Лекция 8. Вероятность: первые шаги.

Дискретная математика, ВШЭ, факультет компьютерных наук

(Осень 2014 – весна 2015)

О вероятности говорят, когда мы не знаем, каким будет исход того или иного события, но тем не менее хотим что-то о нем сказать. Простейшей моделью является модель с равновозможными исходами. В этой модели, мы считаем, что есть несколько возможных исходов события и все они одинаково возможны. Например, если мы подбрасываем монетку, то можно считать, что "орел" 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 произойдет. 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 орел", 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.

Теперь скажем все то же самое более формально и в общем случае. Вероятностным пространством называется конечное множество U возможных исходов. На вероятностном пространстве задана функция $Pr\colon U \to [0,1]$, такая что $\sum_{x\in U} Pr[x] = 1$. Число $Pr[x]$ называется вероятностью исхода $x \in U$. Событием называется произвольное подмножество $A \subseteq U$. Вероятностью события A называется число $Pr[A] = \sum_{x\in A} Pr[x]$.

Пример 1. В модели с равновозможными исходами функция p задается формулой $Pr[x] = 1/|U|$ для всякого $x \in U$. Тогда вероятность события A равна $Pr[A] = |A|/|U|$.

Кроме равновозможной модели мы также можем говорить и о произвольной вероятностной модели. В ней также 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 подсчета 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.

В большинстве обсуждаемых нами примеров исходы равновозможны.

Лемма 1. *Если $A, B \subseteq U$, то $Pr[A \cup B] \leqslant Pr[A] + Pr[B]$. Если кроме того $A \cap B = \emptyset$, 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 $A \cap B = \emptyset$, называются несовместными. Таким образом вероятность объединения несовместных событий равна сумме их вероятностей.

Заметим, что если речь идет 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 еще разделить его на количество всех исходов.

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

Следствие 1. *В равновозможной модели для произвольных множеств $A_1, \ldots, A_n \subseteq U$ верно*

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

Доказательство. Действительно, по принципу включений и исключений для мощностей, аналогичное равенство верно для мощностей множеств. Чтобы получить аналогичное равенство 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 произвольных.

Лемма 2. *Для всякой вероятностной модели и для произвольных множеств $A_1, \ldots, A_n \subseteq U$ верно*

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

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 разберемся со случаем двух множеств:

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

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 шага, 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 своем месте. В терминах теории вероятности мы можем оценить долю таких перестановок.

Лемма 3 (Задача о беспорядках)**.** *Рассмотрим случайные перестановки n различных объектов, то есть рассмотрим вероятностное пространство всех перестановок на n заданных объектов с равновероятным распределением на них. 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 означающее, что 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 сошлемся на факт, который будет позднее обсуждаться в курсе математического анализа, а именно, что эта сумма очень напоминает начало ряда Тейлора для функции e^x, а более конкретно, мы воспользуемся тем, что $e^{-1} = \sum_{k=0}^{\infty}(-1)^k/k!$. Отсюда

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

□

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 на примере чисел Расмея. 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 1. *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 каждого ребра графа есть два выбора, либо добавить его в граф, либо нет. Всего ребер в графе $\binom{n}{2}$, так что всего 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,

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

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

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

1 Условные вероятности

Кроме того, чтобы говорить о вероятностях тех или иных событий, оказывается нужным говорить и о вероятностях одних событий при условии других. Мы хотим определить вероятность выполнения события A в том случае, если мы знаем, что 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 пересечений событий A и B достаточно найти вероятность события B и условную вероятность события A при условии события B.

Задача 1. 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 2. Приведите примеры, в которых условная вероятность $\Pr[A \mid B]$ больше вероятности $Pr[A]$, меньше её, а также равна ей.

Уже из определения условной вероятности можно получить на первый взгляд неожиданные утверждения.

