

概率与期望

李淳风

长郡中学

2024 年 5 月 5 日

概率

设样本空间为 Ω , 则对于 Ω 中的每一个随机事件 A , 都存在实值函数 $P(A)$, 满足:

- $P(A) \geq 0$;
- $P(\Omega) = 1$;
- 对于若干个两两互斥事件 A_1, A_2, \dots , 有 $\sigma P(A_i) = P(\bigcup A_i)$, 其中 \bigcup 表示并集。

则称 $P(A)$ 为随机事件 A 发生的概率。通俗地讲, 概率是对随机事件发生可能性的度量, 是一个 $0 \sim 1$ 之间的实数。

期望

若随机变量 X 的取值有 x_1, x_2, \dots , 一个随机事件可表示为 $X = X_i$, 其概率为 $P(X = X_i) = p_i$, 则称 $E(X) = \sum p_i x_i$ 为随机变量 X 的数学期望。

通俗地讲, 数学期望是随机变量取值与概率的乘积之和。

期望

若随机变量 X 的取值有 x_1, x_2, \dots , 一个随机事件可表示为 $X = X_i$, 其概率为 $P(X = X_i) = p_i$, 则称 $E(X) = \sum p_i x_i$ 为随机变量 X 的数学期望。

通俗地讲, 数学期望是随机变量取值与概率的乘积之和。

例如, 设掷两次骰子的点数之和为 X , 则随机变量 X 的取值为 $2 \sim 12$, 则掷出 8 点的概率 $P(X = 8) = \frac{5}{36}$ 。

其数学期望 $E(X)$ 为:

期望

若随机变量 X 的取值有 x_1, x_2, \dots , 一个随机事件可表示为 $X = X_i$, 其概率为 $P(X = X_i) = p_i$, 则称 $E(X) = \sum p_i x_i$ 为随机变量 X 的数学期望。

通俗地讲, 数学期望是随机变量取值与概率的乘积之和。

例如, 设掷两次骰子的点数之和为 X , 则随机变量 X 的取值为 $2 \sim 12$, 则掷出 8 点的概率 $P(X = 8) = \frac{5}{36}$ 。

其数学期望 $E(X)$ 为:

$$\frac{1}{36} * (2+12) + \frac{1}{18} * (3+11) + \frac{1}{12} * (4+10) + \frac{1}{9} * (5+9) + \frac{5}{36} * (6+8) + \frac{1}{6} * 7 = 7$$

期望

数学期望是线性函数，满足 $E(aX + bY) = aE(X) + bE(Y)$ 。
该性质是我们能够对数学期望进行递推求解的基本依据。例如，设随机变量 X 表示掷一次骰子的点数，则“掷两次骰子的点数”的数学期望可以用随机变量 $2X$ 来表示，于是 $E(2X) = 2E(X) = 7$ 。

例题

Rainbow 的信号

Rainbow 发明的信号可以用 N 个自然数 A_1, A_2, \dots, A_N 表示。为了防止被偷听, Rainbow 把对话分成了 A, B, C 三部分, 分别用 a, b, c 三个密码加密。现在了解到, 这三部分的密码计算方式如下:

在 $1 \sim N$ 这 N 个数中, 等概率地选取两个数 l, r , 如果 $l > r$, 则交换 l, r 。把信号中的第 l 个数到第 r 个数取出来, 构成一个数列 P 。 A, B, C 三部分对话的密码是数列 P 的 xor 和、 and 和、 or 和的数学期望。现在请你计算这三个密码。

$1 \leq N \leq 10^5$, N 个自然数均不超过 10^9 。

例题

Rainbow 的信号

Rainbow 发明的信号可以用 N 个自然数 A_1, A_2, \dots, A_N 表示。为了防止被偷听, Rainbow 把对话分成了 A, B, C 三部分, 分别用 a, b, c 三个密码加密。现在了解到, 这三部分的密码计算方式如下:

