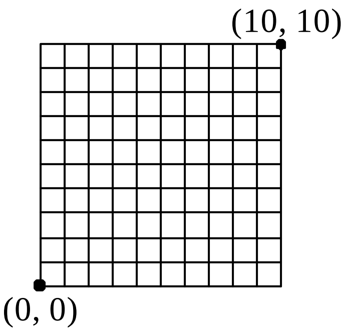
Catalan Numbers 簡介

許介彥

大葉大學　通訊與計算機工程學系

暖身題

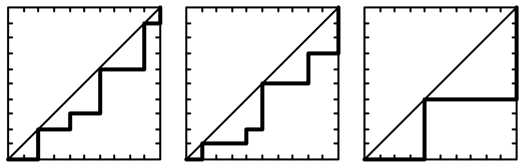
小明的家及上班的地點分別位於下圖中坐標為 (0,0) 及 (10,10) 的位置，圖中的直線與橫線代表著道路。假設小明想由家裡出發，循最短的路徑到達上班地點，也就是說，他隨時不是向東就是向北，請問他總共有幾種走法？



這是一個簡單的問題。不管小明怎麼走，他都須向東及向北各走 10 格，如果我們將向東一格及向北一格分別用一個 E 及一個 N 來表示，那麼小明的任何一種走法都可以用一個包含了 10 個 E 及 10 個 N 的字串來描述；反過來說，任何一個由 10 個 E 及 10 個 N 組成的字串就對應到一種可能的走法。因此，由 10 個 E 及 10 個 N 可組成多少個字串就對應到小明有多少種走法，而這個數很顯然是由 20 個位置中選出 10 個來作為 E 的位置的方法數，也就是 $C(20,10)$；一旦選定了 10 個 E 的位置，剩下的位置也就是 10 個 N 的位置。

加上一條限制

假設有一條河流筆直地流經小明的家及上班的地點，這兩個處所都位於河的東側；由於河上沒有任何橋樑可供兩岸往來，因此小明上班的路徑無法跨過這條河流（也就是圖中連接 (0,0) 與 (10,10) 的對角線）；每當小明向北碰到對角線就不得不轉向東行。在這樣的限制下，我們想要知道小明有多少種走法。下圖顯示了三種可能的路徑：



如果我們沿用前述以 E 及 N 來描述路徑的方式，那麼上面三個圖的最左邊的圖顯示的路徑相當於

$$EENNEENEENNNEEENNNEN$$

中間的圖顯示的路徑相當於

$$ENEEENENNNEEENNEENNN$$

最右邊的圖顯示的路徑則相當於

$$EEEENNNNEEEEEENNNNNN$$

讀者不難看出這種字串的開頭第一個字母一定是 E，最後一個字母一定是 N，也就是說，小明一開始一定是向東，而抵達終點時一定是朝北。

如果沒有河流的限制，我們已經知道小

明總共有 $C(20,10)$ 種走法，這些走法可以分為兩類，一類是走的過程中沒有跨過河流的（也就是符合要求的），另一類則是會跨過河流的（也就是「不可行」的）。

想求出符合要求的走法的總數，有了暖身題的經驗，我們還是可以試著由字串著手。對任意一個由 10 個 E 及 10 個 N 組成的字串，我們能否由字串本身判斷它對應到的路徑是可行的或是不可行的呢？答案是肯定的。如果小明上班的路徑沒有跨過河流，那麼由起點出發後，走的過程中不管在任何時候，他已經走過的向東的格子數絕對不會少於已經走過的向北的格子數，頂多兩數相等，也就是小明正位於對角線上（來到河邊）的情形。

因此對任意一個由 10 個 E 及 10 個 N 組成的字串，我們只須由最左邊一個一個字母往右看，並隨時分別注意已經看過的 E 及 N 的個數；如果檢視過程中 E 的個數時時都不小於 N 的個數，這個字串所對應的路徑就是可行的，否則就是不可行的。

因此，我們的工作只剩下找出在全部 $C(20,10)$ 個由 10 個 E 及 10 個 N 組成的字串裡面，由左往右看的過程中 E 的個數時時都不小於 N 的個數的字串有幾個就行了。雖然目標明確，這個問題其實由正面並不容易解決，以下我們試著由反面著手，先設法求出不可行的走法總共有幾種，一旦知道了，再用 $C(20,10)$ 減去不可行的走法數就是符合要求的走法數了。

