Catalan数

重庆南开信竞基础课程

什么是Catalan数

卡特兰数是组合数学中一个常在各种计数问题中出现的数列。

由以比利时的数学家欧仁·查理·卡特兰 (Eugene Charles Catalan 1814–1894)命名。

其前几项为 : **1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862**, 16796, 58786, 208012, 742900, 2674440, 9694845, 35357670, 129644790, 477638700, 1767263190,......

Catalan数的定义

令h(0)=1, Catalan数的第n项为h(n), 满足递归式:

$$h(n) = h(0) * h(n-1) + h(1) * h(n-2) + \ldots + h(n-1) * h(0)$$

即 $h(n) = \sum_{i=0}^{n-1} h(i) * h(n-1-i)$

还可以表示成: $h(n) = \frac{C_{2n}^{n}}{(n+1)}$

也可以表示成: $h(n) = C_{2n}^{n} - C_{2n}^{n-1}$

也可以表示为 : h(n) = h(n - 1) * (4 * n - 2) / (n + 1)

神奇之处：很多看上去毫不相干的组合计数问题的最终表达式都是卡特兰数的形式。

问题1：括号匹配

n对括号有多少种有效配对方式？

比如n=2是，合法的配对方式有下列两种：

“()()” ，“(())”

下列方式是不合法的：

“)()(”

“))((”

“))()”

“()))”

……

问题1：括号匹配

思路1：

n对括号相当于有2n个符号，n个左括号、n个右括号，可以设问题的解为f(2n)。
即f(2n)表示n个左括号和n个右括号配对的合法方案数

第1个符号肯定为左括号，假设与之匹配的右括号为第i个符号，那么**i显然是偶数。**

f(2n) = f(0)*f(2n-2) + f(2)*f(2n - 4) + ... + f(2n - 4)*f(2) + f(2n-2)*f(0)

f(0) * f(2n-2)表示第1个括号与第2个括号配对，它们间包含的其他括号个数为0，剩余为2n-2个括号。

f(2)*f(2n-4)表示第1个括号与第4个括号配对，它们间包含了2个括号，相当于（ （ ） ），剩余为2n-4个数字括号。依次类推。

假设f(0) = 1，计算一下开始几项，f(2) = 1, f(4) = 2, f(6) = 5， f(8) =14 ......

结合递归式，不难发现f(2n) 等于h(n)。

问题1：括号匹配

思路2：

对于一个长度为2n括号序列：
设CntL[i]表示[1,i]区间中左括号的数量
设CntR[i]表示[1,i]区间中右括号的数量
观察可得，一个合法的括号序列**始终满足**CntL[i]>=CntR[i]　　1<=i<=2n

总的括号方案数为C_{2n}^{n} 即2n个位置中放n个左括号，剩余的全放右括号

其中包括了不合法的方案，不合法方案数是多少呢?

我们采取**反转映射**的方式，从左往右找到第一个不合法的地方k,即CntL[k]<CntR[k]，把[1,k]中的左括号变右括号，右括号变左括号，得到的括号序列中一定有n+1个左括号，n-1个右括号。即2n个位置中有n+1个左括号n-1个右括号的方案数，不合法的方案数为C_{2n}^{n-1}

也就是：对于任意含n+1个左括号和n-1个右括号的序列，我们都能用上述操作的逆操作，还原一个不合法的由n个左括号和n个右括号构成的方案。

所以总的方案数为$C_{2n}^{n} - C_{2n}^{n-1}$　恰好就是卡特兰数的另一种表达式

问题2：链乘

P=A1×A2×A3×......×An，依据乘法结合律，不改变其顺序，只用括号表示**成对**的乘积，试问有几种括号化的方案？

例如，A1×A2×A3×A4 有下列5种方案：

(A1×(A2×(A3×A4)))

(A1×((A2×A3)×A4))

((A1×A2)×(A3×A4))

(((A1×A2)×A3)×A4)

((A1×(A2×A3))×A4)

问题2：链乘

思路：可以这样考虑，首先通过括号化，将P分成两个部分，然后分别对两个部分进行括号化。比如分成(A1)×(A2×a3×......×An)，然后再对(A1)和(A2×A3×......×An)分别括号化；又如分成(A1×A2)×(A3×......×An)，然后再对(A1×A2)和(A3×......×An)括号化。

