Дискретная математика

Конспекты лекций М. Н. Вялого,
ПМИ ФКН ВШЭ, 1 курс, основной поток, 2020/21 уч.г.

Версия от 19 марта 2021 г.

Файл пополняется по мере чтения лекций. Просьба о замеченных ошибках и неточностях сообщать автору (адрес `vyalyi`⟨стандартный суффикс gmail⟩).

Оглавление

Лекция 1

Алгебра логики

Для занятий математикой важно отличать осмысленные утверждения от бессмысленных, а среди осмысленных утверждений различать истинные и ложные.

Для первого математики дают ясные и недвусмысленные определения используемых в утверждениях слов. Каждый математический термин означает лишь то, что указано в его определении. Например, совершенное число равно сумме всех своих делителей, которые меньше самого числа (например, $6 = 1+2+3$). Никаких отсылок к смыслу слова «совершенный» в обыденном языке в математических рассуждениях о совершенных числах не допускается.

Для второго математики используют доказательства — рассуждения по правилам формальной логики, которые объясняют, почему из истинности одних утверждений следует истинность других утверждений.

Мы сейчас познакомимся с примерами доказательств и с самыми началами формализма, придуманного математиками для доказательств.

1.1. Булевы функции–1. Логические связки и доказательства

В элементарной алгебре очень много доказательств имеют вид вычислений.

Пример 1.1. Тождество $x^2 - y^2 = (x - y)(x + y)$ доказывается цепочкой равенств:

$$(x - y)(x + y) = (x - y)x + (x - y)y = x^2 - yx + xy - y^2 = x^2 - y^2$$

(на самом деле, часть равенств пропущена). Каждое равенство использует одно из основных свойств арифметических операций с числами. □

Аналогично этим доказательствам-вычислениям, можно строить «алгебраические» доказательства и в более общем случае так называемой алгебры высказываний.

Мы считаем, что высказывание обязательно либо истинно, либо ложно (в противном случае фраза не является высказыванием). Из высказываний можно строить составные высказывания, используя *логические связки*.

Значения составных высказываний определяются в зависимости от связки по таблицам истинности. Приведём таблицы истинности самых употребительных связок: конъюнкции $\wedge$ «A и B», дизъюнкции $\vee$ «A или B», импликации (логическое следование: «если A, то B», «из A следует B», обозначение $\to$) и равносильности $\equiv$ «A равносильно B».

A	B	$A \wedge B$	$A \vee B$	$A \to B$	$A \equiv B$
Л	Л	Л	Л	И	И
Л	И	Л	И	И	Л
И	Л	Л	И	Л	Л
И	И	И	И	И	И

Заметим, что аргументы в импликации не симметричны, в отличие от остальных связок. Им поэтому присвоены специальные названия: в импликации $A \to B$ высказывание A называется *посылкой*, а высказывание B — *заключением*.

Таблица истинности импликации требует комментариев. На впечатлительные умы производит сильное впечатление тот факт, что из лжи следует истина. Однако, более точно сказать, что из лжи следует что угодно, не только истина, но и ложь.

Таблица истинности импликации вполне согласуется с бытовым применением конструкции «если ..., то ...». Представим себе, что кто-то делает следующее заявление: «Если я завтра заболею, то не приду на занятия». Если на следующий день этот человек не заболел и пришёл на занятия, следует ли считать, что он соврал? А если он не заболел и не пришёл? И то, и другое было бы странно, ведь человек и вовсе ничего не говорил про этот случай. Нечто содержательное было сказано только для той ситуации, когда он заболел. Так что, если посылка импликации ложна, то и в обычной жизни утверждение считают истинным.

Приведём аналогичный пример из математики. Рассмотрим утверждение «если число n делится на 4, то n чётно». Это утверждение верно. И оно не перестанет быть верным, если вместо произвольного числа подставить какое-нибудь конкретное, например 6. Посылка утверждения становится ложной, 6 не делится на 4, но само утверждение остаётся истинным.

Ещё одна постоянно используемая связка: отрицание «не A» (обозначение $\neg$). Таблица отрицания очень простая: $\neg Л = И$, $\neg И = Л$.

Логические тождества выражают законы логики. Если истинно утверждение, выражающееся левой частью тождества, то истинно и утверждение, выражающееся правой частью тождества. И наоборот. Части логического тождества соединяют знаком $\equiv$ и называют *равносильными* утверждениями.

Давайте рассмотрим пример такого закона логики, который применяется во всех доказательствах. По таблицам истинности легко проверить, что

$$(A \wedge (A \to B)) \to B \equiv И. \tag{1.1}$$

Это тождество мы используем всякий раз, когда из истинности высказывания A и составного высказывания «если A, то B» мы заключаем, что истинно B. Это

логическое правило, на латыни называемое modus ponens, лежит в основе всех математических рассуждений.

Составные высказывания, которые истинны при любых значениях входящих в них элементарных высказываний, называются *тавтологиями*. Высказывание в левой части (1.1) — пример тавтологии.

Помимо механической проверки тождества по таблицам истинности можно применить такой приём доказательств, как *разбор случаев*. Тут важно не пропустить какой-нибудь случай, иначе доказательство будет неполным.

Доказательство тождества 1.1 разбором случаев. Пусть $B = \mathsf{И}$. Из таблицы импликации видим, что $X \to \mathsf{И} \equiv \mathsf{И}$. Поэтому левая часть истинна.

Пусть $B = \mathsf{Л}$. Из таблицы импликации видим, что $X \to \mathsf{Л} = \neg X$. Поэтому левая часть приобретает вид $(A \wedge \neg A) \to \mathsf{Л} = \neg(A \wedge \neg A)$. Поскольку любое высказывание либо истинно, либо ложно, $A \wedge \neg A = \mathsf{Л}$, а $\neg(A \wedge \neg A) = \mathsf{И}$. □

В этом доказательстве возникло тождественно ложное высказывание $A \wedge \neg A$, которое называется *противоречием*. Если мы имеем противоречие — два противоположных высказывания «A» и «не A», — то по modus ponens мы можем вывести ложь, так как $(A \wedge \neg A) \to \mathsf{Л} \equiv \mathsf{И}$. А из лжи мы уже выведем что угодно. Это объясняет, почему математики так беспокоятся об отсутствии противоречий. Любое противоречие позволяет утверждать, что истинно любое высказывание.

Одним из популярных приёмов доказательств являются *доказательства от противного*. Допустим, мы хотим доказать утверждение A. Для этого мы выводим из его отрицания $\neg A$ противоречие, то есть доказываем, что некоторое высказывание B одновременно истинно и ложно. Другими словами, мы доказываем истинность составного высказывания $\neg A \to (B \wedge \neg B)$. Заключение этой импликации ложно. Поскольку сама импликация истинна, то и её посылка ложна. То есть, A истинно, что и требовалось доказать.

Пример 1.2. Приведём простой пример доказательства от противного. Докажем такое утверждение[1)]: если $a_1 + \cdots + a_n > n$, то какое-то из этих чисел больше 1.

Действительно, пусть $a_i \leqslant 1$ для всех i. Складывая эти неравенства, получаем, что $a_1 + \cdots + a_n \leqslant n$. Пришли к противоречию. □

Одной из важнейших тавтологий является *транзитивность* импликации:

$$\big((A \to B) \wedge (B \to C)\big) \to (A \to C) \equiv \mathsf{И}. \tag{1.2}$$

Это логическое тождество позволяет строить длинные цепочки вывода, аналогичные вычислениям в примере 1.1. Однако во многих случаях структура математических доказательств сложнее, она нелинейная.

[1)] Утверждение кажется очевидным, но мы уже говорили, что математики панически боятся противоречий. Поэтому «очевидные» утверждения нуждаются в доказательствах. В какой-то момент придётся остановиться и признать некоторые утверждения истинными, не давая им доказательств. В данном случае мы принимаем без доказательств свойства арифметических операций и сравнений чисел.

Доказательство (1.2) *от противного.* Предположим ложность левой части (1.2). Из таблицы истинности импликации видим, что заключение $A \to C$ внешней импликации ложно, а посылка $(A \to B) \wedge (B \to C)$ истинна.

Из ложности $A \to C$ заключаем, что $A = \mathsf{И}$, $C = \mathsf{Л}$. Истинность конъюнкции означает, что истинны оба члена конъюнкции, в частности $B \to C = B \to \mathsf{Л} = \mathsf{И}$. Это возможно лишь при $B = \mathsf{Л}$. Но тогда $A \to B = \mathsf{И} \to \mathsf{Л} = \mathsf{Л}$, а мы уже установили, что $A \to B = \mathsf{И}$. Пришли к противоречию. □

Ещё один важный приём доказательств основан на тавтологии, которая называется *законом контрапозиции*:

$$A \to B \equiv \neg B \to \neg A. \tag{1.3}$$

Доказательство тождества 1.3. Левая часть ложна тогда и только тогда, когда $A = \mathsf{И}$, $B = \mathsf{Л}$.

Правая часть ложна тогда и только тогда, когда $\neg B = \mathsf{И}$, $\neg A = \mathsf{Л}$. Но это равносильно первым двум условиям. Поэтому значения левой и правой части всегда одинаковы. □

Контрапозиция используется как логический шаг в очень многих доказательствах, у нас дальше будет много таких примеров. Пока приведём один несложный пример.

Пример 1.3. Докажем такое утверждение: если неотрицательное число x иррациональное, то и $\sqrt{x}$ иррациональное.

Закон контрапозиции говорит, что это утверждение равносильно такому: если $\sqrt{x}$ рациональное, то и x рациональное. А это уже легко доказать: пусть $\sqrt{x} = n/m$, где числитель и знаменатель целые. Тогда

$$x = \sqrt{x} \cdot \sqrt{x} = \frac{n}{m} \cdot \frac{n}{m} = \frac{n^2}{m^2}$$

также рациональное. □

Важным примером тавтологий являются *законы де Моргана.* Они позволяют строить высказывания, равносильные отрицанию конъюнкции или дизъюнкции:

$$\neg(A \wedge B) \equiv \neg A \vee \neg B; \quad \neg(A \vee B) \equiv \neg A \wedge \neg B. \tag{1.4}$$

Доказательство тождеств (1.4). Первое. Отрицание конъюнкции ложно тогда и только тогда, когда конъюнкция истинна, то есть $A = B = \mathsf{И}$. Дизъюнкция ложна тогда и только тогда, когда каждый её член ложен, то есть $\neg A = \neg B = \mathsf{Л}$. Эти условия равносильны.

Второе. Отрицание дизъюнкции истинно тогда и только тогда, когда дизъюнкция ложна, то есть $A = B = \mathsf{Л}$. Конъюнкция истинна тогда и только тогда, когда каждый её член истинен, то есть $\neg A = \neg B = \mathsf{И}$. Эти условия равносильны. □

Для конъюнкции и дизъюнкции выполняется много тавтологий, напоминающих свойства обычных арифметических операций. Все они легко доказываются аналогично рассмотренным выше тавтологиям.

Коммутативность конъюнкции, дизъюнкции:

$$A \wedge B \equiv B \wedge A, \quad A \vee B \equiv B \wedge A.$$

Ассоциативность тех же связок:

$$(A \wedge B) \wedge C \equiv A \wedge (B \wedge C), \quad (A \vee B) \vee C \equiv A \vee (B \vee C).$$

Дистрибутивность (тут даже лучше, чем с числами — есть две тавтологии):

$$(A \wedge B) \vee C \equiv (A \vee C) \wedge (B \vee C), \quad (A \vee B) \wedge C \equiv (A \wedge C) \vee (B \wedge C).$$

Этими простыми логическими тождествами законы логики не исчерпываются. В математике нужны ещё законы логики, позволяющие оперировать с кванторами: *квантором всеобщности* $\forall x A(x)$ («для всех x истинно утверждение $A(x)$» и *квантором существования* $\exists x A(x)$ («существует такое x, что $A(x)$ истинно». Систематическое изучение логики с кванторами мы откладываем до второго курса.

Пока лишь скажем, что квантор $\forall$ — это в сущности конъюнкция (по возможным значениям x), а квантор $\exists$ — дизъюнкция. Законы де Моргана позволяют строить утверждения, равносильные отрицаниям утверждений с кванторами

$$\neg\forall x A(x) \equiv \exists x\, \neg A(x), \qquad \neg\exists x A(x) \equiv \forall x\, \neg A(x).^{2)}$$

Эти логические правила вам потребуются всюду в математике.

[2)] Спасибо Никите Лукьяненко за замеченную опечатку.

Лекция 2

Множества, индукция

2.1. Множества–1. Операции с множествами

Множества лежат в основе формального языка математики. Мы сейчас дадим краткие и неформальные объяснения свойств множеств и операций с ними. Таких неформальных объяснений будет достаточно почти для всех тем этого курса. Для лучшего понимания советуем прочитать более подробные книги о множествах. Хорошим введением в теорию множеств является книга Верещагина и Шеня «Начала теории множеств».

Множество — это совокупность каких-то *элементов*. Природа элементов неважна. Никакие взаимоотношения между элементами не важны. Единственное, что существенно — какие элементы в входят в множество, а какие — нет.

Множество полностью определяется своими элементами. Два множества A и B называются *равными*, если каждый элемент множества A является элементом множества B, а каждый элемент множества B является элементом множества A. Множество A является *подмножеством* множества B, если каждый элемент множества A принадлежит множеству B (обозначение $A \subseteq B$). Высказывание «элемент x *принадлежит* множеству A» (обозначение $x \in A$) истинно, если в A есть элемент x, и ложно в противном случае.

Пример 2.1 (Пустое множество)**.** Есть такое множество, которому не принадлежит ни один элемент. Оно называется *пустым множеством* и обозначается $\varnothing$. Высказывание $x \in \varnothing$ ложно для любого x.

Конечное множество можно задать списком его элементов. Хотя на письме мы вынуждены записывать множества в каком-то порядке, этот порядок не имеет значения, как и возможные повторения элементов. Принято заключать список элементов в фигурные скобки.

Пример 2.2. Множество чётных цифр может быть задано такими способами

$$\{0, 2, 4, 6, 8\} = \{4, 2, 0, 8, 6\} = \{2, 2, 4, 6, 8, 6, 0, 2\}.$$

Списки разные, но задают они одно и то же множество. □

Из уже имеющегося множества можно выделить подмножество, указав некоторое свойство элементов. При таком определении множеств обычно используется запись вида

$$\{x \in \mathbb{Z} \mid x = 2y,\ y \in \mathbb{Z}\}$$

или

$$\{x \in \mathbb{Z} : x = 2y,\ y \in \mathbb{Z}\}.$$

В этих формулах $\mathbb{Z}$ обозначает множество целых чисел. Обе формулы определяют множество чётных чисел.

Ещё один способ строить новые множества из уже имеющихся состоит в применении к множествам *операций*.

Перечислим основные операции с множествами.

Объединение множеств. Обозначение $A \cup B$. Это множество, состоящее в точности из тех элементов, которые принадлежат хотя бы одному из множеств A и B. В формальной записи это определение выглядит так:

$$A \cup B = \{x : (x \in A) \vee (x \in B)\}. \tag{2.1}$$

Пересечение множеств. Обозначение $A \cap B$. Это множество, состоящее в точности из тех элементов, которые принадлежат обоим множествам A и B. В формальной записи это определение выглядит так:

$$A \cap B = \{x : (x \in A) \wedge (x \in B)\}. \tag{2.2}$$

Разность множеств. Обозначение $A \setminus B$. Это множество, состоящее в точности из тех элементов, которые принадлежат множеству A, но не принадлежат множеству B. В формальной записи это определение выглядит так:

$$A \setminus B = \{x : (x \in A) \wedge \neg(x \in B)\}. \tag{2.3}$$

Симметрическая разность множеств. Обозначение $A \triangle B$. Это множество, состоящее в точности из тех элементов, которые принадлежат ровно одному из множеств: либо A, либо B. В формальной записи это определение выглядит так:

$$A \triangle B = \Big\{x : \big((x \in A) \wedge \neg(x \in B)\big) \vee \big(\neg(x \in A) \wedge (x \in B)\big)\Big\}. \tag{2.4}$$

Помимо словесных определений, приведённых выше, есть наглядный графический способ иллюстрировать операции с множествами: круги Венна (иногда говорят Эйлера–Венна). При этом способе множество изображается условным кругом (или другой геометрической фигурой) и предполагается, что внутренность круга изображает элементы множества.

На паре кругов легко изобразить объединение, пересечение, разность и симметрическую разность множеств, как это сделано на рисунке 2.1. Для этого результат применения операции выделяют штриховкой или цветом.

Есть ещё одна часто используемая операция с множествами — дополнение к множеству (обозначение $\bar{A}$) определяют обычно как те элементы, которые не входят в

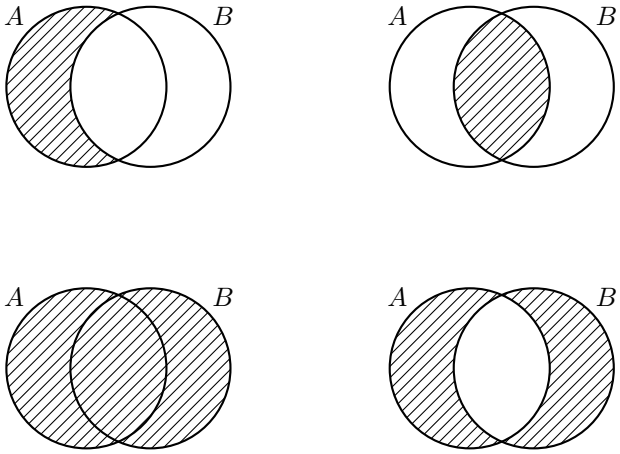


Рис. 2.1: Теоретико-множественные операции на кругах Венна

множество A. Однако буквально такое определение несодержательно. Операция дополнения всегда используется в рассуждениях о подмножествах одного множества U, которое будем называть *универсумом*. В таком контексте дополнение $\bar{A}$ даёт сокращённую запись для разности $U \setminus A$.

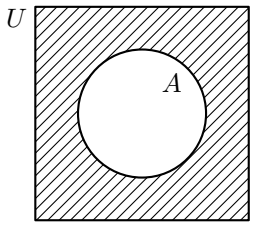


Рис. 2.2: Дополнение $\bar{A} = U \setminus A$ на рисунке заштриховано

2.2. Множества–2. Тавтологии и теоретико-множественные тождества

Между операциями с множествами и логическими связками есть соответствие. Предположим, что все рассматриваемые множества являются подножествами универсума U. Каждому множеству A и каждому элементу U сопоставляется высказывание $x \in A$. Это высказывание истинно для элементов A и ложно для остальных элементов универсума. Из формальных определений операций с множествами (2.1),

(2.2), (2.3) получаем такие эквивалентности

$$(x \in A \cup B) \equiv (x \in A) \vee (x \in B),$$
$$(x \in A \cap B) \equiv (x \in A) \wedge (x \in B),$$
$$(x \in A \setminus B) \equiv (x \in A) \wedge \neg(x \in B),$$
$$(x \in \bar{A}) \equiv \neg(x \in A).$$

Равенство множеств A и B, как уже говорилось, равносильно эквивалентности высказываний $x \in A$ и $x \in B$. Поэтому любой тавтологии взаимно однозначно отвечает некоторое теоретико-множественное тождество, то есть утверждение о том, что применение теоретико-множественных операций в разном порядке даёт одно и то же множество.

Пример 2.3. Докажем, что равенство

$$(A \cap B) \setminus C = (A \setminus C) \cap B$$

выполняется для любых множеств. Для этого запишем соответствующую логическую формулу:

$$(A \wedge B) \wedge \neg C \equiv (A \wedge \neg C) \wedge B.$$

Эта формула является тождеством, потому что конъюнкция коммутативна и ассоциативна. □

Что соответствует импликации на языке множеств? Посмотрев на таблицу значений, видим, что импликация истинна тогда и только тогда, когда элемент принадлежит второму множеству ИЛИ не принадлежит первому. То есть, теоретико-множественная операция, отвечающая импликации, записывается как $\bar{A} \cup B$.

У импликации есть ещё один важный смысл на языке множеств. Утверждение «если $x \in A$, то $x \in B$» (любой элемент A входит в B) выполняется для всех x тогда и только тогда, когда $A \subseteq B$, мы так определяли подмножества. Поэтому импликации соответствует включение множеств (а равносильности соответствует равенство множеств).

Пример 2.4. Что означает тавтология транзитивности импликации

$$(A \to B) \wedge (B \to C) \to (A \to C) \equiv \mathsf{И}$$

на языке множеств? Заменим часть импликаций на включение, получим утверждение «если A — подмножество B и B — подмножество C, то A — подмножество C», истинное для всех множеств.

Другой способ понимать эту тавтологию менее интересный: для любых подмножеств A, B, C множества U выполняется равенство

$$\overline{(\bar{A} \cup B) \cap (\bar{B} \cup C)} \cup (\bar{A} \cup C) = U.$$

Конечно, это равенство можно проверить, исходя из определений или на кругах Венна. □

2.3. Индукция–1

Одним из важнейших способов доказательства в дискретной математике являются доказательства по индукции. Это способ доказать *бесконечно много утверждений* одним рассуждением.

Рассмотрим простой пример.

Пример 2.5. Пусть нужно вычислить $1+2+\cdots+1000$. Неужели придётся выполнить 999 арифметических операций? Не обязательно, достаточно трёх операций:

$$1 + 2 + \cdots + 1000 = \frac{1000 \cdot (1000 + 1)}{2}\,.$$

Как убедиться, что правая часть даёт искомую сумму? Способов много. Все они так или иначе дают доказательство общей формулы

$$1 + 2 + \cdots + n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}\,. \tag{2.5}$$

Мы сейчас будем рассуждать так. Заметим, что формула (2.5) выполняется для первых нескольких сумм:

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

Теперь докажем *условное* утверждение: если формула (2.5) выполняется для n, то она выполняется и для $n+1$. Для этого делаем вычисления, как обычно в элементарной алгебре:

$$1 + 2 + \cdots + n + (n+1) = \Big(1 + 2 + \cdots + n\Big) + (n+1) =$$
$$= \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}\,.$$

При переходе ко второй строчке мы используем формулу (2.5), которая по условию верна. В результате этого вычисления получаем, что формула (2.5) верна и для $n+1$.

И это всё! Завершает доказательство магическая фраза «по принципу математической индукции формула (2.5) верна для всех n». □

Обсудим суть этого принципа подробнее. Он основан на двух простых фактах. Во-первых, мы в прошлый раз обсуждали транзитивность импликации. Поэтому из сколь угодно длинной цепочки логических равенств

$$A_1 = \mathsf{И},$$
$$A_1 \to A_2 = \mathsf{И},$$
$$\cdots$$
$$A_n \to A_{n+1} = \mathsf{И}$$

следует, что $A_{n+1} = \mathsf{И}$.

Во-вторых, у целых положительных чисел есть такое свойство: прибавлением 1 можно получить любое целое положительное число. Отсюда и получается тот самый

> **Принцип математической индукции**. Пусть для последовательности утверждений
>
> $$A_1, A_2, A_3, \ldots, A_n, \ldots,$$
>
> занумерованных целыми положительными числами, верны утверждения
>
> **База индукции:** A_1 истинно.
>
> **Шаг индукции:** $A_n \to A_{n+1}$ истинно для любого n. Посылку импликации A_n называют *индуктивным предположением*.
>
> Тогда A_n истинно для любого n.

В примере 2.5 базой индукции является равенство $1 = 1 \cdot 2/2$, а шагом индукции — утверждение «для любого n, если (2.5) верна для n, то (2.5) верна и для $n+1$». Мы доказали оба эти утверждения. Значит, (2.5) выполняется для всех n.

Приведём ещё один простой пример доказательства по индукции.

Пример 2.6. Докажем, что для любых $q \neq 1$ и целого положительного n выполняется равенство

$$1 + q + q^2 + \cdots + q^n = \sum_{i=0}^{n} q^i = \frac{q^{n+1} - 1}{q - 1}\,. \tag{2.6}$$

База индукции (здесь первое число равно 0, так бывает и ничего не меняет в рассуждении, кроме номеров утверждений):

$$1 = \frac{q^{0+1} - 1}{q - 1}\,.$$

Тут и доказывать нечего, конечно, это верное равенство.

Индуктивный переход: «если (2.6) верно для n, то оно верно для $n+1$». Доказательство вычислением

$$\begin{aligned} &1 + q + \cdots + q^n + q^{n+1} = \left(1 + q + \cdots + q^n\right) + q^{n+1} = \\ =&\frac{q^{n+1} - 1}{q - 1} + q^{n+1} = \frac{q^{n+1} - 1 + q^{n+1}(q-1)}{q-1} = \\ =&\frac{q^{n+1} - 1 + q^{n+2} - q^{n+1}}{q-1} = \frac{q^{n+2} - 1}{q-1}\,. \end{aligned}$$

По принципу математической индукции формула (2.6) выполняется для всех целых неотрицательных n. □

Ещё один пример применения индукции: доказательство «принципа кроликов» или, как было раньше принято говорить в русской литературе, «принципа Дирихле».

Пример 2.7 (Принцип кроликов)**.** Если $k > n$ и k кроликов рассажены по n клеткам, то хотя бы в одной клетке сидит хотя бы два кролика.

Давайте запишем это утверждение в числах. Занумеруем клетки и пусть в клетку с номером i посажено r_i кроликов. Тогда принцип формулируется так: если $k > n$ и $r_1 + \cdots + r_n = k$, то для какого-то i выполняется неравенство $r_i > 1$.

Доказываем индукцией по n.

База: если $k > 1$ и $r_1 = k$, то $r_1 > 1$. Очевидно выполняется.

Индуктивный переход: пусть принцип кроликов верен для n. Докажем его для $n+1$. Пусть $k > n+1$ и $r_1 + \cdots + r_n + r_{n+1} = k$. Доказательство разбором случаев.

Если $r_{n+1} > 1$, то всё доказано: искомая клетка имеет номер $n+1$.

Если $r_{n+1} \leqslant 1$, то $r_1 + \cdots + r_n = k - r_{n+1} \geqslant k - 1 > n$. Значит, по принципу кроликов для n клеток найдётся такое i, что $r_i > 1$. □

Замечание 2.1. Принцип кроликов можно доказать иначе. Вспомним, что в прошлый раз мы доказали, что среднее не больше максимума. То есть из неравенства $r_1 + \cdots + r_n = k > n$ следует требуемое: для какого-то i выполняется $r_i > 1$.

У рассуждений, основанных на индукции, есть много вариантов. Вот поучительный пример.

Пример 2.8 (Числа Фибоначчи)**.** Рассмотрим *последовательность Фибоначчи*

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34, \ldots,$$

в которой первые два числа равны единице, а каждое следующее равно сумме двух предыдущих. Обычный способ задания таких *рекуррентных* последовательностей выглядит так: $F_0 = 1$; $F_1 = 1$; $F_{n+2} = F_{n+1} + F_n$ для всех $n \geqslant 2$.

Докажем для чисел Фибоначчи формулу:

$$F_n = \frac{\varphi^{n+1} - \psi^{n+1}}{\sqrt{5}}, \quad \text{где} \quad \varphi = \frac{1+\sqrt{5}}{2}, \quad \psi = \frac{1-\sqrt{5}}{2}. \tag{2.7}$$

Эта формула верна для $n = 0$:

$$1 = F_0 = \frac{\varphi^{0+1} - \psi^{0+1}}{\sqrt{5}} = \frac{(1+\sqrt{5}) - (1-\sqrt{5})}{2\sqrt{5}}\,.$$

Верна она и для $n = 1$:

$$1 = F_1 = \frac{\varphi^{0+2} - \psi^{0+2}}{\sqrt{5}} = \frac{(\varphi-\psi)(\varphi+\psi)}{\sqrt{5}} = \frac{\sqrt{5}\cdot 1}{\sqrt{5}}\,.$$

Как доказать формулу (2.7) для всех n? Давайте для любого значения n докажем такое условное утверждение: если (2.7) верна для n и для $n+1$, то (2.7) верна для $n+2$.

Для этого заметим, что φ, ψ — корни квадратного уравнения $x^2 - x - 1 = 0$, то есть $\varphi^2 = \varphi + 1$ и $\psi^2 = \psi + 1$. Используя это наблюдение, получаем:

$$\begin{aligned} F_{n+2} = F_{n+1} + F_n &= \frac{\varphi^{n+2} - \psi^{n+2}}{\sqrt{5}} + \frac{\varphi^{n+1} - \psi^{n+1}}{\sqrt{5}} = \\ &= \frac{1}{\sqrt{5}}\Big(\varphi^{n+2} + \varphi^{n+1} - \psi^{n+2} - \psi^{n+1}\Big) = \frac{1}{\sqrt{5}}\Big(\varphi^{n+1}(\varphi+1) - \psi^{n+1}(\psi+1)\Big) = \\ &= \frac{1}{\sqrt{5}}\big(\varphi^{n+3} - \psi^{n+3}\big). \end{aligned}$$

Как теперь закончить рассуждение? Буквально принцип математической индукции не говорит о такой ситуации. Чтобы закончить рассуждение, обозначим через A_n высказывание «(2.7) верна для n» и рассмотрим серию составных высказываний $B_n = A_n \wedge A_{n+1} \to A_{n+2}$. Мы доказали истинность B_1 и для любого n истинность импликации $B_n \to A_{n+2}$. Но последняя равносильна импликации $B_n \to B_{n+1}$ (проверьте!). Значит, по принципу математической индукции все B_n истинные. Но тогда и все A_n истинные. То есть формула (2.7) верна для всех n. □

Мы будем использовать более общие формы индукции.

Принцип полной математической индукции. Пусть для последовательности утверждений

$$A_1, A_2, A_3, \ldots, A_n, \ldots,$$

занумерованных целыми положительными числами, истинны утверждения

База индукции: A_1 истинно.

Шаг индукции: $(A_1 \wedge \ldots \wedge A_n) \to A_{n+1}$ истинно для любого n.

Тогда A_n истинно для любого n.

Разница с обычным принципом математической индукции в том, что на шаге индукции предполагается, что все предыдущие утверждения истинны, а не только самое последнее.

Справедливость принципа полной математической индукции вытекает из справедливости обычного принципа математической индукции. Аккуратное доказательство этого факта оставляется в качестве упражнения. Удобно при этом использовать ещё одну равносильную формулировку принципа математической индукции.

Принцип наименьшего числа. Любое непустое подмножество натуральных чисел содержит наименьший элемент.

Лекция 3

Комбинаторика–1

В прошлый раз мы доказали формулу (2.5) для суммы первых n целых положительных чисел. Но мы не обсуждали, откуда эта формула берётся. Это общая проблема с доказательствами по индукции: если уже есть удачное утверждение, то его доказательство по индукции обычно не очень трудно провести. Но как догадаться до правильного утверждения?

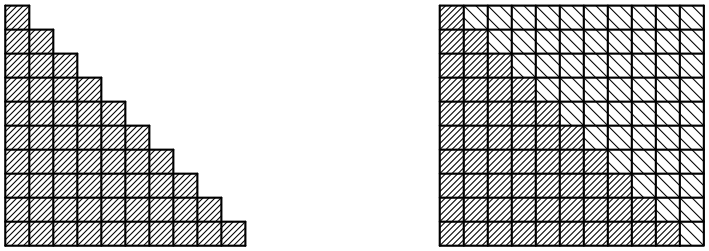
В случае формулы (2.5) есть несколько способов догадаться.

Группировка. Давайте запишем сумму двумя способами — как обычно и задом наперёд, — располагая слагаемые друг под другом:

$$\begin{array}{lll} 1+2 & + \cdots + & n \\ n+(n-1) & + \cdots + & 1 \end{array}$$

Видно, что соответственные слагаемые всегда дают в сумме $n+1$, а всего таких слагаемых n. Значит, удвоенное значение суммы равно $n(n+1)$.

Картинка. Есть наглядный способ нарисовать это рассуждение. Сумму $1+2+\cdots+n$ нарисуем в виде рядов квадратиков, поставленных один на другой. Получается «зубчатый треугольник», см. Рис. 3.1(a). Если приложить к этому треугольнику точно такой же, но повернутый треугольник, получится прямоугольник со сторонами n и $n+1$. В нём $n(n+1)$ клеточек и это ровно в два раза больше количества клеточек в треугольнике.



(a) (b)

Рис. 3.1: Дополнение треугольника до прямоугольника

Это простейший пример подсчётов, которыми занимается *перечислительная комбинаторика*. Мы сегодня разберём основные правила таких подсчётов и простые примеры их использования.

3.1. Правило произведения. Декартово произведение множеств

В примере выше использовано одно из главных правил перечислительной комбинаторики: *правило произведения*. Оно по сути говорит, что в клетчатом прямоугольнике со сторонами n и k ровно nk клеточек.

Во многих случаях нужно ещё увидеть прямоугольник среди данных задачи. Поэтому удобнее сформулировать это правило более общим образом. Для такой формулировки нам потребуется новая операция с множествами.

Декартово произведение множеств. Обозначение $A \times B$. Это множество, состоящее в точности из всех упорядоченных пар (a, b), где $a \in A$; $b \in B$.

Если множества конечны, то декартово произведение можно нарисовать в виде прямоугольника: столбцы — элементы A, строки — элементы B, на пересечении столбца a и строки b расположена пара $(a, b) \in A \times B$.

Отсюда получаем формулировку для правила произведения. Договоримся обозначать через $|A|$ количество элементов в множестве A, обычно оно называется *мощностью множества* (иногда — размером множества).

> **Правило произведения**. Для конечных множеств A, B выполняется равенство $|A \times B| = |A| \cdot |B|$.

Доказать правило произведения можно индукцией по мощности A.

База: $|A| = 0$, то есть $A = \varnothing$. По определению $\varnothing \times B = \varnothing$ (нет ни одной пары, первый член которой был бы элементом пустого множества). Поэтому $|\varnothing \times B| = |\varnothing| = 0 = |\varnothing| \cdot |B|$ — верное равенство.

Шаг индукции: пусть правило произведения справедливо для множеств мощности n. Рассмотрим пару A, B, где $|A| = n + 1$. Выделим в A какой-нибудь элемент a_{n+1} и обозначим через $A' = A \setminus \{a_{n+1}\}$ множество остальных элементов. Из каких пар состоит декартово произведение $A \times B$? Это либо пары вида (a, b), $a \in A'$, $b \in B$, либо пары (a_{n+1}, b), $b \in B$. Первых по предположению индукции $n \cdot |B|$ штук, а вторых — $|B|$ штук. Значит, всего в $A \times B$ входит $n \cdot |B| + |B| = (n+1) \cdot |B| = |A| \cdot |B|$ пар.

Шаг индукции доказан. По принципу математической индукции правило произведения выполняется для всех (конечных) множеств A, B.

3.2. Последовательности

В определении декартова произведения множеств появился новый объект, который будет постоянно возникать как в этом курсе, так и в других математических курсах.

Что такое упорядоченная пара? Это *последовательность* длины 2. Последовательности отличаются от множеств тем, что они образуют список своих элементов (ещё говорят, *членов* последовательности). Последовательности обозначают либо перечислением через запятую, либо, для наглядности, окружают их круглыми скобками. А иногда и попросту записывают элементы подряд, если можно отличить один от другого. Положение в списке важно: последовательности совпадают в том и только том случае, когда их длины (количество членов) равны и на каждом месте в обеих последовательностях стоит один и тот же элемент.

Пример 3.1. $(1,2) \neq (2,1)$; $(1,1,1) \neq (1,1)$. Последний пример показывает, что в последовательности элементы могут повторяться.

Важнейший для дискретной математики пример последовательностей: *слова в алфавите* A. Это произвольные последовательности конечной длины, члены которых принадлежат множеству A.

Всё множество слов в алфавите A обозначается A^*, а множество слов длины n обозначается A^n. В любом алфавите есть одно особое слово длины 0 — *пустое слово*, оно обычно обозначается ε. Это последовательность, в которой нет ни одного члена.

Теорема 3.1. *Количество слов длины n в конечном алфавите A равно $|A|^n$.*

Доказательство. Индукция по n.

База: $n = 0$. Пустое слово одно, $|A^0| = 1 = |A|^0$.

Шаг индукции. Пусть уже доказано, что $|A^n| = |A|^n$.

Рассмотрим слова длины $n+1$ и докажем, что они по сути совпадают с декартовым произведением $A \times A^n$.

Для любого слова α однозначно определены «голова» — первый член α_1 и «хвост» — остальные члены последовательности, они образуют последовательность длины на единицу меньше. Скажем, у слова 1012 головой является 1, а хвостом — 012.

И наоборот: по символу $a \in A$ и слову α однозначно получается слово $a\alpha$, длина которого на единицу больше: запишем слово α после символа a, то есть составим последовательность, первый член которой равен a, второй равен первому члену слова α и т.д.

Значит, мы можем представить слова длины $n+1$ как клеточки прямоугольника, строкам которого отвечают символы алфавита, а столбцам — слова длины n. Поэтому $|A^{n+1}| = |A \times A^n|$.

Применяя правило произведения и индуктивное предположение, получаем

$$|A^{n+1}| = |A| \cdot |A^n| = |A| \cdot |A|^n = |A|^{n+1},$$

что и требовалось доказать для шага индукции.

По принципу математической индукции количество слов длины n равно $|A|^n$ для любого n. □

3.3. Правило суммы

Всем ясно, что если в кошельке лежит 5 пятирублёвых монет и 6 десятирублёвых (и больше ничего), то всего в кошельке 11 монет.

Столь же ясно, что если в группе студентов 20 человек ездят на электричке и 20 человек подключены к Теле2, то это не означает, что в группе по крайней мере 40 человек. Может быть намного меньше.

Различать эти два случая при подсчётах важно. Сформулируем первый из них в виде правила.

> **Правило суммы**. Для конечных *непересекающихся* множеств A, B (то есть $A \cap B = \varnothing$) выполняется равенство $|A \cup B| = |A| + |B|$.

Заметим, что это правило мы неявно применили в доказательстве правила произведения, но не упомянули об этом.

Можно ли *доказать* правило суммы или нужно признать его исходным постулатом (аксиомой, как ещё говорят)? По существу ясно, что доказывать тут нечего: правило суммы фактически говорит о том, как *определяется* сложение целых неотрицательных чисел.

Правило суммы позволяет разделять подсчёт на случаи. Как и в доказательствах разбором случаев важно, чтобы все варианты были учтены. Но теперь не менее важно, чтобы каждый вариант учитывался ровно один раз.

3.4. Монотонные пути по прямой

Рассмотрим такую общую ситуацию. Есть клетчатая лента, по которой можно двигать фишку. Клетки пронумерованы целыми числами. В начале фишка находится в клетке 0. Далее её можно сдвигать вправо. Нужно подсчитать, сколько есть различных способов попасть в клетку с номером n.

Ответ зависит от того, какие ходы разрешены. Например, если разрешено сдвигать фишку только на одну клетку, то способ единственный. Для каждого набора разрешённых ходов получается своя задача подсчёта. Такие задачи в большинстве своём решаются однотипно. Правило суммы приводит к рекуррентному соотношению. Если угадать ответ, то доказать его можно методом математической индукции.

Пример 3.2. Пусть разрешены ходы на одну и на две клетки.

Задачу можно переформулировать на языке последовательностей целых чисел. А именно, нужно подсчитать количество монотонно возрастающих последовательностей целых чисел, первый член которых равен 0, последний равен n, а разность между двумя соседними принимает только значения 1 или 2. Действительно, такие последовательности — это протоколы движения фишки по клеточкам. Каждому способу движения отвечает ровно одна последовательность и по ней этот способ движения так же однозначно определяется.

Обозначим количество таких последовательностей F_n. Заметим, что $F_{n+2} = F_{n+1} + F_n$. Действительно, все последовательности, заканчивающиеся на $n+2$, разделяются на две непересекающиеся группы:

$$0, \ldots, n, n+2;$$
$$0, \ldots, n+1, n+2$$

(в клетку $n+2$ можно попасть либо с клетки n, либо с клетки $n+1$, на месте многоточий возможно вставить любую последовательность чисел, в которой разности между соседними числами равны 1 или 2).

Нетрудно подсчитать количество таких последовательностей при небольших n.

$\boldsymbol{n = 0}$**.** Единственная монотонная последовательность, начинающаяся и заканчивающаяся на 0: это 0. Поэтому $F_0 = 1$.

$\boldsymbol{n = 1}$**.** Единственная монотонная последовательность, начинающаяся на 0 и заканчивающаяся на 1: это 0,1. Поэтому $F_1 = 1$.

$\boldsymbol{n = 2}$**.** Монотонных последовательностей, начинающихся на 0 и заканчивающаяся на 2 уже две: это 0,1, 2 и 0, 2. Поэтому $F_2 = 2$.

Теперь становится понятно обозначение: мы получили рекурренту и начальные условия для чисел Фибоначчи. Для любого n количество способов попасть в n ходами длины 1 или 2 равно F_n.

Пример 3.3. Пусть разрешены ходы на любое количество клеток.

На языке последовательностей целых чисел мы хотим найти количество всех монотонно возрастающих последовательностей целых чисел, первый член которых равен 0, а последний равен n. Обозначим это количество $T(n)$.

Количество таких последовательностей при $n \leqslant 2$ такое же, как и в предыдущем примере: $T(0) = 1$; $T(1) = 1$; $T(2) = 2$.[1)] (Ограничения на длину шага выполняются в этих случаях для любой последовательности.)

При росте n количество вариантов растёт и уже легко ошибиться в подсчёте. Однако можно доказать общий факт:

$$T(n) = T(n-1) + T(n-2) + \cdots + T(0) \tag{3.1}$$

для любого n.

Обозначим множество всех последовательностей длины n через X. Разделим последовательности на группы, в зависимости от последнего хода. То есть группу X_i образуют те монотонные последовательности, которые имеют вид

$$0, \ldots, i, n.$$

Ясно, что каждая последовательность попала ровно в одну группу и группы не пересекаются (смотрим на последний ход или на предпоследний член последовательности). По правилу суммы получаем

$$|X| = |X_0| + |X_2| + \cdots + |X_{n-1}|.$$

[1)] Спасибо Рите Аруновой за замеченную опечатку.

Но, с другой стороны, $|X_i| = T(i)$ (монотонные последовательности, начинающиеся в 0 и заканчивающиеся в i). Отсюда и получается формула (3.1).

Используя формулу (3.1), докажем индукцией по n явную формулу для $T(n)$:

$$T(n) = 2^{n-1} \qquad \text{при } n \geqslant 1. \tag{3.2}$$

База уже доказана.

Индуктивный переход. Если равенство $T(n) = 2^{n-1}$ верно, то

$$\begin{aligned} T(n+1) &= T(n) + T(n-1) + T(n-2) + \cdots + T(0) = \\ &= T(n) + T(n) = 2^{n-1} + 2^{n-1} = 2^n = \\ &= 2^{(n+1)-1}. \end{aligned}$$

В первом равенстве использована формула (3.1), которая верна для всех n.

3.5. Формула включений–исключений

Теперь рассмотрим общую ситуацию, в которой варианты не взаимно исключающие. Другими словами, мы хотим определить мощность объединения множеств. В общем случае она не определяется однозначно размерами самих множеств.

Для случая двух множеств ответ такой.

Утверждение 3.2. $|A \cup B| = |A| + |B| - |A \cap B|$.

Доказательство. Пересчитаем все элементы A и все элементы B. Получим $|A|+|B|$. Элементы $|A \cap B|$, и только они, подсчитаны дважды. Значит, $|A \cup B|$ меньше $|A|+|B|$ на $|A \cap B|$. □

Для трёх множеств формула сложнее.

Утверждение 3.3. $|A \cup B \cup C| = |A|+|B|+|C|-|A \cap B|-|A \cap B|-|B \cap C|+|A \cap B \cap C|$.

Доказательство. Несколько раз применим формулу для объединения двух множеств.

Представим $A \cup B \cup C$ как $(A \cup B) \cup C$, получим

$$|A \cup B \cup C| = |A \cup B| + |C| - |(A \cup B) \cap C|. \tag{3.3}$$

Воспользуемся дистрибутивностью объединения и пересечения (дизъюнкции и конъюнкции): $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. Применяем ещё раз формулу для объединения двух множеств:

$$|(A \cup B) \cap C| = |(A \cap C) \cup (B \cap C)| = |A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)|.$$

Легко увидеть, что $(A \cap C) \cap (B \cap C) = A \cap B \cap C$. Теперь подставляем полученное в (3.3) и получаем

$$|A \cup B \cup C| = |A \cup B| + |C| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Теперь применим ещё раз формулу для $|A \cup B|$ и получим искомое выражение. □

Формулы для двух и трёх множеств подсказывают общий вид формулы включений-исключений. Нужно взять сумму мощностей всех множеств. Некоторые элементы при этом посчитаны более одного раза. Поэтому нужно вычесть мощности попарных пересечений множеств, после чего некоторые элементы объединения вообще не будут посчитаны. Далее нужно последовательно прибавлять и вычитать мощности тройных, четверных и т.д. пересечений, включая их в итоговую сумму с чередующимися знаками.

Теорема 3.4 (Формула включений–исключений)**.**

$$\begin{aligned}|A_1 \cup A_2 \cup \ldots \cup A_n| = |A_1| + \ldots + |A_n| - \\ - |A_1 \cap A_2| - |A_1 \cap A_3| - \ldots \\ + |A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + \ldots \\ \ldots \\ + (-1)^{n+1}|A_1 \cap A_2 \cap \ldots \cap A_n|\end{aligned} \tag{3.4}$$

В первой строчке правой части равенства выписаны мощности всех множеств. Во второй — мощности всех попарных пересечений множеств (со знаком минус). Далее выписываем пересечения троек, четвёрок и т.д. множеств с чередующимися знаками.

Доказательство формулы включений-исключений. Индукция по количеству множеств. База: формула для двух множеств (да и для одного годится).

Шаг индукции в точности следует доказательству формулы для трёх множеств (и использует формулу для объединения двух множеств). Предполагаем, что (3.4) верна для любых n множеств. Рассмотрим набор из $n+1$ множеств: $A_1, \ldots, A_{n+1}$. Нам нужно доказать для этого набора равенство (3.4), что означает

$$\begin{aligned}|A_1 \cup A_2 \cup \ldots \cup A_n \cup A_{n+1}| = |A_1| + \ldots + |A_n| + |A_{n+1}| - \\ - |A_1 \cap A_2| - |A_1 \cap A_3| - \ldots - |A_n \cap A_{n+1}| \\ + |A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + \ldots \\ \ldots \\ (-1)^{n+2}|A_1 \cap A_2 \cap \ldots \cap A_n \cap A_{n+1}|.\end{aligned} \tag{3.5}$$

Запишем формулу для объединения двух множеств:

$$\begin{aligned}|A_1 \cup A_2 \cup \ldots \cup A_{n+1}| = |(A_1 \cup A_2 \cup \ldots \cup A_n) \cup A_{n+1}| = \\ = |(A_1 \cup A_2 \cup \ldots \cup A_n)| + |A_{n+1}| - |(A_1 \cup A_2 \cup \ldots \cup A_n) \cap A_{n+1}|.\end{aligned} \tag{3.6}$$

Применим формулу для объединения n множеств к первому слагаемому в (3.6). Получим те слагаемые из (3.5), в которые входят только множества $A_1, \ldots, A_n$.

Второе слагаемое в (3.6) равно $|A_{n+1}|$, такое же слагаемое есть и в (3.5).

Остальные слагаемые в (3.5) равны мощностям пересечения A_{n+1} с какими-то из множеств $A_1, \ldots, A_n$. Нужно убедиться, что третье слагаемое в (3.6) содержит те же слагаемые с теми же знаками.

В силу дистрибутивности пересечения и объединения множество в третьем слагаемом записывается как[2)]

$$(A_1 \cup A_2 \cup \ldots \cup A_n) \cap A_{n+1} = (A_1 \cap A_{n+1}) \cup (A_2 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1}).$$

Обозначим $B_i = A_i \cap A_{n+1}$. Тогда третье слагаемое в (3.6) равно $-|B_1 \cup \cdots \cup B_n|$. Применим индуктивное предположение (с учётом перемены знака) и получим

$$\begin{aligned} -|B_1 \cup B_2 \cup \ldots \cup B_n| = &-|B_1| - \ldots - |B_n| + \\ &+ |B_1 \cap B_2| + |B_1 \cap B_3| + \ldots + |B_{n-1} \cap B_n| - \\ &- |B_1 \cap B_2 \cap B_3| - |B_1 \cap B_2 \cap B_4| - \ldots \\ &\ldots \\ &- (-1)^{n+1} |B_1 \cap B_2 \cap \ldots \cap B_n|. \end{aligned} \tag{3.7}$$

Нам нужно ещё одно теоретико-множественное тождество

$$(A_1 \cap A_k) \cap (A_2 \cap A_k) \cap \cdots \cap (A_{k-1} \cap A_k) = A_1 \cap A_2 \cap \cdots \cap A_k,$$

которое легко проверяется раскрытием скобок и применением очевидного из определения равенства $A \cap A = A$.

В силу этого тождества в правой части (3.7) написаны пересечения каких-то множеств $A_1, \ldots, A_n$ и (обязательно) множества A_{n+1}. Это в точности те слагаемые, которых нам недоставало в равенстве (3.5).

Проверим, что они входят с правильными знаками.

Так как $B_{i_1} \cup \cdots \cup B_{i_s} = A_{i_1} \cup \cdots \cup A_{i_s} \cup A_{n+1}$, то мощность этого множества входит в (3.7) со знаком $-(-1)^{s+1}$, а в (3.5) — со знаком $(-1)^{(s+1)+1} = -(-1)^{s+1}$. Знаки совпадают. □

[2)] Спасибо Рите Аруновой за замеченные опечатки в формулах на этой странице.

Лекция 4

Графы–1

4.1. Определения

Графы — один из важнейших математических объектов для дискретной математики. Есть много разных видов графов. Мы начнём с *неориентированных графов*. Наглядно такой граф можно изобразить набором точек («вершин»), соединённых линиями («рёбрами»). При этом важно, какие точки соединены, а как именно нарисована линия, соединяющая точки, не имеет значения. Примеры графов показаны на рисунках 4.1–4.3.

