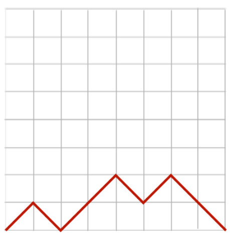
Ranking and unranking Dyck paths

Contents

1 Ranking and unranking Dyck paths

1.1 Dyck paths to words

Recall from Lecture 5 that a Dyck path of lengh $2n$ is a lattice walk from $(0,0)$ to $(2n,0)$ using up-steps $\nearrow = (1,1)$ and down-steps $\searrow = (1,-1)$. We can represent a Dyck path as a binary string of length $2n$ by writing $\nearrow$ as 0 and $\searrow$ as 1.

Example.  $\longleftrightarrow (\nearrow, \searrow, \nearrow, \nearrow, \searrow, \nearrow, \searrow, \searrow) \longleftrightarrow (\,0\,,\,1\,,\,0\,,\,0\,,\,1\,,\,0\,,\,1\,,\,1\,)$

Which binary strings do we get?

Definition. A binary string of length $2n$ is totally balanced if

- The string contains n zeros and n ones.
- For T$1 \le i \le 2n$, the first i elements of the string include at least as many zeros and ones.

Dyck paths correspond to totally balanced binary strings. The first condition ensures that a Dyck path finishes on the x-axis. The second condition ensures that it never goes strictly below the x-axis.

Viewing Dyck paths as binary strings gives them a natural lexicographic order. In this lecture, we find how to lexicographically rank and unrank Dyck paths (i.e. totally balanced binary words).

Example. When $n = 3$, the five Dyck paths of length 6 correspond to balanced binary words lexicographically ordered as follows. In particular, we have $\mathrm{Rank}(\nearrow, \nearrow, \searrow, \searrow, \nearrow, \searrow) = 3$.

$$(0,0,0,1,1,1),\ (0,0,1,0,1,1),\ (0,0,1,0,1,1),\ (0,0,1,1,0,1),\ (0,1,0,1,0,1)$$

We can rank Dyck paths by using the general lexicographic ranking formula from Lecture 8. This formula requires that, for any totally balanced word $W = (w_1, w_2, \ldots, w_{2n})$, any index $1 \le i \le 2n$ and any entry a with $0 \le a \le w_i - 1$, we know to compute the number $P(W;i,a)$ of totally balanced words that lexicographically precede W and begin with the sequence $(w_1, w_2, \ldots, w_{i-1}, a)$. The condition $0 \le a < w_i \le 1$ implies $a = 0$ and $w_i = 1$, so we can simplify the inner sum of the general ranking formula.

$$\mathrm{Rank}(W) = \sum_{i=1}^{2n} \sum_{a=0}^{w_i - 1} P(W;i,a) = \sum_{i=1}^{2n} w_i P(W;i,0). \tag{1}$$

We turn to the problem of evaluating $P(W;i,0)$. We must count the Dyck paths p that correspond to a totally balanced word of the form $(w_1, w_2, \ldots, w_{i-1}, 0, *, *, \ldots, *)$. The first i steps of p will bring us to a lattice point (x,y) where $x = i$. We need to be able to count the paths from (x,y) to $(2n,0)$ that do not go strictly below the x-axis.

1.2 Counting suffixes of Dyck paths

Definition. Let $\mathcal{D}_{2n}(x,y)$ be the set of paths from (x,y) to $(2n,0)$ using the steps $(1,1)$ and $(1,-1)$ and which never go strictly below the x-axis.

Let $d_{2n}(x,y) = |\mathcal{D}_{2n}(x,y)|$.

Such paths are ends (suffixes if you think of them as words) of Dyck paths.

Proposition. *Let x, y, and n be integers with $x+y$ even and $x+y \leq 2n$. Then*

$$d_{2n}(x,y) = \binom{2n-x}{n-\frac{x+y}{2}} - \binom{2n-x}{n-1-\frac{x+y}{2}}$$

Proof. Let $\mathcal{P}_{2n}(x,y)$ be the set of all paths from (x,y) to $(2n,0)$ that use steps $(1,1)$ and $(1,-1)$, including those that go below the x-axis. Let $\mathcal{B}_{2n}$ be the set of paths in $\mathcal{P}_{2n}(x,y)$ be the set of paths in $\mathcal{P}_{2n}(x,y)$ that go strictly below the x-axis. Then

$$\mathcal{D}_{2n}(x,y) = \mathcal{P}_{2n}(x,y) - \mathcal{B}_{2n}(x,y).$$

We aim to count the paths in $\mathcal{P}_{2n}(x,y)$ and the paths in $\mathcal{B}_{2n}(x,y)$.

