长郡中学春季交流

概率与期望

郭晓旭 (@ftiasch)

2018 年 3 月 29 日

我

郭晓旭 @ftiasch a.k.a. 叉姐

Dice (HDU 4652)

m 面骰子，求

1. 出现 n 个相同的停止;
2. 出现 n 个不同的停止

的期望次数。

$(n, m \leq 10^6)$

设 f_x, g_x 表示出现 x 个相同、不同时的期望，则

$$f_x = \frac{1}{m} f_{x+1} + \frac{m-1}{m} f_1 + 1$$

$$g_x = \frac{m-x}{m} g_{x+1} + \frac{1}{m} \sum_{i=1}^{x} g_i + 1$$

设 f_x, g_x 表示出现 x 个相同、不同时的期望，则

$$f_x = \frac{1}{m} f_{x+1} + \frac{m-1}{m} f_1 + 1$$

$$g_x = \frac{m-x}{m} g_{x+1} + \frac{1}{m} \sum_{i=1}^{x} g_i + 1$$

1. 差分
2. 设 f_0，g_0 为主元

Dice' (Folklore)

6 面骰子，

- 出现 $1, 2, 3$ 计数器清零，
- 出现 $4, 5, 6$ 计数器 $+1$，
- 出现 6 停止。

求计数器的期望。

设 f_x 是计数器为 x 时的期望，则

$$f_x = \frac{x+1}{6} + \frac{f_{x+1}}{3} + \frac{f_0}{2}$$

设 f_x 是计数器为 x 时的期望，则

$$f_x = \frac{x+1}{6} + \frac{f_{x+1}}{3} + \frac{f_0}{2}$$

$$f_0 = \sum_{k=0}^{\infty} \frac{1}{3^k} \left(\frac{k+1}{6} + \frac{f_0}{2} \right)$$

New Year and Arbitrary Arrangement (Good Bye 2017)

- 字符串 S 初始为空，
- 以 p_a 概率在后面接 `a`，
- 以 p_b 概率接 `b`，
- 当有 k 个子序列 `ab` 时停止。

求期望长度。

$(k \leq 1000)$

Card Collector (HDU 4336)

n 种卡片，第 i 种出现的概率是 p_i，问收集所有卡片的期望次数。

$(n \leq 20)$

Card Collector (HDU 4336)

n 种卡片，第 i 种出现的概率是 p_i，问收集所有卡片的期望次数。

$(n \leq 20)$

1. The hard way
2. Maximum-minimum identity

Gambler Ruin (Folklore)

初始，Bobo 在（数轴）a 点。每次

- 以 p 的概率 $+1$,
- 以 $(1-p)$ 的概率 -1.

求到达 0 前到达 $(a+b)$ 点的概率。

设 f_x 是从点 x 出发所求的概率，则 $f_0 = 0, f_{a+b} = 1$.

$$f_x = p \cdot f_{x+1} + (1-p) \cdot f_{x-1}$$

$$\implies f_{x+1} - f_x = \frac{1-p}{p} \cdot (f_x - f_{x-1})$$

$$\implies f_a = \frac{f_a - f_0}{f_{a+b} - f_0} = \frac{1 - \left(\frac{1-p}{p}\right)^a}{1 - \left(\frac{1-p}{p}\right)^{a+b}}$$

注意 $p = \frac{1}{2}$ 的情况

CatsOnTheCircle (SRM 667)

n 个 Bobo 形成环状。初始，球在 0 号 Bobo 处。每次

- 以 $\frac{p}{10^9}$ 概率顺时针传递，
- 以 $1-\frac{p}{10^9}$ 概率逆时针传递。
- 最后获得球的 Bobo 胜利。

求 k 号 Bobo 获胜的概率。

$(n, p \leq 10^9)$

夹克赌坊 (51nod 1653)

给定 a, b, c, p, q,

- 当 $x \leq a$ 时，移动到 $(x+1), (x-1)$ 的概率分别是 $p, (1-p)$.
- 当 $x > a$ 时，移动到 $(x+1), (x-1)$ 的概率分别是 $q, (1-q)$.