Рис. 4.1: Путь P_5

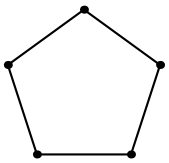


Рис. 4.2: Цикл C_5

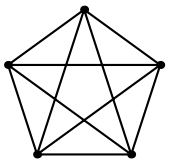


Рис. 4.3: Полный граф K_5

Наглядные картинки помогают понять рассуждения про графы, но для доказательств нужны более точные определения. Мы дадим сразу довольно много определений.

Простой неориентированный граф — это множество вершин V и множество рёбер E. Рёбрами являются 2-элементные подмножества множества V.

Ближайшие две лекции мы будем говорить только о простых неориентированных графах. Поэтому для краткости слова «простой неориентированный» будем пропускать.

Если $e = \{u, v\} \in E$, то вершины u, v называются *концами* ребра e. Концы ребра называются *смежными вершинами* или *соседями*.

Говорят также, что ребро $e = \{u, v\}$ *инцидентно* вершине u (как и вершине v). Ребра с общим концом также называются инцидентными.

Помимо картинок, с графами связаны две таблицы или, как принято говорить в математике, *матрицы*. Эти матрицы содержат только 0 или 1. Чтобы говорить о

матрице графа, нужно перенумеровать вершины и рёбра.

Матрица смежности графа: на пересечении i-й строки и j-го столбца стоит 1, если вершины i, j соседние; иначе там стоит 0.

Таким образом, матрица смежности — это квадратная матрица порядка n, где n — количество вершин графа.

Матрица инцидентности графа: на пересечении i-й строки и j-го столбца стоит 1, если вершина i инцидентна ребру j; иначе там стоит 0.

Таким образом, матрица инцидентности — это матрица размера $n \times m$, где n — количество вершин графа, m — количество рёбер графа.

4.2. Степени вершин

Количество соседей вершины v (оно же количество инцидентных её рёбер) называется *степенью* вершины, обозначать будем $d(v)$.

Теорема 4.1. *Сумма степеней всех вершин графа равна удвоенному числу его рёбер.*

Доказательство. Здесь применим *метод двойного подсчёта*. Под таким громким названием скрывается очень простой факт: если посчитать сумму элементов матрицы по строкам, то получится такое же число, как и при суммировании элементов матрицы по столбцам.

Давайте посчитаем количество 1 в матрице инцидентности графа. В строке i количество 1 равно количеству инцидентных вершине i рёбер, то есть степени этой вершины. Значит, сумма 1 по строкам равна сумме степеней вершин.

В каждом столбце матрицы инцидентности ровно две 1, так как у ребра ровно два конца. Значит, сумма 1 по столбцам равна удвоенному количеству рёбер.

Обе суммы равны общему количеству 1 в матрице инцидентности, а значит, равны между собой. □

Пример 4.1. Из этой теоремы получаются интересные следствия. Давайте докажем, что не существует графа, у которого 77 вершин, а степень каждой вершины равна 15.

Перебрать все возможные графы на 77 вершинах немыслимо.[1)] Однако это необязательно делать.

Доказательство от противного. Пусть такой граф существует. Заметим, что сумма степеней всех его вершин равна $77 \cdot 15$. Это нечётное число. С другой стороны, теорема утверждает, что эта сумма равна удвоенному количеству рёбер графа, то есть чётному числу. Противоречие.

Подсчёт сумм степеней вершин помогает и в более сложных ситуациях. Приведём пример.

Большие графы удобно задавать не матрицами и не рисунками, а точным определением множества вершин и рёбер.

[1)] Это множество конечно, но очень велико.

Пример 4.2 (Булев куб)**.** Вершины булева куба Q_n (булев куб размерности n) — двоичные слова длины n. Два слова u и v соседние в булевом кубе, если и только если одно можно получить из другого инвертированием бита ровно в одной позиции.

Скажем, 00010 и 01010 соседние в Q_5, а 00010 и 10000 — нет.

Как мы уже знаем, количество вершин в булевом кубе Q_n равно 2^n.

Степень каждой вершины равна n: есть ровно n позиций, инвертирование бита в каждой даёт соседа.

Подмножество I множества вершин графа называется *независимым*, если ни одна пара вершин в этом множестве не связана ребром.

Пример 4.3. В булевом кубе Q_n есть независимое множество размера 2^{n-1} (половина всех вершин).

Количество единиц в слове называется *(хэмминговым) весом* слова. При инвертировании бита в одной позиции количество единиц в слове изменяется ровно на 1. Поэтому концами каждого ребра являются слова, у которых чётность веса разная — у одного вес чётный, у другого — нечётный.

Значит, между словами чётного веса рёбер нет (как и между словами нечётного веса).

Как посчитать количество слов чётного и количество слов нечётного веса? Для этого опять применим двойной подсчёт. У каждого ребра один конец имеет чётный вес, а другой — нечётный. Значит, сумма степеней вершин чётного веса равна количеству рёбер и сумма степеней вершин нечётного веса равна количеству рёбер. Поскольку степени всех вершин булева куба одинаковы и равны n, получаем равенство $nV_{\text{odd}} = nV_{\text{even}}$. Поэтому $V_{\text{odd}} = V_{\text{even}} = 2^{n-1}$.

Пример 4.4. Докажем, что в булевом кубе Q_n нет независимого множества размера больше 2^{n-1}.

Обратите внимание, что предыдущего примера множеств вершин чётного и нечётного веса недостаточно. В независимое множество, вообще говоря, могут входить вершины с весами разной чётности.

Пусть I — независимое множество. Количество рёбер, инцидентных этому множеству, равно $n|I|$. Это число не больше общего количества рёбер $n2^{n-1}$. Отсюда и получаем неравенство $|I| \leqslant 2^{n-1}$.

4.3. Связность, компоненты связности

Путь по графу — это такая последовательность вершин $v_1, v_2, v_3, \ldots, v_k$, в которой стоящие рядом члены (вершины v_i и v_{i+1} при всех i) соединены ребром. Вершина v_1 называется *началом* пути, вершина v_k — его *концом*. *Длиной* пути называется число рёбер, то есть $k-1$. Мы будем разрешать также и пути длины 0, то есть последовательности из одной вершины. У такого пути начало совпадает с концом. Рёбер в таком пути нет, но вершина (одна) есть.

Вершины v и v' называются *связанными*, если существует путь с началом в v и концом v'. Граф называется *связным*, если любые две его вершины связаны.

Неформальное понимание связности графа: если представить вершины как города, а рёбра как дороги, то из любого города в любой можно проехать по дорогам.

Пример 4.5. Пример связного графа: *полный граф* K_n. В этом графе n вершин, и каждая пара вершин соединена ребром. Поэтому любая последовательность из двух вершин является путём в этом графе.

Пример 4.6. Пример несвязного графа: граф с n вершинами и 0 рёбер. В этом графе все пути имеют длину 0.

Если граф удачно нарисован, то увидеть его связность или несвязность легко.[2)] Но без картинки, только из определения или списка рёбер, это сделать уже труднее.

Пример 4.7. Вершины графа $Q_{n,3}$ — двоичные слова длины n. Два слова u и v соседние в этом графе, если и только если одно можно получить из другого инвертированием битов ровно в трёх позициях.

Связен ли граф $Q_{n,3}$? Ответ неочевиден и его обоснование требует некоторых усилий.

Для вершины v обозначим через $C(v)$ множество вершин, связанных с v. Будем называть это множество *областью достижимости вершины* v. Если граф связный, то для каждой вершины $C(v)$ совпадает со всем множеством вершин графа. В общем случае это не так. Но всегда выполняется несколько простых свойств для областей достижимости.

Лемма 4.2. *Для любого графа и любых его вершин v_1, v_2, v_3 выполняются следующие свойства:*

1. $v_1 \in C(v_1)$ *(вершина достижима из себя самой);*
2. $v_1 \in C(v_2)$ *равносильно* $v_2 \in C(v_1)$ *(если v_1 достижима из v_2, то v_2 достижима из v_1);*
3. *если* $v_1 \in C(v_2)$ *и* $v_2 \in C(v_3)$, *то* $v_1 \in C(v_3)$.

Доказательство. Эти свойства вполне очевидны из неформального представления о достижимости. Формальные их доказательства также очень просты.

v_1 — путь в любом графе, поэтому v_1 связанная с самой собой.

Если $v_1u_1 \dots u_sv_2$ — путь в графе, то $v_2u_s \dots u_1v_1$ — также путь. Поэтому достижимость v_2 из v_1 равносильная достижимости v_1 из v_2.

Наконец, если в графе есть пути $v_3u_1 \dots u_sv_2$ и $v_2w_1 \dots w_tv_1$ (то есть $v_1 \in C(v_2)$ и $v_2 \in C(v_3)$), то в это графе есть также и путь $v_3u_1 \dots u_sv_2w_1 \dots w_tv_1$, то есть v_1 достижима из v_3. □

[2)] Однако можно так нарисовать граф, что понять, связен ли он, трудно: это популярная разновидность головоломок.

На этой очевидной лемме основан важный способ доказательства связности графа. А именно, достаточно доказать, что из какой-то вершины достижимы все. Это эквивалентно в силу свойства 2 леммы 4.2, что из любой вершины достижима какая-то выделенная. А свойство 3 гарантирует, что года из каждой вершины достижима любая.

Утверждение 4.3. *Граф $Q_{n,3}$ связный при $n \geqslant 4$.*

Доказательство. Докажем, что любое слово длины $\geqslant 4$ последовательными инвертированиями битов в тройках позиций можно превратить в нулевое.

Первый шаг: если единиц в слове больше 2, уменьшаем их количество, инвертируя какие-нибудь три позиции, содержащие только единицы. Этот шаг заканчивается, когда в слове не больше 2 единиц.

Если единиц нет, то мы достигли требуемого — построили путь из начального слова в нулевое.

Если единица одна, то инвертируя позицию, в которой она стоит, и ещё две позиции (с нулями), получаем слово с двумя единицами ровно.

Если единиц две, то инвертируем позицию одной из них и ещё две позиции, содержащие нули (длина слова не меньше 4, так что позиций с нулями хотя бы две). Получаем слово ровно с тремя единицами. Инвертируя их, получаем нулевое слово. □

Лемма 4.4. *Если $w \in C(v_1) \cap C(v_2)$, то $C(v_1) = C(v_2)$. Области достижимости не пересекаются или совпадают.*

Доказательство. Поскольку w достижима из v_1 (определение), а v_2 достижима из w (свойство 2 леммы 4.2), то v_2 достижима из v_1 (свойство 3 леммы 4.2). Значит, и v_1 достижима из v_2 (свойство 2).

Пусть $x \in C(v_1)$. Тогда $x \in C(v_2)$, свидетельством тому путь из v_2 в v_1, продолженный путём из v_1 в x. Значит, $C(v_1) \subseteq C(v_2)$.

Верно и обратное: $C(v_2) \subseteq C(v_1)$ (поменяем индексы в рассуждении). Значит, $C(v_1) = C(v_2)$. □

Из свойства 1 леммы 4.2 следует

$$V = \bigcup_{v \in V} C(v).$$

Лемма 4.4 говорит, что области достижимости не пересекаются или совпадают. Таким образом, мы получаем *разбиение* множества вершин графа на подмножества. Каждое такое подмножество может быть представлено как область достижимости любого элемента из этого множества.

Удобно не упоминать этот случайно выбранный элемент. Поэтому используется такая терминология. *Компонента связности* графа — это область достижимости некоторой вершины этого графа. У связного графа одна компонента связности. Как мы обсудили выше, вершины графа разбиваются на компоненты связности.

У компонент связности есть другое описание. Для него нам потребуется новое понятие: *индуцированный подграф* графа. Пусть $U \subseteq V$ — подмножество вершин графа $G(V, E)$. Тогда $\langle U \rangle$ — это граф с множеством вершин U и множеством рёбер

$$E(\langle U \rangle) = \big\{\{x, y\} : \{x, y\} \in E(G),\ x, y \in U\big\}.$$

Теорема 4.5. *Компоненты связности — это в точности максимальные по включению множества вершин, индуцирующие связный граф.*

Условие теоремы означает фактически два утверждения: (а) любое множество, строго содержащее компоненту связности (все вершины компоненты и ещё какие-то, хотя бы одну), индуцирует несвязный граф; (б) любое множество, индуцирующее связный граф, содержится в какой-то компоненте связности.

Доказательство. Если $X \supset C(v)$ (этот знак обозначает строгое включение), то $\langle X \rangle$ несвязный: из вершины $x \in X \setminus C(v)$ недостижима v. Значит, компоненты связности — максимальные по включению множества, индуцирующие связные графы. Утверждение (а) доказано.

В обратную сторону. Пусть X — множество, индуцирующее связный граф, $v \in X$. Поскольку $\langle X \rangle$ связный, то $C(v) \supseteq X$ (вершины из X достижимы из v даже если запрещено выходить за пределы X). Утверждение (б) доказано. □

Лекция 5

Графы–2. Деревья

Мы охарактеризовали компоненты связности как максимальные по включению подмножества вершин графа, которые индуцируют связный граф.

Ясно, что добавление рёбер к связному графу сохраняет связность: путей становится только больше. И наоборот, выбрасывание рёбер из несвязного графа сохраняет несвязность: путей становится только меньше.

Сегодня рассмотрим *минимальные по включению рёбер* связные графы. Они называются *деревьями*. Примеры деревьев изображены на рисунке 5.1.

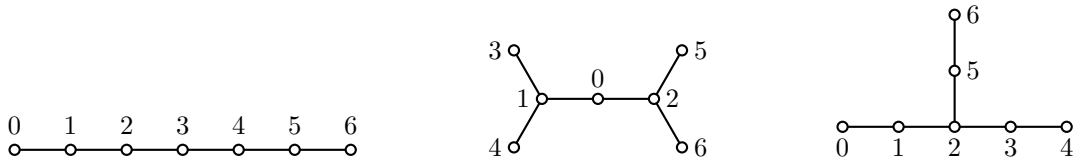


Рис. 5.1: Примеры деревьев

Точное определение дерева такое.

Определение 5.1. *Дерево* — такой связный граф, что выбрасывание любого ребра даёт несвязный граф.

Можно переформулировать это определение, используя понятие *моста*. Мост — это такое ребро в графе, что его удаление увеличивает количество компонент связности. Поэтому деревья — это связные графы, каждое ребро которых мост. А произвольные графы, у которых каждое ребро является мостом, называются *лесами*.

5.1. Простые пути и циклы

Глядя на рисунок 5.1, хочется сформулировать такое свойство деревьев: в них нет циклов. Однако мы ещё не определили, что такое цикл. К сожалению, в теории графов есть много путаницы в терминологии. В частности, слово «цикл» в разных

книгах может означать три разных понятия. Единственный выход: помнить об этом и внимательно следить за определениями, которые даются в книжках.

Мы называем *циклом* замкнутый путь по графу, то есть такой путь, у которого начало совпадает с концом. При таком определении сформулированное выше свойство неверно. Например, в левом графе на рисунке 5.1 есть цикл

$$(0, 1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1, 0).$$

То понятие, которое нужно для описания свойств деревьев, мы будем называть *простым циклом*. Это такой цикл, в котором все вершины различны (за исключением начала и конца, которые совпадают по определению цикла).

Однако и простые циклы бывают во всех графах, которые содержат хотя бы одно ребро. Если в графе G есть ребро $\{u, v\}$, то в этом графе есть цикл (u, v, u).

Мы докажем чуть позже, что в деревьях нет простых циклов длины больше 2.

Нам также будет нужно понятие *простого пути*: это такой путь, в котором все вершины различны. Когда мы говорим о связанности вершин, достаточно рассматривать только простые пути. Неформально это означает очень простое наблюдение: если вы едете из города A в город B, то можно так проехать, чтобы побывать в каждом промежуточном городе ровно один раз.

Утверждение 5.2. *Если две вершины x, y связаны в графе G, то в этом графе существует простой путь с началом x и концом y.*

Доказательство. Мы будем использовать принцип наименьшего числа. В соответствии с этим принципом если есть хотя бы один путь с началом x и концом y, то есть и путь наименьшей длины (нет ни одного пути короче).

Рассмотрим такой кратчайший путь $x = v_1, \ldots, v_k = y$ и докажем, что он простой. Для этого применим контрапозицию. Докажем, что если путь $x = u_1, \ldots, u_t = y$ не простой, то он не кратчайший (есть путь короче). Действительно, пусть $u_i = u_j$, $i < j$. Тогда последовательность $x = u_1, \ldots, u_i, u_{j+1}, u_t = y$ также является путём с началом x и концом y, а длина этого пути меньше. □

Замечание 5.1. В доказательстве есть одно тонкое место. Что если $j = t$? Тогда u_{j+1} не определена. В этом случае более короткий путь имеет вид $x = u_1, \ldots, u_i = u_t = y$.

5.2. Мосты, простые пути и простые циклы

Для деревьев и лесов есть несколько критериев (свойств, равносильных определению дерева). Два из них связаны с простыми циклами и простыми путями в графе.

Теорема 5.3. *Равносильны следующие свойства простых неориентированных графов:*

(1) *каждое ребро — мост;*

(2) *для любых связанных вершин u, v существует единственный простой путь из u в v;*

(3) *нет простых циклов длины больше* 2.

Свойства для деревьев получатся из этой теоремы, если потребовать связности графа.

Следствие 5.4. *Равносильны следующие свойства связных простых неориентированных графов:*

(1) *граф — дерево;*

(2) *для любых двух вершин u, v существует единственный простой путь из u в v;*

(3) *нет простых циклов длины больше* 2.

Будем доказывать утверждения теоремы 5.3 по очереди.

Доказательство (2) $\Rightarrow$ (3). Равносильно контрапозиции $\neg(3) \Rightarrow \neg(2)$. Пусть в графе G есть простой цикл $v_0, v_1, \ldots, v_\ell = v_0$, $\ell > 2$.

Вершины v_0 и v_1 связанные в этом графе, причём есть по крайней мере два разных пути с концами в этих вершинах: (v_0, v_1) (путь из одного ребра) и путь по остальным рёбрам цикла $(v_0 = v_\ell, v_{\ell-1}, \ldots, v_2, v_1)$ (здесь важно, что длина цикла больше 2). □

Доказательство (3) $\Rightarrow$ (1). Равносильно контрапозиции $\neg(1) \Rightarrow \neg(3)$.

Пусть ребро $e = \{v_0, v_1\}$ можно удалить из графа G и полученный граф $G - e$ остаётся связным. Это значит, что вершины v_0, v_1 связанные в G'. По утверждению 5.2 в графе G' есть простой путь $v_1, v_2, \ldots, v_\ell = v_0$. Все вершины этого пути различные.

Но тогда в графе G есть простой цикл $v_0, v_1, v_2, \ldots, v_\ell = v_0$ и, так как $v_0 \neq v_1$, длины этого цикла больше 2. □

Доказательство (1) $\Rightarrow$ (2). Равносильно контрапозиции $\neg(2) \Rightarrow \neg(1)$. Пусть между вершинами u и v есть два разных пути

$$u = u_0 u_1 \ldots u_s = v \quad \text{и} \quad u = v_0 v_1 \ldots v_t = v.$$

Начинаются эти пути в одной вершине, но полностью совпадать не могут. Выделим ребро, которое входит только в один из этих путей. Без ограничения общности это ребро $\{u_i, u_{i+1}\}$ на первом пути.

Докажем, что ребро $\{u_i, u_{i+1}\}$ — не мост. При удалении этого ребра из графа вершины u_i, u_{i+1} остаются в одной компоненте связности: они связаны (не обязательно простым) путём

$$u_i u_{i-1} \ldots u_0 = v_0 v_1 \ldots v_t = u_s u_{s-1} \ldots u_{i+1}.$$

Области достижимости вершин не из $C(u_i)$ не изменяются: пути из таких вершин не проходят через ребро $\{u_i, u_{i+1}\}$. □

Завершение доказательства теоремы 5.3. Поскольку мы доказали циклическую цепочку импликаций (2) ⇒ (1) ⇒ (3) ⇒ (2), все эти утверждения равносильны (если хотя бы одно истинно, остальные два тоже истинны). □

5.3. Размерность графа

Обозначим количество вершина графа G через n, количество рёбер через m, а количество компонент связности через c. *Размерностью* графа называется величина $\dim G = m - n + c$. Это и впрямь размерность некоторого векторного пространства. Мы ограничимся лишь комбинаторными свойствами этой размерности.

Теорема 5.5. *Графы, у которых размерность равна* 0, *— это в точности леса, то есть графы без мостов.*

Отсюда получаем ещё один критерий дерева:

Следствие 5.6. *Связный граф является деревом тогда и только тогда, когда число рёбер в нём на единицу меньше числа вершин.*

Доказательство теоремы основано на таком утверждении.

Утверждение 5.7. *Пусть граф $G' = G + e$ получается из графа G добавлением ребра $e = \{x, y\}$ к множеству рёбер, а вершины у него те же.*

Тогда $\dim G' = \dim G$, если концы ребра x, y лежат в разных компонентах связности графа G, и $\dim G' = \dim G + 1$, если x, y лежат в одной компоненте связности графа G.

Доказательство. Рассмотрим два случая, указанных в утверждении.

Вершины x, y лежат в одной компоненте связности C графа G. Тогда количество компонент связности не изменилось: для любой вершины x область достижимости в графе G' та же самая, что и в G (поскольку y достижима из x и в графе G). Количество рёбер увеличилось на 1, количество вершин не изменилось. Значит, и размерность увеличилась на 1.

Вершины x, y лежат в разных компонентах связности графа G. Тогда в графе G' в область достижимости вершины x добавляется $C(y)$, поскольку в G' вершина y достижима из x. Аналогично рассуждаем про y, получаем

$$C'(x) = C'(y) = C(x) \cup C(y),^{1)}$$

то есть области достижимости x и y в графе G' равны объединению областей достижимости этих вершин в графе G. Области достижимости вершин из других компонент связности G остаются теми же самыми. Значит, количество компонент связности уменьшилось на 1. Количество рёбер увеличилось на 1, количество вершин не изменилось. Поэтому размерность не изменилась. □

[1)] Спасибо Яну Максимову за замеченную опечатку.

Для доказательства теоремы 5.5 мы используем индукцию по числу рёбер графа. Базой индукции будут графы без рёбер (с произвольным количеством вершин). В таком графе размерность равна нулю: рёбер нет, каждая вершина является компонентой связности. И такой граф является лесом, так как каждое его ребро — мост (рёбер вообще нет, так что это утверждение верно).

Лемма 5.8. *Размерность графа неотрицательная.*

Доказательство. Индукция по количеству рёбер. База проверена выше.

Пусть для графа G размерность неотрицательная, то есть $m-n+c \geqslant 0$. Рассмотрим граф $G' = G + e$, $e = \{x, y\}$. По утверждению 5.7 его размерность не меньше размерности графа G, то есть неотрицательная. Шаг индукции доказан, лемма выполняется в силу принципа математической индукции. □

Доказательство теоремы 5.5. Пусть размерность графа G равна 0. Поскольку размерность любого графа неотрицательная, каждое ребро G обязано быть мостом, так как удаление не моста уменьшает размерность.

В другую сторону нужно доказать, что размерность всякого леса равна 0. Доказываем индукцией по количеству рёбер. Базой, как уже говорилось, являются графы без рёбер.

Шаг индукции. Предположим, что утверждение теоремы выполняется для графов с количеством рёбер меньше m. Рассмотрим лес G с m рёбрами, и граф $G' = G - e$, полученный из G удалением ребра $e = \{x, y\}$. Это ребро является мостом, так что размерность графа G' равна размерности графа G (утверждение 5.7).

В графе G' нет простых циклов длины больше 2, так как любой такой цикл был бы и простым циклом в лесу G. Значит, по теореме 5.3, граф G' — лес. По индуктивному предположению, его размерность равна 0. Поэтому и размерность леса G равна 0.

Шаг индукции доказан. По принципу математической индукции, размерность любого леса равна 0. □

5.4. Висячие вершины в деревьях

Вершины степени 0 называются *изолированными*, а вершины степени 1 — *висячими*. В дереве из одной вершины висячих вершин нет вообще, единственная вершина изолированная.

Теорема 5.9. *В дереве с хотя бы двумя вершинами есть по крайней мере две висячие вершины.*

Доказательство. Используем следствие (5.6). Пусть в дереве $n \geqslant 2$ вершин. Тогда в этом дереве $n - 1$ ребро.

Обозначим степени вершин дерева $d_1, \ldots, d_n$. Поскольку $n \geqslant 2$, изолированных вершин нет (каждая изолированная вершина является компонентой связности). Из

теоремы 4.1 получаем равенство

$$d_1 + \cdots + d_n = 2(n-1).$$

Перепишем его в виде

$$(d_1 - 2) + (d_2 - 2) + \cdots + (d_n - 2) = -2.$$

Так как $d_i > 0$, каждое слагаемое в левой части не меньше -1. Значит, хотя бы два должны равняться -1, они отвечают висячим вершинам, для которых $d_i - 2 = 1 - 2 = -1$. □

5.5. Остовные деревья

У нас уже появлялись подграфы, индуцированные подмножеством вершин. В общем случае *подграф* графа получается так: выбираем некоторое подмножество вершин и некоторое подмножество рёбер с концами в выбранных вершинах. Подграф называется *остовным*, если его множество вершин совпадает с множеством вершин самого графа. Из следствия 5.4. получаем такой важный факт.

Теорема 5.10. *В любом связном графе есть остовное дерево.*

Доказательство. Удаляем не мосты графа, пока это возможно. При удалении не моста связный граф остаётся связным. В итоге получится связный граф, в котором каждое ребро — мост, то есть дерево. Оно остовное — вершины те же самые, что в исходном графе. □

Сформулируем без доказательства замечательную теорему.

Теорема 5.11 (Кэли)**.** *Количество остовных деревьев в полном графе на n вершинах равно n^{n-2} при $n \geqslant 2$.*

Лекция 6

Графы–3. Ориентированные графы

Простой ориентированный граф (орграф) — это множество вершин V и множество рёбер E. Рёбрами являются упорядоченные пары вершин.

Если $e = (u, v) \in E$, то вершина u называется *началом* ребра e, а вершина v — *концом*. Говорят также, что ребро $e = (u, v)$ орграфа *выходит* из вершины u и *входит* в вершину v.

На рисунках ориентированные рёбра изображаются линиями со стрелками. Стрелка направлена от начала к концу ребра, см. рис. 6.1.

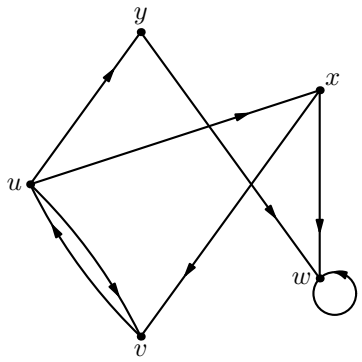


Рис. 6.1: Пример ориентированного графа

В случае неориентированных графов ребро $\{u, v\}$ однозначно определяется своими концами. В случае ориентированных графов между парой вершин u, v возможны два ребра: (u, v) и (v, u). В графе на рис. 6.1 есть оба таких ребра.

Кроме того, наше определение разрешает *петли* в орграфе. На рис. 6.1 есть пример петли: упорядоченная пара (w, w). У петли начало и конец совпадают.

Орграфы, как и обычные неориентированные графы, можно задавать матрицей смежности. Как и раньше, чтобы говорить о матрице смежности графа, нужно перенумеровать его вершины и рёбра.

Матрица смежности орграфа — квадратная матрица порядка n, где n — количество вершин графа. В этой матрице на пересечении i-й строки и j-го столбца

стоит 1, если в орграфе есть ребро (i, j); иначе там стоит 0. Матрица смежности орграфа уже не обязательно симметрическая: возможно, что $A_{ij} = 1$, $A_{ji} = 0$, $i \neq j$.

6.1. Степени вершин

В неориентированных графах степень вершины равна количеству инцидентных ей рёбер. В орграфах часть рёбер входит в вершину, часть — выходит. Их считают по отдельности. *Исходящая степень* вершин равна числу рёбер, выходящих из этой вершины. *Входящая степень* равна числу рёбер, входящих в вершину. Если в вершине v есть петля $e = (v, v)$, то вершина v является и началом и концом петли e. Поэтому петля даёт вклад 1 и в исходящую степень, и во входящую.

Для степеней вершин графа есть соотношение, аналогичное соотношению для суммы степеней вершин неориентированного графа.

Теорема 6.1. *Сумма исходящих степеней всех вершин равна сумме входящих степеней всех вершин: обе суммы равны числу рёбер графа.*

Доказательство. Каждое ребро имеет одно начало (выходит из какой-то вершины) и поэтому учитывается по разу, когда мы складываем исходящие степени всех вершин. Аналогично для концов рёбер. □

У каких орграфов и входящая, и исходящая степени каждой вершины равны 1? Пример такого графа изображён на рис. 6.2. Из этого рисунка легко догадаться, как устроены такие графы в общем случае. Они разбиваются на ориентированные циклы.

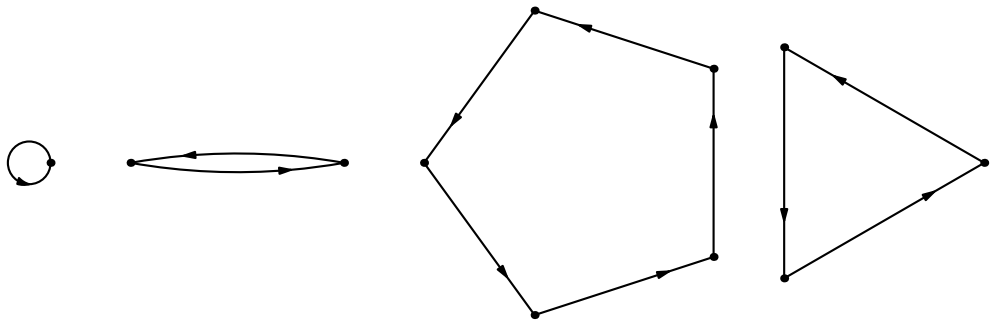


Рис. 6.2: Пример ориентированного графа с входящими и исходящими степенями 1 в каждой вершине

Определения путей и циклов для орграфов похожи на определения для обычных графов. *Путь* по орграфу — это такая последовательность вершин $v_1, v_2, v_3, \ldots, v_k$, в которой стоящие рядом члены (вершины v_i и v_{i+1} при всех i) соединены ребром, причём v_i — начало ребра, а v_{i+1} — его конец. Цикл — это путь, у которого первая и последняя вершины совпадают. Простые пути и циклы определяются точно так же, как в неориентированном случае: все вершины должны быть различны.

Ориентированным циклом называется множество рёбер любого простого цикла (обратите внимание на разницу: ориентированный цикл — множество рёбер, а цикл — последовательность вершин и рёбер).

Теорема 6.2. *Если в орграфе G каждая вершина имеет исходящую и входящую степень* 1*, то рёбра такого графа разбиваются на несколько ориентированных циклов: каждое ребро принадлежит в точности одному из этих циклов.*

Доказательство. Рассмотрим орграф G на множестве вершин V, который удовлетворяет условиям теоремы.

Выберем вершину $v = v_0$ и построим путь $v_0, v_1, \ldots, v_i, \ldots$ по орграфу G. Этот путь однозначно определён, так как из каждой вершины v_i выходит ровно одно ребро (v_i, v_{i+1}).

Построенный путь бесконечный, а вершин в графе конечное количество. Поэтому рано или поздно на этом пути какая-то вершина повторится. Выберем самое первое повторение: $v_i = v_\ell$, $i < \ell$; все вершины $v_0, \ldots, v_{\ell-1}$ различны.

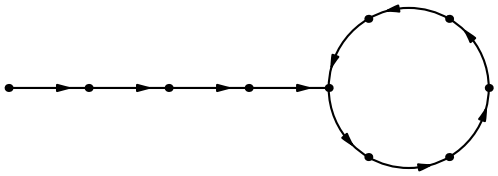


Рис. 6.3: Повтор не в начальной вершине противоречит входящей степени 1

Докажем от противного, что повторится именно вершина v_0. Пусть $i > 0$. Тогда в графе есть два ребра (v_{i-1}, v_i) и $(v_{\ell-1}, v_\ell)$, входящие в $v_i = v_\ell$. Поскольку входящая степень равна 1, то $v_{i-1} = v_{\ell-1}$. Это противоречит тому, что $v_i = v_\ell$ первое повторение. (Рис. 6.3.)

Получили простой цикл $C(v) = (v_0, v_1, \ldots, v_\ell = v_0)$. Из условий теоремы следует, что никакие другие рёбра не входят и не выходят из вершин v_i.

Различные циклы вида $C(v)$ не пересекаются, так как исходящие степени вершин равны 1. С другой стороны, каждая вершина v лежит на цикле $C(v)$. Значит, рёбра этих циклов и задают разбиение рёбер орграфа на ориентированные циклы. □

Теперь опишем графы, у которых исходящая степень в каждой вершине равна 1. Они не исчерпываются наборами ориентированных циклов, как показывает рис. 6.3. Оказывается, в этих графах помимо ориентированных циклов есть ещё *ориентированные деревья*. Определим формально, что это такое.

Иногда в дереве выделяют особую вершину — *корень*. В этом случае дерево называют корневым. (Заметим, что висячие вершины, отличные от корня, называют *листьями*.)

По любому простому неориентированному графу можно построить орграф, выбрав для каждого ребра одну из двух возможных ориентаций. Для любого корневого дерева есть две естественные ориентации: по направлению к корню и по направлению от корня.

Будем называть корневое дерево с корнем r *ориентированным к корню*, если каждое ребро (u, v) удовлетворяет такому условию: вершина v лежит на простом пути от u к r. (Напомним, что в дереве такой путь единственный.)

Аналогично определяется *ориентация от корня*: вершина u лежит на простом пути от v к r. Но нам она сейчас не понадобится.

Теорема 6.3. *Если в орграфе G каждая вершина имеет исходящую степень 1, то рёбра такого графа разбиваются на несколько ориентированных циклов и несколько ориентированных корневых деревьев, корень каждого такого дерева принадлежит одному из циклов.*

Доказательство. Как и доказательстве предыдущей теоремы, рассмотрим путь

$$v = v_0, v_1, \ldots, v_i, \ldots$$

по орграфу G, начинающийся в вершине v. Этот путь однозначно определён, вершины в нём обязаны повторяться.

Рассмотрим первое повторение $v_i = v_j$. Тогда $v_{i+1} = v_{j+1}$ и так далее. Поэтому, начиная с v_i, этот путь периодичен, и наименьший период является простым циклом. Обозначим этот цикл $C(v)$, а первую повторяющуюся вершину — $r(v)$.

Различные циклы вида $C(v)$ не пересекаются. Но теперь, в отличие от случая когда и все входящие степени равны 1, будут также и вершины, которые не лежат на этих циклах. (Поскольку теперь v необязательно лежит на цикле $C(v)$.

Пусть r лежит на одном из циклов. Рассмотрим граф, индуцированный множеством вершин $V_r = \{v : r = r(v)\}$ (за исключением возможной петли r, r). Докажем, что это ориентированное дерево с корнем r.

Заметим, что если убрать ориентации рёбер, то полученный неориентированный граф G_r является связным (из любой вершины достижима вершина r).

Обозначим через n количество вершин в графе G_r, а через m — количество рёбер. Из каждой вершины, кроме r выходит ровно одно ребро (ребро, выходящее из r лежит на цикле). Значит, $m = n - 1$. По следствию 5.6 из теоремы 5.5 граф G_r является деревом.

Осталось доказать, что всё ребра в дереве G_r ориентированы в исходном орграфе к корню. Действительно, из любой вершины $v \in V_r$ есть путь в $r(v)$, то есть корень дерева. Значит ребро (v, v') (единственное ребро, исходящее из v) ориентировано к корню. □

6.2. Сильная связность, компоненты сильной связности

Как и в случае неориентированных графов, обозначим через $R(u)$ множество тех вершин v, которые *достижимы* из u, то есть существует путь с началом u и концом v.

Будем говорить, что вершина u *сильно связана* с вершиной v, если v достижима из u и наоборот, то есть если есть путь из u в v, а также путь из v в u.

Через $C(u)$ обозначим множество тех вершин v, которые сильно связаны с u. Эти множества обладают теми же свойствами, что и компоненты связности обычного ориентированного графа и называются *компонентами сильной связности*.

Лемма 6.4. *Для любого графа и любых его вершин v_1, v_2, v_3 выполняются следующие свойства:*

1. $v_1 \in C(v_1)$ *(вершина сильно связана сама с собой);*
2. $v_1 \in C(v_2)$ *равносильно* $v_2 \in C(v_1)$*;*
3. *если* $v_1 \in C(v_2)$ *и* $v_2 \in C(v_3)$*, то* $v_1 \in C(v_3)$*.*

Доказательство. v_1 — путь в любом графе, поэтому v_1 сильно связана с самой собой.

Определение сильной связанности симметрично, отсюда свойство 2.

Наконец, если в графе есть пути из v_3 в v_2, из v_2 в v_3, из v_2 в v_1, из v_1 в v_2, то обязательно есть и пути из v_1 в v_3 (соединяем путь из v_1 в v_2 с путём из v_2 в v_3), а также из v_3 в v_1 (соединяем путь из v_3 в v_2 с путём из v_2 в v_1). Это доказывает свойство 3. □

Лемма 6.5. *Если $w \in C(v_1) \cap C(v_2)$, то $C(v_1) = C(v_2)$. Компоненты сильной связности не пересекаются или совпадают.*

Доказательство. Поскольку w сильно связана с v_1 и с v_2, то v_2 достижима из v_1 (путь из v_1 в w, соединённый с путём из w в v_2). Аналогично, v_1 достижима из v_2.

Значит, $C(v_2) \subseteq C(v_1)$ и $C(v_1) \subseteq C(v_2)$. То есть $C(v_1) = C(v_2)$. □

Из этих свойств, аналогично случаю неориентированных графов, следует, что компоненты сильной связности орграфа задают разбиение его вершин.

Если всё множество вершин орграфа образует компоненту сильной связности, такой орграф называется *сильно связным*. Примером сильно связного графа является ориентированный цикл.

6.3. Ациклические орграфы

Орграф называется *ациклическим*, если каждая компонента сильной связности состоит из одной вершины. Другими словами, никакие две различные вершины не являются сильно связанными. Название объясняется следующей теоремой. В ней мы требуем отсутствия петель в орграфе.

Теорема 6.6. *Следующие свойства ориентированного графа без петель равносильны:*

(1) *Каждая сильно связная компонента состоит из одной вершины.*

(2) *В орграфе нет циклов длины больше* 0*.*

(3) *Вершины орграфа можно пронумеровать натуральными числами таким образом, чтобы все рёбра вели «вверх»: из вершины с меньшим номером в вершину с бóльшим.*

Для доказательства этой теоремы нам понадобится такая полезная лемма.

Лемма 6.7. *В орграфе без циклов есть вершина, из которой не выходит ни одного ребра, а также есть вершина, в которую не входит ни одного ребра.*

Доказательство. От противного. Если из каждой вершины выходит хотя бы одно ребро, то оставим по одному ребру, исходящему из каждой вершины. Получаем граф, в котором исходящие степени вершин равны 1. Как мы уже доказали в теореме 6.3, в таком графе есть цикл.

Чтобы доказать второе утверждение, перейдём к графу, в котором ориентации всех рёбер изменены на противоположные. Если исходный граф был ациклическим, то граф с инвертированными ориентациями также будет ациклическим. Но исходящие и входящие степени вершин переставляются. □

Доказательство теоремы 6.6. (1) $\Rightarrow$ (2) равносильно контрапозиции $\neg(2) \Rightarrow \neg(1)$. Докажем вторую импликацию. Раз в орграфе нет петель, в нём нет циклов длины 1. Если в орграфе есть цикл с $n > 1$ вершинами, то вершины этого цикла сильно связаны (из любой можно попасть в любую по циклу).

(2) $\Rightarrow$ (1) равносильно контрапозиции $\neg(1) \Rightarrow \neg(2)$. Докажем вторую импликацию. Если вершины $a \neq b$ сильно связаны, то существуют пути из a в b и из b в a. Соединением этих путей получается цикл.

(3) $\Rightarrow$ (2): если возможна нумерация вершин, при которой все рёбра идут из меньшей вершины в бóльшую, то циклов нет: вдоль любого пути номера вершин строго возрастают, что невозможно при возвращении в исходную вершину.

(2) $\Rightarrow$ (3) докажем индукцией по числу вершин усиленный вариант: нумерация использует числа от 1 до n, где n — число вершин в орграфе.

База индукции: граф без петель на одной вершине. Он ациклический и требуемая нумерация существует (это очевидно, так как рёбер нет).

Шаг индукции: пусть (2) $\Rightarrow$ (3) выполняется для графов с $< n$ вершинами. Рассмотрим граф без циклов на n вершинах. Выберем вершину v_n исходящей степени 0, которая существует в таком орграфе по лемме 6.7. Ей присвоим номер n. Удалив v_n и все входящие в неё рёбра, получим граф без циклов. (Циклы в нём были бы циклами и в исходном графе.) По предположению индукции его вершины можно пронумеровать числами от 1 до $n-1$ с соблюдением условия. Объединяя эту нумерацию с номером n вершины v_n, получаем искомую нумерацию. Шаг индукции доказан. □

6.4. Эйлеровы (ор-)графы

Цикл (в неориентированном или ориентированном графе) называется *эйлеровым*, если он проходит по всем рёбрам графа по ровно одному разу (любое ребро входит

в цикл, и никакое ребро не входит дважды).

Граф называется *эйлеровым*, если в нём есть эйлеров цикл.

Есть простой критерий эйлеровости графов и орграфов. Прежде всего заметим, что добавление и удаление *изолированных вершин*, т.е. тех вершин, из которых не входит и в которые не выходит ни одно ребра, не изменяет свойство эйлеровости графа.

Теорема 6.8. *В ориентированном графе*[1)] *без изолированных вершин существует эйлеров цикл тогда и только тогда, когда граф сильно связен и у любой вершины входящая степень равна исходящей.*

Доказательство. Пусть эйлеров цикл в орграфе есть. Тогда он проходит через все вершины (поскольку они имеют ненулевую степень), и по нему можно дойти от любой вершины до любой. Значит, орграф сильно связен.

Возьмём какую-то вершину v, пусть она встречается в эйлеровом цикле k раз. Двигаясь по циклу, мы приходим в неё k раз и уходим k раз, значит, использовали k входящих и k исходящих рёбер. При этом, раз цикл эйлеров, других рёбер у этой вершины нет, так что в ориентированном графе её входящая и исходящая степени равны k.

В обратную сторону чуть сложнее. Пусть орграф сильно связен и в каждой вершине исходящая степень равна входящей. Рассмотрим пути, которые не проходят дважды по одному ребру. Выберем среди таких путей самый длинный (его длина не больше общего количества рёбер)

$$\tau = (v_1,\ v_2,\ v_3,\ \ldots,\ v_{t-1},\ v_t)$$

и докажем, что этот путь и является искомым циклом, то есть что $v_1 = v_t$ и этот путь содержит все рёбра орграфа.

В самом деле, если τ самый длинный, то добавить к нему ребро (v_t, v_{t+1}) невозможно. Это означает, что все выходящие из v_t рёбра уже входят в τ. Это возможно, лишь если $v_1 = v_t$: если вершина v_t встречалась только внутри пути (пусть она входит k раз внутри пути и ещё раз в конце пути), то мы использовали $k+1$ входящих рёбер и k выходящих, и больше выходящих нет. Это противоречит равенству входящей и исходящей степени.

Итак, мы имеем цикл, и осталось доказать, что в него входят все рёбра. В самом деле, если во всех вершинах цикла использованы все рёбра, то из вершин этого цикла нельзя попасть в вершины, не принадлежащие циклу, то есть использованы все вершины (так как орграф сильно связен) и, следовательно, все рёбра. С другой стороны, если из какой-то вершины v_i выходит ребро (v_i, v), то путь можно удлинить до

$$(v_{i+1},\ \ldots, v_t = v_1,\ \ldots,\ v_i,\ v)$$

вопреки нашему выбору самого длинного пути. Аналогично можно получить противоречие и для входящего ребра (v, v_i), добавив его в начало. □

[1)] Спасибо Олегу Мошкину за замеченную опечатку.

Для неориентированных графов критерий аналогичен.

Теорема 6.9. *Неориентированный граф без вершин нулевой степени содержит эйлеров цикл тогда и только тогда, когда он связен и степени всех вершин чётны.*

Доказательство полностью аналогично доказательству в ориентированном случае. Кратко повторим его.

Пусть эйлеров цикл в графе есть. Он проходит по всем вершинам, так что граф связен. В каждую вершину эйлеров цикл k раз заходит и k раз выходит. Значит, степень вершины $k + k = 2k$ чётна.

В обратную сторону опять рассматриваем самый длинный путь, в котором каждое ребро встречается не больше одного раза. Это цикл, так как иначе есть вершина нечётной степени.

Этот цикл обязан содержать все рёбра графа, так как в противном случае его можно удлинить.

Лекция 7

Графы–4. Раскраски. Комбинаторика–2

7.1. Раскраски графов

Правильной раскраской вершин неориентированного графа $G(V, E)$ в k цветов называется такое присваивание вершинам графа чисел (цветов) от 1 до k, что присвоенные смежным вершинам числа различны. Если для графа существует хотя бы одна правильная раскраска в k цветов, граф называется *k-раскрашиваемым*.

Очень легко понять, какие графы 1-раскрашиваемые. Это в точности графы без рёбер. Действительно, если вершинам графа без рёбер присвоить одно и то же число (цвет), то условие правильной раскраски выполняется. И наоборот: если в графе есть ребро $\{u, v\}$, то в правильной раскраске вершинам u, v присвоены разные цвета, поэтому количество цветов хотя бы 2.

Случай правильных раскрасок в 2 цвета интереснее.

Теорема 7.1. *2-раскрашиваемые графы это в точности графы, в которых длины всех циклов чётные.*

Доказательство. Достаточно доказать утверждение для связных графов, так как несвязный граф 2-раскрашиваемый тогда и только тогда, когда все его компоненты связности 2-раскрашиваемые и то же самое верно для свойства «длины всех циклов чётные».

Если в графе есть цикл нечётной длины, то его нельзя правильно раскрасить в 2 цвета. Соседние вершины должны быть противоположных цветов, поэтому количество вершин одного цвета должно равняться количеству вершин другого цвета.

Теперь докажем обратное. Пусть в графе длины всех циклов чётные. Докажем, что тогда для любых двух вершин u, v длины путей из u в v имеют одинаковую чётность.

Если в графе есть путь $\alpha = (u, \ldots, v)$ с чётным числом вершин (нечётной длины), а также другой путь $\beta = (v, \ldots, u)$ с нечётным числом вершин (чётной длины), то соединение этих путей $\alpha\beta$ (идём по первому пути из u в v, затем по второму пути из v в u) даёт цикл с нечётным числом вершин (нечётной длины): вершины u и

v считаются дважды в путях α, β и по одному разу в цикле. Это противоречит сделанному предположению, что все циклы в графе имеют чётную длину.

Теперь укажем искомую правильную раскраску в 2 цвета. Выберем вершину u и раскрасим вершину x графа в цвет 0, если длины путей из u в x чётные; в цвет 1, если длины путей из u в x нечётные. Это правило корректно по доказанному выше утверждению про одинаковую чётность длин циклов в графе без нечётных циклов (вспомним также, что мы доказываем утверждение теоремы для связных графов).

Заметим, что смежные в графе вершины не могут быть покрашены в один цвет: если $\{x, y\}$ — ребро графа, то для каждого пути $(u, \ldots, x)$ существует путь в y противоположной чётности, а именно, $(u, \ldots, x, y)$. □

Пример 7.1. Докажем, что булев куб 2-раскрашиваемый. Вершинами булева куба Q_n являются двоичные слова длины n. Покрасим вершины с чётным количеством единиц в цвет 0; вершины с нечётным количеством единиц в цвет 1.

Ребро булева куба связывает вершины, которые отличаются ровно в одной позиции. Одна вершина на этой позиции содержит 0, другая — 1. Чётность количества единиц в таких вершинах разная, то есть они покрашены в разные цвета.

Замечание 7.1. Эта теорема обосновывает простой и эффективный алгоритм проверки 2-раскрашиваемости графа. Закрасим какую-нибудь вершину в цвет 0. Далее действуем так.

Если есть непокрашенная вершина u, соединённая с уже покрашенной v, красим u в цвет, противоположный v.

Если все непокрашенные вершины несмежны уже покрашенным, красим какую-нибудь из непокрашенных в цвет 0.

Если в процессе такой раскраски встретилась вершина, которая смежна с вершинами двух противоположных цветов, то объявляем, что граф не является 2-раскрашиваемым. В противном случае получаем правильную 2-раскраску.

Корректность алгоритма доказывается так: нужно индукцией по числу действий проверить, что в 0 красятся вершины, которые соединены с начальной путями чётной длины, а в 1 — те, которые соединены путями нечётной длины.

Вопрос о существовании правильной раскраски в 3 и более цветов гораздо сложнее. Простого способа проверить 3-раскрашиваемость нет и не предвидится.

7.2. Двудольные графы

Двудольным графом называется неориентированный граф, в котором вершины заранее разделены на две *доли — левую* и *правую*, и все рёбра соединяют вершины из разных долей (нет рёбер, соединяющих вершины одной доли). Другими словами, чтобы задать двудольный граф, надо указать два конечных множества L (левую долю) и R (правую долю) и указать, какие вершины левой доли соединены с какими вершинами правой доли.

Разделение вершин на левые и правые задаёт правильную раскраску двудольного графа. Таким образом, граф можно представить как двудольный, если в нём нет циклов нечётной длины.

Пример 7.2. *Паросочетанием* называется двудольный граф, у которого степени всех вершин не больше 1.

Пример паросочетания приведён на рис. 7.1. Здесь доли нарисованы как верхняя и нижняя. Ясно, что нет никакой проблемы объявить одну из них левой, а другую — правой.

