概率与期望

Part 0 概念

- 我想你们都会。
- 而且也不重要。

Part 1 高斯消元

- 三方的肯定都会。
- f[u]=sum (p[u][v]* (f[v]+w[u][v]))
- 概率：计算期望在这个节点上，停留多少步。
- f[u]=sum (p[v][u]* f[v]) + (u==S)

HNOI 2013 游走

- 有一个图，你要随机游走，从1到n。
- 求出每条边被经过的期望步数。
- n<=500

经典题

- 有n个点，你每次随机选择两个点连边。
- 求将整个图连通的期望。
- n<=50

SDOI 2012 走迷宫

- 有n个点m条边的有向图，求s到t的期望。
- n<=10000，每个SCC大小<=100

1.2 Band Matrix

- 看图
- O(n d^3)

例题

- WF 2014 H (大概就是会向上下左右转移的那种)
- 网格图的生成树
- CF 475 E (随机游走离原点超过R期望步数)

1.3 主元法(代入消元法)

- 小学生例题：n*m的网格，每个灯控制包括自己的周围5格，求最少步数把所有灯都熄灭。(n<=20)
- 核心思想：设比较少的变量，将其他所有变量都用这个表出。

例题

- TorusSailing (n*m的网格图，(x,y)等概率走向((x+1)%n,y),(x,(y+1)%m)，问到某点的期望步数)
- CF 475 E (随机游走离原点超过R期望步数)

HDU 4035 Maze

- 给一棵n 个点的树，从节点1 开始，每个时间点，若他在点 i，则他有ki 的概率回到1 号点，有ei 的概率结束行程，有 1-ki-ei 的概率在相邻点中等概率随机选一个走过去。
- 问期望行程时间。

Fox Rocks(FHC)

- 有一个n个点的图，a,b在0<=floor(b/4)-floor(a/4)<=1的条件下有边。
- m个操作，修改一个点出边的概率。
- 询问从x点开始不停的随机走经过y点的概率。
- n 50000,m 20000

Solution

- 按4个分块，块内搞出每个点开始转移到下个块中每个点的概率。
- 线段树维护a这个块i点出发进入b这个块是j的概率。

1.4 概率生成函数

- 无情的复读机器人，通过生成函数的方法简化运算。
- $$F(z) = \mathbb{E}(z^X) = \sum_{i=0}^{\infty} \Pr(X = i) z^i$$
- $$E(x) = F'(1) = \sum_{i=0}^{\infty} i \Pr(X = i)$$
- $$Var(X) = F''(1) + F'(1) - (F'(1))^2$$

歌唱王国

- 给定一个长度为L的序列A。然后每次掷一个标有1到m的公平骰子并将其上的数字加入到初始为空的序列B的末尾，如果序列B中已经出现了给定序列A，即A是B的子串，则停止，求序列B的期望长度。$L \le 10^5$。

Solution

令a_i表示$A[1,i]$是否是A的一个border，a_i为1表示是，0为不是。

令f_i为结束时随机序列长度为i的概率，其概率生成函数为$F(x)$。令辅助数列g_i为随机序列长度达到i且还未结束的概率，其普通生成函数为$G(x)$。

我们需要求的便是$F'(1)$。

通过分析可以得到如下等式：

$$F(x) + G(x) = 1 + G(x) \cdot x \quad (1)$$

$$G(x) \cdot \left(\frac{1}{m}x\right)^L = \sum_{i=1}^{L} a_i \cdot F(x) \cdot \left(\frac{1}{m}x\right)^{L-i} \quad (2)$$

- (1)的意义是在一个未结束的序列后加一个数字，可能结束也可能没结束，1是初始时随机序列为空的情况。

  (2)的意义是在一个未结束的序列后加上给定序列，则必定会结束，但是有可能在未加到L个数字时就已经结束了，这时一定要满足已经添加的序列是给定序列的一个border。

思考

- 优点：处理简洁，易扩展，比如可以改成求方差，或者说求某一项的值。
- 缺点：列方程比较不直观，需要一定的套路积累和练习(+一个/+一组)。

Part 2 期望线性性

- $E(X+Y)=E(X)+E(Y)$
- 可以分开计算期望。
- 转化成概率的计算。
- Warm up problems
- https://codeforces.com/blog/entry/62690
- https://codeforces.com/blog/entry/62792