Лемма 4 (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. Например, рассмотрим некоторую болезнь, и предположим, что для ее обнаружения мы можем делать недорогой анализ, который при этом с заметной вероятностью выдает неправильный результат, а также есть дорогостоящее исследование, которое уже наверняка сообщает, болен ли человек. Мы хотим попробовать обнаруживать болезнь следующим образом: сначала для человек сдает недорогой анализ, а если он дал положительный результат, то проверяем его с помощью дорогостоящего исследования. При таком подходе возможно, что больной человек не будет обнаружен нашим анализом. Можем ли мы 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». Тогда мы можем узнать $Pr[A]$ – вероятность для человека получить положительный анализ, мы можем собрать статистику. Также можно примерно предсказать $Pr[B]$ – долю больных рассматриваемой болезнью. Наконец, можно приблизительно предсказать 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. 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 дает формулу Байеса. □

Понятие условной вероятности позволяет нам также говорить о независимых событиях. Неформально, событие A не зависит от события B, если информация о выполнении события

B не влияет на вероятность выполнения события A. Формально, 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.

Задача 3. Пусть события A и B независимы, и вероятность события $\overline{B}$ положительна. Докажите, что события A И $\overline{B}$ независимы.

Другим полезным утверждением об условных вероятностях является формула полной вероятности.

Лемма 5 (Формула полной вероятности)**.** *Пусть $B_1, \ldots, B_n$ – разбиение вероятностного пространства U, то есть $U = B_1 \sqcup \ldots \sqcup B_n$, и для всякого i $Pr[B_i] > 0$. 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.

□

Задача 4. 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. **а)** С какой вероятностью в случайной редакционной колонке обнаружится фактическая ошибка? **б)** В редакционной колонке обнаружена фактическая ошибка. С какой вероятностью ее написал журналист X?

Формула полной вероятности полезна при работе с вероятностными распределениями и мы ее на самом деле уже использовали при решении задач на семинарах. Проиллюстрируем это простым примером.

Пример 2. Пусть G – простой неориентированный граф с n вершинами $v_1, \ldots, v_n$, такой что степени всех вершин равны некоторому числу d (такой граф называется регулярным). Как мы уже доказывали у него $nd/2$ ребер. Рассмотрим два вероятностных распределения на его ребрах. Первое – равновозможное, то есть каждое из $nd/2$ ребер 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 не сложно – у нас задано

равновероятное распределение на $nd/2$ ребрах, так что вероятность равна $2/nd$. Чтобы посчитать вероятность во втором случае рассмотрим событие B_i для всякого $i \in \{1, \ldots, n\}$, 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 ребер 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.

Прием с заменой вероятностного распределения эквивалентным, подобно тому, как в только что разобранном примере, бывает очень полезен на практике.

2 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 ее 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 свое число – это функция).

Важным параметром случайной величины является ее 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 событие 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 мы повторяем эксперимент по выбору случайного элемента из U 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. 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 $1/6$. Ее 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 6. *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*

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

Другими словами, математическое ожидание линейно.

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

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

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

□

Указанная лемма во многих случаях заметно упрощает вычисление математического ожидания.

Пример 3 (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?

Для простоты предположим, что все дни в году могут оказаться днями рождения с равными вероятностями. Обозначим людей через $x_1, \ldots, x_n$, а 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 разобьем сложную случайную величину f в сумму нескольких простых случайных величин. Тогда мы сможем подсчитать отдельно математические ожидания всех простых величин, а затем, пользуясь предыдущей леммой, 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 $1 - 1/365$ – равна 0. Так что $E[g_{ij}] = 1/365$ (для всякой пары i, j). Так что для математического ожидания f мы 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 около одного совпадения дней рождений, что может показаться противоречащим интуиции (поэтому эту задачу иногда называют «парадоксом дней рождения»).

С помощью математического ожидания можно обобщить вероятностный метод.

Лемма 7. *Пусть для какой-то случайной величины $f\colon U \to \mathbb{R}$ верно $E[f] = C$. Тогда существует $u \in U$, такой что $f(u) \geqslant C$. Аналогично, существует $u \in U$, такое что $f(u) \leqslant C$.*

Доказательство. Докажем первое утверждение леммы, второе доказывается аналогично.

На самом деле, доказательство довольно простое. Предположим, что утверждение не верно, 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 5. Объясните, почему старая формулировка вероятностного метода является частным случаем новой формулировки.

