概率期望与组合计数

陈雨昕

Removing Blocks

AGC028B

- 现在有 N 条线段排成一行，其中第 i 条线段的长度为 A_i
- 现在要按照一个顺序删除所有线段，以下定义其代价
- 相邻的没有被删除的线段会连在一起
- 删除连起来的一些线段中的任意一条线段，花费的代价为这些线段的总长
- 一个删除顺序的代价，就是每条线段的删除代价总和
- 对于所有 $1, 2, \ldots, N$ 的排列 $p_1, p_2, \ldots, p_N$, 求依次删除 $p_1, p_2, \ldots, p_N$ 的代价之和
- 答案对 $10^9 + 7$ 取模
- $N \le 10^5,\ 1 \le A_i \le 10^9$, 2s, 1GB

解

- 均匀随机一个顺序，计算代价的期望，再乘以 $N!$ 就是答案
- 建立删除时间的笛卡尔树（小根）
- 一条线段 A_i 对代价的贡献倍率，就是其在笛卡尔树上的深度 h_i（根的深度为 1）
- 现在只需知道 i 的期望深度，又转化为对于 $j = 1, 2, \ldots, n$, j 为 i 的祖先的概率之和
- 若 $j < i$, 那么 j 是 i 的祖先，当且仅当 $j+1, j+2, \ldots, i$ 均迟于 j, 因此概率为 $\frac{1}{i-j+1}$
- 同理若 $j > i$, 概率为 $\frac{1}{j-i+1}$
- 因此设调和级数 $H_n = \sum_{i=1}^{n} \frac{1}{i}$, 有 $E(h_i) = H_i + H_{n-i+1} - 1$
- 时间复杂度：$O(n)$

例题

- 在一棵 N 个结点的 Treap 中，各结点的键值为 1, 2, ..., N
- 设键值为 i 的点权值为 p_i, p_1, p_2, ..., p_N 从 1, 2, ..., N 的排列中均匀随机
- 根的深度为 1, 求所有结点的深度平均值的期望
- 保留 7 位小数
- $N \le 10^{15}$, 1s, 128MB

解

- 直接利用刚才的结论，期望为
  $\frac{1}{n}\sum_{i=1}^{n}(H_i + H_{n-i+1} - 1)$
  $= \frac{2}{n}\sum_{i=1}^{n} H_i - 1$
- $= \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{1}{j} - 1$
  $= \frac{2}{n}\sum_{j=1}^{n}\frac{n-j+1}{j} - 1$
  $= 2\left(1 + \frac{1}{n}\right)H_n - 3$
- 利用不等式 $\ln\frac{n+1}{n} \le \frac{1}{n} \le \ln\frac{n}{n-1}$
- 于是有 $\ln\frac{n+1}{m} \le H_n - H_m \le \ln\frac{n}{m-1}$
- 取上下界平均值作为近似值
- 取阈值 $M = 5000000$
- 对 M 以内的部分打表
- 对 M 以上的部分利用近似公式计算
- 时间复杂度 $O(1)$

残缺的算式

FJWC2018

- 给定一个算术表达式 S, 其中含有加号、减号、乘号、左右括号和若干空位
- 保证将空位填上操作数后是一个合法表达式
- 现在给空位均匀随机，但不重复地填入 1, 2, ..., n, 求该算式的结果的期望值
- 模 10^9+7
- $n \le 10^9$, $|S| \le 10^4$, 2s, 512MB

解

- 利用期望的线性性，我们可以将括号展开
- 记由 m 个操作数连乘组成的项系数和为 $f(m)$, 可以通过树形 DP 求出
- 记 $g(m)$ 表示从 1, 2, ..., n 中均匀（无序）随机的 m 个不同数的连乘积之和
- 则 $g(m)$ 为第一类斯特林数 $s(n, n-m)$
- 答案即为 $\sum_{m=0}^{|S|} \frac{f(m)g(m)}{\binom{n}{m}}$
- $g(m)$ 显然是关于 n 的 $2m$ 次多项式，可以用下降幂多项式来递推
- 也可以默写组合恒等式
- 时间复杂度 $O(|S|^2)$