Problem Set 1

1. **Aces** We choose 10 cards at random from a standard deck of 52 cards. Find EV of the number of aces.
2. **Inflation** The price of a tv is 1000$. Each of the next N days, the prices goes up by 5$ or 10$ (each with p-bility 50%). Find EV of the final price.
   Bonus: What if it goes up by 1% or 2% each day?
3. **Max of two** You roll a 6-sided die twice. Find EV of the bigger of two scores.
4. **Max of N** You roll a 6-sided die N times. Find EV of the biggest score.
   A few possible solutions.
5. **Birthdays** You teach informatics in a class with 20 students. When someone has a birthday, you must let the whole class to play games instead of learning algorithms and using Excel. Maybe some students have birthday on the same day, so there would be fewer than 20 wasted days during a year? Find EV of the number of days when at least one student has birthday.
   Bonus/fact: Birthday paradox.
6. **First heads** Find EV of the number of coin tosses until you get heads.
   How to check your answer with a program?
7. **Two heads** Find EV of the number of coin tosses until you get heads two times in total.
8. **Two heads in a row** Find EV of the number of coin tosses until you get heads two times in a row.
9. **Volleyball** 12 teams, including Poland, play in the volleyball tournament. Teams are divided into 4 groups, each with 3 teams. In each group, every team plays against every other team, and then two best teams advance to the elimination stage. In case of a perfect tie, two random teams advance. The elimination stage has quarterfinals, semifinals and the final match. In every match, a random of two teams wins (50% each).
   Find p-bility that Poland will win the whole tournament. Find EV of the number of matches won by Poland. Find EV of the number of matches won by Poland, assuming that they won the whole tournament (in other words, find EV of the number of matches won by the winner of the whole tournament).

Problem Set 2

1. **Hills** Given a sequence of length N ($N \le 10^5$), count triples of indices $i < j < k$ that $a_i < a_j > a_k$.
   Bonus: Count zig-zags of length 10, i.e. $i_1 < i_2 < ... < i_{10}$ that $a[i_1] < a[i_2] > a[i_3] < a[i_4] > ... < a[i_{10}]$.
2. **Paths in tree** Given a tree of length N ($N \le 10^5$), find the sum of lengths of all paths. (Solve this without not-trivial dp or centroids.)
   Bonus: Find the sum over squares of lengths of paths.
3. **Sum over subsets** There are N competitions, the i-th with prize a_i. You're quite good and you will win each competition with p-bility 50%. Find EV of the total won prizes.
   Equivalently: Find the sum over total prize over all 2^N possibilities, and print answer modulo $10^9 + 7$. (This answer divided by 2^N gives the answer from the first version.).
4. **Math encoder (Code Jam Kickstart 2017 round B)** You are given a sequence of N numbers ($N \le 2000$ or $N \le 10^5$). We're to choose one of $2^N - 1$ non-empty subsets, uniformly at random. Find EV of the difference between the maximum and minimum element in the subset.
   Equivalently: Find the sum over the difference over all subsets, and print the answer modulo $10^9 + 7$.
5. **Imbalanced array (CF Educational Round 23)** Same as the previous problem, but choose a random of $N \cdot (N+1) / 2$ intervals.
   Bonus: Do it in $O(N)$ if the input contains a permutation of numbers 1 through N.
6. **Randomizer (CF 313 D)** You're given a convex polygon with N vertices ($N \le 2000$ or $N \le 10^5$). We choose a random subset of vertices, what gives us a new (small) convex polygon. Find EV of the perimeter of the new polygon.
   Well, the CF problem was a bit harder, with computing area and using Pick's theorem.
7. **Random CH** You're given N points ($N \le 200$ or $N \le 2000$), no three are collinear. Each point disappears with p-bility 50%. Find EV of the size of the convex hull of remaining points.
   (The size of CH is the number of its vertices.).
8. **Eating ends** You're given a sequence of length N ($N \le 2000$ or $N \le 10^5$). $N - 1$ times we will remove the first or the last element, each with p-bility 50%. Find EV of the last remaining number.
   Well, the implementation is hard because of precision issues.
9. **Sum-length** Given a sequence of length N ($N \le 10^5$), find the sum over $sum \cdot len^3$ over all intervals. Print the answer modulo $10^9 + 7$.
   There are a few possible solutions, including one that we will discuss in the next part.

