概率期望计数

吴清月

基础概念

- 概率：
  - 一件事情发生的可能性
- 期望：
  - 平均值，$E = \sum P_i w_i$。
- 两者的联系：
  - 概率为p的事件期望$\frac{1}{p}$次后发生。
- 独立事件：
  - 对于两个独立事件A, B，$P(AB) = P(A)P(B)$

基础概念

- **期望的线性性：**
  - $E(x+y)=E(x)+E(y)$，无论x和y是否独立。
- $E(xy)=E(x)E(y)$，只在x和y独立的时候成立。

方差

- 随机变量的方差定义为$V(x) = E\left(\left(x - E(x)\right)^2\right)$
- 推一下式子：
- $V(x) = E\left(\left(x - E(x)\right)^2\right) = E(x^2 - 2xE(x) + E(x)^2) = E(x^2) - E(x)^2$
- 也就是方差的期望等于平方的期望减去期望的平方。

小练习

- 每次随机一个$[1,n]$中的数，求期望几次能随机出所有的数。

- 首先根据期望的线性性，我们只需要考虑要想把第i个数随机出来，期望需要随机几次。
- 对于第$i+1$个数，它被随机到的概率是$\frac{n-i}{n}$，期望就是$\frac{n}{n-i}$。
- 所以$E=\frac{n}{n}+\frac{n}{n-1}+\frac{n}{n-2}+\cdots+\frac{n}{1}=O(n\ln n)$

小练习

- 随机一个排列，求p_i是$p_1, p_2, \cdots, p_i$中最大的数的概率。

- 首先所有的数是等价的，所以p_1到p_i中每一个数成为最大值的概率是相等的。
- 所以概率就是$\frac{1}{i}$。

小练习

- 给你一张从S到T的DAG，求从S随机游走到T的期望步数。
- 这里的随机游走表示每次随机选择一条出边走过去。

- 设f_i表示从i走到T的期望步数，那么拓扑排序后可以DP，对于每一个点枚举它的出边然后转移：
- $f_u = \frac{1}{d_u} \sum (f_v + 1)$

随机游走——链

- 有一条长度为n的链，求从一端走到另一端的期望步数。
- 每次随机向左或者向右走，如果当前在一端那么就只能向另一端走。

- 设f_i表示从i走到$i+1$的期望步数，则答案就是$f_0+f_1+f_2\cdots+f_{n-1}$。
- 首先$f_0=1$
- 我们考虑第i个节点。如果它第一步往右走，那么步数就是1，概率是$\frac{1}{2}$，而如果它往左走，那么步数就是$f_{i-1}+f_i+1$，概率还是$\frac{1}{2}$，所以$f_i=\frac{1}{2}\cdot 1+\frac{1}{2}\cdot(f_{i-1}+f_i+1)$，解得$f_i=f_{i-1}+2$。
- 所以答案就是$1+3+5+\cdots+2n-1=n^2$

随机游走——团

- n个点的完全图，求从S随机游走到T的期望步数。

- 首先所有的点是等价的。
- 设E表示从一个点走向另一个点的期望步数，则
- $E = \frac{1}{n-1} \cdot 1 + \frac{n-2}{n-1} \cdot (E+1)$
- 即$E = n - 1$

构造

- 构造一个200个节点的无向图，使得从1随机游走到n的期望步数超过10^6。

- 100个点的完全图伸出来一条长度为100的链。
- 设E_0表示从完全图里面的一个点走向链头的期望步数，则$E_0 = 99$
- 再设f_i表示在链上从i走向$i+1$的期望步数，则$f_0 = \frac{1}{100} + \frac{99}{100} \cdot (99 + 1 + f_0)$，解得$f_0 = 9901$
- 然后$f_i = f_{i-1} + 2$，所以最终答案就是$9901 + 9903 + 9905 + \cdots + 10099 = 1000000$

