Часть I

Начальные примеры

Лекция 1

Математическая индукция

Принцип математической индукции — метод доказательства, в котором какое-то свойство натуральных чисел доказывается сначала для какого-то числа (например, нуля), потом для следующего (единицы), потом для двойки и так далее. При этом, доказывая очередное утверждение, мы опираемся на предыдущие (ранее доказанные). Возможно, вы уже встречались с этим принципом — тогда можно быстро проглядеть примеры из этой главы и убедиться, что они вам хорошо знакомы. Если же нет, то не страшно — ничего сложного тут нет, надо только немного привыкнуть.

1.1 Задача о раскраске плоскости

Начнем с такой задачи.

Задача 1.1. На плоскости проведено несколько прямых. Они делят плоскость на области. Докажите, что области можно так раскрасить в два цвета, что соседние (имеющие общий участок границы) области покрашены в разные цвета.

Обратите внимание, что области, имеющие общий угол, разрешается красить в один цвет (иначе уже с двумя пересекающимися прямыми были бы проблемы).

На рисунке 1.1 показана искомая раскраска для нескольких конкретных примеров. Но нам нужно доказать, что это возможно всегда, то есть нужно какое-то общее рассуждение.

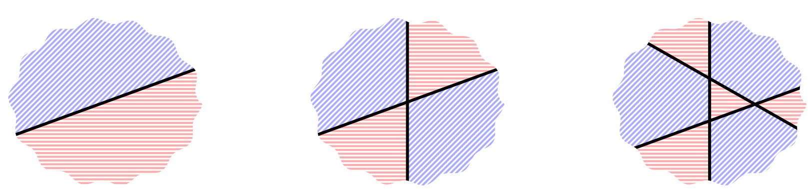


Рис. 1.1: Примеры раскрасок

Чем больше прямых, тем сложнее выглядит наша задача. Начнём с простого случая, скажем, с одной прямой. В этом случае всё понятно: она делит плоскость на две полуплоскости разного цвета. Будем теперь добавлять прямые по очереди.

Что происходит при добавлении одной прямой? Пусть есть набор прямых с требуемой раскраской областей, и мы проводим добавляем ещё одну прямую (пунктир), как на рисунке 1.2.

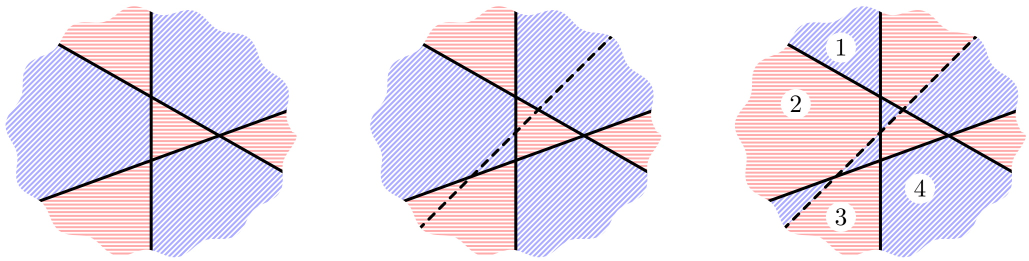


Рис. 1.2: Добавление прямой

Как получить раскраску для новой конфигурации? Старая раскраска не годится: новая прямая разрезает некоторые куски на две части, и эти части одного цвета — а должны быть разного. Если мы перекрашиваем одну из частей, возникают новые проблемы у её краёв. Ясно, что нужно действовать более глобально. Как? Перекрасим в другой цвет все области по одну сторону от новой прямой (рис. 1.2, правая часть). Тогда и у новой прямой будет всё в порядке (потому что с одной стороны мы перекрасили, а с другой нет), и у старых тоже ничего не испортится.

Заметим, что если исходная раскраска была правильной, то и новая раскраска правильная. Действительно, те соседние области, которые не граничат по новой прямой, либо одновременно меняют цвет (как области 1 и 2 на рисунке), либо одновременно не меняют (как области 3 и 4 на рисунке). Поэтому совпасть цвета у них не могут (в исходной раскраске всё было правильно). Те же соседние области, которые граничат по новой прямой, до перекраски были одного цвета. Поскольку перекрашивается только одна из этих областей, то после перекраски эти области будут иметь разные цвета.

По существу мы уже всё сказали, но давайте повторим решение аккуратно, как если бы мы собирались отдавать работу на проверку педантичному и придирчивому экзаменатору.

В нашей задаче есть параметр: число прямых (совсем педантично: число различных прямых). Нам надо доказать возможность раскраски для любого значения параметра, то есть для любого числа прямых. Обозначив наш параметр, скажем, n, получим такую серию утверждений:

> $A(n)$: *если плоскость разделена n прямыми на части, то эти части можно раскрасить в два цвета таким образом, чтобы части, имеющие общий участок границы, были разного цвета.*

Утверждение $A(1)$ говорит про одну прямую и очевидно (она делит плоскость на две полуплоскости, которые можно покрасить в разные цвета). Утверждение $A(2)$ тоже несложное: две прямые либо пересекаются (и можно покрасить части крест-накрест), либо параллельны (и тогда можно часть между ними покрасить в один цвет, а две полуплоскости — в другой). Но дальше ситуация усложняется, вариантов много, и нужно какое-то общее рассуждение.

Рассуждая по индукции, будем предполагать, что $A(n)$ уже доказано, и доказывать $A(n+1)$. Пусть есть какая-то конфигурация из $n+1$ прямых. Выбросим временно из неё одну прямую. Получится конфигурация из n прямых. По предположению её можно раскрасить требуемым образом. Теперь восстановим выброшенную прямую и поменяем все цвета с одной стороны от неё на противоположные. Докажем, что получилась правильная раскраска — и тем самым $A(n+1)$ тоже верно. В самом деле, рассмотрим какой-то граничный отрезок в этой новой конфигурации. Есть два случая:

- Это отрезок добавленной прямой (которую мы выбрасывали). Тогда до её добавления по обеим сторонам был один цвет, а после добавления мы с одной стороны перекрасили.

- Это отрезок одной из старых прямых. Тогда разделяемые им области лежат по одну сторону от новой прямой. Значит, они либо обе сохранили цвет (если перекрашивали по другую сторону), либо обе перекрашены. В любом случае они как были разного цвета до перекраски, так и остались.

Таким образом, мы доказали $A(n+1)$, предполагая $A(n)$ верным (вывели $A(n+1)$ из $A(n)$, доказали утверждение «$A(n)$ влечёт $A(n+1)$» — всё это синонимы). Мы знаем также, что $A(1)$ и $A(2)$ верны. Тогда можно вывести $A(3)$ из $A(2)$, установив его истинность, потом вывести $A(4)$ из $A(3)$ и так далее. Применяя принцип математической индукции, мы заключаем, что все утверждения $A(n)$ верны, что и требовалось доказать.

Что в этом доказательстве лишнее? Можно было не разбирать случай $A(2)$: мы всё равно потом выводим $A(n+1)$ из $A(n)$, и это рассуждение применимо при $n = 1$. В принципе можно было бы даже начинать с нуля (считая, что ноль прямых делит плоскость на одну область, которую мы можем покрасить в любой цвет), хоть это и звучит странно — доказательство $A(n) \Rightarrow A(n+1)$ имеет смысл и при $n = 0$.

Совсем придирчивый экзаменатор потребовал бы определить математически, что такое область, участок границы и т. п. Но это уже нечестно — поскольку эти понятия употребляются в условии задачи без уточнений, то и от решающих задачу этого требовать нельзя.

Мы старались изложить индуктивное рассуждение максимально подробно и формально. Но можно было бы, напротив, постараться его замаскировать и рассказать такое решение: представим себе полуплоскости из прозрачной плёнки, границами которых являются наши прямые (полуплоскость можно класть с любой стороны от прямой). Теперь покрасим в один цвет те точки, где чётное число слоёв, а в другой

— где нечётное. Пересекая границу, мы теряем (или приобретаем) один слой, так что чётный цвет сменяется нечётным.

1.2 Формулировка принципа математической индукции

В общем виде использованная нами схема рассуждения выглядит так.

> **Принцип математической индукции.** *Если некоторое утверждение $A(n)$, зависящее от натурального параметра n, верно для $n = 1$, и для любого n верно утверждение «если $A(n)$ верно, то и $A(n + 1)$ верно», то утверждение $A(n)$ верно при любом $n \geq 1$.*

Утверждение $A(1)$ называют *базисом* (или *базой*) *индукции*), а условное утверждение «если $A(n)$ верно, то и $A(n + 1)$ верно» — *индуктивным переходом* (или *шагом индукции*). В этом условном утверждении *посылка* $A(n)$ называется *индуктивным предположением*. Иногда — так будет в нашем первом примере — цепочка утверждений начинается не с $n = 1$, а с другого числа. В нашем случае это будет число 2. Тогда базисом индукции будет $A(2)$, а шаг индукции $A(n) \Rightarrow A(n+1)$ нужно проводить при $n \geq 2$.

Чтобы привыкнуть к терминологии, разберём ещё несколько задач — сначала рассуждая «на пальцах», а потом более формально.

Задача 1.2. В последовательности чисел $1, 1, 2, 3, 5, 8, \ldots$ два первых числа равны единице, а каждое следующее равно сумме двух предыдущих. Докажите, что в этой последовательности нет двух подряд стоящих чётных членов.

Решение этой задачи можно объяснить, не упоминая индукцию. Будем рассуждать «от противного»: пусть где-то два чётных числа a, b стоят рядом (b после a). Тогда перед ними стоит число $c = b - a$, которое тоже чётно (как разница двух чётных чисел: $2k - 2l = 2(k - l)$). Но тогда и перед c будет чётное число $a - c$ и так далее — двигаясь к началу, мы постепенно обнаружим, что все числа чётны, а это ведь не так. Полученное противоречие показывает, что двух чётных чисел подряд быть не может.

Теперь изложим по существу то же решение более формально. Для начала обозначим наши числа $F_1, F_2, \ldots$ и запишем их определение:

$$F_1 = 1; \quad F_2 = 1; \quad F_{n+1} = F_{n-1} + F_n \quad \text{при } n \geq 1$$

(Эти числа известны как *числа Фибоначчи*, поэтому мы обозначаем их буквой F.)

Утверждение $A(n)$ при $n \geq 2$ говорит, что *числа F_{n-1} и F_n одновременно не могут быть чётными*. Другими словами, оно говорит, что *хотя бы одно из чисел F_{n-1} и F_n нечётно* (могут и оба).

Базис индукции $A(2)$: оба числа F_1 и F_2 равны 1 и нечётны.

Шаг индукции. Пусть мы уже знаем $A(n)$: одно из чисел F_{n-1} и F_n нечётно. Докажем, исходя из этого, что верно $A(n + 1)$: одно из чисел F_n и F_{n+1} нечётно. Возможны два случая.

- Число F_n нечётно. Тогда доказывать нечего: мы уже знаем, что по крайней мере одно из чисел F_n и F_{n+1} (а именно, первое) нечётно.

- Число F_n чётно. По индуктивному предположению одно из чисел F_{n-1} и F_n нечётно, и это может быть только F_{n-1}. Тогда число $F_{n+1} = F_{n-1} + F_n$ будет суммой нечётного и чётного чисел, а такая сумма всегда нечётна. (Нечётное число имеет вид $2k+1$, складывая его с чётным числом $2l$, получим нечётное число $2k+1+2l = 2(k+l)+1$.)

Рассуждение по индукции закончено. (Понятно ли, что это то же самое рассуждение?)

Раз уж зашла речь про числа Фибоначчи, разберём ещё один пример рассуждения по индукции с их участием и оценим, насколько быстро они растут.

Задача 1.3. Докажите, что $F_n \le 1{,}7^n$ при всех $n = 1, 2, \ldots$.

При $n = 1$ и при $n = 2$ это очевидно: оба числа F_1 и F_2 равны 1. Дальше мы рассуждаем так: если $F_{n-1} \le 1{,}7^{n-1}$ и $F_n \le 1{,}7^n$, то

$$F_{n+1} = F_{n-1} + F_n \le 1{,}7^{n-1} + 1{,}7^n = 1{,}7^{n-1} \cdot (1 + 1{,}7) \le \\ \le 1{,}7^{n-1} \cdot 2{,}89 = 1{,}7^{n-1} \cdot 1{,}7^2 = 1{,}7^{n+1}.$$

Поэтому (взяв $n = 2$ в этом вычислении) получим, что и $F_3 \le 1{,}7^3$, затем (взяв $n = 3$) получим, что $F_4 \le 1{,}7^4$, и так далее.

Разница с предыдущими рассуждениями в том, что мы использовали два неравенства (для $n-1$ и n), чтобы получить неравенство для $n+1$. В архитектуре это бы соответствовало башне, в который каждый следующий этаж опирается на два предыдущих. Ничего страшного в таком рассуждении нет — мы могли бы даже использовать все утверждения $A(k)$ для меньших значений k сразу.[1]

Выбор числа 1,7 в этой задаче выглядит довольно загадочно — но всё, что нам от него нужно, это неравенство $1 + 1{,}7 \le 1{,}7^2$ ($2{,}7 \le 2{,}89$). Скажем, число 1,5 бы не подошло, потому что $1+1{,}5 = 2{,}5$, а $1{,}5^2 = 2{,}25$. И действительно, для 1,5 аналогичное неравенство нарушается, хотя и не сразу — можно проверить, что $1{,}5^{11} \approx 86{,}49\ldots$ меньше $F_{11} = 89$, а $1,5^{12} \approx 129{,}74\ldots$ меньше $F_{12} = 144$.

Вообще, при больших n последовательность F_n растёт примерно как геометрическая прогрессия со знаменателем $(1+\sqrt{5})/2 \approx 1{,}61\ldots$; это число называют иногда «золотым сечением», поскольку при отрезании квадрата от прямоугольника таких пропорций остаётся прямоугольник, подобный исходному.[2]

Вернёмся к примерам индуктивных рассуждений.

---

[1] Логики бы сказали здесь, что мы применяем обычный принцип индукции к утверждению $B(n) = [A(n-1)$ и $A(n)]$ и утверждению $B(n) = [A(k)$ при всех $k \le n]$ соответственно.

[2] Оценить рост F_n и доказать сформулированное утверждение можно с помощью линейной алгебры, найдя собственные значения матрицы $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ (преобразующей F_{n-1} и F_n в F_n и F_{n+1}) и установив, что большее из них равно золотому сечению, а меньшее меньше единицы.

Задача 1.4. Докажите, что число $111\ldots111$, в десятичной записи которого 3^n единиц, делится нацело на 3^n.

Например, 111 делится на 3 (и частное равно 37), а 111 111 111 делится на 9: частное тут вычислить сложнее, но можно воспользоваться признаком делимости на 9: сумма цифр равна 9 и делится на 9, значит, и само число делится на 9. (Кто сомневается, может разделить уголком или на калькуляторе.) Следующее утверждение: число из 27 единиц делится на 27. По аналогии хочется сказать, что у него сумма цифр 27 и делится на 27, и поэтому число само делится на 27 по признаку делимости на 27. Вот только признака такого нет (причём в обе стороны: у числа 27 сумма цифр не делится на 27, хотя само оно делится, а бывает и наоборот — можете привести пример?), так что нужно какое-то другое рассуждение.

Давайте попробуем разделить $111\ldots111$ (27 единиц) на 111111111 (9 единиц). Представим себе деление уголком: каждый раз в частном мы пишем единицу, а потом восемь нулей, всего получается $100\ldots0100\ldots1$, где между единицами стоят группы из 8 нулей. (Неопытные люди иногда забывают про эти нули и получают в частном 111, но это, конечно, абсурдно мало.) Если это вызывает сомнения, можно разделить уголком число из 9 единиц на число из 3 единиц, и получить 1001001. Результат можно проверить умножением столбиком.

Теперь внимание: полученное нами частное состоит из трёх единиц, разделённых нулями, поэтому его сумма цифр равна 3 и оно делится на 3 (такой признак делимости есть!), поэтому число из 27 единиц равно произведению числа, делящегося на 9 (состоящего из 9 единиц) и числа, делящегося на 3 (состоящего из трёх единиц и нулей между ними). А такое произведение делится на 27 (деление на 27 сводится к делению первого сомножителя на 9 и второго сомножителя на 3).

Мы доказали, что число из 27 единиц делится на 27. Теперь число из 81 единиц можно разделить на него и получится число с тремя единицами (и многими нулями между ними), которое делится на 3. А если один сомножитель делится нацело на 27, а второй на 3, то произведение делится на $27 \cdot 3 = 81$. И так далее.

Теперь запишем то же самое решение более формально. Докажем индукцией по n утверждение $A(n)$: *число из* 3^n *единиц делится на* 3^n. При $n = 1$ это верно, так как $111 = 3 \cdot 37$. Рассуждая по индукции, предположим, что верно $A(n)$, то есть число $x = 111\ldots111$ из 3^n единиц равно $3^n \cdot k$ при целом k. Число $10^{3^n}x$ получается дописыванием к x нулей в количестве 3^n штук, а число $10^{2\cdot 3^n}$ получается дописыванием ещё 3^n нулей. Сложив эти три числа столбиком, получим число $(10^{2\cdot 3^n} + 10^{3^n} + 1)x$ из 3^{n+1} единиц. Множитель $(10^{2\cdot 3^n} + 10^{3^n} + 1)$ записывается тремя единицами, разделёнными нулями, поэтому делится на 3 по признаку делимости на 3. Записав его как $3l$, получим, что число из 3^{n+1} единиц равно $3^n k \cdot 3l = 3^{n+1}kl$, то есть делится на 3^{n+1}. Таким образом, индуктивный переход $A(n) \Rightarrow A(n+1)$ выполнен, и остаётся воспользоваться принципом математической индукции.

Ещё раз посмотрим на общую схему доказательства по индукции. Мы доказываем, что $A(n)$ верно при всех натуральных n, начиная с некоторого начального значения. В наших примерах мы использовали начальные значения 1 и 2; часто в качестве начального значения используется нуль.[3] Доказательство состоит из двух частей: базиса индукции (мы доказываем утверждение для начального значения параметра) и шага индукции (мы доказываем следующее утверждение $A(n+1)$, используя индуктивное предположение $A(n)$).

Обе эти части существенны, как видно из простых примеров. Скажем, если кто-то решил доказать по индукции, что квадрат любого натурального числа n больше 10, то с шагом индукции проблем не будет: если $n^2 > 10$, то $(n+1)^2 = n^2 + 2n + 1 > n^2 > 10$. Но вот с базисом индукции, конечно, ничего не выйдет.

А вот пример, где ошибка чуть лучше замаскирована. «Докажем», что все натуральные числа равны друг другу. Для этого докажем по индукции утверждение $A(n)$: «в любом наборе из n натуральных чисел все числа равны».

Базис индукции $A(1)$ — тривиальное истинное утверждение «каждое число равно самому себе». Проведём индуктивный переход. Пусть $A(n)$ справедливо, докажем $A(n+1)$. Рассмотрим набор из $(n+1)$ чисел

$$(a_1,\ a_2,\ \ldots, a_n,\ a_{n+1}).$$

Применим утверждение $A(n)$ к наборам

$$(a_1,\ a_2,\ \ldots, a_n) \quad \text{и} \quad (a_2,\ \ldots, a_n,\ a_{n+1})$$

из n чисел. Индуктивное предположение гарантирует, что числа и в одном, и в другом наборе равны. Получаем цепочки равенств

$$\begin{gathered} a_1 = a_2 = \ldots = a_n \quad \text{и} \\ a_2 = a_3 = \ldots = a_{n+1}. \end{gathered}$$

Отсюда уже очевидно, что

$$a_1 = a_2 = \ldots = a_n = a_{n+1},$$

то есть утверждение $A(n+1)$ верно. Применяя принцип математической индукции, получаем, что $A(n)$ верно для всех n.

Видите ошибку в этом рассуждении? Как часто бывает, она ровно в том месте, где написано «очевидно».[4]

Для $n = 1$ цепочки равенств вырождаются: первая имеет вид $a_1 = a_1$, а вторая — $a_2 = a_2$. Из этих двух тривиальных равенств, конечно же, не следует, что $a_1 = a_2$.

Для $n > 1$ ситуация другая: в первой цепочке есть равенство $a_1 = a_2$, а во второй следует, что $a_2 = \ldots = a_{n+1}$. Из этих двух равенств следует, что $a_1 = a_2 = \ldots = a_{n+1}$.

Итак, ошибка в доказательстве индуктивного перехода возникает только в случае $n = 1$. Но само утверждение $A(n)$ неверно для всех $n > 1$.

---

[3] Мы будем считать нуль натуральным числом: удобно считать, что число жирафов в зоопарке натуральное, не зная заранее, есть они там или нет.

[4] Про разных людей рассказывают, что в тексте их выступления была пометка «повысить голос — аргумент слаб». Слово «очевидно» иногда играет ту же роль в математических текстах.

В этом примере, конечно, сразу было ясно, что мы доказываем неверное утверждение (и легко понять, начиная с какого n оно неверно). Но подобная ошибка может случиться и в индуктивном доказательстве более сложного утверждения, и тогда её можно проглядеть — а между тем единственный разрыв в индуктивной цепи полностью лишает силы доказательство по индукции.

В следующих разделах мы рассмотрим несколько типичных примеров индуктивных рассуждений.

1.3 Доказательство формул суммирования по индукции

Индукция часто используется для доказательств тождеств с суммами. Давайте посмотрим, как это делается (а заодно и познакомимся с обозначениями для сумм).

Вот формулы для сумм двух прогрессий: арифметической и геометрической.

$$1 + 2 + \ldots + n = \sum_{k=1}^{n} k = \frac{n(n+1)}{2}; \tag{1.1}$$

$$1 + 2^1 + 2^2 + \ldots + 2^n = \sum_{k=0}^{n} 2^k = 2^{n+1} - 1. \tag{1.2}$$

Сначала сумма записана с многоточием (подразумевающим все целые числа или все степени двойки в соответствующих промежутках), а потом с помощью стандартных обозначений. В выражении $\sum_{k=1}^{n}$ буква k называется «переменной суммирования». Имеется в виду, что в стоящее за этим знаком выражение нужно поочерёдно подставить все значения $k = 1, 2, 3, \ldots, n$, и получившиеся числа сложить.

