组合计数与概率期望

冉雨杭

January 16, 2023

组合计数

加法原理

集合S划分为部分$S_1, S_2, \ldots, S_m$。则S的元素个数可以通过找出它的每一部分元素个数来确定：$|S| = |S_1| + |S_2| + \ldots |S_m|$

乘法原理

S是元素的序偶(a, b)的集合，其中第一个元素a来自大小为p的集合，而对于任意a的选择，b都存在q种选择，于是：$|S| = p \times q$

组合计数

排列数

- n个元素中m个元素的有序摆放称为n个元素的一个m排列
- $A_{n,m} = n \times (n-1) \times (n-2) \times \cdots \times (n-m+1)$
- $A_{n,m} = \frac{n!}{(n-m)!}$

组合数

- n个元素中m个元素的无序摆放称为n个元素的一个m组合
- $C_{n,m} = \frac{n!}{m!(n-m)!}$

- 当上述计数在模质数意义下进行的时候，可以预处理阶乘及其逆元，每次$O(1)$计算
- 当n大m小时，每个组合数可以$O(m)$进行计算

组合计数

- 组合恒等式： $C_n^m = C_{n-1}^m + C_{n-1}^{m-1}$
- 令S是一个多重集，它有k个类型不同的元素，每一种元素都可以重复无数次，那么S的r排列个数是k^r
- 令S是一个多重集，它有k个类型不同的元素，各元素重复次数为$n_1, n_2, \ldots n_k$。令$n = n_1 + n_2 + \cdots + n_k$，那么$S$的排列个数是$\frac{n!}{n_1!n_2!\ldots n_k!}$

组合计数

- n个数的排列方案数是$n!$
- 从$(0,0)$走到(n,m)，每次只能向上或者向右走，方案数为C_{n+m}^{n}
- n个元素的圆排列方案数是$(n-1)!$
- n个球分到m个盒子里，每个盒子不为空的方案数为C_{n-1}^{m-1}
- n个球分到m个盒子里，每个盒子可为空的方案数为C_{n+m-1}^{m-1}
- n个正整数求和小于等于m的方案数为C_{m}^{n}

怎么做好组合计数问题

- 掌握好基础的模型
- 找到组合问题中的计数突破口，有时候需要换角度看问题
- 熟练利用计算"贡献"的思想解决问题

隔板法模型

- n个正整数的和为m的方案数
- n个非负数的和为m的方案数
- $0 \leq a_1 \leq a_2 \leq \cdots \leq a_n \leq m$的方案数

排队问题

- 有n人站在m站，排成m队(队伍带编号)。你不知道哪个人在哪个站，也不知道他们是按照什么顺序排队的，所以你想统计不同方案的数量。
- $T \leq 100, 1 \leq m \leq n \leq 10^5$

排队问题

- 先强制大家按一个顺序站好，方案数是$n!$
- 再分成非空的m组，方案数是C_{n-1}^{m-1}
- 最后答案就是$n!C_{n-1}^{m-1}$

双数组计数

- 给定两个正整数n和m，计算满足以下条件的数组对(a, b)的个数:
    - 两个数组的长度都是n;
    - 两个数组中的每个数都是$[1, m]$中的整数;
    - 对于每个$1 \le i \le n$，都有$a_i \le b_i$;
    - 数组a中的元素从左到右不降;
    - 数组b中的元素从左到右不升,
- $1 \le n, m \le 10^6$

双数组计数

- 两条折线最多在一个y值上相交
- 枚举a_n的值x
- 要求$1 \le a_1 \le a_2 \le \cdots \le a_{n-1} \le x$
- 要求$1 \le b_1 \le b_2 \le \cdots \le b_n \le m - x + 1$
- 差分，变成求n个数的和不超过m的模型
- 答案为$\sum_{x=1}^{m} f(n-1, i) f(n, m-i+1)$，其中$f(n, m) = C_{m+n-1}^{n}$

斯特林数

第一类斯特林数

- n个元素组成m个圆排列的方案数
- $S(n,m) = S(n-1,m-1) + (n-1)S(n-1,m)$

第二类斯特林数

- n个元素放到m个相同的盒子里的方案数
- $S(n,m) = S(n-1,m-1) + mS(n-1,m)$

第二类斯特林数

- $S(n,m)=S(n-1,m-1)+mS(n-1,m)$
- $S(n,m)=\frac{1}{m!}\sum_{k=0}^{m}(-1)^k C(m,k)(m-k)^n$
- 可以在$O(mlogn)$的时间求出$S(n,m)$
- 用fft优化可以在$O(nlogn)$的时间求出一个n的所有$S(n,m)$

