數學傳播 33卷1期, pp. 65-75

棘手, 但很迷人
——從有序樹的計數看數學模型

柳柏濂

一. 問題的提出—從細胞繁殖到有序樹

細胞繁殖 (cell-growth problem) 是生物數學中的一個著名問題: 從一個正方形的「細胞」開始, 每次從其中一邊生長出一個新的細胞。如此繼續, 所形成的「動物」不含空洞。那麼, n 個細胞 (正方形) 能形成多少個不同的動物?

我們試考察 $n=4$ 的情形。四個細胞所形成的動物如下:

$n=4$

(1) (2) (3) (4)

圖1

對於一般的 n。設 $f(n)$ 表示 n 個細胞的不同動物數。迄今, 只知道如下的結果 ($n \le 24$) (見 [1])

n	1	2	3	4	5	6	7	8	9	10	$\cdots$	23	24
$f(n)$	1	1	2	5	12	35	108	369	1285	4655	$\cdots$	168,047,007,728	654,999,200,403

我們所說的不同的動物, 指的是在旋轉, 反射變換下不同的動物式樣。

如果我們把細胞的生成限制在平面上, 則所述的問題變成平面細胞繁殖問題。

同樣, 對於 $n=4$, 考察平面細胞的不同式樣, 則圖 1 的 (1), (3), (4) 圖無變化, 但圖 (2) 應變成下列兩個圖。

(1)　　　　(2)

圖2

也就是說, 在平面上用旋轉變換, 圖2中的圖 (1) 無論如何是不能變成圖 (2) 的。

處於三維空間的你, 習慣於三維空間中的條件變換, 往往不經意地把三維空間中的反射, 用於限制在二維空間的圖形中, 因而誤認爲圖 2 中的圖 (1), (2) 是同一個圖。

正是利用你習慣性思維造成誤判, 電子商炮製出流行的俄羅斯方塊遊戲。當方塊在平面上翻著“筋斗”下落時 (圖3), 你能準確地判斷出它能否嵌入下方的凹洞中嗎?

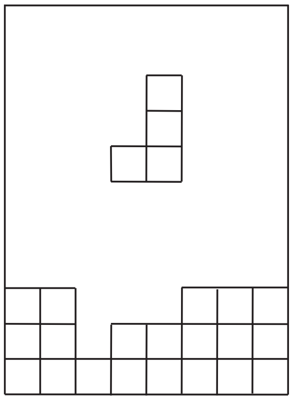
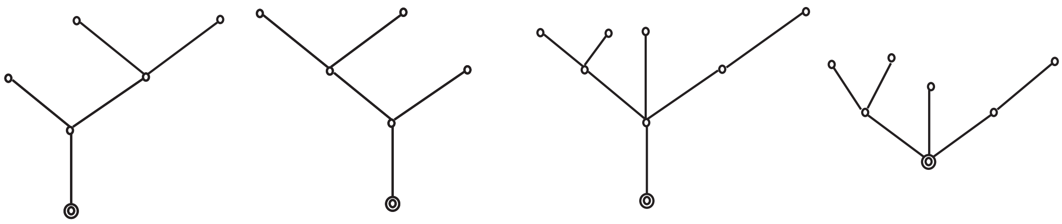


圖3

同樣限制在平面上, 我們考慮在計算機演算法中不可或缺的樹圖 (tree)。

所謂 n 階樹, 就是 n 個頂點的無圈連通圖, 而每個頂點所連接的邊數稱爲該頂點的度 (degree)。易知: n 階樹有 $n-1$ 條邊, 且至少有兩個度爲 1 的點, 即懸掛點 (pendent vertex), 或稱爲葉點。

限制在平面上變換的樹, 稱爲平面樹。把樹的某一點作特殊標號, 這樣的樹稱爲根樹 (rooted tree)。此特殊的點稱爲根點。根點是懸掛點的根樹稱爲植樹 (plant tree), 如同上述俄羅斯方塊的原理, 圖 4 中兩棵平面植樹 (1) 和 (2) 是不同的。