树的计数

NOI2013

- 已知树有 n 个结点，并已知树的DFS、BFS序
- 求所有符合条件的树的高度平均值，四舍五入保留三位小数
- $2 \le n \le 2 \times 10^5$
- 1s, 256MB

解

- 不妨将BFS序重排为 $1, 2, ..., n$, 考虑分层的方式
- 一种分层法在DFS序限制下至多能对应一棵树，考虑怎样的分层方式能对应到树
  - 根独占一层
  - 连续段内DFS序号递增
  - DFS序中后一位的深度至多比前一位大 1
- 如果 i, $i+1$ 不同层，称 i 为分段点
  - 1 是分段点，下称必分段点
  - 若 i 比 $i+1$ 在DFS序中迟出现，则 i 是分段点，下也称必分段点
  - 若DFS序中 u, v 相邻，且 $u < v$, 则 $[u, v)$ 中至多一个分段点，下称限制区间

解

- 对于限制区间 $[u, v)$, 若有一个位置为必分段点，则其余的位置都不能分段
- 若没有位置为必分段点，则 $u \neq 1$ 且 $u, u+1, \ldots, v$ 在DFS序中顺序出现
- 所以只有 $v = u + 1$, 此限制无效
- 因此用差分维护这些点有没有被有效限制区间覆盖
- 若没有，则每个非必分段点，有 $\frac{1}{2}$ 可能性属于分段点
- 因此答案就是 必分段点数 $+ \frac{\text{未覆盖的非必分段点数}}{2}$
- 时空复杂度 $O(n)$

Tree Array

CF1540B

- 给一棵 n 个点的树，初始都是白色的
- 首先均匀随机一个点涂黑
- 然后每次从所有与黑点相邻的白点中均匀随机一个点涂黑
- 点的编号按涂黑时间顺序构成的排列中，逆序对数的期望
- 模 $10^9 + 7$
- $n \leq 200$
- 2s, 256MB

解

- 枚举 $a < b$, 求 a 比 b 迟涂黑的概率
- 把第一个涂黑的点提根，设为 r
- 设提根后LCA为 c, 那么首先会涂黑 $r \to c$, 然后分别涂黑 $c \to a,\ c \to b$
- 设长度分别为 $x,\ y$, 注意每次在两条路径上走出一步的概率是相等的
- 因此可预处理这一概率 $f(x,\ y) = \frac{1}{2}\big(f(x-1,\ y) + f(x,\ y-1)\big)$
- 时间复杂度 $O(n^3)$, 空间复杂度 $O(n^2)$

Games on DAG

AGC016F

- 给一个 N 个点、M 条边的 DAG, 点已经按照拓扑序编号好
- 在 1, 2 号点上各放置一枚棋子，每次可以选恰好一枚棋子沿边走恰好一步
- Alice与Bob轮流操作
- 双方都可以动两颗棋子，棋子可重叠
- 不能操作者判负
- 求原 DAG 的 2^M 张生成子图中有多少是先手必胜的，模 10^9+7
- $N \leq 17$

解

- 应用博弈论的结论，有 $SG(1) \neq SG(2)$, 只需求 $SG(1) = SG(2)$ 的情况数
- 从 SG 入手，从小到大不好做，考虑从大到小确定各个 SG 值的点集
- 若 S 是 $SG > k$ 的点集，T 是 $SG = k$ 的点集，考察新增的连边状况：
  - $S \to T$: 每个 S 中结点至少连一条边
  - $T \to S$: 任意连边
  - $T \to T$: 不准连边
- k 是不必要的
- 预处理点到集合的边数
- 时间复杂度 $O(n3^n)$, 空间复杂度 $O(2^n)$