棋盘问题

- 给定$n \times n$的网格图，要求放n个棋子。一个格子能被某个棋子攻击到当且仅当有棋子跟它同行或者同列且它们之间没有其他棋子。问所有空格都能被攻击，且恰好有k对棋子能互相攻击的方案数有多少
- $n \leq 2 \times 10^5, k \leq \frac{n(n-1)}{2}$

棋盘问题

- 所有空格都被攻击→每一行都有棋子或者每一列都有棋子，只考虑行的情况，列情况是对称的。
- 一行填x个可以造成$x-1$次互相攻击
- 可以发现一共有$n-k$行要填上棋子，可以直接用第二类斯特林数进行计算。
- 最后答案为$C_n^k S(n,n-k)(n-k)!$

一类计算贡献的技巧

- 对于每种情况x，有$f(x)$个东西满足给定条件，求所有情况下满足条件东西的和
- 可以变成枚举每个东西y，求有多少种情况$f(y)$使得它能满足条件，这也就是常说的单独计算每个点对于答案的"贡献"

一类计算贡献的技巧

- 求所有排列的逆序对之和
- 考虑一共有$\frac{n(n-1)}{2}$对数，任何一对有$\frac{n(n-1)}{2}$种放法
- 有$(n-2)!$情况能让它们称为逆序对
- 答案为$\frac{n(n-1)}{4}n!$

矩阵计数1

- 给定一个$n \times m$的01矩阵，记f_i为恰好有i个1的子矩阵个数，求$\sum_i i * f_i$
- $n, m \leq 2000$

矩阵计数1

- 考虑每个位置的1在每个包含它的矩形中都会被算到一次
- 答案为$\sum_i \sum_j i(n-i+1)j(m-j+1)[a_{i,j} == 1]$

一类计算贡献的技巧

- 对于每种情况x，有$f(x)$个东西满足给定条件，求所有情况下$f^k(x)$的和
- $f^k(x) = (1+1+1+1+\cdots+1)^k$
- 选中其中k个东西（有序且可重复），然后计算有多少种情况使得这k个物品的并能够同时满足条件

开关问题

- 有n盏灯，m个开关，每个开关会把某个子集中所有灯的开关状态翻转，问2^m种开关状态下开着灯数量的三次方和是多少?
- $n, m \leq 50$

开关问题

- x^3可以转换为枚举3个灯的位置，然后求有多少种方案使得这三个灯最后都是开着的
- 后者可以用一个dp完成，记录三盏灯的开关情况即可
- 复杂度$O(n^3m)$

背包装球

- 有给定三个正整数n, m, k有n个盒子，每个盒子有m个标号$1 - m$的球，现从每个盒子选出恰好一个球，将奇数编号的球的个数记为x，求所有方案的x^k的和
- $T \leq 5000, 0 \leq n, m < 998244353, k \leq 2000$

背包装球

- n, m都特别大，只能考虑从比较小的k下手
- 发现如果用贡献的方法选k个球出来，那么最多有k组
- 考虑选最后k个球一共在i个组里面，使这k都选的是奇数的球，剩下组任意选
- 方案数为$C_n^i S_{k,i}(\lceil \frac{m}{2} \rceil)^i m^{n-i}$
- 预处理第二类斯特林数即可，每次可以$O(k)$回答一组询问

矩阵计数2

- 给定一个$n \times m$的01矩阵，记f_i为恰好有i个1的子矩阵个数，求$\sum_i i^2 * f_i$
- $n, m \le 2000$

矩阵计数2

- 枚举两个位置$(x_1, y_1), (x_2, y_2)$
- 贡献为$x_1(n+1-x_2)y_1(m+1-y_2)$
- 枚举第二个位置，维护第一个位置乘积的前缀和即可

进位问题

- 给定n, k，问所有$0 \leq a, b < 2^n$的(a, b)中，有多少对(a, b)满足$a + b$的进位次数为k
- $n, k \leq 10^6$

进位问题

- x个数分到y个盒子里，每个盒子不为空的方案数为C_{x-1}^{y-1}
- 令c_i表示第i位的进位情况
- 当$c_i = c_{i-1}$，有三种填法，否则只有一种填法
- 当$c_i \neq c_{i-1}$记成一次切换，枚举有q次切换，令$c_i = 0$的连续段有x个，$c_i = 1$的连续段有y个，这样的x, y只有常数个
- $c_i = 0$的连续段有x个，$c_i = 1$的连续段有y个，此时方案数$C_{k-1}^{x-1} C_{n-k-1}^{y-1}$