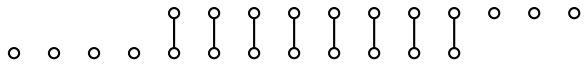


Рис. 7.1: Паросочетание

Такой граф устанавливает взаимно однозначное соответствие между частью вершин левой доли и частью вершин правой доли. В частности, если паросочетание *совершенное* (нет изолированных вершин, то есть вершин степени 0), то рёбра паросочетания устанавливают взаимно однозначное соответствие между вершинами левой и правой долей. Такое возможно, если размеры долей одинаковы.

Рассмотрим такую задачу. Дан двудольный граф G, нужно найти среди его рёбер максимальное по количеству рёбер паросочетание.

Размер максимального паросочетания не превосходит размера любой из долей. Однако бывают графы, в которых размер максимального паросочетания намного меньше, см. рис. 7.2.

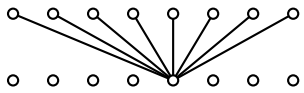


Рис. 7.2: Размер максимального паросочетания в таком графе равен 1

Для решения задачи о максимальном паросочетании есть эффективные алгоритмы. Нас больше интересуют теоремы, а не алгоритмы. Поэтому сформулируем и докажем критерий того, что размер максимального паросочетания достигает верхней границы, то есть совпадает с размером (меньшей) доли.

Без ограничения общности считаем, что в левой (нижней) доле вершин не больше, чем в правой (верхней). Для множества вершин $X \subseteq L$ левой доли обозначим $G(X) \subseteq R$ множество всех соседей этих вершин (они все лежат в правой доле, так как мы рассматриваем двудольные графы).

Теорема 7.2 (теорема Холла)**.** *Пусть для двудольного графа $G = (L \cup R, E)$ выполнено $|L| \leqslant |R|$.*

Если для каждого множества $X \subseteq L$ множество соседей $G(X) \subseteq R$ содержит не меньше вершин, чем X, то в графе G есть паросочетание размера $|L|$.

Доказательство. В доле L нет изолированных вершин, так как для изолированной вершины $v \in L$ нарушается условие теоремы: $1 = |\{v\}| > |G(\{v\})| = |\varnothing| = 0$. Поэтому паросочетание размера 1 точно есть.

Пусть имеется паросочетание P с концами в множестве $X \subset L$ и $Y \subset R$. Если оно не максимальное, то существует вершина $v \in L \setminus X$. Докажем, что тогда существует паросочетание с концами в левой доле $\{v\} \cup X$. Отсюда будет следовать утверждение теоремы.

Определим вспомогательный ориентированный граф $G' = (L \cup R, E')$ на том же множестве вершин. Его рёбра имеют вид

$$E' = \{(x, y) : \{x, y\} \in E(G),\ x \in L,\ y \in R; \quad \{y, x\} \in P,\ y \in R,\ x \in L\}.$$

Другими словами, мы разрешаем идти из левой доли в правую по любом ребру графа G, а из правой в левую — только по рёбрам паросочетания P.

Обозначим через $C(v) = L' \cup R'$, $L' \subseteq L$, $R' \subseteq R$, область достижимости вершины v в этом графе. Заметим, что $G(L') \subseteq R'$: ведь в правую долю разрешается идти по любым рёбрам G. Поэтому $|R'| \geqslant |G(L')|$.

В обратную сторону разрешается ходить по рёбрам P. Поэтому, если $y \in Y \cap R'$, то есть существует ребро $\{y, x\} \in P$, то $x \in L'$. Множество L' заведомо содержит вершину v, остальные вершины этого множества лежат в X (множество левых концов паросочетания P). По условию теоремы $|G(L')| \geqslant |L'|$. Значит, $|R'| \geqslant |L'|$. Так как $|L' \cap X| = |R' \cap Y|$, а $L' \setminus X = \{v\} \neq \varnothing$, то множество $R' \setminus Y$ непусто. Тут полезно посмотреть на рис. 7.3.

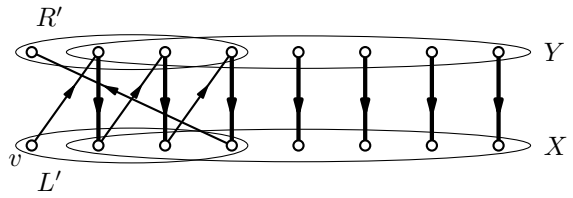


Рис. 7.3: Область достижимости «внешней» вершины v

Выберем в $R' \setminus Y$ вершину w. По определению области достижимости существует ориентированный путь из v в w, причём из правой доли в левую этот путь идёт по (ориентированным) рёбрам паросочетания P.

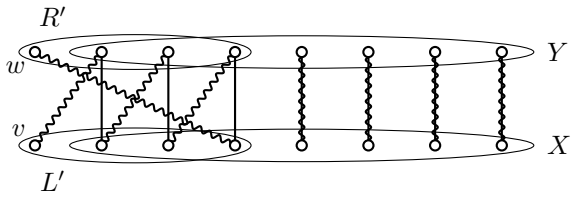


Рис. 7.4: Увеличение паросочетания

Рёбра этого пути разбиваются на две группы: «старые» рёбра из P и «новые» рёбра не из P (но эти рёбра являются рёбрами исходного графа G). Причём рёбра разных групп чередуются, а начало и конец пути инцидентны «новым рёбрам». Это означает, что новые рёбра образуют паросочетание в G, размер которого на 1 больше $|R' \cap Y|$. Заменяя «старые» рёбра, исходящие из $R' \cap Y$, на «новые», и добавляя остальные рёбра исходного паросочетания (с концами в $Y \setminus R'$) получаем паросочетание P', в котором рёбер на одно больше, чем в P (см. рис. 7.4, рёбра увеличенного паросочетания нарисованы волнистыми линиями). □

7.3. Возвращение к комбинаторике

7.3.1 Паросочетания и взаимно однозначные соответствия

Теперь вернёмся к перечислительной комбинаторике. Мы уже отмечали, что паросочетание задаёт взаимно однозначное соответствие между левыми и правыми концами рёбер. Поэтому, в частности, левых концов ровно столько же, сколько правых.

Это позволяет «сводить» одну комбинаторную задачу к другой, если построить паросочетание, в котором левые концы — это элементы множества, размер которого уже известен, а правые концы — элементы того множества, размер которого мы хотим найти.

Приведём пару простых примеров, в последующем таких примеров станет больше.

Пример 7.3. Рассмотрим все подмножества n-элементного множества. Обозначим семейство[1)] таких множеств через $\mathcal{P}(n)$. Сколько их?

Элементы n-элементного множества всегда возможно занумеровать числами от 1 до n и считать, что рассматриваем подмножества множества $[n] = \{1, 2, \ldots, n\}$.

Построим такой двудольный граф: вершины левой доли — двоичные слова длины n, вершины правой доли — подмножества из $\mathcal{P}(n)$. Рёбра этого графа соединяют пары $(w = w_1 w_2 \ldots w_n, S)$, где $i \in S$ тогда и только тогда, когда $w_i = 1$.

Этот граф — паросочетание: у каждого слова есть ровно один сосед в этом графе (множество тех позиций, на которых стоят единицы), и у каждого множества есть ровно один сосед (слово, в котором 1 стоит на тех позициях, номера которых входят в множество).

Значит, подмножеств n-элементного множества столько же, сколько двоичных слов длины n. А это число мы уже подсчитали, оно равно 2^n.

Пример 7.4. Мы уже подсчитывали количество монотонных путей из 0 в n (строго возрастающих последовательностей целых чисел, начинающихся с 0 и заканчивающихся n). При $n \geqslant 1$ оно равно 2^{n-1}. Мы получили его решением рекуррентного соотношения. А сейчас приведём «натуральное» доказательство этого факта.

[1)] Слово «семейство» используется как синоним слова «множество» для благозвучия.

Построим такой двудольный граф: вершины левой доли — подмножества $(n-1)$-элементного множества $[n-1] = \{1, \ldots, n-1\}$, вершины правой доли — монотонные пути из 0 в n. Ребра этого графа соединяют пары (S, τ), где промежуточные вершины пути τ — это в точности числа из множества S.

Есть ровно один монотонный путь, в котором множество промежуточных вершин совпадает с данным множеством S: из-за монотонности нужно сначала посетить минимальное в S число, затем следующее по величине и т.д. Поэтому степени вершин в левой доле равны 1.

У каждого монотонного пути однозначно определено множество промежуточных вершин. Поэтому степени вершин в правой доле также равны 1.

Значит, этот граф — паросочетание. Поэтому количество монотонных путей в точности равно количеству подмножеств $(n-1)$-элементного множества, то есть равно 2^{n-1}.

7.3.2 Комбинаторное «правило деления»

Двудольные графы дают перечислительной комбинаторике не только паросочетания. Рассмотрим более подробно метод двойного подсчёта, который мы уже применяли раньше.

Суть этого метода в том, что число рёбер в двудольном графе равно сумме степеней вершин левой доли (каждое ребро имеет ровно один конец в левой доле). Но оно же равно и сумме степеней вершин в правой доле. Получаем равенство, которое можно использовать для выражения одних величин через другие.

Для перечислительной комбинаторики особенно важен случай, когда степени вершин в каждой доле одинаковы. Пусть в левой доле L вершин, в правой — R; степень каждой вершины в левой доле равна d_1, а в правой — d_2. Тогда выполняется равенство

$$Ld_1 = Rd_2, \tag{7.1}$$

которое мы образно будем называть «правилом деления». С помощью этого правила можно найти R, если известны все три величины: $R = Ld_1/d_2$ (вот оно, деление). Мы использовали частный случай этого правила для паросочетаний ($d_1 = d_2 = 1$).

Теперь рассмотрим примеры, когда одна из степеней отлична от 1.

Те слова длины k в n-символьном алфавите, в которых все символы разные, называют *упорядоченными выборками* или *размещениями* из n по k. (Представьте, что в лототроне лежат шары, помеченные всеми символами B, мы по очереди достаём шары из лототрона и фиксируем порядок их появления.)

Количество размещений из n по k обозначим $A_{n,k}$. Для краткости обозначений используем то же обозначение и для самого множества размещений. Из контекста нужно понимать, идёт ли речь о множестве или о его мощности.

Ясно, что $A_{n,k} = 0$, если $k > n$ (принцип кроликов, сейчас мы «сажаем» позиции слова в клетки, которые являются символами алфавита). В общем случае выполняется рекуррентное соотношение.

Утверждение 7.3. $A_{n,k} = (n-k+1)A_{n,k-1}$, $1 \leqslant k \leqslant n$.

Доказательство. Построим двудольный граф с долями $A_{n,k}$ и $A_{n,k-1}$, рёбра этого графа имеют вид

$$(b_1 b_2 \dots b_{k-1} b_k,\ b_1 b_2 \dots b_{k-1}).$$

Другими словами, мы соединяем ребром размещение длины k с его началом длины $k-1$. Поэтому степень любой вершины левой доли равна 1.

Степень любой вершины правой доли равна $n-k+1$: продолжить размещение можно любым из ещё неиспользованных $n-k+1$ символов алфавита.

Равенство утверждения теперь становится частным случаем равенства (7.1). □

Теорема 7.4. $A_{n,k} = n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)$.

Доказательство. Применяя последовательно утверждение 7.3, получаем

$$\begin{aligned} A_{n,k} &= (n-k+1)A_{n,k-1} = (n-k+1)(n-k+2)A_{n,k-2} = \dots \\ &= (n-k+1)\cdot\ldots\cdot(n-1)nA_{n,0}. \end{aligned}$$

А размещение длины 0 ровно одно. □

Формулу из теоремы 7.4 можно записать через факториалы. По определению $n! = 1\cdot 2\cdot\ldots\cdot n$, $0! = 1$ (последнее не вполне очевидно из общего случая). Поэтому

$$A_{n,k} = n(n-1)(n-2)\cdot\ldots\cdot(n-k+1) = \frac{n!}{(n-k)!}\,.$$

Важный частный случай размещений получается при $k = n$. Получаем последовательности чисел из $[n]$ длины n. Такие последовательности называются *перестановками*. Количество перестановок находится по теореме 7.4, оно равно

$$A_{n,n} = \frac{n!}{0!} = n!\,.$$

Теперь рассмотрим *сочетания* или *неупорядоченные выборки*. В примере с лототроном это означает, что нас интересует лишь то, какие шары были вытащены, а не порядок их появления.

На обычном математическом языке это означает, что нас интересует *подмножество* вытащенных шаров. Вот и будем называть сочетанием из n элементов по k подмножество n-элементного множества, в котором ровно k элементов. Количество сочетаний из n по k будем обозначать C_n^k (обратите внимание на порядок индексов).

Теорема 7.5. $C_n^k = \dfrac{n!}{k!(n-k)!}\,.$

Доказательство. Как нетрудно догадаться, чтобы доказать равенство, в котором используется деление, нужно применить комбинаторное правило деления.

Используя формулу для размещений, перепишем это равенство в другом виде:

$$C_n^k \cdot k! = A_{n,k}\,. \tag{7.2}$$

Построим двудольный граф. Вершины левой доли — сочетания из n по k, то есть k-элементные подмножества множества $[n]$; вершины правой доли — размещения $A_{n,k}$. Рёбра графа соединяют пары (S, w), где S — это множество символов, которые встречаются в размещении w.

Из определения ясно, что степени вершин правой доли равны 1 (по размещению множество символов определяется однозначно).

Найдём степень вершины левой доли. Сколько есть размещений длины k, которые используют ровно k заданных символов? Это $A_{k,k} = k!$, то есть количество перестановок k символов.

Подставляя полученные значения в (7.1), получаем (7.2). □

Комбинируя уже известный приём с паросочетаниями, отсюда можно получить решения многих перечислительных задач.

Пример 7.5. Найдём количество двоичных слов длины n, в которых ровно k единиц.

Для этого ограничим граф из примера 7.3 на слова, у которых ровно k единиц. Соседями в этом графе будут как раз k-элементные множества. Значит, количество таких двоичных слов равно C_n^k.

7.4. Биномиальные коэффициенты

Рассмотрим *бином* $(x + y)^n$. Как известно из алгебры, целое алгебраическое выражение можно записать в виде многочлена. Поэтому

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = x^n + \binom{n}{1} x^{n-1} y + \cdots + \binom{n}{n-1} x y^{n-1} + y^n. \qquad (7.3)$$

Здесь $\binom{n}{k}$ — числа, которые называются *биномиальными коэффициентами*. (Обратите внимание на порядок индексов.) Оказывается, это в точности числа сочетаний из n по k.

Теорема 7.6. $\binom{n}{k} = C_n^k$.

Доказательство. Давайте переходить от левой части (7.3) к правой в два этапа. Сначала раскроем скобки. Получим сумму выражений вида $xyyx\ldots$, где всего сомножителей n, а каждый из них — это x или y. Количество таких сомножителей равно количеству слов длины n в алфавите $\{x, y\}$, то есть равно 2^n (от замены символов $0, 1$ на x, y количество слов не изменяется).

Теперь «приведём подобные». Мы знаем, что сложение коммутативно и ассоциативно. Поэтому все слагаемые с одинаковым количеством x и y равны $x^{n-k}y^k$, где k — количество символов y (а количество символов x равно $n-k$, потому что других символов в этих выражениях нет).

Итак, $\binom{n}{k}$ равен количеству слагаемых с k символами y и $n-k$ символами x, а это количество равно количеству двоичных слов с k единицами, то есть C_n^k (см. пример 7.5). □

Лекция 8

Комбинаторика–3. Биномиальные коэффициенты и их друзья

8.1. Монотонные пути в квадранте

Мы уже рассматривали задачу подсчёта монотонных путей на прямой. Если двигаться можно только на 1 вправо, то ответ в такой задаче тривиальный: для любого n есть ровно один способ попасть из 0 в n.

Задача становится интереснее, если рассматривать монотонные пути на плоскости. Теперь мы двигаем фишку по точкам плоскости с целыми координатами. За один шаг можно увеличить абсциссу на 1 или ординату на 1. Сколько есть различных монотонных путей из точки $(0,0)$ в точку (a,b)? Обозначим это количество $T(a,b)$. Из правила суммы следует рекуррентное соотношение

$$T(a,b) = T(a-1,b) + T(a,b-1). \tag{8.1}$$

Действительно, все пути в (a,b) разбиваются на две группы: те, в которых на последнем шаге увеличивалась абсцисса, и те, в которых на последнем шаге увеличивалась ордината. Это первое и второе слагаемое в (8.1) соответственно.

Одной формулы (8.1) недостаточно для вычисления числа монотонных путей. Нужны ещё «граничные условия»:

$$T(0,b) = T(a,0) = 1 \tag{8.2}$$

(если возможно изменять лишь одну координату, путь единственный).

Пользуясь (8.1) и (8.2), можно посчитать количество монотонных путей для заданной пары (a,b) (при этом придётся решить ту же задачу и для всех пар (x,y), $x \leqslant a$, $y \leqslant b$). Вот несколько первых значений:

1	5	15	35	70
1	4	10	20	35
1	3	6	10	15
1	2	3	4	5
1	1	1	1	1

Есть и обычная, нерекуррентная формула для $T(a,b)$.

Теорема 8.1. $T(a,b) = \binom{a+b}{a} = \dfrac{(a+b)!}{a!b!}$.

Доказательство. Построим, как мы уже не раз делали, двудольный граф. Вершины левой доли — двоичные слова в алфавите $\{x,y\}$ длины $a+b$, а вершины правой доли — монотонные пути из $(0,0)$ в (a,b). Рёбра графа соединяют слово с путём, в котором на i-м шаге увеличивается та координата, которая записана в слове на i-м месте.

Этот граф — паросочетание: слово однозначно определяет путь, путь однозначно определяет слово. Количество слов в двоичном алфавите длины $a+b$, в которых a букв одного вида (и b букв другого) мы уже подсчитали: это как раз биномиальный коэффициент из $a+b$ по a. □

Биномиальные коэффициенты часто записывают в виде *треугольника Паскаля*

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

Легко увидеть, что это те же числа, что мы писали для точек квадранта, только теперь квадрант повёрнут вверх ногами (точнее, на угол 135°).

8.2. Свойства биномиальных коэффициентов

У нас появилось несколько равносильных способов определять биномиальные коэффициенты (они же числа сочетаний, они же — количество монотонных путей в квадранте). Это даёт возможность доказывать свойства биномиальных коэффициентов разными способами: привлекая комбинаторику, алгебру и даже анализ.

Рассмотрим начальную серию таких примеров.

Утверждение 8.2. *Каждая строка треугольника Паскаля симметрична относительно середины.*

Доказательство. В n-й строке треугольника Паскаля записаны биномиальные коэффициенты $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$.

Симметрия относительно середины означает равенство

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

Проще всего это равенство увидеть из формулы для числа сочетаний

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

(переставим сомножители в знаменателе). □

Не сложнее «натуральное доказательство» этого утверждения, которое сопоставляет k-элементному подмножеству n-элементного множества его дополнение, в котором как раз $n-k$ элементов.

Утверждение 8.3. *В первой половине строки треугольника Паскаля числа возрастают.*

Доказательство. И здесь нетрудно воспользоваться формулой для числа сочетаний. Но даже легче использовать двойной подсчёт.

Построим двудольный граф. Вершины левой доли — k-элементные подмножества n-элементного множества $[n]$; вершины правой доли — $(k+1)$-элементные подмножества того же множества. Рёбра — это пары (A, B), для которых $A \subset B$.

Каждое k-элементное подмножество A возможно расширить до $(k+1)$-элементного подмножества $n-k$ способами (столько элементов не входят в A). Поэтому степени вершин левой доли равны $n-k$.

В каждом $(k+1)$-элементном подмножестве B есть $k+1$ подмножество размера k (чтобы получить такое нужно удалить один из элементов подмножества B, а их $k+1$). Значит,

$$\binom{n}{k} \cdot (n-k) = \binom{n}{k+1} \cdot (k+1),$$

что равносильно

$$\binom{n}{k} \Big/ \binom{n}{k+1} = \frac{k+1}{n-k} < 1$$

при $2k < n-1$.

Таким образом, биномиальные коэффициенты растут до тех пор, пока не выполнится неравенство $k \geqslant (n-1)/2$, а это случится как раз в середине строки треугольника Паскаля. □

Утверждение 8.4. *Каждое число в треугольнике Паскаля по крайней мере в* 2 *раза меньше, чем число, которое стоит под ним.*

Доказательство. Тут полезно представление биномиальных коэффициентов как числа монотонных путей в квадранте и рекуррентная формула для этого числа.

Под числом $\binom{n}{k}$ стоит число в $(n+2)$-й строке треугольника Паскаля, это биномиальный коэффициент $\binom{n+2}{k+1}$.

Первое число — $\binom{n}{k}$ — это количество монотонных путей из $(0,0)$ в $A = (k, n-k)$; второе — $\binom{n+2}{k+1}$ — из $(0,0)$ в $B = (k+1, n-k+1)$. Из точки A в точку B ведёт два монотонных пути (увеличиваем сначала абсциссу, а потом ординату или наоборот). Поэтому каждый путь в A продолжается двумя способами до пути в B (а есть ещё, конечно, пути в B, которые вообще не проходят через A). □

Утверждение 8.5. *Количество подмножеств n-элементного множества с нечётным количеством элементов равно количеству подмножеств n-элементного множества с чётным количеством элементов.*

Доказательство. Поскольку количество k-элементных подмножеств n-элементного множества равно биномиальному коэффициенту, по сути речь идёт о том, что знакочередующаяся сумма чисел в строке треугольника Паскаля равна 0.

Запишем формулу бинома

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

и подставим в неё $1 = x = -y$. В левой части получится 0. А в правой

$$\sum_{k=0}^{n} \binom{n}{k} 1^{n-k}(-1)^k = \sum_{k \text{ чётное}} \binom{n}{k} - \sum_{k \text{ нечётное}} \binom{n}{k},$$

т.е. как раз знакопеременная сумма чисел в строке треугольника Паскаля. □

8.3. Мультиномиальные коэффициенты

Вместо бинома рассмотрим n-ю степень суммы нескольких переменных. Она также раскладывается в сумму мономов. Но теперь это мономы от нескольких переменных. Записывать такие мономы сложнее. Мы используем следующую запись: моном $x_1^{a_1} x_2^{a_2} \dots x_k^{a_k}$ однозначно определяется последовательностью показателей $\alpha = (a_1, a_2, \dots, a_k)$, мы будем обозначать такой моном сокращённо как x^α. Разложение

$$(x_1 + x_2 + \dots + x_k)^n = \sum_{\alpha} \binom{n}{\alpha} x^\alpha \tag{8.3}$$

приводит сразу к двум интересным задачам перечислительной комбинаторики.

Первая состоит в нахождении формулы для *мультиномиальных коэффициентов*, т.е. коэффициентов в разложении 8.3. Приведём сразу ответ, а потом дадим два разных доказательства этой формулы.

Теорема 8.6. $\displaystyle\binom{n}{a_1, a_2, \dots, a_k} = \frac{n!}{a_1! a_2! \dots a_k!}$.

Эту теорему можно доказать двумя способами: алгебраическим и комбинаторным. Приведём сначала комбинаторное доказательство.

У мультиномиальных коэффициентов есть внятный комбинаторный смысл. Вспомним, как мы находили формулу для биномиальных коэффициентов. Мы раскладывали бином в два шага: сначала раскрывали скобки, а затем приводили подобные.

Что даёт раскрытие скобок в выражении (8.3)? Получаются слагаемые, каждое из которых имеет вид $x_1 x_2 x_2 \dots$. Это слово ξ в алфавите $\{x_1, \dots, x_k\}$, в котором n букв. После перестановок переменных из этого слова получается моном $x_1^{a_1} x_2^{a_2} \dots x_k^{a_k}$, где a_i — количество букв x_i в слове ξ.

Значит, мультиномиальный коэффициент $\binom{n}{a_1,\dots,a_n}$ равен количеству слов в алфавите $\{x_1, \dots, x_k\}$, длина которых равна n, а количество вхождений каждого символа задаётся числами $a_1, \dots, a_n$.

Комбинаторное доказательство теоремы 8.6. Построим двудольный граф. Вершины левой доли — слова в алфавите $\{x_1, \ldots, x_k\}$, в которых a_1 букв x_1, a_2 букв x_2, и т.д., $a_1 + \cdots + a_k = n$. Вершины правой доли — перестановки чисел от 1 до n.

Чтобы определить рёбра этого графа, разделим числа от 1 до n на k групп так, чтобы в j-й группе было ровно a_j чисел. Пусть первая группа состоит из чисел от 1 до a_1; вторая — от $a_1 + 1$ до $a_1 + a_2$; $\ldots$; k-я группа — числа от $a_1 + \cdots + a_{k-1} + 1$ до n.

Рёбрами в нашем графе будут пары (ξ, π), где слово ξ получается из перестановки $\pi = (\pi_1, \ldots, \pi_n)$ заменой каждого числа π_i на x_j, где j — номер группы, в которую входит это число.

Степень каждой вершины в правой доле этого графа равна 1 (по перестановке однозначно определяется смежное с ней слово).

Чтобы найти степень слова (вершины левой доли), обратим внимание на то, что если (ξ, π) — ребро графа, то вхождения буквы x_j в слово ξ заменяются в перестановке π на числа из j-й группы. Порядок этих чисел произвольный и каждое число должно быть использовано ровно один раз.

Всего есть $a_j!$ перестановок a_j чисел j-й группы, сохраняющих их места в перестановке π. По правилу произведения общее количество перестановок, связанных ребром со словом ξ равно $a_1!a_2! \ldots a_k!$. Подсчитывая число рёбер в построенном графе двумя способами, получаем равенство

$$\binom{n}{a_1 \ldots a_k} \big(a_1! \ldots a_k!\big) = n!\,,$$

что и требовалось. □

Теперь приведём доказательство, основанное на индукции по числу слагаемых в формуле 8.6.

Алгебраическое доказательство теоремы 8.6. Индукция по числу переменных k.

База индукции: $k = 2$. Это обычная формула для биномиальных коэффициентов, см. теорему 7.5. Поскольку $a_1 + a_2 = n$, то

$$\binom{n}{a_1, a_2} = C_n^{a_1} = \frac{n!}{a_1!(n - a_1)!} = \frac{n!}{a_1!a_2!}\ .$$

Шаг индукции. Предполагаем, что формула из теоремы 8.6 доказана для k. Представим $x_1 + \cdots + x_k + x_{k+1}$ как $(x_1 + \cdots + x_k) + x_{k+1}$ и запишем разложение n-й степени суммы $k + 1$ переменной как бинома от этих двух слагаемых:

$$(x_1 + x_2 + \cdots + x_k + x_{k+1})^n = \sum_{j=0}^{n} \binom{n}{j} (x_1 + x_2 + \cdots + x_k)^j x_{k+1}^{n-j}.$$

Теперь раскрываем скобки в множителях вида $(x_1+x_2+\cdots+x_k)^j$ и из индуктивного предположения получаем равенство

$$\binom{n}{a_1,\ldots,a_k,a_{k+1}} = \binom{n}{j}\binom{j}{a_1,\ldots,a_k} = \\ = \frac{n!}{j!(n-j)!}\cdot\frac{j!}{a_1!a_2!\ldots a_k!} = \frac{n!}{a_1!a_2!\ldots a_k!(n-j)!}\,.$$

Осталось заметить, что $n-j = n-a_1-\cdots-a_k = a_{k+1}$. □

Рассмотрим типичный пример применения формулы для мультиномиальных коэффициентов.

Пример 8.1. Три человека составляют график дежурств на 6 дней. Каждый день дежурит кто-то один, каждый должен дежурить ровно 2 дня. Сколько есть вариантов составления графика?

Обозначим людей А, Б, В. Каждый график дежурств представляется словом в алфавите $\{\text{А},\text{Б},\text{В}\}$: слово ААББВВ означает, что в первые два дня дежурит А, в следующий день — Б и т.д. Таким образом, вариантов составления графика столько же, сколько слов в 3-элементном алфавите длины 6, в которые каждая буква входит ровно два раза. Это и есть мультиномиальный коэффициент

$$\binom{6!}{2!2!2!} = 90.$$

8.4. Сочетания с повторениями

Вторая перечислительная задача, связанная с разложением степени суммы (8.3) состоит в подсчёте числа различных мономов в этом разложении. Степень монома — это сумма степеней переменных, в него входящих. Степень монома $x_1x_2x_3$ равна 3, степень монома $x_1^2x_2$ также равна 3. А сколько всего мономов от трёх переменных степени 3? На такой вопрос нетрудно ответить, выписав все нужные мономы:

$$x_1^3,\ x_2^3,\ x_3^3,\ x_1^2x_2,\ x_1^2x_3,\ x_1x_2^2,\ x_2^2x_3,\ x_1x_3^2,\ x_2x_3^2,\ x_1x_2x_3.$$

Но даже в этом случае нужно проверить, что мы ничего не пропустили и эти 10 мономов и есть все возможные варианты.

Сформулируем задачу точно. Моном $x_1^{a_1}x_2^{a_2}\ldots x_k^{a_k}$ имеет степень $a_1+\cdots+a_k$ и мономы совпадают тогда и только тогда, когда соответствующие последовательности показателей равны. Поэтому нам нужно найти количество решений уравнения

$$a_1+\cdots+a_k = n \tag{8.4}$$

в неотрицательных целых числах. Это число традиционно называется *числом сочетаний с повторениями* из n по k. Обозначим его

$$\left(\!\binom{n}{k}\!\right)\,.$$

Название можно объяснить примером с лототроном, который рассматривался выше. Теперь мы каждый раз возвращаем выпавший шар в лототрон, поэтому шары могут повторяться. Но нам неважно, как и раньше, в каком порядке выпадали эти шары.

Ещё одно популярное представление чисел с повторениями даётся в формулировке следующей задачи.

Задача 8.2. Сколько есть вариантов разделить n одинаковых монет между k людьми?

Это то же самое число сочетаний с повторениями из n по k. Действительно, каждый вариант делёжки задаётся указанием числа a_1 монет, которые получает первый человек, числа a_2 монет, которые получает второй и т.д.

Теорема 8.7. $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k-1}$.

Доказательство. Установим взаимно однозначное соответствие между решениями уравнения (8.4) и $(k-1)$-элементными подмножествами $(n+k-1)$-элементного множества. Будем делать это, используя терминологию задачи о разделе монет.

Выстроим монеты в ряд и разделим их перегородками, чтобы указать, кому какие монеты отходят. Первый получает монеты, которые расположены до первой перегородки, второй — те, которые лежат между первой и второй, и т.д. Например, на рисунке 8.1 показан раздел 7 монет между 7 людьми, при котором первому не

Рис. 8.1: $a_1 = 0$, $a_2 = 2$, $a_3 = 0$, $a_4 = 2$, $a_5 = 0$, $a_6 = 1$, $a_7 = 2$

достаётся ничего, второму — две монеты, третьему — ничего, четвёртому — две монеты, пятому — ничего, шестому — одна, а седьмому — две монеты.

Такой раздел отвечает решению уравнения (8.4), которое указано в подписи под рисунком.

Последний человек получает монеты, которые лежат после последней перегородки. Поэтому для 7 людей нужно всего 6 перегородок. А в общем случае, когда людей k, нужна $k-1$ перегородка.

Итак, у нас есть позиции, на каждую из которых можно поставить либо монету, либо перегородку. Всего позиций $n+k-1$, а перегородок — $k-1$. Любой выбор $(k-1)$-элементного подмножества позиций, на котором стоят перегородки, возможен, и каждому такому выбору отвечает ровно одно решение уравнения (8.4). Получаем искомое соответствие. □

Если монеты выглядят неубедительно, то можно использовать то же по сути рассуждение, но в другой комбинаторной ситуации.

Задача 8.3. Сколько есть монотонных путей длины k по прямой из 0 в n?

Решение. Монотонный путь длины k по прямой из 0 в n — это другое название такой строго возрастающей последовательности целых чисел $x_1 < \cdots < x_{k+1}$, что $x_1 = 0$, $x_{k+1} = n$ (длина пути — это количество шагов, оно на 1 меньше количества членов последовательности).

Такой монотонный путь однозначно задаётся выбором $k-1$ числа в интервале от 1 до $n-1$ (путь монотонный, поэтому эти числа он обязан проходить в порядке возрастания).

Поэтому число таких путей равно числу $(k-1)$-элементных подмножеств $(n-1)$-элементного множества, то есть $\binom{n-1}{k-1}$. □

А какое отношение имеют эти пути к нашей задаче о числе решений уравнения (8.4)? Заметим, что путь однозначно задаётся последовательностью длин ходов: $\ell_1 = x_2 - x_1 = x_2$, $\ldots \ell_k = x_{k+1} - x_k = n - x_k$. В сумме эти числа обязаны давать n:

$$\ell_1 + \cdots + \ell_k = n. \tag{8.5}$$

Получаем то же самое уравнение. Вот только ответ почему-то другой получился...

Разница между уравнениями (8.4) и (8.5) в том, какие решения мы подсчитываем. В первом случае мы искали решения в *неотрицательных целых* числах. А во втором нам нужны решения в *положительных целых* числах: ведь стоять на месте не разрешается, все длины ходов должны быть положительными.

Чтобы связать эти два числа, заметим, что если $a_1 + \cdots + a_k = n$ — решение уравнения (8.4) в неотрицательных числах, то ему однозначно соответствует решение уравнения

$$\ell_1 + \cdots + \ell_k = n + k \tag{8.6}$$

в положительных числах: $\ell_i = a_i + 1$. И наоборот, по решению $\ell_1, \ldots, \ell_k$ уравнения (8.6) в положительных целых числах однозначно строится решение уравнения (8.4) в неотрицательных целых числах: $a_i = \ell_i - 1$.

Из решения задачи 8.3 получаем ту же самую формулу

$$\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k-1}$$

для числа сочетаний с повторениями.

8.5. Числа Каталана

Приведём ещё один пример перечислительной задачи, где в ответе возникают биномиальные коэффициенты.

Задача 8.4. Робот ходит по целочисленным точкам координатной плоскости, удовлетворяющим условиям $x \geqslant 0$, $y \geqslant 0$, $x \geqslant y$. На каждом шаге он может увеличить одну из координат на 1. Докажите, что количество способов, которыми можно переместить Робота из точки $(0,0)$ в точку (n,n), равно

$$\frac{1}{n+1}\binom{2n}{n}.$$

Числа, которые возникают в ответе этой задачи, называются *числами Каталана*. Это последовательность чисел является ответом в очень многих задачах перечислительной комбинаторики. Монотонные пути по треугольнику — не самая известная из них.

Подробнее про числа Каталана можно прочитать в учебнике «Лекции по дискретной математике» (М. Вялый, В. Подольский, А. Рубцов, Д. Шварц, А. Шень), раздел 2.10; черновой вариант доступен по ссылке
`publications.hse.ru/mirror/pubs/share/direct/393719078.pdf`.

Ещё больше различных комбинаторных представлений чисел Каталана (более 30) можно найти в материалах 14-й летней конференции международного математического Турнира городов, `olympiads.mccme.ru/lktg/2002/problem2.ru/index.htm`.

Но и этот список далеко не полный.

Лекция 9

Отношения и функции–1

9.1. Бинарные отношения

Язык двудольных графов не всегда удобен. Оказывается полезным ввести более общее понятие.

Определение 9.1. *Бинарным отношением на множествах A и B* называется любое подмножество R декартова произведения $A \times B$.

Двудольный граф на долях L и R задаёт бинарное отношение на множествах A и B. Но в общем случае множества A и B не обязаны быть дизъюнктными (без общих элементов). Во многих интересных случаях это попросту одно и то множество. Тогда говорим о бинарном отношении на множестве.

Пример 9.1. Рассмотрим множество пар действительных чисел (x, y), для которых $x < y$. Это множество по определению задаёт бинарное отношение «строго меньше» на действительных числах.

Пример 9.2. «x является родителем y» — это бинарное отношение на множестве людей.

«x является братом y» — это бинарное отношение на множестве мужчин и множестве людей. В данном случае множества не совпадают, но пересекаются.

Часто бинарные отношения обозначаются как $R(x, y)$ или даже как $x\ R\ y$. Эта запись указывает на утверждение «x и y находятся в отношении R» или, более формально, $(x, y) \in R$. В общем случае такое утверждение истинно для одних пар (x, y) и ложно для других.

Отношения на конечных множествах, как и графы, удобно задавать матрицами. Строки такой матрицы обычно отвечают первому множеству в отношении, а столбцы — второму. На пересечении строки a и столбца b стоит 1, если $(a, b) \in R$, и 0 в противном случае. Часто удобно считать, что это не числа, а логические значения: 1 отвечает истинности высказывания $(a, b) \in R$, а 0 — ложности.

Разумеется, нет принципиальной разницы между строками и столбцами. Если транспонировать матрицу отношения, получим *транспонированное отношение*: $R^T = \{(b,a) : (a,b) \in R\} \subseteq B \times A$.

Пример 9.3. «x является ребёнком y» — это бинарное отношение на множестве людей, которое транспонировано к отношению «x является родителем y».

Из одних отношений можно строить другие как с помощью теоретико-множественных операций, так и с помощью операции композиции.

Определение 9.2. Пусть даны два отношения $R \subset A \times B$ и $S \subset B \times C$. Их *композицией* называется отношение $S \circ R \subseteq A \times C$, определяемое так:

$$(x,z) \in S \circ R \;\Leftrightarrow\; \text{существует такой } y \in B, \text{ что } (x,y) \in R \text{ и } (y,z) \in S.$$

Обратите внимание, что (1) композиция определена не для всех отношений, необходимо совпадение второго множества в первом отношении и первого во втором; (2) порядок записи отношений в композиции важен, ниже станет ясно, почему выбран именно такой порядок.

Если задавать отношения на конечных множествах матрицами, композиция отношений задаётся формулой, напоминающей формулу произведения числовых матриц, но операции в этой формуле логические:

$$(S \circ R)(x,y) = \bigvee_{z \in B} \big(R(x,z) \wedge S(z,y)\big).^{1)}$$

Как и умножение числовых матриц, композиция отношения обладает свойством ассоциативности.

Лемма 9.3. *Если $R \subset A \times B$, $S \subset B \times C$ и $T \subset C \times D$, то $(R \circ S) \circ T = R \circ (S \circ T)$.*

Доказательство. Обе части равенства задают отношение M, для которого $M(x,t)$ равносильно тому, что найдутся такие $y \in B$ и $z \in C$, что одновременно $R(x,y)$, $S(y,z)$ и $T(z,t)$:

$$M(x,t) \Leftrightarrow (\exists\, y \in B)(\exists\, z \in C)[R(x,y) \wedge S(y,z) \wedge T(z,t)].$$

Обе импликации непосредственно следуют из определения композиции. □

Рассмотрим несколько примеров.

Пример 9.4. Чему равна композиция отношения $P =$«x является родителем y» с самим собой? Это отношение «x является дедом или бабушкой y». Проверим это формально, исходя из определения.

$(x,y) \in P \circ P$ тогда и только тогда, когда найдётся такой z, что $(x,z) \in P$ и $(z,y) \in P$. Это означает, что x — родитель z, а z — родитель y. В обычном языке такое свойство выражается как раз словами «дед» или «бабушка», в зависимости от пола x.

[1)] Спасибо Ксении Петренко за замеченную опечатку.

Пример 9.5. Для отношения $x < y$ строгого сравнения действительных чисел выполняется $< \circ < = <$. Проверим это формально.

$(x, y) \in < \circ <$ тогда и только тогда, когда найдётся такой z, что $x < z$ и $z < y$. По свойству сравнения действительных чисел получаем $x < y$, откуда следует включение $< \circ < \subseteq <$.

Чтобы доказать равенство, нужно проверить обратное включение. Для этого нужно доказать, что если $x < y$, то найдется такой z, что $x < z$, $z < y$. В качестве числа z можно взять, например, полусумму $(x + y)/2$.

Пример 9.6. Композиция отношения $B =$«x является братом y» с самим собой не определена, поскольку это отношение на разных множествах. Но определена композиция $B \circ B^T$. Это отношение на множестве мужчин. Совпадает ли оно совпадает с отношением братства, если ограничить его на множество мужчин?

Будем обозначать такое ограниченное отношение той же буквой B. Легко видеть, что B *симмметрично*: $B^T = B$ (на множестве мужчин справедливо утверждение «если x брат y, то y — брат x»).

Легко видеть, что $B \circ B \subseteq B$. Если $(x, y) \in B \circ B$, то существует такой z, что x брат z, а z брат y. Но брат моего брата — мой брат[2)]. Поэтому $(x, y) \in B$.

Обратное включение может и не выполняться. Допустим x и y — братья, но других братьев у них нет. Тогда $(x, y) \in B$, но $(x, y) \notin B \circ B$. Здесь мы считаем, следуя смыслу слов в естественном языке, что отношение братства *антирефлексивно*: $(x, x) \notin B$ для любого x (я не считаю себя своим братом).

9.2. Более общие отношения

Мы рассмотрели бинарные отношения. Есть более общее понятие отношения R на нескольких множествах A_1, A_2,..., A_k. Это по определению подмножество декартова произведения $A_1 \times A_2 \times \cdots \times A_k$. Такое отношение называют k-арным или отношением валентности k.

Наиболее важны *унарные* отношения, то есть подмножества множества и бинарные, которые мы рассмотрели выше. Однако встречаются и отношения большей арности, как в математике, так и в приложениях.

Пример 9.7. Тернарное отношение $S(x, y, z)$ на действительных числах — это множество таких троек (x, y, z), что $x = y + z$. Последняя запись уже может рассматриваться как задание отношения. Это равенство истинно или ложно для любой тройки чисел.

Аналогично можно рассматривать отношение «(x, y, z) — координаты точки на единичной сфере» и т.п.

Пример 9.8. Результаты проверки домашнего задания являются тернарным отношением «x получил за задачу y оценку z». на множествах `Студенты`, `Задачи`, `Оценки`.

[2)] Такое свойство отношений называется *транзитивностью*.

В реляционных базах данных аналогичным образом используются и отношения бóльшей арности.

Мы не будем подробно разбираться с отношениями большой арности. Ограничимся несколькими важными классами бинарных отношений.

9.3. Функции

Важнейшее для математики понятие функции естественно определяется в терминах бинарных отношений.

Неформально под функцией мы понимаем соответствие: элементам одного множества сопоставляются элементы другого множества, причём каждому элементу ставится в соответствие не более одного элемента второго множества.

Дадим формальные определения понятий, связанных с функциями.

Определение 9.4. *Функцией* f из множества A в множество B (обозначение $f\colon A \to B$) называется такое бинарное отношение $f \subseteq A \times B$, что для каждого $a \in A$ есть не более одной пары $(a, b) \in \Gamma_f$. Если такая пара существует, используется также обозначение $b = f(a)$.

Элементы множества A называются *аргументами* функции, элементы множества B — значениями функции.

Область определения $\operatorname{Dom} f$ функции из A в B — это множество тех a, для которых существует такой b, что $(a, b) \in f$. Формальная запись этого определения:

$$\operatorname{Dom}(f) = \{x \in A \mid \exists y \in B\colon y = f(x)\}.$$

Если $\operatorname{Dom}(f) = A$, то функция называется *тотальной* (всюду определённой). Нетотальные функции называют *частичными*.

Область значений $\operatorname{Range} f$ — это множество тех b, для которых существует такой a, что $(a, b) \in \Gamma_f$. Формальная запись этого определения:

$$\operatorname{Range}(f) = \{y \in B \mid \exists x \in A\colon y = f(x)\}.$$

Замечание 9.1. Отношение f называют ещё *графиком функции* и иногда обозначают Γ_f, чтобы различать график и саму функцию. При нашем определении функция и график — это одно и то же. Но в рассуждениях о функциях исторически принято использовать оба термина.

Неформально, мы говорим о функции, если речь идёт о соответствии («значение функции от данного аргумента»), и о графике, если речь идёт об отношении целиком («график линейной функции — прямая»).

Замечание 9.2. Тотальные функции мы будем также называть *отображениями*. Функция и отображение, в сущности, синонимы, но в разных областях математики преимущественно употребляется тот или иной термин.

Замечание 9.3. Если применить операцию композиции к функциям, то получится функция. В привычных обозначениях значение композиции $(f \circ g)(x)$ равно $f\big(g(x)\big)$ (и требуется, чтобы $x \in \operatorname{Dom} g$, а $g(x) \in \operatorname{Dom} f$. Наш выбор порядка отношений в композиции согласован с обозначением $f\big(g(x)\big)$.

С функцией из A в B можно связать двудольный граф с долями A и B. Ребро такого графа связывает a и $f(a)$. Двудольные графы, отвечающие функциям, и только они обладают таким свойством: из каждой вершины доли A выходит не более одного ребра.

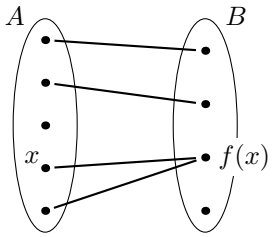


Рис. 9.1: функция

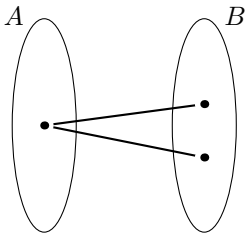


Рис. 9.2: не функция

Эта конструкция корректна, если A и B — непересекающиеся множества. Если у них есть общие элементы при построении графа нужно мысленно «дублировать» общие элементы A и B, отправляя один в левую долю, а другой — в правую. Именно поэтому удобнее использовать язык бинарных отношений.

С функциями из A в A (множество аргументов и значений совпадают) можно также связать ориентированный граф, рёбрами которого являются пары $(x, y) \in f$. Исходящая степень вершины в таком графе не больше 1.

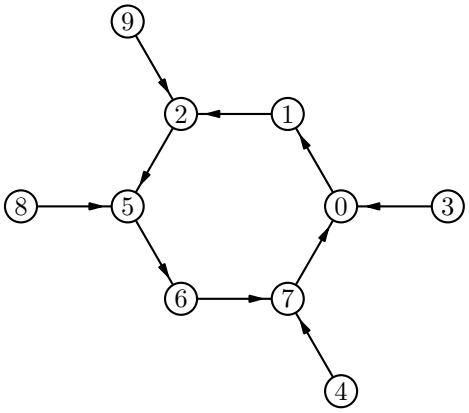


Рис. 9.3: Ориентированный граф функции

9.4. Сюръекции, инъекции, биекции

Из тотальных функций $A \to B$ выделяются некоторые функции с особыми свойствами.

Транспонированное к графику отношение не является в общем случае функцией. Если $y = f(x_1) = f(x_2)$ и $x_1 \neq x_2$, то $(x_1, y) \in \Gamma_f$, $(x_2, y) \in f$. Поэтому $(y, f_1) \in f^T$ и $(y, f_2) \in f^T$, что противоречит определению (графика) функции.

Тотальная функция $f\colon A \to B$ называется *инъекцией*, если транспонированное отношение f^T также является функцией. Другими словами, значения инъекции в различных точках различны. Пересказывая определения функции и транспонированного отношения, получаем такое равносильное определение: f — инъекция, если $x_1 \neq x_2$ влечёт $f(x_1) \neq f(x_2)$. (Или контрапозиция: f — инъекция, если $f(x_1) = f(x_2)$ влечёт $x_1 = x_2$.)

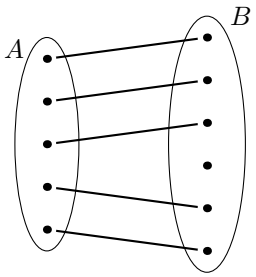


Рис. 9.4: инъекция

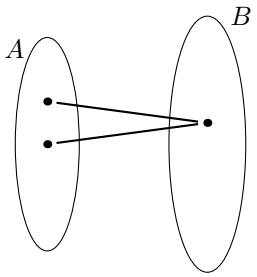


Рис. 9.5: не инъекция

Тотальная функция $f\colon A \to B$ называется *сюръекцией*, если область значений совпадает со всем множеством B, то есть если для всякого элемента $y \in B$ найдётся элемент $x \in A$, для которого $f(x) = y$.

Пример сюръекции изображён на рис. 9.6. Примеры не сюръекций изображены на рисунке 9.7 (а также на рисунке 9.4): для сюръективной функции не должно быть точек справа, в которые не ведёт ни одного ребра.

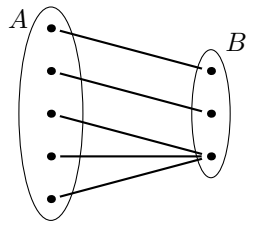


Рис. 9.6: сюръекция

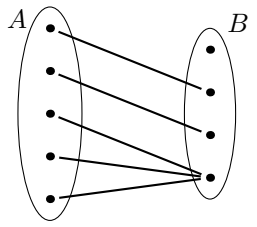


Рис. 9.7: не сюръекция

Наконец, тотальная функция $f\colon A \to B$ называется *биекцией*, если она одновременно является и инъекцией, и сюръекцией. Другими словами, функция является

биекцией, если всякому элементу из B соответствует ровно один элемент из A.

Лемма 9.5. *Для тотальных функций из конечного множества в конечное выполняются такие свойства:*

1. *если $f\colon A \to B$ сюръекция, то $|A| \geqslant |B|$;*
2. *если $f\colon A \to B$ инъекция, то $|A| \leqslant |B|$;*
3. *если $f\colon A \to B$ биекция, то $|A| = |B|$.*

Доказательство. Первый факт утверждает неформально, что если рассадить кроликов по клеткам и в каждой клетке есть хотя бы один кролик, то кроликов не меньше, чем клеток.

Формально: обозначим через a_i, $i \in B$, количество тех элементов $a \in A$, для которых $f(a) = i$ (размер *полного прообраза* множества $\{i\}$). Для сюръекции $a_i \geqslant 1$ для любого i. Поэтому

$$|A| = \sum_i a_i \geqslant \sum_i 1 = |B|.$$

Первое равенство выполняется потому, что a_i подсчитывают элементы в непересекающихся подмножествах A (функция сопоставляет каждому элементу A только один элемент B).

Второй факт означает неформально, что если в комнате есть люди и стулья, и каждый человек сел на стул, то людей не больше, чем стульев.