求从 $x = 0$ 开始，在到 $(-b)$ 之前到 c 的概率。

Wandering Robots (HDOJ 6229)

$n \times n$ 的棋盘上有 k 个障碍。

初始，Bobo 在 $(1, 1)$，每秒等概率地留在原地或移向相邻格子。

设无穷久后 Bobo 在 (x, y) 的概率是 $\pi(x, y)$. 求

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \pi(x, y)[x + y > n]$$

$(n \leq 10^4, k \leq 10^3)$

考虑一般图的情况，设 $\pi(u)$ 是无穷久后在点 u 的概率，则

$$\pi(u) = \frac{\deg(u) + 1}{|V| + 2|E|}$$

具体实现?

Pachinko (World Finals 2014)

$h \times w$ 的棋盘，格子有空地、障碍、目标 3 种类型。

初始，Bobo 等概率地出现在首行的空地格子，每秒向上、下、左、右移动的概率分别是 u, d, l, r. 如果对应方向越界或是障碍，则不动。

到达目标格子后停止，求到达每个目标的概率。

$(w \leq 20, h \leq 10^4)$

怎么算

Smart

设 M 是转移矩阵，π_0 是初始概率分布，要求

$$\pi = \sum_{k=0}^{\infty} M^k \pi_0 = (I - M)^{-1} \pi_0$$

等价于 $(I - M)\pi = \pi_0$

Stupid

假设所有目标的集合是 T，考虑某个目标 $t \in T$.
如果要求到达 t 的概率，自然地设 x_u 表示点 u 出发，到达目标 t 的概率。
那么 $x_t = p_t = 1$, $x_{t'} = p_{t'} = 0$ $(\forall t' \neq t)$.
同时 $\mathbf{x}^T(I - M) = \mathbf{p}$，即 $\mathbf{x}^T = \mathbf{p}(I - M)^{-1}$. 最终答案是 $\mathbf{x}^T\pi_0$.
如果考虑函数 $f(\mathbf{p}) = \mathbf{p}(I - M)^{-1}\pi_0$，那么

$$\frac{\partial f}{\partial p_t} = \left((I - M)^{-1}\pi_0\right)_t$$

怎么解

- Adhoc
- Band matrices

Rainbow Balls (CF 432)

n 种颜色的球，第 i 种有 a_i 个，每次随机选出两个球，把第一个染成第二个的颜色放回，问所有球同色次数的期望。

$n \le 2500, a_i \le 10^5$

设 $p_{i,k}$ 是第 k 次后都是颜色 i 的概率，答案是

$$\sum_{i=1}^{n}\sum_{k=1}^{\infty} k \cdot p_{i,k}$$

单独考虑第 i 种球。设 $S = a_1 + a_2 + \cdots + a_n$.

设 f_x 是当前有 x 个球的期望，要求 x 在到达 0 前到达 S. 则

$$f_x = pf_{x-1} + pf_{x+1} + (1-2p)f_x + \frac{x}{S}$$

其中 $p = \frac{x(S-x)}{S(S-1)}$.

注意最后的常数项是 $\frac{x}{S}$ 而不是 1，因为只有 $\frac{x}{S}$ 的后续状态可以以颜色 i 结束（考虑 x 的随机游走）.

答案就是 $\sum_{i=1}^{n} f_{a_i}$. 具体的值可以仿照前面的方法求解。

k-th point (HDOJ 4653)

在 p 维单位球中随机 n 个点，求到球心第 k 远点的距离的期望。

$(n, d \leq 10^4)$

根据对称性只考虑 $r_1 \geq r_2 \geq \cdots \geq r_k \geq \max\{r_{k+1}, r_{k+2}, \ldots, r_n\}$. 最后把答案乘上 $\frac{n!}{(n-k)!}$.