不可行的走法數也就是由左往右看的過程中 N 的個數會在某個時候大於 E 的個數的字串個數。舉個例子，如果某個由 10 個 E 及 10 個 N 組成的字串為

$$EENNNENEENNNEEENNNEE$$

當我們看到第五個字母時，我們已經可以確定這個字串對應到的路徑是不可行的，因為字串最前面的五個字母包含了三個 N 及兩個 E，在第五個字母（也就是第三個 N）出現時，小明已經跨過了河流。以上字串中，在第三個 N 之前有兩個 E 及兩個 N，而第三個 N 之後還有 8 個 E 及 7 個 N；如果我們將第三個 N 之後所有的 E 以 N 取代，所有的 N 以 E 取代，所得的結果為

$$EENNNNENNEEENNNEEENN$$

其中讓小明跨過河流的第一個 N 我們特別以陰影顯示。上面的字串包含了 9 個 E 及 11 個 N，也就是 N 比 E 多兩個，這是合理的，因為原來的字串中，由最左邊看到第三個 N 為止，N 的個數已經比 E 多一個；第三個 N 之後，N 原來比 E 少一個，因此第三個 N 之後的字串經過轉換後，N 會比 E 再多一個，所以整體而言 N 會比 E 多兩個。當然，小明如果真的依上述轉換後的字串來走，即使允許他可以跨河，終點也已經不再是他上班的地點了，不過這不是我們目前的重點，請讀者安心往後看就會明白了。

再舉一個例子；以下是另一個由 10 個 E 及 10 個 N 組成的「不可行」的字串：

$$ENENENENNENENEENENNE$$

我們同樣先由左而右找到第一個會使得小明跨過河流的 N（以陰影顯示），然後將其後所有的 E 以 N 取代，所有的 N 以 E 取代，得如下結果：

$ENENENENNNENENNNENEEN$

這又是另一個由 9 個 E 及 11 個 N 組成的字串。讀者不難看出，對任何一個由 10 個 E 及 10 個 N 組成的不可行的字串而言，經過上述步驟一定會得到一個包含 9 個 E 及 11 個 N 的字串，而且任意兩個不同的不可行的字串經過轉換後得到的一定是兩個不同的各自包含 9 個 E 及 11 個 N 的字串；換句話說，這種轉換是「一對一」(one-to-one）的。

另一方面，是不是任意一個由 9 個 E 及 11 個 N 組成的字串都一定是某個由 10 個 E 及 10 個 N 組成的不可行的字串轉換後的結果呢？是的。對任意一個由 9 個 E 及 11 個 N 組成的字串，如果我們由左往右看，N 的個數一定會在某個時候第一次超過 E 的個數（因為整個字串的 N 比 E 還多），我們如果將這個 N 之後所有的 E 以 N 取代，所有的 N 以 E 取代，所得結果將是一個由 10 個 E 及 10 個 N 組成的不可行的字串。舉例來說，以下是一個由 9 個 E 及 11 個 N 組成的字串：

$ENENEENENENENNNENENNNE$

找到第一個使得 N 的個數比 E 還多的 N（以陰影顯示）之後，將其後所有的 E 以 N 取代，所有的 N 以 E 取代，可得一個由 10 個 E 及 10 個 N 組成的不可行的字串：

$ENENEENENENENNNNENENEEEN$

因此，前述的轉換不只是一對一的，而且還是「映成」(onto）的，也就是說，所有包含 10 個 E 及 10 個 N 的不可行的字串與所有包含 9 個 E 及 11 個 N 的字串之間存在著一個「映射」(bijection）關係，由 10 個 E 及 10 個 N 可以組成的不可行字串的總數就等於由 9 個 E 及 11 個 N 可以組成的字串總數，而這個數很顯然是 $C(20,9)$，因此小明不跨過河流上班的走法總共有

$$C(20,10)-C(20,9)=\frac{20!}{(10!)(10!)}-\frac{20!}{(9!)(11!)}\text{ 種。}$$

將問題一般化

我們可以很容易地將上述討論一般化，以下我們假設小明的家及上班地點分別位於 $(0,0)$ 及 (n,n) 的位置，這裡的 n 是任意正整數；前面的討論中，n 的值是 10。