树上随机游走

- 给你一棵树，求从S随机游走到T的期望步数。

树上随机游走

- 根据期望的线性性，我们只需要求出从一个点随机游走到达它的父节点的期望步数f_x和从父节点随机游走到它的期望步数g_x。设x的度数是d_x。
- 对于f，我们枚举它下一步是往上走还是往下走，得到$f_x = \frac{1}{d_x} + \sum_v \frac{1}{d_x} \cdot (f_v + f_x + 1)$，解得$f_x = d_x + \sum_v f_v$
- 对于g，它有$\frac{1}{d}$的概率直接走到，$\frac{1}{d}$的概率往上走，其余的情况是走到了别的子树内。
- 所以$g_x = \frac{1}{d} + \frac{1}{d}\left(1 + g_{fa_x} + g_x\right) + \sum_{fa_y = fa_x} \frac{1}{d}\left(1 + f_y + g_x\right)$，解得$g_x = d_{fa_x} + g_{fa_x} + \sum_{fa_y = fa_x} f_y = d_{fa_x} + g_{fa_x} + f_{fa_x} - d_{fa_x} - f_x = g_{fa_x} + f_{fa_x} - f_x$。
- 然后就可以递推了，时间复杂度$O(n)$。

NOIP2016 换教室

- 一天一共有n节课，第i节课默认在第c_i间教室上可以申请第i节课到第d_i间教室上，申请有k_i的概率通过，从第x间教室到第y间教室有距离$dis_{x,y}$，你最多可以申请m门课，要求期望跑的距离尽量少。
- $n, m \leq 2000$

NOIP2016 换教室

- 根据期望的线性性，跑动距离的期望等于每一天下课后跑动距离的期望之和。
- 而每一天下课后跑动的距离只和这一天和下一天是否申请换教室有关。
- 所以我们先预处理一个数组e[i][0/1][0/1]，表示第i天申请/不申请，第i+1天申请/不申请的情况下，期望需要跑动的距离。
- 接下来就可以DP了。设f[i][j][0/1]表示前i天，申请了j天，第i天申请了/没有申请的最短期望距离。
- 转移如下：
  - f[i][j][1]=min(f[i-1][j-1][0]+e[i-1][0][1],f[i-1][j-1][1]+e[i-1][1][1])
  - f[i][j][0]=min(f[i-1][j][0]+e[i-1][0][0],f[i-1][j][1]+e[i-1][1][0])
- 时间复杂度$O(nm)$。

BZOJ4318 OSU!

- 有一个长度为n的序列，第i个位置有p_i的概率为1，$1-p_i$的概率为0，一个序列的分数是所有极长连续的1的长度的三次方和。求期望分数。
- $n \leq 100000$

BZOJ4318 OSU!

- 1次方怎么求?
- 就是$\sum p_i$啊。
- 2次方怎么求?
- 设f_i表示以i结尾的长度的期望，再设g_i表示以i结尾的长度的平方的期望。
- 那么$f_{i+1} = p_{i+1}(1 + f_i)$， $g_{i+1} = E\left((x+1)^2\right) = E(x^2 + 2x + 1) = p_{i+1}(g_i + 2f_i + 1)$
- 同理，三次方怎么求?
- 设h_i表示以i结尾的三次方的期望。那么$h_{i+1} = E\left((x+1)^3\right) = E(x^3 + 3x^2 + 3x + 1) = p_{i+1}(h_i + 3g_i + 3f_i + 1)$
- 最后$\sum h_i(1 - p_{i+1})$ 就是答案。

SHOI2017 分手是祝愿

- 有n盏灯，给定每一盏灯的初始状态。每次操作可以选择一盏灯，将它的所有因数和它翻转。每次操作策略如下：
  - 如果可以通过小于等于k次操作把所有灯灭掉，那么直接按照最优策略进行。
  - 否则，从1到n中随机选择一盏灯进行操作。
- 求把所有灯灭掉的期望步数。
- $1 \le n \le 10^5$

