順序木の数え上げ

3 本の辺を持つ順序木は, 図 1 に示す 5 個です. 4 本の辺を持つ順序木は, 14 個あります. 5 本の辺を持つ順序木は, 42 個あります. 一般に, m 本の辺を持つ順序木の個数は何個でしょうか. 答は, カタラン数

$$C(m) = {}_{2m+1}C_m/(2m+1)$$

個です. 以下, 解説します.

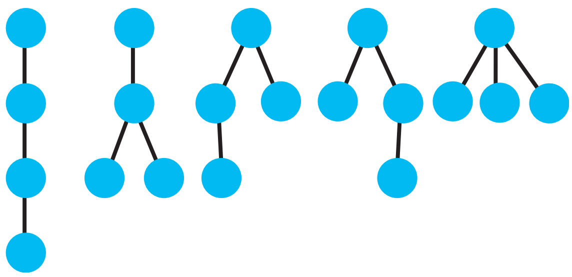

図 1: 3 本の辺を持つ順序木一覧

木とは閉路のない連結グラフです. 1 点を根として指定した木を根つき木といいます. 木を図示するとき, 通常の植物とは逆に, 根を一番上に描きます. 下にむかって枝分かれしていきます. 枝の先端の点を葉といいます.

根つき木は, 家系図のようにもみえます. 一族の創始者が根に対応します. ある点とある点が, 親子であるとか, 兄弟であるというようにいいます. 例えば, 図 2 において, 点 a は根です. 点 b の親は点 a です. 点 c は点 b の弟です. 子供がいない点は葉といいます. 点 d,e,f が葉です.

順序木が与えられたとき, 根からスタートして, この順序木のまわりを一周たどりましょう. 図 3 に例を示します. 小くなった自分が左手を壁につけたまま一周することを想像しましょう. 深さ優先探索といいます.

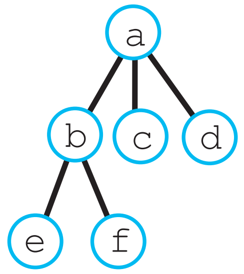

図 2: 順序木の例

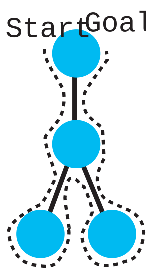

図 3: 深さ優先探索

辺を下にたどるとき D と書き, 辺を上にたどるとき U と書くことにします. Up と Down の意味です. すると, 図 1 のそれぞれの順序木に対して, 図 4 に示す文字列が得られます.

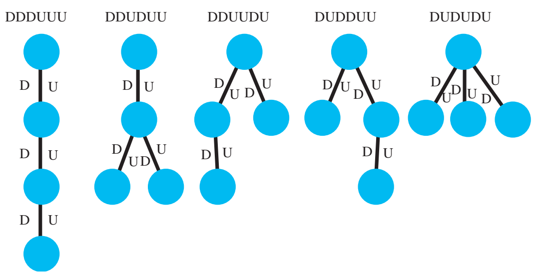

図 4: 文字列

各辺は下向きと上向きの 2 回たどられます. よって, 3 本の辺を持つ順序木から, 長さ $3 \times 2 = 6$ の文字列が得られます. この文字列は, 3 個の U と 3 個の D を含んでい

ます．また, この文字列から元の木を復元することもできます.

反対に, 3 個の U と 3 個の D からなる長さ 6 の文字列が与えられたとき, これに対応するような, 3 本の辺を持つ木は存在するでしょうか.

たくさんの反例があります．DDUUUD や,DUUUDD などです．なぜ, これらの文字列は順序木に対応しないのでしょうか.

上記の文字列を座標平面上のパスで表現しましょう．Dyck パスといいます．このパスは原点 $(0,0)$ からスタートします．文字 U は, x 方向に $+1$ および y 方向に $+1$ 進む線分に対応させます．文字 D は, x 方向に $+1$ および y 方向に -1 進む線分に対応させます．Up と Down の感じがでているでしょうか．図 5 に例を示します．この順序木に対応する文字列は DDUDUU であり, 対応する Dyck パスが図示されています．3 個の U と 3 個の D からなる長さ 6 の文字列から得られる Dyck パスは, 原点 $(0,0)$ からスタートし, 座標 $(6,0)$ がゴールです.

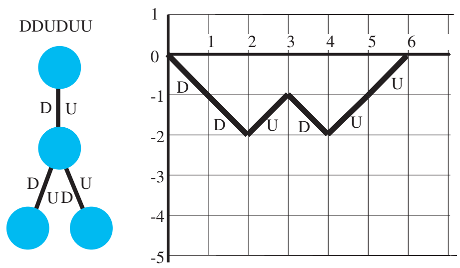

図 5: Dyck パス

順序木に対応する文字列から得られる Dyck パスは, y 座標が正の点を含みません．これに対し, 順序木に対応しない文字列から得られる Dyck パスは, y 座標が正の点を含みます．これは順序木のまわりを 1 周するときに, 根から (存在しない) 根の親へたどろうとして失敗することを意味します．それゆえに, 対応する順序木がないということになります.