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

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

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

Доказательство. Рассмотрим случайный разрез графа G. То есть мы берем равновероятное распределение на всех разрезах. 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 $1/2^n$. Из этого видно, что взять равновероятно случайно подмножество – это то же самое, что задать случайное подмножество следующим образом: для каждого $x \in V$ добавляем x в S случайно с вероятностью $1/2$, причем для каждого x делаем это независимо. Можно заметить, что вероятность получить любое конкретное множество есть снова $1/2^n$, так что мы только что вторым способом задали то же самое распределение. Отметим, что не то чтобы рассуждение этого абзаца как-то принципиально важно для этой конкретной теоремы и тесно с ней связано. Но оно полезно в целом, а в этой теореме нашелся повод его рассказать.

Итак, рассмотрим случайный разрез и рассмотрим случайную величину f, равную размеру разреза. Посчитаем ее 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, 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 не трудно посчитать. Действительно, для всякого фиксированного ребра e вероятность, что оно попадет в разрез равна $1/2$. А значит для всякого $e \in E$ $E[f_e] = 1/2$, откуда

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

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

На самом деле, этот результат не сильно удивителен, его можно доказать и “руками”.

Задача 6. 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 3. *Рассмотрим граф $G = (V, E)$, в котором количество вершин $|V| = 2n$ – четно. 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 7. Докажите, что если в графе $G = (V, E)$ число вершин $|V| = 2n+1$ – нечетно, то есть разрез размера не меньше $\frac{|E|(n+1)}{2n+1}$.

Математическое ожидание позволяет давать оценки вероятностей некоторых событий.

Лемма 8 (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 ее 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 остается 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 8. Где в нашем доказательстве мы использовали неотрицательность случайной величины? Остается ли лемма верной, если убрать условие неотрицательности случайной величины?

Задача 9. 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 4. Приведем 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 могут 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 время $O(n^2)$, где n – размер входных данных. Для наших практических целей, нам бы хотелось, чтобы алгоритм всегда (то есть независимо от случайных чисел) заканчивал свою работу за время $O(n^2)$, и 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 ждем пока он сделает $10000 \cdot T$ шагов. Если алгоритм успел выдать ответ, прекрасно. Если нет, выдаем 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 $0,01\%$.

3 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: это неявно предполагается в простонародном понимании «честности». Если кто-то будет подбрасывать монету и после выпадения четвёртого орла подряд переворачивать его, 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). Оно равно биномиальному коэффициенту

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

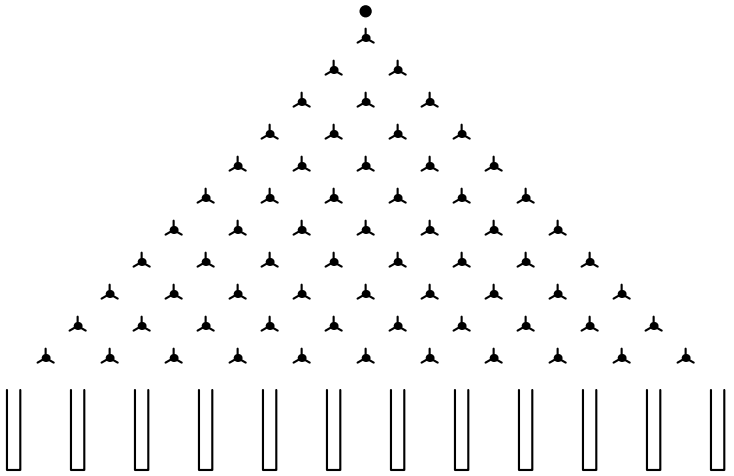

Рис. 1: Принципиальная схема доски Гальтона

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.

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 1.

Если набросать через такую решетку 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 все больше будет напоминать кривую, изображенную на рисунке 2.