Sequence Growing Hard

AGC024E

- 给出 N, K
- 从空序列 A_0 开始，往 A_i 中插入一个 1 至 K 的整数，得到 A_{i+1}, 重复 N 轮
- 求字典序递增的 $A_0, A_1, \ldots, A_N$ 的方案数，模给定的数 M
- $N, K \le 300$, 2s, 1GB

解

- 设第 i 步将 v_i 插入 A_{i-1} 的第 p_i 个位置前（或者 $p_i = i$ 即放在最后一个）
- 则需要保证的就是 $v_i > A_{i-1,p_i}$
- 在 A_N 中，考虑每个数的加入时刻，构成笛卡尔树（小根）
- 那么意思就是，一个结点的左子树内所有结点的值，要大于它自身的值
- 因此将序列的计数转化为了笛卡尔树的计数
- 记 $f(i, j)$ 表示一棵 i 个点的合法带权二叉树，所有权值属于 $[j, K]$ 的方案数
- 列出递推 $f(i, j) = \sum_{x=j}^{K} \sum_{l=0}^{i-1} f(l, x+1) f(i-l-1, j)$
- 后缀和优化后时间复杂度 $O(N^2K)$

Placing Squares

AGC013E

- 给出一条长度为 N 的线段，将它划分为若干条长度为整数的子线段
- 给出 M 个特殊位置，规定它们不能作为分割点
- 一种划分方式的权值，是各子线段的长度平方之积
- 求所有划分方式的权值和
- $1 \le N \le 10^9, 0 \le M \le 10^5$

解

- 记特殊点的集合为 X
- 记 $f(i)$ 表示对线段的前 i 单位长度进行划分，且 i 作为划分点的权值和
- $f(0)=1,\ f(i)=[i\notin X]\sum_{l=1}^{i} l^2 f(i-l)$
- 难以通过，考虑优化

解

- 为 l^2 赋予组合意义
- 将长度为 l 的线段均分为 l 格，并在其中放置黑白棋子各一枚的方案数为 l^2
- 记 $g(i,\ j)$ 表示对线段的前 i 单位长度进行划分，i 不一定是划分点，使得第 i 单位长度所在线段现在摆放了 j 枚棋子的方案数
- 转移：$g(i,\ j)=[i-1\notin X]g(i-1,\ 2)+\sum_{k=0}^{j}\binom{j}{k}g(i-1,\ k)$
- 对于 $i-1\notin X$ 的连续一段，将 j 维写成矩阵乘法快速转移
- $O(M\log N)$

Gem Island

WF2018

- n 个人，每个人一颗宝石
- 每天晚上均匀随机一颗宝石分裂成两颗
- d 个晚上后，求持有宝石前 r 多的人的总宝石数期望
- 输出小数，绝对或相对误差不超过 10^{-6}
- $n, d \leq 500$
- 3s, 1GB

解

- 首先总分裂方案数是 $n(n+1)\cdots(n+d-1)$, 可预处理阶乘的对数
- 每次分裂时将宝石主人记下来，构成一个序列
- 若 i 被记 a_i 次，则有 $a_i!$ 种方法选择宝石来分裂，与总排法数 $\frac{d!}{\prod a_i!}$ 相乘只剩下 $d!$
- 作一步和式变换，若 $a_i \geq x$ 的有 b_x 个，则对答案贡献为 $\min\{b_x, r\}$
- 现在问题转化为：已知 $\sum_{i=1}^{n} a_i = d$, 求所有 $\sum_{x=0}^{d} \min\{b_x, r\}$ 的和
- 记 $f(i, j)$ 表示 j 个数的和为 i 的方案数, $g(i, j)$ 表示在此情况下 $\sum_{x=0}^{d} \min\{b_x, r\}$ 的和
- 枚举 $b_1 = k$: $f(i, j) = \sum_{k=0}^{j} \binom{j}{k} f(i-k, k)$, $g(i, j)$ 相应转移
- 时间复杂度 $O(n^3)$, 空间复杂度 $O(n^2)$, 实验表明能够胜任精度要求