则

$$\mathbb{E}[r_1] = \int_0^1 r \cdot r^{p(n-1)} \mathrm{d} r^p = \frac{p}{pn+1}$$

设 $E(R, n, k)$ 表示在半径为 R 的球中的答案，容易想象有 $E(R, n, k) = E(1, n, k)R$. 从而

$$\begin{aligned} E(1, n, k) = &\quad \int_0^1 E(r, n-1, k-1) r^{p(n-1)} \mathrm{d} r^p \\ = &\quad E(1, n-1, k-1) E(1, n, 1) \end{aligned}$$

最终答案是 $\prod_{i=n-k+1}^{n} \frac{pi}{pi+1}$

Second Maximum (GP of Peterhof)

n 个连续随机变量 $x_1, x_2, \ldots, x_n$ 其中 x_i 在 $[l_i, r_i]$ 种均匀随机。

求第二大值的期望。

$(n \leq 500)$

Solution 1

暴力积分，枚举最大值是 i，次大值是 j，即要求

$$\int_{l_i}^{r_i} x \Pr[x_j > x] \prod_{k \neq i,j} \Pr[x_k < x] \mathrm{d}x$$

上式是分段多项式，可以 dp 优化，需要注意精度

Solution 1

暴力积分，枚举最大值是 i，次大值是 j，即要求

$$\int_{l_i}^{r_i} x\Pr[x_j > x] \prod_{k \neq i,j} \Pr[x_k < x]\mathrm{d}x$$

上式是分段多项式，可以 dp 优化，需要注意精度

Solution 2

离散化，枚举次大值所在区间，设 $f_{i,j,k}$ 表示考虑了前 i 个变量，有 j 在次大值区间，k 个大于次大值区间。

如果有 $k = 0$，那么次大值在区间的 $\frac{j-1}{j+1}$ 分点处；如果 $k = 1$，那么次大值在 $\frac{j}{j+1}$ 分点处。

新年的五维几何 (UOJ 352)

n 个连续随机变量 $x_1, x_2, \ldots, x_n$ 其中 x_i 在 $[l_i, r_i]$ 种均匀随机。

问 $x_i - x_j \geq A_{i,j}$ 全部满足的概率。

$n \leq 5, 0 \leq l_i \leq r_i \leq 10,\ l_i, r_i$ 是整数

先 $O(\prod_{i=1}^{n}(r_i - l_i))$ 枚举整数部分，分数部分只有大小关系

Arcs on a Circle (AGC 020)

长度为 m 的圆环上，随机放长度为 $l_1, l_2, \ldots, l_n$ 的线段，问盖住的概率。

$m \leq 50, n \leq 6$

先 $O((n-1)!)$ 枚举分数部分，整数部分可以 $O(2^n(nm)^2)$ dp.

Dont Exceed (CF Hello 2018)

n 个 $[0,1]$ 上的连续均匀随机变量 $x_1, x_2, \ldots, x_n$，给出 $a_1, a_2, \ldots, a_n$，问 $\sum_{i=1}^{k} x_i \leq a_k$ 同时成立的概率

$(n \leq 30)$

设 $P_k(x), C_k(x)$ 是 $\sum_{i=1}^k x_i$ 的 pdf, cdf. 则

$$C_k(x) = \int_0^x P_k(t)\mathrm{d}t = \int_0^x (C_{k-1}(t) - C_{k-1}(t-1))\mathrm{d}t.$$

$C_k(x)$ 是个 $O(n^2)$ 段的分段函数，每段是多项式。

$O(n^5)$ to $O(n^4)$

Monkey at the Keyboard (Timus 1677)

Bobo 等概率地输入前 n 个字母，求字符串 S 出现的期望时间。

$n \le 26, |S| \le 30000$

Generator (HDOJ 5850)

给 n 个字符串 $s_1, s_2, \ldots, s_n$，以 p_i 概率输入第 i 个字符，求 n 个字符串都出现的期望时间。

$n \leq 15, |\Sigma| \leq 100, |s_1| + |s_2| + \cdots + |s_n| \leq 10^5$

Recurrent Event

设 $P(x)=\sum_{k=1}^{\infty} p_k x^k$ 满足

- $x_k \geq 0$,
- $P(1)=1$,
- 所有满足 $p_k>0$ 的 k 的最大公约数是 1.