Формальное доказательство получается из выражения свойства инъекции в виде $a_i \leqslant 1$ для любого i. Тогда

$$|A| = \sum_i a_i \leqslant \sum_i 1 = |B|.$$

Третий факт непосредственно следует из первых двух. □

Когда мы строили взаимно однозначные соответствия, мы фактически задавали биекции. В отличие от двудольных графов, в общем случае нет нужды следить за тем, чтобы множества, между которыми строится биекция, не пересекались. Они даже могут совпадать.

Что такое биекция множества A на себя? Посмотрим на ориентированный граф такой функции. Исходящая степень каждой вершины равна 1 (функция тотальная) и входящая степень каждой вершины равна 1 (сюръективна и инъективна). Мы такие графы полностью расклассифицировали: это объединения ориентированных циклов.

Если $A = [n]$, то биекциям $[n] \to [n]$ взаимно однозначно соответствуют перестановки. Биекции $f\colon [n] \to [n]$ сопоставляем перестановку $f(1)f(2)\dots f(n)$ (каждое число использовано и ровно по одному разу). Поэтому биекции $[n] \to [n]$ часто также называют перестановками. Ориентированный граф такой перестановки называется *цикловым разложением.*

9.5. Индикаторные функции

Если множество значений функции — действительные числа, с такими функциями можно выполнять арифметические действия «поточечно»:

$$(f+g)(x) = f(x) + g(x), \quad (fg)(x) = f(x)g(x).$$

Приведём один важный пример. *Индикаторная функция* $\chi_S \colon X \to \mathbb{R}$ подмножества S множества X опредедяется как

$$\chi_S(x) = \begin{cases} 1, & \text{если } x \in S, \\ 0 & \text{в противном случае.} \end{cases}$$

Индикаторные функции χ_{S_1} и χ_{S_2} равны тогда и только тогда, когда подмножества S_1 и S_2 равны. Таким образом, индикаторные функции и подмножества — это два способа говорить об одном и том же.

Покажем, как на языке индикаторных функций доказывается формула включений-исключений. Будем считать, что все рассматриваемые множества лежат в каком-то объемлющем множестве (*универсуме*), например, в объединении всех множеств, для которых доказывается формула.

Теоретико-множественные операции выражаются через индикаторные функции. Для пересечения

$$\chi_{A\cap B}(x) = \chi_A(x) \cdot \chi_B(x)$$

(x принадлежит пересечению тогда и только тогда, когда $x \in A$ и $x \in B$). Для дополнения получаем

$$\chi_{\overline{A}}(x) = 1 - \chi_A(x),$$

для разности

$$\chi_{A\setminus B}(x) = \chi_A(x) \cdot \big(1 - \chi_B(x)\big),$$

для объединения

$$\chi_{A\cup B}(x) = 1 - \big(1 - \chi_A(x)\big)\big(1 - \chi_B(x)\big). \tag{9.1}$$

Доказать последнюю формулу можно как разбором случаев, так и с помощью логических формул де Моргана: $A \cup B = \overline{\overline{A} \cap \overline{B}}$. Формула (9.1) получается из предыдущих заменой левой части формулы де Моргана на правую.

Аналогично можно выразить и объединение нескольких множеств через дополнения и переесечения:

$$\bigcup_{i=1}^{n} A_i = \overline{\bigcap_{i=1}^{n} \overline{A}_i} \tag{9.2}$$

(в объединение множеств A_i входят в точности те элементы, которые не входят в пересечение всех дополнений).

Обозначим $A = \bigcup_{i=1}^{n} A_i$. Для индикаторных функций из (9.2) получаем

$$\chi_A(x) = 1 - \big(1 - \chi_{A_1}(x)\big)\big(1 - \chi_{A_2}(x)\big) \ldots \big(1 - \chi_{A_n}(x)\big). \tag{9.3}$$

Раскроем скобки в (9.3) и заменим произведения индикаторных функций на индикаторные функции пересечений. Для удобства записи через A_S обозначим пересечение всех множеств, входящих в семейство S, то есть

$$A_S = \bigcap_{i \in S} A_i.$$

Получаем

$$\chi_A(x) = \sum_{S \neq \varnothing} (-1)^{|S|+1} \chi_{A_S}(x)$$

(перед произведением стоит минус, в каждой скобке стоит минус; поэтому коэффициент при произведении k множителей равен $(-1)^{k+1}$).

Количество элементов в множестве выражается как сумма индикаторной функции по всему универсуму:

$$|A| = \sum_u \chi_A(u)$$

(каждый элемент множества даёт вклад 1 в сумму, остальные элемементы дают вклад 0).

Поэтому для размера объединения получается формула

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{S \neq \varnothing} (-1)^{|S|+1} |A_S|,$$

это и есть формула включений-исключений.

Лекция 10

Отношения и функции–2

10.1. Композиции функций

Функции мы определяем как особый вид отношений. Ограничение операции композиции отношений на функции даёт операцию *композиции функций*.

Разворачивая определения функции и композиции отношений, получаем: для функции f из множества A в множество B и функции g из множества B в множество C *композицией* $g \circ f$ этих функций является такая функция из A в C, которая определена на тех x из области определения функции f, для которых $f(x)$ принадлежит области определения функции g, и равна $g(f(x))$. Это действительно функция, а не более общее отношение, так как каждое значение $x \in A$ аргумента находится в отношении не более чем с одним элементом из C.

Отсюда становится понятным порядок записи отношений в композиции. Мы хотим, чтобы он был согласован с порядком записи функций в привычном обозначении $g(f(x))$.

Как и в общем случае, композиция функций ассоциативна:

$$(f \circ g) \circ h = f \circ (g \circ h).$$

10.2. Обратная функция

Тождественной функцией на множестве A (или тождественным отображением множества A в себя) называется функция $\mathrm{id}_A \colon A \to A$, которая отображает всякий элемент $x \in A$ в себя: $\mathrm{id}_A(x) = x$. При композиции тождественные функции ведут себя, как единица при умножении: для любого отображения $f \colon A \to B$ выполнены равенства

$$\mathrm{id}_B \circ f = f \circ \mathrm{id}_A = f.$$

(Обратите внимание, что здесь две тождественные функции — одна на A, другая на B, иначе композицию нельзя определить.)

Для биекции $f \colon A \to B$ (взаимно однозначного отображения) определена *обратная функция* (или обратное отображение) f^{-1}: если f отображает x в y, то обратная

функция f^{-1} отображает y в x. Инъективность f гарантирует, что это действительно функция, а сюръективность f гарантирует, что эта функция определена на всём B.

Заметим, что определение обратной функции симметрично: если g обратна к f, то и f обратна к g.

Свойство биективности может быть выражено алгебраическими средствами.

Теорема 10.1. *Если для отображений $f\colon A \to B$ и $g\colon B \to A$ выполнены два равенства $g \circ f = \mathrm{id}_A$ и $f \circ g = \mathrm{id}_B$, то функция f является биекцией и g обратна к f.*

Доказательство. Пусть $f(x_1) = f(x_2)$. Тогда из первого условия на композиции получаем:

$$x_1 = (g \circ f)(x_1) = g(f(x_1)) = g(f(x_2)) = (g \circ f)(x_2) = x_2.$$

Значит, функция f инъективна.

Для любого $y \in B$ из второго условия на композиции следует, что $y = f(g(y))$, то есть y принадлежит множеству значений f. Значит, функция f сюръективна.

Итак, f биекция.

Если $y = f(x)$, то из первого условия на композиции получаем $g(y) = g(f(x)) = x$. Значит, g обратна к f. □

Выполнения одного условия $g \circ f = \mathrm{id}_A$ недостаточно, чтобы утверждать, что f биекция.

Пример 10.1. Определим две функции на множестве неотрицательных целых чисел: $f\colon x \mapsto 2x$ и $g\colon x \mapsto \lfloor x/2 \rfloor$. Композиция $g \circ f$ — тождественное отображение:

$$x \xrightarrow{f} 2x \xrightarrow{g} x$$

для любого x. Однако f не биекция: у числа 1 нет прообраза.

Аналогичное замечание справедливо и для второго условия $f \circ g = \mathrm{id}_B$.

Ещё один важный факт: композиции сохраняют биекции. Другими словами,

Теорема 10.2. *Если отображения $f\colon A \to B$ и $g\colon B \to C$ — биекции, то и их композиция $g \circ f\colon A \to C$ является биекцией.*

Доказательство. Разные элементы A переходят в разные элементы B, потому что f инъекция, и эти разные элементы B переходят в разные элементы C, потому что g инъекция. Таким образом, при $a \neq a'$ получаем $f(a) \neq f(a')$ и затем $g(f(a)) \neq g(f(a'))$, так что $g \circ f$ — инъекция. Ещё надо проверить, что $g \circ f$ — сюръекция, то есть что в каждый элемент $c \in C$ что-то переходит. Но мы знаем, что g — сюръекция, так что $c = g(b)$ для некоторого $b \in B$; поскольку f — сюръекция, то $b = f(a)$[1)] для некоторого $a \in A$, так что $c = g(b) = g(f(a)) = (g \circ f)(a)$. □

[1)] Спасибо Нине Челышевой за замеченную опечатку.

10.3. Подсчёты числа функций

Пример 10.2. Сколько есть тотальных функций из k-элементного множества в n-элементное?

Ответ: ровно столько же, сколько есть слов длины k в алфавите из n символов.

Пусть $|A| = k$, $|B| = n$. Занумеруем элементы A: $a_1, a_2, \ldots, a_k$.

Теперь сопоставим тотальной функции $f\colon A \to B$ слово $\beta(f) = b_1 b_2 \ldots b_k$ длины k в алфавите B по правилу: $b_i = f(a_i)$. Это инъекция: разным функциям сопоставлены разные слова, так как из $f(a_j) \neq g(a_j)$ следует, что $\beta(f)_j \neq \beta(g)_j$.

Но это и сюръекция: по слову $b = b_1 b_2 \ldots b_k$ однозначно определяется такая функция f, что $\beta(f) = b$. Нужно положить $f(a_j) = b_j$.

Значит, количество тотальных функций из A в B равно количеству слов длины k в алфавите из n символов и равно n^k.

Заметим, что множество тотальных функций из A в B обозначается B^A для любых множеств. Этот пример показывает причину такого обозначения.

Пример 10.3. Сколько есть всех функций из k-элементного множества в n-элементное?

Ответ теперь другой, поскольку нужно посчитать и частичные функции. Это нетрудно сделать. Давайте добавим элемент $\texttt{void} \notin B$.

Тотальные функции из A в $B \cup \{\texttt{void}\}$ находятся во очевидном взаимно однозначном соответствии с функциями из A в B: значение `void` мы рассматриваем как указание на то, что функция из A в B не определена.

Ответ: $(n+1)^k$.

Пример 10.4. Сколько есть *инъекций* из k-элементного множества в n-элементное?

Достаточно рассмотреть случай $A = [k]$, $B = [n]$.

Сопоставим инъекции $f\colon A \to B$ слово в алфавите B длины k: $f(1)f(2)\ldots f(k)$. Поскольку f — инъекция, все символы в этом слове разные. Это соответствие взаимно однозначно. Поэтому искомое количество инъекций равно числу размещений $A_{n,k} = n!/(n-k)!$.

Пример 10.5. Сколько есть *биекций* из n-элементного множества в n-элементное? Уже обсуждалось, что их столько же, сколько перестановок, то есть $n!$.

Труднее подсчитать количество сюръекций.

Теорема 10.3. *Количество сюръекций k-элементного множества в n-элементное при $k \geqslant n$ равно*

$$\sum_{p=0}^{n} (-1)^p \binom{n}{p} (n-p)^k = n^k - \sum_{p=1}^{n} (-1)^{p+1} \binom{n}{p} (n-p)^k$$

и равно нулю при $k < n$.[2)]

[2)] Спасибо Анне Балашовой за замеченную опечатку.

Доказательство. Воспользоваться формулой включений и исключений. Для этого удобно выделить первое слагаемое суммы. Получаем правую часть равенства теоремы. Чтобы найти количество сюръекций, нужно из всего количества тотальных функций, их n^k, вычесть количество не-сюръекций.

Пусть B состоит из n элементов $b_1, \ldots, b_n$. Не-сюръекции $A \to B$ — это те тотальные функции, область значений которых не содержит одного из элементов $b_1, \ldots, b_n$, то есть объединение множеств

$$A(b_1) \cup A(b_2) \cup \ldots \cup A(b_n),$$

где через $A(b)$ обозначается множество тех функций, которые не принимают значения b.

Все множества $A(b)$ имеют размер $(n-1)^k$ (мы выбросили одно из возможных значений функции, поэтому количество таких функций равно количеству тотальных функций из k-элементного множества в $(n-1)$-элементное).

Для формулы включений и исключений нужно ещё подсчитать размер пересечений таких множеств. Рассмотрим пересечение p множеств вида $A(b)$. Это функции, которые не принимают некоторые p значений. Таких функций столько же, сколько тотальных функций из k-элементного множества в $(n-p)$-элементное.

А всего разных наборов из p множеств вида $A(b)$ столько же, сколько p-элементных подмножеств n-элементного множества, то есть $\binom{n}{p}$. Поэтому формула включений и исключений для данного семейства множеств приобретает вид, указанный в теореме. □

10.4. Отношения эквивалентности

Определение 10.4. Отношение R на некотором множестве A, которое одновременно *рефлексивно*: xRx для всех $x \in A$, *симметрично*: если xRy, то yRx для всех $x, y \in A$ и *транзитивно*: если xRy и yRz, то xRz для всех $x, y, z \in A$, называют *отношением эквивалентности*.

Пример 10.6. Отношение $x \in C(v)$, где $C(v)$ — область достижимости вершины v простого неориентированного графа, является отношением эквивалентности на множестве вершин графа. Это мы уже проверяли.

Пример 10.7. Отношение, состоящее из пар (u, v) сильно связанных вершин в орграфе, также является отношением эквивалентности. Это мы уже проверяли.

Пример 10.8. Отношение, состоящее из пар (x, x), $x \in A$, является отношением эквивалентности. Рефлексивность, симметричность и транзитивность очевидны.

Последний пример — отношение *равенства* — является прототипом всех отношений эквивалентности. Его можно обобщить.

Пример 10.9. Пусть A разбито в дизъюнктное объединение семейства множеств A_i:

$$A = \bigcup_i A_i, \qquad A_i \cap A_j = \varnothing, \text{ если } i \neq j. \quad {}^{3)}$$

Тогда пары (x, y), для которых выполняется условие $x \in A_i$, $y \in A_i$ (то есть эти элементы лежат в одном множестве разбиения), образуют отношение эквивалентности.

Рефлексивность и симметричность очевидны из определения. Транзитивность легко проверяется. Пусть $x, y \in A_i$; $y, z \in A_j$. Так как $A_i \cap A_j \supseteq \{y\} \neq \varnothing$, то $A_i = A_j$. Значит (x, z) также находится в отношении.

Последний пример исчерпывает все возможные отношения эквивалентности.

Теорема 10.5. *Любое отношение R, являющееся отношением эквивалентности на множестве A, делит A на классы эквивалентности — непересекающиеся подмножества множества A,*[4)] *при этом любые два элемента одного класса находятся в отношении R, а любые два элемента разных классов не находятся в отношении R.*

Доказательство. По сути мы повторим в общем виде доказательство аналогичного утверждения для случая компонент связности.

Для каждого $x \in A$ рассмотрим множество $C(x) = \{y : xRy\}$ тех y, для которых верно xRy. Это и есть обещанные классы эквивалентности. Чтобы это доказать, нужно проверить три условия:

1. Объединение всех множеств вида $C(x)$ совпадает с множеством A.
2. Два множества $C(x)$ и $C(y)$ либо не пересекаются, либо совпадают;
3. $C(x) = C(y)$ в том и только том случае, когда xRy (то есть R совпадает с отношением «принадлежать одному классу», как в примере 10.9).

1. В силу рефлексивности множество $C(x)$ содержит x в качестве своего элемента: $x \in C(x)$, поскольку xRx. Отсюда следует, что объединение всех этих множеств совпадает с A.

2. Пусть $z \in C(x) \cap C(y)$, то есть верно xRz и yRz. Симметричность даёт zRy. Теперь применим транзитивность к xRz и zRy, заключаем, что xRy и по симметричности yRx.

Пусть $t \in C(y)$, то есть yRt. Применим транзитивность к xRy и yRt, заключаем, что xRt, то есть $t \in C(x)$. Значит, $C(y) \subseteq C(x)$. Аналогично доказывается, что $C(x) \subseteq C(y)$, так что $C(x) = C(y)$.

3. Если для каких-то x, y верно xRy, то x и y оба лежат в одном классе, а именно, в $C(x)$. Обратно, если x и y лежат в каком-то $C(z)$, то по определению имеем zRx и zRy. Симметричность даёт xRz, после чего транзитивность даёт xRy. □

[3)] Спасибо Валерии Булановой за замеченную опечатку.

[4)] Спасибо Анне Балашовой за замеченную опечатку.

10.5. Изоморфизм графов

Отношения эквивалентности встречаются повсюду в математике. Очень часто они имеют вид отношений «изоморфизма», когда две структуры объявляются «по сути одинаковыми, с точностью до переобозначений».

Давайте рассмотрим пример такого отношения из теории графов.

Посмотрим на рис. 10.1. Там изображены два связных графа. Одинаковы ли они? Рисунки выглядят совершенно по-разному, но с точки зрения теории графов геометрия неважна, важны лишь связи между вершинами. С этой точки зрения оба графа выглядят одинаково: их рёбра являются рёбрами простого цикла длины 7 и других рёбер в этих графах нет.

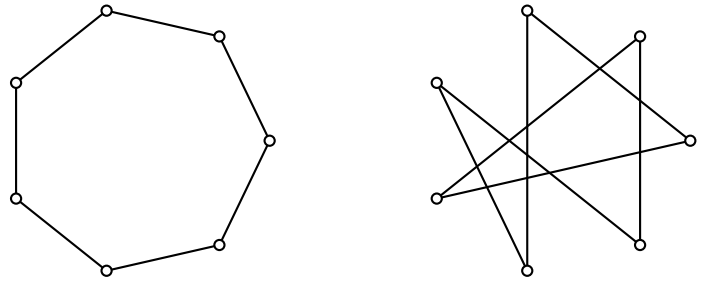


Рис. 10.1: Одинаковы ли эти графы?

Эта «похожесть» фиксируется определением отношения изоморфизма.

Определение 10.6. Графы $G_1 = (V_1, E_1)$ и $G_2 = (V_2, E_2)$ называются *изоморфными* (обозначение $G_1 \cong G_2$), если существует такая биекция $\pi\colon V_1 \to V_2$ на множествах их вершин, которая переводит множество рёбер первого графа в множество рёбер второго графа, т.е. $\{u, v\} \in E_1$ равносильно $\{\pi(u), \pi(v)\} \in E_2$.

Неформально это определение можно пересказать так: графы изоморфны, если можно так отождествить их вершины, чтобы рёбра этих графов совпали.

Пример 10.10. Зададим изоморфизм графов на рисунке 10.1. Для этого нужно как-то обозначить их вершины. Занумеруем их как на рис. 10.2.

Один из возможных изоморфизмов левого графа на правый задаётся как

i	0	1	2	3	4	5	6
$\pi(i)$	0	2	5	3	6	1	4

Заметим, что в верхней строки вершины выписаны в порядке прохождения цикла длины 7 в левом графе, а к нижней строке — в порядке прохождения цикла длины 7 в правом графе. Поэтому рёбра левого графа переходят при таком отображении в рёбра правого графа.

Утверждение 10.7. *Отношение изоморфизма — отношение эквивалентности.*

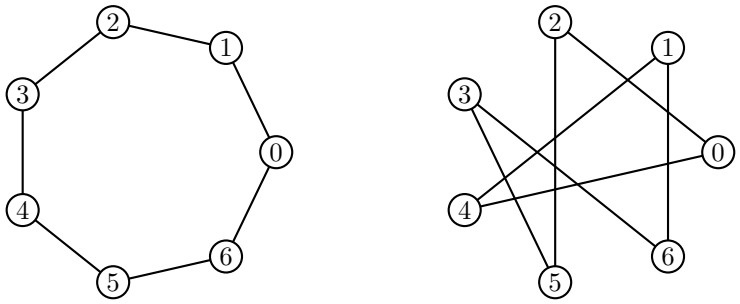


Рис. 10.2: Нумерация вершин графов

Доказательство. Рефлексивность: тождественное отображение задаёт изоморфизм графа с самим собой.

Симметричность. Докажем, что если $\pi\colon V_1 \to V_2$ — изоморфизм, то $\pi^{-1}\colon V_2 \to V_1$ — также изоморфизм. По определению нужно проверить, что $\{u, v\} \in E_2$ равносильно $\{\pi^{-1}(u), \pi^{-1}(v)\} \in E_1$. Так как π — изоморфизм (V_1, E_1) на (V_2, E_2), второе равносильно $\{\pi \circ \pi^{-1}(u), \pi \circ \pi^{-1}(v)\} \in E_2$, что равносильно $\{u, v\} \in E_2$, так как $\pi \circ \pi^{-1} = \mathrm{id}_{V_2}$.

Транзитивность. Пусть $\pi_1\colon V_1 \to V_2$ — изоморфизм (V_1, E_1) на (V_2, E_2); а $\pi_2\colon V_2 \to V_3$ — изоморфизм (V_2, E_2) на (V_3, E_3). Докажем, что тогда $\pi = \pi_2 \circ \pi_1$ — изоморфизм (V_1, E_1) на (V_3, E_3). Условие $\{u, v\} \in E_1$ равносильно $\{\pi_1(u), \pi_1(v)\} \in E_2$, так как π_1 — изоморфизм, а это условие в свою очередь равносильно $\{\pi_2 \circ \pi_1(u), \pi_2 \circ \pi_1(v)\} \in E_3$. Осталось применить транзитивность импликации. □

Изоморфизм сохраняет все свойства графов, которые выражаются в терминах связей между вершинами и рёбрами и не ссылаются на конкретные имена вершин. Такие свойства называются *инвариантами изоморфизма*.

Для доказательства неизоморфности графов достаточно указать какое-нибудь инвариантное свойство, которое есть у одного графа и которого нет у другого графа.

Например, число вершин в графе — инвариант изоморфизма, так как биекция возможна лишь между множествами одинакового размера.

Число рёбер в графе также является инвариантом изоморфизма, так как изоморфизм устанавливает взаимно однозначное соответствие между рёбрами.

Степень вершины сама по себе инвариантом изоморфизма не является. Но степень вершины сохраняется при изоморфизме: соседей у v ровно столько же, сколько у $\pi(v)$: соседи вершины обязаны переходить в соседей её образа.

Отсюда следует, что если в графе G_1 есть вершина v степени 5, а в графе G_2 все вершины имеют степень отличную от 5, то такие графы неизоморфны: при любой биекции π степень вершины $\pi(v)$ будет отличаться от 5.

Поэтому множества степеней вершин в графах должны совпадать. Но есть и более сильное условие. Если в графе G_1 есть ровно две вершины степени 5, а в графе G_2 таких вершин три, то графы неизоморфны: при любой биекции π в одну из вершин степени 5 в графе G_2 перейдёт вершина другой степени и условие изоморфизма

будет нарушено.

Значит, должны совпадать *наборы степеней*. Что это такое? Это множество таких пар (d, k), что в графе есть ровно k вершин степени d. Обычно набор степеней задают, записывая степени всех вершин графа в невозрастающем порядке.

Однако и совпадения наборов степеней недостаточно для изоморфизма.

Пример 10.11. Два графа на рис. 10.3 неизоморфны. Оба графа — деревья, наборы

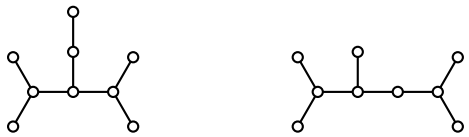


Рис. 10.3: два неизоморфных дерева

степеней вершин которых одинаковы и равны $(3, 3, 3, 2, 1, 1, 1, 1, 1)$. Однако соседи у вершины степени 2 в правом графе имеют степени 3, 3, а в левом — 3, 1. Изоморфизм должен сохранять степени вершин, поэтому вершина степени 2 перейдёт при изоморфизме в вершину степени 2. Её соседи также должны перейти в соседей образа, а это невозможно.

В общем случае проверка изоморфизма графов является трудной задачей, требующей перебора многих вариантов. Прямолинейный подход, основанный на определении, требует перебора $n!$ возможных биекций, если речь идёт об изоморфизме n-вершинных графов. Есть более изощрённые алгоритмы проверки изоморфизма, которым работают за время $n^{\log^k n}$, здесь k — некоторая константа.

Лекция 11

Отношения–3. Частичные порядки

11.1. Определения отношений частичного порядка

Определение 11.1. Бинарное отношение R на множестве X является *строгим частичным порядком*, если выполнены такие свойства:

- если aRb и bRc, то aRc (*транзитивность*);
- aRa всегда ложно (*антирефлексивность*).

Утверждение 11.2 (антисимметричность строгого порядка)**.** *Если R — строгий частичный порядок, то aRb влечёт ложность bRa.*

Доказательство. Пусть одновременно истинны aRb и bRa. Тогда по транзитивности истинно aRa, противоречие. □

Примером порядка является сравнение чисел по величине: $x < y$ обозначает, что число x меньше числа y. В дальнейшем для наглядности мы часто используем в качестве имени отношения строго порядка знак «<», обозначая им не только сравнение чисел. Для сравнения чисел выполняется дополнительное свойство *линейности* порядка: для любых $a \neq b$ истинно одно из двух: aRb или bRa. Но так бывает не всегда (отсюда слово «частичный» в названии).

Пример 11.1 (порядок по включению подмножеств)**.** На множестве $\mathcal{P}(n)$ всех подмножеств n-элементного множества определим порядок $X \subset Y$, включая в него те пары подмножеств (X, Y), для которых все элементы подмножества X содержатся в Y и есть хотя бы один элемент в Y, который не содержится в X.

Это отношение строгого частичного порядка. При этом свойство линейности порядка не выполняется: одновременно $\{1\} \not\subset \{2\}$ и $\{2\} \not\subset \{1\}$. (Перечёркнутый знак отношения означает, что пара не находится в отношении.)

Для чисел помимо отношения строго сравнения используется также отношение $\leqslant$ («меньше или равно») и часто оно удобнее строгого порядка.

То же самое верно и для частичных порядков.

Определение 11.3. Бинарное отношение $\leqslant$ на множестве X является *нестрогим частичным порядком*, если выполнены такие свойства:

- $a \leqslant a$ (*рефлексивность*);
- $(a \leqslant b)$ и $(b \leqslant a)$ влечёт $(a = b)$ (*антисимметричность*);
- $(a \leqslant b)$ и $(b \leqslant c)$ влечёт $(a \leqslant c)$ (*транзитивность*).

Между отношениями строго и нестрого порядка есть очевидная связь.

Утверждение 11.4. *Пусть $<$ — отношение строгого частичного порядка. Тогда отношение*

$$a \leqslant b \text{ равносильно } (a < b) \vee (a = b)$$

является отношением нестрогого частичного порядка.

Доказательство. Рефлексивность записана в определении.

Антисимметричность: предположим, что $a \leqslant b$, $b \leqslant a$, но $a \neq b$. Тогда по построению должно быть $a < b$ и $b < a$, что невозможно (см. утв. 11.2).

Транзитивность доказывается разбором случаев: если $a = b$ или $b = c$, то транзитивность очевидно выполняется, а если $a < b$ и $b < c$, применяем транзитивность для исходного отношения строгого частичного порядка. □

Утверждение 11.5. *Пусть $\leqslant$ — отношение нестрогого частичного порядка. Тогда отношение*

$$a < b \text{ равносильно } (a \leqslant b) \wedge (a \neq b)$$

является отношением строгого частичного порядка.

Доказательство. Антирефлексивность ясна по построению. Надо проверить транзитивность: пусть $a < b$ и $b < c$, то есть (согласно определению порядка $<$) $a \leqslant b$, $a \neq b$, $b \leqslant c$, $b \neq c$. Надо получить $a \leqslant c$ и $a \neq c$. Первое сразу следует из транзитивности отношения $\leqslant$. Докажем второе: если $a = c$, то получаем $a \leqslant b$ и $b \leqslant a$, откуда следует $a = b$ (антисимметричность порядка $\leqslant$) в противоречии с предположением. □

Таким образом, имеется взаимно однозначное соответствие между строгими и нестрогими порядками: один получается из другого выбрасыванием (или добавлением) *диагонального множества* $\{(a, a) : a \in X\}$.

В силу этих утверждений мы будем пользоваться и одним, и другим видом отношений частичного порядка, выбирая тот, который удобнее.

Частично упорядоченным множеством $(X, <)$ называется множество X с отношением частичного порядка $<$ на нём. Для краткости такое множество часто называют частичным порядком или даже просто порядком. Если отношение на записи часто опускают, если оно ясно из контекста.

11.2. Частичные порядки и ориентированные графы

Пример 11.2. На множестве чисел $\{0, 1, 2, \ldots, 8, 9\}$ рассмотрим обычное отношение порядка. Изобразим это упорядоченное множество, помещая меньшие элементы левее бóльших, см. рис. 11.1. Это линейный порядок: любые два числа сравнимы.

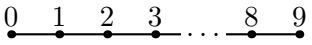


Рис. 11.1: Линейный порядок на $\{0, 1, 2, \ldots, 8, 9\}$, сравнение по величине

Аналогично можно рисовать и другие порядки.

Пример 11.3. Порядок по включению на подмножествах 2-элементного множества изображён на рис. 11.2.

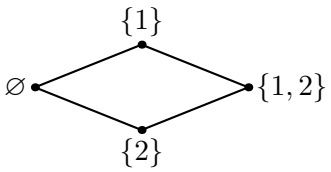


Рис. 11.2: $(\mathcal{P}(2), \subset)$

Этот порядок, как уже обсуждалось, частичный. Подмножества $\{0\}$ и $\{1\}$ несравнимы. Поэтому между ними нет линии. Но и между $\varnothing$ и $\{0, 1\}$ нет линии, хотя они сравнимы. Мы предполагаем при таком изображении, что одно множество меньше другого, если на рисунке из первого можно пройти во второе, двигаясь по линиям вправо.

Соблюдать такое соглашение не всегда легко. Поэтому лучше проводить проводить стрелки на рёбрах в направлении возрастания. Таким образом, частичный порядок удобнее изображать ориентированным графом. Разберём подробнее соответствие между ориентированными графами и частичными порядками.

Во-первых, заметим, что любой частичный порядок на множестве V задаёт ориентированный граф, рёбра которого — это в точности те пары (x, y), для которых $x < y$. Видно, что на риунках выше изображены какие-то другие графы: рёбер на рисунках меньше, чем пар сравнимых элементов.

Говорят, что элементы x, y частично упорядоченного множества $(X, <)$ *соседние*, элемент x *непосредственно предшествует* y, элемент y *непосредственно следует за* x, если $x < y$ и нет такого z, что $x < z < y$.[1)]

По частичному порядку $<$ на множестве X построим ориентированный граф $N(<)$ с множеством вершин X и множеством рёбер

$$E = \{(u, v) : u \text{ непосредственно предшествует } v\}.$$

[1)] Это три названия для одного и того же.

Утверждение 11.6. *Для всякого частичного порядка $<$ граф $N(<)$ ациклический.*

Доказательство. Если в $N(<)$ есть цикл $(a_1 a_2 \dots a_n a_1)$, то по определению графа $N(<)$ выполняются сравнения:

$$a_1 < a_2 < \dots < a_n < a_1$$

(соседние в цикле непосредственно предшествуют в порядке). По транзитивности получаем $a_1 < a_1$ и приходим к противоречию с антирефлексивностью. □

В другую сторону. Любой ациклический граф $G = (V, E)$ задаёт на множестве вершин V частичный порядок по следующему правилу:

$$u \leqslant_G v \text{ равносильно } v \in R(u),$$

где $R(u)$ обозначает область достижимости вершины u.

Утверждение 11.7. *Для любого ациклического графа отношение $\leqslant_G$ является отношением нестрогого частичного порядка.*

Доказательство. Проверим свойства частичного порядка для отношения $\leqslant_G$. Рефлексивность и транзитивность мы уже проверяли, когда рассматривали ориентированные графы.

Если $u \leqslant_G v$ и $v \leqslant_G u$, то по определению $v \in R(u)$ и $u \in R(v)$, то есть вершины u, v сильно связаны. Но в ациклическом графе каждая компонента сильной связности состоит из одной вершины. Поэтому $u = v$. □

Обратите внимание, что построенное соответствие не взаимно однозначно.

Пример 11.4. В порядке $(\mathbb{R}, <)$ на действительных числах соседних пар нет вообще. Это называется свойством *плотности* порядка: между любыми двумя различными числами $a < b$ есть хотя бы одно число, которое больше a и меньше b (например, полусумма $(a + b)/2$).

Поэтому граф $N(<)$ не содержит рёбер. Но такой граф задаёт тривиальный частичный порядок, в котором нет ни одной пары сравнимых элементов.

11.3. Операции с порядками

Пример 11.5. На множестве векторов $\mathbb{R}^n$ определено отношение *покоординатного порядка*:

$$x = (x_1, \dots, x_n) \leqslant (y_1, \dots, y_n) = y \quad \Leftrightarrow \quad x_i \leqslant y_i \text{ для всех } i.$$

Числа в правой части сравниваются обычным способом.

Свойства частичного порядка легко проверяются (упражнение для самостоятельной работы).

Этот порядок частичный. Например, векторы $(0, 2)$ и $(1, 1)$ несравнимы.

Этот пример — частный случай операции *покоординатного произведения частичных порядков*. Пусть P, Q — два частичных порядка. Тогда покоординатный порядок на декартовом произведении $P \times Q$ задается правилом:

$$(p_1, q_1) \leqslant (p_2, q_2) \text{ по определению означает } p_1 \leqslant_P p_2 \text{ и } \ q_1 \leqslant_Q q_2.$$

Свойства частичного порядка для такого отношения также легко проверяются, их доказательство оставляется для самостоятельной работы.

Как мы видели в примере 11.5, покоординатное произведение линейных порядков не всегда линейно. Есть другие операции с порядками, которые сохраняют свойство линейности.

Определение 11.8. Пусть P, Q — два частичных порядка. *Лексикографический порядок* на $P \times Q$ задаётся правилом:

$(p_1, q_1) \leqslant (p_2, q_2)$ по определению означает, что $(p_1 <_P p_2)$ или $\ (p_1 = p_2)$ и $(q_1 \leqslant_Q q_2)$.

Лемма 11.9. *Лексикографический порядок на $P \times Q$ является отношением частичного порядка. Если P и Q — линейные порядки, то лексикографический порядок также линейный.*

Доказательство. Рефлексивность сразу следует из определения и рефлексивности порядка Q.

Антисимметричность. Пусть $(p_1, q_1) \leqslant (p_2, q_2)$ и $(p_2, q_2) \leqslant (p_1, q_1)$. Случай $p_1 \neq p_2$ невозможен, так как тогда $p_1 < p_2 < p_1$ в порядке P, что противоречит антирефлексивности строгого порядка. Если же $p_1 = p_2$, то получаем из определения $q_1 \leqslant q_2 \leqslant q_1$, то есть $q_1 = q_2$ в силу антисимметричности порядка Q.

Транзитивность. $(p_1, q_1) \leqslant (p_2, q_2)$ и $(p_2, q_2) \leqslant (p_3, q_3)$. Из определения лексикографического порядка видим, что $p_1 \leqslant p_2 \leqslant p_3$. Если $p_1 < p_2 < p_3$, то $p_1 < p_3$ по транзитивности порядка P и потому $(p_1, q_1) < (p_3, q_3)$. Если $p_1 = p_2 < p_3$ или $p_1 < p_2 = p_3$, то также $p_1 < p_3$ и $(p_1, q_1) < (p_3, q_3)$. Если же $p_1 = p_2 = p_3$, то из определения лексикографического порядка получаем $q_1 \leqslant q_2 \leqslant q_3$, в силу транзитивности порядка Q и определения лексикографического порядка получаем $(p_1, q_1) \leqslant (p_3, q_3)$.

Последнее свойство очевидно из определения. □

Определение 11.10 (неудачное)**.** Пусть P, Q — два частичных порядка, $P \cap Q = \varnothing$. *Суммой* $P + Q$ называется порядок на $P \cup Q$, в котором все элементы P меньше всех элементов Q.

Легко проверить (оставляется в качестве самостоятельного упражнения), что сумма частичных порядков является частичным порядком и сумма линейных порядков является линейным порядком.

Обратите внимание, что сумма порядков некоммутативна. Порядки $\mathbb{N} + \mathbb{Z}$ и $\mathbb{Z} + \mathbb{N}$ — существенно разные порядки. Но что это утверждение означает? Ведь $\mathbb{N} \subset \mathbb{Z}$, а в данном выше определении суммы порядков такое запрещено. Определение нужно поправить.

11.4. Изоморфизм порядков

Определение 11.11. Порядки P и Q называются *изоморфными* (обозначение $P \cong Q$), если есть такая биекция $\varphi\colon P \to Q$, что $x \leqslant y$ равносильно $\varphi(x) \leqslant \varphi(y)$ для всех пар x, y.

Пример 11.6. Рассмотрим два порядка: порядок $(\mathcal{P}_n, \subset)$ на подмножествах n-элементного множества по включению и покоординатный порядок на двоичных словах длины n. Это два разных отношения, заданные на разных множествах.

Изоморфизм между ними устанавливает стандартная биекция: подмножеству S сопоставляется слово x_S, в котором на i-й позиции стоит 1 тогда и только тогда, когда $i \in S$.

Если $S \subseteq T$, то на всех позициях, в которых в слове x_S стоит 1, в слове x_T также стоит 1.

Теперь уже ясно, как определить сумму порядков в общем случае.

Определение 11.10 (окончательное)**.** Пусть P, Q — два частичных порядка, $P' \cong P$, $Q' \cong Q$, $P' \cap Q' = \varnothing$. *Суммой* $P + Q$ называется порядок на $P' \cup Q'$, в котором все элементы P' меньше всех элементов Q'.

Как и в случае изоморфизма графов, проще устанавливать неизоморфизм порядков, указывая различающее их *инвариантное свойство*, то есть свойство, которое не изменяется при изоморфизме. Такими свойствами являются любые свойства, которые выражаются в терминах сравнения элементов без отсылок к конкретным именам элементов.

Пример 11.7. Докажем, что $\mathbb{N} + \mathbb{Z}$ и $\mathbb{Z} + \mathbb{N}$ неизоморфны. Чтобы перейти к непересекающимся множествам, рассмотрим обычные целые числа и «штрихованные натуральные»: числа вида $0', 1', \ldots$. Сравниваются эти числа так же, как нештрихованные. Но теперь множества не пересекаются (штрих либо есть, либо его нет).

В $\mathbb{N}+\mathbb{Z}$ есть *наименьший элемент* — $0'$ меньше всех остальных элементов суммы порядков. А в $\mathbb{Z} + \mathbb{N}$ такого элемента нет (для каждого целого числа есть меньшее его).

Из этого примера можно заключить, что на бесконечном множестве существуют неизоморфные линейные порядки. Для конечных множеств это не так, если в множестве одинаковое количество элементов.

Теорема 11.12. *Пусть $(X, \leqslant)$ и $(Y, \leqslant)$ — два линейных порядка на конечных множествах и $|X| = |Y|$. Тогда эти порядки изоморфны.*

Сначала докажем лемму.

Лемма 11.13. *В конечном линейном порядке есть наибольший и наименьший элементы.*

Доказательство. Рассмотрим *убывающие цепи*: последовательности элементов порядка $x_1 > x_2 > \ldots$, в которой каждый следующий элемент меньше предыдущего. Будем дополнять убывающую цепь новыми элементами, пока это возможно. Поскольку всего элементов конечное число, процесс рано или поздно остановится. Последний элемент такой убывающей цепи обязан быть наименьшим: в противном случае её можно было бы продолжить.

Аналогичное рассуждение с *возрастающими цепями* показывает существование наибольшего элемента. □

Доказательство теоремы 11.12. Индукция по числу элементов. База — один элемент в порядке — очевидна.

Индуктивный переход. Предположим, что все линейные порядки с n элементами изоморфны. Рассмотрим два линейных порядка P и Q с $n+1$ элементом. Выделим в них наименьшие элементы p_0, q_0. Порядки на оставшихся элементах изоморфны по предположению индукции. Продолжая этот изоморфизм соответствием $p_0 \mapsto q_0$, получаем искомый изоморфизм порядков P и Q. □

11.5. Цепи и антицепи

Среди частичных порядков на конечном множестве есть два крайних случая. Первый: линейный порядок (каждая пара элементов сравнима). Второй: пустой порядок (никакая пара различных элементов несравнима).

В остальных случаях можно выделять подмножества частично упорядоченного множества, которые образуют эти два крайних случая при ограничении исходного частичного порядка на это подмножество. *Цепь* — это такое подмножество частично упорядоченного множества, которое образует линейный порядок. *Антицепь* — это такое подмножество, в котором элементы попарно несравнимы.

Пример 11.8. Рассмотрим порядок $(\mathcal{P}_3, \subset)$ на подмножествах 3-элементного мно-

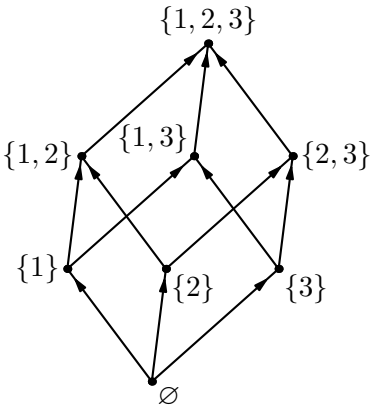


Рис. 11.3: $(\mathcal{P}(3), \subset)$

жества, упорядоченных по включению. Он изображён на рис. 11.3. Легко увидеть антицепь размера 3: $\{\{1\},\{2\},\{3\}\}$. Антицепей большего размера в этом порядке нет. Но как это доказать? Перебор вариантов кажется очень трудоёмким.

Разобьём элементы этого порядка на три цепи: $\varnothing \subset \{1\} \subset \{1,2\} \subset \{1,2,3\}$; $\{2\} \subset \{2,3\}$; $\{3\} \subset \{1,3\}$.

Любая антицепь пересекает цепь не более чем по одному элементу (любая пара сравнима в цепи и несравнима в антицепи). Значит, любая антицепь в этом порядке содержит не более трёх элементов.

Оказывается, это соотношение точное.

Теорема 11.14 (Дилуорс)**.** *Наибольший размер антицепи в порядке равен наименьшему количеству цепей в разбиениях порядка на непересекающиеся цепи.*

Известно много доказательств этой теоремы. Приведём свежее.

Доказательство Фёдора Куянова, ФКН ВШЭ, 1 курс. В одну сторону нужно повторить рассуждение из примера 11.8: если порядок разбит на k непересекающихся цепей, то любая антицепь пересекается с каждой из цепей не более чем по одному элементу и в антицепи не больше k элементов.

Теперь предположим, что в частично упорядоченном множестве P размер максимальной антицепи равен k. Нужно доказать, что найдётся разбиение P на k цепей.

Добавим к этому множеству k несравнимых между собой элементов, которые больше всех остальных элементов порядка (k дополнительных максимумов) и k несравнимых между собой минимумов, которые меньше всех остальных элементов порядка (включая добавленные максимумы). Получаем новый порядок G. Размер максимальной антицепи не изменился: антицепь в G либо содержит только добавленные элементы (причём либо максимумы, либо минимумы), либо только элементы порядка P.

Разбиение порядка G на цепи урезается до разбиения P на цепи выбрасыванием добавленных элементов. Поэтому достаточно доказать, что существует разбиение G на k цепей. Для этого применим теорему Холла.[2)]

Рассмотрим двудольный граф $F = (L, R, E)$, в котором L — множество элементов G без добавленных минимумов, R — множество элементов G без добавленных максимумов, а множество рёбер $E = \{(a,b) : a > b\}$ (сравнение в порядке G). Из построения видно, что $|L| = |R|$. В таком графе элементы исходного порядка P входят в обе доли. Но мы хотим рассматривать этот граф как двудольный. Нужно себе представить, что сделаны две копии P', P'' порядка P, одна отправлена в левую (верхнюю на рисунке 11.4) долю, вторая — в правую (нижнюю).

Как видно из рисунка 11.4, цепям в порядке G отвечают паросочетания в графе F. Совершенному паросочетанию в F будет соответствовать разбиение порядка G на цепи (на рисунке выделена одна из таких цепей, остальные строятся из других вершин области maxima).

[2)] Спасибо Ивану Ершову за замеченную неточность.

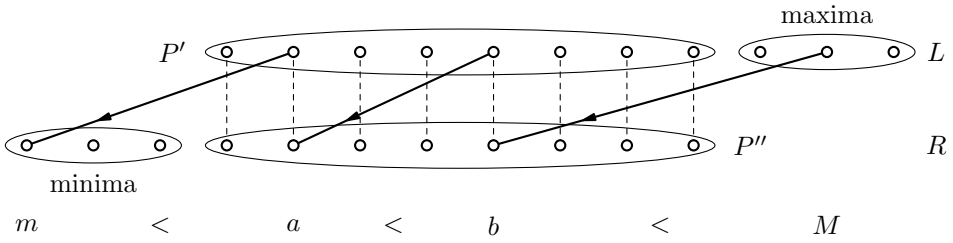


Рис. 11.4: Цепи превращаются в паросочетания во вспомогательном графе

Осталось доказать, что в графе F существует совершенное паросочетание. Проверим выполнение условия теоремы Холла. Нужно доказать $|G(S')| \geqslant |S'|$ для любого $S' \subseteq L$. Здесь $G(S')$ обозначает множество вершин доли R, соединённых ребром хотя бы с одной вершиной множества S'. Заметим, что в $G(S')$ входят все минимумы, их k штук. Обозначим A' подмножество S', состоящее из тех вершин, копии которых не принадлежат $G(S')$ или которые являются максимумами. Это антицепь, так как из $x > y$ и $x' \in S'$ следует, что $y'' \in G(S')$. Поэтому $|A'| \leqslant k$. Но тогда

$$|G(S')| \geqslant k + |S' \setminus A'| \geqslant k + |S'| - k = |S'|,$$

что и требовалось доказать. Первое слагаемое в первом неравенстве отвечает минимумам в правой доле, второе — тем вершинам S', копии которых в правой доле лежат в $G(S')$. □

Лекция 12

Числа–1. Деление с остатком. Алгоритм Евклида

12.1. Деление с остатком

Деление (операция, обратная к умножению) не всегда возможно, если ограничиваться только целыми числами. Например, уравнение $2x = 1$ не имеет решений в целых числах.

Целое число a *делится на* целое число b, если $a = bk$ для некоторого целого числа k. В этом случае говорят также «a *кратно* b», и «b *является делителем числа* a».

Получаем бинарное отношение на множестве целых чисел (иногда будем ограничивать его на множество положительных целых чисел). Принятое в современных математических книгах обозначение этого отношения $b \mid a$ (первое число — делитель второго, второе — кратное первого.

Пример 12.1. $2 \mid 4$; $5 \nmid 7$ (перечёркнутая черта означает, что отношение делимости не выполняется); $0 \nmid 1$; $n \mid 0$ для любого целого n: ведь $0 = 0 \cdot n$.

Теперь определим деление с остатком. Разделить целое a на целое положительное b означает найти такое целое q (*частное*) и такое целое r (остаток), что

$$a = b \cdot q + r; \quad 0 \leqslant r < b.$$

Пример 12.2. Остаток от деления 2020 на 100 равен 20, так как $2020 = 20 \cdot 100 + 20$.

Остаток от деления -5 на 7 равен 2: $-5 = (-1) \cdot 7 + 2$.

Благодаря ограничению на r справедливо следующее утверждение.

Утверждение 12.1. *Деление с остатком всегда возможно, притом единственным образом.*

Доказательство. Единственность. Если $a = bq + r = bq' + r'$, то $r - r' = b(q' - q)$ и потому $r - r'$ делится на b. Но оба числа r, r' находятся в интервале $0, 1, \ldots, b-1$, так что их разность (если из большего вычесть меньшее) не больше $b-1$ и делится на b. Поэтому $r = r'$, откуда и $q = q'$.

Существование для неотрицательных чисел можно доказать индукцией по a. Для $a = 0$ частное и остаток равны нулю: $0 = 0 \cdot b + 0$. Если $a = bq + r$, то $a + 1 = bq + (r + 1)$. При этом $r + 1 \leqslant b$, так как $r < b$. Если $r + 1 < b$, то для $a + 1$ получаем частное q и остаток $r + 1$. Если же $r + 1 = b$, то $a + 1 = bq + b = b(q + 1) + 0$, получаем частное $q + 1$ и остаток 0.

Для отрицательных чисел: разделим $-a$ на b с остатком, получим $-a = bq + r$, $0 \leqslant r < b$. Тогда в случае $r = 0$ получаем $a = -bq$, а в случае $r > 0$ получаем

$$a = -bq - r = b(-q - 1) + (b - r),$$

в этом случае $0 < b - r < b$. □

12.2. Арифметика остатков. Вычеты

Как найти остаток суммы или произведения чисел x, y от деления на некоторое заданное число b? Из свойств арифметических операций следует, что нужно взять остаток от деления суммы (произведения) остатков на b:

$$\begin{aligned} x &= q_1 b + r_1, \\ y &= q_2 b + r_2, \\ (x + y) &= (q_1 + q_2)b + (r_1 + r_2) = (q_1 + q_2 + q_3)b + r_3, \\ xy &= (q_1 q_2 + r_1 q_2 + r_2 q_1)b + r_1 r_2 = (q_1 q_2 + r_1 q_2 + r_2 q_1 + q_4)b + r_4. \end{aligned}$$

Чтобы найти остаток от деления 6^{1000} на 7, придётся выполнить 999 таких преобразований. Есть более удобный способ выполнять арифметические операции с остатками, из которого сразу очевидно, что этот остаток равен 1.

Если два числа a и b дают одинаковые остатки при делении на положительное число N, то говорят, что они *сравнимы* по модулю N, и пишут $a \equiv b \pmod N$.