•

CF 280C

- 给出一棵含n个白点的有根树，每次随机选择一个还没有被染黑的节点吗，将这个节点和这个节点子树中的所有点染黑。
- 问期望操作多少次后所有点都被染黑。

经典题

- 给你一棵树，随机游走，求s到t的期望步数。
- q组询问。

CF 1153 F

- 给一个长度为l的线段，随机选了n条子线段，求被覆盖了k次的期望长度。
- n<=2000

CF 235 D

- 给你一个n个点，n条边的连通图。每次随机选一个点删除，然后递归所有连通分量，代价是每次连通块大小。
- 问代价期望。

CF 457 D

- 你有一个n*n的棋盘，你先往这个棋盘里面随机填了1…m中n*n个不同的数，然后你随机选了1…m中的k个数。
- 你的分数是2^(被完全选中的行和列总个数)，求期望。

random

- 有个长度为n的排列，每次随机选择一对相邻逆序对，然后交换，代价是下标，问期望代价。

CTSC 没头脑与不高兴

- 随机一个排列，指定若干个位置会排好序。问逆序对个数的期望。
- 每次会选择覆盖一个区间作为排好序的位置或者取消一段区间。

Projecteuler 584

- 每个人生日在1到365天内独立随机，依次加入。
- 问期望加入多少个人，存在4个人他们的生日相差不超过7天。

Projecteuler 522

- n个点，每个点随机连向和它不同的点。
- 问最少改几个点出边变成环的期望。
- n 1e7

2.2 期望的平方/方差

Problem Set 3

1. **Inflation 2** The price of a tv is 1000\$. Each of the next N days, the prices goes up by 1% or 2% (each with p-bility 50%). Find EV of the final price.

2. **Square of wins** You bought N ($N \le 10^5$) lottery tickets. The i-th of them is winning with p-bility p_i. The event are independent (important!). Find EV of the square of the number of winning tickets.

3. **Cube of wins** Same but find EV of the 3-rd or 4-th power.

4. **Bulbs** There are N ($N \le 50$) light bulbs and M ($M \le 50$) switches. For every switch, we know a set of bulbs such that its state is flipped (on to off, off to on) when the switch is clicked. All bulbs are off, and then we click each switch with p-bility 50%. Find EV of the cube of the number of bulbs that are on.

5. **Sum-length** Given a sequence of length N ($N \le 10^5$), find the sum over $sum \cdot len^3$ over all intervals. Print the answer modulo $10^9 + 7$. There are a few possible solutions.

6. **Small power of subtree** You're given a tree of size N ($N \le 10^5$) and an integer k ($k \le 10$). Find the sum of $size^k$ over all "subtrees", i.e. connected subgraphs. Print the answer modulo $10^9 + 7$.

7. **DoubleLive (Topcoder SRM 727)** Your army consists of $B + H$ soldiers: B bears and H humans ($B, H \le 2000$). A bear has two hit points, a human has one hit point.
The enemy archer has T arrows ($T \le 2 \cdot B + H$). Every next arrow will hit a random alive soldier in your army, taking one hit point. When someone has zero hit points, he dies.
The strength of your army is defined as $bears \cdot humans \cdot all$, e.g. $3 \cdot 7 \cdot 10 = 210$ for an army of 3 bears and 7 humans. It doesn't matter if some bears have only one hit point.
Find the EV of the strength of your army after the enemy archer uses all his arrows.

