概率论

rp

- 先抄一大段Yasar GG的课件（

概率

定义

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur, or how likely it is that a proposition is true. (Wikipedia)
概率是对于事件发生或命题正确的可能性的数学描述，属于数学的一个分支。事件 A 的概率写作 $P(A)$。

例如

令事件 A 为“抛一枚硬币，正面朝上”，则 $P(A)=\frac{1}{2}$，$P(\bar{A})=\frac{1}{2}$。
这是显而易见的，但是我们可以更加完备地定义概率。

定义

- 样本空间 Ω：随机试验所有可能结果构成的集合。
- 事件 A：样本空间的一个子集，即 $A \subset P(\Omega)$。试验结果在这个集合内，我们则称事件 A 发生。
- 概率 P：给每个事件一个 $[0,1]$ 之间的数，满足 $P(\Omega) = 1$ 且对于一个不重集合 A，$P(A) = \sum_{x \in A} P(\{x\})$。
- 事件的对立 $\bar{A}$：试验结果不在 A 内，即 $\bar{A} = \Omega \setminus A$，则 $P(\bar{A}) = 1 - P(A)$。
- 事件的交和并：根据集合论的定义，我们可以一样的定义事件的交和并。

定义

- 对于一般的事件 A 和 B，则 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$。
- 若 A 和 B 是独立事件，则 $P(A \cap B) = P(A) \cdot P(B)$。
- 若 A 和 B 是互斥事件，则 $P(A \cap B) = 0$ 且 $P(A \cup B) = P(A) + P(B)$。
- 条件概率 $P(A|B)$：在已知 B 发生的情况下，A 发生的概率。有 $P(A|B) = \frac{P(A \cap B)}{P(B)}$，变形可得 $P(A \cap B) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A)$。

全概率公式（Law of total probability）

定义 Ω 的一个划分 $B_i(i = 1, 2, 3 \cdots)$，$\forall i \neq j, B_i \cap B_j = \emptyset$ 且 $\bigcup_i B_i = \Omega$，则有：

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(B_i) \cdot P(A|B_i)$$

贝叶斯公式（Bayes' theorem）

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{P(A)} = \frac{P(B_i) \cdot P(A|B_i)}{\sum_j P(B_j) \cdot P(A|B_j)}$$

四.经典公式

1. 全概率公式

$$P(B) = \sum_{i}^{n} P(A_i) P(B|A_i).$$ 其中 $\sum P(A_i) = 1$

2. 贝叶斯公式.

$\sum P(A_i) = 1$

$$P(A_j|B) = \frac{P(A_j) \cdot P(B|A_j)}{\sum P(A_j) P(B|A_j)}$$

$\Rightarrow P(A_j B)$

$\Rightarrow P(B)$

练习

1. $P(\bar{A})=0.3$ $P(B)=0.4$ $P(A\bar{B})=0.5$. $P(B|(A\cup\bar{B}))=$

$$P(B \mid A\cup\bar{B}) \cdot P(A\cup\bar{B}) = P(B\cap(A\cup\bar{B}))$$

$$= P(AB)$$

$$P(A) - P(AB) = 0.7 - P(AB) = P(A\bar{B}) = 0.5$$

$$\Rightarrow P(AB) = 0.2$$

$$P(A\cup\bar{B}) = \underset{0.7}{P(A)} + \underset{0.6}{P(\bar{B})} - \underset{0.5}{P(A\bar{B})} = 0.8$$

$$\Rightarrow \frac{0.2}{0.8} = 0.25$$

期望

定义

对于 X 取值有限的情况，设其可能取值为 $x_1, x_2, \cdots, x_n$，定义 X 的期望 $E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$。

整数期望（概率）公式

$$E(X) = \sum_{i=1}^{\infty} i \cdot P(X = i) = \sum_{i=1}^{\infty} P(X = i) \sum_{j=1}^{i} 1$$

$$= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} P(X = i) = \sum_{j=1}^{\infty} P(x \geq j)$$

❶ 抛一枚硬币，硬币正面朝上的次数的期望。

❷ 抛两枚硬币，硬币正面朝上的次数的期望。

我们观察到两枚硬币的期望是一枚的两倍，手算三枚可得是一枚的三倍，这是巧合吗？

性质

1. 期望是线性的：对于任意两个随机变量 X 和 Y，常数 k，有 $E(X+Y)=E(X)+E(Y), E(k\cdot X)=k\cdot E(X)$。
   注： 这里并不要求 X 和 Y 是独立的两个随机变量。
   由此，我们可以知道上面出现的现象并非巧合。