Это равносильно такому условию: разность $a - b$ делится на N. В самом деле, если a, b дают одинаковый остаток r, то $a = kN + r$, $b = lN + r$, и $a - b = kN - lN = (k - l)N$. Наоборот, если $a - b = mN$, и b даёт остаток r, то $b = lN + r$ и $a = (a - b) + b = mN + lN + r = (m + l)N + r$, то есть a даёт тот же остаток r.

Для любого N отношение сравнимости по модулю N является отношением эквивалентности. Классы эквивалентности — множества чисел, имеющих одинаковый остаток от деления на N, — называются классами вычетов или просто *вычетами* по модулю N.

Пример 12.3. Числа $\ldots, -11, -5, 1, 7, \ldots$ дают остаток 1 по модулю 6. Они образуют класс вычетов по модулю 6.

Когда мы говорим о вычете 1 по модулю 6, то имеем в виду всё это множество чисел.

Лемма 12.2. *Класс суммы, разности или произведения чисел зависит только от классов слагаемых.*

Доказательство. Если к одному из слагаемых прибавить kN, то к сумме тоже прибавится kN, аналогично для разности. С произведением: $(a+kN)b = ab+kbN \equiv ab \pmod N$.

Поэтому для любых чисел, лежащих в одном классе вычетов, класс вычетов суммы, разности или произведения один и тот же. □

В силу этой леммы на вычетах корректно определены операции сложения и умножения:

$$\{a+kN : k\in\mathbb{Z}\} + \{b+kN : k\in\mathbb{Z}\} = \{a+b+kN : k\in\mathbb{Z}\};$$
$$\{a+kN : k\in\mathbb{Z}\} \cdot \{b+kN : k\in\mathbb{Z}\} = \{ab+kN : k\in\mathbb{Z}\}.$$

Теперь легко обосновать, почему остаток от деления 6^{1000} на 7 равен 1:

$$6^{1000} \equiv (-1)^{1000} = 1 \pmod 7.$$

В силу леммы 12.2 класс вычетов можно задавать любым числом, в него входящим. Мы так и будем дальше делать, выбирая представителей класса как удобно (не обязательно это будут числа от 0 до $N-1$).

12.3. Свойства арифметики остатков

Для каждого N определены сложение и умножение вычетов по модулю N. В силу леммы 12.2 эти операции наследуют обычные свойства арифметических операций:

- $a+(b+c) = (a+b)+c$ (ассоциативность сложения);
- $ab = ba$ (коммутативность умножения)
- $a(bc) = (ab)c$ (ассоциативность умножения)
- $a(b+c) = ab+ac$ (дистрибутивность)

Числа 0 и 1 (точнее, классы вычетов, содержащие 0 и 1) при делении на N обладают обычными свойствами:

- $0+a = a$;[1)]
- $1\cdot a = a$;
- $0\cdot a = 0$.

[1)] Спасибо Денису Осташову за замеченную опечатку.

Для каждого вычета a есть противоположный, который в сумме с a равен нулю: это вычет, содержащий $-a$. Другими словами, противоположный к вычету $\{a+kN : k \in \mathbb{Z}\}$ это вычет $\{-a + kN : k \in \mathbb{Z}\}$. Поэтому на множестве вычетов определено вычитание: $a - b = a + (-b)$.

А вот деление возможно не всегда. Вычет по модулю N называется *обратимым*, если в произведении с каким-то другим вычетом он даёт 1. Другими словами, a обратим, если уравнение $ax = 1$ имеет решение в арифметике вычетов. Чтобы помнить о модуле, мы будем записывать это уравнение как $ax \equiv 1 \pmod N$.

Пример 12.4. Вычет 5 обратим по модулю 12: $5 \cdot 5 = 2 \cdot 12 + 1$.

А вычет 8 необратим. Если $8x \equiv 1 \pmod{12}$, то в обычных целых числах выполняется равенство $8x = 12q + 1$, то есть $4 \cdot (2x - 3q) = 1$, что невозможно, так как 1 не кратен 4.

На обратимые элементы можно делить: если a обратим, то уравнение $ax \equiv b \pmod N$ имеет в вычетах[2)] единственное решение при любом b. В самом деле, обозначим через a^{-1} *какой-то* вычет, который в произведении с a даёт единицу (по модулю N). Положим $x = a^{-1}b$. Тогда $ax \equiv a(a^{-1}b) \equiv (aa^{-1})b \equiv 1 \cdot b = b \pmod N$, так что одно решение уравнения $ax \equiv b \pmod N$ мы нашли. Оно будет единственным: если $ax \equiv b \pmod N$, то $a^{-1}(ax) \equiv a^{-1}b \pmod N$, но $a^{-1}(ax) \equiv (a^{-1}a)x \equiv 1 \cdot x = x \pmod N$, так что для x есть только одна возможность.

Определение 12.3. *Взаимно простыми* называются числа, которые не имеют общего положительного делителя, не считая 1.

Если a взаимно прост с N, то и любое число $a + kN$ взаимно просто с N: общий делитель $a+kN$ и N является общим делителем a и N. Поэтому корректно говорить о вычетах по модулю N, взаимно простых с N.

Теорема 12.4. *Обратимыми по модулю N являются те и только те вычеты, которые взаимно просты с N.*

Доказательство. Пусть N и a имеют общий положительный делитель $k > 1$: $a = a'k$, $N = N'k$. Тогда, как в примере 12.4, из $ax \equiv 1 \pmod N$ получаем в обычных целых числах $ax = qN + 1$, откуда $(a'x - qN')k = 1$, что невозможно при $k > 1$.

Доказательство второй части теоремы будет удобно иллюстрировать на числовой окружности. Теперь мы предполагаем, что N и a взаимно просты.

Зафиксируем модуль N и вычет a. Рассмотрим *множество кратных вычета* a, то есть множество $S_a = \{x : x \equiv ka \pmod N,\ k \in \mathbb{Z}\}$. На рис. 12.1 показано несколько кратных a.

Выделим наименьший ненулевой остаток, входящий в множество S_a, обозначим его d (см. рис. 12.1 справа).

Обратимость вычета в точности означает, что $d = 1$ (то есть, что найдётся кратное a, которое даёт остаток 1 по модулю N). Все кратные d входят в множество кратных a: $kd \equiv kja \pmod N$ (см. обозначения на рис. 12.1 справа).

[2)] Напомним, что вычет — это не число, а множество чисел.

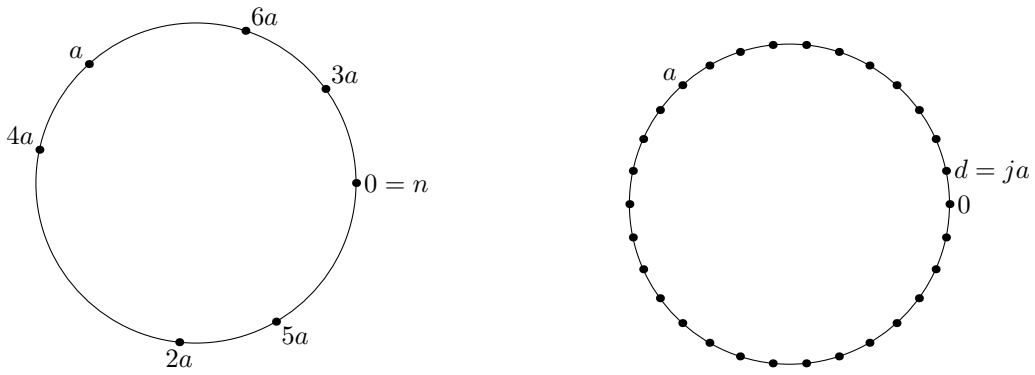


Рис. 12.1: Вычеты, кратные вычету a

Никаких других вычетов в S_a нет. Предположим, что $x \in S_a$, $dy < x < d(y+1)$. Так как $x = x'a$, то в множество S_a входит также $r = x - dy$, который меньше d.

Докажем, что d делит N. В противном случае $N = qd + r$ и $r < d$ также входит в множество кратных: $r \equiv -qd \pmod N$,[3)] а все кратные d входят в S_a.

Аналогично проверяем, что a делится на d. Разделим a на d с остатком: $a = qd+r$. Так как $a \in S_a$ и $qd \in S_a$, то $r = a - qd \in S_a$ и $0 \leqslant r < d$. По выбору d получаем $r = 0$.

Поскольку a и N взаимно просты, $d = 1$. □

12.4. Наибольший общий делитель. Алгоритм Евклида

Число d, появившееся в доказательстве теоремы 12.4, называется *наибольшим общим делителем* (НОД) чисел a, N, обозначение $d = \text{НОД}(a, N)$. Заметим, что из построения в доказательстве теоремы 12.4 следует такое утверждение.

Утверждение 12.5. НОД(a, b) *является целочисленной линейной комбинацией* a, b, *то есть* НОД$(a, b) = xa + yb$, $x, y \in \mathbb{Z}$.

Доказательство. Используем то же самое множество S_a вычетов по модулю b, кратных a. НОД принадлежит этому множеству по построению, $d \equiv ja \pmod b$. В целых числах получаем искомое равенство НОД$(a, b) = d = ja + kb$, $j, k \in \mathbb{Z}$. □

Название НОД оправдывается такой леммой.

Лемма 12.6. *Любой общий делитель* d' *чисел* a, N *является делителем числа* d.

Доказательство. Пусть $N = N'd'$, $a = a'd'$. Поскольку $d = xa + yN$, получаем в целых числах равенство $d = xa'd' + yN'd' = (xa' + yN')d'$. □

Теперь можно выяснить количество решений произвольного сравнения $ax \equiv b \pmod N$ в вычетах.

[3)] Спасибо Арсению Чеканову за замеченную опечатку.

Задача 12.5. [4)] Докажите, что сравнение $ax \equiv b \pmod n$ либо не имеет решений, если $\text{НОД}(a,n) \nmid b$, либо имеет ровно $\text{НОД}(a,n)$ решений в вычетах.

Решение. Первая часть утверждения очевидна: если $ax \equiv b \pmod n$, то $ax = b + nt$ для некоторого целого t и потому $b = ax - nt$ обязано делиться на $\text{НОД}(a,n)$.

Далее считаем, что $\text{НОД}(a,n) \mid b$. Обозначим $d = \text{НОД}(a,n)$, $a' = a/d$; $b' = b/d$; $n' = n/d$. Числа a' и n' взаимно просты: если $d' \mid a'$, $d' \mid n'$, то $dd' \mid n = dn'$ и $da' \mid a = da'$. Поэтому из теоремы 12.4 следует обратимость a' по модулю n' и существование единственного решения сравнения $a'x \equiv b' \pmod{n'}$ в вычетах. Целых решений у такого сравнения бесконечного много: они имеют вид $x_0 + n't$, $t \in \mathbb{Z}$.

Равенства

$$ax = b + nt \quad \text{и} \quad a'x = b' + n't$$

равносильны (одно получается из другого умножением на d). Поэтому те же самые числа являются и решениями сравнения $ax \equiv b \pmod n$.

Осталось понять, сколько классов вычетов по модулю n содержат такие числа. Если $x_0 + n't \equiv x_0 + n'u \pmod n$, то $n'(t-u) \equiv 0 \pmod n$. Сокращая на общий множитель n' получаем $u - v \equiv 0 \pmod d$. Поэтому ровно d классов вычетов содержат числа $x_0 + n't$, это и означает, что в вычетах по модулю n сравнение $ax \equiv b \pmod n$ имеет ровно $d = \text{НОД}\,a, n$ решений. □

Как находить НОД двух чисел? Для этого есть алгоритм Евклида. Мы рассмотрим расширенную версию алгоритма Евклида, которая одновременно находит и $\text{НОД}(a,b)$, и его представление в виде целочисленной линейной комбинации a, b. Такое представление позволяет найти обратный к обратимому вычету. Если a и N взаимно просты, то $1 = xa + yN$ для некоторых целых x, y. И тогда x — обратный к a по модулю N.

Корректность алгоритма Евклида основана на таком утверждении.

Лемма 12.7. $\text{НОД}(a,b) = \text{НОД}(a - qb, b)$ *для любого целого* q.

Доказательство. Если $d \mid a$ и $d \mid b$, то по свойствам делимости $d \mid (a - qb)$.

И в обратную сторону: если $d \mid (a - qb)$ и $d \mid b$, то по свойствам делимости $d \mid \big((a - qb) + qb\big)) = a$.

Значит, множества общих делителей у этих пар чисел совпадают. □

Расширенный алгоритм Евклида рекуррентно вычисляет три такие последовательности чисел a_i, x_i, y_i, что для каждого i выполняется соотношение (инвариант цикла)

$$a_i = x_i a + y_i b. \tag{12.1}$$

Начальные члены этих последовательностей:

$$\begin{aligned} a_0 &= a, \quad x_0 = 1, \quad y_0 = 0, \\ a_1 &= b, \quad x_1 = 0, \quad y_1 = 1, \end{aligned}$$

[4)] Это задача 5 из классного листка 14.

для них инвариант цикла (12.1) выполняется очевидным образом.

Чтобы найти a_i, x_i, y_i при $i \geqslant 2$, делим a_{i-2} на a_{i-1} с остатком, это и есть a_i. Если $a_{i-1} = 0$, то алгоритм останавливается. В противном случае получаем $a_i = a_{i-2} - q_{i-1}a_{i-1}$. Остальные числа вычисляем по аналогичной формуле, используя найденное неполное частное q_{i-1}:

$$x_i = x_{i-2} - q_{i-1}x_{i-1},$$
$$y_i = y_{i-2} - q_{i-1}y_{i-1}.$$

По индукции докажем, что инвариант цикла (12.1) выполняется на всех шагах алгоритма. База индукции проверена выше.

Шаг индукции. Пусть (12.1) выполняется для $i-2$ и $i-1$. Тогда подставим эти равенства в выражение для a_i и перегруппируем слагаемые:

$$\begin{aligned} a_i = a_{i-2} - q_{i-1}a_{i-1} =& x_{i-2}a + y_{i-2}b - q_{i-1}(x_{i-1}a + y_{i-1}b) = \\ =& (x_{i-2} - q_{i-1}x_{i-1})a + (y_{i-2} - q_{i-1}y_{i-1})b = x_ia + y_ib. \end{aligned}$$

Значит, (12.1) выполняется и для i.

Последовательность a_i уменьшается, начиная со второго шага. Поэтому алгоритм рано или поздно остановится: в некоторый момент a_{k-1} будет делится на a_k, поэтому $a_{k+1} = 0$ и алгоритм остановится при попытке разделить с остатком на a_{k+1}. Будем называть a_k *последним числом* в алгоритме Евклида.

Утверждение 12.8. *Последнее число a_k в алгоритме Евклида является НОД чисел a, b.*

Доказательство. По индукции с помощью леммы 12.7 проверяется, что

$$\text{НОД}(a_i, a_{i+1}) = \text{НОД}(a, b), \quad i + 1 \leqslant k.$$

База $i = 0$ следует из построения. Шаг индукции: так как $a_{i+1} = a_{i-1} - q_{i-1}a_i$, то по лемме 12.7 $\text{НОД}(a_i, a_{i+1}) = \text{НОД}(a_{i-1}, a_i)$. По предположению индукции второе число равно $\text{НОД}(a, b)$.

Поскольку $a_k \mid a_{k-1}$, то $a_k = \text{НОД}(a_k, a_{k-1}) = \text{НОД}(a, b)$. □

Замечание 12.1. Обычный алгоритм Евклида строит только последовательность a_i, последнее число в этой последовательности и будет $\text{НОД}(a_0, a_1)$.

Лекция 13

Числа–2

13.1. Линейные диофантовы уравнения

Уравнение

$$ax + by = c, \tag{13.1}$$

где a, b, c — целые числа, называется *линейным диофантовым уравнением*. Решить такое уравнение — найти все пары (x, y) целых чисел, которые удовлетворяют этому равенству.

Из соображений делимости легко следует, что уравнение (13.1) имеет решение только если c делится на НОД(a, b). В противном случае левая часть равенства делится на НОД(a, b), а правая — нет.

В случае НОД$(a, b) \mid c$ найти какое-то одно решение (13.1) можно с помощью расширенного алгоритма Евклида. Напомним, что расширенный алгоритм Евклида находит $d =$ НОД(a, b) и такие числа x_0, y_0, что $ax_0 + by_0 = d$. Пусть $c = k\,$НОД(a, b). Тогда $(\tilde{x}_0, \tilde{y}_0) = (kx_0, ky_0)$ является решением уравнения (13.1):

$$a(kx_0) + b(ky_0) = k(ax_0 + by_0) = kd = c.$$

Как найти **все** решения уравнения (13.1)? Задача разбивается на две, как показывает следующее утверждение.

Утверждение 13.1. *Пусть $(\tilde{x}_0, \tilde{y}_0)$ — решение уравнения* (13.1)*. Тогда все решения этого уравнения имеют вид $(\tilde{x}_0 + x, \tilde{y}_0 + y)$, где пара (x, y) является решением однородного линейного уравнения*

$$ax + by = 0. \tag{13.2}$$

Доказательство. По сути утверждается, что условия

$$ax + by = 0 \quad \text{и} \quad a(x_0 + x) + b(y_0 + y) = c$$

равносильны, если $ax_0 + by_0 = c$. Проверка равносильности состоит в применении равносильных преобразований к равенствам. □

Частное решение уравнения (13.1) мы уже научились находить. Теперь нужно найти общее решение уравнения (13.2).

Лемма 13.2. *Решениями однородного линейного уравнения* (13.2) *являются в точности такие пары* (x, y)*, что*

$$x = t \cdot \frac{b}{d},\ y = -t \cdot \frac{a}{d}, \quad d = \text{НОД}(a, b),\ t \in \mathbb{Z}.^{1)}$$

Доказательство. Разделив обе части уравнения на наибольший общий делитель (a, b), получим равносильное уравнение, коэффициенты которого взаимно просты. Поэтому достаточно решить уравнение $ax + by = 0$ со взаимно простыми коэффициентами.

Если $\text{НОД}(a, b) = 1$, то $au + bv = 1$ для некоторых $u, v \in \mathbb{Z}$. Умножим равенство $ax + by = 0$ на u:

$$u(ax + by) = uax + uby = (1 - bv)x + uby = x + b(uy - vx) = 0.$$

Поэтому $x = tb$ при некотором $t \in \mathbb{Z}$. Но тогда $by = -ax = -tab$ и $y = -ta$. С другой стороны, любая пара $(tb, -ta)$ является решением уравнения $ax + by = 0$. □

Из этой леммы и предыдущего утверждения получаем общую формулу для решений линейного диофантова уравнения.

Теорема 13.3. *Пусть* $\text{НОД}(a, b) \mid c$*,* $a\tilde{x}_0 + b\tilde{y}_0 = c$*. Тогда множество решений уравнения* (13.1) *— это множество пар*

$$\big(\tilde{x}_0 + tb/\,\text{НОД}(a, b), \tilde{y}_0 - ta/\,\text{НОД}(a, b)\big), \quad t \in \mathbb{Z}.$$

13.2. Свойства отношения делимости

Отношение делимости $a \mid b$ является отношением нестрогого частичного порядка на положительных целых числах (будем обозначать это множество $\mathbb{N}_+$). Антисимметричность: если $a \mid b$ и $b \mid a$, то $a = kb = k\ell a$, откуда получаем $k\ell = 1$, то есть $k = \ell = 1$. Транзитивность: если $c = kb$ и $b = \ell a$, то $c = k\ell a$.

Число 1 является минимальным элементом в этом порядке: $1 \mid n$ для любого n. А какие числа непосредственно следуют за 1 в этом порядке?

Определение 13.4. Целое положительное число называется *простым*, если оно больше 1 и делится только на 1 и на само себя.

Числа, которые не являются простыми и не равны 1, называются *составными*.

Число 1 исключительное: не простое и не составное.[2)]

[1)] Спасибо Дарье Гвоздевой за замеченную опечатку.

[2)] Спасибо Амиру Камалову за замеченную неточность.

Это определение на языке порядков как раз и означает, что каждое простое число непосредственно следует за 1 в порядке делимости, а каждое составное — нет (если $n = k\ell$, $k, \ell > 1$, то $1 \mid k \mid n$).

Среди маленьких чисел простых довольно много, вот первые из них:

$$2,\ 3,\ 5,\ 7,\ 9,\ 11,\ 13,\ 15,\ 17,\ 19,\ \ldots$$

Нетрудно понять, что далее лакуны между простыми числами могут становиться сколь угодно большими.

Утверждение 13.5. *Для любого n найдётся такое k, что все числа k, $k+1$, …, $k+n$ составные.*

Доказательство. Возьмём $k = 2 + (n+2)! = 2 + 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n+2)$. Тогда $k + i$ делится на $2 + i$ для любого i от 0 до n. □

Однако простых чисел бесконечно много. (Это одна из самых старых теорем в математике.)

Теорема 13.6 (Евклид(?))**.** *Простых чисел бесконечно много.*

Доказательство. Нам нужен такой факт: любое целое число > 1 делится на простое. Доказательство полной индукцией (по всем меньшим числам) по величине числа. База $n = 2$ очевидна, а шаг индукции состоит в том, что либо число n простое, либо делится на какое-то меньшее число k. Применяя индуктивное предположение к числу k, получаем простой делитель для n.

Теперь рассмотрим любое конечное множество простых чисел p_1, p_2, …, p_s. Число $p_1 \cdot p_2 \cdot \ldots \cdot p_s + 1$ даёт остаток 1 при делении на p_1, p_2, …, p_s. Значит, его простые делители (а они существуют, как мы показали выше) не принадлежат этому множеству. □

13.2.1 Основная теорема арифметики

Чтобы увидеть более тонкие свойства порядка делимости, нужно использовать знаменитый факт об однозначности разложения на простые. В силу важности это утверждение называется основной теоремой арифметики.

Теорема 13.7 (Основная теорема арифметики)**.** *Всякое целое положительное число, большее 1, разлагается на простые множители, причём единственным образом: любые два разложения отличаются только перестановкой сомножителей.*

В доказательстве потребуется такое утверждение.

Лемма 13.8. *Если p — простое число, то из $p \mid xy$ следует, что $p \mid x$ или $p \mid y$.*

Доказательство. Из теоремы об обратимости вычетов следует, что если $x \not\equiv 0 \pmod p$, то x обратим. Пусть $xz \equiv 1 \pmod p$. Тогда из $xy \equiv 0 \pmod p$ следует $0 \equiv xyz \equiv 1 \cdot y = y \pmod p$.

Итак, если $p \nmid x$, то $p \mid y$, что и требовалось. □

Замечание 13.1. В этой лемме простота p существенна. Например, $4 \mid 6 \cdot 10$, но $4 \nmid 6$, $4 \nmid 10$.

Доказательство теоремы 13.7. Существование разложения. (Полная) индукция по величине числа. База $n = 2$ очевидна. Шаг индукции. Как мы уже проверяли в доказательстве бесконечности простых чисел, каждое число > 1 делится на простое. Если число n простое, то получилось разложение (из одного сомножителя). Иначе $n = k\ell$, $1 < k, \ell < n$. Индуктивное предположение говорит, что у k и ℓ есть разложения на простые множители. Соединяя их, получим разложение n на простые.

Единственность разложения. Пусть некоторое число имеет два *различных* разложения на простые множители (то есть, разложения отличаются не только порядком множителей. Приравняем эти разложения и сократим общие множители. По предположению сократится не всё и получаем равенство

$$p_1 \cdot p_2 \cdot \ldots \cdot p_r = q_1 \cdot q_2 \cdot \ldots \cdot q_s,$$

здесь числа p_i отличаются от q_j (иначе возможно сокращение на общий множитель).

Левая часть делится на p_1. А правая часть равна произведению чисел, ни одно из которых не делится на p_1: они ведь простые и p_1 среди них по предположению нет. Теперь применим лемму 13.8 и придём к противоречию. □

13.2.2 Целые положительные числа с точки зрения умножения

Из основной теоремы арифметики становится понятно, как устроены целые положительные числа относительно умножения. Упорядочим простые числа по возрастанию:

$$2 = p_1 < p_2 < \cdots < p_k < \ldots$$

Возьмём любое целое положительное число n, разложим его в произведение простых и сгруппируем одинаковые множители. Простое число p_i встречается в этом разложении $a_i \geqslant 0$ раз. Если $a_i = 0$, то p_i не делит n, такое будет выполняться для всех i, начиная с некоторого (наверняка с того i, для которого $p_i > n$, но вообще-то раньше). Получаем *каноническое разложение* — формально бесконечное произведение

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_k^{a_k} \cdot \ldots, \tag{13.3}$$

в котором показатели a_i, за исключением конечного числа, равны 0. Последовательности (a_i) с таким свойством называются *финитными*.

Каждой финитной последовательности (a_i) целых неотрицательных чисел можно сопоставить целое положительное число n. По основной теореме арифметики это соответствие взаимно однозначно.

Отношение делимости в терминах последовательности показателей простых в разложении (13.3) выражается очень просто.

Утверждение 13.9. *Пусть числу n соответствует последовательность показателей (a_i), а числу k — (b_i). Тогда $k \mid n$ равносильно $b_i \leqslant a_i$ для всех i.*

Доказательство. В одну сторону: если $b_i \leqslant a_i$ для всех i, то

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_k^{a_k} \cdot \ldots = \left(p_1^{b_1} \cdot p_2^{b_2} \cdot \ldots \cdot p_k^{b_k} \cdot \ldots\right) \cdot \left(p_1^{a_1-b_1} \cdot p_2^{a_2-b_2} \cdot \ldots \cdot p_k^{a_k-b_k} \cdot \ldots\right)$$ [3)]

и $k \mid n$.

В другую сторону: если $k \mid n$, то из разложения n на простые возможно выделить разложение k на простые (перегруппируем множители). Поэтому $b_i \leqslant a_i$ для всех i. □

Условие на последовательности в этом утверждении говорит, что $(b_i) \leqslant (a_i)$ в отношении покоординатного порядка. Собирая все эти рассуждения вместе, получаем такую теорему.

Теорема 13.10. *Порядок делимости на целых положительных числах изоморфен покоординатному порядку на финитных последовательностях целых неотрицательных чисел.*

Как следствие получаем простые формулы для НОД и НОК в терминах последовательностей показателей в разложении на простые.

Лемма 13.11. *Пусть числу n соответствует последовательность показателей (a_i), а числу k — последовательность (b_i). Тогда* НОД(n,k) *соответствует последовательность* $(\min(a_i,b_i))$*, а* НОК(n,k) *— последовательность* $(\max(a_i,b_i))$*.*

Заметьте, что мы пока даже не доказали существование наименьшего общего кратного (в формулировке леммы оно обозначено НОК(n,k)). Это будет частью доказательства леммы.

Доказательство. Пользуемся изоморфизмом порядков (теорема 13.10).

Если $x_i \leqslant a_i$ и $x_i \leqslant b_i$ для всех i, то $x_i \leqslant \min(a_i,b_i)$ для всех i. Поскольку $\min(a_i,b_i) \leqslant a_i$ и $\min(a_i,b_i) \leqslant b_i$ для всех i, эта последовательность и будет последовательностью показателей для НОД(n,k).

Точно так же рассуждаем про кратные, знак $\leqslant$ нужно всюду заменить на $\geqslant$, а $\min$ — на $\max$. □

Следствие 13.12. НОД$(n,k) \cdot$ НОК$(n,k) = kn$.

Доказательство. Достаточно проверить, что $\max(a,b)+\min(a,b) = a+b$. Поскольку a, b входят в равенство симметрично, считаем без ограничения общности, что $a \leqslant b$. Тогда $\min(a,b) = a$, $\max(a,b) = b$. □

[3)] Спасибо Михаилу Сазонову за замеченную опечатку.

13.3. Малая теорема Ферма

Теорема 13.13. *Если p — простое число, то*

$$a^{p-1} \equiv 1 \pmod{p}$$

при любом a, не делящемся на p.

Доказательство. На ненулевых вычетах по модулю p рассмотрим функцию $x \mapsto ax$. Так как a взаимно просто с p, эта функция обратима, обратная задаётся как $x \mapsto bx$, где $ab \equiv 1 \pmod{p}$. Действительно, композиция этих функций переводит x в $abx \equiv x \pmod{p}$.

Рассмотрим ориентированный граф этой функции (биекции). Как мы уже знаем, этот граф распадается на циклы. Возьмём цикл, содержащий 1:

$$1 \mapsto a \mapsto a^2 \mapsto a^{k-1} \mapsto a^k \equiv 1 \pmod{p}.$$

Здесь k — минимальная степень a, равная 1 по модулю p. Если взять другой цикл, то получим

$$b \mapsto ab \mapsto a^2b \mapsto a^{k-1}b \mapsto a^kb \equiv 1 \cdot b = b \pmod{p}.^{4)}$$

Почему этот цикл на замкнётся раньше? Если $a^rb \equiv b \pmod{p}$, то $a^r \equiv 1 \pmod{p}$: вычет b обратим, $bc \equiv 1 \pmod{p}$; домножая первое сравнение на c, получаем второе сравнение. Поскольку k — минимальная степень a, равная 1 по модулю p, то $r = k$.

Итак, $p-1$ ненулевых вычетов по модулю p разбиваются на циклы одинакового размера. Поэтому $p-1$ делится на k, то есть $p-1 = km$ при каком-то m. Тогда

$$a^{p-1} \equiv a^{km} \equiv (a^k)^m \equiv 1^m = 1 \pmod{p},$$

что и требовалось доказать. □

13.4. Теорема Эйлера

Аналогичное малой теореме Ферма утверждение выполняется для любого модуля. Но теперь нужно учесть, что некоторые вычеты необратимы.

Определение 13.14. *Функция Эйлера* $\varphi(n)$ равна количеству остатков по модулю n, взаимно простых с n.

Пример 13.1. $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(4) = 2$. Проверим последнее равенство. Остатки $0, 2$ не взаимно просты с 4, а остатки $1, 3$ — взаимно просты.

Теорема 13.15. *Пусть $n > 1$ — произвольное целое положительное число, а взаимно прост с n. Тогда*

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

[4)] Спасибо Яну Максимову за замеченную опечатку.

Доказательство. Рассуждаем так же, как при доказательства малой теоремы Ферма. Функция $x \mapsto ax$ является биекцией на множестве вычетов, взаимно простых с n. Циклы ориентированного графа этой функции имеют одинаковую длину (буквально повторяем аргументы из доказательства малой теоремы Ферма). Эта длина k — наименьшая степень a, равная 1 по модулю n.

Значит, $k \mid \varphi(n)$ и $a^{\varphi(n)} \equiv 1 \pmod{n}$ в силу точно того же вычисления, что в доказательстве малой теоремы Ферма. □

Лекция 14

Числа–3

14.1. Китайская теорема об остатках

Так по традиции называется следующее утверждение.

Теорема 14.1. *Пусть числа u и v взаимно просты, и пусть a и b — любые целые числа. Тогда можно найти число x, для которого $x \equiv a \pmod{u}$ и одновременно $x \equiv b \pmod{v}$. В промежутке от 0 до $uv - 1$ такое число единственное.*

Для некоторых u, v, a и b подобрать число x из китайской теоремы нетрудно.

Пример 14.1. Найдём положительное целое число, которое даёт остаток 1 при делении на 27 и остаток 1 при делении на 31. Ясно, что этим условиям удовлетворяет число 1.

Пример 14.2. Найдём положительное целое число, которое даёт остаток 26 при делении на 27 и остаток 30 при делении на 31. Для этого заметим, что $26 \equiv -1 \pmod{27}$ и $30 \equiv -1 \pmod{31}$. Значит, число $27 \cdot 31 - 1$ даёт нужные остатки по обоим модулям и лежит в промежутке от 0 до $27 \cdot 31 - 1$.

Пример 14.3. А как найти число, которое даёт остаток 1 при делении на 27 и остаток 30 при делении на 31? Другими словами, по модулю 27 это число сравнимо с 1, а по модулю 31 оно сравнимо с -1.

Если перебрать все числа в диапазоне от 0 до $27 \cdot 31 - 1$, такое число обязательно найдётся. Можно ли обойтись без перебора?

Искомое число x имеет вид $27k + 1$, k — целое. С другой стороны, оно же имеет вид $31t - 1$, t — целое. Найдём решение сравнения $27k + 1 \equiv -1 \pmod{31}$. Для этого можно использовать расширенный алгоритм Евклида. Однако для «ручных вычислений» удобнее более гибкая его версия, в которой мы умножаем обе части сравнения на подходящие числа, взаимно простые с модулем (такие вычеты обратимы и потому получается равносильное сравнение).

Получаем такую цепочку равносильных сравнений:

$$\begin{aligned} &27k \equiv -2 \pmod{31} && (27 \equiv -4 \pmod{31}),\\ &-4k \equiv -2 \pmod{31} && (-4 \cdot 8 \equiv -1 \pmod{31}),\\ &-k \equiv -16 \pmod{31}, \quad k \equiv 16 \pmod{31}. \end{aligned}$$

Значит, $k = 16+31s$, $x = 27k+1 = 27\cdot 16+1+27\cdot 31\cdot s$. Искомые числа различаются по крайней мере на $27\cdot 31$, поэтому в интервале от 1 до $27\cdot 31$ такое число единственное: $27 \cdot 16 + 1 = 433$.

Вычисления этого примера легко обобщаются до доказательств существования и единственности в китайской теореме.

Доказательство китайской теоремы. **Существование.** Поскольку $\text{НОД}(u, v) = 1$, существует такое $\widetilde{u}$, что $u \cdot \widetilde{u} \equiv 1 \pmod{v}$. Поэтому числа вида

$$x = a + u\widetilde{u}(b - a) + suv, \quad s \in \mathbb{Z},$$

удовлетворяют обоим сравнениям: $x \equiv a \pmod{u}$ [1] и $x \equiv a + 1 \cdot (b - a) = b \pmod{v}$.

Единственность. Пусть $x \equiv a \pmod{u}$, $x \equiv b \pmod{v}$ и $y \equiv a \pmod{u}$, $y \equiv b \pmod{v}$, а также $0 \leqslant x, y < uv$. Тогда $(x - y) \equiv 0 \pmod{u}$ и $(x - y) \equiv 0 \pmod{v}$. То есть $x - y = ku = tv$. Но $\text{НОД}(u, v) = 1$, поэтому k кратно v, а t кратно u (мы уже знаем по крайней мере два способа доказательства этого факта: с помощью решения однородных линейных диофантовых уравнений и с помощью основной теоремы арифметики). Таким образом, $x - y$ делится на uv. В промежутке от 0 до $uv - 1$ такое число единственное — это 0. Поэтому $x = y$, что и требовалось доказать. □

14.2. Мультипликативность функции Эйлера

Напомним, что $\varphi(n)$ — это количество вычетов по модулю n, взаимно простых с n.

Теорема 14.2 (мультипликативность функции Эйлера)**.** *$\varphi(uv) = \varphi(u)\varphi(v)$, если u и v взаимно просты.*

Доказательство. Китайская теорема гарантирует, что для любой пары остатков a (по модулю u) и b (по модулю v) существует ровно один такой остаток c по модулю uv, что $c \equiv a \pmod{u}$ и $c \equiv b \pmod{v}$. То есть имеется биекция между парами остатков по модулям u, v и остатками по модулю их произведения.

Докажем, что та же биекция устанавливает взаимно однозначное соответствие между вычетами по модулю uv, взаимно простыми с uv, и парами вычетов (a, b) по модулям u, v соответственно, которые взаимно просты с u (соответственно, с v).

Если a взаимно просто с u, то и c взаимно просто с u (они лежат в одном классе выечтов по модулю u; аналогично, если b взаимно просто с v, то и c взаимно просто

[1] Спасибо Михаилу Сазонову за замеченную опечатку.

с v. Но тогда c взаимно просто просто с uv: все простые делители uv являются делителями либо u, либо v и потому не делят c; поэтому НОД$(c, uv) = 1$.

Верно и обратное: если НОД$(c, uv) = 1$, то НОД$(a, u) = 1$ и НОД$(b, v) = 1$. Чтобы доказать это утверждение, запишем 1 как целочисленную линейную комбинацию c и uv:

$$1 = xc + yuv.$$

Но $c = ku + a$, поэтому $1 = xa + (kx + vy)u$, откуда следует, что НОД$(a, u) = 1$. Аналогично доказывается, что НОД$(b, v) = 1$.

Чтобы закончить доказательство, осталось заметить, что пар остатков, взаимно простых с модулями u, v, ровно $\varphi(u)\varphi(v)$ штук; остатков, взаимно простых с uv, ровно $\varphi(uv)$ штук. Построенная биекция доказывает, что эти числа равны. □

Из свойства мультипликативности вытекает сравнительно простой способ вычисления функции Эйлера $\varphi(n)$, если известно разложение числа[2)] n на простые множители (такое разложение для небольших n легко находится).

Теорема 14.3. *Пусть* $n = p_1^{a_1} p_2^{a_2} \cdot \ldots \cdot p_s^{a_s}$, $a_i > 0$, p_i *— различные простые. Тогда*

$$\varphi(n) = \prod_{i=1}^{s} (p_i^{a_i} - p_i^{a_i - 1}) = n \prod_{i=1}^{s} \left(1 - \frac{1}{p_i}\right).$$

Поскольку только степени простых невозможно представить в виде произведения меньших взаимно простых чисел, свойство мультипликативности сводит вычисление функции Эйлера к вычислению функции Эйлера для степеней простых. Но для степеней простых нужно отдельное рассуждение.

Утверждение 14.4. $\varphi(p^n) = p^n(1 - 1/p) = p^{n-1}(p - 1)$ *для простого* p.

Доказательство. Какие остатки не взаимно просты с p^n? В точности те, которые делятся на p (других простых делителей у p^n нет). Всего остатков p^n, а делится на p ровно доля $1/p$: в каждой группе из p последовательных остатков на p делится только один. Поэтому взаимно простых остатков $p^n - p^n/p = p^n(1 - 1/p)$. □

Заметьте, что

$$\varphi(p^2) = p^2 - p \neq (p - 1)^2 = \varphi(p) \cdot \varphi(p)$$

для простого p. То есть условие взаимной простоты множителей существенно для свойства мультипликативности.

Доказательство теоремы 14.3. Применяя мультипликативность функции Эйлера и формулу для функции Эйлера от степени простого, получаем:

$$\varphi(n) = \prod_{i=1}^{s} \varphi(p_i^{a_i}) = \prod_{i=1}^{s} (p_i^{a_i} - p_i^{a_i - 1}) = n \prod_{i=1}^{s} \left(1 - \frac{1}{p_i}\right),$$

что и требовалось. □

[2)] Спасибо Артёму Максаеву за замеченную опечатку.

Пример 14.4. Вычислим $\varphi(360)$. Разложение на простые множители для такого маленького числа легко находится: $360 = 2^3 \cdot 3^2 \cdot 5$. Поэтому

$$\varphi(360) = \varphi(2^3) \cdot \varphi(3^2) \cdot \varphi(5) = (8-4) \cdot (9-3) \cdot (5-1) = 96.$$

14.3. Оценки количества простых чисел

Выполнять сложение и умножение остатков по небольшим модулям легче, чем выполнять эти операции с большими числами. Поэтому есть такой способ выполнения арифметических действий с большими целыми числами: представлять числа наборами остатков по небольшим модулям и выполнять вычисления по этим модулям независимо. Конечно, если нужно сравнивать числа, придётся преобразовывать их в позиционную систему, а это уже долгое дело. Однако, если требуется выполнять большое количество арифметических операций без сравнений чисел (например, как при решении систем линейных уравнений), этот способ выглядит привлекательно.

Этот способ хорош, если достаточно длинные числа можно представлять по модулям сравнительно маленьких взаимно простых чисел. А для этого требуется иметь хороший запас небольших простых чисел. Та же потребность — значительный запас небольших простых чисел — возникает и во многих других алгоритмических вопросах. Да и с чисто математической точки зрения интересно, сколько же простых чисел. Да, мы уже знаем 2500 лет, что их бесконечно много. Но факториалов чисел тоже бесконечно много, однако последовательность факториалов растёт очень быстро.

Обозначим через $\pi(x)$ количество простых чисел, не превосходящих x. Тогда *асимптотический закон распределения простых чисел* утверждает, что

$$\pi(x) \sim \frac{x}{\ln x}$$

(логарифм натуральный). Асимптотический закон — одна из жемчужин математики. Он доказан 124 года назад и самое естественное его доказательство использует анализ функций комплексной переменной.

Для теоретической информатики такая высокая точность, в общем-то, не нужна. Нам достаточно оценок Чебышёва

$$C_1 \frac{x}{\log x} \leqslant \pi(x) \leqslant C_2 \frac{x}{\log x} \tag{14.1}$$

для некоторых констант C_1, C_2. Здесь уже основание логарифма неважно, оно влияет только на константы.

Доказательство оценок Чебышёва не требует изощрённого анализа. Но зато нужно больше узнать про факториалы и биномиальные коэффициенты. Идея доказательства состоит в том, чтобы изучить разложение «среднего» биномиального коэффициента на простые множители.

Формула бинома даёт равенство

$$(1+1)^{2n} = \binom{2n}{0} + \binom{2n}{1} + \cdots + \binom{2n}{n} + \cdots + \binom{2n}{2n-1} + \binom{2n}{2n},$$

причём самый большой биномиальный коэффициент — средний. Получаем оценки среднего коэффициента

$$\frac{2^{2n}}{2n+1} < \binom{2n}{n} < 2^{2n}$$

(первое неравенство строгое, потому что $2n+1$ нечётное и потому $2n+1 \nmid 2^{2n}$). Далее мы получим оценки количества простых сомножителей в среднем коэффициенте сверху и снизу, откуда и будут следовать оценки Чебышёва. Эти оценки следуют из таких утверждений.

Для верхней оценки нужно такое утверждение.

Утверждение 14.5. *Любое простое p, для которого $n < p \leqslant 2n$, делит $\binom{2n}{n}$.*

Доказательство. Формула для биномиальных коэффициентов

$$\binom{2n}{n} = \frac{(2n)!}{n!\,n!}$$

показывает, что такое p входит в числитель дроби, но не входит в знаменатель. □

Замечание 14.1. В теории чисел известен так называемый постулат Бертрана: для любого n между n и $2n$ есть простое число. Это ненастоящий постулат, а теорема с таким странным названием. Утверждение 14.5 на него не опирается. Постулат Бертрана можно вывести из оценок Чебышёва, это нетрудное упражнение.

Для нижней оценки нужно более сложное утверждение. Пусть

$$\binom{2n}{n} = p_1^{t_1} \cdot \ldots \cdot p_s^{t_s} \; — \tag{14.2}$$

разложение среднего биномиального коэффициента на простые, здесь p_i все различны. Для нижней оценки нужно ограничить сверху величину каждого множителя $p_i^{t_i}$.

Утверждение 14.6. $p_i^{t_i} \leqslant 2n^2$.

Это уже совсем неочевидное утверждение. Сам биномиальный коэффициент намного больше $2n^2$ и почему он не делится на большую степень 2, например, непонятно на первый взгляд. Тут и потребуются некоторые дополнительные факты о делителях факториалов и биномиальных коэффициентов.

Лемма 14.7. *Пусть p — простое число. Наибольшая степень p, которая делит $n!$, равна*

$$\left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \cdots + \left\lfloor \frac{n}{p^k} \right\rfloor + \ldots$$

Хотя формула содержит бесконечное количество слагаемых, для любого n лишь конечное количество слагаемых отлично от нуля: $n < p^k$ для достаточно больших k, и тогда $\lfloor n/p^k \rfloor = 0$.

Доказательство. Как известно, $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$. Каждый множитель kp, $1 \leqslant k \leqslant \lfloor n/p \rfloor$ делится на p. Поэтому степень p в разложении $n!$ по степеням простых не меньше $\lfloor n/p \rfloor$.

Однако каждый множитель kp^2, $1 \leqslant k \leqslant \lfloor n/p^2 \rfloor$ делится на p^2. Это даёт прибавку к степени $\lfloor n/p^2 \rfloor$.

Продолжая это рассуждение, получаем формулу из условия леммы. □

Формулу из леммы 14.7 можно записать в конечном виде, воспользовавшись представлением n в p-ичной системе счисления. А именно, пусть

$$n = n_k p^k + n_{k-1} p^{k-1} + \cdots + n_1 p + n_0 = \overline{(n_k n_{k-1} \ldots n_1 n_0)}_p, \quad 0 \leqslant n_i < p.$$

Тогда наибольшая степень p в разложении $n!$ по степеням простых равна

$$\begin{aligned} &\overline{(n_k n_{k-1} \ldots n_2 n_1)}_p + \\ &\overline{(n_k n_{k-1} \ldots n_2)}_p + \\ &\ldots + \\ &\overline{(n_k n_{k-1})}_p + \\ &\overline{(n_k)}_p \,, \end{aligned} \tag{14.3}$$

так как i-е сверху слагаемое равно $\lfloor n/p^i \rfloor$ (запись в p-ичной системе счисления).

Наша следующая цель — найти наибольшую степень p, на которую делится биномиальный коэффициент, который нам будет удобно представлять как

$$\binom{a+b}{a} = \frac{c!}{a!b!}\,, \quad c = a + b.$$

Это вариант теоремы Люка, которая выражает через цифры p-ичных представлений a и b остаток при делении биномиального коэффициента на p.

Теорема 14.8 (теорема Люка)**.** *Пусть p — простое число. Наибольшая степень p, которая делит $\binom{a+b}{a}$, равна количеству переносов при сложении чисел a и b, записанных в p-ичной системе счисления.*

Доказательство. Обозначим $c = a + b$. По лемме 14.7 искомая степень равна

$$\begin{pmatrix} \overline{(c_k c_{k-1} \ldots c_2 c_1)}_p + \\ \overline{(c_k c_{k-1} \ldots c_2)}_p + \\ \ldots + \\ \overline{(c_k c_{k-1})}_p + \\ \overline{(c_k)}_p \end{pmatrix} - \begin{pmatrix} \overline{(a_k a_{t-1} \ldots a_2 a_1)}_p + \\ \overline{(a_k a_{t-1} \ldots a_2)}_p + \\ \ldots + \\ \overline{(a_k a_{k-1})}_p + \\ \overline{(a_k)}_p \end{pmatrix} - \begin{pmatrix} \overline{(b_k b_{k-1} \ldots b_2 b_1)}_p + \\ \overline{(b_k b_{k-1} \ldots b_2)}_p + \\ \ldots + \\ \overline{(b_k b_{k-1})}_p + \\ \overline{(b_k)}_p \end{pmatrix} \tag{14.4}$$

Сравним числа в первой строке формулы (14.4), то есть

$$\overline{(c_k c_{k-1} \ldots c_2 c_1)}_p \quad \text{и} \quad \overline{(a_k a_{k-1} \ldots a_2 a_1)}_p + \overline{(b_k b_{k-1} \ldots b_2 b_1)}_p$$

Если при сложении a и b в p-ичной системе счисления переноса в первый разряд не было, эти числа равны. Если перенос был, то первое на единицу больше второго (при сложении в следующий разряд переносится не больше 1).

Аналогично сравниваются и остальные слагаемые в (14.4): в i-й строчке получается 0, если не было переноса в i-й разряд при сложении a и b в p-ичной системе счисления; и получается 1, если перенос был.

Таким образом, значение формулы (14.4) равно общему количеству переносов при сложении a и b в p-ичной системе счисления. □

Доказательство утверждения 14.6. Пусть p^t — наибольшая степень простого числа p, которая делит $\binom{2n}{n}$. По теореме Люка число t равно количеству переносов при сложении числа n с числом n в p-ичной системе счисления.

Если $n = \overline{(n_k n_{k-1} \dots n_2 n_1 n_0)}_p$, $n_k > 0$, количество переносов не больше $k+1$ (всего $k+1$ разряд используется). Поэтому $p^t \leqslant p^{k+1} \leqslant pn \leqslant 2n^2$ (последнее неравенство следует из того, что простой делитель биномиального коэффициента $\binom{2n}{n}$ не превосходит $2n$). □

Теперь мы готовы к доказательству оценок Чебышёва. Будем доказывать оценки для чётных чисел вида $2n$. Этого достаточно, так как $\pi(2n+1) - \pi(2n)$ не больше 1.

Доказательство нижней оценки в (14.1). Утверждение 14.6 говорит, что вклад в биномиальный коэффициент каждой степени простого не превосходит $2n^2$. Поэтому

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

Логарифмируя, получаем нижнюю оценку Чебышёва. □

Доказательство верхней оценки в (14.1). Используем утверждение 14.5. Средний биномиальный коэффициент меньше 2^{2n}. Поэтому

$$n^{\pi(2n)-\pi(n)} < \prod_{n<p\leqslant 2n} p < \binom{2n}{n} < 2^{2n},$$

то есть

$$\pi(2n) - \pi(n) < \frac{2n}{\log_2 n} \,. \tag{14.5}$$

Складывая несколько неравенств вида (14.5) для $2n = 2^t, 2^{t-1}, \dots, 2^2$, получаем

$$\pi(2^t) - \pi(2) < \frac{2^t}{t-1} + \frac{2^{t-1}}{t-2} + \dots + \frac{4}{1} \,.$$

Отсюда уже следует верхняя оценка Чебышёва, нужно только аккуратно оценить сверху сумму в правой части. Достаточно это сделать при больших t, ведь нас в итоге интересует асимптотическая оценка. Давайте оценим сверху первые $\lfloor 0.1t \rfloor$

слагаемых в этой сумме как $2^{t-i}/(0.9t)$, а в оценке остальных слагаемых заменим знаменатели на 1. Получаем

$$\frac{2^t}{t-1}+\frac{2^{t-1}}{t-2}+\cdots+\frac{4}{1}<\frac{2^t+\cdots+2^{t-\lfloor 0.1t\rfloor+1}}{0.9t}+(2^{t-\lfloor 0.1t\rfloor}+\cdots+2^2)<$$
$$<\frac{10}{9}\cdot\frac{2^{t+1}}{t}+2^{t-\lfloor 0.1t\rfloor}\ .$$

Второе слагаемое $o(2^t)$, поэтому получаем искомую верхнюю оценку Чебышёва. $\square$

Лекция 15

Элементарная теория вероятностей–1

Понятие вероятности является фундаментальным не только для математики и информатики, но и для всех естественных наук. Это сложное понятие и мы не будем изучать его во всей общности. Мы здесь ограничимся очень узким взглядом на вероятность, который, тем не менее, очень важен для многих разделов информатики. Как станет ясно дальше, мы по сути продолжм изучение комбинаторики. Но теперь будем считать не в штуках, а в долях.