设n个数字链乘的括号化方案的种数为f(n)，那么问题的解为：

f(n) = f(1)*f(n-1) + f(2)*f(n-2) + f(3)*f(n-3) + f(n-1)*f(1)

f(1)*f(n-1)表示分成(A1)×(A2×A3......×An)两部分，然后分别括号化

f(2)*f(n-2)表示分成(A1×A2)×(A3×......×An)两部分，然后分别括号化

计算开始几项，f(1) = 1, f(2) = 1, f(3) = 2, f(4) = 5，f(5)=14 ……
结合递归式，不难发现f(n)等于h(n)。

问题3：进出栈

有n个整数1,2,...,n和一个栈，进栈顺序为1,2,...,n,每个数字都要出入一次栈，用s表示数字入栈，x表示数字出栈，那么那么合法的序列有多少个。

比如：n=2,合法的出入栈顺序有：

S1 S2 X2 X1

S1 X1 S2 X2

这里我们只关心出入栈操作构成的序列，不关心具体的数字，所以，合法的序列有ssxx和sxsx两种。

问题3：进出栈

用S表示数字入栈，X表示数字出栈

比如：n=2,合法的出入栈顺序有：

S1 S2 X2 X1

S1 X1 S2 X2

这里我们只关心出入栈操作构成的序列，不关心具体的数字，所以，合法的序列有SSXX和SXSX两种。

我们把进栈看成左括号，出栈看成右括号，显然，这就是一个问题1所述的**括号配对**问题。答案就是h(n)

例题4：排队

12个**高矮不同**的人，排成两排，每排必须是**从矮到高**排列，而且**第二排比对应的第一排的人高**，问排列方式有多少种？

以6个人排队为例，下面是一种合法的排队：
第一排：1 5 7
第二排：2 6 8

例题4：排队

我们先把这12个人从低到高排列,然后,选择6个人排在第一排,那么剩下的6个肯定是在第二排.

用0表示对应的人在第一排,用1表示对应的人在第二排,那么含有6个0,6个1的序列,就对应一种方案.

比如000000111111就对应着

第一排:0 1 2 3 4 5

第二排:6 7 8 9 10 11

再比如010101010101就对应着

第一排:0 2 4 6 8 10

第二排:1 3 5 7 9 11

问题转换为，这样的满足条件的01序列有多少个。

观察1的出现，我们考虑一个人能不能放在第二排，显然，在这个1之前出现的那些0,1对应的人要么是在这个1左边，要么是在这个1前面。而肯定要有一个0的，在这个1前面，统计在这个1之前的0和1的个数。

也就是要求序列中，**数字1左侧0的个数要大于等于1的个数。**

例题4：排队

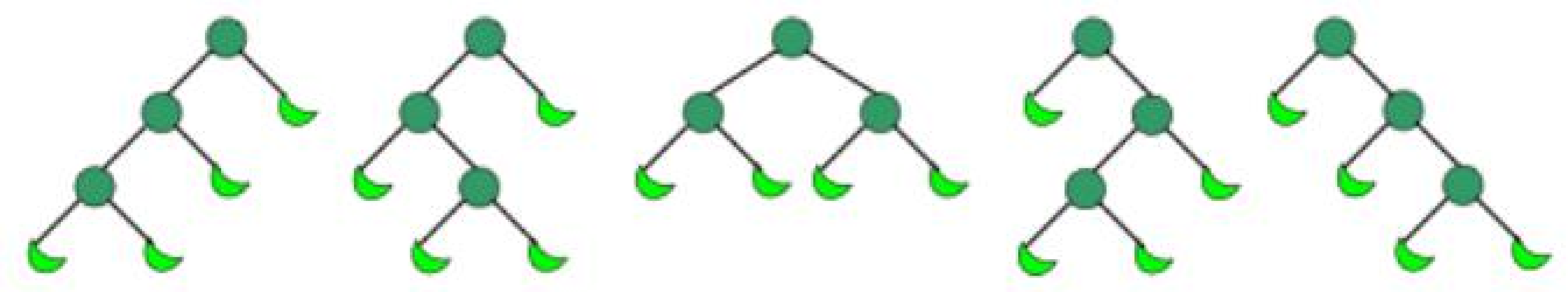
如果**把0看成入栈**操作，**1看成出栈**操作，就是说给定6个元素，**合法的入栈出栈序列有多少个。**这就是前面应用中分析的catalan数,这里只是用于栈。