We start with $\mathcal{P}_{2n}(x,y)$. Consider a path $p \in \mathcal{P}_{2n}(x,y)$ go from (x,y) to $(2n,0)$. Let u be the number of up-steps in p, and let d be the number of down-steps in p. Since p goes from (x,y) to $(2n,0)$ we must have

$$d+u = 2n-x \qquad \text{and} \qquad d-u = y.$$

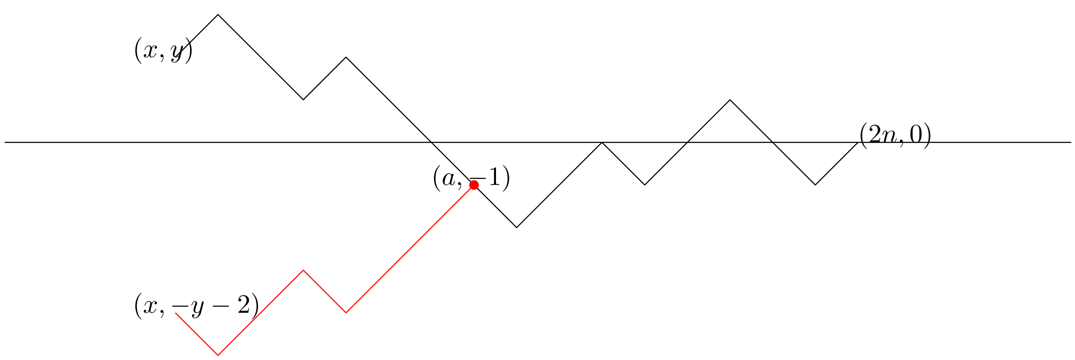
Solving these two equations, we find that

$$u = n - \frac{x+y}{2} \qquad \text{and} \qquad d = n - \frac{x-y}{2}.$$

Every path in $\mathcal{P}_{2n}(x,y)$ corresponds bijectively to a sequence of length $2n-x$ containing exactly u up-steps (the remaining steps are down-steps). There are $\binom{2n-x}{u}$ ways to choose the locations of the up-steps. Therefore

$$\mathcal{P}_{2n}(x,y) = \binom{2n-x}{u} = \binom{2n-x}{n-\frac{x+y}{2}}. \tag{2}$$

We now count $\mathcal{B}_{2n}(x,y)$.To determine $d_{2n}(x,y)$ we just need to subtract from (2) the number paths which *do* go strictly below the x axis. Let w be such a path, and let $(a,-1)$ be the first point along w which touches the line $y=-1$. We reflect the portion of w before $(a,-1)$ across the line $y=-1$, as illustrated, to obtain a new path w'.



Note that w' goes from $(x,-y-2)$ to $(2n,0)$. Note also that the construction $w \mapsto w'$ is reversible. That is, for every path w' from $(x,-y-2)$ to $(2n,0)$ there is a unique path w which produces w' in this way. Specifically, there is a unique point where it may have been flipped according to this rule: namely the first place where the path reaches $(a,-1)$ for some a.

Thus the number of paths from (x, y) to $(2n, 0)$ which go strictly below the x-axis is the same as the number of paths from $(x, -y-2)$ to $(2n, 0)$ with no restrictions. This number is given by (2) with y replaced by $-y-2$. we leave as a little exercise to show this equals

$$\binom{2n-x}{n-1-\frac{x+y}{2}}$$

Therefore the number of paths from (x, y) to $(2n, 0)$ which never go strictly below the x-axis is

$$\binom{2n-x}{n-\frac{x+y}{2}} - \binom{2n-x}{n-1-\frac{x+y}{2}}$$

$\square$

Corollary. *The number of Dyck paths of length $2n$ is the n-th Catalan number* $\dfrac{1}{n+1}\dbinom{2n}{n}$.

