$) и $e^{-\lambda}$ (если $w_i = 0$). Значит, это те же самые слагаемые, которые получаются из бинома $(e^\lambda + e^{-\lambda})^n$ после раскрытия скобок (и до приведения подобных). Поэтому правая часть равенства равна $2^{-n}(e^\lambda + e^{-\lambda})^n$, что совпадает с $(\operatorname{ch}\lambda)^n$. □

Доказательство формулы (11.8). Тут нужно использовать разложение экспоненты в ряд Тейлора:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}\,.$$

Ряд для гиперболического косинуса получается отсюда почленным сложением рядов. Остаются только слагаемые в чётных степенях:

$$\operatorname{ch} x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}\,.$$

Второй ряд получается подстановкой $x^2/2$ в ряд для экспоненты. Опять есть только слагаемые для чётных степеней:

$$e^{-x^2/2} = \sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!}\,.$$

Осталось заметить, что при каждом k выполняется

$$\frac{1}{(2k)!} < \frac{1}{2^k k!}\,,$$

формула (11.8) получается почленным сравнением рядов. □

11.8 Подробности для любознательных

11.8.1 Ещё одна элементарная оценка отношения биномиальных коэффициентов

Лемма 11.14 даёт достаточно хорошее приближение к скорости убывания биномиальных коэффициентов, близких к среднему. Мы сейчас приведём доказательство оценки в другую сторону, которая не использует формулы Стирлинга. В некоторых случаях такие оценки предпочтительнее, так как не зависят от скорости сходимости в формуле Стирлинга.

Будем оценивать сверху величину

$$a_t = \frac{\binom{n}{n/2}}{\binom{n}{n/2-t}},$$

предполагая n чётным (для нечётных всё аналогично). Как и раньше, запишем отношение биномиальных коэффициентов в виде произведения дробей (полагаем $k = n/2$):

$$a_t = \binom{n}{n/2} \Big/ \binom{n}{n/2-t} = \frac{n-n/2+t}{n/2} \cdot \frac{n-n/2+t-1}{n/2-1} \cdot \ldots \cdot \frac{n-n/2+1}{n/2-t+1}.$$

Теперь нас интересует верхняя оценка, поэтому заменим все дроби на наибольшую — последнюю. Получаем

$$a_t < \left(1 + \frac{t}{n/2-t+1}\right)^t. \tag{11.9}$$

Пусть $t < \sqrt{n/4}$. Тогда можно заметить, что слагаемые в разложении бинома

$$\left(1 + \frac{t}{n/2-t+1}\right)^t = \\ = 1 + \binom{t}{1}\frac{t}{n/2-t+1} + \binom{t}{2}\left(\frac{t}{n/2-t+1}\right)^2 + \ldots + \binom{t}{j}\left(\frac{t}{n/2-t+1}\right)^j + \ldots$$

убывают быстрее геометрической прогрессии со знаменателем $1/2$.

Задача 11.34. Докажите это утверждение.

Поэтому при $t < \sqrt{n/4}$ и $n > 2$ получаем неравенство

$$a_t < 1 + \frac{2t^2}{n/2-t+1}. \tag{11.10}$$

(Сравните с оценкой из леммы 11.14.)

11.8.2 Другое доказательство неравенства Чернова

Идея этого доказательства состоит в том, чтобы использовать другую монету, у которой вероятность выпадения орла $p = \frac{1}{2} + \varepsilon$, а вероятность выпадения решки $1 - p = \frac{1}{2} - \varepsilon$. Обозначим через $X_{n,\varepsilon}$ случайную величину, равную количеству выпавших орлов после n независимых подбрасываний этой «испорченной» монеты.

У испорченной монеты вероятности выпадения орлов больше. Оказывается, если сравнить вероятности событий $X_n = k$ (честная монета дала k орлов) и $X_{n,\varepsilon} = k$ (испорченная монета дала k орлов) при $k \geqslant pn$, то первая вероятность намного меньше второй при больших n. Но тогда сумма всех таких вероятностей для честной монеты намного меньше суммы тех же вероятностей для испорченной монеты. А эта вторая сумма уж точно не больше 1. Отсюда и получим верхнюю оценку на вероятность больших уклонений величины X_n.