Неформальное понимание вероятности некоторого события состоит в том, что при многократном повторении экспериментов частота события примерно равна некоторому числу, которое и называется *вероятностью* события. Мы сейчас опишем простую математическую модель, в которой можно определять вероятности строго. Соответствие между этой моделью и неформальным пониманием вероятности есть, но оно не столь просто и мы его подробно изучать не будем.

15.1. Определения

Вероятностным пространством называется конечное множество U, его элементы называются *возможными исходами* (для краткости слово «возможный» будем пропускать). Функция $\mathbf{Pr}\colon U \to [0,1]$, удовлетворяющая соотношению $\sum_{x\in U} \mathbf{Pr}[x] = 1$, называется *вероятностным распределением*, а число $\mathbf{Pr}[x]$ 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 $\mathbf{Pr}[A] = \sum_{x\in A} Pr[x]$.

Вероятностное пространство с заданным на нём вероятностным распределением называется *вероятностной моделью*.

В *модели с равновозможными исходами* функция $\mathbf{Pr}$ задается формулой $\mathbf{Pr}[x] = 1/|U|$ для всякого $x \in U$ (такое распределение называют также *равномерным*). Тогда вероятность события A равна $\mathbf{Pr}[A] = |A|/|U|$. Подсчёт вероятностей в такой модели сводится к подсчёту количества элементов в множестве A. Это отличается от перечислительной комбинаторики лишь тем, что мы делим количество элементов в A на общее количество элементов во всём множестве исходов U.

Приведём несколько примеров равномерных распределений на разных множествах и вычисления вероятностей событий. Для некоторых примеров есть устоявшиеся названия: подбрасывание монеты, подбрасывание игральной кости и т.п. Нужно понимать, что это условные названия, которые приняты для наглядности. Насколько подбрасывания реальной монеты соответствуют математической модели, — трудный вопрос и мы его здесь не обсуждаем.

Пример 15.1 («Подбрасывание монеты»)**.** В случае подбрасывания монетки исходом является выпадение одной из её сторон, орла или решки. Всего исходов, таким образом, два, и они считаются равновозможными. Обычно удобнее говорить не о сторонах монетки, а отождествить их с числами 0 и 1.

Итак, вероятностное пространство: числа 0 и 1. Все исходы равновозможны.

Пример 15.2 («Подбрасывание 6 монет»)**.** Вероятностное пространство: двоичные последовательности длины 6. Все исходы равновозможны. Можно представить, что эти слова являются записью результатов 6 подбрасываний монеты.

Пример события, то есть множества в этом пространстве: ровно три элемента последовательности равны 1 (на неформальном жаргоне говорят «выпало три орла»). Чтобы найти вероятность этого события, нужно подсчитать количество исходов в нём и поделить на общее количество исходов. Общее количество двоичных последовательностей длины 6 равно 2^6. Количество последовательностей, в которых ровно три единицы, равно $\binom{6}{3}$.

Поэтому вероятность события равна

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

Пример 15.3 («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 15.4. Пусть вероятностное пространство состоит из двоичных слов длины n и вероятности всех событий «на i-й позиции в слове стоит 1» равны $1/2$. Следует ли из этого, что распределение равномерное?

Легко понять, что ответ отрицательный. Например, вероятности слов из одних нулей и одних единиц могут равняться $1/2$, а вероятности остальных слов равняться нулю.

Обобщая примеры 15.1–15.3, можно взять в качестве вероятностного пространства множество всех слов длины n в алфавите размера k. Такое вероятностное пространство содержательно соответствует тому, что из лототрона, в котором лежит k

различных шаров, n раз вытаскивается шар, который затем возвращается в лототрон. Ещё одно часто используемое название для такого эксперимента: «выборка с возвращением».

Пример 15.5 («Подбрасывание двух игральных костей»)**.** Если $k = 6$, то можно представить подбрасывание игральной кости (у кубика 6 граней). Вероятностное пространство в данном случае: последовательности (x_1, x_2) длины 2, состоящие из целых чисел в диапазоне от 1 до 6. Все исходы равновозможны.

Найдём 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 $6/36 = 1/6$.

В предыдущем примере с лототроном мы предполагали, что шары возвращаются в лототрон. Это не отвечает обычной практике розыгрышей лотерей. В них вытащенные шары не возвращаются. Таким розыгрышам отвечает другое вероятностное пространство: «выборка без возвращения». Это размещения, то есть множество последовательностей длины k в алфавите из n символов, в которых нет повторяющихся символов.

Пример 15.6 («Монотонный результат»)**.** Пусть $k = 3$, $n = 10$. Будем считать, что алфавит — это множество $[10] = \{1, 2, \ldots, 10\}$. Исходы — это последовательности длины 3 из различных чисел этого диапазона.

Найдём вероятность события «последовательность монотонно убывающая». Всего исходов $A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$. Монотонно убывающие последовательности находятся во взаимно однозначном соответствии с 3-элементными подмножествами множества $[10]$. Поэтому их $\binom{10}{3} = 10 \cdot 9 \cdot 8/(1 \cdot 2 \cdot 3) = 720/6$. Вероятность равна доле таких исходов, то есть $1/6$.

Пример 15.7 («Случайная перестановка»)**.** Это случай выборки без возвращения[1)] при $k = n$. Вероятностное пространство: перестановки $(a_1, a_2, \ldots, a_n)$ чисел от 1 до n. Все исходы равновозможны.

Посчитаем вероятность события «$a_1 = 1$». Благоприятные исходы образуют множество перестановок, в которых 1 стоит на первом месте. Остальные элементы стоят как угодно, поэтому благоприятные исходы находятся во взаимно однозначном соответствии с перестановками $n - 1$ элемента. Вероятность равна $(n-1)!/n! = 1/n$.

Вероятность события «$a_n = 1$» такая же. Теперь благоприятные исходы образуют множество перестановок, в которых 1 стоит на последнем месте. И в этом случае благоприятные исходы находятся во взаимно однозначном соответствии с перестановками $n - 1$ элемента.

[1)] Спасибо Сергею Кожаринову за замеченную опечатку.

15.2. Оценка объединения

События по определению являются множествами, поэтому для них определены все теоретико-множественные операции. В частности, дополнительное событие $\bar{A}$ это просто разность $U \setminus A$. Легко увидеть из определения распределения вероятностей, что всегда $\mathbf{Pr}[A] + \mathbf{Pr}[\bar{A}] = 1$:

$$1 = \sum_{x \in U} \mathbf{Pr}[x] = \sum_{x \in A} \mathbf{Pr}[x] + \sum_{x \notin A} \mathbf{Pr}[x] = \mathbf{Pr}[A] + \mathbf{Pr}[\bar{A}].$$

Эту формулу можно обобщить. События A и B, которые не могут произойти одновременно, то есть для которых $\mathbf{Pr}[A \cap B] = 0$, называются *несовместными*.

Лемма 15.1. *Если собятия A_i попарно несовместны, то*

$$\mathbf{Pr}\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbf{Pr}[A_i].$$

Доказательство. Содержательно это просто комбинаторное правило суммы на новом языке. Сумма вероятностей по объединению семейства попарно несовместных событий после перегруппировки слагаемых превращается в сумму по событиям вероятностей события:

$$\mathbf{Pr}\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{x \in \bigcup_{i=1}^{n} A_i} \mathbf{Pr}[x] = \sum_{i=1}^{n} \sum_{x \in A_i} \mathbf{Pr}[x] = \sum_{i=1}^{n} \mathbf{Pr}[A_i].$$

При переходе от второй суммы к третьей возможно, что некоторые исходы попадут в несколько слагаемых с разными значениями i. Однако события несовместны, поэтому вероятности таких исходов равны 0, так что равенство выполняется. □

Для произвольных событий равенство уже не выполняется. Но сохраняется неравенство в одну сторону.

Лемма 15.2 (оценка объединения)**.** *Для любых событий $A_1, \ldots, A_n \subseteq U$ выполняется неравенство*

$$\mathbf{Pr}\left[\bigcup_{i=1}^{n} A_i\right] \leqslant \sum_{i=1}^{n} \mathbf{Pr}[A_i].$$

Доказательство. И в левой, и в правой части стоит сумма вероятностей исходов. Каждый исход в левой сумме встречается и в правой (возможно, не один раз). Неравенство выолняется, так как вероятности неотрицательные. □

Это простое неравенство имеет множество применений. В основном они относятся к *вероятностному методу* доказательства существования объектов с заданными свойствами. Общая идея метода состоит в том, что если вероятность некоторого события положительная, то хотя бы один исход в событии есть. Оценка объединения часто позволяет оценивать вероятность дополнительного события.

Приведём один из простейших примеров такого рассуждения.

Подмножество вершин графа называется *кликой*, если любая пара вершин этого подмножества связана ребром. Подмножество вершин графа называется *независимым*, если любая пара вершин этого подмножества не связана ребром.

Теорема 15.3. *Для всякого $k \geqslant 3$ существует граф $G = (V, E)$ на $n = \lfloor 2^{(k-1)/2} \rfloor$ вершинах, в котором нет ни клики размера k, ни независимого множества размера k.*

Доказательство. Рассмотрим случайный граф на n вершинах. Вероятностное пространство U в данном случае состоит из всех графов на этом множестве вершин, все исходы равновозможны.

Всего таких графов $2^{\binom{n}{2}}$, потому что такие графы находятся во взаимно однозначном соответствии с подмножествами множества пар вершин. Всего пар вершин $\binom{n}{2}$, так что графов как раз $2^{\binom{n}{2}}$.

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 вероятность дополнительного собятия положительная и потому это дополнительное событие непусто, то есть существует граф без клик и независимых множеств размера k.

Разобьем событие A в объединение нескольких событий. Для этого для всякого подмножества $W \subseteq V$ множества из k вершин рассмотрим событие A_W: «в случайном графе множество W образует клику или независимое множество». Нетрудно видеть, что

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

а значит

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

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

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

Таким образом, при $k \geqslant 3$, $n = \lfloor 2^{(k-1)/2} \rfloor$ получаем неравенство

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

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

Это доказательство неконструктивное: примера графа, в котором нет клики и независимого множества, оно не даёт. Построить «явно» такой граф для всех k — давно известная и безнадёжно трудная комбинаторная задача.

15.3. Формула включений и исключений для вероятностей

Теорема 15.4. *Для всякой вероятностной модели и для произвольных множеств $A_1, \dots, A_n \subseteq U$ верно*

$$\begin{aligned}\mathbf{Pr}[A_1 \cup A_2 \cup \dots \cup A_n] &= \sum_i \mathbf{Pr}[A_i] - \sum_{i<j} \mathbf{Pr}[A_i \cap A_j] + \dots = \\ &= \sum_{\varnothing \neq S \subset \{1,2,\dots,n\}} (-1)^{|S|+1} \mathbf{Pr}\left[\bigcap_{i\in S} A_i\right].\end{aligned} \tag{15.1}$$

Если распределение равномерное, это та же самая формула включений и исключений, которая была в комбинаторике. Только теперь каждый член формулы является дробью с общим знаменателем $|U|$. Однако теорема верна для любого распределения.

Доказательство. Вспомним доказательство формулы включений и исключений с помощью индикаторных функций. В нём мы использовали равенство

$$\chi_A(x) = 1 - \big(1 - \chi_{A_1}(x)\big)\big(1 - \chi_{A_2}(x)\big) \dots \big(1 - \chi_{A_n}(x)\big),$$

которое переписывается как

$$\chi_A(x) = \sum_{S\neq\varnothing} (-1)^{|S|+1} \chi_{A_S}(x), \tag{15.2}$$

где

$$A_S = \bigcap_{i\in S} A_i.$$

Вероятность A выражается как сумма по всему вероятностному пространству

$$\mathbf{Pr}[A] = \sum_{u\in U} \mathbf{Pr}[u]\chi_A(u)$$

(каждый благоприятный исход даёт вклад 1 в сумму, неблагоприятные исходы дают вклад 0).

Для $A = A_1 \cup A_2 \cup \dots \cup A_n$ подставим в эту сумму формулу (15.2) и получим

$$\begin{aligned}\mathbf{Pr}[A] &= \sum_{x\in U} \mathbf{Pr}[x]\chi_A(x) = \sum_{x\in U} \mathbf{Pr}[x] \sum_{S\neq\varnothing} (-1)^{|S|+1} \chi_{A_S}(x) = \\ &= \sum_{S\neq\varnothing} (-1)^{|S|+1} \sum_{x\in U} \mathbf{Pr}[x]\chi_{A_S}(x),\end{aligned}$$

а это и есть формула включений и исключений. □

Одним из стандартных приложений формулы включений и исключений является следующая задача.

Лемма 15.5 (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 $\lim_{n\to\infty} \mathbf{Pr}[A_n] = 1/e$.*

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 посчитать вероятности событий $\bigcap_{i\in S} B_i$ для всевозможных $S \subseteq [n]$. Перестановки из этого события — это в точности перестановки, оставляющие на месте элементы из S, и переставляющие остальные элементы произвольным образом. Таких перестановок $(n - |S|)!$. Таким образом, для всякого S

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

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

$$\mathbf{Pr}\left[B_1 \cup \ldots \cup B_n\right] = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}\ .$$

Тогда для вероятности события A_n мы получаем формулу

$$\mathbf{Pr}[A_n] = 1 - \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}\ .$$

Эта сумма совпадает с началом ряда Тейлора для функции e^x в точке $x = -1$. Поскольку $e^{-1} = \sum_{k=0}^{\infty} (-1)^k/k!$, то

$$\lim_{n\to\infty} \mathbf{Pr}[A_n] = \frac{1}{e}\ ,$$

что и требовалось доказать. □

Лекция 16

Элементарная теория вероятностей–2

16.1. 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 произошло. 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

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

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

Определение условной вероятности можно переписать следующим образом:

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

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

При использовании условных вероятностей важно явно обозначать вероятностное пространство и распределение на нём, после чего пользоваться формальным определением условной вероятности. Интуитивные представления о «шансах» часто вводят в заблуждение.

Пример 16.1. 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?

Исходами в данном случае являются монеты и все шесть исходов равновозможны. Событие $A =$«выбранная монета золотая» состоит из 3 исходов. Событие $B =$ «выбран мешок с двумя золотыми монетами» состоит из двух исходов, причём оба они содержатся в A (это в точности монеты из этого мешка). Значит, $A \cap B = B$ и поэтому

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

Пример 16.2. Есть десять 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. Какова вероятность того, что в последней коробке шарик есть при условии, что в остальных коробках его нет?

Какое здесь вероятностное пространство? Перенумеруем коробки числами от 0 до 9 в порядке их открывания. Исход — это пара (i, j), где i — номер выбранной коробки, а j равно 1 (шарик положили в коробку) или 0 (не положили). Все исходы равновозможны.

Событие-условие «в коробках с номерами от 0 до 8 шарика нет» состоит из 11 исходов:

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

(первые 9 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 $1/11$.

16.2. Независимые события

Условная вероятность $\mathbf{Pr}[A \mid B]$ может быть как больше вероятности $\mathbf{Pr}[A]$, так и меньше её. В примерах 16.1 и 16.2 условная вероятность была больше. Неформально говоря, событие-условие «благоприятствовало» нашему событию. Легко придумать пример обратного. Пусть равновозможно выбирается число от 0 до 9. Вероятность выбрать нечётное число при условии, что выбрано чётное число, равна 0.

Бывает и так, что вероятность события A совпадает с условной вероятностью $\mathbf{Pr}[A|B]$. Неформально, событие B является «нейтральным» по отношению к A. Точный термин другой.

Событие A *не зависит от события* B, если

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

Из определения условной вероятности получаем эквивалентное определение независимости событий

$$\mathbf{Pr}[A \cap B] = \mathbf{Pr}[B] \cdot \mathbf{Pr}[A|B] = \mathbf{Pr}[B] \cdot \mathbf{Pr}[A], \tag{16.1}$$

откуда становится ясно, что отношение независимости симметрично. Второе достоинство этого определения состоит в том, что оно применимо и к событиям нулевой вероятности. Однако мы будем предполагать, если не оговорено противное, что отношение независимости определено для событий положительной вероятности.

Формула (16.1) обобщает правило произведения в комбинаторике. Если вероятностное пространство является декартовым произведением $U \times V$, а распределение равномерное, то события $A \times V$ и $U \times B$ независимы для любых $A \subseteq U$, $B \subseteq V$:

$$\frac{|A \times B|}{|U| \cdot |V|} = \frac{|A \times V|}{|U| \cdot |V|} \cdot \frac{|U \times B|}{|U| \cdot |V|} .$$

Пример 16.3. Пусть «честная» монета подбрасывается 2000 раз. События «первые 1000 подбрасываний дают в два раза больше орлов, чем решек» и «среди последние 1000 подбрасываний дают в два раза больше орлов, чем решек» независимы, хотя азартные игроки обычно в это не верят.

Условие равномерности распределения существенно.

Пример 16.4. В лототроне 36 шаров, пронумерованных числами от 1 до 36. Вытаскиваем два шара без возвращения. То есть вероятностное пространство — размещения из 36 по 2. Событие A =«первый шар чётный», событие B =«второй шар чётный». Независимы ли они?

Вместо размещений в качестве вероятностного пространства можно рассматривать любые упорядоченные пары чисел от 1 до 36, но при этом нужно считать, что вероятности пар (a, a) равны 0, а вероятности всех остальных пар одинаковы.

Вероятности событий A и B равны, что ясно из симметрии, и при это равны $1/2$ (на нужное место выбираем один из 18 чётных шаров, на второе место ставим какой угодно).

Равенство (16.1) не выполняется:

$$\mathbf{Pr}[A \cap B] = \frac{18 \cdot 17}{36 \cdot 35} \neq \frac{18}{36} \cdot \frac{18}{36} = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B],$$

поэтому события не являются независимыми. Вероятность того, что второй шар чётный при условии, что первый шар чётный, меньше вероятности, что второй шар чётный.

Пример 16.5. Подбрасываются две игральных кости. Независимы ли события $A =$ «сумма очков делится на 2» и $B =$«сумма очков делится на 3»?

Вероятностное пространство: упорядоченные пары чисел от 1 до 6. Вероятность первого события $1/2 \cdot 1/2 + 1/2 \cdot 1/2 = 1/2$ (первое слагаемое отвечает случаю, когда количество очков на обеих костях чётно, второе — на обеих костях нечётно). Вероятность второго события $1/3 \cdot 1/3 + 1/3 \cdot 1/3 + 1/3 \cdot 1/3 = 1/3$ (первое слагаемое отвечает случаю, когда количество очков на обеих костях делится на 3; второе — на первой остаток 1 от деления на 3, на второй — остаток 2 от деления на 3; третье — на первой остаток 2 от деления на 3, на второй — остаток 1 от деления на 3).

Теперь найдём вероятность пересечения этих событий. Если число делится на 2 и на 3, то оно делится на 6 (тогда и только тогда, как говорит китайская теорема). Проделывая аналогичный предыдущим двум случаям подсчёт, получаем

$$\mathbf{Pr}[A \cap B] = \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{6}.$$

Здесь i-е слагаемое отвечает исходу, в котором $i \bmod 6$ на первой кости и $-i \bmod 6$ на второй. Равенство (16.1) выполняется:

$$\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$$

поэтому события независимы.

16.3. Формула полной вероятности и формула Байеса

Эти две простые формулы очень популярны в приложениях теории вероятностей.

Лемма 16.1 (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 $\mathbf{Pr}[B_i] > 0$ для всякого i. Тогда для всякого события A*

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

Доказательство. Согласно свойству аддитивности вероятности,

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

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

Лемма 16.2 (формула Байеса)**.** *Если вероятность событий A и B положительна, то*

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

Доказательство. Выразим вероятность события $A \cap B$ через условные вероятности двумя способами:

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

Формула Байеса получается из второго равенства. □

Польза от формулы Байеса в том, что практическая оценка условных вероятностей несимметрична.

Пример 16.6. Болезнью К больны 1% людей в городе М. Тест П даёт положительный результат для 99% больных и для 1% здоровых. Я сдал тест и тот оказался положительным. Какова вероятность, что я болен?

События, о которых идёт речь: $A =$«болен», $B =$«тест дал положительный результат». Из данных о тесте мы знаем, что $\mathbf{Pr}[B|A] = 0.99$. Вероятность $\mathbf{Pr}[A]$ равна 0.01. Чтобы найти $\mathbf{Pr}[B]$, нужно использовать формулу полной вероятности:

$$\begin{aligned}\mathbf{Pr}[B] &= \mathbf{Pr}[A] \cdot \mathbf{Pr}[B|\text{«болен»}] + \mathbf{Pr}[\bar{A}] \cdot \mathbf{Pr}[B|\text{«здоров»}] = \\ &= 0.01 \cdot 0.99 + 0.99 \cdot 0.01 = 0.0198\end{aligned}$$

По формуле Байеса получаем ответ: $\mathbf{Pr}[A|B] = 0.01 \cdot 0.99/0.0198 = 0.5$.

Формула полной вероятности часто используется для обоснования рассуждений, основанных на симметрии задачи. Приведём два примера.

Пример 16.7. Пусть G — простой неориентированный граф с n вершинами $v_1, \ldots, v_n$, степень каждой из которых равна d. У такого графа $nd/2$ 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 равновозможно выбирается вершина, а затем равновозможно выбирается ребро, инцидентное этой вершине. Оказывается, эти распределения одинаковы.

Рассмотрим произвольное ребро e и событие A, означающее, что выбрано это ребро. Подсчитаем $\mathbf{Pr}[A]$ в каждом из распределений. В первом случае вероятность равна $2/nd$, так как распределение равномерное. Чтобы посчитать вероятность во втором случае, рассмотрим события B_v, состоящие в том, что была выбрана вершина v. Эти события образуют разбиение вероятностного пространства. Так что по формуле полной вероятности получаем

$$\mathbf{Pr}[A] = \sum_{v \in V} \mathbf{Pr}[A|B_v] \cdot \mathbf{Pr}[B_v].$$

На вершинах у нас задано равномерное распределение, так что $\mathbf{Pr}[B_v] = 1/n$ для каждого v. Теперь подсчитаем условную вероятность $\mathbf{Pr}[A|B_v]$. Если вершина v не является концом ребра e, то ребро никак не может быть выбрано, так что условная вероятность в этом случае равна нулю. Если же v является концом ребра e, то

вероятность, что мы выберем его равна $1/d$: мы равновероятно выбираем одно из d рёбер с концом в v. 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 16.8. 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 «$X \cap Y = \varnothing \mid Y = A$» и «$X \cap Y = \varnothing \mid Y = B$» одинаковы для любых A и B.

Из формулы полной вероятности получаем для любого множества A, $|A| = k$:

$$\mathbf{Pr}[X \cap Y = \varnothing] = \sum_{|S|=k} \mathbf{Pr}[X \cap Y = \varnothing \mid Y = S]\,\mathbf{Pr}[Y = S] =$$
$$= \mathbf{Pr}[X \cap Y = \varnothing \mid Y = A] \cdot \sum_{|S|=k} \mathbf{Pr}[Y = S] = \mathbf{Pr}[X \cap Y = \varnothing \mid Y = A].$$

Пусть $A = \{1, \ldots, k\}$. Вероятность выбрать k-элементное множество X, 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}. Получаем неожиданный с точки зрения интуиции факт: весьма малые случайные подмножества большого конечного множества почти заведомо пересекаются.

Лекция 17

Элементарная теория вероятностей–3

17.1. Случайная величина

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

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

Формальное определение: *математическим ожиданием* случайной величины $f\colon U \to \mathbb{R}$ называется число

$$\mathbf{E}[f] = \sum_{x \in U} f(x)\,\mathbf{Pr}[x].$$

Пример 17.1. Бросания игральной кости, случайная величина: количество выпавших очков.

Вероятностное пространство состоит из чисел $1, 2, 3, 4, 5, 6$, такой случайной величине отвечает тождественная функция. Распределение равномерное, поэтому математическое ожидание этой величины равно

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

Другими словами, мы *ожидаем*, что среднее арифметическое многих результатов бросания кости будет близко к числу 3.5 . Насколько такие ожидания оправданы — трудный вопрос.

Лемма 17.1 (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*

$$\mathbf{E}[f + g] = \mathbf{E}[f] + \mathbf{E}[g].$$

Доказательство. Запишем определение и перегруппируем слагаемые:

$$\mathbf{E}[f+g] = \sum_{x \in U} (f+g)(x)\,\mathbf{Pr}[x] = \sum_{x \in U} f(x)\,\mathbf{Pr}[x] + \sum_{x \in U} g(x)\,\mathbf{Pr}[x] = \mathbf{E}[f] + \mathbf{E}[g]. \quad \square$$

Линейность математического ожидания во многих случаях заметно упрощает вычисления.

Пример 17.2. Студент решает домашнее задание, в котором 10 задач. Он решает первую задачу, а затем, решив очередную задачу, он с вероятностью $4/5$ переходит к следующей задаче, а с вероятностью $1/5$ бросает решение домашнего задания. Каково математическое ожидание количества решённых задач?

Обозначим количество решённых задач через f. Эта случайная величина разлагается в сумму случайных величин $f_1 + f_2 + \ldots + f_{10}$, где

$$f_k = \begin{cases} 1, \text{ если } k\text{-я задача решена;} \\ 0, \text{ иначе.} \end{cases}$$

Действительно, если в исходе u было решено ровно m задач, то $f_k(u) = 1$ при $k \leqslant m$ и $f_k(u) = 0$, при $k > m$, а потому $\sum f_k(u) = m = f(u)$.

Математическое ожидание случайной величины f_k выражается как

$$\mathbf{E}[f_k] = 1 \times \mathbf{Pr}[A_k] + 0 \times \mathbf{Pr}[\bar{A}_k] = \mathbf{Pr}[A_k],$$

где A_k — событие, состоящее в решении k-й задачи. Найдём математическое ожидание каждой случайной величины f_k. Заметим, что k-я задача решается только после всех предыдущих $k-1$ задачи. Поэтому

$$\mathbf{Pr}[A_k] = \mathbf{Pr}[A_1] \cdot \mathbf{Pr}[A_2|A_1] \cdot \ldots \cdot \mathbf{Pr}[A_k|A_{k-1}] = (4/5)^{k-1}$$

(первую задачу студент решает с вероятностью 1). Теперь, пользуясь линейностью математического ожидание, осталось просуммировать:

$$\mathbf{E}[f] = \mathbf{E}\left[\sum_{k=1}^{10} f_k\right] = \sum_{k=1}^{10} \mathbf{E}[f_k] = \sum_{k=1}^{10} (4/5)^{k-1} = \frac{1-(4/5)^{10}}{1-(4/5)} = 5 - 5\left(\frac{4}{5}\right)^{10}.$$

Пример 17.3 («парадокс дней 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?

Уточним вопрос и упростим его. Предполагаем, что дни рождения у разных людей независимы, а в году 365 дней. Другими словами, вероятностное пространство: всюду определённые функция из n-элементного множества людей $\{x_1, \ldots, x_n\}$ в 365-элементное множество дней в году. Все исходы равновозможные.

Обозначим случайную величину, равную количеству пар людей с совпадающими днями рождения, через f. Нам требуется посчитать математическое ожидание случайной величины f.

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

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

Подсчитаем математическое ожидание случайной величины g_{ij}. Нетрудно видеть, 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 $\mathbf{E}[g_{ij}] = 1/365$ (для всякой пары i, j). Для математического ожидания f из линейности получаем

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

Если число людей n больше 27, то $\mathbf{E}[f] > 1$, то есть стоит ожидать[1)], что будет не меньше одного совпадения дней рождений, что может показаться противоречащим интуиции.

17.2. Обобщение вероятностного метода

С помощью математического ожидания можно обобщить вероятностный метод.

Лемма 17.2 («среднее не больше максимума и не меньше минимума»)**.** *Пусть* $\mathbf{E}[f] = C$ *для какой-то случайной величины* $f\colon U \to \mathbb{R}$. *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 от противного первое утверждение леммы, второе доказывается аналогично. Предположим, что утверждение неверно, а значит для всякого $u \in U$ верно $f(u) < C$. Тогда

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

противоречие. □

Эта лемма обобщает предыдущий вариант вероятностного метода. Если в качестве случайной величины взять индикаторную функцию события A, то $\mathbf{E}[\chi_A] = \mathbf{Pr}[A]$ (мы это уже несколько раз использовали). Поэтому при $\mathbf{Pr}[A] > 0$ существует исход x, для которого $\chi_A(x) = 1$, то есть $x \in A$.

Приведём примеры использования этой леммы.

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

[1)] Если посчитать точно, то при $n = 28$ вероятность того, что будет не меньше одного совпадения дней рождения, равна ≈ 0.65; математическое ожидание числа совпадений ≈ 1.03.

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 $1/2^n$. Нетрудно 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 $1/4$. (Сопоставьте случайным множествам двоичные строки длины $|V|$, тогда интересующие нас события состоят в том, что на позициях x, y записаны конкретные значения.)

Рассмотрим случайный разрез и 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

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

Однако, для случайной величины f_e математическое ожидание легко вычисляется. Для всякого фиксированного ребра e вероятность, что оно попадёт в разрез равна $1/2$. А значит, $\mathbf{E}[f_e] = 1/2$ для всякого $e \in E$, откуда

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

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

Эта оценка не слишком удивительна, её можно доказать и без вероятностей. Нужный разрез находится «локальным поиском», который перебрасывает вершину из одной доли разреза в другую, если при этом величина разреза увеличивается. (Проверьте!)

Вероятностным рассуждением нетрудно получить чуть более сильную оценку на размер разреза.

Теорема 17.4. *Рассмотрим граф $G = (V, E)$ с чётным количеством вершин, $|V| = 2n$. 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 так же, 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

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

Тогда, аналогично предыдущему, получаем

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

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

Такой разрез также можно построить напрямую, но это уже заметно сложнее. А вероятностное рассуждение технически не слишком сложное.

17.3. Неравенство Маркова

Математическое ожидание позволяет давать оценки вероятностей некоторых событий, связанных со значениями случайной величины.

Лемма 17.5 (неравенство Маркова)**.** *Пусть f — случайная величина, принимающая только неотрицательные значения. Тогда для всякого $\alpha > 0$ верно*

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

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

Доказательство. Доказываем равносильное неравенство

$$\mathbf{E}[f] \geqslant \alpha \cdot \mathbf{Pr}[f \geqslant \alpha].$$

Запишем математическое ожидание как сумму и разобьём эту сумму в два слагаемых:

$$\mathbf{E}[f] = \sum_{x \in U} f(x)\,\mathbf{Pr}[x] = \sum_{x : f(x) \geqslant \alpha} f(x)\,\mathbf{Pr}[x] + \sum_{x : f(x) < \alpha} f(x)\,\mathbf{Pr}[x].$$

Заменим в первом слагаемом $f(x)$ на α, от этого сумма может лишь уменьшится. Во втором слагаемом заменим $f(x)$ на 0, от этого сумма также может лишь уменьшится. Получаем

$$\mathbf{E}[f] \geqslant \alpha \cdot \sum_{x : f(x) \geqslant \alpha} \mathbf{Pr}[x] + 0 \cdot \sum_{x : f(x) < \alpha} \mathbf{Pr}[x] = \alpha \cdot \mathbf{Pr}[f(x) \geqslant \alpha],$$

что и требовалось. □

Неравенство Маркова позволяет оценивать вероятность того, что значение случайной величины заметно отличается от её математического ожидания. Для этого часто используется характеристика, называемая *дисперсией*:

$$\mathbf{D}[f] = \mathbf{E}\left[(f - \mathbf{E}[f])^2\right] = \mathbf{E}[f^2] - \mathbf{E}[f]^2.$$

Лемма 17.6 (неравенство Чебышёва)**.** $\mathbf{Pr}\left[\left|f - \mathbf{E}[f]\right| \geqslant \alpha\right] \leqslant \dfrac{\mathbf{D}[f]}{\alpha^2}$.

Доказательство. Событие $\left|f - \mathbf{E}[f]\right| \geqslant \alpha$ совпадает с событием $\big(f - \mathbf{E}[f]\big)^2 \geqslant \alpha^2$. Применение неравенства Маркова к этому второму событию и даёт неравенство Чебышёва. □

Дисперсия очень популярна в теории вероятностей и приложениях. Однако аналогичные неравенства можно написать и для других функций, например, экспоненты. На таком пути можно получать очень сильные оценки вероятности отклонения среднего случайной величины от математического ожидания. Очень важный пример — неравенство Чернова — разобран в черновике нашего учебника.

Лекция 18

Разрешающие деревья

18.1. Задача об угадывании числа

Начнём с анализа хорошо известной игры. Алиса загадывает натуральное число от 1 до N, а Боб пытается это число отгадать. При этом Бобу разрешается задавать вопросы, на которые Алиса может ответить «да» или «нет», и Алиса должна на эти вопросы давать правильные ответы. Цель Боба состоит в том, чтобы задать как можно меньше вопросов. При этом мы не хотим полагаться на удачу, то есть нужно, чтобы число вопросов было гарантировано небольшим. Другими словами, мы хотим найти такое минимальное k, что у Боба есть алгоритм, позволяющий отгадать число за не более чем k вопросов, какое бы число ни загадала Алиса.

Утверждение 18.1. *Бобу достаточно задать не более* $\lceil \log_2 N \rceil$ *вопросов.*

Доказательство. Алгоритм Боба основан на *методе деления пополам*. Боб каждым своим вопросом будет сокращать количество оставшихся возможных чисел примерно в два раза.

Более точно, Боб будет поддерживать такой инвариант: после q вопросов он сможет указать такой отрезок $\{y : a_q \leqslant y \leqslant b_q\}$, которому заведомо принадлежит задуманное Алисой число x, причём длина этого интервала $l_q = b_q - a_q + 1$ примерно $N/2^q$. Чтобы поддерживать такой инвариант, Бобу следующим вопросом нужно спросить, верно ли, что $x < (b_q + a_q)/2$ (сравнить x с серединой отрезка). Если ответ «да», то $a_q \leqslant x \leqslant \lceil (b_q + a_q)/2 - 1 \rceil$, если «нет», то $\lceil (b_q + a_q)/2 \rceil \leqslant x \leqslant b_q$.

Для анализа удобна другая реализация этой идеи. Боб сначала округляет N до степени двойки вверх, то есть находит такое s, что $2^{s-1} < N \leqslant 2^s$. Нетрудно видеть, что $s = \lceil \log_2 N \rceil$.

Далее Боб считает, что Алиса задумала число на отрезке от 1 до 2^s. Это только усложняет ему задачу (как увидим далее, несущественно). Но зато длина отрезка после каждого вопроса уменьшается ровно вдвое. Поэтому длина отрезка после q вопросов равна 2^{s-q}. То есть, после s вопросов длина отрезка станет равной 1 и Боб в состоянии отгадать число. □

Эта оценка точная.

Утверждение 18.2. *Бобу необходимо задать не менее* $\lceil \log_2 N \rceil$ *вопросов.*

Доказательство. Для доказательства нижней оценки мы применим так называемый *мощностной метод*.

Пусть алгоритм Боба задаёт не более k вопросов. Каждое число, задуманное Алисой, порождает последовательность ответов. Длина этой последовательности не больше k. Поэтому возможно не более 2^k вариантов ответов Алисы на вопросы Боба (на каждый вопрос есть два варианта ответа).

Если $2^k < N$, то для двух каких-то чисел $x_1 \neq x_2$ все ответы Алисы на вопросы Боба совпадают. Но это означает, что алгоритм Боба работает некорректно. Если Алиса задумала x_1, то Боб, следуя этому алгоритму, не сможет в конце отгадать число (поскольку x_2 и x_1 неразличимы по ответам на заданные вопросы).

Значит, $2^k \geqslant N$, то есть $k \geqslant \lceil \log_2 N \rceil$. □

18.2. Модель разрешающих деревьев (decision trees)

Теперь опишем более общую ситуацию. Пусть фиксирована некоторая функция $f\colon A \to B$, где A, B — какие-то конечные множества. Требуется вычислить $f(x)$, но x явно не сообщается. Разрешается задавать вопросы вида $x \in S$ для некоторых подмножеств S множества A. Список таких разрешённых вопросов также считаем фиксированным.

Пример с угадыванием числа укладывается в эту постановку: любой вопрос с ответом «да» или «нет» сводится к вопросу о том, принадлежит ли загаданное число некоторому подмножеству, — а именно, подмножеству тех чисел, для которых ответ «да». Семейство разрешённых множеств в примере с угадыванием числа максимальное: любое подмножество разрешённое.

Зафиксируем модель вычислений: *разрешающее дерево* (или протокол вычисления функции). Это двоичное дерево, каждая промежуточная вершина которого (не лист) помечена некоторым разрешённым подмножеством $S \subseteq A$. Каждый лист помечен элементом $b \in B$. Из каждой промежуточной вершины, выходит три ребра: одно к корню и два к листьям. Для каждой промежуточной вершины одно из рёбер, ведущих к листьям, помечено единицей, а другое — нулём.

Вычисление согласно протоколу происходит следующим образом. Строим путь из корня дерева в какой-то из листьев. Элемент множества B, которым помечен лист, будет результатом вычисления.

В начале пути мы находимся в корне дерева. Корень, как и всякая промежуточная вершина, помечен каким-то подмножеством множества A. Если вход x принадлежит этому подмножеству, то мы переходим по ребру, помеченному единицей, иначе — по ребру помеченному нулём. Если следующая вершина — лист, то путь построен. Если же она является промежуточной вершиной, то мы повторяем процедуру: спрашиваем, лежит ли x в подмножестве, которым помечена текущая вершина, и переходим по ребру, помеченному единицей, если x лежит в подмножестве, и по ребру, помеченному нулём, иначе. Мы говорим, что протокол вычисляет функцию f, если для всякого $x \in A$ результат вычисления равен $f(x)$.

Сложностью протокола будем называть глубину дерева: максимальную длину пути из корня в лист. Нетрудно видеть, что она равна числу вопросов, которое потребуется задать в худшем случае.

Сложность вычисления функции равна минимуму сложности протоколов, вычисляющих эту функцию.

18.3. Сортировка

Задача о сортировке неформально описывается так. Дано n объектов, все разного веса. За один шаг разрешается сравнить веса двух объектов (мы узнаем, какой из этих объектов тяжелее). Требуется расположить эти объекты в порядке возрастания веса.

Опишем задачу формально. Будем считать, что объекты расположены в виде последовательности. Обозначим в этой последовательности самый тяжёлый объект единицей, второй по тяжести — двойкой, и так далее, самый лёгкий объект обозначим n. Таким образом, на вход по существу подаётся перестановка n-элементного множества. Чтобы упорядочить объекты по возрастанию нам нужно найти данную перестановку. Таким образом, в этом примере A — множество перестановок n-элементного множества и требуется вычислить тождественную функцию на A, то есть $f(x) = x$ для всякого $x \in A$.

При этом разрешается задавать не любые вопросы, а только вопросы о сравнении двух элементов перестановки. Формально это означает, что в вершинах разрешающего дерева могут стоять не любые подмножества множества перестановок, а только множества $S_{i,j}$ для $i, j = 1, \ldots, n$, состоящие из всех перестановок $(a_1, \ldots, a_n)$, в которых $a_i > a_j$.

Заметим, что если бы не было ограничения на вид множеств, то задача была бы полностью аналогична задаче об угадывании числа: на вход подаётся один из $n!$ объектов и требуется угадать, какой именно. Поскольку у нас добавляется ограничение на тип вопросов, то наша задача усложняется, а значит в задаче о сортировке требуется не меньше вопросов.

Следствие 18.3. *Сложность задачи о сортировке n объектов не меньше* $\lceil \log_2 n! \rceil$.

Верхнюю оценку нужно искать заново: теперь не все вопросы возможны.

Теорема 18.4. *Сложность задачи о сортировке n объектов не больше* $\sum_{k=1}^{n} \lceil \log_2 k \rceil$.

Доказательство. Доказательство будем вести индукцией по n. Для $n = 1$ оценка верна — никаких сравнений не требуется.

Пусть утверждение доказано для n, докажем его для $n + 1$. Сначала возьмём первые n объектов и упорядочим их, пользуясь предположением индукции. После этого у нас остаётся $\lceil \log_2(n + 1) \rceil$ сравнений, и нам нужно один оставшийся объект поместить в уже упорядоченный список из n объектов. То есть, для $(n+1)$-го объекта есть $n+1$ место среди упорядоченного списка из n объектов и нам нужно это место найти. Пользуясь методом деления пополам, решаем эту задачу за $\lceil \log_2(n + 1) \rceil$

сравнений: на каждом шаге уменьшаем размер отрезка возможных мест (примерно) вдвое, сравнивая со средним объектом из отрезка. □

Полученные верхняя и нижняя оценки для сложности задачи сортировки, обозначим её $\mathrm{Sort}(n)$, различаются и обе неточны. Например,

$$\lceil \log_2 5! \rceil = 7 = \mathrm{Sort}(5) < 8, \quad \Big(8 = \lceil \log_2 1 \rceil + \lceil \log_2 2 \rceil + \lceil \log_2 3 \rceil + \lceil \log_2 4 \rceil + \lceil \log_2 5 \rceil\Big)$$

$$\lceil \log_2 12! \rceil = 29 < \mathrm{Sort}(12).$$

(Первое неравенство ещё можно доказать руками, для второго разумнее использовать грамотный компьютерный перебор.)

Однако асимптотически эти оценки не слишком различаются.

Утверждение 18.5. $\displaystyle\sum_{k=1}^{n} \lceil \log_2 k \rceil - \lceil \log_2 n! \rceil \leqslant n = o\big(\lceil \log_2 n! \rceil\big).$

Доказательство. Неравенство получается так:

$$\sum_{k=1}^{n} \lceil \log_2 k \rceil \leqslant \sum_{k=1}^{n} (\log_2 k + 1) \leqslant n + \log_2 n! \leqslant n + \lceil \log_2 n! \rceil.$$

Асимптотика легко получается из оценки факториала $n! > (n/2)^{n/2}$ (ограничимся лишь теми множителями в факториале, которые больше $n/2$). □

Точное значение $\mathrm{Sort}(n)$ до сих пор неизвестно.

18.4. Адаптивные и неадаптивные алгоритмы

В нашей модели следующие вопросы могут зависеть от ответов на предыдущие. Такие вычислительные модели обычно называют *адаптивными*. Можно рассмотреть и неадаптивную постановку той же задачи: в ней вопросы не должны зависеть от ответов на предыдущие вопросы.

Пример 18.1 (Неадаптивное угадывание числа)**.** Боб должен составить на бумаге список вопросов и передать его Алисе. Алиса должна ответить на все эти вопросы и вернуть список Бобу. После этого Боб должен назвать загаданное число.

Неадаптивная модель слабее адаптивной: задача Боба стала только сложнее. Это значит, что в неадаптивной модели Бобу потребуется не меньше вопросов, чем в адаптивной.

Заключаем, что для угадывания числа из отрезка от 1 до N Бобу потребуется не менее $\lceil \log_2 N \rceil$ вопросов.

Оказывается, этого количества вопросов и достаточно: i-м вопросом Боб спрашивает, верно ли, что i-й бит двоичной записи загаданного числа x равен 1. После $k = \lceil \log_2 N \rceil$ вопросов Боб знает всю двоичную запись числа x, а значит знает само загаданное число.

Пример 18.2 (Неадаптивная сортировка)**.** В задаче о сортировке ситуация намного хуже. Нетрудно доказать, что Бобу необходимо не менее $\binom{n}{2} = n(n-1)/2$ вопросов. Этого количества, разумеется, и достаточно: если Боб знает сравнения между всеми объектами, то он в состоянии их упорядочить.

Докажем нижнюю оценку. Модель неадаптивная, так что протокол — это просто список всех вопросов. Раз вопросов в нём меньше $\binom{n}{2}$, значит какие-то два объекта между собой не сравниваются. Рассмотрим два таких входа, при котором эти два объекта тяжелее всех остальных и в первом случае, первый объект тяжелее, а во втором — второй (а все остальные объекты сравниваются друг с другом одинаково в обоих входах). Тогда наш протокол на этих входах получит одни и те же ответы, а значит выдаст один и тот же результат. Поскольку самые тяжёлые объекты в этих двух входах разные, наш протокол на одном из них ошибается, а значит мы пришли к противоречию.

18.5. Взвешивания монет

Это по сути ограниченные варианты сортировки, когда требуется не восстановить всю перестановку, а узнать о ней что-то меньшее.

Пример 18.3 (Поиск самого тяжёлого объекта)**.** Есть n объектов, веса попарно различны. За один ход разрешается сравнить по весу два из них. Требуется найти самый тяжёлый объект. Формальная модель отличается от задачи сортировки функцией, которую нужно вычислить. Теперь по данной перестановке мы хотим найти номер позиции, в которой стоит число 1.

Сложность этой задачи равна в точности $n-1$.

Докажем, что $n-1$ взвешивания достаточно. Доказательство индукцией по n. База $n=1$: ничего взвешивать не нужно, достаточно $n-1=0$ mвзвешиваний.

Шаг индукции. Пусть мы доказали утверждение для $n-1$. Рассмотрим n объектов. Возьмём любые два и сравним их. Заметим, что более лёгкий из них не может быть самым тяжёлым, так что его можно выбросить из рассмотрения. Таким образом у нас остаётся $n-1$ объект и по предположению индукции мы можем найти самый тяжёлый из них за $n-2$ оставшихся взвешивания.

Докажем, что $n-1$ взвешивание необходимо. Пусть мы сделали $\leqslant n-2$ взвешивания. Рассмотрим следующий граф. Его вершинами будут наши объекты, и мы соединяем рёбрами те из них, которые мы сравнили в одном из взвешиваний. Тогда в этом графе n вершин и $\leqslant n-2$ ребра. Значит этот граф не связен. Рассмотрим множество V_1 объектов в одной из его компонент связности и множество V_2 всех остальных объектов. Предположим, для определённости, что самый тяжёлый объект находится в V_1. Увеличим вес всех объектов в V_2 на одно и то же очень большое число, такое чтобы все объекты в V_2 стали тяжелее всех объектов в V_1. При этом результаты всех взвешиваний не изменятся, поскольку все сравнения были либо внутри V_1, либо внутри V_2, а самая тяжёлая монета станет другой (теперь она будет в V_2). Таким образом, все взвешивания дадут один и тот же результат в обеих ситуациях, а самый тяжёлый объект будет разным. Значит в одной из двух ситуаций

наш протокол выдаёт неправильный ответ. Мы пришли к противоречию, а значит для нахождения самого тяжёлого объекта требуется не меньше $n-1$ сравнения.

18.6. Метод противника (adversary method)

Для доказательства верхних оценок обычно предъявляют алгоритм, проверяют его корректность и оценивают его сложность (в нашем случае, наибольшее количество вопросов).

Для доказательства нижних оценок, если действовать прямолинейно, нужно рассматривать все возможные алгоритмы и оценивать сложность каждого из них. Это намного труднее. Однако, есть приёмы, которые сводят доказательство нижних оценок к построению алгоритма (не для самой задачи, а для чего-то, с ней связанного). Для разрешающих деревьев есть очень мощный и удобный приём такого вида, который называется *метод противника* (adversary method).

Представим, что вход для алгоритма вычисления функции с помощью разрешающих деревьев выбирается не произвольно, а есть ещё один участник (противник), который его выбирает. При этом противник может выбирать вход не заранее, а по ходу работы протокола, вычисляющего функцию. То есть, изначально противник вход не фиксирует, а по мере поступления запросов от протокола даёт на них ответы так, чтобы ответы были согласованы хотя бы с одним входом. При этом цель противника в том, чтобы заставить протокол задать как можно больше вопросов.

Если у противника есть стратегия (функция выбора ответов в зависимости от истории вопросов и ответов), которая гарантирует, что после k вопросов и ответов на них есть хотя бы два входа, согласованных со всеми вопросами и ответами, то никакой протокол не может решить задачу за $\leqslant k$ вопросов.

Действительно, пусть стратегия противника работает против протокола, задающего $\leqslant k$ вопросов. После $\leqslant k$ остаются по крайней мере два входа, согласованных со всеми ответами и протокол не в состоянии корректно завершиться.

Во многих случаях стратегия противника очень простая (и поэтому мы до сих пор без неё обходились). Скажем, в задаче о нахождении самого тяжёлого объекта из n объектов (адаптивной) противник выбирает изначально какой-то вход, даёт ответы в соответствии с ним и ждёт пока протокол (задающий не более $n-2$ вопросов) выдаст какой-то ответ, а затем изменяет свой вход, добавив большой вес ко всем объектам в одной из компонент связности (см. пример 18.3). При этом все ответы согласованы и с новым входом тоже.

Рассмотрим более сложный пример.

Пример 18.4 (Поиск самого тяжёлого и самого лёгкого)**.** Есть $n = 2k$ объектов, веса попарно различны. За один ход разрешается сравнить по весу два из них. Требуется найти самый тяжёлый объект и самый лёгкий объект. Формально по данной перестановке нужно найти номера позиций, в которых стоят числа 1 и n.

Для решения этой задачи достаточно $3n/2 - 2$ взвешиваний. Алгоритм такой: разбиваем на пары. Проводим k взвешиваний внутри пар. Самый тяжёлый обязан

оказаться среди тех k, которые оказались тяжелее. Самый лёгкий — среди тех, кто оказались легче, их тоже k штук. Теперь находим самый тяжёлый за $k-1$ взвешивание и самый лёгкий за $k-1$ взвешивание. В итоге

$$3k-2 \leqslant 3n/2-2$$

взвешиваний.