在 $1 \sim N$ 这 N 个数中, 等概率地选取两个数 l, r , 如果 $l > r$, 则交换 l, r 。把信号中的第 l 个数到第 r 个数取出来, 构成一个数列 P 。 A, B, C 三部分对话的密码是数列 P 的 xor 和、 and 和、 or 和的数学期望。现在请你计算这三个密码。

$1 \leq N \leq 10^5$, N 个自然数均不超过 10^9 。

首先, 位运算都是不涉及进位的, 所以我们可以把这 N 个自然数 A_1, A_2, \dots, A_N 都按照二进制分为 31 位, 对于每一位分别进行处理。按照题目中 l, r 的选取方式, 每一个长度为 1 的区间被选出的概率为 $1/N^2$, 其它区间每个被选出的概率为 $2/N^2$ 。

例题

Rainbow 的信号

Rainbow 发明的信号可以用 N 个自然数 A_1, A_2, \dots, A_N 表示。为了防止被偷听, Rainbow 把对话分成了 A, B, C 三部分, 分别用 a, b, c 三个密码加密。现在了解到, 这三部分的密码计算方式如下:

在 $1 \sim N$ 这 N 个数中, 等概率地选取两个数 l, r , 如果 $l > r$, 则交换 l, r 。把信号中的第 l 个数到第 r 个数取出来, 构成一个数列 P 。 A, B, C 三部分对话的密码是数列 P 的 xor 和、 and 和、 or 和的数学期望。现在请你计算这三个密码。

$1 \leq N \leq 10^5$, N 个自然数均不超过 10^9 。

首先, 位运算都是不涉及进位的, 所以我们可以把这 N 个自然数 A_1, A_2, \dots, A_N 都按照二进制分为 31 位, 对于每一位分别进行处理。按照题目中 l, r 的选取方式, 每一个长度为 1 的区间被选出的概率为 $1/N^2$, 其它区间每个被选出的概率为 $2/N^2$ 。

因此, 对于长度为 1 的区间我们可以直接计算答案, 即若 A_i 的第 k 位为 1, 则把答案加上 $2^k * 1/N^2$, 也就是直接把答案加上 $A_i * 1/N^2$ 。

例题

接下来, 我们只需要分别考虑 *xor* 和、*and* 和、*or* 和为 1, 且长度大于等于 2 的区间个数即可。

设 B 是一个 01 序列, B_0, B_1, \dots, B_{31} 分别表示 $A_1 \sim A_N$ 的第 0 到第 31 位, 我们分别考虑每一个 B 序列, 设当前考虑的是 B_k 。

依次枚举右端点 r , 设 $last_j (j = 0, 1)$ 表示数字 j 上一次出现的位置。

对于 *and* 和, 当 B_k 的第 r 位为 1 时, 若 $last_0 < l < r$ 时, B_k 的第 $l \sim r$ 位的 *and* 和为 1。此时对答案的贡献为:

$$2^k * ((r - 1) - (last_0 + 1) + 1) * 2 / N^2$$

对于 *or* 和, 当 B_k 的第 r 位为 1 时 l 可以任取, 此时对答案的贡献为 $2^k * (r - 1) * 2 / N^2$ 。否则若第 r 位不为 1, 需要 $1 \leq l \leq last_1$, 对答案的贡献为 $2^k * last_1 * 2 / N^2$ 。

例题

接下来, 我们只需要分别考虑 xor 和、 and 和、 or 和为 1, 且长度大于等于 2 的区间个数即可。

设 B 是一个 01 序列, B_0, B_1, \dots, B_{31} 分别表示 $A_1 \sim A_N$ 的第 0 到第 31 位, 我们分别考虑每一个 B 序列, 设当前考虑的是 B_k 。

