组合数学第六讲

授课时间: 2018年10月22日 授课教师: 孙晓明 记录人: 王珎 付娆

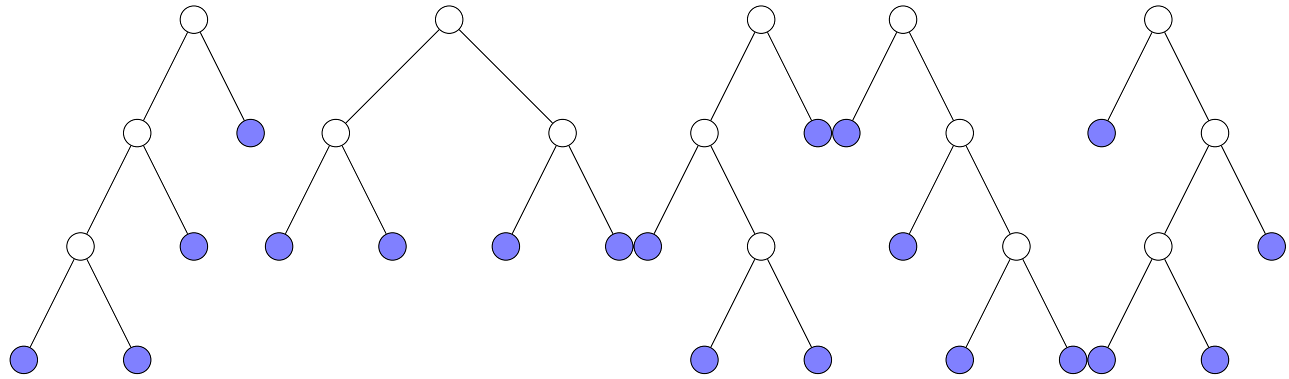
1 卡特兰数（Catalan number）的一些例子

1.1 n个矩阵相乘问题

作为例子，考虑三个矩阵$A_{m\times1}$，$A_{1\times m}$，$A_{m\times1}$相乘的问题$A_{m\times1}\times A_{1\times m}\times A_{m\times1}$。若按顺序相乘，前两个矩阵相乘需要$m^2$次乘法，得到一个$m\times m$的矩阵；它再与第三个矩阵相乘，仍需要$m^2$次乘法。为了得到最终结果，共进行了$2m^2$次乘法运算。若先让后两个矩阵相乘（需要进行m次乘法运算），再与第一个矩阵相乘（需要进行m次乘法运算），则总共进行了$2m$次乘法运算。由此可见，不同的相乘顺序（方案）的运算时间不同。这里我们不考虑如何寻找最优的方案，转而考虑矩阵相乘的方案数。

例 1 对n个矩阵$A_1,A_2,\cdots,A_n$，共有多少种不同的方案计算$A_1\times A_2\times\cdots\times A_n$?

解 记n个矩阵相乘的方案数为T_n。易知$T_2=1$，$T_3=2$，$T_4=5$。观察到，n个矩阵相乘时，每种方案恰好对应于一棵二叉树。以$n=4$为例，五种相乘方案对应于如下五棵二叉树：



(a) (b) (c) (d) (e)

例如，图（a）表示的相乘方案为$((A_1\times A_2)\times A_3)\times A_4$，图（b）表示的相乘方案为$(A_1\times A_2)\times(A_3\times A_4)$，以此类推。上图的二叉树，每一棵都有$2n-1$个结点，其中$n$个叶子结点，$n-1$个中间结点，且中间结点都有两个孩子。每个结点都代表一个矩阵。其中n个叶子结点分别代表$A_1,A_2\cdots,A_n$。中间结点代表的矩阵由它的左孩子代表的矩阵乘以右孩子代表的矩阵得到。因此根节点代表的矩阵就是$A_1\times A_2\times\cdots\times A_n$。现在，求矩阵相乘的方案数的问题转化为了求上述二叉树的个数的问题。

现在对二叉树进行编码。把向左的边标记为1，向右的边标记为0。给定一棵二叉树，对其使用先序遍历[1]。每当遇到一条未曾经过的边时，记录下它的标记。则遍历结束后，得到一个长度为$2n-2$的01串。以$n=4$为例，图（a）可以表示为111000，图（b）为110010，图（c）为110100，图（d）为101010，图（e）为101100。继续观察可发现，若有n片叶子，则01串中有$n-1$个1和$n-1$个0，且01串前i（$1\le i\le 2n-2$）个数中，1的个数不少于0的个数。由于二叉树和01串编码是一一对应的，因此求解矩阵相乘的方案数等价于求如下问题：