(1) 有序植樹且是二叉樹　(2) 有序植樹且是二叉樹　(3) 有序植樹　(4) 有序樹

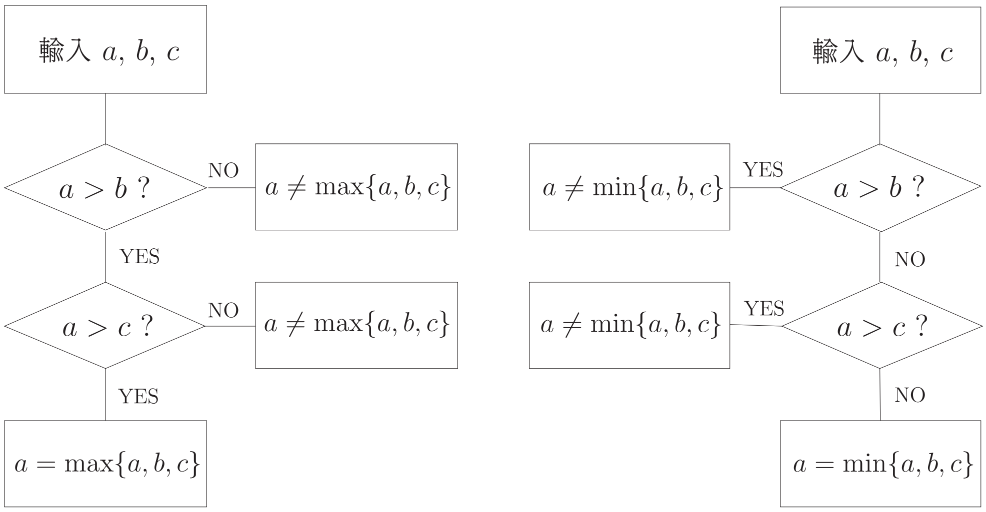
圖4

因爲圖 4 中的樹 (1) 不能通過旋轉變爲樹 (2)。於是, 在平面樹中「枝椏」都是有序的。因此, 平面植樹 (圖 4 中 (1) (2) (3)) 又稱爲有序植樹 (ordered plant tree), 平面根樹 (圖

4(4)) (根點不一定是懸掛點) 稱爲有序樹 (ordered tree), 而像圖 4 的 (1) (2) 那樣僅由 1 度點和 3 度點組成的植樹, 稱爲二叉 (有根) 樹 (2-ray tree)。二叉樹的每個分叉都可以分成左右兩枝。

我們看看以圖 4 中的兩棵二叉樹爲模型的演算法框圖, 就可以瞭解它們爲什麼應該看作不同。

我們約定左枝表示 Yes, 右枝表示 No。以比較 a, b, c 3個不同的實數的大小爲例。



(1)　　　　(2)

圖5

圖 5 中的框圖 (1) 的樹圖模型是圖 4 的根樹 (1) (倒過來), 它判斷是否有 $a = \max\{a, b, c\}$ 而圖 5 中的框圖 (2) 的樹圖模型是圖 4 的根樹 (2)。它判斷是否有 $a = \min\{a, b, c\}$。

一個自然提出的問題是: n 階有序植樹有多少種不同的式樣?

二. 組合模型—枯燥的排隊激發鮮活的靈感

排隊, 在生活中屢見不鮮, 你曾想過, 排隊與有序樹有什麼聯系嗎?

試看下面的一個排隊買票問題。

戲院票房前有 $2n$ 個人買票, 其中 n 個人只有 5 角一張的紙幣, 其餘的 n 個人只有 1 元一張的紙幣。在開始賣票時, 票房裏無任何零錢, 而每個人只能買一張 5 角的票, 問買票的人有多少種不同的排隊方法, 能使每個人都能買到一張 5 角的票?