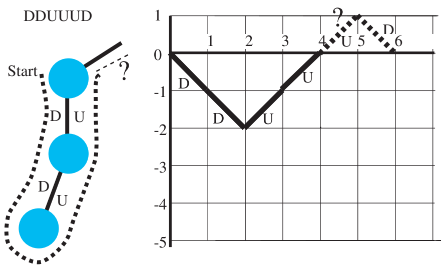

図 6: 順序木に対応しない Dyck パス

順序木に対応しない Dyck パスを, 次のように改造して, 無理矢理に順序木に対応させましょう.

Dyck パスを, 一番高い点 (y 座標の値が一番大きな点) で切り取り, これより左の部分を残りの部分の右に継ぎ足します. これは, 文字列の最初の何文字かを, 文字列の最後に移動することに相当します. サイクルシフトという改造です. 図 7 に例を示します. 文字列 DDUUUD(図上段) の最初の 5 文字 (図中段) を切り取り, 残りの文字列の最後に移動 (図下段) します.

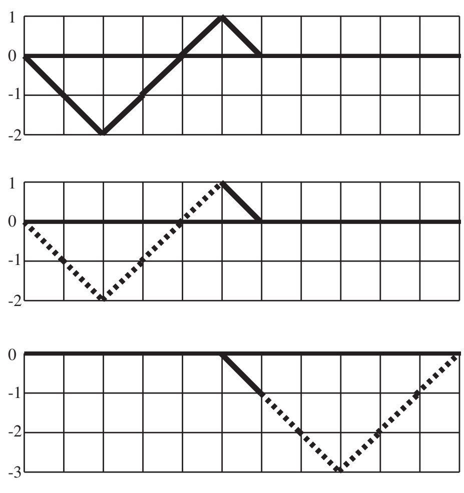

図 7: サイクルシフト

改造後の Dyck パスでは, スタート点が一番高いので必ず順序木に対応します. ただし, 一番高い点が複数あるような Dyck パスには, 上記のサイクルシフトによる改造の

方法が複数個存在します．図 8 に例を示します．

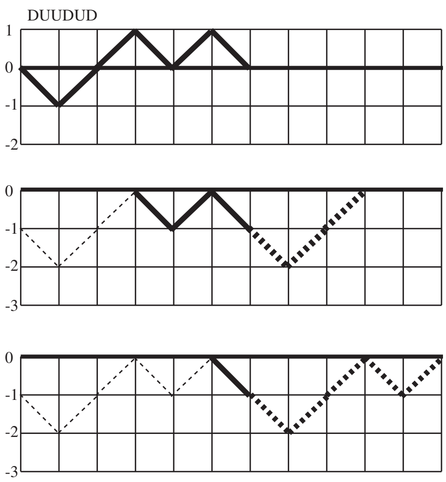

図 8: サイクルシフトによる改造が複数ある例

これを避けるため次のようにひと工夫します．(なぜ避けるのかは後でわかります.)

順序木のまわりを一周たどるとき，スタート点を根の(ダミーの) 親の点とします．ゴールは根のままとします．例えば図 5 の順序木のたどり方は図 9 のようになります．

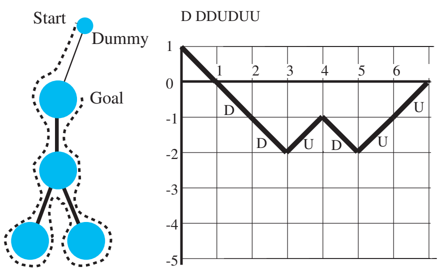

図 9: 新深さ優先探索

得られる文字列は，以前の方法で得られた文字列の先頭に D をつけたものとなります．よって 3 個の U と 3+1=4

個の D からなる長さ 7 の文字列となります. この文字列から得られる Dyck パスのスタート点は座標 $(0,1)$ とします. 追加した先頭の 1 文字 D の分を考慮しています. 座標 $(7,0)$ がゴールとなります. x の値が 1 以上の点はすべて y 座標が 0 以下でなくてはなりません. スタート点が唯一の最も高い点となることに注意して下さい.

さて, 今回も前回と同様に, 3 個の U と 4 個の D からなる長さ 7 の文字列であっても, これに対応するような順序木がない場合があります. 得られる Dyck パスに, x 座標が 1 以上, かつ, y 座標が正である点が存在する場合です. 図 10 に例を示します.