通过暴力枚举的结果是132。
利用卡特兰数计算的结果：

c(12, 6)/7 = 12*11*10*9*8*7/(7*6*5*4*3*2) = 132

例题5：二叉树

n个节点构成的二叉树，共有多少种不同情形？
当n=7时，方案如下图：



例题5：二叉树

思路：

设T(i, j)表示根的左子树含i个结点，右子树含j个结点，的二叉树。总共n个节点，根肯定会占用一个结点，那么剩余的n-1个结点可以有如下的分配方式，T(0, n-1),T(1, n-2),...T(n-1, 0)

设f(n)表示有n个节点的二叉树的形态总数

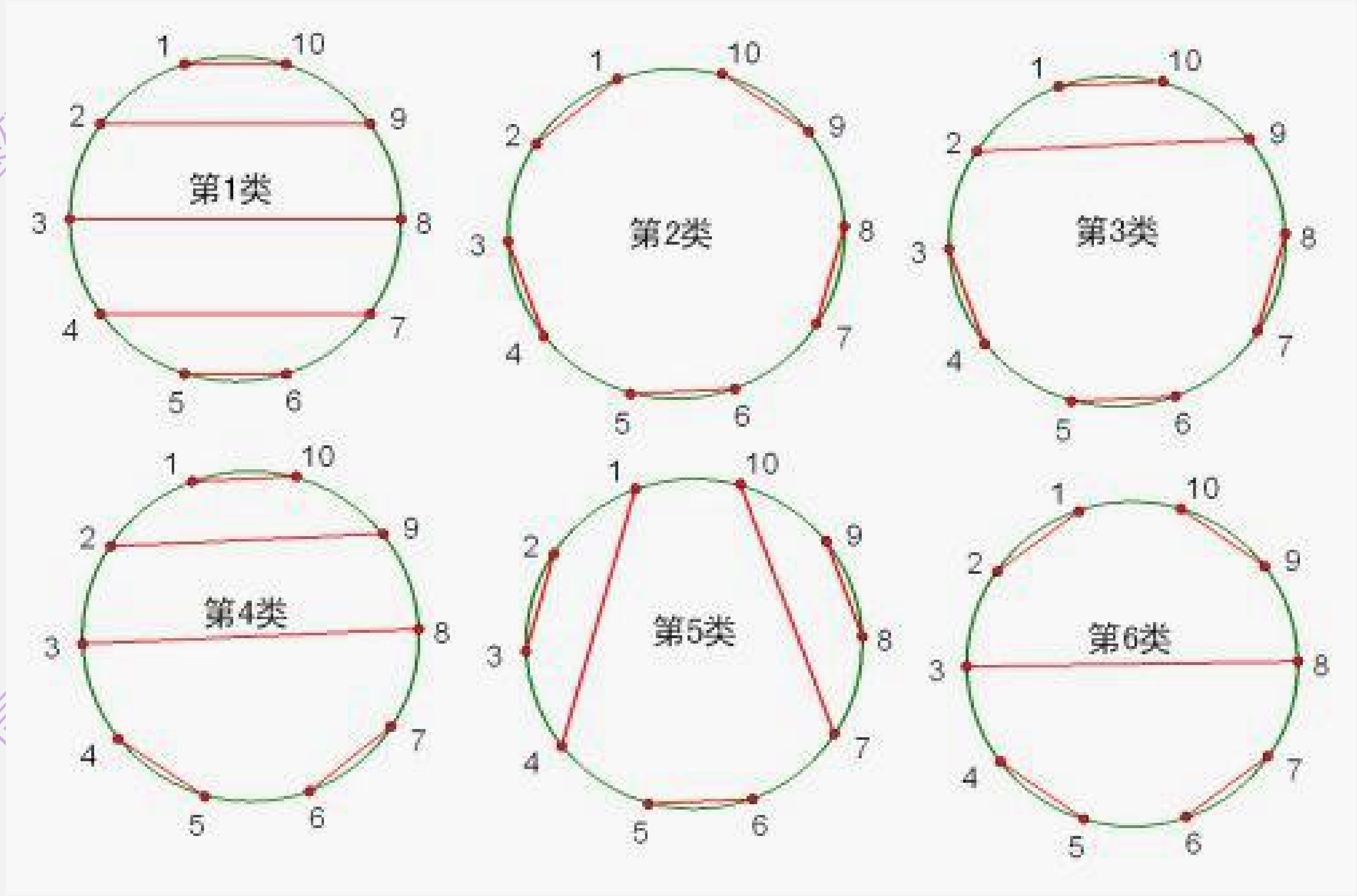
那么f(n) = f(0)*f(n-1) + f(1)*f(n-2) + ...+ f(n-2)*f(1) + f(n-1)*f(0)。