我們分析一下, 這種排隊方法, 它必須具有下列特徵:

(A)
- (A_1)　$2n$ 個人排隊, 每個人手裡不是拿著 5 角就是拿著 1 元
- (A_2)　恰有 n 個人拿 5 角, 另外 n 個人拿 1 元
- (A_3)　爲了保證:「每個人都能買到一張 5 角的票」, 第一個人必須拿 5 角, 第 2 個人可拿 5 角或 1 元。若第 2 個人拿 1 元, 則第 3 個人必須拿 5 角, 否則票就賣不出去; 若第 2 個人拿 5 角, 則第 3 個人 (甚至第 4 個人) 都可拿 5 角或 1 元 $\cdots$。

爲了使上述特徵變得清晰, 我們把上述組合模型數字化。把拿 5 角的人表爲 0, 拿 1 元的人表爲 1, 於是符合要求的排隊, 可表示爲如下特徵:

(B)
- (B_1)　長爲 $2n$ 的 0, 1 序列
- (B_2)　序列中恰有 n 個 0 和 n 個1, 且第 1, 2 個數爲 0, 1
- (B_3)　序列的第 1 項爲 0, 且由首項到任一項所形成的子序列, 0 的個數不少於 1 的個數

我們稱滿足 (B_1), (B_2), (B_3) 特徵的 (0, 1) 序列爲 B 序列。

現在, 回過頭來, 看看我們的有序植樹。(圖6)

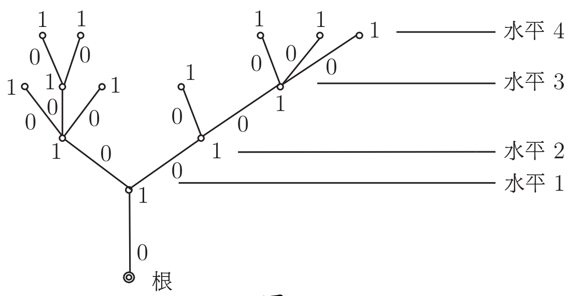


圖6

有序植樹的任一個頂點 v 到根點都有唯一的一條路, 路所含的邊數稱爲點 v 的水平。因此, 任一有序樹都可以分成若干水平的頂點。

我們把一個有序植樹的邊表爲 0, 頂點 (除了根點) 表爲 1, 從根點相連的邊開始, 依次從低水平到高水平, 先點再邊, 同一水平的點從左到右, 兩水平之間的邊亦從左到右, 我們便可讀出一個 0, 1 序列。例如, 圖 6 中所示的樹的 (0, 1) 序列爲

$$0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1$$

因爲任一 $(n+1)$ 階有序植樹, 除了根點外, 恰有 n 個頂點, n 條邊。因此它的 (0, 1) 序列滿足特徵 (B_1), (B_2), 而對於有序樹, 除了根點, 第一項必從邊開始, 即 (0, 1) 序列第 1 項必爲 0 第 2 項必爲 1, 且由於每一水平的頂點都讀在邊之後, 故序列滿足 (B_3)。

容易驗證, 任一個 B 序列, 都可對應畫出唯一一個有序植樹。於是一個 $(n+1)$ 階有序植樹一一對應於一個數字化的組合模型, 即 B 序列。

因此, 求 $(n+1)$ 階有序植樹的個數, 即求 B 序列的個數。

三. 代數模型—此時無聲勝有聲

0 和 1 是世界上最簡單的兩個數, B 序列就是用一個 0, 1 的數組給出了有序植樹的一個組合模型。

可是, 組合模型並沒有直接給出數學性質, 經過分析, 我們給出了 (B_1), (B_2), (B_3) 的描述性的特徵。

下面, 我們把組合模型進一步代數化。

由 B 序列的特徵 (B_1), (B_2), (B_3), 可設數組為 $0, 1, x_1, \ldots x_{2n-2}$。於是, B 序列的特徵表述為下列代數化條件

$$
(\mathrm{C})\begin{cases}
x_i = 0 \ \text{ 或 } \ 1, & i = 1, 2, \ldots, 2n-2 \qquad (C_1)\\
x_1 + x_2 + \cdots + x_{2n-2} = n-1, & n \ \text{ 為正整數} \qquad (C_2)\\
x_1 + x_2 + \cdots + x_j \le \dfrac{1}{2} j, & j = 1, 2, \ldots, 2n-3 \qquad (C_3)
\end{cases}
$$

