Теоретическая информатика - 1

Комбинаторика: выборки, числа Каталана

Типы выборок

Пусть $A = \{a_1, \ldots, a_n\}$ — множество из n элементов.

Набор элементов $a_{i_1}, \ldots, a_{i_r}$, $r \geq 1$, называется выборкой объема r из n элементов, или (n, r)-выборкой.

Выборки бывают:

- упорядоченные и неупорядоченные
  - *упорядоченные* — порядок элементов важен (т.е., две упорядоченные выборки, отличающиеся лишь порядком элементов, считаются различными);
  - *неупорядоченные* — порядок элементов неважен
- с повторениями и без повторений

Правило суммы и правило произведения

Типичная задача: число возможных выборок с определенными свойствами.

Полезны два правила:

- *Правило суммы*: если объект A может быть выбран m способами, а другой объект B — другими n способами, то выбор "A или B" можно осуществить $m + n$ способами.
- *Правило произведения*: если объект A может быть выбран m способами, и после каждого из таких выборов объект B в свою очередь может быть выбран n способами, то выбор пары (A, B) можно осуществить mn способами.

Выборки k элементов из n:

1. Упорядоченная с повторениями:

$$n^k$$

Выборки k элементов из n:

1. Упорядоченная с повторениями:

$$n^k$$

2. Упорядоченная без повторений (размещения):

$$A_n^k = n!/(n-k)! = n(n-1)\dots(n-k+1)$$

Выборки k элементов из n:

1. Упорядоченная с повторениями:

$$n^k$$

2. Упорядоченная без повторений (размещения):

$$A_n^k = n!/(n-k)! = n(n-1)\dots(n-k+1)$$

3. Неупорядоченная без повторений (сочетания):

$$C_n^k = n!/k!(n-k)! = A_n^k/k! = n(n-1)\dots(n-k+1)/k!$$

Выборки k элементов из n:

1. Упорядоченная с повторениями:

$$n^k$$

2. Упорядоченная без повторений (размещения):

$$A_n^k = n!/(n-k)! = n(n-1)\dots(n-k+1)$$

3. Неупорядоченная без повторений (сочетания):

$$C_n^k = n!/k!(n-k)! = A_n^k/k! = n(n-1)\dots(n-k+1)/k!$$

- число перестановок из n элементов (т.е., различных способов упорядочить n элементов):

$$n!$$

  частный случай A_n^k при $k = n$

4. Неупорядоченная с повторениями (сочетания с повторениями): $\hat{C}_n^k = C_{n+k-1}^k$.

4. Неупорядоченная с повторениями (сочетания с повторениями): $\hat{C}_n^k = C_{n+k-1}^k$.

Доказательство: Пусть $A = \{a_1, \dots, a_n\}$. Неупорядоченная выборка с повторениями k элементов из A задается вектором чисел $(x_1, \dots, x_n)$, где x_i — число повторений a_i, и

$$x_1 + \dots + x_n = k.$$

Т.е. число таких выборок равно числу решений этого уравнения в целых неотрицательных числах.

4. Неупорядоченная с повторениями (сочетания с повторениями): $\hat{C}_n^k = C_{n+k-1}^k$.

Доказательство: Пусть $A = \{a_1, \dots, a_n\}$. Неупорядоченная выборка с повторениями k элементов из A задается вектором чисел $(x_1, \dots, x_n)$, где x_i — число повторений a_i, и

$$x_1 + \dots + x_n = k.$$

Т.е. число таких выборок равно числу решений этого уравнения в целых неотрицательных числах.

Закодируем решение бинарным вектором:

- x_i соответствует блок из x_i единиц,
- между x_i и x_{i+1} разделитель 0.

Т.е. получаем вектор длины $(n + k - 1)$, содержащий ровно k единиц (и ровно $(n - 1)$ нулей).

4. Неупорядоченная с повторениями (сочетания с повторениями): $\hat{C}_n^k = C_{n+k-1}^k$.

Доказательство: Пусть $A = \{a_1, \ldots, a_n\}$. Неупорядоченная выборка с повторениями k элементов из A задается вектором чисел $(x_1, \ldots, x_n)$, где x_i — число повторений a_i, и

$$x_1 + \cdots + x_n = k.$$

Т.е. число таких выборок равно числу решений этого уравнения в целых неотрицательных числах.