В этих утверждениях уже есть параметр n, который и будет параметром индукции. Докажем сначала первую формулу, в которой утверждение $A(n)$ говорит, что

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2}.$$

Базис индукции $A(1)$ проверить легко: $1 = 1 \cdot 2/2$. Тут, конечно, надо правильно понимать обозначение $1 + 2 + \ldots + n$: при $n = 1$ в этой сумме (вопреки её внешнему виду) нет слагаемого 2, а слагаемое 1 встречается только один раз.

Шаг индукции. Тут мы предполагаем, что утверждение $A(n)$ верно, то есть что

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2}.$$

Надо доказать $A(n+1)$, то есть

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

Для этого разделим эту сумму на $(1 + 2 + \ldots + n)$, которая по предположению индукции равна $n(n+1)/2$, и последний член $n+1$. Всего получится

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

что и требовалось (разница только в порядке сомножителей). Шаг индукции (и всё доказательство по индукции) на этом завершаются.

Аналогичное рассуждение для второй прогрессии (геометрической со знаменателем 2) запишем более кратко и с использованием обозначений для сумм. Базис индукции: для $n = 1$ получается верное равенство $1 = 2^{0+1} - 1$. Шаг индукции: если

$$\sum_{k=0}^{n} 2^k = 2^{n+1} - 1,$$

то

$$\sum_{k=0}^{n+1} 2^k = \sum_{k=0}^{n} 2^k + 2^{n+1} = 2^{n+1} + 2^{n+1} - 1 = 2^{n+2} - 1.$$

Конечно, у этих формул есть и другие доказательства. Например, первую сумму (обозначим её S) можно удвоить, перевернув:

$$2S = S + S = [1 + 2 + \ldots + n] + [n + (n-1) + \ldots + 1].$$

Теперь, группируя члены в пары, получаем n пар $1+n$, $2+(n-1)$,..., $n+1$, каждая из которых даёт $n+1$, так что $2S = n(n+1)$, как и говорит наша формула.

Многие другие формулы для сумм тоже легко доказываются с помощью индукции — правда, надо как-то догадаться заранее, что именно доказывать. Скажем, несложно доказать по индукции формулу для сумм квадратов

$$1 + 4 + 9 + 16 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Но с чего бы нам могло прийти в голову, что нужно доказывать именно эту формулу?

В данном случае может помочь такой общий факт: *для любого d сумма $1^d + 2^d + \cdots + n^d$ является многочленом от n степени $d+1$* (раздел 2.9, с. 66). Если мы это знаем, то остаётся лишь подобрать коэффициенты у многочлена (для сумм квадратов это многочлен третьей степени, у него четыре коэффициента, и можно составить систему уравнений на них). При этом знать доказательство этого факта не обязательно: после того как формула доказана по индукции, мы не обязаны рассказывать, как мы до неё догадались.

1.4 Доказательство неравенств

Теорема 1.1 (неравенство Бернулли)**.** *Для всех $n \geqslant 1$ верно неравенство $(1+h)^n \geqslant 1 + hn$, при $1 + h \geqslant 0$.*

Это неравенство говорит, что получая в течении 100 лет по одному проценту по вкладу, мы как минимум удвоим вклад. Собственно говоря, это почти что доказательство неравенства, по крайней мере для положительного h, но мы изложим его более формально со ссылкой на индукцию, и заодно убедимся, что условия $1+h \geq 0$, то есть $h \geq -1$, нам достаточно.

Доказательство. База индукции ($n=1$): $1+h \geqslant 1+h$.

Индуктивный переход: пусть $(1+h)^n \geqslant 1+hn$ (предположение индукции). Умножим обе части этого неравенства на неотрицательное число $(1+h)$. Получится

$$(1+h)(1+h)^n \geqslant (1+h)(1+hn) = 1+(n+1)h+h^2n \geqslant 1+(n+1)h,$$

что и требовалось (в последнем переходе мы отбросили число $h^2n \geq 0$). □

Более сложный пример — неравенство

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2} < 2.$$

(при любом натуральном $n \geq 1$). С основанием индукции проблем нет, но индуктивный шаг выполнить напрямую не удается (попробуйте!).

Здесь оказывается полезной такая идея: *доказывать по индукции более сильное утверждение.* На первый взгляд неясно, как это может помочь, ведь доказывать более сильное утверждение сложнее. Но дело тут в том, что в индуктивном шаге станет не только рассуждение, но и посылка, и это усиление можно будет использовать. (Как при строительстве дома: если делать стены толще, то они будут тяжелее — но и стены нижних этажей, на которые падёт эта тяжесть, будут прочнее!)

Как и раньше, основная трудность тут — догадаться, какое более сильное утверждение надо доказывать, но после того, как мы догадались (или кто-то нам сказал), дальше уже легко. В этом примере таким более сильным утверждением будет

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2} \leq 2-\frac{1}{n}.$$

Базис индукции ($n=1$): $1 \leq 2-\frac{1}{1}$ (здесь неравенство обращается в равенство).

Шаг индукции: если мы уже знаем, что

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2} \leq 2-\frac{1}{n},$$

то можно прибавить к обеим частям неравенства $1/(n+1)^2$. Получится

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}+\frac{1}{(n+1)^2} \leq 2-\frac{1}{n}+\frac{1}{(n+1)^2},$$

и достаточно доказать, что

$$-\frac{1}{n}+\frac{1}{(n+1)^2} \leq -\frac{1}{n+1}$$

(мы сократили член 2 с обеих сторон). Это неравенство можно переписать как

$$\frac{1}{n} - \frac{1}{n+1} \geq \frac{1}{(n+1)^2},$$

а это уже понятно, так как в левой части получается $1/n(n+1)$, и у этой дроби знаменатель меньше $(n+1)^2$.

На этом доказательство завершается. Мы привели его скорее в качестве примера индуктивного рассуждения, чем ради результата: на самом деле примерно то же рассуждение можно изложить проще и короче, написав

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{n^2} \leq 1 + \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \ldots + \frac{1}{(n-1)n} =$$
$$= 1 + \bigl(1 - \frac{1}{2}\bigr) + \bigl(\frac{1}{2} - \frac{1}{3}\bigr) + \ldots + \bigl(\frac{1}{n-1} - \frac{1}{n}\bigr) = 2 - \frac{1}{n}$$

(все остальные члены сокращаются).

1.5 Ненулевые решения однородной линейной системы

Нашим следующим примером будет факт, лежащий в основе всей линейной алгебры (с его помощью определяется центральное понятие «размерности»). Но сначала напомним определения. *Однородным линейным уравнением* называется уравнение вида

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0,$$

где x_i — переменные (неизвестные), а a_i — числовые коэффициенты (произвольные действительные числа). Слово «однородный» означает здесь (хоть это и странно с точки зрения русского языка), что в правой части уравнения стоит нуль.

Система однородных линейных уравнений состоит из нескольких таких уравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0. \end{cases}$$

Как принято в линейной алгебре, мы обозначаем числовые коэффициенты буквами с двумя индексами. В линейной алгебре их записывают в таблицу, которую называют матрицей этой системы уравнений. В данном случае у нас m уравнений и n переменных, так что матрица получается ширины n и высоты m.

Как обычно, *решением* называется такой набор числовых значений переменных $(x_1, \ldots, x_n)$, для которого все уравнения системы обращаются в верные равенства.

У системы однородных уравнений решение есть всегда: достаточно положить

$$x_1 = x_2 = \cdots = x_n = 0.$$

Поэтому интересен вопрос о существовании *ненулевого* решения системы однородных линейных уравнений, где не все переменные обращаются в нуль (хотя некоторые могут). Следующая теорема (тот самый факт, о котором мы говорили) даёт достаточное условие существования такого решения.

Теорема 1.2. *Если количество уравнений m в однородной системе линейных уравнений меньше количества переменных n, то система имеет ненулевое решение.*

Это не так удивительно: чем меньше уравнений, тем меньше требований, и чем больше переменных, тем больше свобода выбора при поиске решений. Но эти разговоры не заменяют формального доказательства; его мы будем проводить индукцией *по числу уравнений.*

Базис индукции: случай одного уравнения (с $n > 1$ переменными: число переменных ведь больше числа уравнений). Тут есть два варианта.

- Все коэффициенты уравнения нулевые:

  $$0 \cdot x_1 + 0 \cdot x_2 + \dots + 0 \cdot x_n = 0.$$

  Любой набор значений переменных является решением такого уравнения, и можно выбрать любой ненулевой набор.

- В уравнении

  $$a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$$

  есть хотя бы один ненулевой коэффициент. Пусть, скажем, $a_1 \neq 0$ (в противном случае перенумеруем переменные).

  Выберем какие-нибудь ненулевые значения для переменных $x_2, \dots, x_n$, скажем,

  $$x_2 = \ldots = x_n = 1.$$

  Значение x_1 выразим из уравнения:

  $$x_1 = \frac{1}{a_1}\left(-a_2 - a_3 - \dots - a_n\right).$$

  Получили ненулевой набор значений переменных, для которых уравнение выполняется (из этого расчёта мы выбирали x_1).

Контрольный вопрос 1.1. Где в этом рассуждении использовано условие на количество переменных?

Шаг индукции. Нам надо доказать, что если теорема выполняется для систем из m уравнений с $n > m$ переменными, то она выполняется и для систем из $m + 1$ уравнений с $n > m + 1$ переменными.

В одной фразе можно сказать так: выберем какой-то ненулевой коэффициент при какой-то переменной в каком-то уравнении, выразим эту переменную из этого уравнения и подставим это выражение в остальные, уменьшив на единицу число уравнений и число переменных. Объясним подробнее, что это означает.

В системе из $m+1$ уравнений с $n > m+1$ переменными возьмём первое уравнение

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0.$$

Если все его коэффициенты нулевые, то любой набор значений переменных подходит, и это уравнение можно удалить из системы без изменения её смысла. Получится система из m уравнений с $n > m + 1 > m$ переменными, и у неё есть ненулевое решение в силу индуктивного предположения.

Второй случай: не все коэффициенты первого уравнения нулевыые. Пусть, скажем, $a_{11} \neq 0$ (иначе перенумеруем переменные). Тогда первое уравнение равносильно уравнению

$$x_1 = -\frac{1}{a_{11}}\left(a_{12}x_2 + a_{13}x_2 + \ldots + a_{1n}x_n\right). \tag{1.3}$$

Из этого уравнения значение переменной x_1 однозначно выражается через значения остальных переменных. Это выражение теперь можно подставить вместо x_1 остальные уравнения, получив равносильную систему. Для порядка выпишем, что получится после этого из уравнения с каким-то номером $i > 1$:

$$\begin{aligned}0 = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n &= \\ = -\frac{a_{i1}}{a_{11}}\left(a_{12}x_2 + a_{13}x_2 + \ldots + a_{1n}x_n\right) &+ a_{i2}x_2 + \ldots + a_{in}x_n = \\ = \left(a_{i2} - \frac{a_{i1}}{a_{11}}a_{12}\right)x_2 + \ldots + \left(a_{in} - \frac{a_{i1}}{a_{11}}a_{1n}\right)x_n &= a'_{i2}x_2 + \ldots + a'_{in}x_n.\end{aligned}$$

(буквами со штрихами мы обозначили новые коэффициенты). У системы уравнений с m уравнениями и $n - 1 > m$ переменными

$$\begin{cases} a'_{22}x_2 + \dots + a'_{2n}x_n = 0, \\ a'_{32}x_2 + \dots + a'_{3n}x_n = 0, \\ \dots \\ a'_{(m+1)2}x_2 + \dots + a'_{(m+1)n}x_n = 0. \end{cases}$$

по предположению индукции есть ненулевое решение. Добавим к нему значение переменной x_1, вычисленное из уравнения (1.3). Получим ненулевое решение исходной системы.

Теорема 1.2 доказана.

1.6 Коды Грея

Алфавитом размера q называется любой набор из q различных символов. *Слово длины n* в алфавите из q символов — это последовательность n символов из этого алфавита.

Пример 1.5 (Двоичные слова)**.** Пусть в алфавите всего два символа — 0 и 1. Такой алфавит называется *двоичным* и слова в таком алфавите называются двоичными.[5] Тогда есть два слова длины 1 (0 и 1). Двоичных слов длины 2 четыре:

$$00,\ 01,\ 10,\ 11.$$

Пример 1.6. Пусть алфавит состоит из 33 русских букв от А до Я. Словами длины 3 в таком алфавите будут, например,

СЫР, МАЙ, ЪЪФ.

Последний пример показывает, что словами мы считаем любые последовательности букв — не обязательно осмысленные (даже если буквы берутся из русского алфавита).

В примере 1.5 перечислены все двоичные слова длины 2. Они перечислены здесь в порядке возрастания чисел в двоичной системе. Заметим, что при переходе от второго слова к третьему изменяются все символы (все биты). Это иногда неудобно — пусть, скажем, мы хотим перепробовать все положения n выключателей по одному разу, переключая каждый раз по одному выключателю. Положения выключателей задаются двоичными словами, а условие означает, что соседние слова должны отличаться ровно в одном бите.

Оказывается, что такой порядок действительно можно найти (он называется *кодом Грея*). Это несложно сделать в конкретных примерах, скажем, двоичные слова длины 2 можно перечислить в таком порядке:

$$00,\ 01,\ 11,\ 10.$$

(по сравнению с предыдущим порядом мы переставили местами два последних слова в предыдущем списке). Но как доказать, что это возможно всегда? Снова нам поможет рассуждение по индукции, которое работает для любого алфавита, не только двухбуквенного.

Теорема 1.3. *Можно так упорядочить все слова длины n в алфавите из q символов, что любые два соседних слова различаются только в одной позиции.*

Доказательство проведём по индукции.

База индукции — слова длины 1, то есть просто одиночные символы. Тут вообще никаких ограничений нет, их можно перечислять в любом порядке (каждый раз меняется один символ — единственный возможный).

Шаг индукции. Пусть задача для слов длины n уже решена. Можно представлять себе дело так: есть человек, который пишет на доске слово длины n, потом меняет в нём один символ, потом ещё один, и так далее, пока по очереди не будут выписаны все слова длины n.

[5] Программисты назвали бы их *битовыми строками* — они вообще называют слова строками.

Придадим ему ассистента. После того, как человек напишет какое-то слово X, его ассистент по очереди дописывает в конец этого слова все буквы алфавита. Если, скажем, алфавит состоит из букв a, b, c, d, то сначала появится слово Xa, потом ассистент сотрёт a и напишет b, появится Xb, потом Xc, потом Xd. Когда алфавит заканчивается, снова приходит гражданин начальник и, не обращая внимания на действия ассистента, заменяет слово X на следующее слово Y в своём списке. На доске будет Yd, и после этого помощник на место d по очереди пишет буквы a, b, c (в любом порядке, это не важно). Затем снова приходит начальник, который заменяет слово Y на следующее слово Z, и так далее.

Другими словами, вся последовательность разбивается на участки из q символов. Каждый из этих участков устроен так: меняется только последняя буква (по разу встречается каждый символ). Между участками последняя буква не меняется, а первые $n-1$ букв меняются согласно предположению индукции.

Остаётся заметить две вещи. Во-первых, в итоге будут выписаны все слова из $n-1$ букв в комбинации со всеми последними буквами, то есть все слова из n букв (разбитые на группы, в которых меняется только последняя буква). Каждое слово встречается по одному разу.

Во-вторых, на каждом шаге изменения делает либо начальник, либо ассистент. По предположению индукции начальник меняет только одну букву, а у ассистента вообще нет иной возможности, так как он меняет только последнюю букву.

Пример. Для алфавита из четырёх букв при $n = 2$ получаем такую последовательность:

$$aa, ab, ac, ad,\ \ bd, bc, bb, ba,\ \ ca, cb, cc, cd,\ \ dd, dc, db, ad$$

(ассистент меняет буквы от a до d и обратно, но в принципе порядок букв у него может быть любым, не обязательно таким).

Для $n = 3$ к каждому из этих слов приписываются по очереди четыре варианта последней буквы:

$$aaa, aab, aac, aad,\ \ abd, abc, abb, aba,\ \ aca, acb, acc, acd,\ \ add, adc, adb, ada,$$
$$bda, bda, bdc, bdd,\ \ bcd, bcc, bcb, bca, bba, bbb, bbc, bbd, \ldots$$

1.7 Ханойская башня

«Ханойская башня» — это старинная головоломка.[6] Есть три штырька и n дисков разного размера, которые на них можно надевать. Изначально все диски на одном штырьке в порядке убывания размера (рис. 1.3). За один ход разрешается снимать верхний диск на любом штырьке и перекладывать его на другой штырёк, только нельзя класть бóльший диск на меньший (рис. 1.4). Требуется перенести всю стопку дисков с одного штырька на другой. Возможно ли это?

[6] Опубликована — вместе с рассказом про браминов, перекладывающих диски в ожидании конца света — в книге французского преподавателя математики Эдуарда Люка в 1892 году.

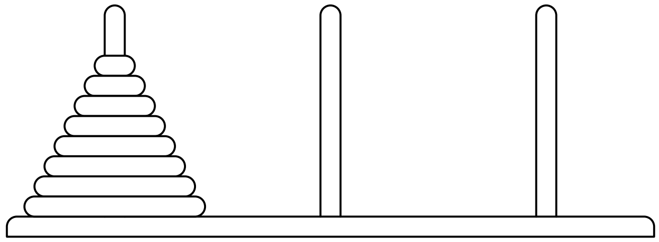


Рис. 1.3: В оригинальном варианте было 8 дисков

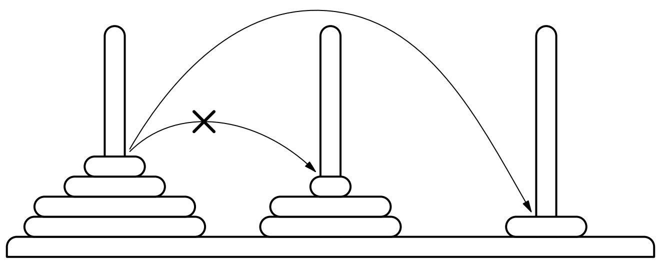


Рис. 1.4: Разрешённый ход и запрещённый ход (перечёркнут)

Оказывается, что да, и это можно доказать по индукции. Посмотрим для начала, что будет при небольших значениях n. Если $n = 1$, то есть диск всего один, никаких препятствий к переносу нет. При $n = 2$ мы сначала перекладываем меньший диск на свободный штырёк (а что ещё делать?) — затем можно переложить бóльший диск на третий штырёк и, наконец, положить меньший диск на бóльший (всего 3 хода).

Попробуйте отложить книгу и решить задачу для случая 3 дисков. Общее решение после этого станет гораздо понятнее.

Решение в общем случае получается по индукции. Базис индукции уже установлен. Осталось сделать индуктивный переход. Предположим, что n дисков возможно перенести с одного штырька на другой. Как перенести $n+1$ диск? Это можно сделать в три стадии (рис. 1.5)

(I) Перенесём верхние n дисков с первого штырька на второй: это возможно по предположению индукции, так как самый большой диск в это время лежит на первом штырьке в самом низу и остальным не мешает, так что мы находимся в ситуации n дисков.

(II) Перенесём самый большой диск с первого штырька на третий.

(III) Перенесём n дисков со второго штырька на третий, снова воспользовавшись предположением индукции (снова самый большой диск никому не мешает).

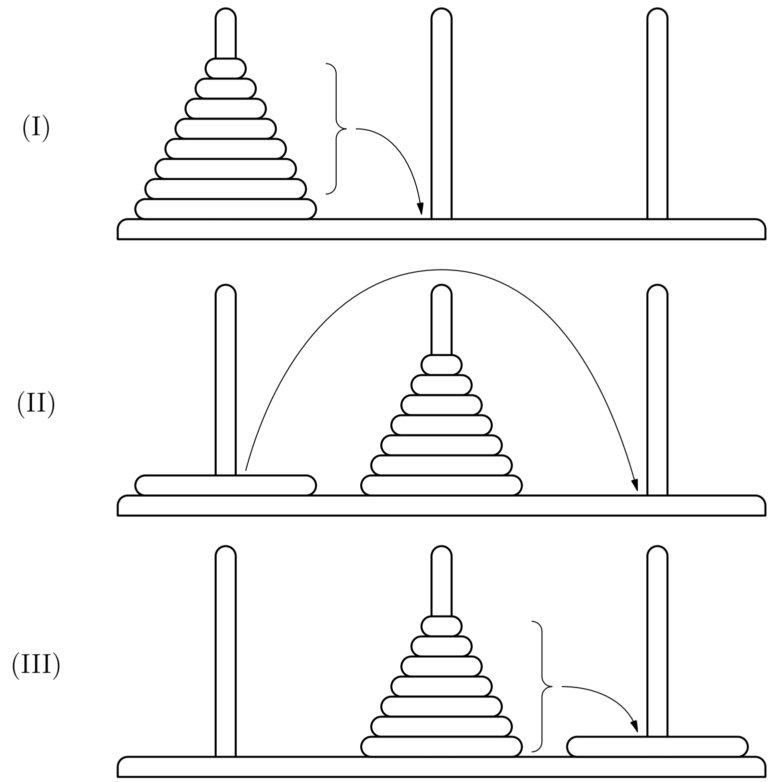


Рис. 1.5: Индуктивное решение головоломки «Ханойская башня»

После этого, как мы и хотели, все диски будут на третьем штырьке (самый большой там уже был).

Программисты записывают это решение в виде рекурсивной процедуры, которая дважды вызывает себя при меньших значениях параметра. (Сведение к меньшим значениям параметра, которое в математике называется индукцией, в программировании называют рекурсией.) Можно оценить количество ходов, которое в этом рекурсивном решении делается. Обозначив его за $T(n)$, заметим, что

$$T(1) = 1; \quad T(n+1) = 2T(n) + 1.$$

Такая запись традиционно называется *рекуррентным соотношением*, и из него легко получить явную формулу — снова воспользовавшись индукцией.

Задача 1.7. Докажите, что $T(n) = 2^n - 1$.

Решив эту задачу, вы оценили число действий описанного нами рекурсивного алгоритма. Естественный вопрос в таких ситуациях — а будет ли он оптимальным? В этой задаче да, и это несложно доказать.

Задача 1.8. Покажите, что *любой* способ перекладывания n дисков с одного штырька на другой требует не меньше $2^n - 1$ операций. (Указание: самый большой диск переложить так или иначе придётся, и в этот момент остальные диски должны быть собраны в пирамидку на другом штырьке.)