[1]先序遍历的方式如下：先访问根，然后递归地访问左子树，最后递归地访问右子树。

例 2 在由n个1和n个0组成的长为$2n$的所有序列$a_1a_2\cdots a_{2n}$中，满足其任意前缀中1的个数不少于0的个数的有多少？

这与上次课的汉诺塔问题变种类似。盘子必须先进栈（对应于1），而后再出栈（对应于0），且在任意时刻，进栈的次数不少于出栈的次数。因此例2的答案为第n项卡特兰数C_n，而例1的答案为$T_n = C_{n-1} = \binom{2(n-1)}{n-1}/n$。 □

与例2类似的还有如下两个问题，它们的答案与例2一样。

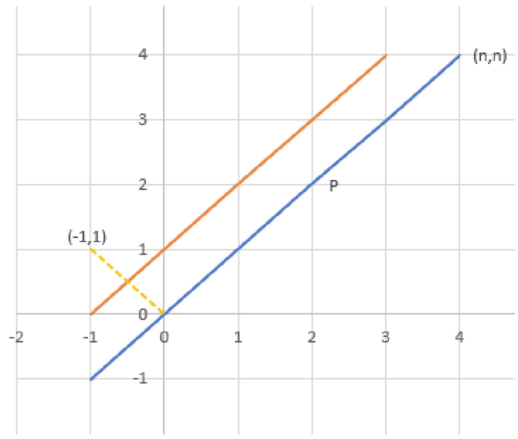
例 3 进公园问题。有$2n$个人，n个人有10元面额的现金，n个人有5元面额的现金，售票员处没有零钱，公园的门票是5元。这$2n$个人排队进入公园，一旦售票处找不开零钱，则售票系统瘫痪。问有多少种不同的排队顺序，使得这$2n$个人都能够顺利进入公园？

例 4 n对括号共能组成多少种合法的括号序列？

1.2 坐标图上最短路径问题

例 5 一个质点需要从$(0,0)$处沿整点走到(n,n)处，每次只能移动单位距离。那么在所有最短路径中，质点始终位于直线$y=x$下方（包括$y=x$）的有多少条?

解 若没有任何限制，最短路线需要走$2n$步，其中n步向右，n步向上，因此有$\binom{2n}{n}$条最短路径。现在要求质点始终位于$y=x$下方。反过来，我们计算穿过直线$y=x$的最短路径的数目。考虑某条穿过直线$y=x$的最短路径S。设它在$y=x$上的点$P=(x_P,x_P)$处第一次穿过$y=x$。由于质点在整点上移动，这条路径此时到达了$y=x+1$上的点$P'=(x_P,x_P+1)$。记路径S中从$(0,0)$到P'的部分为$S(P')$，从P'到(n,n)的部分为$\bar{S}(P')$。令$T(P')$表示由$S(P')$关于直线$y=x+1$做对称得到的路径。则$T(P')$为由$(-1,1)$到P'的路径。那么$T=T(P')+\bar{S}(P')$为由$(-1,1)$到(n,n)的路径。因此，一条穿过$y=x$且由$(0,0)$到(n,n)的最短路径S，对应于一条穿过$y=x+1$且由$(-1,1)$到(n,n)的路径T。对于后者，由于由$(-1,1)$到(n,n)的路径一定会穿过$y=x+1$，因此“穿过$y=x+1$”这一限制是多余的。那么，路径T（或等价地，路径S）的个数为$\binom{2n}{n+1}$。因此，从$(0,0)$到(n,n)且始终在$y=x$下方的最短路径的个数为$\binom{2n}{n}-\binom{2n}{n+1}=\binom{2n}{n}-\frac{n}{n+1}\binom{2n}{n}=\frac{1}{n+1}\binom{2n}{n}$。



□

例 6 考虑进公园问题的变种。若初始时，售票处有5元零钱。这时合法的排队顺序有多少种？

解 这等价于，在由n个1和n个0组成的长为$2n$的所有序列$a_1a_2\cdots a_{2n}$中，满足对任意的k（$1\le k\le 2n$），$|\{i:i\le k,a_i=1\}|+1\ge|\{i:i\le k:a_i=0\}|$的有多少？在例5的背景下，这等价于计算从$(0,0)$到$(n,n)$且始终位于直线$y=x+1$下方（包括$y=x+1$）的最短路径的数目。利用与例5相同的论证，可以计算出答案为$\binom{2n}{n}-\binom{2n}{n+2}$。 □

1.3 1-3-2-avoid排列问题

例 7 一个排列是1-3-2-avoid，当且仅当对任意满足$a<b<c$的三个数a,b,c，它们在排列中的相对顺序不能是a在c的前面且c在b的前面。在$1,2,\cdots,n$的所有排列中，共有多少个满足1-3-2-avoid？