SHOI2017 分手是祝愿

- 首先策略是唯一的。
- （为什么？每次考虑最大的元素，它只能通过操作自己灭掉）
- 所以我们可以得到一个01串，里面的每一个位置表示这个位置是否需要翻转。
- 然后所有的位置就等价了。
- 设f_i表示有i盏灯亮着的时候，灭掉其中一盏灯的期望步数。则当$i > k$时，有$f_i = \frac{i}{n} + \frac{n-i}{n}(1 + f_{i+1} + f_i)$，即$f_i = \frac{n}{i} + \frac{n-i}{i} f_{i+1}$，当$i \le k$时$f_i = 0$。
- 所以可以$O(n)$递推。

组合数

- C_n^m或$C(n,m)$或$\binom{n}{m}$，读作n选m。
- 组合意义是从n个元素中选择m个元素的方案数。
- 二项式定理：$(x+y)^n=\sum_{0\le k\le n}\binom{n}{k}x^k y^{n-k}$。
- 组合数的相关性质：
  - $\binom{n}{m}=\binom{n-1}{m}+\binom{n-1}{m-1}$
  - $\binom{n}{m}=\frac{n(n-1)(n-2)\cdots(n-m+1)}{m(m-1)(m-2)\cdots 1}=\frac{n!}{(n-m)!m!}$
  - $\sum_{i=0}^{n}\binom{n}{i}=2^n$

小练习

- 有一张$n \times m$的网格纸，求从左下走到右上的方案数，每次只能往右走或者往上走。

- 答案就是$\binom{n+m}{n}$，也就是说总共需要走$n+m$次，其中有n次是往上走的。

AGC001E

- 给定两个长度为n的数组a,b，求下面这个式子：

$$\sum_{i=1}^{n}\sum_{j=i+1}^{n}\binom{a_i+a_j+b_i+b_j}{a_i+a_j}$$

- $1 \le n \le 200000, 1 \le a_i, b_i \le 2000$

AGC001E

- 不要被眼前的困难吓倒。
- 考虑组合意义。
- $\binom{n+m}{n}$表示从$(0,0)$走到(n,m)的方案数。
- 所以$\binom{n_1+n_2+m_1+m_2}{n_1+n_2}$就表示从$(-n_1,-m_1)$走到$(n_2,m_2)$的方案数。
- 所以原式就表示从任意一个$\left(-a_i,-a_j\right)$出发走到任意一个$\left(b_i,b_j\right)$的方案数之和。
- 直接让f[-a[i]][-a[j]]=1，然后递推。
- f[i][j]=f[i-1][j]+f[i][j-1]。

组合问题

- 一个很有用的计数原则：
- 一个由计数对象组成的集合S，要计算它的大小$|S|$，考虑如果我们找到一个集合T，使得S的元素与T的元素一一对应，那么$|S| = |T|$
- 一个推广：
- 如果S的每个元素对应到a个T中的元素，T的每个元素对应到b个S中的元素，那么有：
- $a|S| = b|T|, |S| = |T| \cdot \frac{b}{a}$
- 举个例子：n个元素的圆排列不好计算，但是一个圆排列枚举端点后会对应n个全排列，所以n个元素的圆排列数量就是$(n-1)!$

Twelve Fold Way

- 你需要解决12个组合计数问题。
- n个有标号/无标号的球放到m个有标号/无标号的盒子
- 盒子有三种限制：
- A、无限制
- B、每个盒子至少有一个球
- C、每个盒子至多有一个球
- 共有12种问题。