1.8 Теорема Холла о представителях

Мы старались выбрать разнообразные примеры рассуждений по индукции, но все они были достаточно простыми. Но в заключение надо всё-таки привести пример более сложного рассуждения, в качестве которого мы выбрали знаменитую «теорему Холла о представителях». Она играет важную роль в теории графов, но её можно переформулировать и ничего не говоря о графах, почти как олимпиадную задачу для школьников, и сейчас мы приведём утверждение и доказательство этой теоремы в таких терминах.

Пусть в школе имеется несколько кружков (на разные темы). Администрация хочет назначить в каждом кружке старосту из числа участников этого кружка. При этом нельзя, чтобы один и тот же школьник был старостой сразу в нескольких кружках. Ясно, что это не всегда возможно: скажем, если школьник Петя ходит в два кружка, а больше никто в эти кружки не ходит, то требования невыполнимы — он не может быть одновременно старостой в обоих, а других школьников там нет. Другой (ещё более вырожденный) случай — когда есть один кружок, в который вообще никто не ходит. Вообще, если имеется k кружков, в которые всего ходят меньше k школьников (если собрать все кружки в одном помещении, в нём будет меньше k человек), то задача неразрешима — кандидатов меньше, чем должностей. Оказывается, что это единственное препятствие.

Теорема 1.4 (Теорема Холла о представителях)**.** *Если для любых k кружков общее число школьников, которые ходят хотя бы в один из них, не меньше k, то назначение старост возможно.*

«Представители» — это более научный термин для старост. По-английски это утверждение называют Hall's marriage theorem (понятно, почему?) — а доказал её английский математик Филипп Холл в 1935 году.

Доказательство. Индукция по общему числу кружков. Если кружок один, то всё понятно (назначаем старостой любого его участника). Пусть кружков n, и для меньшего количества кружков утверждение теоремы верно. Попробуем волюнтаристский подход: выберем какой-то кружок, назначим там старосту произвольно — пусть это будет Лена, — и предложим остальным $s-1$ кружкам после этого самим разобраться со своими назначениями — разумеется, уже не назначая Лену старостой.

Если им это удастся, то всё хорошо. Если же не удастся, то мы знаем (индуктивное предположение) причину: есть некоторые s кружков, у которых вместе меньше s школьников (не считая Лены). А раньше? Раньше — с Леной — у них было *ровно*

s школьников (больше быть не могло, раз стало меньше — а меньше быть не могло по условию).

Итак, достаточно рассмотреть случай, когда *некоторые $s < n$ кружков в объединении включают ровно s школьников.* Назовём эти кружки «особыми»: в s особых кружков ходят ровно s школьников. Будем теперь решать задачу отдельно для особых и неособых кружков. И тех, и других меньше n, так что можно воспользоваться предположением индукции. Сначала сделаем это для особых — назначим им старост любым разрешённым способом (таковой существует по предположению индукции). После этого с неособыми кружками сложнее: мы не можем просто так воспользоваться предположением индукции, так как s участников особых кружков уже назначили старостами и их назначать нельзя. Надо сначала обосновать возможность применения индуктивного предположения, то есть надо убедиться, что в любые t неособых кружков ходят не меньше t школьников, не посещающих особые кружки (которых назначать нельзя). В самом деле, если бы их было меньше t, то вместе с s особыми кружками мы получили бы $s + t$ кружков, в которые ходит меньше $s + t$ школьников, что противоречит условию теоремы. Шаг индукции завершён. □

Это доказательство, хотя и не очень сложное, производит странное впечатление волшебства (или жульничества, если выражаться менее деликатно) — вроде мы ничего интересного не делаем, скорее переливаем из пустого в порожнее, а почему-то в итоге всё получается. Это бывает с индуктивными рассуждениями, и иногда можно придумать более наглядное (хотя, возможно, и более длинное) доказательство, которое лучше объясняет, «почему» теорема верна. В случае с теоремой Холла о представителях такое доказательство получается, если выводить эту теорему из общего утверждения о потоках в сетях (перевозке грузов по сети дорог с ограниченной пропускной способностью), но мы сейчас про это говорить не будем — мы хотели проиллюстрировать возможности индуктивных рассуждений.

1.9 Задачи для самостоятельного решения

9. Докажите, что для любого целого положительного n выполняется

a) $1 + 3 + 5 + \cdots + (2n-1) = n^2$;

b) $1 \cdot 2^1 + 2 \cdot 2^2 + 3 \cdot 2^3 + \cdots + n \cdot 2^n = (n-1) \cdot 2^{n+1} + 2$;

c) $1 + \dfrac{1}{2} + \dfrac{1}{3} + \cdots + \dfrac{1}{2^n} > \dfrac{n}{2} + 1$.

10. Докажите равенства

a) $1 \cdot (n-1) + 2 \cdot (n-2) + \cdots + (n-1) \cdot 1 = \dfrac{(n-1)n(n+1)}{6}$;

b) $\cos x + \cos 2x + \cdots + \cos nx = \dfrac{\sin(n+\frac{1}{2})x}{2\sin\frac{x}{2}} - \dfrac{1}{2}$.

11. Докажите неравенства

a) $\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n} > \dfrac{2}{3} \cdot n\sqrt{n}$;

b) $\dfrac{1}{2} \cdot \dfrac{3}{4} \cdot \dfrac{5}{6} \cdot \ldots \cdot \dfrac{2n-1}{2n} < \dfrac{1}{\sqrt{n}}$.

12. Докажите, что 1 можно представить в виде суммы 2014 различных обыкновенных дробей с числителем 1 и положительным знаменателем.

13. В зачете участвовало несколько студентов и преподавателей. Известно, что в комнату, где происходил зачет, каждый участник зачета вошел лишь однажды и что каждый преподаватель поговорил с каждым студентом. Докажите, что в какой-то момент зачета в комнате присутствовали либо все студенты (и, может быть, кто-то из преподавателей), либо все преподаватели (и, может быть, кто-то из студентов).

14. В прямоугольнике $3 \times n$ стоят фишки трех цветов, по n штук каждого цвета. Докажите, что можно переставить фишки в каждой строке так, чтобы в каждом столбце были фишки всех цветов.

15. Из целых чисел от 1 до $2n$ выбрано $n + 1$ число. Докажите, что среди выбранных чисел найдутся два, одно из которых делится на другое.

16. На доске написаны сто цифр — нули и единицы (в любой комбинации). Разрешается выполнять два действия:

1. заменять первую цифру (нуль на единицу и наоборот);
2. заменять цифру, стоящую после первой единицы.

Докажите, что после нескольких таких замен можно получить любую комбинацию из 100 нулей и единиц.

17. На краю пустыни имеется большой запас бензина и машина, которая при полной заправке может проехать 50 километров. Имеются (в неограниченном количестве) канистры, в которые можно сливать бензин из бензобака машины и оставлять на хранение (в любой точке пустыни). Доказать, что машина может проехать любое расстояние. (Канистры с бензином возить не разрешается, пустые можно возить в любом количестве.)

18. На кольцевой дороге стоит некоторое количество одинаковых автомобилей. Суммарное количество бензина в их бензобаках достаточно, чтобы один автомобиль мог совершить полный круг. Докажите, что найдется автомобиль, который, начав двигаться против часовой стрелки и забирая бензин по ходу движения у стоящих на дороге автомобилей, сможет совершить полный круг.

19. a) Докажите, что любой квадрат $2^n \times 2^n$, из которого вырезана угловая клетка, можно разрезать на уголки из трех клеток.

b) Докажите, что на уголки можно разрезать любой квадрат $2^n \times 2^n$, из которого вырезана любая (не обязательно угловая) клетка.

20*. Целые положительные числа a_1, a_2, …, a_n таковы, что $a_k \leqslant k$ и сумма всех этих чисел четна и равна $2S$. Докажите, что эти числа можно разбить на две группы, сумма по каждой из которых равна S.

21*. Лабиринтом называется клетчатый квадрат 10×10, некоторые пары соседних узлов в котором соединены отрезком — «стеной» — таким образом, что переходя из клетки в соседнюю по стороне клетку и не проходя через стены, можно посетить все клетки квадрата. Границу квадрата будем также считать обнесенной стеной. В некоторой клетке некоторого лабиринта стоит робот. Он понимает 4 команды — Л, П, В, Н, по которым соответственно идет влево, вправо, вверх и вниз, а если перед ним «стена», то стоит на месте. Как написать программу для робота, выполняя которую он обойдет все клетки независимо от лабиринта и от своего начального положения?

Лекция 2

Подсчёты

Сколько можно составить слов, переставляя буквы в слове КОТ? Нетрудно выписать все такие слова:

КОТ КТО ОКТ ОТК ТКО ТОК

(говоря о словах, мы не имеем в виду что-то осмысленное — просто цепочки букв). Конечно, возникает вопрос, почему мы ничего не пропустили — но это легко сообразить. Сначала написаны два варианта с буквой К на первом месте (две другие могут стоять в том или другом порядке), потом два с буквой О, потом с буквой Т, всего 6 вариантов.

Если бы слова были длиннее, то вручную их выписать было бы труднее, понадобилась бы компьютерная программа (не такая уж простая, кстати), а для совсем длинного слова и программа бы не помогла: вариантов было бы больше, чем можно перечислить в обозримое время. Но можно *подсчитать число вариантов, не перечисляя их все*, и в этой лекции мы будем как раз изучать разные способы такого подсчёта. Научное название для такого рода вопросов — «перечислительная комбинаторика», и это большой раздел (дискретной) математики с очень нетривиальными результатами и методами; мы разберём только самые простые. Начнём мы с совсем тривиальных вещей.

2.1 Правило суммы

Сколько существует четырёхзначных чисел? Это числа от 1000 (предыдущее — трёхзначное 999) до 9999 (дальше идёт пятизначное 10000). Несложно сообразить, что их 9000. В самом деле, из списка

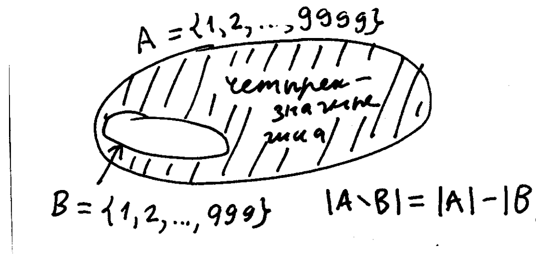
$$1, 2, 3, 4, 5, \ldots, 9999$$

в котором 9999 чисел (одно, два,..., 9999), мы должны вычеркнуть числа

$$1, 2, 3, 4, 5, \ldots, 999$$

(первые 999), остаётся $9999 - 999 = 9000$.

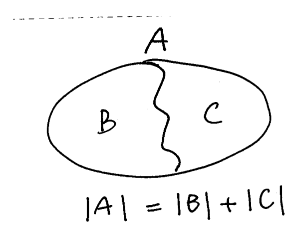
Это рассуждение — очень простой вариант более общего рассуждения, которое будет использоваться дальше очень часто. Мы рассмотрели *множество*[1] A, состоящее из чисел $1, 2, \ldots, 9999$, и его *подмножество* B, состоящее из чисел $1, 2, \ldots, 999$, и затем подсчитали, сколько чисел входит в их *разность* $A \setminus B$, множество всех четырёхзначных чисел. Это количество равно разности количеств чисел в A и в B, см. рисунок.



Мы обозначаем количество чисел в множестве X через $|X|$

> **Правило суммы.** *Если какое-то множество A разделено на две части B и C, не имеющие общих элементов, то $|A| = |B| + |C|$.*

Другими словами, если надо подсчитать количество объектов какого-то вида, и это объекты можно поделить на непересекающиеся типы, то общее количество объектов равно сумме количеств объектов каждого типа.[2]



[1] Подробно множества и операции с ними обсуждаются в лекции 5.

[2] Число способов добиться цели «не мытьём, так катаньем» равно сумме числа способов добиться её мытьём и числа способов добиться её катаньем — при условии, что нет пересечений и мытьё с катаньем не спутаешь.

Философы могут спросить: как доказать это утверждение? или это не теорема, а аксиома, и доказывать её не надо? Не углубляясь в основания математики, можно ответить им, что это и не теорема, и не аксиома, а *определение* сложения: когда в начальной школе учили складывать числа с помощью «счётных палочек», то предлагали отсчитать три палочки, потом отдельно четыре палочки, а потом подсчитать их все и получить семь палочек — это и значит, что четыре плюс три равно семи.

Задача 2.1. Какое наименьшее четырёхзначное число делится на 7? какое наибольшее? Сколько четырёхзначных чисел делятся на 7?

Задача 2.2. Сколько четырёхзначных чисел делятся на 123?

Задача 2.3. Сколько четырёхзначных чисел *не* делятся на 123?

Задача 2.4. Каких четырёхзначных чисел больше — чётных или нечётных? Почему?

В правиле суммы важно, что

- всякий элемент множества A входит либо в B, либо в C (мы никого не пропустили);
- в B и C нет общих элементов.

Если в классе m мальчиков и n девочек, то всего в нём $m+n$ школьников. Но если в классе m знающих английский язык и n знающих немецкий, то отсюда не следует, что всего там $m+n$ человек: во-первых, кто-то может не знать ни того, ни другого (и мы его не посчитаем в $m+n$), во-вторых, кто-то может знать оба языка (и мы его посчитаем дважды). Если мы хотим получить общее количество людей, знающих хотя бы один язык (из эти двух), то надо сложить число знающих английский и число знающих немецкий *и вычесть число посчитанных дважды*, то есть число знающих оба языка.

Другой пример: посчитаем, *сколько чисел от* 1 *до* 1000 *делятся на* 2 *или на* 3 (то есть делятся хотя бы на одно из этих двух чисел). Подсчитать чётные (делящиеся на 2) просто: числа

$$1,2,\ 3,4,\ 5,6,\ \ldots,\ 999,1000$$

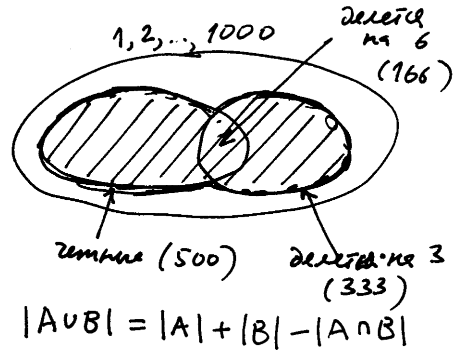
делятся на пары (нечётное, чётное), и этих пар $500 = 1000/2$, то есть чётных чисел 500. Немного сложнее подсчитать делящиеся на 3; если делить на тройки, то одно число останется:

$$1,2,3,\ 4,5,6,\ 7,8,9,\ \ldots,\ 997,998,999,\ 1000.$$

Будет $333 = 999/3$ полные тройки и ещё одно число (не делящееся на 3), то есть всего есть 333 числа, делящиеся на 3. Сколько же чисел, делящихся на два или на три? Если мы сложим 500 и 333, то некоторые числа мы подсчитаем дважды. Это числа, которые делятся и на 2, и на 3, то есть делятся на 6, их в каждой шестёрке

$$1,2,3,4,5,6,\ 7,8,9,10,11,12,\ \ldots,\ 991,992,993,994,995,996,\ 997,998,999,1000.$$

по одному.[3] Делим 1000 на 6 с остатком: $1000 = 166 \times 6 + 4$, видим, что шестёрок будет 166. Теперь можно посчитать, сколько чисел делятся на 2 или на 3:



Если обозначить множество чётных чисел до 1000 за A, делящихся на 3 — за B, то можно записать наш подсчёт так:

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

Здесь $A \cup B$ — это *объединение множества* A *и* B (всё, что входит туда или сюда), $A \cap B$ — их *пересечение* (всё, что входит и туда, и сюда), а $|X|$ — число элементов в множестве X.

Задача 2.5. Сколько чисел от 1 до 1000 не делятся ни на 2, ни на 3?

Задача 2.6. Доля чисел, не делящихся ни на 2, ни на 3 в предыдущей задаче, близка к 1/3. Как это объяснить, глядя на шестёрки чисел?

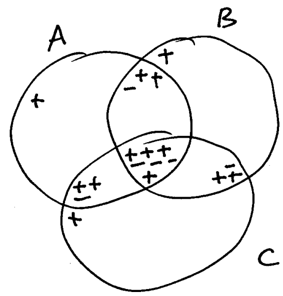
Аналогичные подсчёты возможны и для большего числа множеств. Скажем, пусть в классе кто-то знает немецкий, кто-то английский, кто-то французский, и мы хотим подсчитать, сколько человек знают хотя бы один из этих трёх языков. Можно сложить все три числа, но это будет явно с избытком: тех, кто знают два языка из трёх, мы посчитаем дважды. Чтобы скомпенсировать это, можно вычесть три пересечения (знающих английский и немецкий, знающих английский и французский, и знающих немецкий и французский). Тогда получится правильно? Тоже нет: человека, знающего три языка, мы сначала три раза посчитали, а потом три раза вычли — и в итоге ни разу не посчитали, значит, его надо добавить. В итоге получаем такую формулу:

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

Её называют *формулой включений-исключений* для трёх множеств — сначала включаем все множества, потом исключаем попарные пересечения, потом снова включаем пересечение всех трёх. По существу мы её уже доказали, убедившись, что в

[3]Бдительный читатель спросит здесь, почему свойства «делиться на 2 и делиться на 3» и «делиться на 6» равносильны. Тут можно сослаться на единственность разложения на множители (см. лекцию 4) или просто рассмотреть всевозможные остатки при делении на 6.

правой части любой элемент (входящий ровно в одно из трёх множеств, входящий ровно в два из трёх множеств, и входящий во все три) посчитан правильно (по разу).



На картинке мы ставили плюсы, когда элемент соответствующей части попал в подсчёт, и минусы, когда он вычитался. Правда, там виден только результат, а не процесс, но можете проделать это сами и убедиться, что в каждой части плюсов на один больше, чем минусов.

Задача 2.7. Сколько чисел от 1 до 1000 делятся хотя бы на одно из чисел $2, 3, 5$?

Задача 2.8. Как могла бы выглядеть аналогичная формула для четырёх множеств? Как её можно было бы доказать?

Мы ещё вернёмся к формуле включений-исключений для произвольного числа множеств и докажем её (даже разными способами).

Задача 2.9. Два колокола начали бить одновременно. Удары одного следуют через 2 секунды, а другого — через три секунды. Сколько ударов слышно в минуту, если одновременные удары двух колоколов считать за один?

В музыке такое встречается, когда в одном голосе идут триоли, а в другом обычные восьмые (скажем). Это не так просто сыграть без подготовки. Попробуйте двумя руками стучать по столу: левой на три счёта, правой — на два (каждая рука — как метроном, но частоты отличаются в полтора раза). Получилось? Если да, можно попробовать отношения $3:4$ или $3:5$.

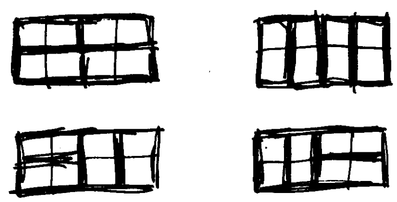
Задача 2.10. Тот же вопрос, если есть ещё и третий колокол, который бьёт раз в 5 секунд (и начал бить в то же время).

2.2 Рекуррентное соотношение: пример

Если нужно подсчитать какие-то предметы (в жизни или в математике), можно разбить их на группы, подсчитать отдельно в каждой группе и потом полученные числа сложить (правило суммы). Собственно, мы с этого начинали, переставляя буквы в слове КОТ: все варианты делятся на три группы (начиная с К, начиная с

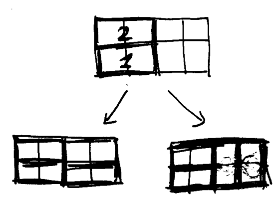
О, начиная с Т), а в каждой группе по два слова (две оставшиеся буквы идут в том или другом порядке). Всего получается $2+2+2=3\times 2=6$ вариантов.

Тот же самый принцип действует и в более сложных случаях. Пусть *мы хотим замостить прямоугольник* 2×4 *доминошками* 1×2. *Сколькими способами это можно сделать*? Можно класть их все горизонтально, или все вертикально, или часть положить горизонтально, а часть вертикально. На рисунке показано четыре варианта, но все ли это варианты? не пропустили ли мы чего-нибудь?

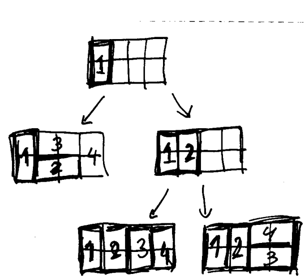


Задача 2.11. Укажите пропущенный вариант.

А теперь? ну хорошо, есть пятый вариант (с краёв вертикально, в середине горизонтально) — но, может, ещё что-то пропущено? Как это узнать? Давайте рассуждать логически, как советовал профессор Стравинский Ивану Бездомному у Булгакова. Посмотрим на левое нижнее поле ($a1$, сказали бы шахматисты). Оно должно быть покрыто доминошкой, которая может быть (а) горизонтальной или (б) вертикальной. Если она горизонтальна (плитка 1), то над ней неизбежно должна быть вторая горизонтальная (плитка 2). Остаётся замостить квадрат 2×2, и это можно сделать двумя способами.



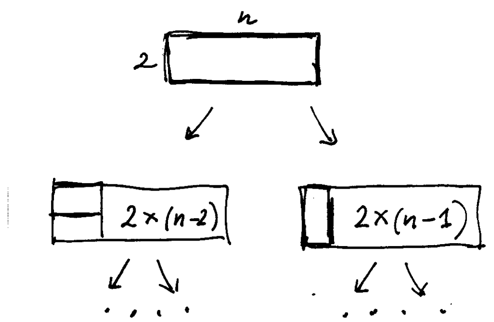
Остался случай (б), когда первая доминошка вертикальна. Тогда остаётся замостить прямоугольник 3×2. Тут снова приходится разбирать случаи.



Левый нижний угол может быть покрыт горизонтальной доминошкой 2 (левый вариант на картинке), тогда над ней будет однозначно 3, а справа 4. А может быть покрыт вертикальной: тогда остаётся квадрат 2×2, который можно замостить двумя способами.

Вот теперь мы не только нашли все возможные способы, но и убедились, что других нет: в группе (а) есть два варианта, в группе (б) есть три варианта, а всего пять вариантов.

Аналогичное рассуждение можно применить и к полю размера $2 \times n$.