Proof. By its definition, $\mathcal{D}_{2n}(0, 0)$ is precisely the set of Dyck paths of length $2n$, so the number of Dyck paths of length $2n$ is

$$d_{2n}(0,0) = \binom{2n}{n} - \binom{2n}{n-1} = \binom{2n}{n} - \frac{n}{n+1}\binom{2n}{n} = \left(1 - \frac{n}{n+1}\right)\binom{2n}{n}.$$

$\square$

1.3 Ranking and unranking

The formula for counting suffixes tells us about the rank. Suppose $w = w_1w_2\cdots w_{2n}$ is a Dyck path represented as a totally balanced binary string. As discussed in Lecture 8, we step through $i = 1, 2, \ldots, 2n$, each time accounting for those predecessors of begin with $w_1w_2\cdots w_{i-1}a$, where $a < w_i$. This only happens when $a = 0$ and $w_i = 1$. Suppose that after the first i steps the path is at the point (x_i, y_i). Then $x_i = i$. If $w_i = 0$, then no accounting is needed. If $w_i = 1$, then the previous step was down, and there is exactly one totally balanced word which begins $w_1w_2\cdots w_{i-1}0$ for every path from (x_i, y_i+2) to $(2n, 0)$ which does not go below the x-axis. There are exactly $d_{2n}(i, y_i+2)$ such predecessors (the $+2$ comes because their ith step is an up step, instead of a down step). Thus we have

$$\mathrm{Rank}(w) = \sum_{i=1}^{2n} w_i d_{2n}(i, y_i+2) = \sum_{i=1}^{2n} w_i\left(\binom{2n-i}{n-\frac{i+y_i+2}{2}} - \binom{2n-i}{n-1-\frac{i+y_i+2}{2}}\right)$$

In the following we keep track of the pair $(x, y) = (i, y_i)$.

Algorithm: RankDyck

```
input: n, w.   w is a totally balanced word of length 2n
y = 0
r = 0
for x from 1 to 2n
  y = y+1
  if w(x) = 1
     r = r + binom( 2n-x, n-(x+y)/2 ) - binom( 2n-x , n-1-(x+y+2)/2 )
     y = y - 2
output: r
```

The following unranking algorithm is an adaptation of algorithm `UnrankGeneral` from Lecture 8.

Algorithm: UnrankDyck

```
input: n,r.
y = 0
R = 0
for x from 1 to 2n
   w(x) = 0
```

```
    y = y+1
    P = binom( 2n-x, n-(x+y)/2 ) - binom( 2n-x, n-(x+y)/2 )
    if  R + P <= r
      w(x) = 1
      y = y - 2
      R = R + P
output: w
```

Here's an example of each algorithm which is adapted from Kreher and Stinson, *Combinatorial Algorithms*, Section 3.4.

Suppose we want to compute the rank of $w = 0010110101$. We calculate as follows.

x	$w(x)$	y	$d_{10}(x, y+2)$	r
1	0	1		0
2	0	2		0
3	1	1	$\binom{7}{2} - \binom{7}{1} = 14$	14
4	0	2		14
5	1	1	$\binom{5}{1} - \binom{5}{0} = 4$	18
6	1	0	$\binom{4}{1} - \binom{4}{0} = 3$	21
7	0	1		21
8	1	0	$\binom{2}{1} - \binom{2}{0} = 4$	22
9	0	1		22
10	1	0	$\binom{0}{-1} - \binom{0}{-2} = 0$	22

So the rank is 22. Notice that steps $x = 1$ and $x = 2n$ can is always be omited.

Now we calculate $\mathrm{Unrank}(22)$

$r = 22$

x	y	$d_{10}(x, y) = P$	R	$w(x)$
1	1	$\binom{9}{4} - \binom{9}{3} = 42$	0	0
2	2	$\binom{8}{3} - \binom{8}{2} = 28$	0	0
3	1	$\binom{7}{3} - \binom{7}{2} = 14$	14	1
4	2	$\binom{6}{2} - \binom{6}{1} = 9$	14	0
5	1	$\binom{5}{1} - \binom{5}{0} = 4$	18	1
6	0	$\binom{4}{1} - \binom{4}{0} = 3$	21	1
7	1	$\binom{3}{1} - \binom{3}{0} = 2$	21	0
8	0	$\binom{2}{0} - \binom{2}{-1} = 1$	22	1
9	1	$\binom{1}{0} - \binom{1}{-1} = 1$	22	0
10	1	$\binom{1}{-1} - \binom{0}{-2} = 0$	22	1

so $\mathrm{Unrank}(22) = 0010110101$ as expected.