條件 (C) 沒有任何描述性的語句, 但卻簡明地刻畫了 B 序列的數學本質, 特別是 (C_3)。也許, 從性質 (B) 寫出式子 (C) 並非難事。然而, 在閱讀數學書籍時, 能夠把式子 (C) 描述成性質 (B), 卻需要一定的數學修養。

於是, 要求出 $(n+1)$ 階有序植樹的個數。只須求出滿足條件 (C) 的所有數組 $(x_1 \ \ x_2 \ \cdots \ x_{2n-2})$ 的個數。

代數模型 (C) 提供了我們解決類似問題的思路。試看下列選舉問題。

某次選舉中有侯選人 P 和 Q, P 得票 p 張, Q 得票 q 張, $p>q$。要使開票過程中, P 的得票數恒領先於 Q 的可能情況有多少種?

類似前面的分析, 我們不難得出這一問題的代數模型。

我們把 $p+q$ 張選票按開票先後順序記作 $x_1, x_2, \ldots, x_{p+q}$, 其中選 P 的票記作 $x_j=0$, 選 Q 的票, $x_j=1$, 因 P 得票數恒領先於 Q; 顯見, 選舉的代數模型為:

$$
\begin{cases}
x_i = 0 \ \text{ 或 } \ 1, & i = 1, 2, \ldots, p+q\\
x_1 + x_2 + \cdots + x_{p+q} = q, & \\
x_1 + x_2 + \cdots + x_j < \dfrac{1}{2} j, & j = 1, 2, \ldots, p+q-1
\end{cases}
$$

本文前面提到的二叉 (有根) 樹是研究計算機資訊存儲的一類重要結構。例如, 與二叉樹的計數直接聯繫的一個問題是: 按一定順序排列的 9 個字母 a, b, c, d, e, f, g, h, i, j, 能構

成多少個不同的不可結合積。圖 7二叉樹 (省掉了一個根點及邊) 展示的, 就是其中對二元運算 $*$ 的一個不可結合積。我們注意到 9 個字母按規定次序排在葉點上。邊連字母表示運算 $*$ 。

有序樹與二叉樹的區別在於: 前者的每個頂點不限於 1 度點與 3 度點。可是, 運用代數模型, 我們卻能知道: n 個葉點的二叉樹與 $(n+1)$ 個頂點 (包括根點) 的有序樹之間是1−1對應的。

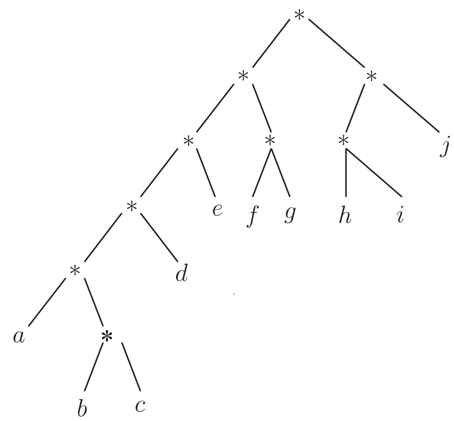


圖7. $a*(b*c)*d*e*(f*g)*((h*i)*j)$

容易證明: 若二叉樹的 3 度頂點個數爲 n, 則葉點的個數必爲 $n+1$

證: (對 n 用數學歸納法) $n=1$, 顯然 (如圖8) 設二叉樹有 n 個 3 度頂點時, 葉點的個數爲 $n+1$。現增加一個 3 度頂點即把一個葉點改爲一個 3 度點, 就必增加兩邊, 於是葉點共有 $(n+1)-1+2=n+2$ 個, 證畢。

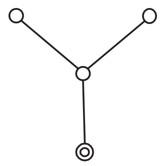


圖8

因此, 有 n 個葉點的二叉樹, 它的 3 度點有 $n-1$ 個, 連同根點二叉樹共有 $2n$ 個頂點, 把這些頂點用 $1,2,\ldots,2n$ 編號。編號的規則是: 從根點開始由低水平到高水平, 同一水平的頂點從左到右。如圖 9 就是一個編了號的 4 葉點的二叉樹。令每一點 i 對應於 0 或 1, 規定 3 度點及根點對應於 0, 葉點對應於 1。於是每一個編號的二叉樹一一對應於一個長爲 $2n$ 的 $(0,1)$ 數組, 其性質是