Если левая нижняя клетка покрыта горизонтальной доминошкой, то над ней тоже горизонтальная, остаётся покрыть прямоугольник $2 \times (n-2)$. А если левая нижняя клетка покрытая вертикальной доминошкой, то остаётся покрыть прямоугольник $2 \times (n-1)$. То есть все способы — обозначим их число через $T(n)$ — делятся на две группы. В первой группе $T(n-2)$ способов, а во второй $T(n-1)$, итого будет

$$T(n) = T(n-2) + T(n-1).$$

Мы начали с прямоугольника 2×2, у которого два варианта. Если угодно, можно сказать, что $T(2) = T(0) + T(1)$, а прямоугольники 2×0 и 2×1 можно замостить единственным способом. Для прямоугольника 2×3 есть $T(1) + T(2) = 2 + 1 = 3$ варианта, для прямоугольника 2×4 есть $T(2)+T(3)$ варианта, и так далее. Разбиение $T(4)$ на $T(2) + T(3)$ мы как раз и обсуждали подробно, и подсчитали, что $T(4) = 5$. Далее $T(5) = T(3) + T(4) = 3 + 5 = 8$, $T(6) = T(4) + T(5) = 5 + 8 = 13$ и так далее.

Задача 2.12. Нарисуйте все 8 вариантов замощения для прямоугольника 2×5. Как они делятся на группы по 3 и 5?

Мы получили, как говорят, *рекуррентную формулу* для $T(n)$. По этой формуле можно легко подсчитать $T(n)$ и дальше, для $n = 7, 8, 9, \ldots$; эту последовательность $1, 1, 2, 3, 5, 8, 13, \ldots$ называют *числами Фибоначчи*, по имени итальянского математика XIII века, хотя вроде бы она была и раньше известна в Индии. (Фибоначчи объяснял её на примере размножающихся кроликов, но весь этот «оживляж», кажется, только запутывает дело.) Для чисел Фибоначчи есть странная формула

$$T(n-1) = \frac{\varphi^n - \psi^n}{\sqrt{5}}, \quad \text{где} \quad \varphi = \frac{1+\sqrt{5}}{2}, \quad \psi = \frac{1-\sqrt{5}}{2}.$$

Задача 2.13. Проверьте, что по этой формуле при $n = 1$ и при $n = 2$ получится 1.

Задача 2.14. Проверьте, что в последовательности, построенной по этой формуле, каждое число равно сумме двух предыдущих. (Указание: числа φ и ψ являются корнями уравнения $1 + x = x^2$.)

Задача 2.15. Докажите по индукции эту формулу для последовательности Фибоначчи.

Задача 2.16. Проверьте, что можно найти числа Фибоначчи, просто округлив $\varphi^n/\sqrt{5}$ до ближайшего целого числа. Объясните, почему так получается.

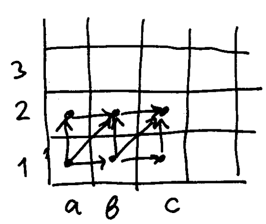
С точки зрения практического вычисления, непонятно, помогает ли эта формула или мешает. С одной стороны, вместо n сложений (если вычислять их подряд) нужно вычислять n-е степени, а это программисты умеют делать за $O(\log n)$ действий, возводя в квадрат и потом перемножая нужные степени двойки, то есть число операций сильно уменьшается. С другой стороны, операции нужно производить с действительными числами, и заранее неясно, с какой точностью, так что возникают дополнительные сложности. Но можно соединить достоинства этих двух способов и ограничиться $O(\log n)$ действиями с целыми числами, но для этого надо знать линейную алгебру и вычислять степени матрицы $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Задача 2.17. (Для знакомых с линейной алгеброй) Как это сделать?

Задача 2.18. Требуется написать строку из n букв А и Б, причём есть такое ограничение: две буквы А не должны идти подряд. Сколькими способами это можно сделать? (Указание: поделим все варианты на две группы — если первая буква А и если первая буква Б.)

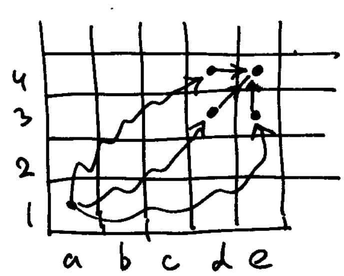
2.3 Рекуррентное соотношение: число путей

Сейчас мы применим тот же метод деления на части и составления рекуррентного соотношения в другой задаче. Она тоже будет на шахматной доске. Представим себе, что на поле $a1$ стоит шахматный король, которому разрешено двигаться только вправо, вверх и по диагонали вправо-вверх. Он может попасть на любое поле, причём не одним способом. Например, на поле $c2$ он может попасть пятью способами.



Задача 2.19. Покажите эти пять способов на картинке. (Они пересекают прямую между первой и второй горизонталью в пяти разных местах.)

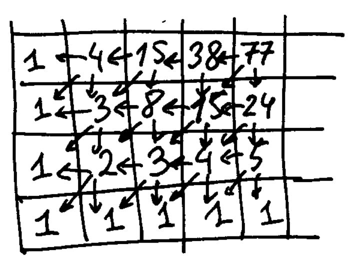
Если речь пойдёт о более далёком поле, скажем, $e4$, то вариантов путей будет много, и перечислять их долго. *Как подсчитать количество вариантов, не перечисляя их*? Давайте пойдём с конца. Откуда король мог попасть на поле $e4$?



Есть три возможных поля: $d4$, $d3$ и $e3$. Соответственно все возможные пути делятся на три (непересекающиеся) группы, и надо сложить количества путей в каждой из этих групп. То есть мы свели задачу к такой же задаче для других клеток:

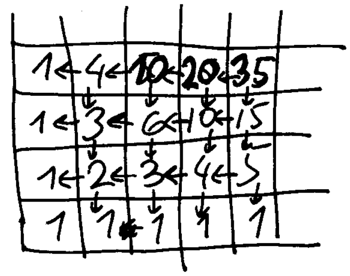
$$T(e4) = T(d4) + T(d3) + T(e3),$$

где $T(X)$ обозначает число способов попасть в клетку X. Аналогичное соотношение можно написать для всех других клеток, кроме клеток на нижней или верхней границе, когда вариант только один (идти вдоль границы — если мы отойдём, то не сможем вернуться). Теперь, пользуясь этим соотношением, мы можем вычислить ответы для всех клеток.



Каждое число, кроме крайних единиц, равно сумме трёх чисел: слева, снизу и слева-снизу (по рекуррентной формуле), и после нескольких сложений получаем ответ: в клетку $e4$ можно попасть 77 способами.

Если мы запретим королю ходить по диагонали, а разрешим ходить только на клетку вправо или вверх (отчего он немного станет ладьёй), то способов станет, естественно, меньше. И рекуррентная формула будет другой: чтобы попасть на клетку $e4$ можно только с $e3$ или $d4$, групп не три, а две. Соответственно число вариантов в каждой клетке равно сумме чисел под ним и слева от него:



Для клетки $e4$ получается 35 вариантов. Знающие люди сразу узна́ют в этой картинке *треугольник Паскаля*, только повёрнутый, и могут получить тот же ответ 35 по формуле

$$\binom{7}{3} = \frac{7!}{4!3!} = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = 35,$$

но и безо всех этих знаний заполнить таблицу можно совсем быстро и прийти к тому же ответу.

Мы ещё вернёмся к этой задаче, но сейчас рассмотрим несколько совсем простых рекуррентных соотношений.

2.4 Слова и правило произведения

Будем рассматривать пятизначные числа, в которых встречаются только цифры 1 *и* 2. Наименьшее из них 11111, наибольшее 22222, но идут они с большими промежутками, так что их на самом деле совсем немного. *Сколько*?

На первом месте может стоять 1 или 2, так что все такие числа делятся на две группы. Сколько чисел в каждой группе? нам остаётся дописать четыре цифры, так что надо подсчитать количество четырёхзначных чисел из цифр 1 и 2. Получаем, что пятизначных чисел вдвое больше, чем четырёхзначных: $T(5) = 2T(4)$, если за $T(n)$ обозначить количество n-значных чисел, составленных из цифр 1 и 2. Другое объяснение: из каждого четырёхзначного числа можно получить два пятизначных, приписав к нему единицу или двойку. Значит, пятизначных чисел вдвое больше.

По тем же причинам $T(4) = 2T(3)$ и так далее:

$$T(5) = 2T(4) = 4T(3) = 8T(2) = 16T(1) = 16 \cdot 2 = 32.$$

Мы использовали здесь, что $T(1) = 2$, поскольку однозначных чисел ровно два (числа 1 и 2). Можно было бы формально написать $T(1) = 2T(0) = 2 \cdot 1 = 2$, считая, что нульзначные числа не содержат ни одной цифры, тут выбора нет, и вариант только один. Так или иначе, $T(5) = 32 = 2^5$.

Задача 2.20. А чему равна сумма этих 32 чисел? (Чтобы её найти, не обязательно их все выписывать и складывать!)

Точно так же можно подсчитать, скажем, *количество семизначных чисел, в которых все цифры нечётны.* На первом месте может стоять любая из пяти нечётных цифр $1, 3, 5, 7, 9$, так что все они делятся на пять групп (в зависимости от первой цифры). Сколько чисел в каждой группе? Столько, сколько есть шестизначных чисел из нечётных цифр, и так далее. Получаем рекуррентное соотношение $T(n) = 5T(n-1)$ с начальным условием $T(1) = 5$, откуда

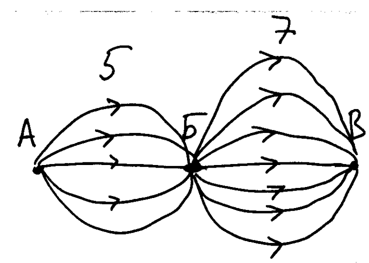
$$T(7) = 5T(6) = 5^2T(5) = 5^3T(4) = 5^4T(3) = 5^5T(2) = 5^6T(1) = 5^6 \cdot 5 = 5^7.$$

Задача 2.21. Найдите сумму всех этих чисел. (Теперь выписывать и складывать их явно слишком долго.)

В общей форме это утверждение можно сформулировать так. Пусть есть какой-то *алфавит*, содержащий n символов (*букв* этого алфавита, как говорят), и мы составляем последовательности (цепочки) из k букв этого алфавита. Их принято называть *словами длины* k в этом алфавите, хотя, конечно, никакого смысла в этих «словах» не ищут. В последнее время всё больше говорят «строки длины k», видимо, переводя английское слово *string*. (Иногда его даже не переводят и говорят о «стрингах длины k», а способы быстрой обработки слов называют «стрингологией», от английского жаргонного слова stringology — но это уже, пожалуй, экстремизм.)

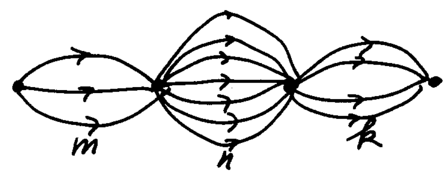
Так вот, мы по существу доказали, разбирая наши примеры, что *число различных слов длины k в алфавите из n символов равно n^k*. Формально это легко доказывается индукцией по k. При $k = 1$ получаем n слов длины n, то есть n букв (база индукции). Слова длины k делятся на k групп по первой букве, в каждой группе к этой первой букве дописывают любое слово из $k-1$ букв, поэтому в группе n^{k-1} слов (предположение индукции), а всего $n \cdot n^{k-1} = n^k$, что и требовалось доказать.

Ту же самую формулу можно объяснить немного другим образом. Представим себе, что *из города А в город Б ведут* 5 (*непересекающихся*) *дорог, а из города Б в город В ведут* 7 (*непересекающихся*) *дорог. Сколькими способами можно проехать из А до В в два приёма* (*сначала Б, и потом в В*)?



Ясно, что есть 35 вариантов: каждая из пяти возможностей на первом шаге может комбинироваться с каждой из семи возможностей на втором. (Если следовать нашей схеме, то все варианты делятся на пять групп, по своему первому шагу, и в каждой группе семь вариантов — по второму.)

Аналогичное утверждение верно и для большего числа этапов: если на первом шаге можно выбрать любой из m способов, на втором шаге (независимо от того, что выбрано на первом) можно выбрать любой из n способов, а на третьем — любой из k способов, то всего путей будет mnk.



Формально говоря, его тоже можно доказывать индукцией по числу этапов, но это и так понятно: если у матери m дочерей, у каждой дочери по n дочерей, а у каждой из последних по k дочерей, то всего будет mnk правнучек. На генеалогическом дереве (по женской линии) у корня будет m дочерей, у каждой из них по n дочерей, у которых по k дочерей — всего mnk листьев (вершин верхнего уровня). Ясно, что коэффициенты ветвления на всех уровнях перемножаются (на каждом следующем уровне больше вершин во столько раз, каков коэффициент ветвления).

Сформулируем ещё раз «правило произведения»: *если объект интересующего нас вида строится в несколько шагов, и на каждом шаге есть выбор из какого-то числа вариантов, то общее количество объектов равно произведению количеств вариантов выбора для каждого из шагов.*

Задача 2.22. В каждый из пяти дней недели один из 20 школьников класса объявляется дежурным. Сколькими способами можно составить таблицу дежурств на неделю? (На каждый день дежурным могут назначить любого школьника, независимо от того, дежурил он уже или нет.)

Задача 2.23. В некоторой стране[4] автомобильные номера состоят из трех цифр и трёх букв, при этом можно использовать любые цифры от 0 до 9 и любые английские буквы от A до Z (их 26). Какое максимальное число различных автомобильных номеров можно составить таким образом?

С другими применениями этой формулы хорошо знакомы программисты: скажем, существует $256 = 2^8$ различных байтов (последовательностей из восьми битов, или слов длины 8 в двоичном алфавите $\{0, 1\}$).

Задача 2.24. Шахматный король, который умеет ходить только вправо и вверх на одну клетку, находится на поле $a1$ и должен попасть на другую диагональ доски (из $a8$ в $h1$). Сколько ходов он для этого должен сделать? Сколькими способами он может это сделать?

[4]В России автомобильные номера образца 1993 года тоже содержат три буквы и три цифры, но разрешается использовать лишь 12 букв, похожих на латинские, набор из трёх нулей не допускается. Ещё в правой части номера указывается код региона (обычно две цифры). Количество этих кодов постепенно увеличивается, когда старых номеров не хватает, и даже появляются трёхзначные коды регионов. Так что жизнь сложнее нашей формулы.

Можно подсчитать и количество различных подмножеств у множества из n элементов. Допустим, *в классе n школьников, и некоторые из них пошли сегодня вечером на концерт. Сколько может быть вариантов* (кто пошёл, а кто не пошёл)? Здесь подмножество состоит из тех школьников, которые были на концерте. Возможно, что никого не было — тогда получается, как говорят, *пустое* множество. Допускается и вариант, когда были все n школьников (само множество тоже считают его подмножеством).

Представим себе, что мы составляем отчёт о посещении концерта, поставив рядом с каждой из n фамилий плюс или минус (был или не был). Получится слово длины n в алфавите из двух символов $\{+,-\}$, и таких слов, как мы знаем, 2^n.

Другое объяснение: выберем одного из n школьников, скажем, Васю, и все варианты посещения концерта поделятся на две категории: те, где Вася был на концерте, и те, где он не был. А в каждой группе столько вариантов, сколько подмножеств у множества с $n-1$ элементами, так что с каждым новым школьником число вариантов увеличивается вдвое.

Или так: у школьников по очереди спрашивают, пойдут ли они на концерт. Получается n этапов, на каждом из них выбор из двух вариантов, по правилу произведения всего 2^n вариантов.

Задача 2.25. Каких подмножеств у множества $\{a, b, c, d, e, f\}$ больше — содержащих букву a или не содержащих? Каких подмножеств больше у того же множества — из чётного числа элементов или из нечётного числа элементов?

Задача 2.26. Будем рассматривать расстановки слонов на шахматной доске, при которой они не бьют друг друга. (Число слонов может быть произвольным, в том числе равным нулю). Покажите, что число расстановок является точным квадратом (квадратом целого числа).

2.5 Выбор с ограничениями

Применять формулу из предыдущего раздела нужно аккуратно. Скажем, в самом начале мы составляли слова длины 3 из букв К, О, Т, но получили не $3^3 = 27$ слов, по нашему теперешнему подсчёту, а только 6. В чём тут дело? Или ещё: подсчитаем четырёхзначные числа: четыре места, десять цифр, по формуле их будет $10^4 = 10000$, а у нас получалось 9000. Что не так?

Дело в том, что в обоих случаях мы считаем лишнее: нас интересуют только комбинации с некоторыми ограничениями. Составляя слова из К,О,Т, мы разрешали только переставлять буквы, то есть каждую букву можно было использовать по разу (не больше и не меньше). В четырёхзначных числах на первом месте не может стоять нуль (тогда будет трёх- или меньше-значное число). Однако и в том, и в другом случае можно получить правильный ответ по формуле произведения.

Для четырёхзначных чисел: на первом месте может стоять любая из 9 цифр $1, 2, \ldots, 9$ (кроме нуля), на втором, третьем и четвёртом — любая из десяти цифр, поэтому в произведении получаем как раз $9 \cdot 10 \cdot 10 \cdot 10 = 9000$. Аналогично с котом:

на первом месте может стоять любая из трёх букв К,О и Т. После того как первая буква выбрана, на втором месте может стоять любая из двух оставшихся букв — ну, а на третьем месте уже никакого выбора не остаётся, ставим ту букву, что не использована. Всего получаем как раз $3 \cdot 2 \cdot 1 = 6$.

Задача 2.27. Забор состоит из 100 вертикальных досок (самая левая, потом вторая слева, потом третья, и так далее до сотой). Маша хочет покрасить каждую доску забора в какой-то цвет. У неё есть три разных краски, и при этом она не хочет красить соседние доски в один и тот же цвет (иначе границы не видно). Сколько вариантов раскраски есть?

Задача 2.28. Сколько существует четырёхзначных чисел, в которые входит (хотя бы раз) цифра 7? (Указание. Эту задачу можно решать, разбивая числа на группы в зависимости от такого, на каких местах появляется семёрка. Но проще подсчитать четырёхзначные числа, в которые цифра 7 *не* входит.)

Важный случай выбора с ограничениями — это перестановки. Мы уже рассматривали все перестановки в слове из трёх букв, и насчитали шесть таких перестановок. А что будет для слова из n букв? *Сколько слов можно получить из него перестановками букв*? (Все буквы в слове считаем разными.)

Первую букву можно выбрать n способами. После этого (для каждого из n вариантов) вторую букву можно выбрать $n-1$ способами, взяв любую из $n-1$ оставшихся букв. На третьем шаге у нас $n-2$ способов, и так далее до последнего шага, когда уже выбора нет (осталась ровно одна буква). Всего получается

$$n(n-1)(n-2)(n-3) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$$

способов, для этого числа есть специальное обозначение $n!$ (читается «эн факториал»).

Задача 2.29. Что больше: $100!$ или 100^{100}?

Эту задачу можно решить комбинаторно, посчитав, сколько всего есть слов длины 100 в 100-буквенном алфавите и сколько их них не содержат повторений букв (и потому каждую букву содержат по одному разу).

Более общая ситуация, чем перестановки — это выбор без повторений. Пусть есть какие-то n предметов, и нужно составить упорядоченный список, взяв k из них. Например, *пусть в соревновании участвуют* 20 *спортсменов, и разыгрываются три медали — золотая, серебряная и бронзовая.* В нашей терминологии это значит, что $n = 20$, а $k = 3$. *Сколько различных возможных исходов у такого соревнования*?

Победителем может стать любой из 20 человек. После того как победитель выбран, остаётся 19 кандидатов на второе место, а после того как один из них выбран, 18 кандидатов на третье место. Всего, таким образом, $20 \cdot 19 \cdot 18$ вариантов.

В общем случае надо перемножить k сомножителей $n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)$ (проверьте, что последний из них написан правильно и получается именно k сомножителей).

В терминах слов последнее утверждение можно сформулировать так: *количество k-буквенных слов в n-буквенном алфавите, не содержащих повторяющихся букв, равно* $n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)$. Это число иногда обозначают $(n)_k$ по аналогии со степенью n^k (которая получится, если не запрещать повторения букв).

Можно ещё записать, что

$$(n)_k = \frac{n!}{(n-k)!}$$

(в этой дроби последние $n-k$ множителей числителя сокращаются со знаменателем, и остаётся ровно то, что мы написали).

Мы предполагали, определяя число $(n)_k$, что $k \le n$. Как вы думаете, как разумно определить $(n)_k$ при $k > n$? По формуле надо умножать числа n, $n-1$, и так далее до 1, потом дойдет дело до нуля (ведь $k > n$), и даже до отрицательных чисел. Но в любом случае в произведении будет нуль (один из сомножителей равен нулю). И с комбинаторным смыслом это тоже согласовано: нельзя составить слова без повторений, у которых длина больше числа букв в алфавите. Значит, количество таких слов равно нулю.

2.6 Подсчёты с кратностью

Вернёмся опять к перестановкам букв в слове КОТ. Что будет, если вместо слова КОТ взять, скажем, слово ТОТ? Тогда получится только три слова: ОТТ, ТОТ и ТТО. И понятно, почему их стало меньше: потому что в слове ТОТ есть две одинаковые буквы. Если в словах

КОТ КТО ОКТ ОТК ТКО ТОК

заменить К на Т, то получится список

ТОТ ТТО ОТТ ОТТ ТТО ТОТ,

в котором каждое слово входит два раза (перестановка букв К и Т давала новое слово, а перестановка двух букв Т — нет), и разных слов только три.

Вот ещё пример аналогичной ситуации. Предположим, что *в турнире* 10 *команд каждая сыграла по одному разу с каждой. Сколько было игр*? Будем считать это число так: каждая команда сыграла девять игр (с каждой из 9 остальных), команд 10, получается $10\cdot 9 = 90$ игр. Правильно это?

Нет: скажем, для трёх команд подобный подсчёт даёт $3\times 2 = 6$ игр, в то время как реально игры будет только три (отдыхать может любая из трёх команд). В чём тут ошибка? Дело в том, что игру между любыми двумя командами А и Б мы посчитали дважды: один раз — когда мы считали игры команды А, и второй раз — когда считали игры команды Б. Значит, полученный результат вдвое больше истинного числа игр, и в случае 10 команд игр было не $10\cdot 9 = 90$, а $10\cdot 9/2 = 45$.

Задача 2.30. Будем считать иначе: когда первая команда отыграет свои $(n-1)$ игр, останется та же задача (подсчёт общего числа игр) для $n-1$ команд. Эта рекуррентная формула даёт ответ $(n-1)+(n-2)+\ldots+3+2+1+0$ (когда останется одна команда, число игр равно нулю). Совпадает ли этот ответ с полученным нами раньше?