f(0) = 1，那么f(1) = 1, f(2) = 2, f(3) = 5，f(4)=14,……

结合递推式，不难发现f(n)等于h(n)。

例题6：圆上连点

在圆上选择2n个点，将这些点成对连接起来使得所得到的n条线段不相交的方法数？



例题6：圆上连点

思路：

以其中一个点为基点，编号为0，然后按顺时针方向将其他点依次编号。

那么与**编号为0相连点的编号一定是奇数**，否则，这两个编号间含有奇数个点，势必会有个点被孤立，即在一条线段的两侧分别有一个孤立点，从而导致两线段相交。

设选中的基点为A，与它连接的点为B，那么A和B将所有点分成两个部分，一部分位于A、B的左边，另一部分位于A、B的右边。然后分别对这两部分求解即可。

例题6：圆上连点

设问题的解f(2n)，那么:

`f(2n) = f(0)*f(2n-2) + f(2)*f(2n-4) + f(4)*f(2n-6) + ......f(2n-4)*f(2) + f(2n-2)*f(0)`

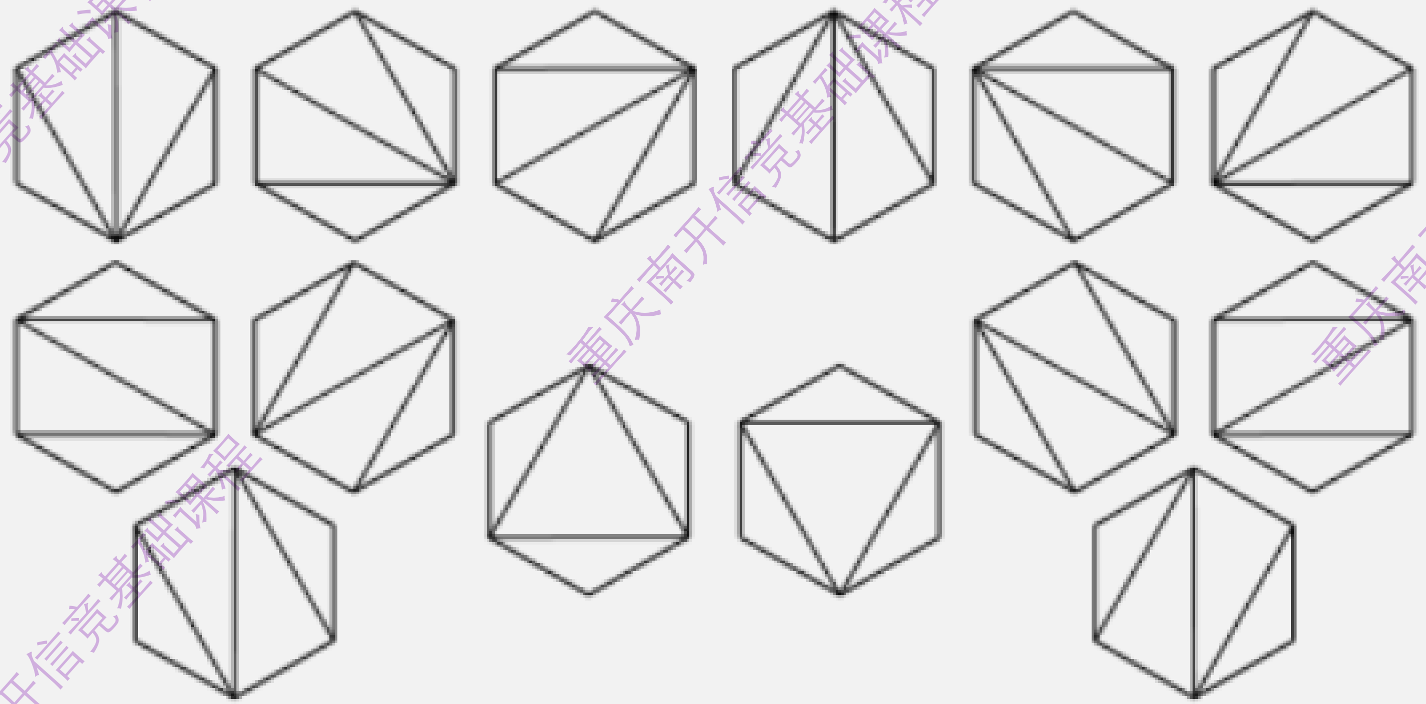
f(0)*f(2n-2)表示编号0的点与编号1的点相连，此时位于它们右边的点的个数为0，而位于它们左边的点为2n-2。