(1) 有 n 個 0, n 個 1

(2) 對任一個 $j<2n$, 由 1 到 j 的 j 個點中 3 度點必多於 1 度點。

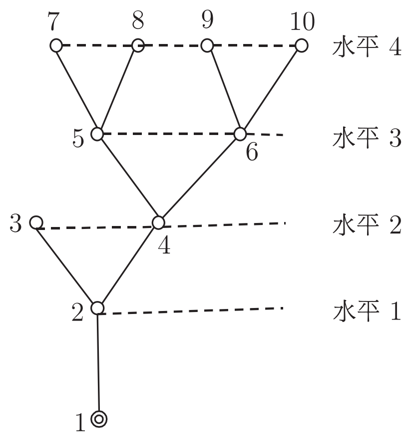


圖9

於是, 有 n 個葉點的二叉樹的代數模型是

$$
\text{(D)}\begin{cases} x_i=0 \ \text{ 或 } \ 1, & i=1,2,\ldots,2n \\ x_1+x_2+\cdots+x_{2n}=n, & n \ \text{ 爲正整數} \\ x_1+x_2+\cdots+x_j<\dfrac{1}{2}j, & j<2n \end{cases}
$$

一個有趣的事實是, 代數模型 (C) 和 (D) 是等價的。因爲如果我們把一個 B 序列的後 $2n-2$ 項, 前面加一個 0, 後面加一個 1。即 $0\ \ x_1\ \ x_2\ \ \cdots\ \ x_{2n-2}\ \ 1$ 就是一個滿足 (D) 的數列, 而任一個滿足 (D) 的數列 $x_1\ \ x_2\ \ \cdots\ \ x_{2n}$ 必須有 $x_1=0$, 且 $x_{2n}=1$, 去掉此 0, 1 後把數列由 x_2 開始重新按自然數序編號, 便得到一個長爲 $2n-2$ 的滿足 (C) 的數列。由此, 對代數模型的討論, 我們知道:

$(n+1)$ 階有序植樹的個數等於 n 個葉點的二叉樹的個數。作爲一個練習, 讀者可寫出 n 階有序樹的代數模型, 由此可證明 n 階有序樹的個數等於 $(n+1)$ 階有序植樹的個數。

四. 幾何模型—照鏡子的啓示.

你每天都照鏡子嗎?

在鏡子前, 你看到了另一個世界, 它和你生活中的世界是一樣的, 這叫做反射。因爲反射, 產生了對稱, 這在數學中稱爲「反射原理」。我們就用這個反射原理, 構造出有序樹的一個幾何模型, 從而達到求解的目的。

在構造幾何模型之前, 先看一個衆所周知的郵遞員問題:

在一個 p 段直路, q 段橫路組成的方格幹道網中 (圖10), 郵遞員從郵局 A 到達客戶 B, 可以走多少條不同的最短路?

易知, 由 A 到 B 走最短路 (簡稱爲 L 路徑) 至少要走 p 段直路, q 段橫路。不同的 L 路徑, 相當於 p 段直路和 q 段橫路不同組合數, 其方法數是:

$$\frac{(p+q)!}{p!q!}$$

的確, 代數模型 (C) 能簡明地刻畫有序數的特徵。然而利用它卻不易算出滿足 (C) 的解的個數, 下面我們構造出 B 序列的幾何模型, 幫助我們直接求出問題的解。

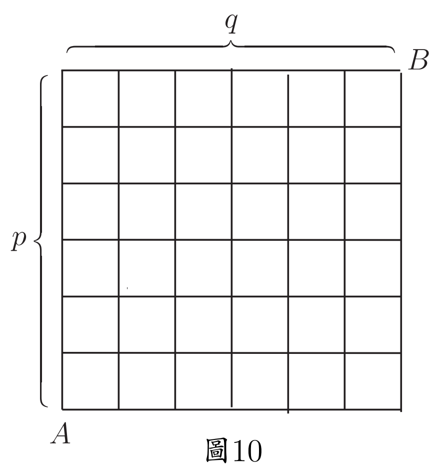
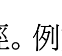