Задача 2.31. Сколько диагоналей в выпуклом n-угольнике? (Указание: из каждой вершины выходят $n-3$ диагонали во все места, кроме самой вершины и двух её соседей.)

Задача 2.32. Сколько есть двухэлементных подмножеств в множестве из n элементов? Почему эта задача повторяет задачу о количестве игр в турнире n команд?

Приведём ещё несколько примеров такого рода.

Вспомним задачу о соревновании, в котором из 20 участников определяются три победителя (золотой, серебряный и бронзовый призёр). Как мы посчитали, есть $20 \times 19 \times 18$ возможных исходов такого соревнования. *Что изменится, если в соревновании по-прежнему три победителя, но они никак не ранжированы, просто выбрано три человека из двадцати*? Тогда в списке из $20 \times 19 \times 18$ исходов каждый вариант будет встречаться 6 раз (ведь если уже фиксированы три победителя, то распределить между ними золотую, серебряную и бронзовую медаль можно $3! = 6$ способами, это мы тоже знаем). Значит, чтобы подсчитать количество возможных исходов в новом типе соревнования, старый ответ нужно поделить на $6 = 3!$, получится

$$\frac{20 \cdot 19 \cdot 18}{1 \cdot 2 \cdot 3} = 1140 \text{ исходов.}$$

Задача 2.33. Сколько существует 3-элементных подмножеств у 20-элементного множества?

Задача 2.34. В классе из 20 человек нужно выбрать старосту и двух его помощников. Сколькими различными способами это можно сделать? (Каким будет это число, если у старосты есть первый помощник и второй помощник, и как оно изменится, если перестать их ранжировать?)

Задача 2.35. Семь человек, среди которых есть Аня, Беня и Ваня, становятся в очередь. Сколькими способами они могут это сделать? В какой доле из них Аня стоит ближе к началу, чем Беня? В какой доле из них Аня стоит дальше от начала, чем Беня, но ближе, чем Ваня? В какой доле из них Аня стоит ближе к началу, чем Беня и чем Ваня?

Другой пример. Пусть *есть n людей, которые хотят образовать хоровод* (встать в круг и кружиться). *Сколькими способами они могут это сделать*? Разница между хороводом, и, скажем, очередью в том, что варианты, отличающиеся лишь поворотом круга, считаются за один (когда они начнут кружиться, уже никто не вспомнит, с чего началось).

Тот же приём позволяет получить ответ. Если не разрешать хороводу кружиться, а просто нарисовать n точек в вершинах правильного n-угольника, и считать, сколькими способами можно расставить в них n человек, то получится, как мы знаем, $n!$ способов. Если вспомнить о вращении, то получатся группы из n вариантов, отличающиеся лишь поворотом (из одного способа можно получить n, рассмотрев все повороты), поэтому этих групп будет $n!/n = (n-1)!$.

Задача 2.36. Как получить тот же самый ответ, описывая хоровод с точки зрения одного из участников?

С последней задачей связан любопытный парадокс (или, точнее сказать, софизм). Представим себе *круг, разбитый на* 7 *равных секторов. Каждый из секторов нужно раскрасить в красный или синий цвет. При этом раскраски, отличающиеся поворотом круга, считаем за одинаковые. Сколькими способами можно это сделать*?

Действуем как раньше: если круг не вращать, то всего есть 7 секторов, для каждого из них надо решить, в какой из двух цветов красить, получаем по правилу произведения $2^7 = 128$ вариантов. Повороты круга, как мы видели, делят их на группы по 7, так что групп будет $128/7 = 18\frac{2}{7}$. Это, конечно, ерунда — групп должно быть целое число. Где мы ошиблись? Постарайтесь догадаться, не читая объяснения в следующем абзаце.

* * *

На самом деле не все группы будут по 7: если все сектора круга красные (или синие), то вращение ничего не меняет. То есть есть две группы, состоящие из единственной раскраски. Остальные группы, как можно проверить (тут важно, что число 7 простое), действительно состоят из 7 элементов, их $(128-2)/7 = 18$, добавляя две особые группы, получаем ответ: 20 вариантов.

Задача 2.37. Почему важно, что число n простое? Пройдёт ли это рассуждение, скажем, для круга из четырёх секторов? Покажите, что при простом n число $2^n - 2$ делится на n, и вообще $a^n - a$ делится на n при любом целом a. (Это утверждение эквивалентно *малой теореме Ферма*, о которой пойдёт речь в главе о целых числах. Можно сказать, таким образом, что эта задача даёт комбинаторное доказательство этой теоремы.)

Задача 2.38. Сколько различных слов можно получить, переставляя буквы в слове КАНАТ?

Задача 2.39. Сколько различных слов можно получить, переставляя буквы в слове КОКОН?

Задача 2.40. Сколько различных слов можно получить, переставляя буквы в слове МАТЕМАТИКА?

2.7 Подмножества и числа сочетаний

Сейчас мы используем метод предыдущего раздела, чтобы получить формулу для числа k-элементных подмножеств n-элементного множества. Представим себе, что *в классе из n учеников нужно сформировать спортивную команду из k учеников* (внутри команды все равны, никаких ролей нет). *Сколькими способами это можно сделать?*

Один способ подсчёта такой. Будем выбирать игроков команды по очереди. Первого можно выбрать n способами, второго $n-1$ способами (годятся все, кроме первого), третьего $n-2$ способами, всего

$$(n)_k = n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)$$

способами (см. выше). Но так мы подсчитали не число возможных команд, а число возможных упорядоченных списков команд (известно, какой игрок первый, какой второй, и так далее). Если не обращать внимание на номера игроков, то много разных списков соответствуют одной команде — они получатся, если игроков переставлять в списке, а это можно сделать $k!$ способами. Значит, число групп равно $(n)_k/k!$; поскольку $(n)_k = n!/(n-k)!$, то можно переписать ответ в более симметричной форме:

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

Это число (число способов выбрать k элементов из n, без учёта порядка, или число k-элементных подмножеств n-элементного множества, если говорить научно) традиционно называют *числом сочетаний из n по k* и обозначают $\binom{n}{k}$ (мы использовали это обозначение в левой части формулы) или C_n^k. Первый вариант сейчас используется всё чаще, но во многих русских (и французских) книгах используется обозначение C_n^k.

Интересным следствием этой формулы является такое наблюдение: произведение любых подряд идущих k натуральных чисел делится на $k!$. В самом деле, это произведение равно $(n)_k$, где n — последнее из этих чисел, и по формуле $(n)_k/k!$ должно получиться целое число.

При $n=2$ получается, что произведение любых двух соседних натуральных чисел чётно (это не удивительно, поскольку одно из них нечётно, а второе чётно). При $n=3$ получается, что произведение любых трёх соседних натуральных чисел делится на $3!=6$. Это тоже не так удивительно: ровно одно из этих чисел делится на 3, поэтому произведение делится на 3, а также оно делится на 2 (для этого достаточно даже двух чисел), значит, оно делится на 6. Дальше объяснения становятся сложнее.

Задача 2.41. Чисто арифметически (не ссылаясь на число сочетаний) докажите, что произведение любых четырёх подряд идущих чисел делится на 24.

Другой способ подсчёта чисел сочетаний — использование рекуррентной формулы. Допустим, мы хотим найти $\binom{n}{k}$, количество способов выбрать команду из k человек, если в классе всего n человек. Выберем какого-нибудь школьника, пусть его зовут Петя, и все способы разделим на две группы: когда Петя входит в команду и когда Петя не входит в команду.

Если мы решили, что Петя не входит в команду, то наша задача состоит в том, чтобы выбрать k человек из оставшихся $n-1$ человек. Это можно сделать $\binom{n-1}{k}$ способами (в наших обозначениях). С другой стороны, если мы решили, что Петя входит в команду, то остаётся выбрать $k-1$ человек из $n-1$ человек (один из k уже есть). По правилу суммы получаем рекуррентное соотношение

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Правда, чтобы это рассуждение и формула были законными, нужно, чтобы $0 < k < n$, иначе в правой части будет бессмыслица (при $k < 0$ или $k > n$ непонятно, как понимать $\binom{n}{k}$).

Эту формулу полезно проиллюстрировать на картинке, расположив числа сочетаний в виде треугольника:

$$\begin{array}{ccccccccccc}
 & & & & & \binom{0}{0} & & & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5}
\end{array}$$

Наша рекуррентная формула означает, что каждое число в этом треугольнике получается сложением двух чисел в предыдущей строке (сочетания из $n-1\ldots$): слева над ним ($\ldots$ по $k-1$) и справа над ним ($\ldots$ по k). Числа с краю таблицы равны единице: если никого выбирать в команду не нужно ($k=0$) или если нужно выбрать всех ($k=n$), какие уж тут варианты. По этим правилам таблицу легко заполнить:

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

Этот треугольник называют *треугольником Паскаля*, в честь Блеза Паскаля, французского математика и философа XVII века (хотя он не был первым — в разных странах он упоминался и раньше, как считают историки[5]).

Теперь, помимо формулы с факториалами, у нас есть и другой способ находить числа сочетаний. Для заполнения таблицы он явно лучше (на каждый элемент

[5]В частности, говорят, что его знал Омар Хайам, которого также считают автором четверостиший-рубаи. Правда, в них вроде бы ничего о числах сочетаний не говорится — напротив, одно из них, в переводе Германа Плисецкого, звучит так: «*Я спросил у мудрейшего: что ты извлёк // Из своих манускриптов? Мудрейший изрёк:// Счастлив тот, кто в объятьях красавицы нежной // По ночам от премудрости книжной далёк*».

уходит одно сложение). Если же нам надо вычислить какое-то одно число в этой таблице, то по сравнению с формулой операций больше (пропорционально n^2, а не n), но каждая из них проще (сложение вместо умножений и делений).

Можно окружить этот треугольник слева и справа нулями, тогда рекуррентная формула будет верной и для крайних чисел (и вообще для всех чисел: добавленные нули тоже будут ей подчиняться).

Задача 2.42. Сколько есть двоичных слов длины n, содержащих k единиц?

Задача 2.43. Слово состоит из n букв А и k букв Б. Сколько различных слов можно составить, переставляя буквы в этом слове?

Задача 2.44. Сколько существует слов длины 10 в алфавите $\{\text{А}, \text{Б}, \text{В}\}$, содержащих ровно 4 буквы А? (Сначала надо выбрать, где будут буквы А, а потом посмотреть, сколькими способами можно выбрать остальные буквы.)

Задача 2.45. Сколько существует четырёхзначных чисел, в которых цифры идут в убывающем порядке?

2.8 Ещё о числах сочетаний

Числа сочетаний обладают разными интересными свойствами. Доказывая эти свойства, мы можем пользоваться и определением (как числа способов), и формулой с факториалами, и свойством треугольника Паскаля. Часто одно и то же свойство можно доказывать по-разному.

2.8.1 Симметрия

Треугольник Паскаля симметричен относительно вертикальной оси: если читать каждую строку справа налево, получится то же самое. Другими словами,

$$\binom{n}{k} = \binom{n}{n-k} \quad \text{при } k = 0, 1, 2, \ldots, n$$

Доказательство с факториалами. В самом деле,

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

Если заменить в этой формуле k на $n-k$, то $(n-k)$ превратится в $n-(n-k)=k$, так что два сомножителя в знаменателе просто поменяются местами. Например,

$$\binom{20}{3} = \frac{20!}{3!17!}, \quad \binom{20}{17} = \frac{20!}{17!3!}.$$

Доказательство с треугольником. Первые несколько строк треугольника симметричны. Достаточно доказать, таким образом, что симметрия сохраняется при переходе к следующей строке. Это следует из того, что и правило симметрично: оба крайних числа равны единице, а между ними каждое равно сумме чисел слева-над и справа-над. Если мы будем следовать этому правилу и выписывать треугольник на стеклянной доске, то стоящий сзади доски коллега с нами согласится, просто для него числа «слева-над» и «справа-над» поменяются местами.[6]

Комбинаторное доказательство. Что такое, скажем, $\binom{19}{9}$? Это количество способов, которым в классе из 19 человек можно выбрать команду из 9 человек. А $\binom{19}{10}$ — это число способов выбрать команду из 10 человек. Но и в том, и в другом случае мы просто делим школьников на две команды по 9 и 10 человек, так что речь идёт об одном и том же числе.

Более формально можно сказать так. Пусть есть n-элементное множество A. Для каждого его k-элементного подмножества B рассмотрим множество из остальных $n-k$ элементов, обозначаемое $A \setminus B$ (те элементы A, которые не входят в B). Таким образом возникает взаимно однозначное соответствие между k-элементными подмножествами A и $(n-k)$-элементными подмножествами A. Значит, тех и других одинаковое количество.

2.8.2 Сумма чисел в строке

Сумма всех чисел в n-й строке треугольника Паскаля равна 2^n:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n.$$

Например, $1+2+1 = 4 = 2^2$, $1+3+3+1 = 8 = 2^3$ и так далее.

Доказательство по индукции. Достаточно показать, что в каждой следующей строке сумма вдвое больше предыдущей. Раскроем каждое число как сумму двух чисел предыдущей строки. Скажем, для четвёртой строки:

$$\begin{array}{ccccccccc} & (1+1) & & (3+3) & & (3+3) & & (1+1) & \\ 1 & & 1+3 & & 3+3 & & 3+1 & & 1 \\ \shortparallel & & \shortparallel & & \shortparallel & & \shortparallel & & \shortparallel \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

Если сгруппировать числа по-другому, то видно, что каждое число предыдущей строки, включая крайние единицы, входит дважды. Поэтому сумма следующей строки будет вдвое больше, чем предыдущей.

Комбинаторное доказательство. Число $\binom{n}{k}$ равно числу k-элементных подмножеств n-элементного множества. Когда мы складываем эти числа при всех k (и фиксированном n), мы получаем число всех подмножеств n-элементного множества, а оно равно 2^n, как мы знаем (для каждого из n элементов есть два варианта: включать его или нет).

[6] Не надо только понимать эту метафору слишком буквально: на самом деле с другой стороны цифры выглядят не так.

Можно пересказать это в терминах двоичных слов. Скажем, при $n = 3$ мы считаем двоичные слова длины 3 по группам: сначала из одних нулей, потом с одной единицей, потом с двумя, и наконец, с тремя:

$$\underbrace{000}_{1}\quad\underbrace{001\ 010\ 100}_{3}\quad\underbrace{110\ 101\ 011}_{3}\quad\underbrace{111}_{1}$$

Ещё одно доказательство мы легко получим в следующем разделе по формуле бинома Ньютона.

2.8.3 Знакочередующаяся сумма

Если складывать числа в строке треугольника Паскаля с чередующимися знаками, то получится нуль:

$$\begin{aligned}
1-1&=0\\
1-2+1&=0\\
1-3+3-1&=0\\
1-4+6-4+1&=0\\
1-5+10-10+5-1&=0\\
&\dots
\end{aligned}$$

или, в общем виде,

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

(множитель $(-1)^k$ как раз и обеспечивает знакочередование: он равен то $+1$, то -1).

Почему это так? Если приглядеться, то в нечётных строках всё понятно: равные числа сокращаются, и ничего не остаётся. Но и в чётных строках, уже более загадочным образом, получается нуль. Почему?

Посмотрим, скажем, как из третьей строки получается четвёртая. При этом нам даже неважно, какие числа стоят в третьей строке, так что заменим их буквами:

$$\begin{array}{ccccccccc}
 & a & & b & & c & & d & \\
a & & a+b & & b+c & & c+d & & d
\end{array}$$

Теперь в знакопеременной сумме четвёртой строки всё сокращается:

$$a-(a+b)+(b+c)-(c+d)+d=0.$$

Так и в общем случае: каждое число в n-й строке войдёт в два соседних числа в следующей, $(n+1)$-й строке, и эти соседние числа будут с разными знаками, так что всё сократится.

Можно доказать равенство нулю знакопеременной суммы и комбинаторно. Перенесём члены с минусами в другую часть. Нам надо доказать теперь, что

$$\binom{n}{1}+\binom{n}{3}+\binom{n}{5}+\ldots=\binom{n}{2}+\binom{n}{4}+\binom{n}{6}+\ldots$$

Другими словами, надо доказать, что у n-элементного множества столько же подмножеств нечётного размера (левая часть), что и чётного (правая). Почему это так? Выберем в нашем n-элементном множестве какой-то элемент, и поделим все подмножества на пары, отличающиеся только добавлением-удалением этого элемента. (Если подмножества представлять как битовые слова, то мы выбираем какую-то позицию и спариваем две строки, различающиеся в этой позиции.) Подмножества, образующие пару, отличаются по количеству элементов на один (тот самый выбранный), поэтому в одном из них чётное число элементов, а в другом нечётное число элементов. Дальше всё понятно (если в мешке белые и чёрные зёрна слиплись в бело-чёрные пары, то белых и чёрных зёрен поровну).

2.8.4 Снова о включениях и исключениях

Общая формула включений и исключений:

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

Чтобы найти общее количество элементов в n пересекающихся множествах, мы складываем размеры всех этих множеств, затем вычитаем размеры попарных пересечений, потом прибавлям размеры тройных, и так далее с чередованием знаков до пересечения всех множеств. Знак перед последним членом зависит от чётности n, отсюда и множитель $(-1)^n$.

Нам надо проверить (как мы делали при $n = 2$ и $n = 3$), что каждый потенциальный кандидат учтён ровно один раз (после всех сокращений). Пусть кандидат входит в k из n множеств. Тогда он войдёт k раз, когда мы будем складывать размеры множеств. Потом мы его вычтем столько раз, во сколько пар множеств он входит. Другими словами, мы разными способами выбираем из содержащих его k множеств пару. Таких способов $\binom{k}{2}$. Затем мы учитываем кандидата по разу для каждой тройки содержащих его множеств, таких троек $\binom{k}{3}$. В итоге он будет учтён

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \binom{k}{4} + \ldots$$

Это и есть знакопеременная сумма, которая равна нулю, только без первого члена и с обратным знаком, поэтому эта сумма равна единице, что нам и требовалось.

Например, если выбранный элемент входит в 4 множества, то сначала он будет посчитан 4 раза с каждым из них, потом вычтен 6 раз с каждой парой (таких пар $\binom{4}{2} = 6$), потом добавлен 4 раза обратно с каждой тройкой, и наконец вычтен при вычитании пересечения четырёх множеств. Всего получится

$$4 - 6 + 4 - 1 = 1$$

(если перенести левую часть в правую, и будет нулевая знакопеременная сумма четвёртой строки треугольника Паскаля).

Задача 2.46. Имеется слово из n различных букв. Сколько слов можно получить перестановками этих букв, если требуется, чтобы никакая буква не осталась на своём месте? (Примените формулу включений и исключений, взяв в качестве A_i множество слов, где i-я буква остаётся на своём месте).

Если вы знакомы с математическим анализом, покажите, что доля таких слов (среди всех $n!$ слов) близка к $1/e$, где $e = 2{,}71828\ldots$ — основание натуральных логарифмов.

2.8.5 Пути, подмножества, слова

Мы рассматривали задачу, в которой надо было попасть в данную клетку шахматной доски, начав с нижней левой клетки и делая шаги вверх и вправо. Для числа способов (в том варианте, когда по диагонали ходить нельзя) получалась та же рекуррентная формула, что и для треугольника Паскаля (только повёрнутая). Отсюда уже следует, что число способов в этой задаче равно соответствующему числу сочетаний.

Но как это объяснить комбинаторно, не используя рекуррентного соотношения и рассуждения по индукции? Это тоже несложно. Будем записывать последовательность ходов как слово в алфавите $\{\text{П}, \text{В}\}$: ход в**п**раво буквой П, а ход на**в**ерх буквой В. Скажем, чтобы попасть из $a1$ в $e4$, надо сделать четыре хода вправо и три хода вверх. Это соответствие взаимно однозначно: любое слово из четырёх букв П и трёх букв В описывает движение из $a1$ в $e4$, надо просто следовать инструкциям, закодированным буквами.

Следовательно,

маршрутов из $a1$ в $e4$ столько же, сколько слов из четырёх букв П и трёх букв В.

А сколько таких слов? Можно сказать, что у нас есть слово длины 7, содержащее четыре буквы П (и остальные В). Чтобы задать такое слово, надо указать, в каких четырёх (из семи) позициях стоят буквы П. Другими словами, надо указать четырёхэлементное подмножество семиэлементного множества. А это можно сделать $\binom{7}{4}$ способами. Значит, и маршрутов из $a1$ в $e4$ будет столько же. То же самое, конечно, будет и для любой другой клетки доски: число маршрутов в эту клетку равно числу сочетаний $\binom{n+m}{n}$, где n и m — числа ходов вправо и вверх.

В нашем рассуждении есть некоторая несимметрия: почему мы говорили о позициях букв П, а не букв В? Можно было бы и наоборот, тогда получались бы трёхэлементные подмножества семиэлементного множества, но их столько же (как раз об этом соответствии между множеством и его дополнением мы говорили выше).

Задача 2.47. Сколько различных слов можно составить из 10 букв А и 15 букв Б, если требуется, чтобы буквы А не шли подряд (а всегда были бы разделены хотя бы одной буквой Б)? (Ответ: $\binom{16}{10}$.)

Зная, что число сочетаний равно количеству слов с данным числом букв В и П, можно обобщить задачу и спросить: пусть у нас есть три буквы, а не две, сколько тогда будет слов с данными количествами букв? Сколько, скажем, есть *десятизначных чисел, в которых две тройки, три пятёрки и пять двоек*? (Здесь вместо букв у нас цифры, но какая разница.) Другая переформулировка того же вопроса: сколько различных слов можно получить, переставляя буквы в слове ААБББВВВВВ?

Для этого количества тоже есть обозначение, аналогичное числам сочетаний: $\binom{10}{2,3,5}$. Сверху стоит общее число букв, а снизу через запятую идут количества букв всех трёх сортов. (Сумма чисел снизу должна быть равна числу сверху, иначе это обозначение не имеет смысла.) Оказывается, что для этого числа тоже есть формула с факториалами:

$$\binom{10}{2,3,5} = \frac{10!}{2!3!5!}$$

(проверьте, что при двух числах в нижней строке эта новая формула даёт то же, что и раньше). Как это доказать?