CF 1187

- 令b[i]为1到a[i]里面的随机数，X为b不同段的段数，求E[X^2]
- n<=10^5

Fibonacci's Nightmare

- $a_0=1$
- $a_n=a_i+a_j$,其中i,j在[0,n-1]之间随机选择。
- 求第n项的方差。
- $n<=1e6$

CF 1236 F

- 给一个仙人掌，每个点1/2删除。
- 问连通块个数的方差。

火车题

- 无向图的权值个数为连通块个数的m次。
- 求n个点有标号连通块个数。
- n<=30000, m<=15

SRM 686 CyclicNumber

- 求n个点的置换循环个数m次方的和。
- n<=100000, m<=500

HDU 5404

- 一个1到n的随机排列，指定其中某些位置排序。
- 将其随机切成k段，问每段的逆序对的乘积的期望。
- $n<=100$

2.3 min-max容斥

- 我真的一点都不想讲这一part

经典题

- 每次产生S的概率为p(S)，问并全集的期望步数。
- |U|<=20

51nod 1355

- 求lcm(fib[a1],fib[a2], …, fib[an])
- n<=50000, an<=10^6

RandomPaintingOnABoard

- n*m的棋盘，每个位置有pi,j。
- (i,j)被选中的概率为pi,j/S。
- 至少几轮后每一行一列至少一个被选中。
- nm<=150,0<=pi,j<=9

PKUSC 随机游走

- 给你一棵树n，给你Q个询问S，问从x点出发随机游走走遍S中所有点的期望。
- n<=18

例题2. 抛骰子游戏

给定n个长度分别为L_i的序列A_i。再给出一个标有1到m的骰子，其中抛出i的概率为P_i。然后每次抛一次骰子将骰子上的数字加入到初始为空的序列B末尾，如果给定的n个序列都是B的子串，则停止，求序列B的期望长度。保证$n \leq 15$，$L \leq 20000$，$m \leq 10^5$。

Solution

- Minmax容斥之后变成每个子集碰到的概率，去除一下后缀的case。
- 用前面类似方法解决。

推广

- k-th max(s) = sum min(T) (-1)^|T| C(|T|-1,k-1)

2.4 真题选讲

- 五个期望+斗地主

PKUWC2018 随机算法

- 我们知道，求任意图的最大独立集是一类NP完全问题，目前还没有准确的多项式算法，但是有许多多项式复杂度的近似算法。

  例如，小 C 常用的一种算法是：

  1. 对于一个 n 个点的无向图，先等概率随机一个 $1 \ldots n$ 的排列 $p[1 \ldots n]$。
  2. 维护答案集合 S，一开始 S 为空集，之后按照 $i = 1 \ldots n$ 的顺序，检查 $\{p[i]\} \cup S$ 是否是一个独立集，如果是的话就令 $S = \{p[i]\} \cup S$。
  3. 最后得到一个独立集 S 作为答案。

  小 C 现在想知道，对于给定的一张图，这个算法的正确率，输出答案对 998244353 取模
- n<=20

Solution

- 改成随机一个排列，然后每次加一个点。

PKUWC2018 猎人杀

- 猎人杀是一款风靡一时的游戏“狼人杀”的民间版本，他的规则是这样的：

  一开始有 n 个猎人，第 i 个猎人有仇恨度 w_i ，每个猎人只有一个固定的技能：死亡后必须开一枪，且被射中的人也会死亡。

  然而向谁开枪也是有讲究的，假设当前还活着的猎人有 $[i_1 \ldots i_m]$，那么有 $\frac{w_{i_k}}{\sum_{j=1}^{m} w_{i_j}}$ 的概率是向猎人 i_k 开枪。

  一开始第一枪由你打响，目标的选择方法和猎人一样（即有 $\frac{w_i}{\sum_{j=1}^{n} w_j}$ 的概率射中第 i 个猎人）。由于开枪导致的连锁反应，所有猎人最终都会死亡，现在 1 号猎人想知道它是最后一个死的的概率。

  答案对 998244353 取模。

- sum w[i]<=100000

Solution

- 改成每次按固定概率向每个人开枪，然后求第一人最后一个死。
- 容斥第一个人死的时候哪些人没有死。

PKUSC 2018 最大前缀和

- 小 C 是一个算法竞赛爱好者，有一天小 C 遇到了一个非常难的问题：求一个序列的最大子段和。

  但是小 C 并不会做这个题，于是小 C 决定把序列随机打乱，然后取序列的最大前缀和作为答案。

  小 C 是一个非常有自知之明的人，他知道自己的算法完全不对，所以并不关心正确率，他只关心求出的解的期望值，现在请你帮他解决这个问题，由于答案可能非常复杂，所以你只需要输出答案乘上 $n!$ 后对 998244353 取模的值，显然这是个整数。