f(2)*f(2n-4)表示编号0的点与编号3的点相连，此时位于它们右边的点的个数为2，而位于它们左边的点为2n-4。

依次类推:f(0) = 1, f(2) = 1, f(4) = 2。结合递归式，不难发现f(2n) 等于h(n)。

例题7：凸多边形的划分

将一个凸多边形划分成三角形总共有多少种方法？
下图是6边形的划分方案：



例题7：凸多边形的划分

思路：以凸多边形的一边为基，设这条边的2个顶点为A和B。从剩余顶点中选1个，可以将凸多边形分成三个部分，中间是一个三角形，左右两边分别是两个凸多边形，然后求解左右两个凸多边形。

设问题的解f(n)，其中n表示顶点数，那么

$$f(n) = f(2)*f(n-1) + f(3)*f(n-2) + ...+f(n-2)*f(3) + f(n-1)*f(2)$$

f(2)*f(n-1)表示三个相邻的顶点构成一个三角形，那么另外两个部分的顶点数分别为2和n-1，如图1所示。

f(3)*f(n-2) 表示中间三角形的左侧有3个点，右侧有n-2个点，如图2所示。

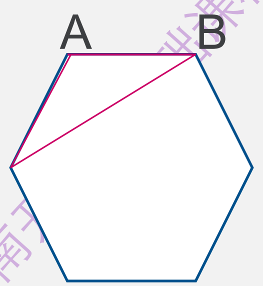


图1

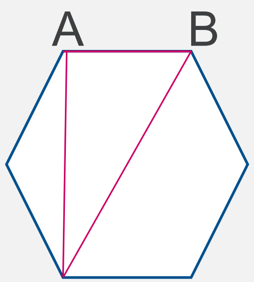


图2

设f(2) = 1，那么f(3) = 1, f(4) = 2, f(5) = 5。结合递推式，不难发现f(n) 等于h(n-2)。

例题8：找零

有2n个人排成一行进入剧场。入场费5元。其中只有n个人有一张5元钞票，另外n人只有10元钞票，问题是剧院售票处现在一分钱也没有（没有任何钞票），问有多少中方法使得只要有10元的人买票，售票处就有5元的钞票找零？