Закодируем решение бинарным вектором:

- x_i соответствует блок из x_i единиц,
- между x_i и x_{i+1} разделитель 0.

Т.е. получаем вектор длины $(n + k - 1)$, содержащий ровно k единиц (и ровно $(n - 1)$ нулей).

Число таких векторов — это число сочетаний C_{n+k-1}^k. ЧТД.

Выборки

Как найти число решений уравнения

$$x_1 + \cdots + x_n = k$$

в целых положительных числах $(x_i > 0,\ k \geq n)$?

Выборки

Как найти число решений уравнения

$$x_1 + \dots + x_n = k$$

в целых положительных числах ($x_i > 0$, $k \geq n$)?

Положим $y_i = x_i - 1$ и найдем число решений

$$y_1 + \dots + y_n = k - n$$

в целых неотрицательных числах ($y_i \geq 0$):
$\hat{C}_n^{k-n} = C_{n+k-n-1}^{k-n} = C_{k-1}^{n-1}$.

Число сочетаний: свойства

Теорема

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}.$$

Доказательство. Преобразованием:

$$C_{n-1}^k + C_{n-1}^{k-1} = \frac{(n-1)!}{(n-k-1)!k!} + \frac{(n-1)!}{(n-k)!(k-1)!} =$$

$$\frac{(n-1)!((n-k)+k)}{(n-k)!k!} = \frac{n!}{(n-k)!k!} = C_n^k$$

Число сочетаний: треугольник Паскаля

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
 & & & & \dots & & & & \\
 & & & C_{n-1}^{k} & & C_{n-1}^{k-1} & & & \\
 & & & & C_n^k & & & &
\end{array}
$$

Число сочетаний: бином Ньютона

Теорема

$$(a + b)^n = \Sigma_{k=0}^{n} C_n^k a^k b^{n-k}.$$

Число сочетаний: бином Ньютона

Теорема

$$(a+b)^n = \Sigma_{k=0}^{n} C_n^k a^k b^{n-k}.$$

Доказательство. Член $a^k b^{n-k}$ участвует в разложении $(a+b)^n$ столько раз, сколько есть способов выбрать a в k множителях из n (и, соответственно, b в $n-k$ множителях из n) — а это C_n^k.

Факториал

Лемма (грубые оценки для n!)

$$(n/e)^n < n! < n^n.$$

Факториал

Лемма (грубые оценки для n!)

$$(n/e)^n < n! < n^n.$$

Доказательство. Верхняя оценка очевидна.
Докажем нижнюю по индукции.

- Базис: $(1/e) < 1!$, верно.
- Шаг индукции: Пусть верно для n: $(n/e)^n < n!$.
  Покажем для $(n+1)$.
  $(n+1)! = (n+1)n! > (n+1)(n/e)^n$ (по предположению индукции);
  Надо показать: $(n+1)(n/e)^n > ((n+1)/e)^{n+1}$;
  Это верно тогда и только тогда, когда $en^n > (n+1)^n$, что в свою очередь равносильно $(1+1/n)^n < e$, что верно (курс матанализа).

формула Стирлинга

Теорема (формула Стирлинга)

$$n! = (1 + o(1))\sqrt{2\pi n}(n/e)^n.$$

Доказательство будет в курсе матанализа.

Язык Дика

Правильная скобочная последовательность (ПСП)

Язык строк над $\Sigma = \{(,)\}$.
Определение по индукции:

- пустая строка ε — ПСП;
- если w — ПСП, то (w) — ПСП;
- если w, u — ПСП, то wu — ПСП.

Множество ПСП называется языком Дика:
ε, (), ()(), (()), (())()...

Язык Дика

Сколько существует ПСП с n парами скобок
(= сколько слов длины $2n$ в языке Дика)?

ε	$D_0 = 1$
$()$	$D_1 = 1$
$()(), (())$	$D_2 = 2$
$()()(), (())(), ()(()), ((())), (()())$	$D_3 = 5$
$\ldots$	

Это задается последовательностью *чисел Каталана* D_n.

Энциклопедия целочисленных последовательностей

Sloane's encyclopedia of integer sequences:

https://oeis.org

Последовательность A000108.