依次枚举右端点 r , 设 $last_j (j = 0, 1)$ 表示数字 j 上一次出现的位置。

对于 and 和, 当 B_k 的第 r 位为 1 时, 若 $last_0 < l < r$ 时, B_k 的第 $l \sim r$ 位的 and 和为 1。此时对答案的贡献为:

$$2^k * ((r - 1) - (last_0 + 1) + 1) * 2 / N^2$$

对于 or 和, 当 B_k 的第 r 位为 1 时 l 可以任取, 此时对答案的贡献为 $2^k * (r - 1) * 2 / N^2$ 。否则若第 r 位不为 1, 需要 $1 \leq l \leq last_1$, 对答案的贡献为 $2^k * last_1 * 2 / N^2$ 。

对于 xor 和则稍微复杂, 我们需要记录

$[1, r - 1], [2, r - 1], \dots, [r - 1, r - 1]$ 这些区间中, xor 和为 0、1 的数量, 可以用 h_0, h_1 表示。若第 r 位为 0, 则加上 $2^k * h_1 * 2 / N^2$ 。右端点 r 增加至 $r + 1$ 时, 若第 r 位为 0, 则 h_0 加 1, 否则交换 h_0, h_1 的值并把 h_1 加 1。

例题

绿豆蛙的归宿

给出一个 N 个点 M 条边的有向无环图，图中每条边都有一个长度。保证从起点出发能到达所有点，从任意点出发都能到达终点。

一只绿豆蛙从起点出发，到终点停止。到达一个顶点时，若有 K 条离开该点的道路，它会等概率地选择一条道路离开该点，即走向每条道路的概率都是 $1/K$ 。求它从起点出发，到达终点的期望长度。

$1 \leq N \leq 10^5$, $1 \leq M \leq 2N$ 。四舍五入保留两位小数输出。

例题

绿豆蛙的归宿

给出一个 N 个点 M 条边的有向无环图，图中每条边都有一个长度。保证从起点出发能到达所有点，从任意点出发都能到达终点。

一只绿豆蛙从起点出发，到终点停止。到达一个顶点时，若有 K 条离开该点的道路，它会等概率地选择一条道路离开该点，即走向每条道路的概率都是 $1/K$ 。求它从起点出发，到达终点的期望长度。

$1 \leq N \leq 10^5$, $1 \leq M \leq 2N$ 。四舍五入保留两位小数输出。

设 f_x 表示从节点 x 出发，走到终点所经过路径的期望长度。若从 x 出发有 k 条边，分别到达 y_1, y_2, \dots, y_k ，长度分别为 z_1, z_2, \dots, z_k ，则有：

$$f_x = \frac{1}{k} \sum_{i=1}^k (f_{y_i} + z_i)$$

显然 $f_N = 0$ ，我们逆拓扑序进行计算即可。

例题

游走

给出一个 n 个点 m 条边的无向连通图，点的编号为 $1 \sim n$ ，边的编号为 $1 \sim m$ 。

小 Z 在这张图上进行随机游走。初始时小 Z 在 1 号点，每一步小 Z 以相等的概率随机选择当前点的某条边，沿着这条边走到下一个顶点，获得等于这条边编号的分数。

当小 Z 到达 n 号点时游走结束，总分为获得的所有分数之和。现在，请你对 m 条边进行编号，使得小 Z 获得的总分期望值最小。

$2 \leq n \leq 500, 1 \leq m \leq 125000$ 。

例题

游走

给出一个 n 个点 m 条边的无向连通图，点的编号为 $1 \sim n$ ，边的编号为 $1 \sim m$ 。

小 Z 在这张图上进行随机游走。初始时小 Z 在 1 号点，每一步小 Z 以相等的概率随机选择当前点的某条边，沿着这条边走到下一个顶点，获得等于这条边编号的分数。