Arcs on a Circle

AGC020F

- 有一个长度为 C 的圆周
- 现有 N 段圆弧，长度分别为 $L_1, L_2, \ldots, L_N$
- 将每个圆弧等概率随机放置
- 求每个位置都至少有一段圆弧覆盖的概率
- 要求绝对误差不超过 10^{-11}
- $2 \le N \le 6,\ 2 \le C \le 50,\ 1 \le L_i < C$, 输入均为整数
- $5s,\ 512MB$

解

- 我们肯定希望破环为链，不妨设逆时针为正方向，第 N 条线段最长并占据 $[0, L_N]$
- 对于余下的部分，设第 i 个圆弧的起始位置为 $x_i, a_i = \lfloor x_i \rfloor, b_i = \{x_i\}$
- 对于 b_i, 只需要考虑其大小关系，可以枚举，每种大小关系均有 $\frac{1}{(N-1)!}$ 概率
- 将所有 $p + b_i \ (0 \le p < C,\ 1 \le i < N)$ 作为关键点
- 记 $f(i, S, j)$ 表示考虑前 i 处关键点，当前放置的圆弧集合为 S, 最远覆盖到了第 j 个关键点的概率
- 时间复杂度 $O\big((N-1)!\,2^N(CN)^2\big)$, 空间复杂度 $O(2^N CN)$

另解

EntropyIncreaser

- 记 $f_i(S, jN+p)$ 表示将 $a \le i$ 的放置完毕，已经放置了集合 S
- 最远的终点的整数部分为 j, 小数部分为已出现的 b 中第 p 名
- 从 $f_{i-1}(*,\ *)$ 转移到 $f_i(*,\ *)$, 放置 $a=i$ 的弧，暂不考虑覆盖圆周的限制
- 记 $g_{i,k}(S, jN+p)$ 表示考虑了编号 $<k$ 的弧的状态
- 那么 $g_{i,k}(S, jN+p) \to g_{i,k+1}(S \cup \{k\}, \max\{jN+p+[q \le p], (i+L_k)N+q\})$
- 当 $i=j$ 时，须扣除所有插入的数都比 $f_{i-1}(*,\ *)$ 中的 $p=p_0$ 大的方案数
- $h_{i,k,p_0}(S, jN+p) \to h_{i,k,p_0}(S \cup \{k\}, \max\{jN+p+[q \le p], (i+L_k)N+q\})\ (q > p_0)$
- $f_i(S, jN+p) = g_{i,N-1}(S, jN+p) - \sum_{p_0=0}^{p} h_{i,N-1,p_0}(S, jN+p)$
- 时间复杂度 $O(2^N C^2 N^3)$, 空间复杂度 $O(2^N CN)$

又一另解

alphaGem

- 不妨设逆时针为正方向，第 N 条线段最长并占据 $[0, L_N]$
- 把每个单位长度平均分成 m 段，它们的端点作为关键点，只能在关键点放置线段
- 记 $f(i, S, j)$ 表示考虑前 i 处关键点，当前放置的圆弧集合为 S, 最远覆盖到了第 j 个关键点的方案数
- 这个 DP 的时间复杂度为 $O(2^N(Cm)^2N)$
- 设总方案数为 $F(m)$, 那么答案即 $\lim_{m\to\infty}\frac{F(m)}{(Cm)^{N-1}}$
- 从组合意义看，$F(m)$ 为关于 m 的多项式，且次数必然不高于 $N-1$
- 算出 $m = 1, 2, \ldots, N$ 处的点值，求其 $N-1$ 次项系数
- 总时间复杂度 $O(2^NC^2N^4)$

Intergalaxy Trips

CF605E

- 给定一张 n 个点的图，目标是花费最小的天数从 1 到 n
- 每一天，都有 p_{ij} 的概率存在边 $\langle i, j\rangle$，花费恰好 1 天
- 每天可以选定一条存在的边出发，也可以选择留在原地不动（即 $p_{ii} = 1$）
- 求最优策略下的期望天数
- $n \le 1000$, 2s, 256MB