圖10

建立平面坐標系 XOY。對任一個 B 序列的後 $2n-2$ 項 $(x_1\ \ x_2\ \ \cdots\ \ x_{2n-2})$。對應於平面的一條從 $(0,0)$ 到 $(2n-2,0)$ 的由點對角線組成的折線。由 x_1 開始, 若 $x_i=0$, 則對應於一段  的線段; 若 $x_i=1$, 對應於一段  的線段。由這兩類線段組成的路徑, 我們稱爲 T 路徑。例如圖 4 中圖 (1) 的有序樹的 B 序列是 0 1 0 0 1 1 0 0 1 1 其後面 8 項對應的 T 路徑如圖 11

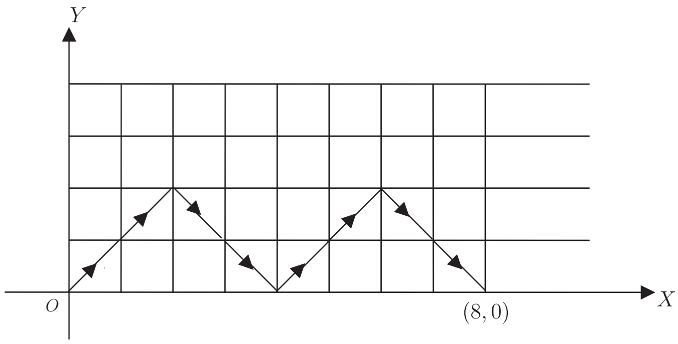


圖11

由 B 序列的代數模型 (C) 我們知道:

(T)
- (1) T 路徑由 和 線段組成 (見性質C_1)
- (2) T 路徑有 $n-1$ 段 及 $n-1$ 段 , 因此由 (0, 0) 開始, 到 $(2(n-1), 0)$ 點結束 (見性質C_2)
- (3) T 路徑由 (0, 0) 開始, 只在 XOY 的上半平面 (第一象限) (見性質C_3)

T 路徑給出了有序樹的一個幾何模型。求出所有滿足性質 (T) 的 T 路徑就能求得 $n+1$ 階有序樹的個數。

我們先計算從 $(0,0)$ 到 $(2(n-1),0)$ 的所有 T 路徑。考察由直線 $y=x$, $y=x-2(n-1)$, $y=-x$, $y=-x+2(n-1)$, 組成的正方形 Q。整點把 Q 分成 $(n-1)\times(n-1)$ 個小正方形 (如圖12)。