Докажем, что необходимо $3n/2-2$ взвешиваний. Опишем стратегию противника. Он поддерживает такую структуру. В множестве L те объекты, которые при всех взвешиваниях оказывались легче и хотя бы раз участвовали во взвешиваниях. Это кандидаты на самого лёгкого. В множестве H те объекты, которые при всех взвешиваниях оказывались тяжелее и хотя бы раз участвовали во взвешиваниях. Это кандидаты на самого тяжёлого.

Стратегия состоит в том, что при очередном взвешивании говорить, что x тяжелее y, если $x \in H$ и $y \notin H$ или $x \notin L$ и $y \in L$. В остальных случаях ответ произвольный, но согласован с предыдущими ответами. Такая стратегия гарантирует, что из множества L удаляется объект лишь в том случае, когда взвешиваются два объекта из L. Аналогично для множества H.

Пусть противник работает против корректного протокола. В конце множества L и H содержат ровно по одному элементу. Отметим для каждого объекта первое взвешивание, в котором он участвовал. Среди этих взвешиваний по крайней мере $n/2$ различных (так как во взвешивании участвуют два объекта). Заметим, что после первого взвешивания объект попадает либо в L, либо в H. Для всех объектов, кроме самого лёгкого и самого тяжёлого, отметим то взвешивание, после которого он выбывает из кандидатов. Все эти $n-2$ взвешиваний различны и отличаются от взвешиваний первого рода. Как уже сказано, удаление из L возможно лишь при взвешивании двух объектов из L. Но при таком взвешивании удаляется лишь один объект. Аналогично рассуждаем для H.

Итак, мы указали не менее $n/2+n-2 = 3n/2-2$ различных взвешиваний, что и требовалось.

Лекция 19

Булевы функции–2

Булева функция — это функция вида $f\colon \{0,1\}^N \to \{0,1\}$. Ближайшие несколько занятий мы будем изучать разные способы вычисления булевых функций.

19.1. Сложность разрешающих деревьев

Определим модель разрешающих деревьев для вычисления булевых функций. Разрешённые множества имеют вид $S_i = \{w \in \{0,1\}^n : w_i = 1\}$. Другими словами, разрешается спрашивать значение переменной.

Сложность булевой функции f в модели разрешающих деревьев обозначается через $D(f)$.

Если известны все значения переменных, значение функции однозначно определено. Поэтому сложность любой булевой функции в модели разрешающих деревьев не превосходит количества переменных N.

Иногда эта сложность меньше. Например, если у функции есть *несущественные* переменные. Переменная x_i называется несущественной, если равенство

$$f(a_1, \ldots, a_{i-1}, 0, a_{i+1}, \ldots, a_N) = f(a_1, \ldots, a_{i-1}, 1, a_{i+1}, \ldots, a_N)$$

выполняется для любых $a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N \in \{0,1\}$.

Запрашивать значения несущественных переменных не имеет смысла: от них значение функции не зависит. Поэтому $D(f)$ не больше числа существенных переменных.

Однако бывают функции, у которых все переменные существенные, а сложность $D(f)$ очень маленькая, порядка $\log_2 N$.

Для многих функций эта сложность в точности равна N. Рассмотрим интересный пример.

Пример 19.1 (Проверка связности графа)**.** Дан неориентированный граф G на вершинах $\{1, \ldots, n\}$. За один ход разрешается спрашивать наличие или отсутствие конкретного ребра. Нужно проверить, является ли граф связным, то есть выдать 1, если является, и 0 иначе.

Эта задача — частный случай модели разрешающих деревьев для булевых функций. Действительно, можно считать, что рассматривается функция от множества переменных $\{x_{ij} : 1 \leqslant i < j \leqslant n\}$, где $x_{ij} = 1$ тогда и только тогда, когда между вершинами i и j есть ребро в графе на множестве вершин $\{1, 2 \ldots, n\}$. Обозначим через CONN функцию, которая равна 1, если заданный значениями переменных граф связен, и 0 в противном случае. Всего переменных у функции $\binom{n}{2} = n(n-1)/2$, поэтому $D(\text{CONN}) \leqslant \binom{n}{2}$.

Докажем, что $D(\text{CONN}) = \binom{n}{2}$ методом противника. Стратегия противника поддерживает два графа MAX и MIN на том же самом множестве вершин $\{1, \ldots, n\}$. В MIN включаем рёбра, про которые уже был дан ответ «да» (то есть, рёбра, которые необходимо присутствуют в графе, поданном на вход), а в MAX — те рёбра, про которые не было ещё сказано «нет». Такие рёбра могут входить во входной граф. Другими словами, MIN — это минимальный возможный граф, согласованный с уже данными ответами, а MAX — максимальный.

Противник придерживается такой стратегии, при которой выполняется инвариант: MAX всегда связен, а если MIN $\neq$ MAX, то MIN не связен.

Выполнение такого инварианта означает, что корректному протоколу придётся задать вопросы про все рёбра. Если MIN = MAX, то про каждое ребро был задан вопрос (с ответом «да» для рёбер из MIN и с ответом «нет» для остальных рёбер). А если MIN $\neq$ MAX, то корректный протокол ещё не может закончиться: с текущими ответами согласован как некоторый связный граф, так и некоторый несвязный.

Стратегия противника: если спрашивают про ребро e, то он отвечает «нет», если граф MAX остаётся связным после удаления e, иначе отвечает «да».

При таком выборе ответов MAX всегда связен. Докажем, что если MIN $\neq$ MAX, то MIN не связен.

Заметим, что если в MAX есть цикл, то ни одно его ребро не принадлежит MIN. Действительно, пусть это не так. Рассмотрим ребро цикла, которое попало в MIN первым. Это означает, что противник ответил на вопрос об этом ребре положительно, то есть граф MAX при удалении этого ребра переставал быть связным. Но этого не может быть, потому что в любом пути это ребро можно заменить обходом по циклу. Противоречие.

В частности, из этого получается, что в MIN нет циклов: иначе такой цикл лежал бы и в MAX и все его ребра были бы при этом в MIN.

Если бы MIN был связен, то он был бы остовным деревом для MAX. При этом MAX $\neq$ MIN, то есть в MAX есть ребро, которого нет в MIN. Тогда это ребро вместе с графом MIN содержит цикл, а значит мы нашли цикл в MAX, часть рёбер которого (все, кроме одного) лежат в MIN. Противоречие. Получается, что MIN не связен.

19.2. Формулы для булевых функций. Полные системы функций

Булева функция $f\colon \{0,1\}^n \to \{0,1\}$ имеет конечную область определения. Поэтому задать булеву функцию можно таблицей значений. Мы уже использовали такое представление булевых функций, когда рассматривали булевы связки: конъюнкцию,

дизъюнкцию и т.п. Однако такое представление становится неудобным для большого числа аргументов: для булевой функции от n переменных есть 2^n разных наборов аргументов и это **очень большое** число уже при $n = 100$.

Есть более компактные способы представления булевых функций. Например, как обычно в математике, функцию можно задавать *формулой*:

$$f(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_2 \wedge x_3).$$

Рассмотрим такой вопрос: какие функции необходимы и достаточны для задания *произвольной* булевой функции в виде формулы?

Уточним вопрос. Выберем некоторый набор булевых функций $B = \{f_1, f_2, \dots\}$ (как обычно говорят, *систему функций*), не обязательно конечный. Будем считать, что имена этих функций можно использовать в формулах. В качестве аргументов этих функций можно использовать имена переменных или другие формулы. Получаем *формулы в системе* B. Каждая формула задаёт некоторую булеву функцию. Функцию f будем называть *выразимой* в системе B, если существует формула в системе B, которая задаёт функцию f.

Определение 19.1. Система B называется *полной*, если любая булева функция выразима в этой системе.

Теперь уточнение вопроса формулируется так: какие системы функций полные, а какие — нет? Мы дадим исчерпывающий ответ на этот вопрос.

Начнём с примеров полных систем. Из булевых связок легко собрать полную систему.

Теорема 19.2. *Система* $\{\wedge, \vee, \neg\}$ *полная.*

Доказательство. Сначала научимся выражать через конъюнкции и отрицания особые функции, которые равны 1 ровно для одного набора аргументов $(a_1, a_2, \dots, a_n)$, а на остальных равны 0. Примером такой функции является конъюнкция переменных: $x_1 \wedge \dots \wedge x_n$. Она равна 1 только лишь на наборе $(1, 1, \dots, 1)$.

Чтобы получить функцию e_a, которая равна 1 лишь на наборе аргументов $a = (a_1, a_2, \dots, a_n)$, нужно в конъюнкции заменить часть переменных на отрицания. А именно, если $a_i = 1$, то включаем в конъюнкцию переменную x_i; если $a_i = 0$, то включаем в конъюнкцию отрицание переменной $\neg x_i$. Это гарантирует, что на наборе $(a_1, a_2, \dots, a_n)$ такая формула принимает значение 1. Для любого другого набора $(b_1, b_2, \dots, b_n)$ при некотором i выполняется неравенство $b_i \neq a_i$. Но тогда соответствующий член конъюнкции обращается в 0 и вся конъюнкция также равна 0.

Теперь выразим произвольную булеву функцию. Для этого возьмём дизъюнкцию формул для функций e_a по всем a, для которых $f(a) = 1$. Такая формула даёт значение 1 если и только если один из членов дизъюнкции равен 1. Но каждый член дизъюнкции e_a равен 1 ровно на одном наборе a. Поэтому построенная формула выражает f. □

В доказательстве теоремы 19.2 мы выразили произвольную булеву функцию формулой очень специального вида: дизъюнкция конъюнкций переменных или их отрицаний. Такие формулы называют *дизъюнктивными нормальными формами*, сокращённо, ДНФ. Переменные или отрицания называются *литералами*, конъюнкции литералов называются *конъюнктами*. ДНФ — это дизъюнкция конъюнктов.

Построенная в доказательстве теоремы 19.2 ДНФ называется *совершенной*. Эта формула во многих случаях очень длинная: если у функции много единиц, то представление в виде совершенной ДНФ не слишком лучше представления в виде таблицы.

Представление функции в виде ДНФ не единственно.

Пример 19.2. Дизъюнкция переменных $f(x) = x_1 \vee x_2 \vee \cdots \vee x_n$ по определению является ДНФ. Однако совершенная ДНФ для этой функции совсем другая. Поскольку $f(x) = 1$ для $2^n - 1$ наборов аргументов (исключением является набор из одних нулей), совершенная ДНФ для f является дизъюнкцией $2^n - 1$ членов, каждый из которых является конъюнкцией n литералов.

Задача нахождения самой короткой ДНФ (как по числу дизъюнктов, так и по общей длине формулы), представляющей данную функцию, в общем случае очень трудна.

Имея одну полную систему B_1 легко строить примеры других полных систем. Если любая функция из полной системы B_1 выражается в формулой в системе B_2, то система B_2 — полная (подставим в формулу в системе B_1, выражающую функцию f, формулы в системе B_2 для функций из системы B_1).

Утверждение 19.3. *Системы* $\{\vee, \neg\}$, $\{\wedge, \neg\}$, $\{1, \oplus, \wedge\}$ *являются полными.*

Доказательство. Полнота первых двух систем следует из тождеств

$$x \wedge y = \neg(\neg x \vee \neg y); \quad x \vee y = \neg(\neg x \wedge \neg y),$$

то есть законов де Моргана.

Для полноты третьей системы (*базис Жегалкина*) нужно тождество, выражающее $\neg$ через 1, $\oplus$:

$$\neg x = 1 \oplus x,$$

это тождество очевидно из определения $\oplus$. Получаем представления обеих функций из полной системы $\{\wedge, \neg\}$ в базисе Жегалкина. □

Базис Жегалкина особенно интересен, потому что конъюнкция аналогична произведению чисел (и мы будем сейчас её записывать как произведение для краткости), а сумма по модулю 2 — обычной сумме. Эта аналогия будет подробно изучена в курсе алгебры.

Для обычных целых алгебраических выражений известно, что любое такое выражение тождественно равно *многочлену*, то есть формуле, которая является суммой произведений переменных и чисел. Аналогичное утверждение выполняется и для

булевых функций. В этом случае коэффициенты равны 0, 1, что упрощает вид формулы.

Многочленом Жегалкина называются сумма по модулю 2 конъюнкций переменных (одночленов). Считаем, что 1 также одночлен: произведение пустого множества переменных. Сумму пустого количества слагаемых полагаем равной 0.

Теорема 19.4. *Каждая булева функция однозначно представляется в виде многочлена Жегалкина.*

Доказательство. **Существование.** Индукция по числу переменных n. База индукции — $n = 0$, то есть константы. По нашим соглашениям они выражаются в виде многочлена Жегалкина.

Шаг индукции. По функции $f(x_1, \ldots, x_{n+1})$ от $n+1$ переменной определим две функции от n переменных: а именно,

$$f_0(x_1, \ldots, x_n) = f(x_1, \ldots, x_n, 0) \quad \text{и} \quad f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n, 1).$$

Тогда

$$f = ((1 \oplus x_{n+1})f_0) \oplus (x_{n+1}f_1) = f_0 \oplus x_{n+1}(f_0 \oplus f_1).$$

Действительно, при $x_{n+1} = 0$ получаем $f = f_0$, а при $x_{n+1} = 1$ — $f = f_1$.

По предположению индукции и f_0, и f_1 выражаются как многочлены Жегалкина. Тогда и $f_0 \oplus f_1$ выражается как многочлен Жегалкина (возьмём симметрическую разность множеств одночленов для f_0 и для f_1).

Раскрывая скобки во втором слагаемом, получим многочлен Жегалкина для f.

Единственность. Сколько всего есть булевых функций от n переменных? Такая функция взаимно однозначно задаётся таблицей значений. Но таблицы значений находятся во взаимно однозначном соответствии с двоичными словами длины 2^n: запишем все 2^n различных наборов значений n булевых переменных в каком-то порядке, после этого таблица значений задаётся двоичным словом длины[1)] 2^n (на i-м месте стоит $f(x^{(i)})$, где $x^{(i)}$ — i-й набор значений переменных). В качестве иллюстрации запишем слово для конъюнкции $x_1 \wedge x_2$, предполагая, что наборы значений переменных упорядочены лексикографически:

00	01	10	11
0	0	0	1

.

Из этой таблицы видно, что конъюнкции соответствует слово 0001.

Двоичных слов длины 2^n ровно 2^{2^n}, поэтому и булевых функций столько же.

А сколько многочленов Жегалкина? В каждом одночлене все переменные разные (это не ограничивает общности, так как $x \wedge x \equiv x$). Поэтому различных одночленов столько же, сколько подмножеств множества n переменных, то есть 2^n. Каждому многочлену однозначно сопоставляется множество одночленов, суммой которых он

[1)] Спасибо Булату Миннахметову за замеченную опечатку.

является (так как $x \oplus x \equiv 0$, это не ограничивает общности). Поэтому всего многочленов Жегалкина столько же, сколько подмножеств 2^n-элементного множества, то есть, тоже 2^{2^n}.

Но мы уже знаем, что соответствие «многочлен f» $\to$ «функция, задаваемая многочленом f» является сюръективной функцией (существование прообраза уже доказано). Осталось заметить, что сюръекция между двумя конечными множествами одинакового размера является также и инъекцией (если два различных элемента из множества A имеют одинаковый образ в множестве B, то в образе $f(A)$ меньше элементов, чем в A), а потому биекцией. Это и означает единственность. □

19.3. Неполные системы. Теорема Поста

Есть важное препятствие к полноте системе функций, не совпадающей со всем множеством булевых функций: *замкнутость*.

Определение 19.5. Система функций F называется *замкнутой*, если любая формула в этой системе выражает функцию из F.

Из определения ясно, что замкнутая система полна лишь в том случае, когда совпадает с множеством всех функций.

Доказательства замкнутости системы основаны на следующей лемме.

Лемма 19.6. *Пусть система F содержит функции-переменные и для любых $f, g_1, g_2, \ldots, g_n \in F$ выполняется*[2)]

$$f(g_1(x^{(1)}), g_2(x^{(2)}), \ldots, g_n(x^{(n)}) \in F,$$

здесь $x^{(i)}$ — какие-то множества переменных.

Тогда F замкнутая.

Доказательство. Докажем, что любая формула в системе F задаёт функцию из F.

Индукция по разбору формулы. Технически это доказательство по длине формулы. Базисом индукции будут формулы x_i, они принадлежат F по условию.

Шаг индукции. Рассмотрим формулу вида

$$\Phi = f(g_1(x^{(1)}), g_2(x^{(2)}), \ldots, g_n(x^{(n)})),$$

где $f \in F$, а g_i — какие-то формулы в системе F. Для формул g_i выполняется индуктивное предположение, то есть $g_i \in F$. Но тогда по условию Φ задаёт функцию из F. □

Приведём примеры замкнутых систем булевых функций (или классов, как ещё говорят). Все они неполные, так как не содержат всех булевых функций.

[2)] Спасибо Артёму Максаеву за замеченную опечатку в выключной формуле.

Функции, сохраняющие 1, класс T_1. Это функции, удовлетворяющие равенству

$$f(1,1,\ldots,1,1)=1.$$

Функции, сохраняющие 0, класс T_0. Это функции, удовлетворяющие равенству

$$f(0,0,\ldots,0,0)=0.$$

Монотонные функции, класс M. Это такие функции, что из $x_i \leqslant y_i$ для всех $1 \leqslant i \leqslant n$ следует $f(x_1,\ldots,x_n) \leqslant f(y_1,\ldots,y_n)$.

Линейные функции, класс L. Это функции вида

$$x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_k} \oplus a,$$

здесь x_{i_α} — переменные, $a \in \{0,1\}$.

Самодвойственные функции, класс S. Это функции удовлетворяющие равенству

$$f(\neg x_1, \neg x_2, \ldots, \neg x_n) = \neg f(x_1, x_2, \ldots, x_n)$$

для всех наборов аргументов $x_1,\ldots,x_n$.

Оказывается, перечисленные выше классы и есть в точности препятствия к полноте системы функций.

Теорема 19.7 (теорема Поста)**.** *Система B является полной тогда и только тогда, когда она не лежит целиком ни в одном из классов T_0, T_1, M, L, S.*[3)]

(Более формально, выполняются пять условий $B\setminus T_0 \neq \varnothing$, $B\setminus T_1 \neq \varnothing$, $B\setminus M \neq \varnothing$, $B \setminus L \neq \varnothing$, $B \setminus S \neq \varnothing$.)

[3)] Спасибо Рите Аруновой за замечание по поводу формулировки теоремы.

Лекция 20

Булевы функции–3. Теорема Поста. Схемы

20.1. Доказательство теоремы Поста

Пусть F — система функций, которая не лежит целиком в одном из классов T_0, T_1, M, L, S. Это значит, что $\{t_0, t_1, m, \ell, s\} \subseteq F$, причём $t_0 \notin T_0$; $t_1 \notin T_1$; $m \notin M$; $\ell \notin L$; $s \notin S$. (Среди указанных функций могут быть одинаковые.)

Докажем, что формулами в системе $\{t_0, t_1, m, \ell, s\}$ выражаются отрицание и конъюнкция, которые образуют полную систему. Поэтому и система F полная.

Первый шаг состоит в том, чтобы выразить константы 0 и 1, а также отрицание.

Рассмотрим формулу $t_1(x, x, \ldots, x)$ (все аргументы равны одной и той же переменной), задающую функцию f_1 от одной переменной. Так как t_1 не сохраняет 1, получаем $f_1(1) = 0$. Если $f_1(0) = 0$, мы выразили константу 0 и будем дальше её использовать. Если $f_1(0) = 1$, то $f_1(x) = \neg x$ и мы выразили отрицание.

Аналогичное рассуждение с функцией t_0 показывает, что либо мы выразили константу 1, либо выразили отрицание.

Теперь нужно рассмотреть два случая.

Выразили обе константы 0 и 1. Чтобы выразить отрицание, используем немонотонную функцию m. Для неё есть такие наборы x, y, что $x \leqslant y$ в покоординатном порядке и $m(x) = 1$, $m(y) = 0$.

Докажем, что можно найти наборы с таким свойством, различающиеся ровно в одной позиции. Пусть x и y различаются в k позициях. Последовательно заменяя в этих позициях 0 на 1, получаем такую последовательность $x = x_0, x_1, \ldots x_k = y$, в которой соседние наборы различаются ровно в одной позиции, причём $m(x_0) = 1$, $m(x_k) = 0$, так что для какого-то i должно выполняться $m(x_i) = 1$, $m(x_{i+1}) = 0$.

Обозначим через j позицию, в которой различаются x_i и x_{i+1}. Рассмотрим формулу $m(b_1, b_2, \ldots, x, \ldots, b_n)$, в которой на j-й позиции стоит переменная x, а на

остальных позициях стоят формулы, выражающие константы 0 и 1 (в соответствии со значениями в наборах x_i, x_{i+1}). Эта формула задаёт $\neg x$.

Выразили отрицание. Чтобы выразить константу, используем несамодвойственную функцию s. Для некоторого набора аргументов $a_1, \ldots, a_n$ выполняется равенство

$$s(a_1, \ldots, a_n) = s(\neg a_1, \ldots, \neg a_n).$$

Рассмотрим формулу $s(\varphi_1, \ldots, \varphi_n)$, в которой $\varphi_i = x$, если $a_i = 0$ и φ_i задаёт $\neg x$, если $a_i = 1$. Эта формула выражает такую функцию $f_2(x)$, что

$$f_2(0) = s(a_1, \ldots, a_n) = s(\neg a_1, \ldots, \neg a_n) = f_2(1),$$

то есть константу. Имея одну константу и отрицание, легко выразить вторую константу.

Второй шаг состоит в том, чтобы выразить конъюнкцию, используя контанты, отрицание и нелинейную функцию ℓ. В многочлене Жегалкина для функции ℓ есть одночлен степени не меньше 2. Подставим вместо части переменных константы так, чтобы получился многочлен Жегалкина степени ровно 2. Для этого выберем минимальный по включению одночлен степени $\geqslant 2$ (нет одночлена степени $\geqslant 2$, в который входят только переменные, входящие в выбранный одночлен). Оставим в нём две какие-то переменные, скажем, x_1, x_2, а вместо остальных переменных этого одночлена подставим 1 (точнее, формулу, выражающую 1). Вместо переменных, не входящих в этот многочлен, подставим 0. Докажем, что полученная формула выражает многочлен степени 2 от переменных x_1, x_2. Действительно, все одночлены, которые содержат хотя бы одну переменную, не входящую в выбранный одночлен, при такой подстановке обращаются в 0. А в силу выбора одночлена, остальные одночлены имеют степень < 2.

Осталось выразить конъюнкцию через константы, отрицание и многочлен

$$f_3(x_1, x_2) = x_1x_2 \oplus ax_1 \oplus bx_2 \oplus c, \qquad a, b, c \in \{0, 1\},$$

построенный на предыдущем шаге. Перепишем этот многочлен как

$$f_3(x_1, x_2) = (x_1 \oplus b)(x_2 \oplus a) \oplus d, \quad d = ab \oplus c.$$

Заметим, что $x \oplus 0 = x$, $x \oplus 1 = \neg x$. Поэтому подставляя вместо x_1 (x_2) формулу, задающую $\neg x_1$ ($\neg x_2$), если $a = 1$ ($b = 1$), получаем формулу, которая выражает либо конъюнкцию (если $d = 0$), либо отрицание конъюнкции (если $d = 1$). Во втором случае подставим эту формулу вместо переменной в формулу, выражающую отрицание. В любом случае получаем формулу, выражающую конъюнкцию.

Замечание 20.1. Заметим одну тонкость в этом доказательстве. Мы не можем таким образом выразить функции от нуля переменных (то есть константы), а лишь функции от хотя бы одной переменной (см. первый шаг доказательства).

Формальных выходов из этой трудности два. Первый: считать, что у булевой функции положительное количество переменных. Тогда при построении базиса Жегалкина нужно оговорить, что 1 — это функция от одной переменной, тождественно равная 1.

Второй выход: считать константы полноценными функциями (от 0 переменных), но разрешить более общее представление функции формулой, допуская несущественные переменные в формуле.

Первый вариант представляется предпочтительнее.

20.2. Схемы

Есть ещё более компактный способ задавать булевы функции: *схемы*. Схемой мы будем называть программу простейшего вида, состоящую из последовательности присваиваний.

Определение 20.1. *Булевой схемой* от переменных $x_1, \dots, x_n$ мы будем называть последовательность булевых функций $g_1, \dots, g_s$, в которой всякая g_i получается из предыдущих функций последовательности и переменных применением отрицания, конъюнкции или дизъюнкции. Другими словами, для всякого i имеет место одно из равенств

$$\begin{array}{llll} g_i = g_j \wedge g_k & (j,k<i), & g_i = g_j \vee g_k & (j,k<i), \\ g_i = g_j \wedge x_k & (j<i), & g_i = g_j \vee x_k & (j<i), \\ g_i = x_j \wedge x_k, & & g_i = x_j \vee x_k, & \\ g_i = \neg g_j & (j<i), & g_i = \neg x_k. & \end{array}$$

Имея в виду эти связи между элементами последовательности (схемы), будем также называть элементы схемы *присваиваниями*.

Для булевой схемы также задано число $m \geqslant 1$ и члены последовательности $g_{s-m+1}, \dots, g_s$ называются *выходами схемы* (их как раз m). Число m называют числом выходов схемы. Схема вычисляет булево отображение $f\colon \{0,1\}^n \to \{0,1\}^m$, если $f(x) = (g_{s-m+1}(x), \dots, g_s(x))$ для всякого $x \in \{0,1\}^n$. *Размером схемы* называют число s, то есть количество присваиваний в схеме.

Если $m = 1$, то схема как раз задаёт булеву функцию.

Пример 20.1 (схема для $\oplus$)**.** Схема с одним выходом

$$\neg x_1,\ \neg x_2,\ x_1 \wedge \neg x_2,\ \neg x_1 \wedge x_2,\ (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2) \tag{20.1}$$

размера 5 вычисляет функцию $x_1 \oplus x_2$. Значение последней функции равно 1, если аргументы различны, в противном случае равно 0.

Схемы часто представляют в виде графов. Схема из примера 20.1 представляется графом на рисунке 20.1.

Вершины графа схемы соответствуют функциям и переменным схемы. Вершина v_i соответствует функции g_i. Вершина x_j соответствует переменной x_j. Всякая

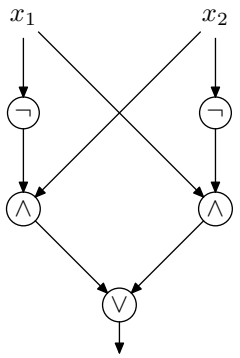


Рис. 20.1: Схема для функции $x_1 \oplus x_2$

вершина v_i помечена логической связкой, с помощью которой функция g_i получена из предыдущих. При этом, если функция g_i была получена из функций g_j и g_k, мы проводим ребра из вершин v_j и v_k в вершину v_i. Если функция g_i получена как отрицание g_j, то мы проводим только ребро из v_j в v_i. Вершины $v_{s-m+1}, \ldots, v_s$ помечены как выходные вершины схемы (из них проведены выходящие стрелки).

Пример 20.2 (схема для функции MAJ_3)**.** Схема с одним выходом

$$x_1 \vee x_2,\ (x_1 \vee x_2) \wedge x_3,\ (x_1 \wedge x_2),\ ((x_1 \vee x_2) \wedge x_3) \vee (x_1 \wedge x_2) \tag{20.2}$$

вычисляет функцию $\mathrm{MAJ}_3(x_1, x_2, x_3)$, которая равна 1 тогда и только тогда, когда больше половины её аргументов равны 1. Размер этой схемы равен 4, она изображена на рисунке 20.2.

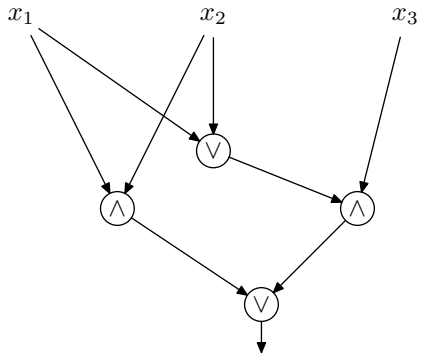


Рис. 20.2: Схема для функции MAJ_3

Пример 20.3 (схема для конъюнкции нескольких переменных)**.** Схема

$$g_1 = x_1 \wedge x_2,\ g_2 = g_1 \wedge x_3, \ldots, g_{n-1} = g_{n-2} \wedge x_n$$

вычисляет конъюнкцию переменных $x_1, \ldots, x_n$. Размер схемы $n-1$. Аналогичную схему размера $n-1$ можно построить и для дизъюнкции n переменных.

Если уже построена схема для функции f, её выход можно использовать в дальнейших присваиваниях. Это позволяет обобщить определение схемы, допуская произвольные булевы функции в присваиваниях. Нужно только правильно считать размер схемы: учитывать, что функции, отличные от базисных, вычисляются схемами некоторого размера.

Пример 20.4 (схема линейного размера для сложения целых чисел)**.** Входами схемы являются n-битовые двоичные записи чисел x и y; $x = x_{n-1}\ldots x_1x_0$, где x_0 — младший разряд двоичной записи (аналогично, $y = y_{n-1}\ldots y_1y_0$). Нужно вычислить двоичную запись их суммы $z = x+y$. В двоичной записи z будет не более $n+1$ разрядов. Так что нужно построить схему с $2n$ входами и $n+1$ выходом.

Схема будет реализовывать обычный школьный алгоритм сложения в столбик. Складываем числа x и y поразрядно, попутно вычисляя биты переноса в следующий разряд.

Для удобства будем обозначать через b_i бит, который переносится в i-й разряд из предыдущих.

Младший разряд ответа легко вычисляется: $z_0 = x_0 \oplus y_0$. Здесь нужно подставить схему из примера 20.1. Бит переноса равен $b_1 = x_0 \wedge y_0$, добавим соответствующее присваивание в схему. Перейдём к следующему разряду. Здесь $z_1 = x_1 \oplus y_1 \oplus b_1$ и $b_2 = \mathrm{MAJ}_3(x_1, y_1, b_1)$. Для вычисления первого добавим сначала подсхему, вычисляющую промежуточную величину $c_1 = x_1 \oplus y_1$, а затем подсхему, вычисляющую $z_1 = c_1 \oplus b_1$. Для вычисления b_2 просто добавим подсхему, вычисляющую функцию MAJ_3. Такая схема также приведена выше. Дальше, случай произвольных z_i и b_i полностью аналогичен случаю z_1 и b_1 и мы можем последовательно вычислить все эти значения.

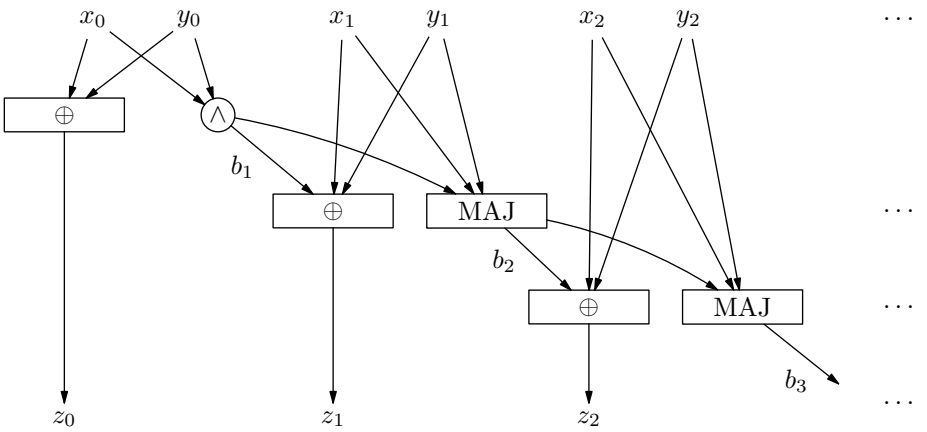


Рис. 20.3: Схема для сложения положительных целых чисел

На рис. 20.3 изображена получившаяся схема, точнее, её начальная часть. На этом рисунке подсхемы изображаются в виде блоков с входящими и выходящими стрелками. Другими словами, мы построили схему, используя более широкий набор присваиваний, а затем реализовали схемой со стандартными связками каждую из более сложных функций (в нашем случае это $x \oplus y$, $x \oplus y \oplus z$, $\mathrm{MAJ}_3(x, y, z)$).

Оценим теперь размер построенной схемы. Для каждого разряда ответа нам нужно не больше двух раз применить подсхему для вычисления функции $\oplus$ и не более одного раза подсхему для вычисления MAJ_3. Все эти схемы имеют фиксированный размер, так что для вычисления каждого разряда z мы используем фиксированное число элементов, не зависящее от числа входных переменных. Поэтому всего в схеме $O(n)$ элементов.

Лекция 21

Булевы функции–4. Схемная сложность

21.1. Примеры схем

Пример 21.1 (схема квадратичного размера для умножения целых чисел)**.** На вход подаются два числа $x = x_{n-1} \dots x_1 x_0$ и $y = y_{n-1} \dots y_1 y_0$. Нужно вычислить $z = x \cdot y$. Заметим, что z имеет не больше $2n$ разрядов. Действительно, $x, y < 2^n$, так что $z = x \cdot y < 2^{2n}$, а значит для его записи достаточно $2n$ разрядов.

Для вычисления произведения, как и в случае суммы, воспользуемся школьным методом. В нем умножение двух чисел сводится к сложению n чисел. Действительно, чтобы умножить x на y достаточно для всякого $i = 0, \dots, n-1$ умножить x на y_i, приписать в конце числа i нулей и затем сложить все полученные числа.

Умножение x на y_i легко реализуется с помощью n конъюнкций. Чтобы приписывать нули, их нужно иметь. В нашем определении не разрешается использовать константы. Поэтому нуль нужно вычислить. Для этого годится, например, такая схема

$$\neg x_1,\ x_1 \wedge \neg x_1 = 0. \tag{21.1}$$

После этого остаётся сложить n чисел длины не более $2n$. Для этого мы можем $n-1$ раз применить схему для сложения, описанную выше. Размер каждой схемы для сложения линейный, так что размер схемы для умножения получается $O(n^2)$.

Имея квадратичную схему для умножения чисел, легко построить схему полиномиального размера для перемножения целочисленных матриц. По определению, $A = BC$ означает равенства для матричных элементов:

$$A_{ij} = \sum_k B_{ik} C_{kj}.$$

Вычисление произведения матриц по этой формуле требует $O(n^3)$ арифметических операций с числами, где n — порядок матриц.[1)]

[1)] Матрицы можно перемножать быстрее. Текущий рекорд $O(n^{2.37\dots})$.

Для комбинаторных приложений важным оказывается *булево произведение* $(0,1)$-матриц (булевых матриц), когда сложение заменяется конъюнкцией, а умножение — дизъюнкцией. Булево произведение $A = B \wedge C$ булевых матриц задаётся формулой

$$A_{ij} = \bigvee_k B_{ik} \wedge C_{kj},$$

которая аналогична формуле для умножения числовых матриц. Это по сути композиция бинарных отношений, заданных матрицами B, C. Поэтому, в частности, булево произведение ассциативно. Ясно, что булево произведение также вычисляется схемой размера $O(n^3)$, где n — порядок матриц.

Пример 21.2 (проверка связности графа)**.** Булево произведение матриц можно использовать для построения схемы вычисления функции CONN, для которой мы раньше нашли сложность в модели разрешающих деревьев.

Напомним, что аргументы этой функции $\{x_{ij} : 1 \leqslant i < j \leqslant n\}$ задают неориентированный граф G на n вершинах $1, 2, \ldots, n$ по правилу: в графе есть ребро $\{i, j\}$ тогда и только тогда, когда $x_{ij} = 1$.

Сейчас удобнее считать, что мы имеем дело с немного моифицированной матрицей смежности графа X, которая симметрическая $x_{ij} = x_{ji}$ для всех i, j, а на диагонали стоят 1. Так что переменные схемы, в которой используются элементы матрицы X легко выражаются через переменные функции CONN. Константу 1 представляем как результат вычисления схемы $\neg x$, $x \vee \neg x$.

Модифицированная матрица смежности обладает таким свойством: $x_{ij} = 1$ тогда и только тогда, когда в графе есть путь длины не больше 1 между вершинами i и j. (Путь длины 1 существует только между смежными вершинами, а из i в i всегда есть путь длины 0.)

По индукции легко проверить, что булева степень X^d этой матрицы обладает аналогичным свойством: $(X^d)_{ij} = 1$ тогда и только тогда, когда в графе есть путь длины не больше d между вершинами i и j. Базу мы уже проверили. Шаг индукции: пусть свойство выполняется для X^d. Рассмотрим $X^{d+1} = X^d \wedge X$. Из определения булева произведения матриц получаем

$$(X^{d+1})_{ij} = \bigvee_k (X^d)_{ik} \wedge X_{kj}.$$

Если $(X^{d+1})_{ij} = 1$, то для какого-то k выполняются равенства $(X^d)_{ik}$ и $X_{kj} = 1$. По предположению индукции это означает, что в графе есть путь из i в k длины $\leqslant d$, и есть ребро $\{k, i\}$. Добавив вершину j к указанному пути, получаем путь из i в j, длина которого не больше $d + 1$.

И обратно, если есть путь из i в j длины не больше $d+1$, то его начальная часть, за исключением последней вершины, является путём из i в такую вершину k, что $(X^d)_{ik} = 1$ (по индуктивному предположению) и k соединена ребром с j, то есть $X_{kj} = 1$. Но тогда $(X^{d+1})_{ij} = 1$.

Осталось заметить, что граф на n вершинах связный тогда и только тогда, когда для любых двух его вершин i, j есть путь из i в j длины меньше n (напомним, что

если существует какой-то путь с данными концами, то существует и простой путь с теми же концами, а в простом пути не более n вершин, а его длина $< n$).

Поэтому граф связный тогда и только тогда, когда X^{n-1} состоит из одних единиц. Схема для проверки связности графа вычисляет X^{n-1}, для чего требуется $O(n^4)$ присваиваний, а затем вычисляет конъюнкцию всех матричных элементов X^{n-1} (ещё $O(n^2)$ присваиваний). Итоговый размер схемы $O(n^4)$.

Замечание 21.1. Аналогичную схему можно построить для проверки сильной связности графа: всё отличие в том, что теперь матрица X не является симметрической, остальные рассуждения остаются в силе.

21.2. Схемная сложность

Основной мерой сложности схемы является её размер.

Определение 21.1. *Схемная сложность* булева отображения $f\colon \{0,1\}^n \to \{0,1\}^m$ (в частности, булевой функции) — это наименьший размер схемы, вычисляющей это отображение.

Наше определение схемной сложности зависит от определения схемы, в котором правые части присваиваний содержат только дизъюнкции, конъюнкции и отрицания. Можно обобщить определение схемы, разрешая в правых частях присваиваний функции из некоторой конечной полной системы функций B (скажем, все функции от не более чем двух переменных). Схемная сложность функции при этом изменится, но не более, чем в константу раз: подставляя в схему, использующую функции из B подсхемы, которые вычисляют функции из B, используя наш стандартный набор, получим схему, которая лишь в $O(1)$ раз больше исходной (тут существенно, что B конечная).

Насколько велика может быть схемная сложность функции от n булевых переменных? Используя ДНФ, можно получить верхнюю оценку.

Теорема 21.2. *Всякую функцию $f\colon \{0,1\}^n \to \{0,1\}$ можно вычислить схемой размера не больше $O(n2^n)$.*

Доказательство. Для всякого $a \in \{0,1\}^n$ рассмотрим такую функцию $f_a\colon \{0,1\}^n \to \{0,1\}$, что $f(x) = 1$ тогда и только тогда, когда $x = a$. Будет удобно ввести обозначения $x^1 = x$ и $x^0 = \neg x$. Тогда функцию f_a можно записать формулой

$$f_a(x) = \bigwedge_{i=1}^{n} x_i^{a_i},$$

где $x = (x_1, \ldots, x_n)$ и $a = (a_1, \ldots, a_n)$.

Произвольная функция f выражается через функции f_a с помощью дизъюнкции:

$$f(x) = \bigvee_{a \in f^{-1}(1)} f_a(x).$$

Эти формулы без труда переделываются в схему. Наша схема сначала будет вычислять отрицания всех переменных, на это нужно n элементов. После этого можно вычислить все функции f_a. Для вычисления каждой такой функции нужно $n-1$ раз применить конъюнкцию. Всего получается $2^n(n-1)$ элемент. Наконец, для вычисления f нужно взять дизъюнкцию нужных функций f_a, на это уйдёт не более 2^n элементов. Суммарно в нашей схеме получается $O(n2^n)$ элементов. □

Мы показали, что всякую функцию можно вычислить схемой, но размер схемы при этом получился большим: экспоненциальным по числу переменных. Оказывается, это неизбежно — существуют функции экспоненциальной схемной сложности.

Теорема 21.3. *Для всякого $n \geqslant 10$ существует функция $f\colon \{0,1\}^n \to \{0,1\}$, которую нельзя вычислить схемой размера меньше $2^n/10n$.*

Доказательство. Для доказательства применим мощностной метод: докажем, что функций больше, чем маленьких схем. Тогда маленьких схем не хватит, чтобы вычислить все функции.

Всего булевых функций от n переменных 2^{2^n}.

Заметим, что если схемная сложность функции не больше S, то существует схема размера ровно S, вычисляющая эту функцию (добавим в схему столько присваиваний $x_1 \wedge x_1$, сколько нужно для выравнивания размера).

Оценим сверху количество схем размера S от n переменных. Для этого заметим, что всякую схему размера S с n переменными можно описать с помощью не больше чем $S \cdot 2(1 + \lceil \log_2(n+S) \rceil)$ битов. Для описания схемы удобно её расширить, добавив в начало все переменные.

Теперь для каждого из S элементов схемы нужно указать его тип (конъюнкция, дизъюнкция, отрицание), на что достаточно потратить два бита. Кроме того, нужно указать, к каким из предыдущих элементов применяется операция. Достаточно указать номера элементов в расширенной последовательности, начинающейся со всех переменных схемы. На это требуется не более $2\lceil \log_2(n+S) \rceil$ битов.

Мы применим эту оценку на длину описания схемы при $S = \lfloor 2^n/10n \rfloor$. В этом случае, как легко проверить, $S > n$ при $n \geqslant 10$. Получаем такую верхнюю оценку на длину описания схемы:

$$S \cdot 2(1 + \lceil \log_2(n+S) \rceil) \leqslant \left(\frac{2^n}{10n} + 1 \right) \cdot 2(1 + \lceil 1 + \log_2 S \rceil) \leqslant \frac{3}{10} \cdot \frac{2^n}{n} \cdot (3 + \log_2 S) \leqslant$$
$$\leqslant \frac{3}{10} \cdot \frac{2^n}{n} \cdot 2 \log_S \leqslant \frac{6}{10} \cdot \frac{2^n}{n} \cdot (n - \log_2(10n)) \leqslant \frac{3}{5} \cdot 2^n.$$

Таким образом, при $n \geqslant 10$ и $S = \lfloor 2^n/10n \rfloor$ всякую схему размера S можно описать строкой из не более $3 \cdot 2^n/5$ битов. Поэтому количество схем размера S не больше, чем количество таких строк, то есть не больше $2^{3 \cdot 2^n/5}$. Это меньше 2^{2^n}, а значит не всякую функцию можно вычислить схемой размера S (или меньшего, как мы заметили с самого начала). □

Нижняя оценка неконструктивна: доказано, что функции экспоненциальной схемной сложности существуют, но не приведено «явного» примера таких функций. Указать явные функции с экспоненциальной схемной сложностью не получается даже при очень слабых требованиях к слову «явно». Это одна из важнейших открытых проблем в теоретической информатике.

Наилучшие нижние оценки для «явных» функций всего лишь линейные.

21.3. Глубина схем

Вторым важным параметром схемы является её глубина. Рассмотрим схему как ориентированный граф. Глубиной вершины в схеме называется длина наибольшего пути из какой-нибудь переменной в эту вершину. Глубиной схемы называется максимальная глубина вершины в этой схеме.

Неформально, глубина схемы характеризует время, необходимое для вычисления значения схемы на данном входе, если вычисление элементов схемы можно производить параллельно: на первом шаге можно параллельно вычислить все элементы глубины 1, на втором — все элементы глубины 2, и так далее, на некотором k-м шаге можно параллельно вычислить все элементы глубины k. Это можно сделать, так как на вход элементам глубины k подаются только элементы меньшей глубины, а их значения уже вычислены.

Пример 21.3. Рассмотрим функцию $x_1 \wedge x_2 \wedge \ldots \wedge x_n$. Её можно вычислить схемой размера $n-1$, просто вычислив последовательно $x_1 \wedge x_2$, $x_1 \wedge x_2 \wedge x_3$ и так далее. Однако глубина такой схемы также равна $n-1$, поскольку каждый следующий элемент получает на вход предыдущий. Ту же функцию можно вычислить схемой глубины $\lceil \log_2 n \rceil$, если организовать схему как двоичное дерево. Разобьём переменные на пары x_1 и x_2, x_3 и x_4 и так далее. Вычислим конъюнкцию каждой пары. Все эти элементы имеют глубину 1. Повторим рассуждение: разобьём полученные элементы на пары и вычислим конъюнкцию парных элементов. Если количество переменных n есть степень двойки, то мы получим полное двоичное дерево глубины $\log_2 n$. Если количество элементов не равно степени двойки, то на некоторых шагах у некоторых элементов не будет парных и мы будем брать их конъюнкцию с самими собой (по существу, мы будем переносить их на следующий шаг). В результате получится поддерево полного двоичного дерева глубины $\lceil \log_2 n \rceil$.

То же самое рассуждение применимо к функции $x_1 \vee x_2 \vee \ldots \vee x_n$.

Из приведённого выше примера и конструкции схемы для произвольной функции из леммы 21.2 видно, что всякую функцию можно вычислить схемой глубины $O(n)$. Действительно, в конструкции с дизъюнктивной нормальной формой возникают конъюнкции n элементов и дизъюнкции 2^n элементов. Если вычислять их так, как описано выше, то глубина схемы будет $O(n)$.

Пример 21.4 (проверка связности графа схемой полиномиального размера и полилогарифмической глубины)**.** В примере 21.2 построены схемы проверки связности

графа на n вершинах. Если для вычисления дизъюнкции n конъюнкций использовать схему глубины $O(\log n)$, как в примере выше, глубина этих схем окажется равной $O(n \log n)$, поскольку мы умножаем матрицу смежности на очередную степень матрицы смежности.

Но есть более экономный способ: последовательное возведение в квадрат. Пусть $t = \lceil \log_2 n \rceil$, тогда $n \leqslant 2^t$. Вычислим последовательность t матриц $Y_1 = X$, $Y_2 = X^{\wedge 2}$, $\ldots$, $Y_k = Y_{k-1}^{\wedge 2}, \ldots$, которые совпадают с последовательностью X^{2^i}, $0 \leqslant i \leqslant t$.

Для проверки связности графа будет достаточно взять конъюнкцию всех элементов Y_t.

В итоге у нас получается схема меньшего размера $O(n^3 \log n)$ (теперь производится всего $O(\log n)$ матричных умножений). Глубина этой схемы оценивается так: предполагаем, что каждая дизъюнкция в формулах матричного умножения вычисляется на глубине $O(\log n)$, и конъюнкция матричных элементов также вычисляется на глубине $O(\log n)$. Тогда общая глубина схемы будет $O(t \log n) = O(\log^2 n)$.

Лекция 22

Множества–2. Схемы и формулы. Сравнение множеств. Счётные множества

22.1. Формулы и схемы

Булеву функцию можно задавать формулой, а можно схемой. Второй способ более общий. Чтобы это увидеть, давайте сопоставим каждой формуле (в стандартной полной системе $\{\wedge, \vee, \neg\}$) схему.

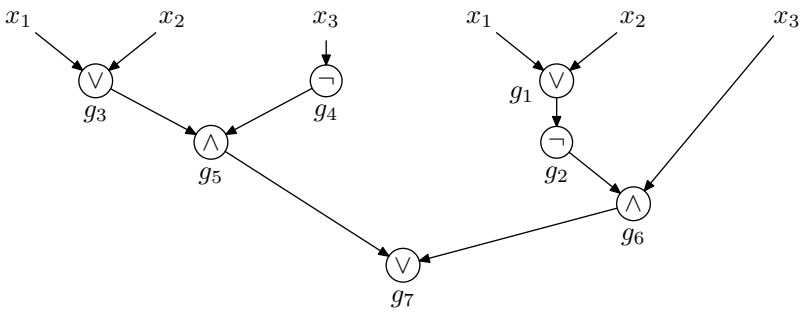
Любая формула в такой системе имеет вид дизъюнкции, конъюнкции или отрицания предыдущих формул или переменных. Запишем соответствующее присваивание, это будет последнее присваивание в схеме. Далее нужно записать присваивания для предыдущих формул и т.д.

Пример 22.1. Рассмотрим формулу

$$(x_1 \vee x_2) \wedge \neg x_3 \vee \neg(x_1 \vee x_2) \wedge x_3.$$

Соответствующая схема показана на рисунке слева.

$g_1 = x_1 \vee x_2$
$g_2 = \neg g_1$
$g_3 = x_1 \vee x_2$
$g_4 = \neg x_3$
$g_5 = g_3 \wedge g_4$
$g_6 = g_2 \wedge x_3$
$g_7 = g_5 \vee g_6$



Справа изображён граф этой схемы. Для удобства в графе есть несколько вершин, помеченных одной переменной. В результате из каждой вершины, кроме последней выходит ровно одно ребро.

На рисунке получилось дерево. Так будет для любой формулы, если размножить вершины, отвечающие переменным. Мы это уже доказывали (добавьте петлю в вершину, отвечающую значению формулы, получится граф, все исходящие степени которого равны 1).

Основное отличие от схем в том, что в формуле нет возможности повторного использования уже вычисленного значения. Скажем, в примере 22.1 приходится повторять вычисление $x_1 \vee x_2$, потому что такое выражение стоит в двух местах формулы.

Верно и обратное. Если каждая функция, вычисленная в схеме, встречается в правых частях присваиваний только один раз, то такая схема превращается в формулу обращением указанной процедуры.

Это оправдывает следующее определение.