容斥原理

容斥原理

令X为一个有限集合，$P_1, P_2, \ldots, P_m$是一些性质的集合，对于任意$S \subseteq [m]$我们定义$N(S) = \{x \in X \mid \forall i \in S : x \text{ has } P_i\}$. 那么$X$中不满足任何一个性质$P_i$的元素个数为$\sum\limits_{S \subseteq [m]} (-1)^{|S|} |N(S)|$

- $|\bigcup_{i=1}^n A_i| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}|$
- $\left|\overline{A_1} \bigcap \overline{A_2} \bigcap \ldots \bigcap \overline{A_n}\right| = N - |A_1 \bigcup A_2 \bigcup \ldots \bigcup A_n|$
- 容斥系数只跟集合大小的奇偶性有关

错位排列

- 求有多少排列满足$\forall i, p_i \neq i$
- $n \leq 10^6$

错位排列

- 令A_i表示位置i满足$i \neq p_i$
- 答案求$|\cap_i A_i|$
- $|\cap_i A_i| = C_n^0 n! - C_n^1 (n-1)! + C_n^2 (n-2)! + \ldots + (-1)^n \cdot C_n^n 0!$

GCD计数

- 给定n, m和序列$(b_1, b_2, \ldots, b_n)$, 问有多少种序列$(a_1, a_2, \ldots, a_n)$满足$1 \leq a_i \leq m$且$gcd(a_1, a_2, \ldots, a_i) = b_i$
- $n \leq 2 \times 10^5, m \leq 10^9$

GCD计数

- $b_i \neq b_{i-1}$最多只会发生log次
- $b_i = b_{i-1}$时，只需要填b_i的倍数即可
- $b_i \neq b_{i-1}$，b_i一定是b_{i-1}的因子，那么令$g = \frac{b_{i-1}}{b_i}$，则填法可以是$[1, \lfloor \frac{m}{b_i} \rfloor]$中与$g$互质的数
- 枚举所有因子进行容斥即可

染色问题

- 给定n, m, k，用m种颜色中的恰好k种给n个物品染色，且相邻元素不能是相同颜色，问方案数
- $1 \le n, m \le 10^9, 1 \le k \le 10^6$

染色问题

- 选k种颜色的方案数是C_m^k
- 枚举有多少种颜色没用，贡献是$(-1)^{k-i+1}C_k^i * i * (i-1)^{n-1}$
- 答案为$C_m^k\sum_{i=0}^{k}(-1)^{k-i+1}C_k^i * i * (i-1)^{n-1}$

树上匹配计数

- 给定$2n$个点的树，要求两两匹配。两个点能匹配当且仅当两点之间没有连边，问有多少种匹配方案
- $n \le 2000$

树上匹配计数

- 发现没有连边的匹配不好算，但是有连边的匹配是好算的
- 可以用dp求出这个树至少有i对有连边的匹配的方案数
- 容斥即可算出答案
- 注意树上背包的写法和复杂度

树上双联通计数

- 给定n个点m条边的无向联通图，问有多少种加边方式使得图变成双联通的（任意两点至少有两条边不相交的路径）
- $n \le 5000, m \le 10000$

树上双联通计数

- 双联通应该怎么判定？（dfs树）
- 如果是树该怎么做?
- 图和树有什么区别

树上双联通计数

- dfs树上，可以将树划分成若干个联通块，里面任意连边，联通快之间只有树边
- 记f_i表示划分成i个联通块的方案数，答案为$f_1 - f_2 + f_3 - f_4 \cdots + (-1)^{i-1} f_i$
- 于是用$dp_{i,j,k}$表示当前在i及其子树，已经划分了j个联通块，最上面的联通块有k个点的方案数
- 容易发现答案之跟j的奇偶性有关，每次生成一个新联通块会乘上-1，于是可以把第二维去掉
- 变成图就是在dfs树上被覆盖的边不能切断，dp的时候特判即可
- 复杂度$O(n^2 + m)$

排序问题

- 给定n，表示一个长度为$3n$的排列。你每次可以对前$2n$个数或者后$2n$个数进行排序，定义一个排列的代价$f(p)$为排列p变为有序所需要花费的最小次数，求所有长度$3n$的排列的$f(p)$之和
- $n \leq 2 \times 10^5$

排序问题

- 任何一种排列的操作次数不会多于3
- 0次和1次操作的比较好计算
- 只用考虑2次和3次的方案数谁比较好算，再用总的减去即可得到另一个的方案数
- 需要3次当且仅当前n个数包含$[2n+1,3n]$范围内的数且后n个数包含$[1,n]$范围内的数