Временно снабдим буквы в слове ААБББВВВВВ нижними индексами, чтобы их можно было различать, получится слово $А_1А_2Б_1Б_2Б_3В_1В_2В_3В_4В_5$ В нём уже все десять букв различны, поэтому можно выписать 10! перестановок этих букв. Все эти перестановки будут разными, если учитывать индексы у букв, но после стирания индексов получатся группы из одинаковых слов (которые раньше отличались только индексами).

Сколько слов в такой группе? Они все получаются, если по-разному расставлять индексы. Сколькими способами это можно сделать? В слове две буквы А, и нужно одну из них назвать $А_1$, а другую $А_2$, это можно сделать двумя способами. Букв Б у нас три, и есть $6 = 3!$ способов снабдить их индексами. Для букв В получается 5! вариантов. По правилу произведения всего способов «объиндексить» данное слово будет 2!3!5!, то есть в каждой группе по 2!3!5! слов. Соответственно, число групп равно 10!/(2!3!5!), как мы и говорили.

Ясно, что конкретный выбор чисел $10, 2, 3, 5$ не играет роли, аналогичная формула верна для других чисел и для другого количества букв:

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

Задача 2.48. Имеется колода из 36 различных карт. Эти карты раздают игрокам, по 9 карт каждому. Сколькими способами это можно сделать? (Способы считаются одинаковыми, если в них любой из четверых игроков получил одни и те же карты — но, возможно, в разном порядке.)

2.8.6 Соседние числа в строке

Глядя на треугольник Паскаля, мы видим, что числа сначала растут (до середины), а потом начинают симметрично убывать — график имеет горб в середине.[7] Как это

[7]Этот горб называют биномиальным распределением, он при разных n имеет примерно один и тот же вид и близок к масштабированному графику $y = e^{-x^2}$, называемому *нормальным распре-*

доказать?

Для этого вычислим отношение соседних чисел в строке. Это можно сделать, исходя из формулы

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

Что происходит с этой дробью, когда k увеличивается на 1? При этом в числителе появляется новый множитель $(n-k)$, а в знаменателе новый множитель $k+1$, так что

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

Это частное больше 1, если $n-k > k+1$, или $k+(k+1) < n$, то есть если середина между k и $k+1$ левее середины строки $n/2$. Так что при чётном n есть максимальное среднее число $\binom{n}{n/2}$, а при нечётном n есть два максимальных числа $\binom{n}{(n-1)/2}$ и $\binom{n}{(n+1)/2}$.

Интересно, что отношение соседних чисел можно вычислить, и не применяя формулы для чисел сочетаний. Мы хотим сравнить число k-элементных и $(k+1)$-элементных подмножеств n-элементного множества. Нарисуем картинку (двудольный граф): слева изобразим точками все k-элементные подмножества, а справа все $(k+1)$-элементные подмножества. Нас интересует соотношение количества точек слева и справа.

Чтобы найти его, соединим линиями точку X слева и точку Y справа, если $X \subset Y$, то есть Y получается из X добавлением одного элемента (ну а X получается из Y удалением одного элемента). Посмотрим, сколько линий выходит из одной точки слева и справа. К данному множеству из k элементов можно присоединить любую из $n-k$ оставшихся точек, поэтому слева из каждой точки выходит $n-k$ линий. Из данного множества с $k+1$ элементами можно удалить любой из них, поэтому справа из каждой точки выходит $k+1$ линий. Это определяет соотношение между левой и правой частями: если каждый мальчик в классе поздравил двух девочек, а каждая девочка получила поздравление от трёх мальчиков, значит, в классе в полтора раза больше мальчиков, чем девочек. Более формально, подсчитаем число левых концов линий и правых концов линий, их поровну, отсюда получаем пропорцию между количествами точек.

Рассматривая этот граф и применяя теорему Холла о представителях, можно вывести интересное следствие:

Задача 2.49. Покажите, что при $k < n/2$ можно к каждому k-элементному подмножеству n-элементного множества так добавить по элементу, что все получившиеся $(k+1)$-элементные множества будут разные. При $k > n/2$ из каждого k-элементного множества можно так выбросить по элементу, что все полученные $(k-1)$-элементные множества будут разные.

делением. Факт этой близости составляет простейший вариант *центральной предельной теоремы* из теории вероятностей.

Задача 2.50. (Продолжение) Покажите, что все подмножества $2n$-элементного множества можно разбить на $\binom{2n}{n}$ цепочек так, что в каждой цепочке соседние множества отличаются добавлением/удалением элемента.

Задача 2.51. (Продолжение) Покажите, что если в некотором семействе подмножеств множества из $2n$ элементов любые два элемента несравнимы по включению, то есть ни один не является подмножеством другого, то в этом семействе не больше $\binom{2n}{n}$ множеств.

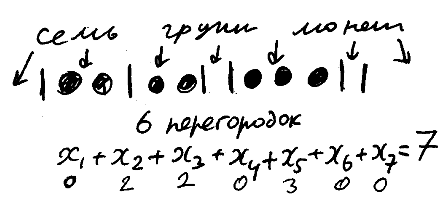
2.8.7 Монеты и перегородки

Есть ещё одна задача, где естественным (но не вполне очевидным) образом возникают числа сочетаний. *Сколько решений имеет уравнение* $x_1 + x_2 + \ldots + x_k = n$ *в целых неотрицательных числах*? Если это недостаточно наглядно, можно спросить иначе: есть k различных человек и n одинаковых монет; сколькими способами можно раздать этим людям эти монеты? Каждый такой способ определяется тем, сколько монет получил каждый (но какие именно монеты — не учитывается, все монеты одинаковые). Это целые неотрицательные числа (допускаются варианты, где некоторые получили по нулю монет).

Задача 2.52. А если бы все монеты были разными и мы учитывали не только, сколько монет досталось, но и то, какие достались — тогда сколько было бы способов? (Ответ: n^k.)

Математики бы сказали, что мы ищем число целых точек в симплексе $x_1 + x_2 + \ldots + x_k = n$, $x_i \geq 0$. При $k = 3$ это получается треугольник (и такие числа называются «треугольными»), при $k = 4$ это тетраэдр и так далее.

Как найти это число? Представим себе, что наши n монет разложены в ряд. Прежде чем раздавать эти монеты, давайте разделим их на k групп перегородками, и договоримся, кому идёт самая левая группа, кому вторая слева и т. д. Заметим, что мы допускаем случай, когда две перегородки оказываются рядом — это значит просто, что человеку, которому была назначена группа между ними, не повезло и в этом раскладе ему ни одной монеты не достанется.



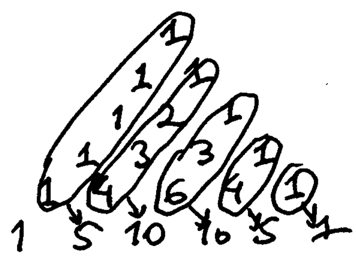
Каждому варианту раздачи (каждому решению уравнения в неотрицательных числах) соответствует последовательность из n монет и $k-1$ перегородок. (Число перегородок на единицу меньше числа группы: первая группа стоит слева от первой перегородки, а последняя — справа от последней.) Наоборот, каждой последовательности из n монет и $k-1$ перегородок соответствует некоторый способ раздачи монет. Поэтому надо подсчитать число способов расставить перегородки.

А это совсем просто — каждый такой способ можно рассматривать как слово в алфавите $\{\text{монета}, \text{перегородка}\}$, содержащее n монет и $k-1$ перегородок. Это количество, как мы знаем, равно $\binom{n+k-1}{n}$ (или $\binom{n+k-1}{k-1}$).

Мы искали решения уравнения в неотрицательных целых числах. А что будет, если искать лишь *решения уравнения в положительных целых числах*, то есть рассматривать условия $x_1 + \ldots + x_k = n$, $x_i > 0$? Легко сообразить, что получится примерно та же самая задача, только симплекс немного уменьшится. В терминах монет мы теперь требуем, чтобы каждый человек получил хотя бы одну монету — так выдадим каждому по монете заранее, до основной раздачи. Тут никакого выбора нет, а после этого остаётся раздать $n-k$ монет по-старому. Получаем ответ $\binom{n-1}{n-k}$ или $\binom{n-1}{k-1}$.

В терминах уравнений: мы вводим новые переменные $y_i = x_i - 1$. Тогда $y_1 + \ldots + y_k = n - k$ и $y_i \geq 0$, так что задача сводится к предыдущей с заменой n на $n-k$.

Задача 2.53. Для числа решений уравнения $x_1 + \ldots + x_k = n$ в неотрицательных целых числах можно написать рекуррентную формулу: последняя переменная может быть равна $0, 1, 2, \ldots, n$, поэтому искомое число равно сумме чисел решений такого же уравнения с $k-1$ переменными и числами $n, n-1, n-2, \ldots, 0$ в правой части. Покажите, что в терминах чисел сочетаний получается такое тождество: сумма чисел на каждой диагонали (см. рисунок) равна числу, стоящему снизу справа под ней. Как объяснить это тождество непосредственно, не вспоминая про уравнения?



2.9 Бином Ньютона и производящие функции

Одно из открытий в перечислительной комбинаторике состоит в том, что «считать нужно не числами, а многочленами» (быть может, многочленами бесконечной степени — степенными рядами). Это значит, что из последовательности чисел $a_0, a_1, a_2 \ldots$ нужно собрать один математический объект — многочлен $a_0 + a_1x + a_2x^2 + \ldots$ и работать с ним.

Это звучит заманчиво, но едва ли понятно, и мы начнём издалека — с бинома Ньютона, названного в честь того самого Ньютона, который изобрёл параллельно с Лейбницем математический анализ, открыл законы механики и закон всемирного тяготения, изучал преломление света в призме, наблюдал кольца Ньютона при контакте выпуклого стекла с плоским, занимался алхимией, выслеживал фальшивомонетчиков, будучи начальником монетного двора, и многое другое, вот только «бином Ньютона» был известен до него (в частности, Паскалю), а Ньютон обобщил его на случай нецелых показателей и бесконечных рядов.

Хорошо известны формулы раскрытия скобок

$$\begin{aligned}(a+b)^2 &= a^2+2ab+b^2\\ (a+b)^3 &= a^3+3a^2b+3ab^2+b^3\end{aligned}$$

(а если и неизвестны, их легко проверить). Оказывается, что есть *формула бинома* для любой степени:

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \ldots + \binom{n}{k}a^{n-k}b^k + \ldots + \binom{n}{n}b^n$$

Или, с употреблением знака суммы:

$$(a+b)^n = \sum_k \binom{n}{k}a^{n-k}b^k = \sum_k \frac{n!}{k!(n-k)!}a^{n-k}b^k.$$

Собственно говоря, сразу ясно, что после раскрытия скобок в $(a+b)(a+b)\ldots(a+b)$ (n раз) получатся произведения n букв (сколько-то a, остальные b), и вопрос только в том, какие будут коэффициенты при этих произведениях (сколько подобных членов). Так вот, формула бинома Ньютона и говорит, какие это будут коэффициенты: это числа сочетаний, и они написаны в n-й строке треугольника Паскаля. Поэтому числа сочетаний также называют *биномиальными коэффициентами*.

Это можно доказать разными способами. Если мы раскроем скобки в $(a+b)^n$ и не будем приводить подобные члены, переставляя сомножители, то получится 2^n слагаемых — всевозможные слова длины n из букв a и b. (В каждой скобке можно взять a или b = на каждом месте в слове может стоять a или b.) Если сгруппировать члены по степеням, то соберутся все слова, в которых данное число букв a и данное число букв b. А сколько их, мы знаем — это как раз числа сочетаний. Слова, в которых $n-k$ букв a и k букв b, войдут в количестве $\binom{n}{k}$ штук в слагаемое $\binom{n}{k}a^{n-k}b^k$, как и предсказывает бином Ньютона.

Другое доказательство — по индукции. Допустим, у нас уже есть формула

$$(a+b)^3 = a^3+3a^2b+3ab^2+b^3,$$

и мы хотим из неё получить формулу для $(a+b)^4$. Для этого надо домножить правую часть на $a+b$:

$$\begin{aligned}(a+b)^4 = (a+b)^3(a+b) &= (a^3+3a^2b+3ab^2+b^3)(a+b) = \\ &= (a^3+3a^2b+3ab^2+b^3)a + \\ &\qquad + (a^3+3a^2b+3ab^2+b^3)b = \\ &= a^4+3a^3b+3a^2b^2+ab^3+ \\ &\qquad + a^3b+3a^2b^2+3ab^3+b^4 = \\ &= a^4+4a^3b+6a^2b^2+3ab^3+b^4.\end{aligned}$$

Видно, что коэффициент в новой формуле получается как раз из двух соседних коэффициентов старой ($3+1=4$, $3+3=6$ и так далее), как в треугольнике Паскаля, и из третьей строки треугольника получается четвёртая. Ясно, что так же будет и дальше.

Что даёт нам эта формула? Сразу же можно получить несколько забавных следствий. Давайте подставим $a=b=1$. Тогда слева получится 2^n, а справа — сумма всех биномиальных коэффициентов, то есть всех чисел в n-ой строке треугольника Паскаля. Это равенство мы уже доказывали (двумя способами — комбинаторно и по индукции), а теперь мы его вывели как простое следствие формулы бинома.

Можно подставить $a=1, b=-1$. Тогда слева получится нуль, а справа получится знакопеременная сумма биномиальных коэффициентов (с плюсом будут те, где степень b чётна, то есть каждый второй). И это мы уже доказывали.

Чтобы получить что-то новое, подставим $a=1, b=2$. Получится, что

$$3^n = \binom{n}{0} + \binom{n}{1}2 + \binom{n}{2}2^2 + \binom{n}{3}2^3 + \ldots \binom{n}{n}2^n.$$

Вроде такого у нас не было — можно радоваться, что бином Ньютона мы доказали не зря. Впрочем, скептики скажут, что это можно доказать и комбинаторно. Давайте подсчитаем, сколько можно составить n-буквенных слов в алфавите $\{a,b,c\}$. Их, конечно, 3^n, но будем считать по-другому, отдельно рассматривая слова с разным количеством букв a. Тех, где n букв a, всего одно. Освободим одно место для букв b/c. Это можно сделать $\binom{n}{1}$ способами, а после этого есть два варианта вписать на это место b или c. Если для букв b/c оставить два места, то это можно сделать $\binom{n}{2}$ способами, а потом на эти места вписать буквы $2^2=4$ способами. И так далее — как раз в соответствии с правой частью формулы.

Можно использовать бином Ньютона и более изысканным образом. Для начала оставим только одну переменную, положив $a=1$, а вместо b будем писать x, так привычнее. Получится такая формула:

$$(1+x)^n = \sum_k \binom{n}{k} x^k = 1 + nx + \frac{n(n-1)}{2}x^2 + \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}x^3 + \ldots$$

Имея эту формулу, мы можем с ней делать разные странные вещи. Например, можно её продифференцировать (почему бы и нет? это просто равенство двух многочленов). Получится

$$n(1+x)^{n-1} = \sum_k k\binom{n}{k}x^{k-1}.$$

А вот теперь подставим единицу, получится ещё одна формула, которой у нас не было:

$$\sum_k k\binom{n}{k} = n2^{n-1}.$$

Задача 2.54. Дайте комбинаторное доказательство этого тождества, начав так: пусть мы хотим выбрать в классе из n человек команду из k человек с капитаном. Мы можем сначала выбрать капитана, а потом решать, кого ещё мы включим в команду, а кого нет, а можем сначала выбрать команду…

Вот ещё одно тождество с числами сочетаний, которое сразу следует из бинома Ньютона. Начнём с очевидного равенства

$$(1+x)^k(1+x)^l = (1+x)^{k+l},$$

в котором левая и правая части — равные многочлены. Значит, как известно из алгебры, у них равные коэффициенты. Коэффициент при x^n равен $\binom{k+l}{n}$, согласно биному Ньютона. А в левой части два многочлена перемножаются, и потому коэффициент при x^n складывается из нескольких произведений (когда степени в сумме дают n). Должно получиться одно и то же, и поэтому мы заключаем, что

$$\binom{k+l}{n} = \binom{k}{0}\binom{l}{n} + \binom{k}{1}\binom{l}{n-1} + \binom{k}{2}\binom{l}{n-2} + \ldots + \binom{k}{n}\binom{l}{0}.$$

Впрочем, и комбинаторно это совсем просто. Пусть множество состоит из двух частей: в одной k элементов, в другой l. Чтобы выбрать n элементов из этого множества, нужно выбрать сколько-то, пусть i, элементов из первой части и $n-i$ элементов из второй. При данном i это можно сделать $\binom{k}{i}\binom{l}{n-i}$ способами, остаётся сложить ответы для разных i и получить то самое тождество.

Ещё одно применение бинома Ньютона: убедимся, что *при больши́х n величина 1.01^n больше, чем n*. В самом деле, разложим $(1+0.01)^n$ по формуле бинома, там все члены положительны, так что можно оставить только второй член:

$$(1+0.01)^n > \frac{n(n-1)}{2}\cdot 0.0001$$

При больши́х n (точно: когда $(n-1)/2\cdot 0.0001 > 1$) будет $1.01^n > n$.

Задача 2.55. Докажите, что $1.01^n > n^2$ при всех достаточно больши́х n.

До сих пор мы имели дело с конечными суммами (многочленами). Но и для бесконечной последовательности $a_0, a_1, a_2, a_3, \ldots$ можно написать выражение

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots,$$

которое называется её *производящей функцией.* Это выражение можно рассматривать по-разному, или просто как «формальный ряд», с которым можно обращаться по формальным правилам, или в самом деле как функцию действительного или комплексного переменного (и тогда, если ряд бесконечный, возникает вопрос о сходимости). На первый взгляд это странно (какая разница, говорить ли просто о последовательности чисел, или о последовательности коэффициентов многочлена или ряда), но дело в том, что алгебраические операции приобретают комбинаторный смысл или наоборот.

Например, определение чисел Фибоначчи ($F_0 = 1$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$ при $n \geq 2$) в переводе на язык рядов говорит, что ряд

$$\varphi(x) = F_0 + F_1x + F_2x^2 + F_3x^3 + \ldots = 1 + x + 2x^2 + 3x^3 + 5x^4 + 8x^5 + \ldots$$

удовлетворяет уравнению

$$\varphi(x) = x\varphi(x) + x^2\varphi(x) + 1$$

(при складывании последовательности коэффициентов, сдвинутой вправо на 1 и на 2, получается исходная последовательность, кроме первого члена). Действуя по формальным алгебраическим правилам, можно заключить, что

$$\varphi(x) = \frac{1}{1 - x - x^2},$$

затем разложить знаменатель на множители и представить дробь в виде суммы, заметив, что

$$\frac{1}{(x-a)(x-b)} = \frac{1}{a-b}\left(\frac{1}{x-a} - \frac{1}{x-b}\right),$$

а потом снова перейти к рядам, используя формулу для суммы бесконечной геометрической прогрессии

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots.$$

Это даст (уже известную нам) формулу для чисел Фибоначчи (и все эти действия можно надлежащим образом узаконить).

Ещё несколько примеров: в произведении

$$(1 + x + x^2 + x^3 + x^4 + \ldots)(1 + x^2 + x^4 + x^6 + x^8 + \ldots)(1 + x^5 + x^{10} + x^{15} + \ldots)$$

коэффициент при x^n равен числу способов заплатить n рублей, имея монеты в 1, 2 и 5 рублей: он складывается из произведений x^k из первой скобки, x^{2l} из второй скобки и x^{5m} из третьей, и каждое такое произведение соответствует варианту с k рублёвыми, l двухрублёвыми и m пятирублёвыми монетами. А если мы перемножим

$$(1 + x)(1 + x^2)(1 + x^3)(1 + x^4)(1 + x^5)\ldots,$$

то это будет соответствовать ситуации, когда у нас есть монеты всех достоинств (рубль, два, три,…), но по одной штуке, то есть коэффициент при x^n будет равен *количеству разбиений числа n в сумму различных слагаемых.* Теперь в этом произведении можно сгруппировать члены: сначала выделить степени двойки

$$(1+x)(1+x^2)(1+x^4)(1+x^8)\ldots,$$

затем степени вида $3\cdot 2^k$, то есть

$$(1+x^3)(1+x^6)(1+x^{12})(1+x^{24})\ldots,$$

затем степени вида $5\cdot 2^k$, то есть

$$(1+x^5)(1+x^{10})(1+x^{20})(1+x^{40})\ldots,$$

и так далее для всех нечётных чисел (ведь каждое целое положительное число однозначно представляется в виде $u2^v$ с нечётным u). Если раскрыть скобки в первом произведении, то получится просто

$$1+x^2+x^3+x^4+x^5+\ldots,$$

поскольку каждое число однозначно представляется в виде суммы степеней двойки (двоичная система). Во втором произведении то же самое, только x надо заменить на x^3, получится

$$1+x^3+x^6+x^9+x^{12}+\ldots;$$

третье даст

$$1+x^5+x^{10}+x^{15}+x^{20}+\ldots,$$

а вместе получится произведение

$$(1+x^2+x^3+x^4+x^5+\ldots)(1+x^3+x^6+x^9+x^{12}+\ldots)(1+x^5+x^{10}+x^{15}+x^{20}+\ldots)\cdot\ldots,$$

в котором, если приглядеться, коэффициент при x^n равен *количеству разбиений n в сумму нечётных слагаемых* (не обязательно различных). Таким образом, наши действия (если считать их законными) доказывают, что *количество разбиений числа n в сумму нечётных слагаемых равно количеству его разбиений в сумму различных слагаемых.* Попробуем, скажем, число 10. Вот его разбиения в сумму нечётных слагаемых:

$$\begin{aligned}
&9+1,\\
&7+3, 7+1+1+1,\\
&5+5, 5+3+1+1, 5+1+1+1+1+1,\\
&3+3+3+1, 3+3+1+1+1+1, 3+1+1+1+1+1+1+1,\\
&1+1+1+1+1+1+1+1+1+1.
\end{aligned}$$

(в каждом разбиении слагаемые мы записываем от больших к меньшим). А вот разбиения в сумму различных слагаемых:

$$\begin{aligned}
&10,\\
&9+1,\\
&8+2,\\
&7+3, 7+2+1,\\
&6+4, 6+3+1,\\
&5+4+1, 5+3+2,\\
&4+3+2+1.
\end{aligned}$$

Видно, что разбиения совершенно разные, но чудесным образом их количество оказывается одинаковым — и это чудо объясняется волшебными манипуляциями с производящими функциями (которые вполне законны, если разобраться). В данном случае механизм волшебства не так сложен, и это рассуждение можно проследить шаг за шагом и превратить в комбинаторное доказательство.