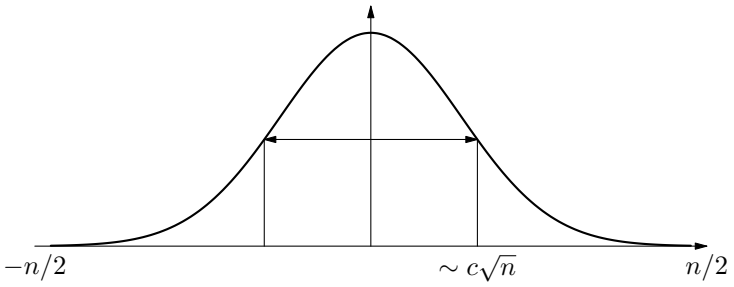

Рис. 2: Биномиальные коэффициенты: взгляд издалека

Задача 10. Рисунок 2 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, более нам близкий, состоит в использовании математики вместо бобов и гвоздиков.

Задача 11. 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^n. 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 $1/(n+1)$.

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

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{3}$$

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

Подставляя в формулу для биномиального коэффициента, получаем

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

Из этой оценки видим, что с ростом n вероятность получить ровно половину орлов убывает к нулю. Если вам показали результаты 10000 подбрасываний, в которых получилось ровно 5000 орлов, это серьёзный 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. Приведем самую популярную пару:

$$\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} < \left(\frac{en}{k}\right)^k. \tag{5}$$

Доказательство левого неравенства в (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

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

Заменяя каждую из этих k дробей на наименьшую среди них (это как раз n/k), получаем левую часть левого неравенства в (5). □

Доказательство правого неравенства в (5). Это неравенство уже требует хоть какого-нибудь анализа (откуда-то должно взяться число e).

Мы используем неравенство

$$e > \left(1+\frac{1}{k}\right)^k, \tag{6}$$

которое сразу следует из самого элементарного определения числа e. Перемножим неравенства (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 $e > 2$ (см. (6) при $k = 1$). □

Приведём ещё одно полезное неравенство для биномиальных коэффициентов.

Лемма 9. *Докажите, что для любых целых чисел k, t, удовлетворяющих условиям $0 \leqslant t \leqslant k \leqslant n/2$, 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 маленька 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. □

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.

4 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.

Теорема 4 (неравенство Чернова)**.** $\Pr\left[|X_n - \frac{n}{2}| > \varepsilon\right] = \Pr\left[|\xi_n - \frac{1}{2}| > \varepsilon\right] < 2e^{-2\varepsilon^2 n}$.

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 все: 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):

$$E[Z_n] = \prod_i E[e^{\lambda y_{n,i}}] = \left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^n = (\operatorname{ch}\lambda)^n. \tag{7}$$

Основания степеней равны в последнем равенстве по определению функции гиперболического косинуса $\operatorname{ch} x$.

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 еще одна порция анализа, а именно, неравенство

$$\operatorname{ch} x \leqslant e^{x^2/2}. \tag{8}$$

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 дает неравенство Чернова. □

Для завершения доказательства нам нужно проверить два технических утверждения.

Доказательство формулы (7). Запишем определение 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 слагамое 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 (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 четных 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 слаагаемые для четных 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 (8) получается почленным сравнением рядов. □

5 Подробности для любознательных

5.1 Ещё одна элементарная оценка отношения биномиальных коэффициентов

Лемма 9 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 $k = n/2$):

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

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

$$a_t < \left(1 + \frac{t}{k-t+1}\right)^t \, . \tag{9}$$

Пусть $t < \sqrt{k/3}$. Тогда можно заметить, что слагаемые в разложении бинома

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

убывают, начиная со второго, быстрее геометрической прогресии со знаменателем $1/2$.

Задача 12. Докажите это утверждение.

Поэтому при $t < \sqrt{k/3}$ и $k > 1$ получаем неравенство

$$a_t < 1 + \frac{2t^2}{k-t+1} < \frac{3t^2}{k} \, , \tag{10}$$

что всего в 3 раза больше нижней оценки из леммы 9.

5.2 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 10. *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

$$\frac{\Pr[X_n = qn]}{\Pr[X_{n,\varepsilon} = qn]} \leqslant 2^{-(h(1/2)-h(p))n} = 2^{-\eta^2 n}. \tag{11}$$

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

Теорема 5. $\Pr\left[|\xi_n - \frac{1}{2}| > \varepsilon\right] < 2 \cdot 2^{-\eta^2 n}$.

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 еще 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), получаем неравенство Чернова

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