排序问题

- 0次的方案数为1
- 1次的方案数$2*(C_{2n}^{n} \times n! \times n! - 1) - (n! - 1)$
- 3次的方案数考虑求其反面，有三种情况，1不包含3，3不含1，13互不包含
- 前两种方案数为$C_{2n}^{n} n! \times (2n)!$
- 互不包含考虑枚举$[n+1, 2n]$中有x个给1填了，方案数为$C_n^x C_n^{n-x} C_{2n-x}^{n} (n!)^3$
- 3的方案数：$(3n)! - (2C_{2n}^{n} n! \times (2n)! - \sum_{x=0}^{n} C_n^x C_n^{n-x} C_{2n-x}^{n} (n!)^3)$

概率与期望

期望

- $E[X] = \sum_x xP(x)$

期望的线性性

- $E[X+Y] = E[X] + E[Y]$
- 各事件和的期望等于各事件期望的和
- 不要求变量之间独立
- 期望的线性性完美契合了计算"贡献"的思想，很多期望的题目都是考虑每一个独立点对答案的"贡献"

例题

- 随机一个长度为n的排列，i和p_i连一条边，问期望有多少个联通块
- $n \le 2 \times 10^5$

例题

- 考虑每一个联通块的贡献，假设长度为i，贡献是$C_n^i(i-1)!(n-i)!\frac{1}{n!}=\frac{1}{i}$
- 总方案数是$\sum_{i=1}^{n}\frac{1}{i}$
- 考虑有更简单的方法吗?
- 每次考虑一个点，它能新建一个联通块的概率是$\frac{1}{n}$，对答案贡献就是$\frac{1}{n}$，问题规模变成$n-1$

小球染色

- 给定n个初始颜色不同的小球，每次操作随机选择两个小球出来，把第二个变成第一个的颜色，然后再放回去。问所有小球同色的期望操作次数
- $n \leq 2 \times 10^5$

小球染色

- 只用考虑某种颜色即可，最后答案乘上n即可
- 令$p_i = \frac{2i(n-i)}{n(n-1)}$
- f_i为当前有i个某种颜色同色的小球到全部同色所需的期望步数，则有$f_i = p_i(f_{i-1} + 1) + p_i(f_{i+1} + 1) + (1 - 2p_i)(f_i + 1)(1 < i < n)$
- $f_n = 0, f_1$需要单独考虑即可
- 答案为nf_1

区间本质不同数

- 给定n个数，随机一对l, r，若$l > r$则先交换l, r，问本质不同的数的期望是多少?
- $n \leq 10^6, a_i \leq 10^6$

区间本质不同数

- 对于一个区间，我们只在每种数第一次出现的位置计数
- 考虑每个点作为某个第一次出现时，包含它的区间的方案数
- 左端点属于$[last_{a_i}+1, i]$，右端点属于$[i, n]$

删树游戏

- 给一个1为根的树，每次随机选一个点，删去其子树，当选到1的时候结束，问期望操作次数
- $n \leq 10^5$

删树游戏

- 期望删除次数= 每个点期望被选的次数之和
- 每个点是否被选只与它到祖先那条链上的所有点有关，这个点能被选当且仅当它是这些点当中第一个被选的
- 假设有x个点，这个概率是$\frac{1}{x}$
- 答案就是所有深度的倒数和

凸包面积

- 给定n个点的凸多边形，随机选k个点，期望面积是多少
- $k \leq n \leq 2500$

凸包面积

- 枚举两个点，考虑这块不被包含的期望面积是多少
- $\frac{C_{n-(j-i+1)}^{k-2}}{C_n^k} * area_{i,j}$
- 答案为：总面积- $\sum_i \sum_j \frac{C_{n-(j-i+1)}^{k-2}}{C_n^k} * area_{i,j}$

猜数

- 给定n, k，一共4种花色，每种花色有$1-n$的n张牌。现在一个人开始抽牌，它会记得最近k张牌的结果（如果不足k张则是全部)。他会选择其中出现次数最少的花色作为他下一次的猜测。问期望能猜对多少次
- $k, n \leq 500$

猜数

- 只用单独考虑每个位置的贡献即可
- 多于k张牌时，不用管前面忘掉的牌是啥，只用考虑最近k张牌出现次数最少的牌是多少即可
- 用$dp_{i,j}$表示抽了i张牌，其中花色最少的牌为j的概率是多少
- 枚举每一个预测位置i，考虑其对答案的贡献即可，求和即是答案