解 记满足1-3-2-avoid的排列数为C_n。易知，$C_1=1$，$C_2=2$，$C_3=5$，$C_4=14$。一般地，若数字n放置在排列的第k个位置上，为了满足题目条件，排列的前$k-1$个位置上的任意一个数字都必须大于排列的后$n-k$个位置上的任意一个数字。由此可知，n左侧的$k-1$个位置上放置的是数字$n-k+1,n-k+2,\cdots,n-1$；n右侧的$n-k$个位置上放置的数字是$1,\cdots,n-k$。这两组数各自需要满足1-3-2-avoid性质，根据乘法原理，共有$C_{k-1}C_{n-k}$种可能。再根据加法原理，有

$$C_n=\sum_{k=1}^{n}C_{k-1}C_{n-k}。$$

由初始条件和递归关系可知，$C_n=\frac{1}{n+1}\binom{2n}{n}$，为卡特兰数。 □

1.4 凸多边形剖分

例 8 使用凸n边形的对角线将其剖分成若干三角形，且这些对角线在凸n边形的内部不相交，求可能的剖分方案数$T(n)$。

解 易知$T(3)=1,T(4)=2,T(5)=5$，为后面讨论需要，令$T(2)=1$。对一般的n，依然可以采取通过分类计数获得递推式并求解得到通项的策略，这里给出两种分类方式。

[*法一*] 第一种方法从顶点的角度考虑。从任意定点开始顺时针将多边形的顶点标号为$v_1,v_2,\ldots,v_n$，以顶点v_1是否引出对角线作为分类标准：若v_1未引出对角线，那么v_2和v_n之间必然有一条对角线，问题退化为凸$n-1$边形的情况，共有$T(n-1)$种方案；若v_1引出了至少一条对角线，那么可以按照如下标准对这一情况分类：记与v_1之间有对角线的顶点中标号最小的点v_i，利用“标号最小”这一限制可知v_1与$v_2,\cdots,v_{i-1}$之间没有对角线，那么，v_2到v_i必然有一条对角线。因此，这一情况下问题退化为两个独立的子问题：由$v_2,\cdots,v_{i-1}$围成的凸$i-1$边形的剖分和由$v_i,v_{i+1},\ldots,v_n,v_1$ 围成的凸$n-i+2$边形的剖分。它们剖分的方案数分别为$T(i-1)$和$T(n-i+2)$，根据加法原理和乘法原理，有

$$T(n)=T(n-1)+\sum_{i=3}^{n-1}T(n-i+2)T(i-1)=\sum_{i=3}^{n}T(n-i+2)T(i-1)。$$

观察到，

$$T(n+2)=\sum_{i=3}^{n+2}T(n-i+4)T(i-1)=\sum_{i=1}^{n}T(n-i+2)T(i+1),$$

若令$C_n=T(n+2)$，则有$C_0=1,C_1=1,C_2=2,C_3=5$，且

$$C_n=\sum_{i=1}^{n}C_{n-i}C_{i-1}。$$

因此C_n恰为卡特兰数。则$T(n)=C_{n-2}$。

[法二] 第二种方法从边的角度考虑。仍然按照法一的方式为顶点编号。考虑边v_1v_2所属的三角形的另一个顶点v_k所在，三角形$v_1v_2v_k$将原凸n边形分为凸$k-1$边形和凸$n-k+2$边形。根据加法原理和乘法原理，我们再次得到了递推式：

$$T(n)=\sum_{i=3}^{n}T(i-1)T(n-i+2)。$$

□

2 随机游走

2.1 一维随机游走

让我们先运用一个实例来说明一维随机游走的应用背景。我们都知道，概率是源于赌场的，假想你现在来到了拉斯维加斯，想要去体验一下当地的赌场，为了避免自己沉迷其中无法自拔，只带一元钱进入娱乐一下，而且你能保持理智不会去借钱，这就意味着一旦你的钱数变为零便无法继续玩下去，你决定参与猜单双，假定你来到了一个十分正经的赌场，你赢和输的概率各占一半，那么我们想知道的是，你是否总会输光？你是否可能保持手中有钱地状态，在赌场中无限地进行这个游戏？我们把以上的问题抽象如下：

例 9 考虑一维网格上的随机游走。初始时刻质点位于$x_0=1$处。每次移动时，质点以p的概率向右移动一步，以$q=1-p$的概率向左移动一步，且各次移动是相互独立的。若质点到达原点$x=0$处，则游走结束。在此随机游走的过程中，质点到达原点的概率是多少？质点游走的期望步数是多少？