首先，如果沒有河流的限制，小明上班總共有 $C(2n,n)$ 種走法；若有河流流經 $(0,0)$ 與 (n,n)，對任意一個包含了 n 個 E 及 n 個 N 的不可行的字串而言，我們都能將它轉換成一個包含 $(n-1)$ 個 E 及 $(n+1)$ 個 N 的字串，而由 $(n-1)$ 個 E 及 $(n+1)$ 個 N 組成的字串總共有 $C(2n,n-1)$ 個，因此如果小明不跨過河流上班的走法總共有 C_n 種，那麼

$$\begin{aligned}
C_n &= C(2n,n)-C(2n,n-1)\\
&= \frac{(2n)!}{(n!)(n!)}-\frac{(2n)!}{(n-1)!\,(n+1)!}\\
&= \frac{(2n)!}{n!\,(n-1)!}\left(\frac{1}{n}-\frac{1}{n+1}\right)\\
&= \frac{(2n)!}{n!\,(n-1)!}\cdot\frac{1}{n(n+1)}\\
&= \frac{(2n)!}{(n!)(n!)}\cdot\frac{1}{n+1}=\frac{1}{n+1}\binom{2n}{n}\ .
\end{aligned}$$

上式在 $n=0$ 時可得 $C_0=1$。由 $C_0, C_1, C_2, \ldots\ldots$ 形成的數列是數學上很有名的一個數列，英文稱為 Catalan numbers，因比利時數學家 Eugène-Charles Catalan（1814–1894）而

得名；這個數列的最前面十二項依序為 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786；對任意正整數 n，此數列的相鄰兩項的比值為

$$\frac{C_n}{C_{n-1}}=\left(\frac{1}{n+1}\cdot\frac{(2n)!}{(n!)(n!)}\right)\Big/\left(\frac{1}{n}\cdot\frac{(2n-2)!}{(n-1)!(n-1)!}\right)$$

$$=\ \frac{n}{n+1}\cdot\frac{(2n)(2n-1)}{n^2}\ =\ \frac{4n-2}{n+1}$$

當 n 趨於無窮大，此比值趨近於 4。

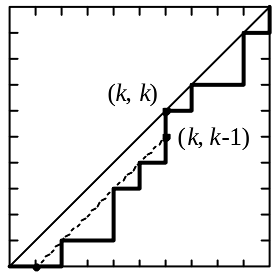
數列的遞迴表示方式

假設由 $(0,0)$ 至 (n,n) 且沒有跨過河流的走法總共有 C_n 種，讓我們試試看能不能用遞迴的方式定義數列 $C_0, C_1, C_2, \ldots\ldots$。

為了簡單起見，還是先考慮 $n = 10$ 的情形。我們可以將所有符合要求的路徑依由 $(0,0)$ 出發後「第一次」碰到對角線的地點分為 10 類；有些路徑第一次碰到對角線的地點為 $(1,1)$，有些為 $(2,2)$，有些為 $(3,3)$，……，有些為 $(10,10)$ 等，每條路徑都有唯一的一個第一次碰到對角線的地點；如果我們能夠知道這 10 類的每一類各有幾種走法，C_{10} 應該就等於這 10 個數的和。

對任意一個大於 1 且小於 10 的正整數 k 而言，由 $(0,0)$ 出發後第一次碰到對角線的地點是 (k,k) 的上班路徑總共有幾種呢？這種路徑可以看成是由前後兩段連接而成；前段是由 $(0,0)$ 至 (k,k)，後段則是由 (k,k) 至 $(10,10)$。如果我們能求得前段與後段各有幾種走法，前後段連起來總共的走法數應該是此兩數相乘的結果（注意：須相乘而不是相加）。

先看前段。由於 (k,k) 是第一次碰到對角線的地點，因此小明在由 $(0,0)$ 至 (k,k) 的途中沒有碰到對角線，也就是沒有跨過連接 $(1,0)$ 與 $(k,k-1)$ 的直線：