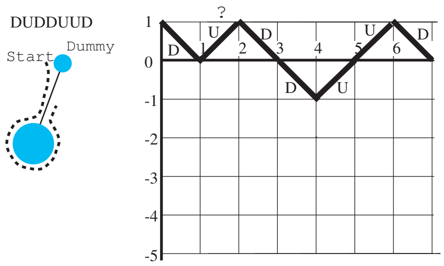

図 10: 順序木に対応しない Dyck パス

今回も, 順序木に対応しない Dyck パスを, また, サイクルシフトにより改造して, 無理矢理に順序木に対応させましょう.

Dyck パスを, (また,) 一番高い点 (y 座標の値が一番大きな点) で切り取り, これより左の部分を残りの部分の右に継ぎ足します. サイクルシフトです. 図 11 を見てください. しかし今回はスタート点とゴール点の y 座標が 1 だけずれているので, 継ぎ足した各点の y 座標は, 切り取り前より 1 だけ小さくなりました.

一番高い点以外で切り取り, 左の部分を残りの部分の右に継ぎ足すときは, 改造後にスタート点が最も高い点にならないので, 順序木に対応しません. 順序木に対応する Dyck パスは, スタート点が唯一の一番高い点であること

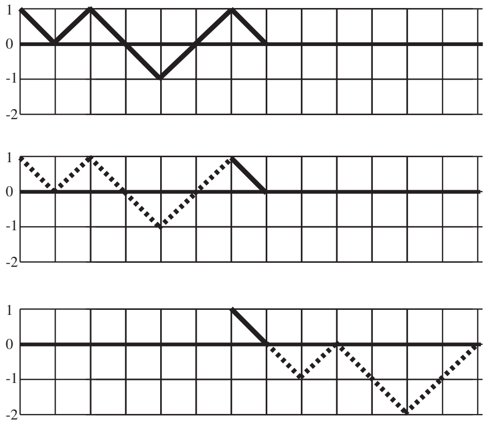

図 11: サイクルシフト

に注意しましょう.

さて, 一番高い点が複数ある Dyck パスは, 今回も上記のサイクルシフトによる改造が複数存在しますが, 順序木に対応するのは, 一番高い点のうち, 一番右の点で切り取った (図 11 の) 場合のみです. 例えば, 一番高い点のうち, 一番左の点で切り取った (図 12 の) 場合は, パスの途中に y 座標が正の点が存在し, 順序木に対応しません.

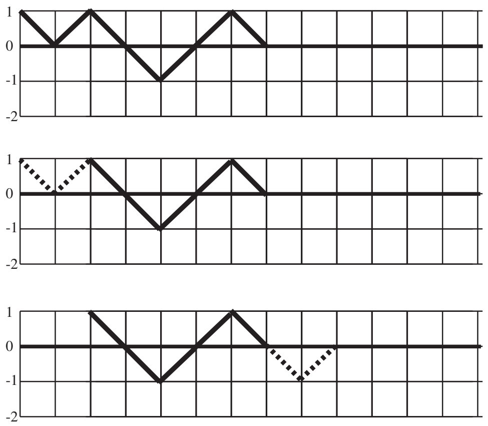

図 12: 順序木に対応しないサイクルシフト

つまり, 3 個の U と 4 個の D からなる長さ 7 の任意の文字列が与えられたとき, サイクルシフトによる改造によ

り, 順序木に対応する Dyck パスがちょうど 1 つだけ得られます.

一方, 次がいえます. 順序木に対応するような 3 個の U と 4 個の D からなる長さ 7 の文字列が与えられたときに, 先頭の 0 から 6 文字をそれぞれ切り取り, 残りの部分の右に貼り付けて得られる 7 つの文字列を考えると, この中には元の文字列以外に順序木に対応する文字列はありません。なぜなら, 一番高い点のうち, 一番右の点で切り取り, サイクルシフトした場合のみに, 順序木に対応する Dyck パスが得られるからです. 元の文字列で一番高い点はスタート点のひとつだけです.

以上の議論から, 次がいえます. 3 個の U と 4 個の D からなる長さ 7 の ${}_7C_3$ 個の文字列のうち, ちょうど 1/7 だけが 3 本の辺を持つ順序木に対応します. よって, $m = 3$ として, カタラン数 $C(m) ={}_{2m+1} C_m/(2m + 1) = 5$ 個の順序木があるのです.

一般の m についても, 同様に説明できます. m 個の U と (m+1) 個の D からなる長さ (2m+1) の ${}_{2m+1}C_m$ 個の文字列のうち, ちょうど $1/(2m+1)$ だけが m 本の辺を持つ順序木に対応します. つまり m 本の辺を持つ順序木は, カタラン数 $C(m) ={}_{2m+1} C_m/(2m + 1)$ 個あるのです.