则对于 $U(x)=\frac{1}{1-P(x)}=\sum_{k=1}^{\infty} u_k x^k$ 有

$$\lim_{n\to\infty} u_k=\frac{1}{\sum_{k=1}^{\infty} k\cdot p_k}$$

Dice Revisited (for Recurrent Event)

m 面骰子，求

1. 出现 n 个相同的停止;
2. 出现 n 个不同的停止

的期望次数。

$(n, m \leq 10^6)$

Tarot Sham Boast (World Finals 2017)

给出 n 个长度相同的字符串 $s_1, s_2, \ldots, s_n$，按照它们在长度为 m 的随机串中出现的概率排序。

$n \le 10, |s_1| = |s_2| = \cdots = |s_n| \le 10^5, m \le 10^6, |\Sigma| = 3$

Birthday (AIM Tech Round)

盒子有 n 个球，Bobo 蒙眼摸球，摸出第 i 个球的概率是 p_i.

每轮 Bobo 摸球，猜摸出的球的编号，没有任何反馈，直到所有球都被猜对至少一次结束。

问轮数期望的最小值。

$(n \le 100, 0.01 \le p_i \le 1)$

何谓决策?

定义决策函数 $f\colon K \to A$，其中 K 是知识 (knowledge)，A 是行动 (action).

题目中，$K = \{(\text{the \# of gusses})\}$, $A = \{\text{guess the } i\text{th} : i \in [n]\}$.

固定决策 $f(1), f(2), \ldots$，计算期望

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \Pr[X \geq k] = \sum_{k=1}^{\infty} \left(1 - \prod_{i=1}^{n} \left(1 - (1 - p_i)^{c_i(k)}\right)\right)$$

其中 $c_i(k) = \sum_{j=1}^{k} [f(j) = i]$.

每轮贪心选使乘积增加最多。

Hey, Better Bettor (World Finals 2013)

Bobo 有无限多的本金，有 $p\%$ 的概率赢 1 元，有 $(100-p)\%$ 的概率输 1 元。离开时如果亏 l 元，赌场会返利 $(l \cdot x\%)$ 元。

问收益期望的最大值。

$(p < 50, x < 100)$

考虑决策函数 $f: K \to A$, 显然 $A = \{\text{exit}, \text{not exit}\}$.

而过去路径不影响未来决策，所以 $K = \{\text{currentwealth}\}$.

只有

$$L = \max\{x : f(x) = \text{exit} \land x < 0\}$$
$$W = \min\{x : f(x) = \text{exit} \land x > 0\}$$

有意义

Q: 如果 $L = -\infty$?

考虑 Gambler ruin，赢的概率是

$$\frac{1-\left(\frac{p}{1-p}\right)^{W}}{1-\left(\frac{p}{1-p}\right)^{W-L}}$$

因为 $\frac{p}{1-p} < 1$，所以当 $L \to -\infty$ 时极限不是 1.

考虑 Gambler ruin，赢的概率是

$$\frac{1-\left(\frac{p}{1-p}\right)^{W}}{1-\left(\frac{p}{1-p}\right)^{W-L}}$$

因为 $\frac{p}{1-p} < 1$，所以当 $L \to -\infty$ 时极限不是 1.

利用 Gambler ruin 可以算出在 W 和 $-L$ 退出概率，暴力枚举 W, L 得到期望的最大值

Compromise or persist (PE 503)

初始时，$S = [n]$ 而 $T = \{\}$. 每轮 Bobo

- 随机选择 $x \in S$,
- 得知 x 在 T 中的排名，
- 选择退出，则得分 x,
- 选择继续，则把 x 从 S 移动到 T.