Рекуррентная формула для чисел Каталана

Теорема (рекуррентная формула)

$$D_0 = 1;\ D_n = \sum_{k=0}^{n-1} D_k D_{n-1-k}.$$

Рекуррентная формула для чисел Каталана

Теорема (рекуррентная формула)

$$D_0 = 1; D_n = \Sigma_{k=0}^{n-1} D_k D_{n-1-k}.$$

Доказательство: Пусть w — произвольная ПСП длины $2n$. Она начинается с открывающейся скобки. Найдем парную ей закрывающуюся скобку и представим последовательность w в виде: $w = (u)v$, где u и v ПСП.

Рекуррентная формула для чисел Каталана

Теорема (рекуррентная формула)

$$D_0 = 1; D_n = \Sigma_{k=0}^{n-1} D_k D_{n-1-k}.$$

Доказательство: Пусть w — произвольная ПСП длины $2n$. Она начинается с открывающейся скобки. Найдем парную ей закрывающуюся скобку и представим последовательность w в виде: $w = (u)v$, где u и v ПСП.

Если длина u равна $2k$, то u можно составить D_k способами.
Тогда длина v равна $2(n-k-1)$ и v можно составить D_{n-k-1} способами.

Рекуррентная формула для чисел Каталана

Теорема (рекуррентная формула)

$$D_0 = 1; D_n = \Sigma_{k=0}^{n-1} D_k D_{n-1-k}.$$

Доказательство: Пусть w — произвольная ПСП длины $2n$. Она начинается с открывающейся скобки. Найдем парную ей закрывающуюся скобку и представим последовательность w в виде: $w = (u)v$, где u и v ПСП.

Если длина u равна $2k$, то u можно составить D_k способами.
Тогда длина v равна $2(n - k - 1)$ и v можно составить D_{n-k-1} способами.

Комбинация любого способа составить u с любым способом составить v даст новую последовательность w.

Рекуррентная формула для чисел Каталана

Теорема (рекуррентная формула)

$$D_0 = 1; D_n = \Sigma_{k=0}^{n-1} D_k D_{n-1-k}.$$

Доказательство: Пусть w — произвольная ПСП длины $2n$. Она начинается с открывающейся скобки. Найдем парную ей закрывающуюся скобку и представим последовательность w в виде: $w = (u)v$, где u и v ПСП.

Если длина u равна $2k$, то u можно составить D_k способами.
Тогда длина v равна $2(n-k-1)$ и v можно составить D_{n-k-1} способами.

Комбинация любого способа составить u с любым способом составить v даст новую последовательность w.

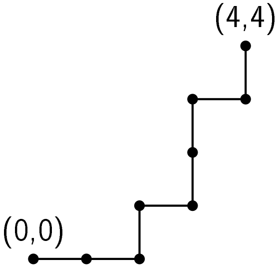
Суммируя по k от 0 до $n-1$, получаем рекуррентную формулу.

Числа Каталана через монотонные пути

ПСП длины $2n$ поставим в соответствие путь в квадрате $[0, n] \times [0, n]$ из точки $(0, 0)$ в точку (n, n).

Открывающей скобке сопоставим горизонтальный отрезок длины 1, а закрывающей — вертикальный.

$(()())() \quad \Leftrightarrow$

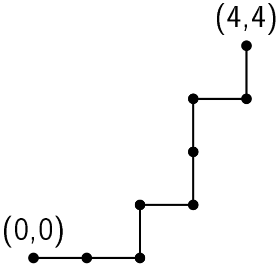


Числа Каталана через монотонные пути

ПСП длины $2n$ поставим в соответствие путь в квадрате $[0, n] \times [0, n]$ из точки $(0, 0)$ в точку (n, n).

Открывающей скобке сопоставим горизонтальный отрезок длины 1, а закрывающей — вертикальный.

$(()())()$ $\Leftrightarrow$



Если путь сопоставлен ПСП, то ни одна его точка не может лежать выше главной диагонали квадрата ("правильный путь"). Обратно, такому пути сопоставляется ПСП.

Аналитическая формула для чисел Каталана

Теорема (Аналитическая формула)

$$D_n = \frac{1}{n+1} C_{2n}^n.$$

Аналитическая формула для чисел Каталана

Теорема (Аналитическая формула)

$$D_n = \frac{1}{n+1} C_{2n}^n.$$

Доказательство: Сместим правильный путь на 1 клетку вниз. Теперь правильный путь идет из $(0, -1)$ в $(n, n-1)$ и не имеет общих точек с прямой $y = x$.