解 设质点到达原点的概率为P，质点在第n步首次到达原点的概率为P_n。观察到质点只有在奇数步时才能到达偶数的坐标点，因此$P_{2k}=0$，$P=\sum_{k\geq 0}P_{2k+1}$。根据P_{2k+1}的定义，在第$2k+1$步，质点一定是由$x=1$处移动到了$x=0$。则前$2k$次移动中，恰有k次向右，k次向左，并且前i次移动中（$1\leq i\leq 2k$），向右的次数不少于向左的次数。由此可知，第$2k+1$步首次到达原点的方案数为卡特兰数$C_k=\frac{1}{k+1}\binom{2k}{k}$，那么$P_{2k+1}=\frac{1}{k+1}\binom{2k}{k}p^k(1-p)^{k+1}$。由此可知，

$$\begin{aligned}
P &= \sum_{k\geq 0}\frac{1}{k+1}\binom{2k}{k}p^k(1-p)^{k+1}\\
&= (1-p)C(p(1-p))\\
&= (1-p)\cdot\frac{1-\sqrt{1-4p(1-p)}}{2p(1-p)}\\
&= \frac{1-|1-2p|}{2p},
\end{aligned}$$

其中，$C(x)=\frac{1-\sqrt{1-4x}}{2x}$为卡特兰数的生成函数。因此，当$p\leq 1/2$时，$P=1$；当$p>1/2$时，$P=(1-p)/p$。

设质点在第T步回到原点，我们想知道质点回到原点时移动步数的期望值$E[T]$。根据之前的计算，

$$E[T]=\sum_{k\geq 0}\frac{2k+1}{k+1}\binom{2k}{k}p^k(1-p)^{k+1}=(1-p)\cdot F(x)|_{x=p(1-p)}。$$

其中$F(x)$为序列$\{\frac{2k+1}{k+1}\binom{2k}{k}\}_{k\geq 0}$的生成函数。注意到$F(x) = 2xC'(x) + C(x)$，$C(x)$为卡特兰数的生成函数，$C'(x)$为它的导数。因此有

$$\begin{aligned} E[T] &= (1-p)\cdot(2xC'(x)+C(x))|_{x=p(1-p)} \\ &= (1-p)\cdot(2p(1-p)C'(p(1-p))+C(p(1-p))) \\ &= \frac{2-2p}{|1-2p|} - \frac{1-|1-2p|}{2p} \\ &= \frac{1-|1-2p|}{2p|1-2p|}。\end{aligned}$$

当$p < 1/2$时，$E[T] = 1/(1-2p)$；当$p = 1/2$时，$E[T] = \infty$，可见虽然此时质点以概率1回到原点，但质点游走的期望时间为无穷大；当$p > 1/2$时，$E[T] = \frac{1-p}{p(2p-1)} < \infty$。这似乎与前一个结论矛盾。$p = 1/2$时质点游走的期望时间已经是无穷大了，那么$p > 1/2$时也应该是无穷大才对。这是因为这里我们计算的$E[T]$为在质点回到原点的条件下的期望，而$p > 1/2$时，质点有一定的概率永远不回到原点，而这一部分概率，我们没有计算在$E[T]$当中。 □

3 选做题

（1）

一个排列是1-2-3-avoid，当且仅当对任意满足$a < b < c$的三个数a, b, c，它们在排列中的相对顺序不能是a在b的前面且b在c的前面。考虑如下两个问题：

a. 在$1, 2, \cdots, n$的所有排列中，共有多少个满足1-2-3-avoid？

b. $1, 2, 3, 4$的排列共能定义24种x-x-x-x-avoid。请对这24种avoid类型讨论同样的问题。

（2）

考虑二维网格上的随机游走$\{P_t\}_{t\geq 0}$。如图所示，初始时$P_0 = (0, 0)$，每次在上下左右四个方向上等概率选择一个方向移动长度1，且各次移动之间相互独立。则P_t满足：

$$\begin{gathered} \Pr(P_t = P_{t-1} + (0,1)) = \Pr(P_t = P_{t-1} + (0,-1)) = \\ \Pr(P_t = P_{t-1} + (-1,0)) = \Pr(P_t = P_{t-1} + (1,0)) = 1/4 \end{gathered}$$

那么，游走的质点回到原点的概率是多少？如果在三维的网格上呢？

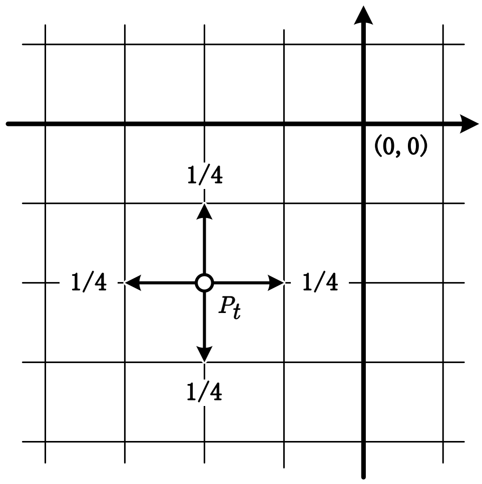


图 1: 二维随机游走