例题8：找零

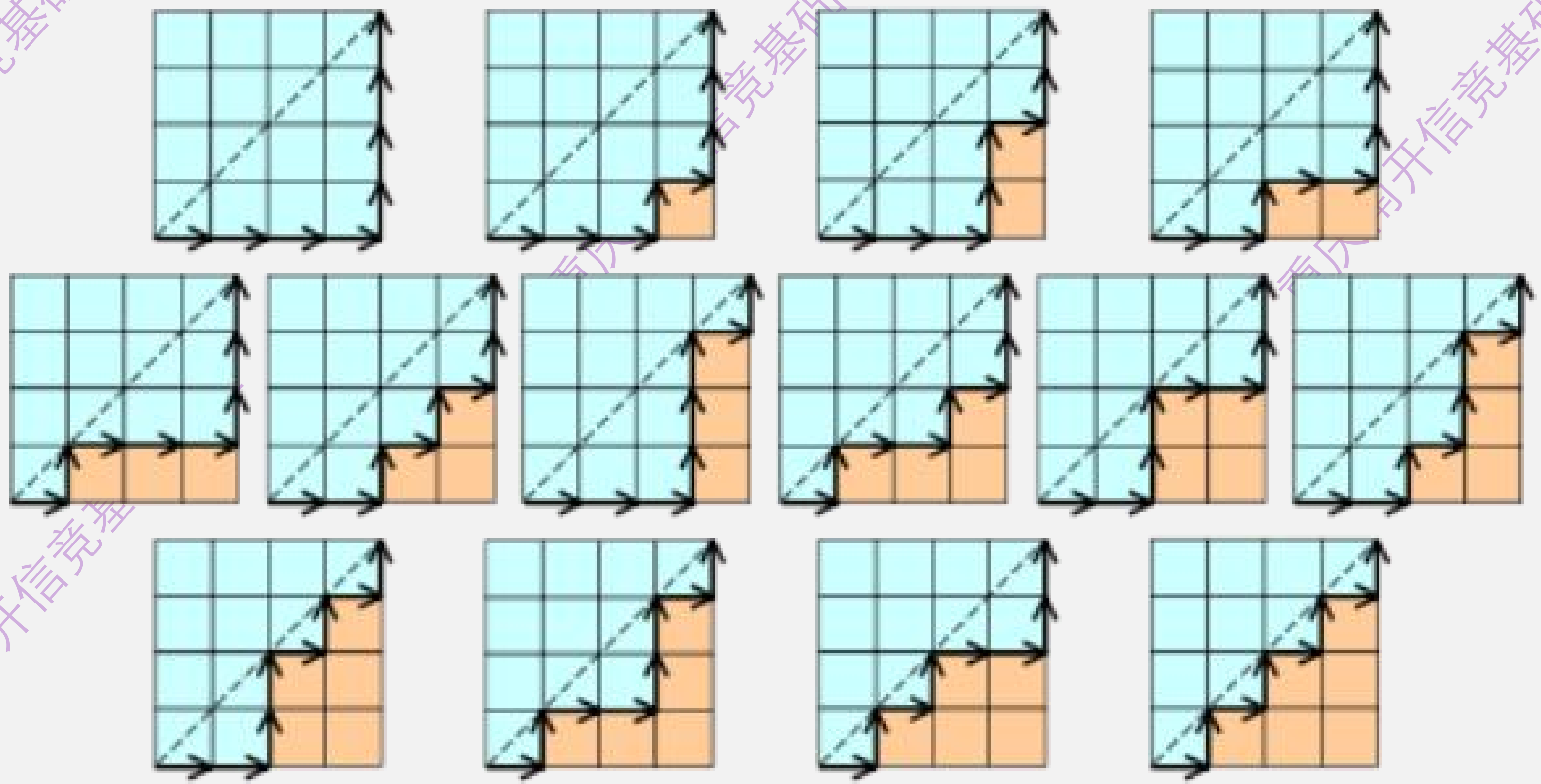
思路：

可以将持**5元买票**视为**进栈**，那么持**10元买票**视为5元的**出栈。**

这个问题就转化成了栈的出栈次序数。由“问题3”的分析直接得到结果，f(2n) 等于h(n)。

例题9：上班

一位大城市的律师在他住所以北n个街区和以东n个街区处工作，每天她走2n个街区去上班。如果他从不穿越（但可以碰到）从家到办公室的对角线，那么有多少条可能的道路？

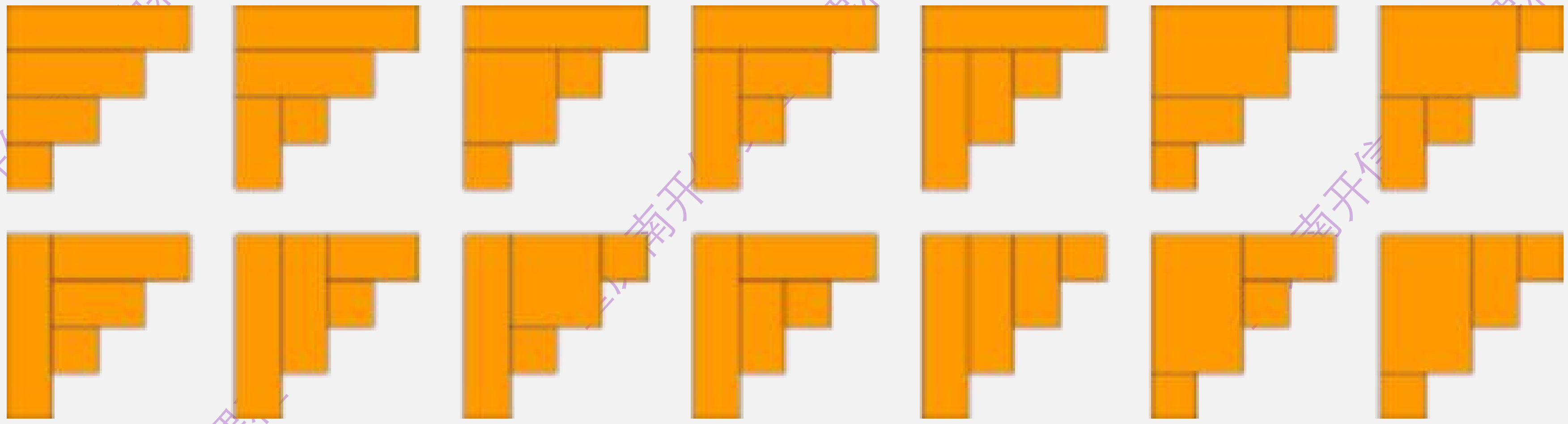


例题9：上班

我们将一条**水平边记为进栈，垂直边记为出栈**，我们所要保证的就是**前k步中水平边的个数不小于垂直边的个数**，换句话说出栈的时候栈内一直有元素，所以从根本上说又回归到Catalan数了。