Задача 2.56. Сделайте это и укажите взаимно однозначное соответствие между разбиениями n в сумму нечётных слагаемых и в сумму различных слагаемых. (Ответ: в разбиении на различные слагаемые нужно каждое чётное число заменять на две его половины, пока не останутся одни нечётные. В другую сторону нужно заменять два одинаковых слагаемых на их сумму.)

Вернёмся к Ньютону: он понял, что эта же формула верна для дробных показателей степени, но с двумя оговорками. Во-первых, для дробных показателей степени числа сочетаний не имеют комбинаторного смысла, и надо просто пользоваться формулой в том виде, в котором она имеет смысл:

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{1\cdot 2}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{1\cdot 2\cdot 3}x^3 + \ldots$$

(вместо n мы написали α в знак того, что теперь n не обязательно целое). Во-вторых, при нецелом α получается бесконечно много членов (потому что множители в числителе не обращаются в нуль, как раньше), и не очень понятно, что эта бесконечная сумма означает.

Здесь мы уже переходим в область математического анализа, но скажем коротко о некоторых вариантах такого понимания. Во-первых, при $|x| < 1$ можно понимать сумму как предел (берём всё больше и больше членов, и смотрим, к чему это сходится), и этот предел будет правильным. Во вторых, можно взять написать приближённую формулу с конечным числом членов, например,

$$(1+x)^\alpha \approx 1 + \alpha x.$$

Про эту формулу можно доказать, что при малых x ошибка (разница между левой и правой частью) мала, и не просто мала, а мала даже по сравнению с x. В учебниках математического анализа пишут $(1+x)^\alpha = 1 + \alpha x + o(x)$. Если взять следующий член, то ошибка будет $o(x^2)$. Для доказательства нужно использовать формулу Тейлора (с остаточным членом в форме Пеано); другой вариант формулы Тейлора (с остаточным членом в форме Лагранжа) позволяет доказать формулу бинома с рядом.

Задача 2.57. Проверьте, что для $\alpha = 1/2$ формула бинома Ньютона даёт

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} + \ldots + \frac{(-1)^n(2n)!}{(1-2n)(n!)^2 4^n} x^n + \ldots$$

(Указание: произведение нечётных чисел $1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)$, иногда обозначаемое $(2n-1)!!$, можно вычислить домножив его на $2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2n) = 2^n n!$.)

В частности, получается хорошо известная приближённая формула $\sqrt{1+x} \approx 1 + \frac{x}{2}$ (попробуйте вычислить корень из 1.000001 на калькуляторе).

Вот ещё одно применение биномиальных коэффициентов с нецелыми числами. В формуле бинома Ньютона для произвольного α мы рассматривали выражение $\binom{\alpha}{k}$ при произвольном α, определив его как

$$\frac{\alpha(\alpha-1) \cdot \ldots \cdot \cdot (\alpha - k + 1)}{1 \cdot 2 \cdot \ldots \cdot k}.$$

Заметим, что при фиксированном k это выражение (как функция α) является многочленом. Обозначим его через $P_k(\alpha)$. Степень этого многочлена равна k, и он принимает целые значения в целых точках (в положительных точках это числа сочетаний, в отрицательных — тоже, только знаки меняются).

Складывая многочлены P_k при различных k с разными целыми коэффициентами, мы получаем многочлен, который в целых точках принимает целые значения. Оказывается, что верно и обратное: *любой многочлен от одной переменной, который во всех целых точках принимает целые значения, можно представить как сумму многочленов P_k с целыми коэффициентами.*

Чтобы доказать этого, введём такое определение: для любого многочлена $P(x)$ рассмотрим его «дискретную производную» $Q(x) = P(x+1) - P(x)$. (Если рассматривать только значения в целых точках, то дискретная производная — это последовательность разностей между соседними членами.) Например, для многочлена x^2 дискретная производная равна $2x+1$ (разности между точными квадратами образуют арифметическую прогрессию), а для многочлена x^3 равна $3x^2 + 3x + 1$ (разложение по биному без первого члена, аналогично и для x^n). Степень дискретной производной на единицу меньше степени многочлена (почему?). Если дискретная производная равна нулю, то многочлен — константа.

Ключевое наблюдение: *дискретная производная многочлена P_{n+1} равна P_n*, то есть

$$P_n(\alpha) = P_{n+1}(\alpha+1) - P_n(\alpha).$$

В самом деле, при целых α это превращается в

$$\binom{\alpha}{n} = \binom{\alpha+1}{n+1} - \binom{\alpha}{n+1},$$

(рекуррентное соотношение для чисел сочетаний), а из алгебры известно, что если многочлены совпадают в бесконечном числе точек (в данном случае в целых α, начиная с $n+1$) , то они равны.

Теперь можно рассуждать по индукции. Пусть некий многочлен P принимает в целых точках целые значения. Если его степень равна нулю, то он константа, и всё ясно. Если его степень больше нуля, то его дискретная производная P' имеет меньшую степень и по предположению индукции представляется в виде целочисленной комбинации наших многочленов. Заменив каждый многочлен P_i в этой комбинации на его «дискретную первообразную» P_{i+1}, мы получим многочлен, у которого дискретная производная такая же, как у P, и потому у их разности дискретная производная равна нулю, значит, эта разность константа (и целая, потому что это разность двух многочленов с целыми значениями). Остаётся только добавить эту константу к целочисленной комбинации и получить P.

Из этого наблюдения возможно, в частности, получить доказательство факта со стр. 18 из лекции 1.

Для этого нам нужно ещё одно равенство, которое легко доказывается как по индукции, так и комбинаторными рассуждениями или вычислениями с производящими функциями.

Задача 2.58. Докажите, что

$$\sum_{j=0}^{n} P_k(j) = \sum_{j=0}^{n} \binom{j}{k} = \binom{n+1}{k+1} \tag{2.1}$$

С точки зрения производящих функций равенство (2.1) содержится в равенстве

$$\sum_{j=0}^{n} (1+t)^j = \frac{(1+t)^{n+1}-1}{t}$$

(сумма геометрической прогрессии).

Комбинаторное доказательство предоставляется найти читателю.

Теперь перейдём к доказательству факта со стр. 18. Выразим n^d как многочлен от n через многочлены $P_k(n)$:

$$n^d = \sum_{i=0}^{d} a_{d,i} P_i(n).$$

Тогда для суммы d-х степеней получаем

$$\sum_{j=0}^{n} j^d = \sum_{j=0}^{n}\sum_{i=0}^{d} a_{d,i} P_i(j) = \sum_{i=0}^{d} a_{d,i} \sum_{j=0}^{n} \binom{j}{i} = \sum_{i=0}^{d} a_{d,i} \binom{n+1}{i+1} = \sum_{i=0}^{d} a_{d,i} P_{i+1}(n+1).$$

В предпоследнем равенстве использовано равенство (2.1). Последнее выражение как функция от n является суммой многочленов степени не выше $d+1$, что и требовалось доказать.

2.10 Числа Каталана

2.10.1 Правильные последовательности скобок

Будем составлять из символов «`(`» и «`)`» последовательности правильно вложенных скобок. «Правильная вложенность» означает, что последовательности `((()))`, `(()())()`, `()()()` допускаются, а, скажем, последовательности `)`, `)(` или `())(` — нет. В правильной последовательности всегда равное число открывающих и закрывающих скобок, но это только необходимое, но не достаточное условие (как показывают два последних примера).

Интуитивно довольно понятно, что мы имеем в виду, говоря о «правильной вложенности» скобок, но как это объяснить формально? Одна из возможностей — дать такое индуктивное определение:

- пустая последовательность (длины 0) правильна;
- если A — правильная последовательность, то `(`A`)` — правильная последовательность;
- если A, B — правильные последовательности, то их конкатенация AB (приписываем B к A справа) — правильная последовательность.

Надо добавить ещё, что других правильных последовательностей нет, кроме тех, которые можно построить по этим правилам: правильные последовательности — это те, которые можно получить из пустой с помощью операций заключения в скобки и конкатенации. Можно сказать, что множество правильных последовательностей — это минимальное множество, обладающее этими тремя свойствами. (Заметим, что оно не единственно — скажем, множество всех последовательностей тоже ими обладает.)

Программисты записывают это определение так:

$$\langle\text{ппп}\rangle ::= \quad \mid \texttt{(}\langle\text{ппп}\rangle\texttt{)} \mid \langle\text{ппп}\rangle\langle\text{ппп}\rangle$$

Здесь $\langle\text{ппп}\rangle$ — сокращение для «правильно построенная последовательность», а черта «|» разделяет разные варианты: пустую последовательность, правильную последовательность в скобках, наконец, конкатенацию двух правильно построенных последовательностей. Такая запись называется *нормальной формой Бэкуса – Наура* (Backus–Naur Form, BNF). Иногда то же самое записывают иначе: говорят, что есть *контекстно-свободная грамматика* с правилами

$$\begin{aligned}
\langle\text{ппп}\rangle &\to \\
\langle\text{ппп}\rangle &\to \texttt{(}\langle\text{ппп}\rangle\texttt{)} \\
\langle\text{ппп}\rangle &\to \langle\text{ппп}\rangle\langle\text{ппп}\rangle
\end{aligned}$$

которые означают то же самое.

Так или иначе, мы определили понятие правильной последовательности скобок, и теперь можем доказать такой факт: *последовательность правильно построена* *тогда и только тогда, когда в ней поровну скобок, а в любом её начале не меньше левых, чем правых.*

Это кажется достаточно очевидным, но давайте это аккуратно докажем. Для удобства введём временное название для последовательностей, обладающих указанным свойством (про количество скобок): будем называть их, скажем, *квазиправильными.* Нам надо показать, что правильные и квазиправильные — это одно и то же. Сначала покажем, что всякая правильная последовательность квазиправильна. Это доказывается «индукцией по построению»[8]: надо доказать, что

- пустая последовательность квазиправильна;
- если A квазиправильна, то (A) квазиправильна;
- если A и B квазиправильны, то их конкатенация AB квазиправильна

Первое очевидно. Второе: начало последовательности (A) либо пусто, либо состоит из открывающей скобки и начала последовательности A, либо есть вся последовательность (A). В первом и третьем случаях скобок поровну, во втором добавление левой скобки только улучшает дело (используем квазиправильность A). Аналогично для конкатенации: её начало может быть либо началом A, либо всей последовательностью A плюс начало B, и в обоих случаях всё очевидно.

В обратную сторону: почему всякая квазиправильная последовательность правильна? Будем идти слева направо, смотря за превышением левых скобок над правыми. (По предположению мы знаем, что оно всё время неотрицательно, а в конце равно нулю.) Обращается ли это превышение в нуль где-то в середине? Если нет, и превышение всегда по крайней мере 1, то выделим первую и последнюю скобки. Первая будет открывающей (иначе условие нарушится сразу же), последняя закрывающая (иначе условие нарушится на предпоследнем шаге). Между ними будет квазиправильная последовательность (потому что отбрасывание первой скобки уменьшает превышение ровно на 1, и оно останется неотрицательным. По индукции заключаем, что эта внутренняя последовательность правильна, а значит, и наша последовательность правильна.

Другой случай: когда превышение обратится в ноль где-то внутри последовательности. Разрежем её в этом месте. Получатся две квазиправильные (проверьте!) последовательности меньшей длины, значит (предположение индукции) они правильны, значит, наша последовательность получена конкатенацией правильных, и потому правильна.

После всей этой подготовки мы можем наконец поставить вопрос: *сколько существует правильных последовательностей из n левых и n правых скобок*?

[8] Формально говоря, тем самым мы докажем, что множество квазиправильных последовательностей удовлетворяет трём свойствам из определения правильных, а множество правильных последовательностей было минимально среди таких.

$n=0$	1	
$n=1$	1	`()`
$n=2$	2	`(())`, `()()`
$n=3$	5	`((()))`, `(()())`, `(())()`, `()(())`, `()()()`

Задача 2.59. Убедитесь, что в этой таблице ничего не пропущено, и выпишите аналогичный список для $n = 4$ (Указание: их будет 14.)

Уже в этой задаче видно, что подсчёт правильных скобочных выражений — дело не такое простое (попробуйте написать программу, которая перечисляет их все для произвольного n). Их количество называют *n-м числом Каталана* (в честь бельгийского математика XIX века) и обозначают C_n. Эти числа удивительным образом появляются в самых разных комбинаторных задачах, и для них есть явная формула с красивыми доказательствами.

2.10.2 Рекуррентная формула

Как считать правильные скобочные выражения? Раз какого-то простого способа сразу не видно, давайте попробуем найти рекуррентную формулу. Её можно составить по грамматике. Сначала мы сделаем это неправильно (попробуйте заметить ошибку до того, как она будет указана), а потом исправим.

Итак, откуда берутся правильные выражения длины $2n$? Из двух источников: можно взять правильное выражение длины $2n-2$ и заключить его в скобки, а также можно соединить правильные выражения длин $2k$ и $2l$, где $k + l = n$ и $0 < k < n$. Первым способом получается C_{n-1} выражений; выражений второго типа будет C_kC_l (при данных k и l), и потом это надо сложить. Отсюда получаем формулу:

$$C_n = C_{n-1} + \sum_{k=1}^{n-1} C_k C_{n-k}$$

Начальное условие $C_0 = 1$ (пустое выражение). Дальше $C_1 = 1$ (сумма не содержит ни одного слагаемого), потом $C_2 = 1 + C_1C_1 = 2$, потом $C_3 = 2 + C_1C_2 + C_2C_1 = 6$ — и тут мы замечаем, что получилось не 5, как должно быть, а больше, мы что-то лишнее посчитали. Что?

* * *

На самом деле мы не то чтобы посчитали лишнее, а посчитали одно и то же выражение дважды: ведь `()()()` можно получить как конкатенацию `()()` и `()` а можно наоборот, и получится одно и то же. Как говорят, наша грамматика *неоднозначна*. Программисты сказали бы, что одно и то же выражение `()()()` имеет два «дерева разбора» в этой грамматике:

Чтобы исправить дело, мы используем другую грамматику для правильных выражений. (Это обычное дело при построении компиляторов: если грамматика из описания языка программирования не обладает хорошими свойствами, её преобразуют в другую, иногда это помогает.)

Вот эта однозначная грамматика в форме Бэкуса–Наура:

$$\langle\text{ппп}\rangle ::= \quad || \; \texttt{(}\langle\text{ппп}\rangle\texttt{)}\langle\text{ппп}\rangle$$

Надо только понять, почему она задаёт тот же самый язык (то же множество правильных выражений) и почему она однозначна.

Заметим, что по этой грамматике получаются только правильные выражения, так как действие $A, B \mapsto \texttt{(}A\texttt{)}B$ состоит из двух разрешённых (окружение скобками и конкатенация). Почему все? Это можно понять, вспомнив рассуждение с разностью левых и правых скобок. Первая скобка левая, после неё эта разность равна 1. Посмотрим на то место, где разность впервые обращается в нуль. (Возможно, это конец слова, если раньше такого не случилось.) Перед ним стоит правая скобка, и внутри этой пары скобок разность 1 или больше. Значит, там стоит правильное выражение (индукция). Оставшаяся часть является правильным выражением, поскольку отрезанное начало на разность не влияет и можно снова воспользоваться предположением индукции.

Почему грамматика однозначна? Потому что место разреза однозначно определяется разностью левых и правых скобок: до этого места она положительна, а здесь обращается в нуль.

Теперь, глядя на эту грамматику, можно написать правильную формулу:

$$C_n = \sum_{k+l=n-1, k,l\geq 0} C_k C_l$$

и по-прежнему $C_0 = 1$. По этой формуле $C_1 = C_0C_0 = 1$, $C_2 = C_0C_1 + C_1C_0 = 2$, $C_3 = C_0C_2 + C_1C_1 + C_2C_0 = 5$, так что в этот раз всё идёт как надо.

Используя эту рекуррентную формулу, можно легко вычислять числа Каталана. Но удивительным образом для них есть более явная формула:

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

Скажем, для C_3: эта формула даёт

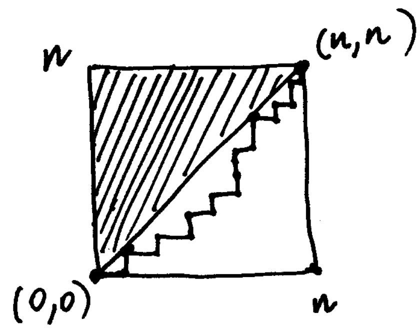
$$\frac{1}{4}\binom{6}{3} = \frac{6\cdot 5\cdot 4}{(1\cdot 2\cdot 3)\cdot 4} = 5,$$

всё в порядке.

У этой формулы есть множество доказательств, но ни одно из них, пожалуй, нельзя назвать простым и естественным. Тем не менее они красивы и поучительны, так что мы приведём несколько из них.

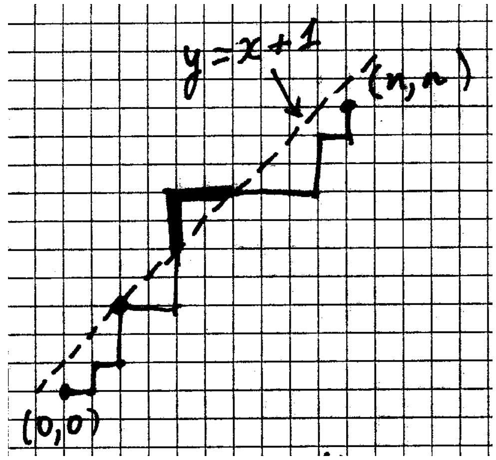
2.10.3 Вычисление с помощью отражений

Удобно представлять себе последовательности скобок как ломаные. Начав с $(0,0)$, мы изображаем открывающую скобку как шаг вправо на 1, а закрывающую — как шаг вверх. (Конечно, можно было бы наоборот.) Если у нас n левых и n правых скобок, то мы заканчиваем движение в точке (n,n). Осталось понять, что означает правильность скобочного выражения. Как мы знаем, она равносильна тому, что в любом начальном отрезке выражения (=пути) левых скобок (=шагов вправа) не меньше, чем правых. Геометрически это значит, что мы никогда не попадаем выше отрезка $y=x$, соединяющего $(0,0)$ с (n,n). Другими словами, нам надо узнать, *сколько есть путей* $(0,0)$ *в* (n,n)*, не попадающих внутрь верхней половины квадрата.* (Говоря о путях, мы имеем в виду последовательность единичных шагов вправо и вверх; можно было бы говорить о «монотонных путях», имея в виду, что влево и вниз идти нельзя.)



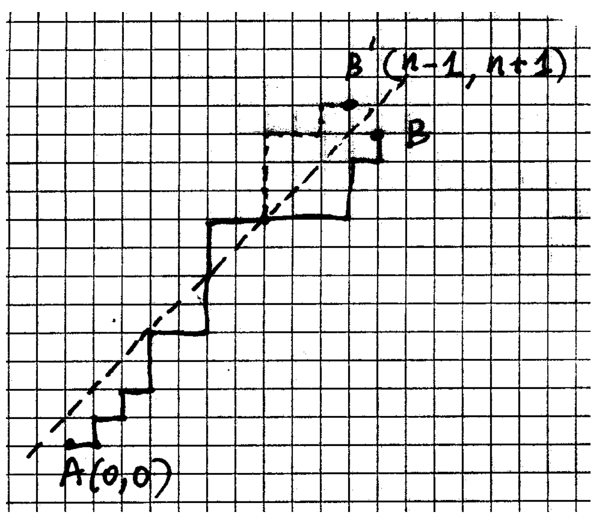
Предполагаемый ответ (который нам надо доказать) гласит, что они составляют $1/(n+1)$-ю долю общего числа путей (которое, как мы знаем, равно $\binom{2n}{n}$). Можно сказать, что если мы рассмотрим случайный путь, то скорее всего (с вероятностью $1-1/(n+1)$) попадём внутрь верхнего треугольника, но всё же с некоторой небольшой вероятностью $(1/(n+1)$ не выйдем из нижнего. (На границу заходить разрешается). Специалисты по теории вероятностей сказали бы, что *если в конце серии из* $2n$ *игр в орлянку мы остались при своих, то вероятность того, что мы ни разу не уходили в минус, равна* $1/(n+1)$.

Но как это доказать? Будем считать плохие пути — те, которые поднимаются выше прямой $y=x$, или, другими словами, доходят до прямой $y=x+1$ (пересекают её или касаются). На рисунке показана эта прямая и один плохой путь: сначала он её касается, а потом даже пересекает (эти участки нарисованы жирнее).



Есть такая задача для школьников: даны две точки A и B с одной стороны прямой реки, надо найти кратчайший путь из A в B с заходом на реку («за водой»). Она решается так: участок пути от набора воды до точки B мы отражаем симметрично относительно реки, получается путь из A в симметричную точку B'. Он имеет ту же длину, и чтобы сделать эту длину минимальной, надо идти по отрезку AB' (который заведомо пересекает реку, так как A и B' по разные стороны).

К чему все эти разговоры? Сделаем примерно то же самое в нашей задаче. А именно, для каждого плохого пути посмотрим на последний момент, когда он попал на запрещённую прямую, и участок пути после этого момента отразим симметрично относительно этой запрещённой прямой:



Получится путь из точки $A(0,0)$ в $B'(n-1,n+1)$; он тоже будет монотонным, так как шаги вправо после симметрии становятся шагами вверх и наоборот. Теперь ключевое наблюдение: *это преобразование задаёт взаимно однозначное соответствие* *между плохими путями и всеми путями из A в B'*. В самом деле, для нового пути последний момент пересечения с прямой тот же самый, что и для старого, поэтому старый путь можно получить из нового, отразив симметрично участок после этого момента. (Кстати, это останется верным, если мы изменим рассуждение и будем рассматривать первый момент пересечения с запрещённой прямой, а не последний.) И любой путь из A в B' волей-неволей пересекает прямую, так что получатся они все.

Остаётся подсчитать: плохих путей $\binom{2n}{n-1}$, а всего путей $\binom{2n}{n}$, так что надо вычислить разность

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n-1)!(n+1)!} = \frac{(2n)!}{n!n!}\left(1 - \frac{n}{n+1}\right) = \frac{1}{n+1}\binom{2n}{n}$$

(Заметим в скобках, что последнее вычисление, независимо от всяких чисел Каталана, показывает, что средний элемент в $2n$-й строке треугольника Паскаля делится на $n+1$, что не сразу очевидно.)