性质

2. 一般情况下，$E(X \cdot Y) \neq E(X) \cdot E(Y)$，特别地，当 X 和 Y 独立时，等式成立。
例：我们进行两次硬币试验，每次抛两枚硬币，求两次实验硬币正面朝上的次数乘积的期望。
由于两次硬币试验是互不干扰的，两次实验硬币正面朝上的次数是独立的随机变量，并且两者是相等的，分别设为 X_1 和 X_2，则 $E(X_1) = E(X_2) = 1$，则
$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2) = 1$。

全期望公式

首先，我们定义条件期望，$E(X|Y=y_i)$ 为当随机变量 $Y=y_i$ 时，X 的期望，由期望的定义可知 $E(X|Y=y_i)=\sum_{x_i} x_i \cdot P(X=x_i|Y=y_i)$。则有：

$$E(E(X|Y))=E(X)$$

在随机变量 X 对应的样本空间 Ω 上取一个划分 $\{A_i\}_i$，定义 $E(X|A_i)$ 为在 A_i 上 X 的期望，则有：

$$E(X)=\sum_i P(A_i)\cdot E(X|A_i)$$

有 n 个随机变量 $\langle X_i \rangle$ ，每个随机变量量都是从 $[1, m]$ 中随机一个整数，求 $\max \langle X_i \rangle$ 的期望。

有 n 个随机变量 $\langle X_i \rangle$ ，每个随机变量量都是从 $[1, m]$ 中随机一个整数，求 $\max \langle X_i \rangle$ 的期望。

设 $S = \max \langle x_i \rangle_{i=1}^{n}$ ，则我们要求的是 $E(S)$ 。根据期望的定义式得到

$$E(S) = \sum_{i=1}^{m} P(S = i)i$$

我们使用前缀和技巧，得到 $P(S = i) = P(S \leq i) - P(S \leq i - 1)$ 。然后我们推一波式子，你发现 $S \leq i$ 等价于 $\langle x_j \rangle_{j=1}^{n} \leq i$ ，那么就可以得到

$$P(S \leq i) = P(\langle x_i \rangle_{i=1}^{n} \leq i) = \left(\frac{i}{m}\right)^n$$
$$P(S = i) = \left(\frac{i}{m}\right)^n - \left(\frac{i-1}{m}\right)^n$$

于是得到

$$E(S) = \sum_{i=1}^{m} i\left(\left(\frac{i}{m}\right)^n - \left(\frac{i-1}{m}\right)^n\right) = m - \frac{1}{m^n}\sum_{i=1}^{m-1} i^n$$

证明：概率为 p 的事件期望 $\frac{1}{p}$ 次发生。

设随机变量 x 表示这个事件在第 x 次发生。于是我们需要求 $E(x)$ 。套公式得到

$$E(x) = \sum_i P(x=i)i$$
$$E(x) = \sum_i (P(x \ge i) - P(x \ge i+1))i$$

于是问题转化为我们要求 $P(x \ge i)$ ，它表示这个事件在第 i 次及以后发生。显然它在前 $i-1$ 次都未发生过，概率为 $(1-p)^{i-1}$ 。因此得到

$$\begin{aligned} E(x) &= \sum_{i=1}^{\infty}((1-p)^{i-1} - (1-p)^i)i \\ &= \sum_{i=0}^{\infty}(1-p)^i \end{aligned}$$

根据小 Trick 得到

$$E(x) = \frac{1}{1-(1-p)} = \frac{1}{p}$$

事实上，对于求期望，可以将其转化为求概率的和：

$$\begin{aligned} E(x) &= \sum_{i=1}^{\infty} P(x=i) \sum_{j=1}^{i} 1 \\ &= \sum_{j=1}^{\infty} \sum_{i \ge j} P(x=i) = \sum_{j=1}^{\infty} P(x \ge j) \end{aligned}$$

箱子里有 n 个球 $1, 2, \cdots, n$ ，你要从里面拿 m 次球，拿了后放回，求取出的数字之和的期望。

- $(n + 1)m/2$

箱子里有 n 个球 $1, 2, \cdots, n$ ，你要从里面拿 m 次球，拿了后不放回，求取出的数字之和的期望。

我们设随机变量 x_i 表示

$$x_i = \begin{cases} i & \text{if i is chosen} \\ 0 & \text{if i isn't chosen} \end{cases}$$