Twelve Fold Way

- 为了方便，我们设U代表无标号，L代表有标号，这样一个问题就可以用三个字母表示，第一个字母表示球是否有标号，第二个字母表示盒子是否有标号，第三个字母表示限制。

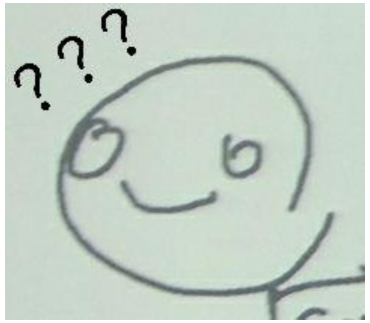
Twelve Fold Way

- 让我们先从简单的情况入手。
- 1、（LLA）n个有标号的球放到m个有标号的盒子里面。
  - 相当于为每一个球选一个盒子放进去，一个球有m种情况。
  - 所以答案就是m^n
- 2、（ULC）n个无标号的球放到m个有标号的盒子里面，每一个盒子最多放一个。
  - 相当于选择n个盒子放球。
  - 所以答案就是$\binom{m}{n}$
- 3、（LLC）n个有标号的球放到m个有标号的盒子里面，每一个盒子最多放一个。
  - 上一个的答案再乘$n!$，也就是$A(m,n)$。

Twelve Fold Way

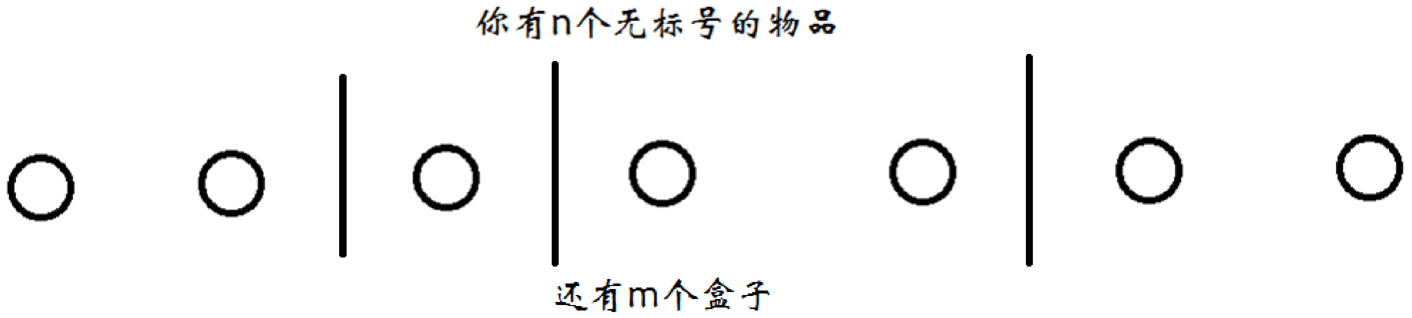
- 4、（LUC）n个有标号的球放到m个无标号的盒子里面，每一个盒子至多放一个。
  - emmmm……
  - 怎么放都是一样的。
  - 答案是$[n \le m]$。

- 5、（UUC）n个无标号的球放到m个无标号的盒子里面，每一个盒子至多放一个。
  - 这……
  - 还是一样的。
  - 答案仍然是$[n \le m]$。



Twelve Fold Way

- 6、（ULB）n个无标号的球放到m个有标号的盒子里面，不允许有空盒。
  - 让我们画一下图。



- 相当于你有$n-1$个间隔，你需要从中选择$m-1$个位置插板，将这n个小球分到m个盒子中。
- 所以答案是$\binom{n-1}{m-1}$。
- **这就是插板法（划重点）。**