Подбрасывания испорченной монеты независимые, поэтому по формуле произведения вероятностей независимых событий вероятность каждого результата, содержащего k единиц, равна $p^k(1-p)^{n-k}$. Суммируя по несовместным событиям (все результаты с k единицами), получаем

$$\Pr[X_{n,\varepsilon} = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

и

$$\frac{\Pr[X_n = k]}{\Pr[X_{n,\varepsilon} = k]} = \frac{\binom{n}{k} 2^{-n}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{1}{2^n (p^{k/n}(1-p)^{1-k/n})^n}\ .$$

Обозначим $q = k/n$ и перепишем это отношение вероятностей в виде

$$\frac{\Pr[X_n = qn]}{\Pr[X_{n,\varepsilon} = qn]} = (2p^q(1-p)^{1-q})^{-n}.$$

Мы хотим доказать, что при $p > 1/2$ основание степени в правой части равенства больше 1 и указать одну общую оценку для всех $p \leqslant q \leqslant 1$. Тогда получим желаемое: вероятности для честной монеты окажутся намного меньше, чем для испорченной (поскольку возводим число, бóльшее 1 в отрицательную степень).

Так как $p/(1-p) = (\frac{1}{2} + \varepsilon)/(\frac{1}{2} - \varepsilon) > 1$, функция

$$p^x(1-p)^{1-x} = (1-p) \cdot \left(\frac{p}{1-p}\right)^x$$

возрастающая. Минимальное значение на луче $[p, +\infty)$ она принимает при $x = p$. Поэтому для оценки отношения вероятностей достаточно сравнить с 1 функцию

$$2p^p(1-p)^{1-p}.$$

Удобнее взять логарифм, т.е. перенести функцию в показатель степени. Пусть это будет двоичный логарифм:

$$\log_2(p^p(1-p)^{1-p}) = p\log_2 p + (1-p)\log_2(1-p) \stackrel{\text{def}}{=} -h(p).$$

Заметим, что $1 = \log_2 2 = h(1/2)$.

Лемма 11.16. *Функция $h(x)$ на интервале $(0,1/2)$ возрастает, а на интервале $(1/2,1)$ убывает. Точка $1/2$ тем самым является точкой максимума.*

Доказательство. Нужно вычислить производную $h(x)$:

$$h'(x) = -\log_2 x - \frac{1}{\ln 2} + \log_2(1-x) + \frac{1}{\ln 2} = \log_2 \frac{1-x}{x}.$$

Так как $1-x > x$ равносильно $x < 1/2$, получаем, что интервале $(0,1/2)$ производная положительная, а на интервале $(1/2,1)$ производная отрицательная. Отсюда и следует утверждение леммы. □

Теперь воспользуемся леммой и перепишем оценку на отношение вероятностей как

$$\frac{\Pr[X_n = qn]}{\Pr[X_{n,\varepsilon} = qn]} \leqslant 2^{-(h(1/2)-h(p))n} = 2^{-\eta^2 n}. \tag{11.11}$$

Число $\eta > 0$ зависит только от выбранного порога частоты ε.

Теорема 11.17. $\Pr\left[|\xi_n - \frac{1}{2}| > \varepsilon\right] < 2 \cdot 2^{-\eta^2 n}$.

Доказательство. Искомая вероятность в два раза больше, чем

$$\sum_{k > n/2+\varepsilon n} \Pr[X_n > k] < \sum_{k > n/2+\varepsilon n} \Pr[X_{n,\varepsilon} > k] 2^{-\eta^2 n} \leqslant 2^{-\eta^2 n}$$

(так как сумма вероятностей не превосходит 1). □

Мы получили, что вероятности больших уклонений убывают экспоненциально быстро.

Применив ещё чуть больше анализа, можно явно выразить η через ε. Вторая производная $h(x)$ на интервале $(1/2,1)$ убывает, так как

$$h''(x) = -\frac{1}{\ln 2} \cdot \frac{1}{x(1-x)}\ .$$

Поэтому функция $h(1/2) - h(x) + \frac{h''(1/2)}{2}(x-1/2)^2$ является выпуклой (вторая производная неотрицательна). Касательная в точке $x = 1/2$ к графику этой функции горизонтальна. Значит, весь график $h(x)$ лежит либо целиком выше графика $h(1/2) + \frac{h''(1/2)}{2}(x-1/2)^2$, либо целиком ниже.

В точке $x = 1$ функция h обращается в 0, а $h(1/2) + \frac{h''(1/2)}{2}(x-1/2)^2 = 1 - 1/2\ln 2 > 0$. Значит, выполняется неравенство

$$h(1/2) - h(1/2+\varepsilon) \geqslant -\frac{h''(1/2)}{2}\varepsilon^2 = \frac{2}{\ln 2}\varepsilon^2.$$

Подставляя в (11.11), получаем неравенство Чернова

$$\Pr\left[|\xi_n - \tfrac{1}{2}| > \varepsilon\right] < 2e^{-2\varepsilon^2 n}.$$