设答案为 S ，得到 $S = \sum_{i=1}^{n} x_i$ ，于是我们要求

$$E(S) = E\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i)$$

展开 $E(x_i)$ 得到

$$E(x_i) = \sum_{j} P(x_i = j)j = P(x_i = i)i$$

显然 $P(x_i = i)$ 表示 i 被取出的概率。而拿球后不放回，相当于在 n 个球中取 m 个，因此概率为 $\frac{m}{n}$ ，于是得到

$$E(x_i) = \frac{m}{n} \cdot i$$

$$E(S) = \sum_{i=1}^{n} \frac{m}{n} \cdot i = \frac{m(n+1)}{2}$$

箱子里有 n 个球 $1,2,\cdots,n$ ，你要从里面拿 m 次球，拿了后以 p_1 的概率放回， p_2 的概率放回两个和这个相同的球（相当于增加一个球），求取出的数字之和的期望。

- 感受一下， 每个球被拿出来概率是一样的。
- 每个球拿出来期望次数一样。
- 随便选个球， 大小期望就是(n+1)/2
- m(n+1)/2

我们仍然设一个随机变量 x_i ，它表示 i 被拿出来的次数乘 i ，即对答案的贡献。形式化地，我们设 y_i 表示 i 被拿了几次，那么 $x_i = y_i \cdot i$ 。假设数字之和为 S ，那么

$$
\begin{aligned}
S &= \sum_{i=1}^{n} x_i \\
E(x_i) &= E(y_i) \cdot i \\
E(S) &= \sum_{i=1}^{n} E(y_i) \cdot i
\end{aligned}
$$

然后我们考虑求 $E(y_i)$ 。首先我们知道，每一个球是平等的，意味着概率是均等的，意味着它们被拿出来的次数的期望是一样的，即 $E(y_1) = E(y_2) = \cdots = E(y_n)$ 。而我们知道，他们被拿出来的次数的和是 m （因为你只拿了 m 个球出来），而 $E(m) = m$ ，即

$$
E\left(\sum_{i=1}^{n} y_i\right) = \sum_{i=1}^{n} E(y_i) = m
$$

因此得到 $E(y_i) = \dfrac{m}{n}$ ，带回上式得到 $E(S) = \dfrac{m(n+1)}{2}$ 。

在一条 n 个点的链上游走，求从一端走到另一端的期望步数。

假设步数是 S ，求 S 的期望。我们定义一个随机变量 x_i 表示从 i 出发随机游走，第一次到 $i+1$ 的步数。

$$S=\sum_{i=1}^{n-1} x_i$$

$$E(S)=E\left(\sum_{i=1}^{n-2} x_i\right)=\sum_{i=1}^{n-1} E(x_i)$$

我们手推一下期望的式子，发现

$$\begin{array}{rcccc}
E(x_1) & = & 1 & & \\
E(x_2) & = & \frac{1}{2}+\frac{1}{2}(1+E(x_1)+E(x_2)) & = & 3 \\
E(x_i) & = & \frac{1}{2}+\frac{1}{2}(1+E(x_{i-1})+E(x_i)) & = & E(x_{i-1})+2
\end{array}$$

就可以递推做了。但算出来就会发现，$E(S)=(S-1)^2$ 。

在一个 n 个点的完全图上游走，求从一个点走到另一个点的期望步数。

设期望步数为 E，得到

$$E = \frac{1}{n-1} + \frac{n-2}{n-1}E + 1$$

解得 $E = n - 1$。

在一个 n 个点的完全二分图上游走，求从一个点走到另一个点的期望步数。

E_1 表示异侧点的期望， E_2 表示同侧点的期望。

$$E_1 = \frac{1}{n} + \frac{n-1}{n}(E_2 + 1)$$
$$E_2 = E_1 + 1$$
$$E_1 = 2n - 1$$

CF1823F

树上从 s 走到 t，随机游走，求走到每个点的期望次数。$2 \le n \le 2 \times 10^5$。

“树上高斯消元”

设 f_u 表示 u 的期望次数，转移要对 s 和 t 特殊处理：

- $$f_t = 1$$
- $$f_u = \sum_{(u,v)\in E, v\neq t} \frac{f_v}{deg_v}$$
- $$f_s = 1 + \sum_{(u,v)\in E, v\neq t} \frac{f_v}{deg_v}$$