因此，由 $(0,0)$ 至 (k,k) 的途中沒有碰到對角線的走法數就等於由 $(1,0)$ 至 $(k,k-1)$ 且沒有跨過上圖中的虛線的走法數；由於 $(k,k-1)$ 的位置相對於 $(1,0)$ 而言是位於其向東及向北各 $(k-1)$ 個格子的位置，這與由 $(0,0)$ 至 (n,n) 的問題在本質上是一樣的，只是距離不同而已，因此前段有 C_{k-1} 種走法。

後段的情形就更簡單了；由於 $(10,10)$ 的位置相對於 (k,k) 而言是位於其向東及向北各 $(10-k)$ 個格子的位置，因此有 C_{10-k} 種走法。前後段一起考慮，我們得知第一次碰到對角線的地點是 (k,k) 的走法總共有 $C_{k-1}\cdot C_{10-k}$ 種。

接著考慮 $k = 1$ 及 $k = 10$ 的情形。當 $k = 1$，很顯然由 $(0,0)$ 至 $(1,1)$ 只有一種走法，而由 $(1,1)$ 至 $(10,10)$ 有 C_9 種走法，因此 $k = 1$ 時有 $1\cdot C_9\ =\ C_{1-1}\cdot C_{10-1}$ 種走法（C_0 的值為 1）。當 $k = 10$，只有前段而沒有後段，而前段的走法有 C_9 種，C_9 又可以寫成 $C_{10-1}\cdot C_{10-10}$。我們的結論是，當 $n = 10$，由 $(0,0)$ 至 $(10,10)$ 總共有

$$\begin{aligned}C_{10} &= \sum_{k=1}^{10} C_{k-1}C_{10-k} \\ &= C_0C_9 + C_1C_8 + \cdots + C_9C_0\end{aligned}$$

種走法；更一般性地說，對任意正整數 n，由 $(0,0)$ 至 (n,n) 總共有

$$\begin{aligned}C_n &= \sum_{k=1}^{n} C_{k-1}C_{n-k} \\ &= C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-1}C_0\end{aligned}$$

種走法；再配合初始條件 $C_0 = 1$ 就構成了數列 $C_0, C_1, C_2, \ldots\ldots$ 完整的遞迴定義：

$$C_n = \begin{cases} 1 & n = 0 \\ \sum_{k=1}^{n} C_{k-1}C_{n-k} & n > 0 \end{cases}$$

這是用遞迴的方式定義 Catalan numbers 最常見的方式，不過當然不是唯一的方式。舉例來說，由我們前面求此數列相鄰兩項比值的式子其實就可以發展出另外一個遞迴的定義：

$$C_n = \begin{cases} 1 & n = 0 \\ \dfrac{4n-2}{n+1}C_{n-1} & n > 0 \end{cases}$$

投票問題

考慮以下問題：在一場選舉中，只有 A 與 B 兩位候選人，開票的結果是 A 與 B 平分秋色，雙方各得了 n 票（假設沒有廢票）。請問：在將票一張一張開出的過程中，(1) A 的得票數一路領先，直到最後一張票才被 B 追上的可能的開票過程有多少種？(2) A 的票數始終不比 B 少的可能的開票過程有多少種？

這兩個小題和前面我們探討的小明的上班問題其實有密切的關聯；您能夠立刻說出這兩個小題的答案嗎？

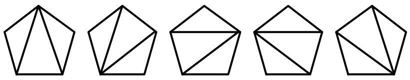
先考慮第二小題。在前面的上班問題中，小明須向東及向北各走 n 格，而且為了不跨過河流，走過的向東的格子數隨時都不能少於走過的向北的格子數；在現在的投票問題中，A 與 B 在開票過程中各得 n 票，而且 A 的票數隨時都不低於 B 的票數；很明顯的，這兩個問題在本質上是一樣的，我們甚至也可以用一個由 n 個 A 及 n 個 B 排列而成的字串來描述開票的過程，一個 A 的票數始終不低於 B 的票數的開票過程就對應到一個由左而右檢視時 A 的累計次數始終不低於 B 的累計次數的字串。