Аналитическая формула для чисел Каталана

Теорема (Аналитическая формула)

$$D_n = \frac{1}{n+1} C_{2n}^{n}.$$

Доказательство: Сместим правильный путь на 1 клетку вниз. Теперь правильный путь идет из $(0, -1)$ в $(n, n-1)$ и не имеет общих точек с прямой $y = x$.

Число правильных путей = общее число путей - число неправильных.

Аналитическая формула для чисел Каталана

Теорема (Аналитическая формула)

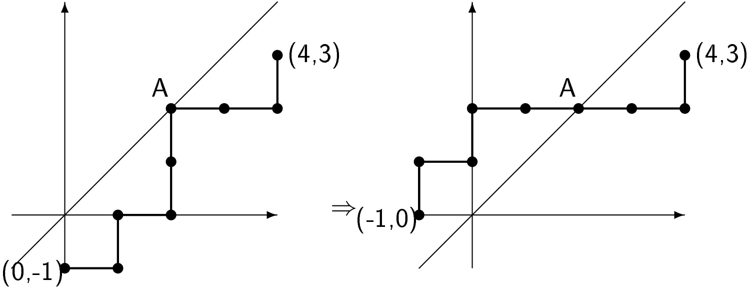
$$D_n = \frac{1}{n+1} C_{2n}^n.$$

Доказательство: Сместим правильный путь на 1 клетку вниз. Теперь правильный путь идет из $(0, -1)$ в $(n, n-1)$ и не имеет общих точек с прямой $y = x$.

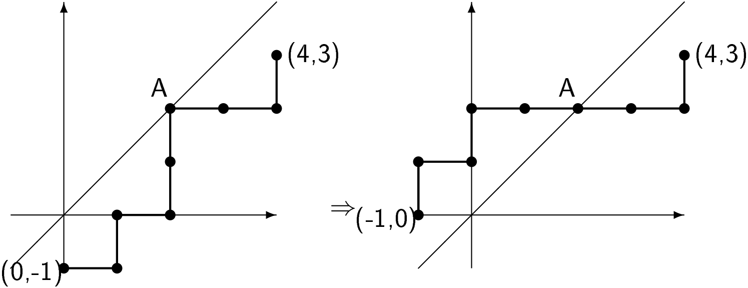
Число правильных путей = общее число путей - число неправильных.

Общее число путей из $(0, -1)$ в $(n, n-1)$ — число способов выбрать n вертикальных сегментов (и n горизонтальных) из общего числа $2n$, т.е. C_{2n}^n.

Рассмотрим неправильный путь и его первую точку на прямой $y = x$ (точка A). Отрезок пути от $(0, -1)$ до A заменим симметричным относительно прямой $y = x$. Получим путь длины $2n$ из $(-1, 0)$ в $(n, n-1)$.

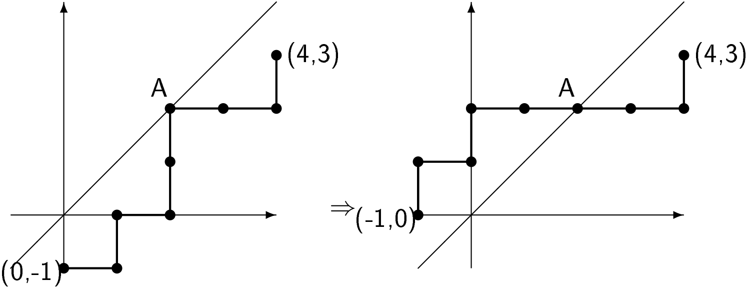


Рассмотрим неправильный путь и его первую точку на прямой $y = x$ (точка A). Отрезок пути от $(0, -1)$ до A заменим симметричным относительно прямой $y = x$. Получим путь длины $2n$ из $(-1, 0)$ в $(n, n-1)$.



Обратно, пусть дан путь длины $2n$ из $(-1, 0)$ в $(n, n-1)$ и пусть A — первая точка этого пути на прямой $y = x$. Заменив участок пути от $(-1, 0)$ до A на симметричный относительно прямой $y = x$, получим неправильный путь из $(0, -1)$ в $(n, n-1)$.