Определение 22.1. Схема, в которой каждый элемент встречается в правых частях присваиваний не более одного раза, называется *формулой*.

Теперь будем говорить о размере и глубине формулы, имея в виду размер и глубину соответствующей схемы.

Формулы — это частный случай схем. При этом всякую схему можно переделать в формулу. Глубина при этом не увеличится, однако размер может вырасти экспоненциально.

Лемма 22.2. *По всякой схеме C от переменных $x_1, \ldots, x_n$ глубины d и размера s можно построить формулу, вычисляющую ту же функцию, глубины не больше d и размера не больше $2^s - 1$.*

Доказательство. Доказательство можно вести индукцией по s и d. По существу мы доказываем утверждения для глубины и размера отдельно, но рассуждение одинаково, так что мы сделаем это параллельно. В качестве базы рассмотрим формулу, вычисляющую одну из переменных. Её размер 1 и глубина равна нулю. Собственно, эта схема уже является формулой и утверждение леммы очевидно.

Предположим, что для схем размера меньше s (глубины меньше d) утверждение уже доказано. Рассмотрим схему размера s (глубины d). Посмотрим на её последний элемент g_s. Он получается применением какой-то логической связки к предыдущим элементам g_i и g_j. Рассмотрим подсхемы нашей схемы, вычисляющие g_i и g_j. Их размер меньше s (глубина меньше d), так что по предположению индукции их можно вычислить формулами размера не больше $2^{s-1} - 1$ (глубины меньше d).

Добавим к этой паре формул новую вершину, соответствующую элементу g_s, и проведём в неё ребра из элементов, вычисляющих g_i и g_j. Полученная схема имеет размер не больше $2 \cdot (2^{s-1} - 1) + 1 = 2^s - 1$ (глубину не больше d). □

Для формул можно доказать и обратное соотношение.

Теорема 22.3. *Для всякой формулы размера s есть эквивалентная глубины не больше $1 + 4 \log s$, вычисляющая ту же функцию.*

Вспомнив, что формула — это дерево, можно заметить, что результат этой теоремы означает, что для всякой формулы есть эквивалентная формула, которая не сильно больше изначальной: размер новой формулы $2^{1+4\log_2 s}$ это $O(s^4)$, но все ветви имеют примерно одинаковую глубину.

В доказательстве теоремы 22.3 нам придётся существенно перестраивать формулу. Для этого потребуется вспомогательное утверждение: нужно выделить «среднюю» вершину в дереве.

Утверждение 22.4. *Пусть формула ϕ имеет размер s. Тогда в ней есть такая подформула ψ размера не меньше $s/2$, что все подформулы формулы ψ имеют размер меньше $s/2$.*

Доказательство. Начнём из выходного элемента спускаться по рёбрам формулы в подформулы так, чтобы каждый раз оставаться в подформуле размера не меньше $s/2$. В выходном элементе размер подформулы равен s, а в переменных размер подформулы равен 1, так что в какой-то момент мы не сможем спуститься из очередной вершины в подформулу. Это как раз и будет означать, что у текущей подформулы размер не меньше $s/2$, а во всех её подформулах уже меньше. □

Доказательство теоремы 22.3. Мы применяем индукцию по s. База индукции — формулы размера 1 — очевидна, так как глубина таких формул уже равна 1.

Теперь предполагаем, что теорема верна для формул размера меньше s. Докажем, что тогда она верна и для формул размера s.

В формуле размера s найдём «среднюю» подформулу ψ, удовлетворяющую свойствам из утверждения 22.4. Удалим из дерева формулы поддерево, которое растёт из вершины, соответствующей подформуле ψ. (Более точно, те вершины дерева, путь из которых в корень проходит через ψ.)

Теперь у дерева образовался новый лист: та вершина, где раньше была подформула ψ. Пометим этот лист константой 0. Мы получили новую формулу ϕ_0. Аналогично сделаем для константы 1: пометим новый лист единицей и обозначим полученную подформулу через ϕ_1.

Описанная конструкция не соответствует нашим определениям: мы не разрешали использовать константы вместо переменных. Однако нетрудно видеть, что от констант можно избавиться, не увеличивая ни размера, ни глубины формулы. Действительно, отрицание константы — это опять константа. Конъюнкция с 0 — это 0, с 1 — это второй аргумент конъюнкции. Аналогично, дизъюнкция с 1 — это 1, с 0 — это второй аргумент дизъюнкции.

Пользуясь этими соотношениями, в схеме с константами можно удалять лишние элементы. Ни размер схемы, ни глубина при таком удалении не увеличиваются, а если исходная схема была формулой, то и преобразованная схема останется формулой. Завершается такое удаление эквивалентной схемой без констант (формулой, если исходная схема была формулой).

Далее под ϕ_0 и ϕ_1 понимаем формулы, полученные избавлением от констант.

Теперь рассмотрим формулу $\alpha = (\phi_1 \wedge \psi) \vee (\phi_0 \wedge \neg\psi)$. Нетрудно видеть, что она вычисляет ту же функцию, что и изначальная формула.

Поскольку размер формулы ψ не меньше $s/2$, то размеры формул ϕ_0 и ϕ_1 не больше $s/2$ и к ним применимо предположение индукции. Для этих формул есть эквивалентные формулы глубины $\leqslant 1 + 4\log_2 s - 4 = 4\log_2 s - 3$.

Для формулы ψ есть эквивалентная формула глубины $4\log_2 s - 2$. Действительно, по предположению индукции для самых больших подформул ψ есть эквивалентные формулы глубины $\leqslant 4\log_2 s - 3$. Ещё единица глубины добавляется из-за последней связки в формуле ψ.

Окончательно получаем, что для формулы α, которая эквивалентна исходной, есть эквивалентная формула глубины $\leqslant 4\log_2 s - 2 + 3 = 1 + 4\log_2 s$ (последнее слагаемое в оценке глубины появляется из-за связок в формуле α, которая имеет глубину 3). Это доказывает шаг индукции. □

Таким образом, мы получаем, что для всякой функции $f\colon \{0,1\}^n \to \{0,1\}$ минимальная глубина схемы и логарифм минимального размера формулы линейно связаны. То есть, изучение размера формул — это по существу то же самое, что изучение глубины схем.

22.2. Сравнение множеств

Мы уже сравнивали конечные множества по размеру. Более того, мы проверили такой факт.

Лемма 22.5 (лемма 9.5)**.** *Для тотальных функций из конечного множества в конечное выполняются такие свойства:*

1. *если $f\colon A \to B$ сюръекция, то $|A| \geqslant |B|$;*
2. *если $f\colon A \to B$ инъекция, то $|A| \leqslant |B|$;*
3. *если $f\colon A \to B$ биекция, то $|A| = |B|$.*

Теперь мы будем сравнивать бесконечные множества. В этом случае уже нет возможности пересчитать элементы и сравнить полученные числа. Лемма 9.5 даёт альтернативный способ.

Определение 22.6. Мощность множества A *не больше* мощности множества B, если существует инъекция $f\colon A \to B$.

Мощность множества A равна мощности множества B, если существует биекция $f\colon A \to B$.

В последнем случае множества называются *равномощными*.

Для бесконечных множеств мощность обобщает количество элементов в (конечном) множестве. Это более тонкое понятие. Неочевидно, что при переходе к бесконечным множествам сохранятся привычные свойства сравнения конечных множеств. Как мы увидим, некоторые свойства сохраняются, а некоторые — нет.

22.3. Свойства равномощности

Лемма 22.7. Рефлексивность: *каждое множество равномощно самому себе.*

Симметричность: *если A равномощно множеству B, то B равномощно A для любых множеств A, B.*

Транзитивность: *если A равномощно B и B равномощно C, то A равномощно C.*

Доказательство. Рефлексивность: тождественное отображение $\mathrm{id}\colon x \mapsto x$ задаёт биекцию между A и A.

Симметричность: для всякой биекции $f\colon A \to B$ существует обратная функция $f^{-1}\colon B \to A$ и она также биекция (теорема 10.1).

Транзитивность: композиция $g \circ f$ биекций $f\colon A \to B$ и $g\colon B \to C$ является биекцией (теорема 10.2). □

Свойства из леммы 22.7 — это в точности свойства отношения эквивалентности. Есть соблазн назвать равномощность отношением. Важно понимать, однако, что это — злоупотребление терминологией. Строго говоря, отношение задаётся на каком-то множестве. На каком множестве задана равномощность? Хочется сказать, что на множестве всех множеств. Но тут возникает неожиданная трудность. Можно доказать, что совокупность всех множеств не является множеством. (Это будет сделано в следующий раз.) Более простое соображение показывает, что считать любые совокупности множествами невозможно: получается противоречие.

Парадокс Рассела. Обозначим через x совокупность всех таких множеств, которые не являются своими элементами. Предположим, что x — множество. Будет ли x своим элементом? По построению $y \in x \Leftrightarrow y \notin y$, при любом y, в том числе при $y = x$. Поэтому $x \in x \Leftrightarrow x \notin x$. Противоречие.

Это рассуждение показывает, что при изучении бесконечных множеств нужно соблюдать особо жёсткую логическую дисциплину. Мы не будем описывать теорию множеств на полностью формальном уровне (это трудно). Вместо этого попробуем рассуждать на содержательном уровне, как и раньше, но тщательнее отслеживая логику рассуждений.

22.4. Счётные множества

Определение 22.8. Множество называется *счётным*, если оно равномощно множеству натуральных чисел $\mathbb{N} = \{0, 1, 2, \dots\}$.

Само множество $\mathbb{N}$, очевидно, счётно (рефлексивность равномощности). Приведём ещё несколько простых примеров.

Пример 22.2. Множество чётных натуральных чисел $2\mathbb{N} = \{x : x = 2y,\ y \in \mathbb{N}\}$ счётно. Биекция задаётся отображением, которое можно прочитать в определении: $x \mapsto 2x$. Это отображение сюръективно в силу определения $2\mathbb{N}$. Но оно и инъективно: из $2x = 2y$ следует $x = y$.

Этот пример показывает разницу между конечными и бесконечными множествами. Никакое собственное подмножество B конечного множества A не равномощно множеству A (в нём меньше элементов).

Пример 22.3. Множество квадратов натуральных чисел $\{x : x = y^2,\ y \in \mathbb{N}\}$ счётно. Биекция задаётся отображением $x \mapsto x^2$. Это отображение сюръективно в силу определения множества квадратов и инъективно, так как из $x^2 = y^2$ следует $x = y$ для любых неотрицательных действительных чисел (а не только целых неотрицательных).

Сравнивая эти два примера, видим, что понятие равномощности не согласовано с понятием «плотности». Доля чётных множеств на больших отрезках натурального ряда стремится к $1/2$, а доля квадратов стремится к нулю. В этом нет ничего удивительного: в определении равномощности множества рассматриваются как совокупности элементов и не учитываются никакие отношения между элементами.

Рассмотрим чуть более сложный пример.

Утверждение 22.9. *Множество целых чисел $\mathbb{Z}$ счётно.*

Доказательство. В этом случае задавать биекцию формулой сложнее, хотя и возможно. Вместо этого мы используем такое соображение. Если из элементов множества A можно составить последовательность (бесконечное слово или, как иногда говорят, *сверхслово*)

$$a_0,\ a_1,\ \ldots,\ a_n,\ \ldots,$$

в которой каждый элемент множества A встречается ровно один раз, то эта последовательность задаёт искомую биекцию: $i \mapsto a_i$.

Для целых чисел такую последовательность построить очень легко. Перечисляем целые числа в порядке возрастания абсолютной величины, положительное число предшествует своему противоположному:

$$0,\ 1,\ -1,\ 2,\ -2,\ 3,\ -3,\ \ldots$$ □

22.5. Свойства счётных множеств

Лемма 22.10. *Всякое подмножество счётного множества конечно или счётно.*

Доказательство. Рассмотрим счётное множество A и его подмножество B. Выпишем элементы A в последовательность

$$a_0, a_1, a_2, a_3, \ldots.$$

Вычеркнем из этой последовательности те элементы, которые не лежат в B. В результате останется последовательность элементов B — конечная или бесконечная. В первом случае множество будет конечным, во втором счётным. □

Лемма 22.11. *Объединение двух счётных множеств счётно.*

Доказательство. Рассмотрим два счётных множества A и B; каждое из них можно записать в последовательность, содержащую каждый элемент множества ровно один раз:

$$\begin{array}{cccccc} a_0, & a_1, & a_2, & a_3, & \ldots \\ b_0, & b_1, & b_2, & b_3, & \ldots \end{array}$$

Теперь построим последовательность элементов $A \cup B$, чередуя элементы из A с элементами из B:

$$a_0,\ b_0,\ a_1,\ b_1,\ a_2,\ b_2,\ \ldots.$$

Ясно, что в этой последовательности встречаются все элементы объединения. Если A и B не пересекаются, то на этом рассуждение заканчивается: все элементы объединения встречаются в последовательности ровно по одному разу. Но если $A \cap B$ непусто, что общие элементы встретятся по два раза. Поэтому изменим последовательность, вычёркивая те её члены, которые уже встречались ранее. В модифицированной последовательности каждый элемент $A \cup B$ будет встречаться ровно один раз. □

Эту лемму можно усилить.

Теорема 22.12. *Объединение конечного или счётного числа конечных или счётных множеств конечно или счётно.*

Доказательство. Пусть есть счётное количество счётных множеств $A_0, A_1, A_2, \ldots$. Расположим элементы каждого в последовательность и объединим эти последовательности в дважды бесконечную таблицу:

$$\begin{array}{lccccc} A_0: & a_{00} & a_{01} & a_{02} & a_{03} & \ldots \\ A_1: & a_{10} & a_{11} & a_{12} & a_{13} & \ldots \\ A_2: & a_{20} & a_{21} & a_{22} & a_{23} & \ldots \\ A_3: & a_{30} & a_{31} & a_{32} & a_{33} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}.$$

Здесь в первой строке мы последовательно выписали элементы A_0, во второй — элементы A_1 и так далее. Если какое-то A_i конечно, дополняем его список до бесконечной последовательности последним элементом списка.

Теперь соединяем эти последовательности в одну, двигаясь по диагоналям:

$$a_{00},\ a_{01},\ a_{10},\ a_{02},\ a_{11},\ a_{20},\ a_{03},\ a_{12},\ a_{21},\ a_{30},\ \ldots.$$

В полученной последовательности присутствуют все элементы объединения, но не обязательно по одному разу. Поэтому выполним прополку: двигаемся по полученной последовательности и вычёркиваем те элементы, которые встречались ранее. В результате получится последовательность (быть может, конечная), которая содержит все элементы объединения ровно по одному разу. □

Следствие 22.13. *Декартово произведение двух счётных множеств $A \times B$ счётно.*

Доказательство. По определению декартово произведение есть множество всех упорядоченных пар вида (a, b), в которых $a \in A$ и $b \in B$. Разделим пары на группы, объединив пары с одинаковой первой компонентой (каждая группа имеет вид $\{a\} \times B$ для какого-то $a \in A$). Каждая группа счётна, поскольку находится во взаимно однозначном соответствии с B (пара определяется своим вторым элементом), и групп столько же, сколько элементов в A, то есть счётное число. $\square$

Лекция 23

Множества–3. Мощность континуум. Теорема Кантора–Бернштейна

23.1. Счётные множества — самые маленькие бесконечные

Лемма 23.1. *Всякое бесконечное множество содержит счётное подмножество.*

Доказательство. Рассмотрим произвольное бесконечное множество X. Нам надо выписать последовательность из некоторых его элементов, не обязательно всех. Будем действовать самым простым образом. Первый элемент a_0 возьмём произвольно. Поскольку X бесконечно, в нем есть ещё элементы (кроме a_0). В качестве a_1 возьмём любой из них. И так далее. В общем случае, когда нам нужно выбрать очередной элемент a_n, мы рассматриваем подмножество $\{a_0, \ldots, a_{n-1}\}$. Оно конечно, а значит, не совпадает со всем множеством X (которое по предположению бесконечно). Значит, в X есть элементы, не лежащие в этом подмножестве — и мы можем взять любой из них в качестве a_n.

Получили бесконечную последовательность из элементов X, и множество элементов этой последовательности образует искомое счётное подмножество множества X. □

Эта лемма по сути утверждает, что счётные множества — это «самые маленькие» бесконечные множества (как уже было доказано, между счётными и конечными нет ничего промежуточного, лемма 22.10).

Ещё одна иллюстрация того, что счётные множества — «самые маленькие»: добавление счётного множества к любому бесконечному не изменяет мощность множества.

Лемма 23.2. *Если X — бесконечное множество, а A — конечное или счётное, то $X \cup A$ равномощно X.*

Доказательство. Удобно доказывать этот факт в случае $A \cap B = \varnothing$. Для этого нужно перейти от B к $B \setminus A$, последнее множество конечно или счётно.

По предыдущей лемме 23.1 в X есть счётное подмножество B. Объединение счётного и конечного или счётного счётно (теорема 22.12). Значит, B равномощно $B \cup A$. Пусть $f\colon B \to B \cup A$ — биекция. Рассмотрим отображение

$$g(x) = \begin{cases} f(x), & \text{если } x \in B; \\ x, & \text{если } x \notin B \end{cases}$$

из X в $X \cup A$. Это биекция: обратное отображение задаётся формулой

$$g^{-1}(x) = \begin{cases} f^{-1}(x), & \text{если } x \in B \cup A; \\ x, & \text{если } x \notin B. \end{cases}$$

□

23.2. Диагональный аргумент

Удивительным открытием, с которого и началась систематическая теория множеств, стало то, что бывают множества, строго бóльшие счётных.

Множество конечных последовательностей из 0 и 1 счётно (задача на прошлом семинаре). Бесконечные последовательности — другое дело.

Определение 23.3. Множество имеет мощность *континуум*, если оно равномощно множеству бесконечных двоичных последовательностей.

Теорема 23.4 (Кантор)**.** *Множество бесконечных последовательностей нулей и единиц несчётно.*

Доказательство. Множество бесконечных двоичных последовательностей обозначим $2^{\mathbb{N}}$. Для любой функции $f\colon \mathbb{N} \to 2^{\mathbb{N}}$ докажем, что f не сюръекция (значит, и не биекция).

Обозначим $a_i = f(i)$, а члены последовательности a_i обозначим $a_{i0}, a_{i1}, \ldots$. Запишем члены последовательностей a_i слева направо, а саму последовательность $a_0, a_1, \ldots$ расположим сверху вниз. Получится бесконечная таблица:

$$\begin{array}{ccccccc} a_0 & = & a_{00} & a_{01} & a_{02} & \ldots \\ a_1 & = & a_{10} & a_{11} & a_{12} & \ldots \\ a_2 & = & a_{20} & a_{21} & a_{22} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Теперь рассмотрим «диагональную» последовательность в этой таблице, то есть последовательность

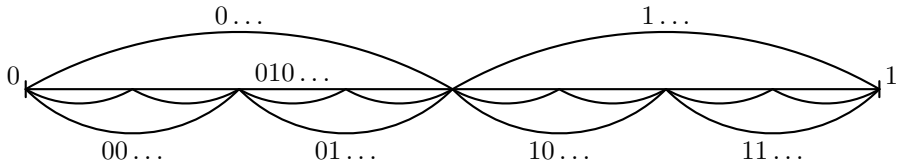
$$a_{00}, a_{11}, a_{22}, \ldots$$

и заменим в ней все биты на противоположные. Другими словами, положим $b_i = 1 - a_{ii}$ и рассмотрим последовательность $b = (b_0, b_1, b_2, \ldots)$. Последовательность b отличается от любой последовательности a_i в i-й позиции, поскольку $b_i = 1 - a_{ii} \neq a_{ii}$. Поэтому $b \notin f(\mathbb{N})$, что и требовалось доказать. □

23.3. Примеры континуальных множеств

Теорема 23.5. *Отрезок* $[0,1]$ *имеет мощность континуум.*

Доказательство. Из курса анализа известно, что каждое число $x \in [0,1]$ можно записать в виде бесконечной двоичной дроби. Напомним, как это делается. Первый знак (бит) после запятой равен 0, если x лежит в левой половине отрезка $[0,1]$, и равен 1, если в правой. Чтобы определить следующий бит, нужно поделить выбранную половину снова пополам. Если x лежит в левой половине, то следующая цифра 0, а если в правой, то 1. И так далее: чтобы определить очередной знак, нужно поделить текущий отрезок пополам и посмотреть, в какую половину попадает x.



Построенное соответствие между числами на отрезке $[0,1]$ и бесконечными двоичными последовательностями не взаимно однозначно. Некоторым числам соответствуют две последовательности. А именно, это происходит, когда точка попадает на границу очередного отрезка. Тогда мы можем относить её как к левой, так и к правой половине. В результате, например, последовательности $0.1001111\ldots$ и $0.101000\ldots$ соответствуют одному и тому же числу.

Назовём *плохими* те последовательности, в которых начиная с некоторого момента все цифры равны 1, за исключением последовательности $0.111\ldots$). Плохих последовательностей счётное множество: каждое двоичное слово однозначно продолжается до последовательности такого вида добавлением бесконечного суффикса $0111\ldots$. Поэтому добавление плохих последовательностей не меняет мощности множества по лемме 23.2. □

Из теоремы 23.5 легко получается континуальность других числовых множеств.

Следствие 23.6. *Интервал* $(0,1)$ *имеет мощность континуум.*

Доказательство. Отрезок является объединением интервала и двухэлементного множества (концы отрезка). Из леммы 23.2 следует искомое. □

Следствие 23.7. *Множество действительных чисел* $\mathbb{R}$ *имеет мощность континуум.*

Доказательство. Все интервалы равномощны, биекция задаётся линейным отображением. Скажем, биекция $(0,1)$ и $(-\pi/2, \pi/2)$ задаётся линейной функцией $f(x) = \pi x - \pi/2$ (монотонной и непрерывной).

Биекция интервала $(-\pi/2, \pi/2)$ и всех действительных чисел задаётся функцией $\operatorname{tg} x$ (монотонной и непрерывной на интервале $(-\pi/2, \pi/2)$). □

Есть ещё одно стандартное множество мощности континуум.

Теорема 23.8. *Подмножества* $\mathbb{N}$ *имеют мощность континуум.*

Доказательство. Стандартная биекция между подмножествами и индикаторными функциями продолжается в случае $\mathbb{N}$ до биекции между подмножествами и бесконечными двоичными последовательностями: индикаторной функции множества $S \subseteq \mathbb{N}$ сопоставляется двоичная последовательность, в которой на месте i стоит $\chi_S(i)$. □

Следствие 23.9. *Множество всех функций* $f\colon \mathbb{N} \to \{0,1\}$ *имеет мощность континуум.*

Теперь докажем довольно неожиданный результат. (Изобретатель теории множеств немецкий математик XIX века Георг Кантор много лет пытался доказать обратное.)

Теорема 23.10. *Отрезок* $[0,1]$ *равномощен квадрату* $[0,1] \times [0,1]$.

Доказательство. Отрезок $[0,1]$ равномощен множеству бесконечных последовательностей нулей и единиц. Тогда квадрат $[0,1] \times [0,1]$ равномощен множеству упорядоченных пар таких последовательностей. Действительно, чтобы получить пару, соответствующую точке (x,y), надо составить пару из последовательности, соответствующей x, и последовательности, соответствующей y.

Осталось установить взаимно однозначное соответствие между бесконечными последовательностями нулей и единиц и парами таких последовательностей (и воспользоваться транзитивностью равномощности). Паре последовательностей

$$(a_0, a_1, a_2, a_3, \dots\ , b_0, b_1, b_2, b_3, \dots)$$

сопоставим последовательность

$$a_0, b_0, a_1, b_1, a_2, b_2, a_3, b_3, \dots.$$

Нетрудно увидеть, что это отображение взаимно однозначное (обратное к нему выделяет из последовательности отдельно чётные и отдельно нечётные члены). □

23.4. Теорема Кантора–Бернштейна

Вернёмся к общим вопросам сравнения множеств. Напомним, что мощность A не больше мощности B, если существует инъекция $f\colon A \to B$. Возможно, что и мощность B не больше мощности A, то есть существует другая инъекция $g\colon B \to A$. Но следует ли из этих двух утверждений, что множества равномощны? Ответ оказывается положительным.

Теорема 23.11. *Если для множеств* A *и* B *существует инъекция из* A *в* B *и инъекция из* B *в* A*, то существует и биекция между* A *и* B.

Доказательство. Будем считать, что $A \cap B = \varnothing$. Этого легко добиться, заменив общую пару множеств A, B на пару $\{0\} \times A$, $\{1\} \times B$.

Пусть $f\colon A \to B$ и $g\colon B \to A$ — инъекции. Рассмотрим (возможно, бесконечный) ориентированный граф с вершинами $A \cup B$. Для точек $x \in A$ и $y \in B$ мы проводим ребро из x в y, если $f(x) = y$, и ребро из y в x, если $g(y) = x$. Если нарисовать множество A слева, а множество B справа, то можно сказать, что мы проводим рёбра слева направо согласно функции f и справа налево согласно функции g.

По построению из каждой точки выходит ровно одно ребро. А сколько рёбер входит? Поскольку функции инъективны, то не больше одного (но может не входить не одного).

Разобьем граф на компоненты связности (забыв для этого об ориентации рёбер) и рассмотрим каждую компоненту отдельно. Как устроены эти компоненты? Есть три возможности. Связная компонента может быть

- циклом из стрелок;
- бесконечной цепочкой стрелок, начинающейся в некоторой вершине (в которую ничего не входит);
- бесконечной в обе стороны цепочкой стрелок.

В самом деле, вперёд всегда можно идти по единственной стрелке, а назад либо можно пойти единственным образом, либо нельзя пойти вовсе. Если, идя вперёд, мы дважды попадём в одну вершину, то образуется цикл (и это возможно, лишь если мы вернёмся в начальную вершину). Если нет, то образуется бесконечная цепочка вперёд; её можно однозначно продолжать назад, при этом либо мы упрёмся в вершину, где назад не пройти, либо получим двустороннюю цепочку.

Это верно для любого ориентированного графа, в котором из каждой вершины выходит ровно одна стрелка и в каждую вершину входит не больше одной стрелки. В нашем конкретном случае есть дополнительная структура: вершины бывают левые и правые (из A и из B). Они чередуются, поэтому цикл может быть только чётной длины и содержит поровну вершин из A и из B. Любое из отображений f и g может быть использовано, чтобы построить биекцию между A- и B-вершинами цикла (так что есть минимум два варианта биекции). То же самое верно для бесконечной в обе стороны цепочки (два варианта). Если же цепочка бесконечна только в одну сторону, то для построения биекции годится только одно из отображений. Скажем, если она начинается с элемента $a \in A$, то годится только функция f (при которой a соответствует $f(a)$, затем $g(f(a))$ соответствует $f(g(f(a)))$ и так далее). Но в любом случае одна из функций f и g годится, так что внутри каждой связной компоненты у нас есть биекция, и остаётся их объединить для всех связных компонент. □

В этом доказательстве теоремы Кантора–Бернштейна искомая биекция h строится только из тех пар, которые отвечают двум инъекциям. В результате множество A разбивается на части A_1 и A_2, а множество B разбивается на части B_1 и B_2, при этом f осуществляет биекцию между A_1 и B_1, а g осуществляет биекцию между B_2 и A_2.

Отсюда получаем интересный результат.

Утверждение 23.12. *Квадрат и круг можно разбить на пары взаимно подобных частей.*

Доказательство. Нужно доказать, что квадрат разбивается в объединение непересекающихся фигур A_1 и A_2, круг разбивается в объединение непересекающихся фигур B_1 и B_2, при этом A_1 подобно B_1, а A_2 подобно B_2.

Возьмём преобразование подобия, которое отображает квадрат внутрь круга, оно будет инъекцией. (Внутри любого круга есть маленький квадрат.) Аналогично можно поместить круг в квадрат с помощью другого преобразования подобия. Остаётся применить доказательство теоремы Кантора–Бернштейна к этим двум инъекциям. □

Теорема Кантора–Бернштейна упрощает доказательства равномощности: пару инъекций найти легче, чем биекцию.

Пример 23.1. Множество $\mathbb{N}^{\mathbb{N}}$ тотальных функций из $\mathbb{N}$ в $\mathbb{N}$ имеет мощность континуум. Это можно доказать и явным построением биекции, но применить теорему Кантора–Бернштейна проще.

Мы уже проверили, что множество $2^{\mathbb{N}}$ всех тотальных функций $\mathbb{N} \to \{0,1\}$ имеет мощность континуум. Это даёт инъекцию $2^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$.

С другой стороны, функция из $\mathbb{N}$ в $\mathbb{N}$ задаётся своим графиком, то есть подмножеством $\mathbb{N} \times \mathbb{N}$. Как мы уже знаем, $\mathbb{N} \times \mathbb{N}$ счётно. Поэтому получаем инъекцию из $\mathbb{N}^{\mathbb{N}}$ в $2^{\mathbb{N}}$.

Теперь осталось применить теорему Кантора–Бернштейна.

23.5. За пределами континуума

Мощность континуума не предел. Обобщением диагонального рассуждения получается следующий результат.

Теорема 23.13 (Кантор)**.** *Никакое множество X не равномощно множеству своих подмножеств.*

Доказательство. Предположим, что множество X и множество всех его подмножеств (оно обозначается 2^X) равномощны. Пусть f — отображение из X в 2^X. Докажем, что оно не сюръекция (значит, и не биекция). Для этого построим множество Y, отличающееся от всех $f(y)$.

А именно, требуем, чтобы в точке y множество Y вело себя не так же, как $f(y)$ (ведь отличия в одной точке достаточно для различия множеств).

Технически закончить рассуждение нужно так. Рассмотрим те элементы x, которые не принадлежат соответствующим им подмножествам, то есть $x \notin f(x)$. Рассмотрим теперь множество всех таких элементов:

$$Y = \{y \in X : y \notin f(y)\}.$$

Это множество не лежит в образе $f(X)$. Пусть $z \in X$. Докажем, что $Y \neq f(z)$. Есть два случая:

(1) $z \in f(z)$. Тогда $z \notin Y$, то есть $f(z) \setminus Y \neq \varnothing$.

(2) $z \notin f(z)$. Тогда $z \in Y$, то есть $Y \setminus f(z) \neq \varnothing$. □

Из этой теоремы Кантора можно заключить, что множества всех множеств не существует. Предположим, что совокупность всех множеств U является множеством. Заметим, что все подмножества множества U (и вообще все подмножества любых множеств!) являются его элементами. Поэтому множество 2^U всех подмножеств U является подмножеством U. Но по теореме Кантора мощность 2^U больше U, приходим к противоречию.

23.6. Индукция по фундированным множествам

Из естественных свойств сравнения множеств по мощности мы не проверили одно, быть может, самое важное: любые два множества сравнимы по мощности. Более точно, для любых множеств A и B выполняется одно из трёх: либо есть инъекция из A в B, но нет инъекции из B в A (B «мощнее» A); либо A и B равномощны; либо есть инъекция из B в A, но нет инъекции из A в B (A «мощнее» B).

По сути нужно доказать существование инъекции одного множества в другое или второго в первое. Но сделать это непросто. Нужны новые инструменты для рассуждений и дополнительная структура на множествах.

Необходимым инструментом является обобщение рассуждений по индукции.

Принцип полной математической индукции можно переформулировать таким образом:

> **Математическая индукция «без базы».** *Пусть для утверждения $A(n)$, зависящего от натурального параметра n, для любого n верно утверждение «если $A(m)$ верно при всех $m < n$, то и $A(n)$ верно». Тогда утверждение $A(n)$ верно при любом n.*

Отсутствие базы индукции тут мнимое. База скрыта в более сложном индуктивном предположении. Действительно, для $n = 0$ посылка условного индуктивного предположения всегда истинна (так как нет натуральных чисел, меньших n). Поэтому $A(0)$ обязано быть истинным.

В такой формулировке принципа математической индукции используется порядок на натуральных числах. Однако аналогичное утверждение справедливо для более широкого класса множеств.

Теорема 23.14. *Следующие свойства порядка P равносильны:*

1°: *каждое непустое подмножество имеет минимальный элемент;*

2°: *любая убывающая цепь конечна;*

3°: *для порядка P справедлив принцип индукции: если для утверждения $A(p)$, зависящего от элемента порядка, для любого p верно утверждение «если $A(q)$ верно при всех $q < p$, то и $A(p)$ верно». Тогда утверждение $A(p)$ верно при любом $p \in P$.*

Доказательство. 1° $\Rightarrow$ 2°. Докажем контрапозицию: из отрицания 2° следует отрицание 1°. Если в порядке есть бесконечная убывающая цепь $y_1 > y_2 > \ldots$, то в ней нет минимального элемента.

2° $\Rightarrow$ 1°. Доказываем контрапозицию этого утверждения. Пусть для порядка 1° не выполняется. Возьмём множество X без минимальных элементов и построим бесконечную убывающую цепь. Выберем какой-нибудь $x_0 \in X$. Он не минимален, поэтому есть $x_1 < x_0$. Аналогично рассуждаем с x_1 и так далее. Получаем бесконечную убывающую цепь, то есть для такого порядка истинно 2° не выполняется.

Теперь выведем принцип индукции для P из существования минимальных элементов. Рассмотрим множество тех x, для которых не выполняется $A(x)$. Если оно непусто, в нем есть минимальный элемент m. Но тогда для всех $y < m$ утверждение $A(y)$ верно и в силу предположения индукции $A(m)$ тоже верно. Получили противоречие, так как по выбору m утверждение $A(m)$ ложно. Значит, имеет место единственная оставшаяся возможность: множество тех x, для которых не выполняется $A(x)$, пусто.

Осталось сделать ещё одну проверку: что из принципа индукции следует существование минимальных элементов в непустых множествах. Предположим, что множество X не имеет минимальных элементов. Возьмём в качестве утверждения $A(p)$ такое: $p \notin X$.

Индуктивное предположение выполняется: если для всех $q < p$ выполняется $q \notin X$, то и $p \notin X$ (иначе p — минимальный элемент). Поэтому $p \notin X$ для всех p, то есть $X = \varnothing$. □

Определение 23.15. Порядок, удовлетворяющий условиям теоремы 23.14, называется *фундированным*. Множество с фундированным порядком для краткости также называется фундированным.

Лекция 24

Множества–4. Вполне упорядоченные множества

В этом лекции мы обсудим, почему любые два множества сравнимы по мощности. Для этого нам потребуется индукция по фундированным множествам особого вида.

Определение 24.1. Множество с порядком называется *вполне упорядоченным*, если оно фундированное и линейно упорядочено (всякие два его элемента сравнимы).

Для линейных порядков нет разницы между минимальными и наименьшими элементами. Поэтому каждое непустое множество вполне упорядоченного множества имеет наименьший элемент.

Индукцию по вполне упорядоченному множеству принято называть *трансфинитной индукцией*. Как и для любого фундированного множества трансфинитная индукция равносильна существованию в каждом непустом подмножестве наименьшего элемента (принцип наименьшего числа в конечном случае) и конечности убывающих цепей (метод бесконечного спуска в конечном случае).

Помимо этого для доказательства сравнимости множеств по мощности требуется *теорема Цермело*.

Теорема 24.2 (теорема Цермело)**.** *Всякое множество можно вполне упорядочить.*

Доказательство теоремы Цермело в очень сильном смысле неконструктивно. Естественного способа вполне упорядочить произвольное множество нет. Даже найти вполне упорядочение множества действительных чисел очень трудно, нет никакого «явного» способа сделать это.

Теорема Цермело равносильна нескольким другим утверждениям, в частности, *аксиоме выбора*: по любому семейству непустых множеств можно построить множество, включив в него по одному элементу из каждого множества, входящего в семейство. Аксиома выбора входит в список аксиом самой популярной аксиоматики теории множеств — Цермело–Френкеля, ZFC. Аксиома выбора позволяет доказать теорему Цермело и много других интересных фактов. Некоторые из них противоречат интуиции. Например, можно доказать с использованием аксиомы выбора, что

шар возможно разделить на конечное количество частей, переместив которые можно собрать два таких же шара. Или шар удвоенного радиуса (разбиения для этих двух трюков разные).

План доказательства сравнимости любой пары множеств состоит в том, чтобы сначала доказать сравнимость вполне упорядоченных множеств, а потом применить теорему Цермело. На самом деле, про вполне упорядоченные множества будет доказан более сильный факт: одно из двух вполне упорядоченных множеств изоморфно как порядок подмножеству другого (или наоборот).

24.1. Начальные отрезки вполне упорядоченного множества

Возьмём какое-нибудь вполне упорядоченное множество X. В нём есть наименьший элемент, обозначим его 0. Если $X\setminus\{0\} \neq \varnothing$, то в нём также есть наименьший элемент, обозначим его 1. Продолжая в том же духе, будем последовательные элементы X и обозначать их $2, 3, \ldots$

В какой-то момент процесс может остановиться. Это произойдёт, если множество X конечное. Мы получим линейный порядок на конечном множестве. Если же множество X бесконечное, то получим такое начало порядка на X:

$$0 < 1 < 2 < \cdots < n < \ldots$$

Будут использованы все натуральные числа и они образуют обычный порядок $\mathbb{N}$. Если $X \setminus \mathbb{N}$ не пусто, то в нём существует наименьший элемент, который мы обозначим ω. Рассуждая аналогично, получим элементы $\omega+1$, $\omega+2$, $\ldots$ (либо множество X исчерпается на одном из шагов).

В совокупности эти элементы образуют порядок $\mathbb{N}+\mathbb{N}$ (одна копия порядка на натуральных числах следует за другой). Но множество X может и не закончиться на этом. Наименьший элемент в дополнении обозначим $\omega \cdot 2$. Далее должен идти $\omega \cdot 2 + 1$ и весь процесс повторится.

Заметим, что никакой «развилки» у нас не было: каждый раз мы получали однозначно определённый порядок. Обобщим все сделанные шаги.

Определение 24.3. *Начальным отрезком* вполне упорядоченного множества X называется такое его подмножество $I \subseteq X$, что $x < y$ для любых $x \in I$, $y \in X \setminus I$.

Пусть I — начальный отрезок вполне упорядоченного множества X. Если $X \setminus I \neq \varnothing$, то в нём есть наименьший элемент x. И тогда $I = [0, x) = \{z : 0 \leqslant z < x\}$, так как все элементы полуинтервала $[0, x)$ должны лежать в I, и ни один другой элемент не лежит в I (все остальные не меньше $x \in X \setminus I$).

Таким образом, начальный отрезок либо является полуинтервалом $[0, x)$, либо совпадает со всем X. Обозначим множество начальных отрезков $\widehat{X}$ и упорядочим его по включению. Получаем вполне упорядоченное множество. Действительно, если непустое множество $\mathcal{I}$ начальных отрезков содержит только X, то X — наименьший элемент в этом множестве. В противном случае непусто множество $\{x \in X : [0, x) \in$

$\mathcal{I}\}$. В нём есть наименьший элемент x_0, соответствующий отрезок будет наименьшим срели отрезков, входящих в множество $\mathcal{I}$.

Порядок $\widehat{X}$ изоморфен порядку $X+1$ (порядок X, к которому добавлен наибольший элемент. Порядковый изоморфизм $X+1 \to \widehat{X}$ устанавливается правилом $x \mapsto [0,x)$, $x \in X$, а наибольший элемент порядка $X+1$ переходит в X.

Лемма 24.4. *Различные начальные отрезки вполне упорядоченного множества неизоморфны как упорядоченные множества.*

Доказательство. От противного. Предположим, что $X \cong [0,a)$ для некоторого $a \in X$. Обозначим соответствующий порядковый изоморфизм $\varphi\colon [0,a) \to X$, а множество неподвижных точек — $S = \{x : \varphi(x) = x\}$. Если $[0,a) \setminus S \neq \varnothing$, то в нём есть наименьший элемент u, $\varphi(u) \neq u$. Для любого $x < u$ выполняется $\varphi(x) = x < \varphi(u)$ (монотонность порядкового изоморфизма), значит, $\varphi(u) > u$. Обозначим $v = \varphi^{-1}(u)$. И опять $v > u$, так как все меньшие элементы остаются на месте. Но $\varphi(v) = u < \varphi(u)$ — противоречие с монотонностью.

Значит, $S = [0,a)$. Но в этом случае у $a \in X \setminus [0,a)$ нет прообраза. Опять приходим к противоречию.

Аналогично рассуждаем в случае $[0,a) \cong [0,b)$, $a < b$ (заменим всюду X на $[0,b)$). □

24.2. Доказательство теоремы о сравнимости вполне упорядоченных множеств

Используя «жёсткость» начальных отрезков вполне упорядоченного множества, докажем такой факт.

Теорема 24.5. *Пусть X, Y — вполне упорядоченные множества. Тогда X изоморфен начальному отрезку Y или Y изоморфен начальному отрезку X.*

«Или» в формулировке этой теоремы неисключающее. Возможно, что X изоморфно Y, тогда выполняются оба члена дизъюнкции.

Порядковый изоморфизм — это биекция с подмножеством, то есть она задаёт инъекцию одного множества в другое. Это как раз то, к чему мы стремимся. Но теорема 24.5 устанавливает этот факт не для всех множеств, а только для вполне упорядоченных.

Если вполне упорядоченное множество Y изоморфно начальному отрезку вполне упорядоченного множества X, то утверждение теоремы 24.5 выполняется.

Теперь рассмотрим второй случай, когда Y неизоморфно никакому начальному отрезку X. В этом случае докажем, что X изоморфно какому-то начальному отрезку Y. Доказательство трансфинитной индукцией по множеству $\widehat{X}$ начальных отрезков порядка X.

Докажем для любого начального отрезка I порядка X условное утверждение: если каждый $I' < I$ изоморфен начальному отрезку Y, то и I изоморфен начальному отрезку Y. Отсюда по трансфинитной индукции будет следовать, что всякий

начальный отрезок X, в том числе и сам X, изоморфен начальному отрезку Y, что и требуется доказать.

В силу леммы 24.4 каждый начальный отрезок $I' < I$ изоморфен ровно одному начальному отрезку Y, причём этот начальный отрезок отличен от Y по сделанному предположению. Поэтому корректно определена (тотальная) функция $\varphi\colon I \to Y$, которая сопоставляет $x \in I$ такое $y = \varphi(x) \in Y$, что $[0, x) \cong [0, y)$. Эта функция инъективна в силу леммы 24.4. Докажем, что она монотонна от противного. Пусть $x' < x''$, но $\varphi(x') > \varphi(x'')$. Тогда

$$[0, x'') \cong [0, \varphi(x'')) \subset [0, \varphi(x')) \cong [0, x'),$$

то есть $[0, x'')$ изоморфен какому-то начальному отрезку $[0, x')$, что невозможно по лемме 24.4.

Итак, φ задаёт порядковый изоморфизм I и $\varphi(I)$. Осталось доказать, что образ $\varphi(X)$ является начальным отрезком Y. Рассмотрим $y < \varphi(x)$. По определению существует порядковый изоморфизм $\alpha\colon [0, x) \to [0, \varphi(x))$. Обозначим $a = \alpha^{-1}(y)$. Для любого $x' < x$ начальный отрезок $[0, x')$ изоморфен $[0, \alpha(x'))$, в силу леммы 24.4 получаем, что $\varphi(x') = \alpha(x')$. Но тогда $\varphi(a) = y$. Поэтому $\varphi(X)$ является начальным отрезком.

24.3. Доказательство теоремы Цермело

Пусть A — некоторое множество. Мы хотим доказать, что его можно вполне упорядочить. Для этого применяем аксиому выбора ко всем непустым подмножествам A. Аксиома выбора гарантирует существование функции $\varphi\colon 2^A \setminus \varnothing \to A$, обладающей свойством $\varphi(S) \in S$ для любого $S \in 2^A \setminus \varnothing$.

На самом деле, нам понадобится другая функция, которая выражается через φ: $f(S) = \varphi(A \setminus S)$. Функция f определена для всех подмножеств A, кроме самого A, и удовлетворяет свойству $f(S) \notin S$ для любого S. Используя эту функцию, мы определим вполне упорядочение на A.

Обозначим $a_0 = f(\varnothing)$, $a_1 = f(\{a_0\})$, $\ldots$, $a_n = f(\{a_0, a_1, \ldots, a_{n-1}\})$, $\ldots$ Получаем последовательность элементов A, она упорядочена как натуральные числа. Если множество A содержит элементы, не входящие в эту последовательность, то определим $a_\omega = f(\{a_0, \ldots, a_n, \ldots\})$. И продолжим этот процесс аналогично построению начальных отрезков вполне упорядоченного множества. Этот процесс и даёт в итоге вполне упорядочение A, но это, конечно, требует доказательства (слово «процесс» слишком неформально, чтобы быть убедительным в рассуждениях об абстрактных множествах).

Определим семейство $\mathcal{R}$ *правильных подмножеств* множества A. Точнее говоря, это семейство будет содержать вполне упорядоченные подмножества множества A, обладающие такими свойствами: если $(S, \leqslant) \in \mathcal{R}$, то

1. $S \subseteq A$, $\leqslant$ — вполне упорядочение S;

2. для любого $x \in S$ выполняется равенство $x = f([0, x))$ (здесь отрезок берётся относительно порядка $\leqslant$).

Из этих свойств сразу следует, что в любом непустом правильном порядке $(S, \leqslant) \in \mathcal{R}$ наименьшим элементов будет a_0 (так как начальный отрезок наименьшего элемента — пустое множество), если S бесконечно, то оно также содержит элементы a_1, a_2, ...

Нам нужно доказать, что $\mathcal{R}$ содержит $(A, \leqslant)$, это и будет означать, что на A есть вполне упорядочение. Заметим, что наряду с каждым множеством $(S, \leqslant)$, $S \neq A$, семейство $\mathcal{R}$ содержит множество $S \cup \{f(S)\}$, порядок на котором продолжает порядок на S таким образом, чтобы $f(S)$ был наибольшим.

Основная часть доказательства состоит в том, чтобы проверить согласованность правильных подмножеств: они линейно упорядочены по включению и порядки на пересечениях совпадают.

Лемма 24.6. *Для любых двух правильных множеств одно является начальным отрезком другого.*

Доказательство. Пусть $(S, \leqslant_S)$ и $(T, \leqslant_T)$ — правильные множества. Определим два множества

$$R_S = \{x \in S : [0, x)_{\leqslant_S} = [0, x)_{\leqslant_T}\},$$
$$R_T = \{x \in T : [0, x)_{\leqslant_S} = [0, x)_{\leqslant_T}\},$$

равенство означает, что начальные отрезки совпадают и как множества, и как порядки. (В частности, эти начальные отрезки принадлежат $S \cap T$.)

Докажем от противного, что $R_S = S$ или $R_T = T$. Пусть эта дизъюнкция не выполняется. Обозначим b_S наименьший элемент в $S \setminus R_S$, b_T — наименьший в $T \setminus R_T$.

Докажем, что $[0, b_S)_S \subseteq T$. Если $x < b_S$, то $x \in R_S$. То есть $[0, x)_{\leqslant_S} = [0, x)_{\leqslant_T}$. Но по определению правильного множества $x = f([0, x)_{\leqslant_S}) = f([0, x)_{\leqslant_T})$. Поэтому $x \in T$ (T — правильное множество).

Аналогично доказывается, что $[0, b_T)_T \subseteq S$.

Если $b_S <_S b_T$, то из определения R_T заключаем, что $[0, b_S)_{\leqslant_S} = [0, b_S)_{\leqslant_T}$, то есть $b_S \in R_S$ — противоречие.

Аналогичное противоречие получается, если $b_T <_S b_S$: тогда $[0, b_T)_{\leqslant_S} = [0, b_T)_{\leqslant_T}$.

Значит, $b_S = b_T = b$. Из определения R_S, R_T следует, как и в рассуждении выше, что порядки $\leqslant_S$ и $\leqslant_T$ совпадают при всех $x <_S y <_S < b$ и при всех $x <_T y <_T < b$. Кроме того, невозможно, чтобы $b <_S u <_T b$: в таком случае $[0, u)_{\leqslant_S} = [0, u)_{\leqslant_T}$ по определению R_T и получаем, что $b = b_S \in R_S$ — противоречие. Аналогично, невозможно $b <_T u <_S b$. Поэтому $[0, b)_{\leqslant_S} = [0, b)_{\leqslant_T}$, приходим к противоречию с определением R_S, R_T.

Далее без ограничения общности считаем, что $R_S = S$. Докажем, что в таком случае $S \setminus T = \varnothing$. От противного: если $S \setminus T \neq \varnothing$, то обозначим s минимальный в $S \setminus T$. Так как $[0, s)_{\leqslant_S} = [0, s)_{\leqslant_T}$, то получаем из определения правильности T, что $s = f([0, s)_{\leqslant_T}) \in T$ — противоречие.

Итак $S \subseteq T$ и на S порядки $\leqslant_S$ и $\leqslant_T$ совпадают. Тогда S — начальный отрезок T: если $x <_T y \in S$, то из совпадения начальных отрезков для y получаем, что $x \in S$. □

Лемма 24.7. *Объединение всех правильных множеств — правильное.*

Доказательство. Обозначим $\widetilde{A}$ объединение всех правильных множеств. Любые два элемента $x, y \in \tilde{A}$ лежат в каком-то правильном множестве (из правильного множества, содержащего x и правильного множества, содержащего y, возьмём бóльшее). Во всех правильных множествах, которые содержат эти два элемента, они сравниваются одинаково. Это задаёт линейный порядок на $\widetilde{A}$. Любая бесконечная убывающая цепь в $\widetilde{A}$ лежит в каком-то правильном множестве и потому конечна. По теореме 23.14 это означает, что $\widetilde{A}$ вполне упорядочено.

Любой $x \in \widetilde{A}$ лежит в каком-то правильном множестве S и потому $x = f([0, x))$ (отрезок один и тот же для S и для $\widetilde{A}$). Это и означает, что $\widetilde{A}$ правильное. □

Чтобы закончить доказательство теоремы Цермело, нужно доказать, что $\widetilde{A} = A$. Действительно, в противном случае множество $\widetilde{A} \cup \{f(\widetilde{A})\}$ тоже правильное, но оно должно лежать в $\widetilde{A}$ как объединении всех правильных множеств. Противоречие.