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 $\qquad D_0 = 1$
() $\qquad D_1 = 1$
()(), (()) $\qquad D_2 = 2$
()()(), (())(), ()(()), ((())), (()()) $\qquad D_3 = 5$
$\dots$

Это задается последовательностью *чисел Каталана* D_n.

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.

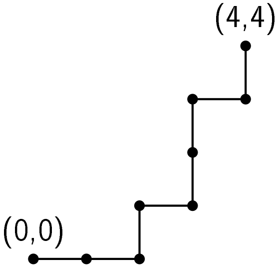
Суммируя по k от 0 до $n-1$, получаем рекуррентную формулу.

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.

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



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)

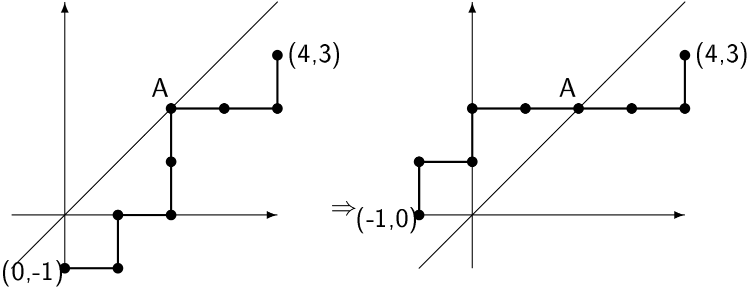
$$D_n = \frac{1}{n+1} C_{2n}^n.$$

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 $(n, n - 1)$.



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 $(n, n - 1)$.

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

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

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