解

- 设点 i 到点 n 的最优策略期望天数为 E_i
- 那么 $E_n = 0$, $E_i = 1 + \sum_{\{i\} \subseteq S \subseteq \{1,2,\dots,n\}} \prod_{j \in S} p_{ij} \prod_{j \notin S} (1 - p_{ij}) \min_{j \in S} \{E_j\}$
- 但是无法确认 E_j 的大小关系
- 可以采用最短路的 Dijkstra 算法来消除后效性，具体地说：
- 假设已经确认了期望最小的 k 个点，依次为 $p_1, p_2, \dots, p_k$
- 期望第 $k+1$ 小的点 i 满足 $\left(1 - \prod_{j=1}^{k} (1 - p_{ij})\right) E_i = 1 + \sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} (1 - p_{ij})\right) p_i E_j$
- 按照此公式算出所有未确认点的期望，取其中最小的
- 这些量可以动态更新，时间复杂度 $O(n^2)$

軍艦ゲーム

ARC016D

- N 个点 M 条边的简单DAG
- 1 号点是母港，N 号点是最终目的地，离开 1 号点时HP为 H
- 每个回合将等概率随机选取一条出边走过去，花费 1 时间，并发生战斗
- 每次到达第 i 个点，战斗都将掉 D_i 点HP, HP必须始终保持为正
- 战斗后有两种选择：传送回母港或继续走
- 如果传送回母港时HP为 C, 则花费 $H-C$ 时间修船
- 求最优策略下的期望最短时间，绝对或相对误差不超过 10^{-6} 即算正确
- $2 \le N \le 100,\ 1 \le H \le 100,\ 0 \le D_i \le 100,\ D_1 = 0,\ D_N \ne 0$, 答案不超过 10^6
- 2s, 64MB

解

- 首先不难列出DP $f(u,\ C)$ 表示位于 u 点，HP为 C 时的期望最短时间
- 转移为 $f(u,\ C) = \min\left\{f(1,\ H) + H - C, 1 + \mathbb{E}_{\langle u,\ v\rangle\in E}\left(f(v,\ C - D_v)\right)\right\}$
- 边界 $f(u,\ C) = \infty\ (C \le 0), f(N,\ C) = 0\ (C > 0)$
- 现在问题就是如何消去后效性，注意后效性只与答案 $A = f(1,\ H)$ 有关
- 把 A 当作变元，考察 $f(1,\ H)$ 所有可能的转移路径，那么无不是 A 的一次函数
- 其最小值是一个上凸壳，且斜率始终不超过 1
- 所以 $f(1,\ H) = A$ 恰好有一个解，通过二分找出

Balance Beam

USACO18DEC

- 现在有一条纸带，纸带上有 n 个格子，第 i 个格子上写有数 v_i
- 初始将棋子放进第 k 格，玩一个游戏，规则是：
  - 每一轮可以选择结束游戏或者继续游戏
  - 如果选择结束游戏，得分为当前棋子所在格子的数
  - 如果选择继续游戏，则棋子会以相同概率走到左侧或右侧相邻格子
  - 如果棋子走出了纸带，游戏立刻结束，得分为 0
- 对于 1, 2, ..., k, 求最优策略下，最大期望得分，要求绝对或相对误差不超过 10^{-3}
- $n \le 2 \times 10^5$, $v_i \le 10^9$, 1s, 256MB

解

- 容易列出动态规划：
- 记 $f(i)$ 为从第 i 格出发的最大期望得分
- $f(0) = f(n+1) = 0,\ f(k) = \max\left\{v_k, \frac{f(k-1)+f(k+1)}{2}\right\}$
- 令 $S = \left\{k | f(k) = \frac{f(k-1)+f(k+1)}{2}\right\}, T = \{0,\ 1,\ \dots,\ n+1\} \setminus S$
- 设 l, r 为 T 中相邻两项，则对于 $l < k < r$ 有 $f(k+1) - f(k) = f(k) - f(k-1)$
- 因此可得 $f(k) = \frac{(r-k)f(l)+(k-l)f(r)}{r-l}$
- 点集 $(i,\ v_i)$ 的上凸壳即为所求