2.10.4 Вычисление с производящей функцией

Ну хорошо, мы нашли изящный приём, позволяющий доказать формулу для чисел Каталана. Но есть ли какой-то более систематический подход, позволяющий получать подобные формулы? Оказывается, что производящие функции тут могут помочь. Вот соответствующее рассуждение (в котором все действия на самом деле могут быть узаконены).

Начнём с того, что поймём, что означает рекуррентное соотношение

$$C_n = \sum_{k+l=n-1,k,l\geq 0} C_k C_l \quad \text{при } n \geq 1$$

в терминах производящей функции, то есть ряда

$$c(x) = C_0 + C_1x + C_2x^2 + C_3x^3 + \ldots = 1 + x + 2x^2 + 5x^3 + \ldots$$

Правая часть будет коэффициентом при x^{n-1} в произведении двух копий этого ряда, то есть в $c^2(x)$. Соответственно, чтобы получить ряд для $c(x)$, мы должны домножить это произведение на x и добавить свободный член 1:

$$c(x) = xc^2(x) + 1.$$

Это можно рассматривать как квадратное уравнение

$$xc^2(x) - c(x) + 1 = 0$$

с неизвестным значением $c(x)$ и известными коэффициентами x, -1 и 1, и написать решение по обычной формуле для квадратного уравнения:

$$c(x) = \frac{1 \pm \sqrt{1-4x}}{2x}$$

Вот только взять плюс или минус? Если взять плюс, то при $x \to 0$ правая часть стремится к бесконечности (числитель стремится к 2, а знаменатель стремится к нулю), а это нехорошо, так как $c(x)$ обращается в 1 при $x = 0$. Поэтому возьмём минус:

$$c(x) = \frac{1 - \sqrt{1-4x}}{2x}$$

Проверим, что поведение около нуля выглядит правдоподобно: $\sqrt{1-4x} \approx 1-2x$ при $x \approx 0$, так что действительно в частном получается примерно 1.

Ну и чего мы добились? мы искали $c(x)$, считая его рядом, а вместо этого получили формулу с квадратным корнем — что это даёт? На самом деле осталось совсем немного: вспомним, что бином Ньютона для показателя степени $1/2$ как раз и представляет корень в виде ряда. Правда, в той формуле для $(1+x)^{1/2}$ надо заменить x на $-4x$, как у нас теперь. Получится

$$\sqrt{1+4x} = 1 - \frac{4x}{2} - \frac{(4x)^2}{8} - \frac{(4x)^3}{16} - \ldots - \frac{(2n)!}{(2n-1)(n!)^2 4^n}(4x)^n + \ldots$$

(теперь все члены одного знака, поскольку ещё одно чередование возникает из-за минуса в $-4x$; в общей формуле поэтому в знаменателе надо изменить знак и написать $2n-1$ вместо $1-2n$). Выпишем явно первые несколько членов и сократим четвёрки в числителе и знаменателе:

$$\sqrt{1-4x} = 1 - 2x - 2x^2 - 4x^3 - \ldots - \frac{(2n)!}{(2n-1)(n!)^2}x^n + \ldots$$

Теперь по формуле корней надо вычесть это из единицы и разность поделить на $2x$. Получится

$$\frac{1-\sqrt{1-4x}}{2x} = 1 + x - x^2 + \ldots + \frac{(2n)!}{2(2n-1)(n!)^2}x^{n-1} + \ldots$$

Число Каталана C_n должно быть коэффициентом при x^n. Видно, что начинается всё правильно (с $1, 1, 2, \ldots$), а в общей формуле надо сдвинуть n на единицу:

$$C_n = \frac{(2n+2)!}{2(2n+1)((n+1)!)^2} = \frac{(2n+1)(2n+2)}{2(2n+1)(n+1)(n+1)} \cdot \frac{(2n)!}{n!n!} = \frac{1}{n+1}\binom{2n}{n},$$

что и требовалось доказать. (Последние слова двусмысленны — строго говоря, мы не доказали того, что требовалось, пока не узаконили все наши действия: операции с бесконечными рядами и применение формулы бинома для показателя $1/2$. Это можно сделать, но требует разных сведений из алгебры и анализа.)

2.10.5 Вычисление с теорией вероятностей и поворотами

В нашем первом (комбинаторном) вычислении мы считали плохие пути. Сейчас мы попробуем подсчитать хорошие, используя терминологию теории вероятностей.[9]

[9] Если вы с этим не сталкивались, то можно прочитать рассуждение просто как правдоподобную байку, но оно на самом деле вполне корректно. В конце раздела мы приводим краткий перевод рассуждения на комбинаторный язык.

Для начала будем представлять себе скобочное выражение как дорогу, в которой есть участки подъёма и спуска (на одну и ту же величину). Левой скобке соответствует подъём, а правой спуск. Последовательностям скобок длины $2n$ соответствуют дороги из $2n$ участков. Равенство тех и других скобок означает, что дорога кончается на той же высоте, что и началась, то есть что её можно считать кольцевой с отмеченным началом (старт и финиш одновременно). Правильность означает, что дорога нигде не опускается ниже своей начальной точки, и вероятность этого события для случайно взятой дороги (все последовательности считаем равновероятными) равна $1/(n+1)$ — это мы хотели бы доказать.

Ничего не изменится, если мы сдвинем точку старта на какой-то фиксированный угол у каждой дороги: распределение вероятностей на дорогах (с выделенной точкой старта) останется тем же. Поэтому и при случайном независимом выборе дороги и сдвига вероятность останется той же. Теперь заметим, что если мы выбираем для данной кольцевой дороги точку старта, то *последовательность подъёмов и спусков будет правильной в том и только том случае, когда мы начинаем с самой низкой точки дороги.* Если бы таких точек всегда была одна, то вероятность равнялась бы $1/(2n)$, но их больше, так что это рассуждение не проходит. (Зато, если знать ответ, оно показывает, что у случайной дороги в среднем $2n/(n+1)$ точек минимума, что совсем не очевидно).

Что же делать? Оказывается, что помогает следующий довольно странный трюк. Давайте ко всем последовательностям из n участков вверх и вниз добавим вначале участок вверх. Получится, что искомое число равно *числу последовательностей из $n+1$ шагов вверх и n шагов вниз, у которых все точки, кроме начальной, строго выше начальной.* (Тут надо заметить, что сформулированное условие гарантирует, что первый шаг будет вверх, так что лишних мы не получим.)

Правда, теперь все наши разговоры о кольцевой дороге утрачивают силу: у нас есть $n+1$ участков, где подъём (скажем, на 1) и n участков спуска на 1, так что в итоге после $2n+1$ участков мы оказываемся на 1 выше, чем начинали. Сделаем такую, на первый взгляд нелепую, вещь: добавим на каждом шаге спуск $1/(2n+1)$, тогда он как раз скомпенсирует разность между числом подъёмов и спусков. Мы получаем случайную кольцевую дорогу из $n+1$ подъёмов на $1-1/(2n+1)$ и n спусков на $1+1/(2n+1)$. Нам надо узнать, какова вероятность того, что мы будем всегда оставаться выше начальной точки — с учётом добавочного спуска или без, это всё равно, так как вначале подъёмы и спуски у нас были целыми числами, и добавка, не достигающая 1 на неполном круге, не может полностью уничтожить даже минимальный возможный подъём на 1. Таким образом, вопрос можно сформулировать так: какова вероятность того, что начало окажется минимумом такой кольцевой дороги? Опять же можно добавить случайный поворот, и заметить, что теперь у каждой дороги минимум единствен, так как благодаря добавке все высоты различаются (по модулю 1, а значит, и вообще). Значит, эта вероятность равна $1/(2n+1)$, а число хороших дорог (равное числу Каталана, как мы видели) есть

$$\frac{1}{2n+1}\binom{2n+1}{n} = \frac{(2n+1)!}{(2n+1)(n+1)!n!} = \frac{1}{n+1}\binom{2n}{n}.$$

Рассуждения со ссылками на теорию вероятностей — дело тонкое, в них легко не заметить какой-то ошибки. Поэтому, чтобы снять с себя подозрения, кратко перескажем то же рассуждение в комбинаторных терминах. Запишем все последовательности из $n+1$ открывающих скобок и n закрывающих в столбик. Рядом с этим столбиком напишем такой же, но только каждую последовательность циклически сдвинем на 1, затем ещё один, с циклическим сдвигом на 2, и так далее. Получим прямоугольную таблицу высоты $\binom{2n+1}{n}$ и ширины $2n+1$. В каждом столбце все комбинации встречаются по разу, значит во всей таблице каждая комбинация встречается $2n+1$ раз, и все они появляются одинаково часто. Благодаря этому доля правильных комбинаций (последовательностей, в которых любое непустое начало содержит больше открывающих скобок, чем закрывающих) во всей таблице такая же, как в первом (или любом другом) столбце. Но во всей таблице эта доля равна $1/(2n+1)$, поскольку наше рассуждение с дорогами, спусками, подъёмами и коррекциями показывает, что в каждой строке ровно одна правильная комбинация.

2.10.6 Доказательство по индукции с дробными параметрами

Если формулу для чисел Каталана мы уже знаем (пусть даже без доказательства), то возникает естественная идея: доказать её по индукции. Другими словами, мы можем определить числа

$$\tilde{C}_n = \frac{1}{n+1} \cdot \frac{2n!}{n!\,n!}$$

безотносительно к их комбинаторному смыслу, и доказать, что для них верна рекуррентная формула

$$\tilde{C}_n = \tilde{C}_{n-1} + \sum_{k=1}^{n-1} \tilde{C}_k \tilde{C}_{n-k}$$

(мы пишем знак тильды, временно различая «комбинаторные» числа Каталана, и полученные по формуле). Эта формула — некоторое тождество для чисел сочетаний, но как его доказать? В него входят только средние числа в разных строках треугольника, поэтому не очень понятно, как бы это можно было комбинаторно истолковать.

Тут помогают дробные аргументы. Несложным вычислением с факториалами можно убедиться, что

$$\binom{1/2}{n} = -\frac{(-1)^n}{2^{2n-1}} \tilde{C}_{n-1}$$

(в левой части возникает произведение всех нечётных чисел $1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-3)$, иногда обозначаемое $(2n-3)!!$; чтобы выразить его через обычные факториалы, надо домножить на произведение чётных чисел $(2n-2)!! = 2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2n-2)$, равное $2^{n-1}(n-1)!$, и получить обычный факториал).

Правда, при $n = 0$ правая часть не имеет смысла, и чтобы придать ей смысл, надо положить $C_{-1} = -1/2$. Теперь мы можем вспомнить тождество

$$\binom{k+l}{n} = \binom{k}{0}\binom{l}{n} + \binom{k}{1}\binom{l}{n-1} + \binom{k}{2}\binom{l}{n-2} + \ldots + \binom{k}{n}\binom{l}{0},$$

которое у нас было для натуральных k и l (выбирая n элементов из $k+l$, надо взять сколько-то из первых k и оставшиеся из остальных l). Но это равенство двух многочленов от k и l, и из алгебры известно, что если оно верно для всех целых чисел, то оно всегда верно. После этого запишем его для $k = l = 1/2$, правая часть обратится в нуль, и получится тождество

$$\tilde{C}_{-1}\tilde{C}_{n-1} + \tilde{C}_0\tilde{C}_{n-2} + \tilde{C}_1\tilde{C}_{n-3} + \ldots + \tilde{C}_{n-2}\tilde{C}_0 + \tilde{C}_{n-1}\tilde{C}_{-1} = 0.$$

Вспоминая о соглашении $C_{-1} = -1/2$, мы получаем искомое рекуррентное соотношение.

2.10.7 Неассоциативные произведения, триангуляции и стековый калькулятор

Мы уже описали четыре способа получения формулы для чисел Каталана — но почему именно они нас интересуют? Дело в том, что эти числа удивительным образом появляются во многих комбинаторных задачах, и мы сейчас рассмотрим два таких примера.

Пусть у нас есть m сомножителей. Вычисляя их произведение, мы можем выполнять действия в разном порядке. Обычно этот порядок указывают скобками. Например, для трёх сомножителей возможны порядка $(ab)c$ и $a(bc)$ — можно сначала перемножить a и b, а потом умножить на c, а можно умножить a на предварительно вычисленное произведение bc. Окончательный результат получится один и тот же (это свойство называют *ассоциативностью*). Для четырёх сомножителей вариантов (если не разрешать переставлять сомножители местами) будет 5:

$$((ab)c)d,\ (a(bc))d,\ (ab)(cd),\ a((bc)d),\ a(b(cd)).$$

Задача 2.60. Выпишите все варианты для пяти сомножителей (каким-то систематическим способом, чтобы ни один не пропустить) и убедитесь, что их будет 14.

Сколько вариантов возможно для m сомножителей? Наверное, вы уже догадались, что ответом к этой задаче (обозначим его P_m) будет число Каталана, только со сдвигом на единицу: $P_m = C_{m-1}$. Для нескольких первых значений m мы это видим своими глазами, но почему это будет верно и дальше?

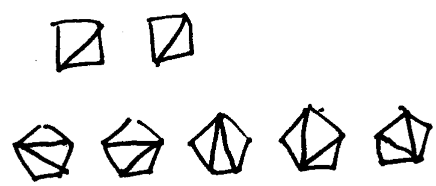
Самый простой способ это доказать — проверить, что для новой последовательности (с учётом сдвига нумерации на единицу) выполняется то же самое рекуррентное соотношение. В самом деле, пусть у нас есть m сомножителей. Посмотрим на последнюю операцию (умножение, которое даёт окончательный ответ). Она может быть в разных местах — после первого сомножителя, после второго сомножителя и так далее. Соответственно по правилу суммы надо посчитать количество вариантов каждого типа и их сложить. Для каждого варианта мы по правилу произведения перемножаем число способов слева и справа от точки деления от точки деления и справа. Скажем, для произведения P_6 мы получаем формулу

$$P_6 = P_1P_5 + P_2P_4 + P_3P_3 + P_4P_2 + P_5P_1$$

(скажем, член P_3P_3 соответствует вариантам, когда на последнем шаге мы перемножаем произведение первых трёх сомножителей и последних трёх, то есть точка разреза посередине). Если переписать это с уменьшенными на единицу индексами, то получится в точности рекуррентное соотношение для чисел Каталана, что и завершает доказательство.

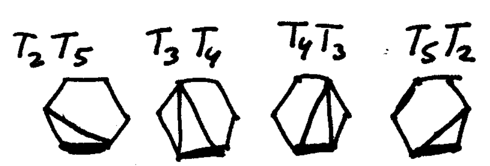
Вот ещё одна похожая задача. Рассмотрим выпуклый m-угольник, и будем разрезать его непересекающимися диагоналями на треугольные части. Можно сразу же заметить (сумма углов), что этих частей будет $m-2$, и что понадобится $m-3$ диагонали, как их ни проводи (каждая новая диагональ добавляет часть). Вопрос: *сколькими способами можно выполнить такую триангуляцию*? (Обозначим их число T_n

Для треугольника способ один, для четырёхугольника, очевидно, два (можно разрезать по любой из диагоналей). Для пятиугольников способов пять (и все они сводятся к тому, что надо из одной вершины провести все диагонали).



Задача 2.61. Выпишите все варианты для шестиугольника (каким-нибудь систематическим образом, чтобы ни одного не пропустить) и посчитайте их (должно получиться 14).

Как вы догадываетесь, мы хотим теперь доказать, что $T_m = C_{m-2}$, и понятно как: получив рекуррентную формулу для числа триангуляций. Вот как это делается. Пусть у нас есть, скажем, шестиугольник. Выберем и зафиксируем какую-то его сторону, назовём её основанием. К этому основанию должен примыкать какой-то треугольник. Его третьей вершиной может быть любая из четырёх оставшихся вершин. После того как мы выбрали положение третьей вершины, мы независимо триангулируем оставшиеся части. (В двух крайних случаях с одной стороны остаётся «двуугольник», и выбора нет, так что мы полагаем $T_2 = 0$.)

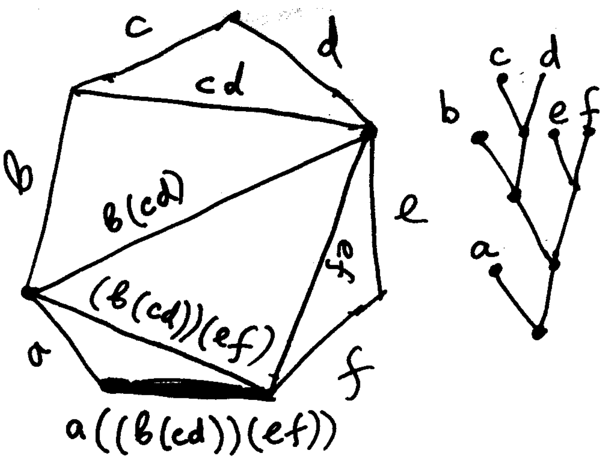


Получаем формулу

$$T_6 = T_2T_5 + T_3T_4 + T_4T_3 + T_5T_2$$

которая опять отличается от каталановской сдвигом индексов (как раз на 2).

Сравнивая две эти задачи, можно заключить, что количество триангуляций $(m+1)$-угольника равно количеству способов расстановки скобок в неассоциативном произведении m элементов. Этому равенству можно дать и комбинаторное объяснение, установив взаимно однозначное соответствие между теми и другими. Для этого на всех сторонах $(m+1)$-угольника, кроме его основания, напишем буквы (переменные, которые надо перемножать, слева направо). Пусть дана некоторая триангуляция этого $(m+1)$-угольника. Тогда на всех диагоналях можно написать промежуточные результаты в процессе умножения, как на рисунке:



Также этой триангуляции и этому произведению можно поставить в соответствие двоичное дерево с корнем, листьями которого являются переменные (как показано справа на том же рисунке), поэтому количество таких деревьев с m вершинами тоже равно C_{m-1}.

Наконец, если уж мы заговорили о комбинаторных доказательствах (их называют ещё «биективными», поскольку взаимно однозначные соответствия называют биекциями), то возникает вопрос: *можно ли поставить в соответствие каждому правильному скобочному выражению* (с которых мы начинали, из n открывающихся и n закрывающихся скобок) *некоторый порядок действий в произведении $n+1$ элементов*? Мы ведь уже знаем, что количество тех и других одинаковое.

Первая идея, которая приходит в голову: может быть, просто стереть все буквы в произведении и оставить одни скобки? получится ведь правильное скобочное выражение. Беда в том, что это никак не будет взаимно однозначным соответствием: без букв в выражениях $a(bc)$ и $(ab)c$ останется одно и то же. Тем не менее такое соответствие можно явно описать, и оно имеет программистский смысл — но скобки в правильном скобочном выражении будут иметь другой смысл, это операции загрузки и умножения в стековом калькуляторе для выражений. Вот что это значит.

Есть такая «обратная польская запись» для арифметических выражений, где знак операции ставится после операндов. При этом скобки не нужны: скажем, произведение $(ab)c$ записывается как $ab{\times}c{\times}$, а произведение $a(bc)$ записывается как

$abc{\times}{\times}$. Достоинство такой записи (из-за чего она использовалась в некоторых программируемых калькуляторах) состоит в том, что записанное таким образом выражение легко вычислять слева направо. Видя букву, мы помещаем соответствующую переменную в стек, а видя операцию ($\times$), мы заменяем два верхних элемента стека на их произведение (тем самым уменьшая число элементов на единицу). Вот как это получается, скажем, для выражения $((ab)(cd))$, которое в польской записи имеет вид $ab{\times}cd{\times}{\times}$; вершина стека справа:

$$\begin{array}{rl}
 & \mathtt{a} \\
a & \\
 & \mathtt{b} \\
a,b & \\
 & \times \\
ab & \\
 & \mathtt{c} \\
ab,c & \\
 & \mathtt{d} \\
ab,c,d & \\
 & \times \\
ab,cd & \\
 & \times \\
(ab)(cd) &
\end{array}$$

В правом столбце сверху вниз написаны символы выражения в обратной польской записи, а слева содержимое стека после соответствующего действия.

Названия переменных среди команд идут в том же порядке, что и в исходном произведении, поэтому они излишни, можно писать одну и ту же букву, скажем, v в смысле «очередная переменная». Тогда расстановке скобок, указывающей порядок умножений в произведении m сомножителей, будет соответствовать последовательность из m букв v (соответствующих сомножителям) и $m-1$ знаков $\times$ (каждое умножение уменьшает число сомножителей на 1). В этой последовательности первые две буквы должны быть v (потому что умножение возможно, только если в стеке как минимум два операнда), а потом в любом начальном отрезке количество букв v должно быть как минимум на одну больше, чем количество знаков умножения (иначе стек опустеет: буквы его удлиняют, а умножения укорачивают). Приходим — если отбросить первую букву v — как раз к правильным скобочным выражениям, где v играет роль левой скобки, а $\times$ — правой.

Конечно, это всё надо аккуратно определить и доказать (мы не сделали ни того, ни другого — попробуйте на досуге), но хотя бы само соответствие мы описали достаточно явно.

2.11 Что дальше?

На этом мы заканчиваем знакомство с перечислительной комбинаторикой — но сама она, конечно, не заканчивается. Чему мы научились — и чему ещё предстоит научиться?

Мы убедились, что во многих случаях можно посчитать число объектов с данным свойством, не пересчитывая эти объекты. Иногда ответ даётся явной формулой; иногда такую явную формулу найти не удаётся, но можно составить рекуррентное соотношение и с его помощью вычислить искомое количество.[10] (Программистам этот приём известен под названием «динамическое программирование».)

Пример с числами Каталана показывает, что при доказательстве комбинаторных тождеств иногда нужны не только остроумные соображения (скажем, идея отражения путей), но и разные сведения из других разделов математики (разложение в ряд, формула Тейлора и многое другое — у нас были только самые простые примеры). Если вам это понравилось, много других интересных вещей вы найдёте в книге «Конкретная математика» (Р. Грэхем, Д. Кнут и О Паташник; один из авторов — тот самый Дональд Кнут, который написал знаменитый многотомник «Искусство программирования» и не менее знаменитую программу TEX, с помощью которой сейчас набираются практически все книги по математике и программированию, включая эту) и в книге «Лекции о производящих функциях» (С. К. Ландо).

[10] Можно спросить, всегда ли для такого рода задач (подсчитать количество объектов с данным просто проверяемым свойством) существует быстрый алгоритм. Этот вопрос естественно формулируется в терминах теории сложности вычислений, но ответ на него неизвестен. Он является некоторым аналогом известной проблемы о равенстве классов P и NP.