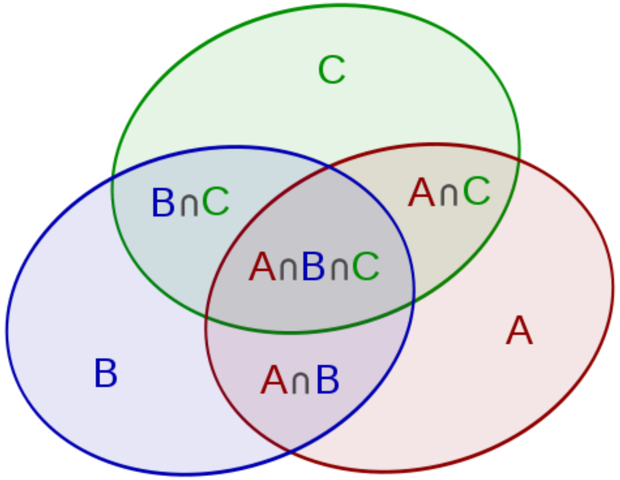
Twelve Fold Way

- 7、（ULA）n个无标号的球放到m个有标号的盒子里面，允许有空盒。
  - 如果再用插板法，板的位置可以重叠，不能直接计算。
  - 上一个问题相当于有m个变量$x_1, x_2, \cdots, x_m$，每一个变量都≥ 1，求它们的和是n的解的数量。
  - 而这一个问题和上一个几乎一样，唯一的区别是每一个变量都≥ 0。
  - 令$y_i = x_i + 1$，则$y_i \geq 1$，要求它们的和是$n + m$，转化为上一个问题。
  - 答案是$\binom{n+m-1}{m-1}$

容斥原理

$$|A \cup B| = |A| + |B| - |A \cap B|,$$

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$



$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{1 \le i < j \le n} |A_i \cap A_j| + \cdots$$
$$\cdots + \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1} |A_1 \cap \cdots \cap A_n|.$$

Twelve Fold Way

- 8、（LLB）n个有标号的球分给m个有标号的盒子，不能有空盒。
  - 用上刚学的容斥原理。
  - 枚举有i个盒子是空的，转化为LLA问题。
  - 答案是$\sum_{i=0}^{m}(-1)^i\binom{m}{i}(m-i)^n$
- 9、（LUB）n个有标号的球分给m个无标号的盒子，不能有空盒。
  - 和上一个几乎一样。
  - 这里面第一个方案标号后会对应上面的$m!$种方案，所以答案是LLB的答案除以$m!$。
  - 即答案等于$\frac{1}{m!}\sum_{i=0}^{m}(-1)^i\binom{m}{i}(m-i)^n$。
- 这就是第二类斯特林数。

第二类斯特林数

- 第二类斯特林数$S(n,k)$，也记作$\left\{ {n \atop k} \right\}$，即LUB的答案。
  - $\left\{ {n \atop m} \right\} = \frac{1}{m!}\sum_{i=0}^{m}(-1)^i\binom{m}{i}(m-i)^n$
- 也可以通过DP来求。我们枚举第n个球是单独放到一个盒子里面还是和前面的球一起放到一个盒子里面：
  - $\left\{ {n \atop k} \right\} = k\left\{ {n-1 \atop k} \right\} + \left\{ {n-1 \atop k-1} \right\}$
  - $\left\{ {0 \atop 0} \right\} = 1$，并且$\left\{ {n \atop 0} \right\} = \left\{ {0 \atop n} \right\} = 0$。

Twelve Fold Way

- 10、（LUA）n个有标号的球分给m个无标号的盒子。
  - 我们枚举有多少个盒子放了球。
  - 那么答案就是$\begin{Bmatrix} n \\ 1 \end{Bmatrix} + \begin{Bmatrix} n \\ 2 \end{Bmatrix} + \begin{Bmatrix} n \\ 3 \end{Bmatrix} + \cdots + \begin{Bmatrix} n \\ m \end{Bmatrix}$

- 12个问题解决了10个，还差两种！
- UUA、UUB

Twelve Fold Way

- 11、（UUB）n个无标号的球放到m个无标号的盒子里面，不能有空盒。
  - 考虑DP。
  - 设$f_{i,j}$表示将i个球放到j个盒子里面的方案数。注意因为盒子没有标号，所以需要强制球的数量单调。
  - 转移分两种情况：
    - 1、新开一个盒子，放入一个球，$f_{i,j} \to f_{i+1,j+1}$
    - 2、将现存的每一个盒子里面都放入一个球，$f_{i,j} \to f_{i+j,j}$
  - 每一个方案都可以通过这两种转移得到，同时每一种方案的操作顺序都唯一，所以这个DP求出的就是解。
- 这个数列也有名字，叫做划分数：
- $p_{n,m} = p_{n-1,m-1} + p_{n-m,m}$