発展問題です. 子供がいない点を葉といいます. m 本の辺を持ち, ℓ 個の葉を持つ順序木は何個あるでしょうか. Narayana 数

$$C(m, \ell) = ({}_mC_{\ell-1} \cdot {}_{m-1} C_{\ell-1})/\ell$$

個であることが知られています. 例えば 4 本の辺を持ち, 2 個の葉を持つ順序木は図 13 に示すように 6 個あります. 確かに

$$C(4, 2) = ({}_4C_1 \cdot {}_3 C_1)/2 = 6$$

です. なぜ 4 本の辺を持ち, 2 個の葉を持つ順序木は 6 個だけであるのかを説明しましょう.

このような順序木を,(また) 根のダミーの親を追加した後に, 深さ優先探索すると, 4 個の U と 5 個の D からなる長さ 9 の文字列が得られます. これに対応する Dyck パスは, 2 個の V 字型で構成されます. 2 個所の “谷底” が葉に対応しますね. “山頂” は両端を同一点とみなすと 2 個所あります.

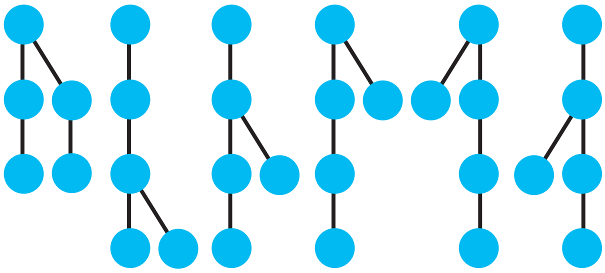

図 13: 4 本の辺を持ち 2 個の葉をもつ順序木一覧

2 個の V 字型からなる Dyck パスが与えられたとき, 一番高い山頂で切り取り, これより左の部分を残りの部分の右に継ぎ足すと, (すなわちサイクルシフトすると,) x 座標が 1 以上の点はすべて y 座標が 0 以下となります. 図 14 参照です. ただし, 一番高い点が 2 個ある場合は, 右の山頂を選びます. 山頂以外で切り取り, サイクルシフトした場合は, 改造後にスタート点が最も高い点にならないので, 順序木に対応することはありません. すなわち, 2 個の V 字型からなる Dyck パスが与えられたとき, これをサイクルシフトして得られる Dyck パスのうち, 順序木に対応するものは, 一番高い山頂でサイクルシフトして得られる 1 個だけです.

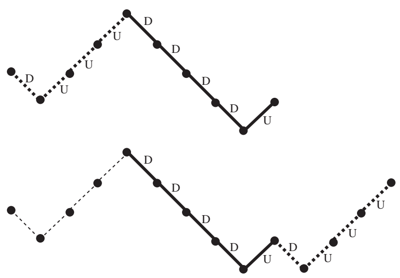

図 14: サイクルシフトによる改造

さて, 2 個の V 字型からなる Dyck パスに対応するような, 4 個の U と 5 個の D からなる長さ 9 の文字列は何個あるでしょうか?

2 個の V 字型からなる Dyck パスは, 左から順に, 下り, 上り, 下り, 上りの 4 つの部分パスで構成されています.

各下りの部分パスは,1 つ以上の文字 D に対応します. D は全部で 5 個ありますから, まず 1 個づつを 2 つの部分パスにそれぞれ与え, のこりの 3 個を 2 つの部分パスに分割して与える方法は

$$ {}_{m+1-\ell+\ell-1}C_{\ell-1} = {}_4 C_1 = 4 $$

個あります.

同様に, 各上りの部分パスは,1 つ以上の文字 U に対応します. U は全部で 4 個ありますから, まず 1 個づつを 2 つの部分パスにそれぞれ与え, のこりの 2 個を 2 つの部分パスに分割して与える方法は

$$ {}_{m-\ell+\ell-1}C_{\ell-1} = {}_3 C_1 = 3 $$

個あります.

よって, 2 個の V 字型からなる Dyck パスに対応するような, 4 個の U と 5 個の D からなる長さ 9 の文字列は $3 \times 4 = 12$ 個あります. これらの文字列のうち, ちょうど 1/2 だけ, つまり 6 個だけが 4 本の辺を持ち, 2 個の葉を持つ順序木に対応します.

この説明を一般化しましょう. m 本の辺を持ち, ℓ 個の葉を持つ順序木は何個あるでしょうか. 上で紹介したように,Narayana 数

$$ C(m,\ell) = ({}_m C_{\ell-1} \cdot {}_{m-1} C_{\ell-1})/\ell $$

個であることが知られています. ぜひ説明に挑戦してみて下さい.

参考文献

N. Dershowitz and S. Zaks, The cycle lemma and some applications, Europ. J. Comb. 11 (1990), 35-40.

中野眞一 群馬大学大学院・工学研究科