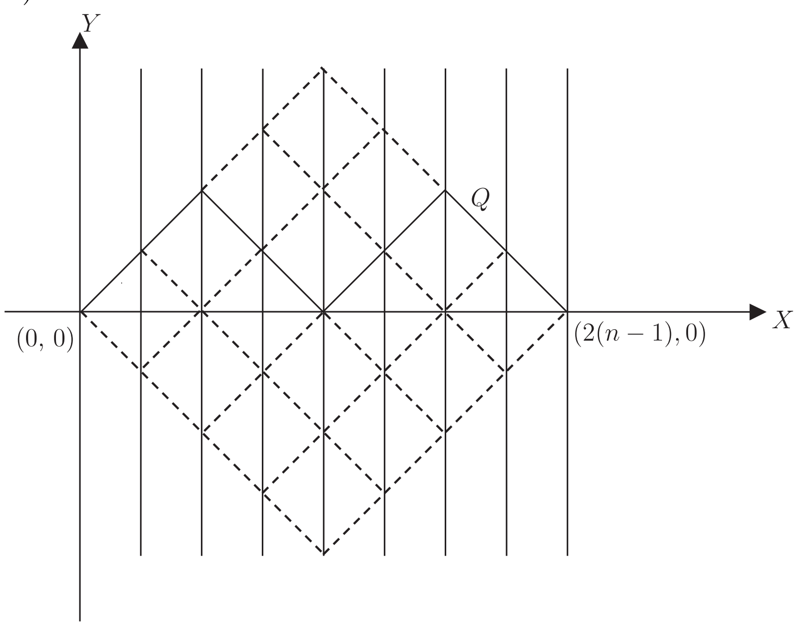


圖12

於是從 $(0,0)$ 到 $(2(n-1),0)$ 的 T 路徑等價於正方形 Q 由 $(0,0)$到 $(2(n-1),0)$ 的 L 路徑, 總數是

$$\frac{(2n-2)!}{(n-1)!(n-1)!}$$

可是, 算出來的結果並非是滿足性質 (T) 的 T 路徑總數。它摻雜了很多不滿足性質 (T) 的 T 路徑, 即從 $(0,0)$ 到 $(2(n-1),0)$ 越過了 x 軸的 T 路徑, 或者說從 $(0,0)$ $(2(n-1),0)$ 接觸到直線 $y=-1$ 的 T 路徑。我們簡稱這些 T 路徑爲「壞 T 路徑」。如果我們把壞 T 路徑的總數算出來, 則我們的問題就告解決。下面, 我們就用所謂反射原理的方法求出所有壞 T 路徑的總數。

試看任一條壞 T 路徑 T_1, 我們把它以直線 $y=-1$ 爲對稱軸反射成一條以 $(-2,0)$ 爲起點, $(2(n-1),0)$ 爲終點的 T_1' 路徑 (如圖13), T_1' 路徑落在是一個由直線 $y=x-2$, $y=x-2(n-1)$, $y=-x-2$, $y=-x+2(n-1)$ 組成的正方形 Q' 內, 是一條 Q' 中由 $(-2,0)$ 到 $(2(n-1),0)$ 的 L 路。注意到 Q' 是一個 $n\times(n-2)$ 的矩形。

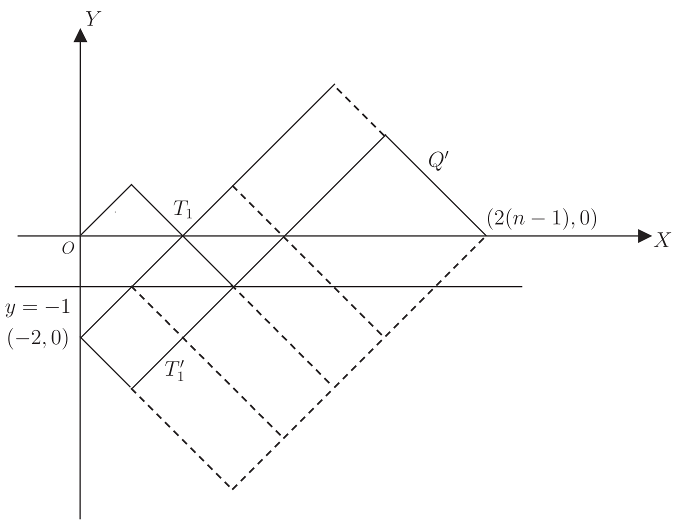


圖13

Q' 所有從 $(-2,0)$ 到 $(2(n-1),0)$ 的 L 路總數爲

$$\frac{(n+n-2)!}{n!(n-2)!}=\frac{(2n-2)!}{n!(n-2)!}$$

於是, 滿足性質 (T) 的 T 路徑總數爲

$$\begin{aligned}&\frac{(2n-2)!}{(n-1)!(n-1)!}-\frac{(2n-2)!}{n!(n-2)!}\\&\quad=\frac{1}{n}\frac{(2n-2)!}{(n-1)!(n-1)!}=\frac{1}{n}\binom{2n-2}{n-1}\end{aligned}\tag{1}$$

在組合數學裏, $\frac{1}{n+1}\binom{2n}{n}=C_n$ 稱為卡塔蘭 (Catalan) 數, 這是 19 世紀比利時數學家 Catalan 在研究 n 個因數的不可結合積的個數問題時提出來的。在計算數學、計算機科學、通訊理論的很多組合問題都歸結為卡塔蘭數。由 (1) 式知: $n+1$ 階有序植樹的個數是卡塔蘭數 C_{n-1}。而因為 n 個因數的不可結合積等價於 n 個葉點的二叉樹。由第三節的結論, n 個因數的不可結合積的個數也就是卡塔蘭數 C_{n-1}。