Twelve Fold Way

- 12、（UUA）n个无标号的球放到m个无标号的盒子里面。
  - 方法1：枚举有几个盒子放了球。答案等于$p_{n,1}+p_{n,2}+\cdots+p_{n,m}$。
  - 方法2：想想插板法是如何扩展到盒子为空的情况的。
  - 每一个盒子都多放进去一个球。答案等于$p_{n+m,m}$

Twelve Fold Way

- 完结撒花！

是否有标号	A	B	C
LL	m^n	$S_{n,m}m!$	$\binom{m}{n}n!$
LU	$\sum_{i=1}^{m} S_{n,i}$	$S_{n,m}$	$[n \le m]$
UL	$\binom{n+m-1}{m-1}$	$\binom{n-1}{m-1}$	$\binom{m}{n}$
UU	$p_{n+m,m}$	$p_{n,m}$	$[n \le m]$

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}$$

$$S_{n,m} = \frac{1}{m!}\sum_{i=0}^{m}(-1)^i(m-i)^n\binom{m}{i}$$

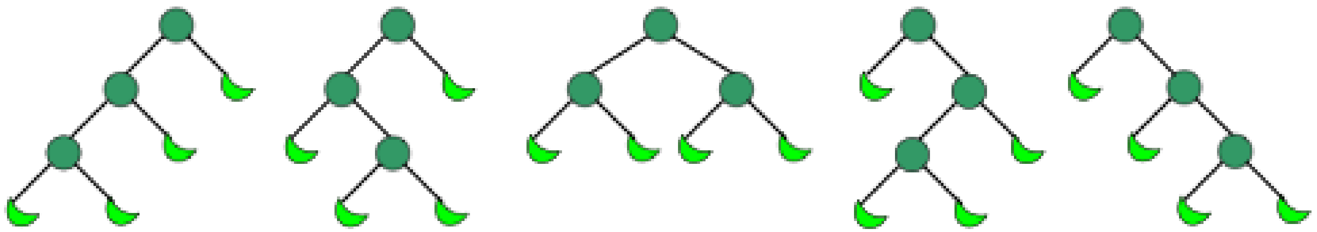
$$p_{n,m} = p_{n-m,m} + p_{n-1,m-1}$$

卡特兰数

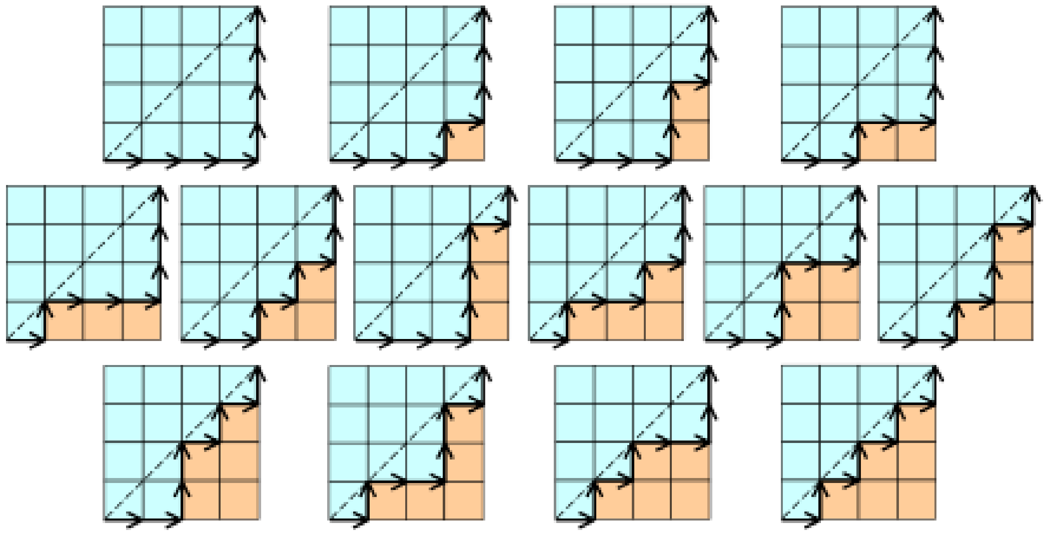
- $C_0 = 1, C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$
- $C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!} = \prod_{k=2}^{n} \frac{n+k}{k}$
- 1,1,2,5,14,42,132,429,1430…
- 组合应用：
- 长度为$2n$的合法的括号序列数量
- $C_3 = 5$，((())),(())(),(()()),()(()),()()()