舉例來說，當 $n = 3$，A 與 B 各得 3 票而且開票過程中 A 的得票數始終不低於 B 的得票數的所有可能情形有 $C_3 = 5$ 種：*ABABAB*, *ABAABB*, *AABBAB*, *AABABB*, *AAABBB*。

因此，第二小題的答案為 C_n，而第一小題我們在前面其實也討論過了，相當於由 $(0,0)$ 出發後第一次碰到對角線的地點為 (n,n) 的情形，因此答案為 C_{n-1}。

凸多邊形的三角化問題

考慮以下動作：在一個凸多邊形的內部畫上一些互不相交的對角線（即頂點間的連線），使得多邊形內部被分割成許多三角形，每個三角形的每一邊皆為多邊形的一邊或是多邊形的一條對角線；以上動作稱為多邊形的「三角化」(triangulation)。舉例來說，下面是將一個凸五邊形三角化的所有可能情形（總共有五種可能）：

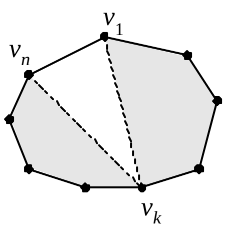


請問：對任意一個大於 2 的正整數 n 而言，將一個凸 n 邊形三角化的方法總共有幾種？

這個問題可以用遞迴的方式來解決。假設 a_n 代表將一個凸 n 邊形三角化的方法數，上圖顯示了 $a_5=5$；我們希望用遞迴的方式定義數列 a_3，a_4，a_5, ...。

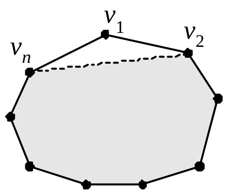
首先，假設我們將一個凸 n 邊形的 n 個頂點依順時鐘方向分別命名為 v_1，v_2, ..., v_n，多邊形的 n 個邊因此是 $\overline{v_1v_2}$，$\overline{v_2v_3}$, ..., $\overline{v_{n-1}v_n}$，$\overline{v_nv_1}$。考慮其中的一邊 $\overline{v_nv_1}$；在我們依某種方式將多邊形三角化之後，$\overline{v_nv_1}$ 必定會是某個三角形的一邊，也就是說，v_1 與 v_n 必定會是某個三角形的三個頂點中的兩個，而這個三角形的第三個頂點有可能是 v_2, v_3, ..., v_{n-1} 等 $(n-2)$ 個點中的任何一點；如果我們能夠知道當第三個頂點是這 $(n-2)$ 個點中的每一點時各有幾種方法可以將多邊形三角化，a_n 應該就等於這 $(n-2)$ 個數的和。

假設以 $\overline{v_nv_1}$ 為一邊的三角形的第三個頂點為 v_k。先考慮 $3\le k\le(n-2)$ 的情形，如下圖所示：



三角形 $v_nv_1v_k$ 將多邊形的內部分隔成兩邊，一邊是一個以 v_1，v_2, ..., v_k 為頂點的凸 k 邊形，另一邊則是一個以 v_k，v_{k+1}, ..., v_n 為頂點的凸 $(n-k+1)$ 邊形，將這兩個多邊形三角化的方法分別有 a_k 與 a_{n-k+1} 種，因此當第三個頂點為 v_k 時，將 n 邊形三角化的方法總共有 $a_k a_{n-k+1}$ 種。

接著考慮 k 等於 2 的情形：



由上圖不難看出，當 k 為 2，將 n 邊形三角化的方法數就等於將由 v_2，v_3, ..., v_n 等 $(n-1)$ 個點形成的凸 $(n-1)$ 邊形三角化的方法數，因此有 a_{n-1} 種；如果我們定義 $a_2=1$，那麼 a_{n-1} 可以寫為 $a_2\cdot a_{n-1}$。同樣地，當 $k=(n-1)$ 時，將 n 邊形三角化的方法數就等於將由 v_1，v_2, ..., v_{n-1} 等 $(n-1)$ 個點形成的凸 $(n-1)$ 邊形三角化的方法數，同樣也有 a_{n-1}（$=a_{n-1}\cdot a_2$）種；因此將凸 n 邊形三角化的方法總共有

$$\begin{aligned} a_n &= \sum_{k=3}^{n-2} a_k a_{n-k+1} + a_2 a_{n-1} + a_{n-1} a_2 \\ &= \sum_{k=2}^{n-1} a_k a_{n-k+1} \end{aligned}$$