抽卡

CCSP2021

- 抽 n 次卡，第 i 次有 p_i% 概率抽出SSR, q_i% 概率抽出SR
- SSR有三种，每次抽出SSR时等概率获得任意一种
- 每集齐三种卡就发动一次结算，若手上有 S 张SSR和 K 张SR则获得 S^2+K 分
- 同种算多张，结算后手中卡牌清零，得分累计
- 抽完后手牌如果没有结算将直接放弃，不结算得分
- 求总得分期望，乘以 300^n 输出
- 还会给出 q 次限时活动，一过性地修改某处 (p_x, q_x)
- $n, q \le 3 \times 10^5$, 1s, 512MB

解

- 期望是线性的，首先处理非线性的 S^2, 加入 S, 1 两项来将它表示成线性变换
- 其实得分 S^2+K 可以合起来计一维，另外除了单次的得分，还需要计算累计得分
- 这样得到一个 4 维向量，并按照手中的SSR种数 0, 1, 2 分成三个状态
- 计算各步的没抽到、抽到SR、抽到已有SSR、抽到新SSR的概率及对应转移
- 为了计算期望，倒过来列递推式
- 为了临时修改，注意这一递推构成 12 阶方阵，写成 $\vec{u}^{\mathrm{T}} M_1 M_2 \cdots M_n \vec{v}$
  - 提示：方阵可以用单位向量各递推一次生成，不需要手工写
- 处理前缀、后缀的矩阵积
- 时间复杂度 $O\left((n+q)S^2\right)$, 空间复杂度 $O(nS)$, 其中 S 表示状态数 12

这是一道非常平凡的计数题

CTT2018

- 给定一个长为 n 的序列 a
- 问存在多少对长为 n 的正整数序列 (b, d) 满足：$d_i | b_i, b_i | a_i, \prod_{i=1}^{n} b_i \geq \prod_{i=1}^{n} d_i^2$
- 模998244353
- $1 \leq n \leq 100, 1 \leq a_i \leq 10^9$

解

- 记 $c_i = \frac{b_i}{d_i}$, 则等价于 $c_i d_i | a_i$（看作大前提），$\prod_{i=1}^{n} c_i \geq \prod_{i=1}^{n} d_i$（看作约束）
- 观察到对称性，用全体方案数与 $\prod_{i=1}^{n} c_i = \prod_{i=1}^{n} d_i$ 的方案数取平均值即得答案
- 这样素因子之间都是独立的
- 每个素因子分别做，关于次数写一个递推

DistancePermutation

TC SRM762 Div1

- 一棵 n 个点的树，恰一个节点藏有宝藏
- 每次随机一个尚未询问的点，询问宝藏与该点的距离，直到确认宝藏位置
- 对于宝藏位于各点的情况，问期望询问次数之和
- 乘以 $n!$ 后模 10^9+7 输出
- $n \le 50$

解

- 首先假设一开始的次数是 n, 此前每个时刻若能唯一确定宝藏则扣除 1 次
- 枚举宝藏位置，再将第一次询问点作为根，现在知道宝藏的深度
- 以后各询问的答复等价于询问点与宝藏的LCA
- LCA一方面确认了宝藏所在子树，另一方面排除了宝藏不在的分支
- 现在转换到宝藏为根的视角，一组询问实际上给出了所有询问点的LCA及其深度
- 因此一个点作为LCA, 它的深度信息必须能够唯一确认根的位置
- 这要求LCA以上部分无歧义，此外还需要限定可能产生歧义的分支内有询问点
- 然后递推求出各时刻LCA恰好为这个点的概率