卡特兰数

- 有n个非叶子节点的满二叉树的个数：

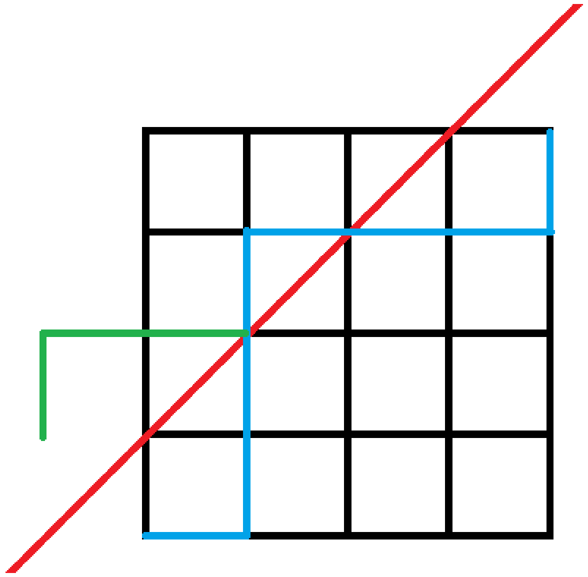


- 不越过对角线的路径条数：



卡特兰数

- 不超过对角线的路径条数
- 怎么求呢?
- 考虑容斥，总方案-超过对角线的路径条数。
- 注意到，每一个超过对角线的路径都会碰到y=x+1这条直线。
- 所以我们把所有不合法的路径沿着这条直线翻过去:
- 这样一条不合法的路径就对应了一条从$(-1,1)$走到(n,m)的路径，所以两者是等价的。
- 所以答案就是$\binom{2n}{n}-\binom{2n}{n+1}=\frac{(2n)!}{n!(n-1)!}\left(\frac{1}{n}-\frac{1}{n+1}\right)=\frac{(2n)!}{n!(n-1)!}\cdot\frac{1}{n(n+1)}=\frac{1}{n+1}\binom{2n}{n}$



容斥原理进阶

- 给定一个正整数n，求下面这个方程的解的个数：
- $\sum x_i = n$，其中$l_i \le x_i \le r_i$，共有m个数。

容斥原理进阶

- 首先可以把$n-\sum l_i$，接下来就变成了每一个元素有一个上界的问题。
- 为了方便，我们设$N=n-\sum l_i , b_i=r_i-l_i$
- 枚举子集进行容斥。首先枚举一个集合里面不满足上界，这样我们就需要强制这个集合里面的数字不满足上界，也就是说上界变成了下界。
- 把N减去这些不满足的上界，用插板法。
- $Ans=\sum_S(-1)^{|S|}\binom{N-\sum_{i\in S} b_i+m-1}{m-1}$
- 时间复杂度$O(m2^m)$，当N比较小的时候可以DP。

容斥原理进阶

- 你有一个1到n的排列p。
- 我们定义一个区间是好的，当且仅当这个区间的值域也是这个区间。
- 我们定义一个排列是不可分割的，当且仅当这个排列不能分成两个或以上的区间，使得这几个区间都是好的区间。
- 求存在多少个长度为n的不可分割的排列。

容斥原理进阶

- 设f_i表示长度为i的不可分割的排列的数量。
- 考虑容斥。注意到任意一个不合法的排列都一定存在一个数x，使得$[1,x]$这个区间是好的，并且对于任意的$y < x$，$[1,y]$都不是好的。
- 所以枚举x，可以得到容斥的公式：
- $f_n = n! - \sum_{i=1}^{n-1} f_i(n-i)!$
- 时间复杂度$O(n^2)$。
- ~~用多项式求逆可以做到$O(n\log n)$。~~