種；再配合初始條件 $a_2=1$，我們有了數列 a_2，a_3，a_4, ...完整的遞迴定義：

$$a_n = \begin{cases} 1 & n=2 \\ \sum_{k=2}^{n-1} a_k a_{n-k+1} & n>2 \end{cases}$$

讀者不難發現這個數列與 Catalan numbers 有密切的關係；事實上，由

$$a_{n+2} = \sum_{k=2}^{n+1} a_k a_{n-k+3} = \sum_{k=1}^{n} a_{k+1} a_{n-k+2}$$

我們很容易可以用數學歸納法証明對所有非負整數 n，$a_{n+2}=C_n$，因為這兩個數列的第

一項相等（$a_2 = C_0 = 1$），而且如果對所有小於 n 的非負整數 k 而言，$a_{k+2} = C_k$ 都成立的話，那麼

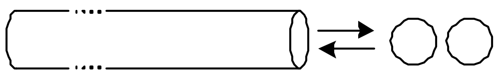
$$a_{n+2} = \sum_{k=1}^{n} a_{k+1} a_{n-k+2} = \sum_{k=1}^{n} C_{k-1} C_{n-k} = C_n$$

因此，對任意一個大於 2 的正整數 n 而言，將一個凸 n 邊形三角化的方法總共有

$$a_n = C_{n-2} = \frac{1}{n-1}\binom{2n-4}{n-2}$$ 種。

堆疊問題

下圖所示為一長條型的裝排球用的袋子（假設袋子很長，不管裝幾粒球都不會滿出來）：



已知在某段時間內，有 n 粒球曾經「進出」此袋，也就是說，每一粒球都曾經在某個時間被放入袋中，並於之後的某個時間被取出。以 $n = 2$ 為例，放入及取出的順序總共有「放入、放入、取出、取出」及「放入、取出、放入、取出」等兩種可能。

請問：當 n 為任意正整數時，這 n 個放入的動作及 n 個取出的動作之間總共有幾種可能的順序？

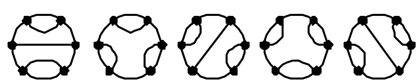
這個問題很明顯又和我們前面探討的小明的上班問題類似，因為這 n 粒球的任何一種可能的進出順序都包含了 n 個放入及 n 個取出的動作（小明須向東及向北各走 n 格），而且過程中已取出的球數隨時都不會超過已放入的球數（小明已走過的向北的格子數隨時都不能超過向東的格子數）。因此，對任意正整數 n 而言，n 粒球進出袋子總共有 C_n 種可能的順序。

讀者如果對「資料結構」（Data Structures）稍有涉獵，不難看出此題中的排球袋其實就是「堆疊」（Stacks）的概念。

三個練習題

以下是另外三個與 Catalan numbers 相關的問題，提供給讀者參考。

1. 有一隻跳蚤在數線上來回跳躍，它每一步的長度都正好是 1。在某段時間內，這隻跳蚤由原點出發後總共跳了 $2n$ 步，其中有 n 步是向右，n 步是向左。
   (1) 試證它最後的位置一定是在原點。
   (2) 有多少種跳法會使得它在跳躍過程中所有的落腳點都是非負整數？
2. 小明帶著 $2n$ 支冰棒到街上販賣，每支賣五元；在一段時間內，有 $2n$ 人陸續向他買走了所有的冰棒。已知小明出門時身上並沒帶錢，而販賣過程中有 n 人是直接拿五元向他買，另外 n 人拿的則是十元（因此須找五元）；所幸販賣過程中始終未曾發生該找五元卻找不出五元的情形。請問可能的販賣過程有多少種？
3. 有 $2n$ 人圍著一張圓桌而坐。請問：有幾種方式可以讓每個人同時伸出一支手與另一人相握，而且不會發生手臂互相重疊的情形？以 $n = 3$ 為例，總共有 5 種可能的握法，如下圖所示：



（下轉第 27 頁）