发现没法直接转移，因为转移关系是互相依赖的。可以将转移式看作线性方程组，然后上高斯消元解出来 f_u，时间复杂度 $O(n^3)$，远远不够。

套路：把 t 当作树根，将每个 f_u 表示成 $A_u f_{fa} + B_u$ 的形式，然后利用 $f_t = 1$ 得到每个 f_u。

具体地，推下式子：

$$
\begin{aligned}
f_u &= [u=s] + \sum_{(u,v)\in E} \frac{f_v}{deg_v} \\
&= \frac{f_{fa}}{deg_{fa}} + [u=s] + \sum_{v\in son} \frac{f_v}{deg_v} \\
&= \frac{f_{fa}}{deg_{fa}} + [u=s] + \sum_{v\in son} \frac{A_v f_u + B_v}{deg_v} \\
&= \frac{f_{fa}}{deg_{fa}} + [u=s] + f_u \sum_{v\in son} \frac{A_u}{deg_v} + \sum_{v\in son} \frac{B_v}{deg_v}
\end{aligned}
$$

把带有 f_u 的项移到一边，得到：

$$(1 - \sum_{v\in son} \frac{A_v}{deg_v}) f_u = \frac{f_{fa}}{deg_{fa}} + [u=s] + \sum_{v\in son} \frac{B_v}{deg_v}$$

已经结束了，两边同时除以 f_u 的系数就可以得到：

$$A_u = \frac{\frac{1}{deg_{fa}}}{1 - \sum_{v\in son} \frac{A_v}{deg_v}},\ B_u = \frac{[u=s] + \sum_{v\in son} \frac{B_b}{deg_v}}{1 - \sum_{v\in son} \frac{A_v}{deg_v}}$$

两者全都可以用一遍 dfs 求出，注意的一点是当 $fa = t$ 时 A_u 应当为 0，因为不能从 f_t 转移。

时间复杂度 $O(n \log p)$，$\log p$ 是求逆元。

每次随机一个 $[1, n]$ 的整数, 问期望几次能凑齐所有数。

设 $f[i]$ 表示凑齐了 i 个数，期望几次凑齐所有数。

$$f[i]=\frac{i}{n}(f[i]+1)+\frac{n-i}{n}(f[i+1]+1)$$
$$f[i]=f[i+1]+\frac{n}{n-i}$$

有一个有趣的问题，就是这个问题的答案为什么不是每个数第一次被凑出的时间的和的期望？如果这样算那么 $E(S)=n^2$ 。但事实上凑出的时间相当于次数的前缀和，因此是不对的。

有 n 堆石头，第 i 堆个数为 a_i 。每次随机选一个石头然后把那一整堆都扔了，求第 1 堆石头期望第几次被扔。

设随机变量 x_i 表示第 i 堆石头是第几个被拿走的，显然我们有

$$x_1 = 1 + \sum_{i=2}^{n} [x_i < x_1]$$

于是我们要求 $E(x_1)$ ，根据期望的线性性转化为 $E([x_i < x_1])$ ，由于艾弗森括号表达式是一个布尔表达式，因此这个期望等价于 $P(x_i < x_1)$ ，于是得到

$$E(x_1) = 1 + \sum_{i=2}^{n} P(x_i < x_1)$$

而根据 E4，我们稍作拓展，其实 $\frac{a_i}{\sum_{i=1}^{n} a_i}$ 表示的就是第 i 堆石子被拿走的概率。因此 $P(x_i < x_1)$ 表示的就是第 i 堆石子比第 1 堆石子先被拿走的概率，显然为 $\frac{a_i}{a_i + a_1}$ ，于是得到

$$E(x_1) = 1 + \sum_{i=2}^{n} \frac{a_i}{a_i + a_1}$$

- 这个想法对应了一个题
- 猎人杀
- 不过可能有点太难

Luogu 1654

- [P1654 OSU! - 洛谷 | 计算机科学教育新生态 (luogu.com.cn)]

- $(x+1)^3 = ?$

分手是祝愿

考虑按灯泡的最优策略。我们显然是先找到最大的 1 把它变成 0，然后再找最大的 1 变成 0……

因此设 $f(i)$ 表示把一个最优按 i 次能全 0 的局面变成按 $i-1$ 次能全 0 的局面的期望步数。显然，当 $i \le k$ 时你会选最优，即 $f(i)=1$；否则，你有 $\frac{i}{n}$ 的概率按到一个应该按的灯，剩余的概率则会转移到 $f(i+1)$。因此得到

$$f(i)=\frac{i}{n}+\frac{n-i}{n}(f(i)+f(i+1)+1)$$

边界情况 $f(n)=1, f(0)=0$。$f(n)=1$ 是因为你不论按哪个，都能按到你应该按的那个灯。

于是我们先求出初始局面在最优情况下需要按的次数，假设为 x。那么答案为

$$n!\sum_{i=1}^{x} f(i)$$

时间复杂度 $O(n\log_2 n)$ 或 $O(n\sqrt{n})$。