当小 Z 到达 n 号点时游走结束，总分为获得的所有分数之和。现在，请你对 m 条边进行编号，使得小 Z 获得的总分期望值最小。

$2 \leq n \leq 500, 1 \leq m \leq 125000$ 。

如果我们可以求出每条边期望的经过次数，那么我们就可以按照经过次数来贪心地给边进行编号了。

但是这不好直接求，于是我们设 f_i 表示 i 号点的经过次数， d_i 表示 i 号点的度数，那么对于一条连接了 u, v 两个点的边，期望经过次数为 $\frac{f_u}{d_u} + \frac{f_v}{d_v}$ 。

例题

我们现在要求出每个点的期望经过次数 f_i ，但是由于不是有向无环图，我们无法直接递推。

但是我们可以把一系列的式子写下来，如果一个点 u 可以连接到 k 个点 v_1, v_2, \dots, v_k ，那么有：

$$f_u = \sum_{i=1}^k \frac{f_{v_i}}{d_{v_i}}$$

注意对 1 号点和 n 号点需要特判。

这样我们就得到了 n 个方程和 n 个变量，使用高斯消元法解方程组即可。

例题

扑克牌

Zhq 把一副扑克牌随机洗匀，然后依次翻开每一张牌，并放到对应花色的牌堆里面。现在要求出，至少得到 a 张黑桃， b 张红桃， c 张梅花， d 张方块所需要翻开的牌的期望 E 。

特别地，如果翻开的牌是大王或者小王，那么 Zhq 可以把它放入任何一种颜色的牌堆，但他会采用最优选择，使得放入之后 E 尽可能小。

例题

扑克牌

Zhq 把一副扑克牌随机洗匀，然后依次翻开每一张牌，并放到对应花色的牌堆里面。现在要求出，至少得到 a 张黑桃， b 张红桃， c 张梅花， d 张方块所需要翻开的牌的期望 E 。

特别地，如果翻开的牌是大王或者小王，那么 Zhq 可以把它放入任何一种颜色的牌堆，但他会采用最优选择，使得放入之后 E 尽可能小。

设 $f_{a,b,c,d,x,y}$ 表示当前已经翻开了 a 张黑桃， b 张红桃， c 张梅花， d 张方块，并且小王状态为 x ，大王状态为 y 的期望值。其中， $x=4$ 表示没用使用过小王， $x=0\sim 3$ 表示把小王放入的牌堆，大王同理。

以黑桃为例，设 sum 表示当前已经翻开的牌的张数，则牌堆中还剩 $54-sum$ 张牌，黑桃还剩 $13-a$ 张，那么就有 $\frac{13-a}{54-sum}$ 的概率翻开一张黑桃，接下来还需要翻开的期望为 $f_{a+1,b,c,d,x,y}$ 。其它花色同理。

而对于大小王，则是取 $f_{a,b,c,d,x',y} (0 \leq x' \leq 3)$ 中最小的一个作为转移。

分数规划

0/1 分数规划模型是指, 给定整数 a_1, a_2, \dots, a_n 和 b_1, b_2, \dots, b_n , 从中选出若干对整数 (a_i, b_i) , 使得所有选出的 a 之和与 b 之和比值最大, 也就是求一组解 $x_i (1 \leq i \leq n, x_i = 0, 1)$, 使得下列式子最大化:

$$\frac{\sum_{i=1}^n a_i * x_i}{\sum_{i=1}^n b_i * x_i}$$

对于这一类问题, 我们不妨先猜测一个答案 L , 问题就转化为了判断是否存在一组解 x , 使得:

$$\frac{\sum_{i=1}^n a_i * x_i}{\sum_{i=1}^n b_i * x_i} \geq L$$

移项之后变为判断 $\sum_{i=1}^n (a_i - L * b_i) * x_i \geq 0$ 能否成立。

可以发现, L 对于上述式子能否成立具有单调性, 我们可以使用二分答案转化为判定性问题进行解决。

也可以使用 Dinkelbach 迭代法求解, 大概思想是每次用上一轮的答案当做新的 L 来输入, 不断地迭代, 直至答案收敛。