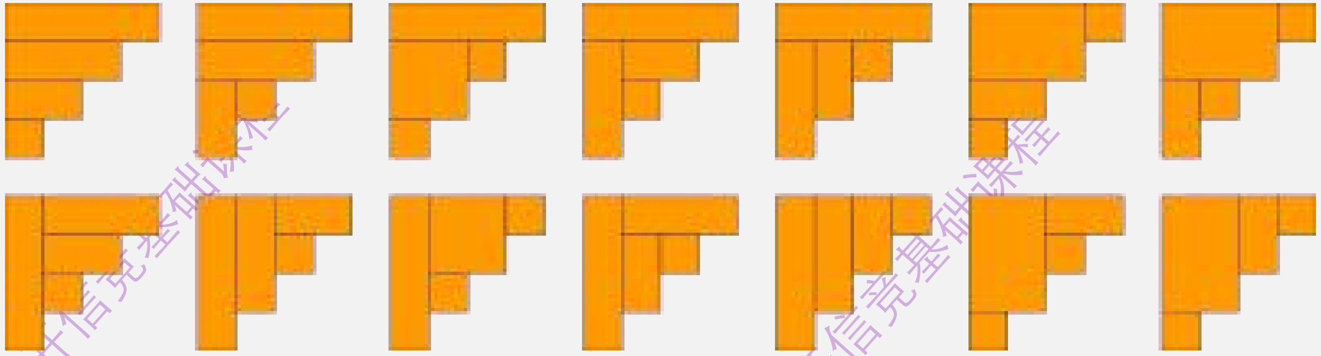
思考10：填充

用n个长方形填充一个高度为n的阶梯状图形的方法个数？



思考10：填充

答案是h(n)



我们注意到每个切割出来的矩形都必需包括一块标示为*的小正方形，那么我们此时枚举每个*与#标示的两角作为矩形对角，剩下的两个小阶梯就是我们的两个更小的子问题了，于是我们的

H(5) = H(0)*H(4) + H(1)*H(3) + H(2)*H(2) + H(3)*H(1) + H(4)*H(0)

满足卡特兰式子的形态

因此这就是我们所求的结果了。

奋斗吧少年

巨大的成功需要付出巨大的代价

no sacrifice, no success