求期望得分的最小值。

$(n \leq 10^6)$

决策和历史排名无关

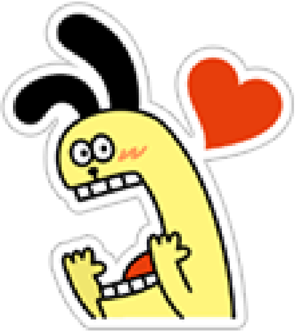
假设第 k 轮得知排名是 r，则 x 的期望是

$$
\begin{aligned}
& \frac{\sum_{i=1}^{n} i\binom{i-1}{r-1}\binom{n-i}{k-r}}{\binom{n}{k}} \\
= \quad & \frac{r\sum_{i=1}^{n} \binom{i}{r}\binom{n-i}{k-r}}{\binom{n}{k}} \\
= \quad & \frac{r\binom{n+1}{k+1}}{\binom{n}{k}} \\
= \quad & \frac{r(n+1)}{k+1}
\end{aligned}
$$

必定存在 $R(k)$ 使得 $r \leq R(k)$ 退出，否则继续，DP 即可。

Bayes' theorem

$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)} = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} = \frac{\Pr(B|A)\Pr(A)}{\sum_{A\in\mathcal{A}}\Pr(B|A)\Pr(A)}$$



Medical Tests (Folklore)

假设

- 某病发生的概率是 0.01%,
- 某病误诊的概率是 1%

如果某人诊断患病，求某人患病的概率。

设 S 表示得病，T 表示诊断。那么

$$\begin{aligned}
\Pr(S|T) &= \frac{\Pr(T|S)\Pr(S)}{\Pr(T|S)\Pr(S)+\Pr(T|\bar{S})\Pr(\bar{S})} \\
&= \frac{99\times 1}{99\times 1+1\times 9999} \\
&\approx 0.98\%
\end{aligned}$$

Good Luck (GCJ 2013 Round 1A)

Bobo

1. 在 $[2, 8]$ 中随机 $n = 12$ 个整数 $A = (a_1, a_2, \dots, a_n)$,
2. 随机 $k = 12$ 个 A 的子集.

给出 k 个子集的乘积 $P = (p_1, p_2, \dots, p_k)$，猜 A.

要求 8000 组中忽略顺序正确至少 1120 组.

依照贝叶斯定理

$$\Pr(A|P) = \frac{\Pr(A)\prod_{i=1}^{n}\Pr(p_i|A)}{\Pr(P)}$$

忽略顺序，A 只有 $\binom{18}{6} = 18564$ 种

$$\Pr(A) = \frac{n!}{7^n \prod_{i=2}^{8} c_i!}$$

其中 c_i 表示 A 中 i 的重数.

$\Pr(p_i|A)$ 可以 $O(2^n)$ 预处理.

$\Pr(P)$ 是定值，输出 $\Pr(A)\prod_{i=1}^{n}\Pr(p_i|A)$ 最大的 A 即可.

Range Estimate (Petr Contest 11)

Bobo

1. 在 $[0, 10^3]$ 中随机实数 r,
2. 在 $[5, 10]$ 中随机整数 n,
3. 在 $[0, r]$ 中随机 n 个实数 $x_1, x_2, \dots, x_n$.

给出 $x_1, x_2, \dots, x_n$，猜 r.

要求 10^5 组数据，绝对误差之和不超过 44.