  注：最大前缀和的定义：$\forall i \in [1,n]$，$\sum_{j=1}^{i} a_j$的最大值。

- n<=20

Solution

- 考虑最大前缀和的充要条件，然后正着反着分开做一下。

PKUSC 2018 PKUSC

- 九条可怜是一个爱玩游戏的女孩子。

  最近她在玩一个无双割草类的游戏，平面上有 n 个敌人，每一个敌人的坐标为 x_i, y_i。可怜有一个技能是在平面上画一个 m 个点的简单多边形，并消灭所有**严格**在多边形内部的敌人。

  不难发现如果想要快速的消灭敌人的话，只要画一个足够大的简单多边形就行了。但是这样的游戏性就太差了。于是可怜打算为游戏增加一定的随机性。

  可怜在平面上随便画了一个 m 个点的简单多边形 (a_i, b_i)。接下来可怜打算按照 $[-\pi, \pi)$ 上的均匀分布随机选取数字 α（可以理解为等概率选取），并把这个简单多边形绕原点逆时针旋转 α 的角度（弧度制）。

  现在可怜给你了每一个点的坐标，多边形的坐标，你的任务是帮助可怜计算在随机旋转后她期望可以消灭多少个敌人。

- n <= 200, m<= 500

2.5 一些连续的问题

- 小结论：
  - 在[0…x]之间随机n个数字第i(1<=i<=n)小期望是ix/(n+1)
  - 把[0…x]分成n段，最短段的期望是x/n^2，次短是x/n(1/(n-1)+1/n)…

方法1：对称性

- 给一个长度为l的线段，随机选了n条子线段，求被覆盖了k次的期望长度。
- 考虑将所有的端点排个序，那么每一段的期望长度是1/(2n+1)，然后计算第i段被覆盖了至少k次的概率。

CF 303 E

- 有n个数字在[li,ri]之间随机，求第i个数是第j大的概率。
- n<=80

Solution

- 首先离散化区间，然后枚举第i个数字在哪个区间，然后求出剩下多少数字在左边，多少数字在右边。
- 对于在中间的数字，概率相同。

新年的五维几何

- 设 $x_1, x_2, \cdots, x_n$ 是 n 个实数变量，其中第 i 个变量 x_i 在区间 $[l_i, r_i]$ 内均匀随机生成，所有 l_i 和 r_i 均为给定的整数且 $l_i \le r_i$（约定 $l_i = r_i$ 时，$[l_i, r_i]$ 表示单元素集合 $\{l_i\}$）。
  给定 $n \times n$ 的整数矩阵，矩阵的每个元素代表一个约束，其中第 i 行第 j 列的元素 $a_{i,j}$ 代表约束
  $$x_i - x_j \ge a_{i,j}$$
  求这 $n \times n$ 个约束同时被满足的概率。
- n<=5, 0<=li<=ri<=10, -10<=a[i][j]<=10，整数

Solution

- 将每个数写成整数部分和实数部分的形式，实数部分对称。

AGC 020 Arcs on a Circle

- 一个周长为C的圆上，随机放了N条长度为li的弧，问全都覆盖的概率。
- N<=6, C<=50

Solution

- 枚举小数部分的顺序，变成离散的问题。

方法2：积分

- n辆车，第i辆车到达时间在[0,ai]之间均匀分布。
- 问期望等待时间。
- n<=2000

Random Spanning Tree

- 给一个图，每条边在[0,1]之间随机，求MST最大边的期望。
- 求MST的期望。

CF 913 H

- s[0]=0, s[i]=s[i-1]+[0…1]的随机数，求s[i]<=x[i]都成立的概率。
- n<=30

方法3：杂

- 你有一个大小为1的披萨，你在上面面独立随机地切了了n刀。
- 然后你想选一段使得与1/3最接近。
- 问与1/3的差的绝对值的期望是多少。

Solution

- 一些简单的分析，就是讲披萨按第一刀切得地方分成[0,1/3), [1/3,2/3), [2/3,1)这么三段。
- 然后可以理解成你在[0,1/3)里面随机放了n-1个点，颜色是随机RGB中的一种，答案就是两个颜色不同点的最近距离。
- 然后答案为最小的距离*答案为最小的概率+第二小的距离*答案为第二小的概率+…