五. 生成函數—又回到代數

幾何模型給我們提供了一條直觀解決問題的途徑。然而, 我們更希望用「數學味」更濃的代數方法。

生成函數, 又給我們打開一個回到代數的新天地。

設 n 階的有序植樹的個數為 p_n。作序列 $\{p_n\}$ 的生成函數 $P(x)=\sum\limits_{n=1}^{\infty}p_nx^n$, 這裏約定 $p_1=0$。

以圖 4(3) 的有序植樹為例。把一個有序植樹的根點及與根相鄰的點去掉, 就得到若干支, 不妨設 r 支 (無根) 的有序植樹。反之, 把 r 支有序植樹的根「粘」在一起, 再添上一個新的懸掛根點, 可得一棵新的有序植樹。這個過程, 用生成函數來表達就是

$$P(x)=x^2\sum_{r=0}^{\infty}\left(\frac{P(x)}{x}\right)^r \tag{2}$$

這裏 $\frac{P(x)}{x}$ 表示每一有序植樹去掉根點, $\left(\frac{P(x)}{x}\right)^r$ 表示把 r 支這種 (無根) 植樹粘起來 (其中 r 可以是 $0,1,2,\ldots$), 而乘 x^2, 就是最後添上粘在一起的一個點, 及另一個新懸掛根點。

把 (2) 式右邊整理 (注意, 生成函數只著眼於它的展開級數各項的系數。級數的求和無須顧及它的收斂性。見《數學傳播》22 卷 2 期拙作「別瞧不起它, 這個中學教材中的公式」) 得

$$P(x)=x^2\cdot\frac{1}{1-\frac{P(x)}{x}}$$

於是得 $P^2(x)-xP(x)+x^3=0$

解關於 $P(x)$ 的二次方程得

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

因 $P(x)$ 中 x 的一次項係數為 0, 故上式的括號中必須取「$-$」號。於是

$$P(x)=\frac{x}{2}(1-\sqrt{1-4x}). \tag{3}$$

在中學數學裏, 我們學過二項式定理

$$(1-x)^m = 1 - mx + \frac{m(m-1)}{2!}x^2 - \frac{m(m-1)(m-2)}{3!}x^3 + \cdots + (-1)^m \frac{m!}{m!}x^m$$

事實上, 上式的 m 不限於正整數, 可以推廣到任意實數。推廣的二項式定理是

$$\begin{aligned}(1-x)^m = 1 - mx + \frac{m(m-1)}{2!}x^2 - \frac{m(m-1)(m-2)}{3!}x^3 + \cdots \\ +(-1)^k \frac{m(m-1)\cdots(m-k+1)}{k!}x^k + \cdots\end{aligned}$$

運用推廣的二項式定理, 我們把式 (3) 展開, 得

$$\begin{aligned}P(x) &= \frac{x}{2}\left(2x + \frac{2}{2}\binom{2}{1}x^2 + \frac{2}{3}\binom{4}{2}x^3 + \cdots + \frac{2}{n}\binom{2n-2}{n-1}x^n + \cdots\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{n}\binom{2n-2}{n-1}x^{n+1}\end{aligned}$$

對照兩邊 x^{n+1} 的係數, 便得

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

生成函數是十九世紀著名數學家拉普拉斯 (Laplace) 爲解決概率問題而首先引進的方法, 它顯然有點難度, 不如其他模型那樣巧妙, 但卻是組合計數中的典型方法。

它是一支帶刺的玫瑰, 棘手, 但很迷人。

參考文獻

1. Redelmeier, D. H., *Counting polyominoes: Yet another attack*, Discrete Math., Vol. 36, p.191-203, 1981.
2. 柳柏濂, 鏡子 · 路子 · 珠子—有趣的卡塔蘭數, 廣東教育出版社, 1994.7 。

———本文作者任教中國廣州華南師範大學數學系———