显然 $R_0 = \max\{x_1, x_2, \dots, x_n\} \le r \le 10^3 = R_1$

依照贝叶斯定理

$$\Pr(r|x_1, x_2, \dots, x_n) = \frac{\Pr(r)\prod_{i=1}^{n}\Pr(x_i|r)}{\Pr(x_1, x_2, \dots, x_n)}$$

其中

- $\Pr(x_i|r) = \frac{1}{r}$
- $\Pr(r)$, $\Pr(x_1, x_2, \dots, x_n)$ 是常数

代入得到

$$\Pr(r|x_1, x_2, \dots, x_n) = \frac{C}{r^n}$$

(C 是常数)

$$\Pr(r|x_1, x_2, \ldots, x_n) = \frac{C}{r^n}$$

最佳猜测 r^* 最小化

$$\int_{R_0}^{R_1} |r - r^*| \Pr(r|x_1, x_2, \ldots, x_n)\mathrm{d}r$$

即 $r^* = \mathrm{median}(r)$

正式地，设

$$F(r_1) = \int_{R_0}^{r_1} \Pr(r|x_1, x_2, \ldots, x_n)\mathrm{d}r = C' \left(\frac{1}{R_0^{n-1}} - \frac{1}{r_1^{n-1}} \right)$$

(C' 是常数)

则 $F(r^*) = \frac{1}{2}$. 可以先利用 $F(R_1) = 1$ 解出 C'，再算出 r^*.

T-shirt (Croc Champ 2012)

n 个人，m 种大小的衣服，第 i 个人是第 j 种大小的概率是 $p_{i,j}$.

选择 n 件衣服，最大化合适衣服的期望。

$(n \leq 3000, m \leq 300)$

假设每种衣服带了 $c_1, c_2, \ldots, c_m$ 件，可能有 $X_1, X_2, \ldots, X_m$ 个人，则要最大化

$$\mathbb{E}[\min\{X_1, c_1\} + \min\{X_2, c_2\} + \cdots + \min\{X_m, c_m\}]$$

根据期望的线性性，只需要分别求出 $X_1, X_2, \ldots, X_m$ 的分布

设 $f_{k,i,j}$ 表示前 i 个人中，适合第 k 种衣服的人有 j 个的概率

复杂度是 $O(n^2 m)$

假设每种衣服带了 $c_1, c_2, \ldots, c_m$ 件，可能有 $X_1, X_2, \ldots, X_m$ 个人，则要最大化

$$\mathbb{E}[\min\{X_1, c_1\} + \min\{X_2, c_2\} + \cdots + \min\{X_m, c_m\}]$$

根据期望的线性性，只需要分别求出 $X_1, X_2, \ldots, X_m$ 的分布

设 $f_{k,i,j}$ 表示前 i 个人中，适合第 k 种衣服的人有 j 个的概率

复杂度是 $O(n^2 m)$

注意到 $\mathbb{E}[\min\{X, c\}] = \sum_{j=1}^{c} \Pr[X \geq j]$

因为 $\Pr[X \geq j]$ 随着 j 递减，贪心地选取，只需要计算 $f_{k,i,j}$ 中的 $O(n)$ 列

Random MST (WJMZBMR)

n 个点的无向图，边权是 $[0, 1]$ 的实数，求 MST 的期望。

$(n \leq 8)$

Random MST (WJMZBMR)

n 个点的无向图，边权是 $[0, 1]$ 的实数，求 MST 的期望。

$(n \leq 8)$

$O(3^n m^3)$ to $O(3^n m^2)$

Random Number (xudyh)

n 个数，每次随机两个数，$\frac{1}{2}$ 概率替换成两数的和，$\frac{1}{2}$ 概率替换成两数之积，问剩下的数的期望。

$(n \leq 2000)$

期望逆序对 (LYDSY 5058)

长度为 n 的排列，每次随机交换两个元素，交换 m 次，求逆序对数的期望。

$(n \leq 5 \times 10^5, m \leq 10^9)$

Endless Spin (HDOJ 4624)

n 个白格子，每次随机一个区间染黑，问所有格子全黑的期望次数。

$(n \leq 50)$

Unlimited Battery Works (Xi'an 2014)

n 个点的有根树，每次随机一个点 i，把 i 向下 a_i 层的点染黑，问所有节点全黑的期望次数。

$(n \leq 50)$

??? (xudyh)

n 个点的图，每次随机 i, j 连边，问连通的期望次数。

$(n \leq 100)$