Рассмотрим неправильный путь и его первую точку на прямой $y = x$ (точка A). Отрезок пути от $(0, -1)$ до A заменим симметричным относительно прямой $y = x$. Получим путь длины $2n$ из $(-1, 0)$ в $(n, n-1)$.



Обратно, пусть дан путь длины $2n$ из $(-1, 0)$ в $(n, n-1)$ и пусть A — первая точка этого пути на прямой $y = x$. Заменив участок пути от $(-1, 0)$ до A на симметричный относительно прямой $y = x$, получим неправильный путь из $(0, -1)$ в $(n, n-1)$.

Следовательно, неправильных путей из $(0, -1)$ в $(n, n-1)$ столько же, сколько путей из точки из $(-1, 0)$ в $(n, n-1)$.

Путь из $(-1, 0)$ в $(n, n-1)$ содержит $n+1$ горизонтальных и $n-1$ вертикальных участков. Поэтому количество таких путей равно C_{2n}^{n-1} .

Путь из $(-1,0)$ в $(n,n-1)$ содержит $n+1$ горизонтальных и $n-1$ вертикальных участков. Поэтому количество таких путей равно C_{2n}^{n-1} .

Значит, количество правильных путей (т.е. число Каталана D_n) равно

$$D_n = C_{2n}^{n} - C_{2n}^{n-1} = \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n+1)!(n-1)!} =$$

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

ЧТД

Асимптотика чисел Каталана

Теорема

$$D_n = (1 + o(1))\frac{4^n}{n^{3/2}\sqrt{\pi}}.$$

Асимптотика чисел Каталана

Теорема

$$D_n = (1 + o(1))\frac{4^n}{n^{3/2}\sqrt{\pi}}.$$

Доказательство: Используем формулу Стирлинга:

$$n! = (1 + o(1))\sqrt{2\pi n}(n/e)^n.$$

Оценим биномиальный коэффициент:

$$C_{2n}^n = \frac{(2n)!}{(n!)^2} = \frac{(1 + o(1))\sqrt{2\pi 2n}(2n/e)^{2n}}{(1 + o(1))2\pi n(n/e)^{2n}} = (1 + o(1))\frac{4^n}{\sqrt{\pi n}}.$$

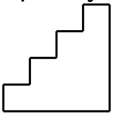
Далее, число Каталана

$$\frac{1}{n+1}C_{2n}^n = (1 + o(1))\frac{1}{n+1}\frac{4^n}{\sqrt{\pi n}} = (1 + o(1))\frac{4^n}{n^{3/2}\sqrt{\pi}}.$$

ЧТД

Другие представления чисел Каталана

- число разбиений выпуклого $(n+2)$-угольника на треугольники непересекающимися диагоналями
- число способов соединения $2n$ точек на окружности n непересекающимися хордами
- число способов заполнить n-лесенку n прямоугольниками



- Число таблиц Юнга размером $2 \times n$. Таблица Юнга — прямоугольник, заполненный последовательными числами так, чтобы они возрастали во всех строках и столбцах:

$$\begin{matrix} 1 & 2 & 4 \\ 3 & 5 & 6 \end{matrix}$$

- число плоских корневых деревьев с $n+1$ вершинами
- число неизоморфных упорядоченных бинарных деревьев с корнем и $(n+1)$ листьями
- число параллеломино (пар путей на клетчатой бумаге с началом $(0,0)$ и концом в одной и той же точке, идущих только вверх и вправо и не имеющих общих точек, кроме начала и конца) периметра $2n+2$
- число последовательностей натуральных чисел $1, a_1, \ldots, a_n, 1$, в которых каждый член является делителем суммы двух соседей:
  1 4 3 2 1 , 1 3 5 2 1, 1 3 2 3 1, 1 2 5 3 1, 1 2 3 4 1
- число наборов из n целых чисел от 0 до n, сумма которых делится на $(n+1)$
  0 0 0, 0 1 3, 0 2 2, 1 1 2, 2 3 3
- число способов разделить скобками $(n+1)$ множитель

И другие: R.P. Stanley, Enumerative Combinatorics, Vol. 2, 1999.