简单离散概率期望

cly_none

前言

在调查的时候好像很多人想要听概率期望的。
我讲课的内容相当简单，其中具体问题的处理方法也不一定是最优的。欢迎各大神仙随时发言。

关于概率

在基本概念中，我们称随机试验 E 的所有基本结果组成的集合为 E 的样本空间。
举个例子，我们扔个正常的六面骰子，会扔出 1 到 6 中的一个数字，那么它的样本空间就是 $\{1,2,3,4,5,6\}$。
我们称样本空间的子集为事件。如果随机试验的结果在某个内，我们称这个事件发生了。
在上面的例子中，$\{1,3,5\}$ 就是一个事件，它发生当且仅当骰子扔出了一个奇数。

离散概率

用来处理样本空间是**可数集**的情况。

设样本空间为 Ω，$f : \Omega \to \mathbb{R}$ 为所有可能结果到其概率的映射，并满足：

1. $\forall x \in \Omega, f(x) \in [0, 1]$。
2. $\sum_{x\in\Omega} f(x) = 1$。

我们定义一个事件 E 的概率为 $P(E) = \sum_{x\in E} f(x)$。
在上面的例子中，骰子每面朝上的概率都是 $\frac{1}{6}$，那么事件 $\{1, 3, 5\}$ 的概率就是 $\frac{1}{2}$。

离散概率 例1

有一个 n 个点的环，一开始你处于 1 号点，你有 50% 概率向前走一步，50% 概率从此停止运动，求停在第 i 个点的概率。

离散概率 例1

这里的样本空间可以视为由若干操作序列构成的集合，每个操作序列记录每一步是向前一步还是就此停下。

走 $\boldsymbol{k}$ 步停下的概率是 $(\frac{1}{2})^{k+1}$。

那么"走的步数模 $\boldsymbol{n}$ 为 $\boldsymbol{i-1}$"的概率就是

$$\left(\frac{1}{2}\right)^{i}+\left(\frac{1}{2}\right)^{i+n}+\left(\frac{1}{2}\right)^{i+2n}+\cdots$$

问题变为无穷项的等比数列求和。

离散概率

从上面的定义，我们可以得到一些推论：

1. 如果事件 E_1 和 E_2 不交，$P(E_1 \bigcup E_2) = P(E_1) + P(E_2)$。
2. 我们称事件 B 一定发生时事件 A 发生的概率为 $P(A|B)$，那么 $P(A|B) = \frac{P(A \bigcap B)}{P(B)}$。
3. 设 $B_1, B_2, \ldots, B_n$ 是样本空间的一个划分，那么 $P(A) = \sum_i P(A|B_i)P(B_i)$。（全概率公式）
4. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$。（贝叶斯公式）

从集合角度不难验证这些推论。

条件概率 例子

你的面前有 4 个盒子，其中某个盒子里有奖品，其余盒子为空。每个盒子有奖品的概率相同。你可以选择一个盒子拿走（但不知道盒子里是什么），在你做出选择之后，主持人会在剩下的空盒子中等概率随机一个，并告诉你这个盒子是空的。问现在每个盒子有奖品的概率是多少。

条件概率 例子

你拿走的那个盒子：设事件 A 是你拿走的盒子有奖品，事件 B 是那个盒子被主持人选中，那么 $P(A|B)=\frac{P(A)P(B|A)}{P(B)}=\frac{\frac{1}{4}\times\frac{1}{3}}{\frac{1}{4}\times\frac{1}{3}+\frac{2}{4}\times\frac{1}{2}}=\frac{1}{4}$。

没被动过的盒子：设事件 A 是这个盒子有奖品，事件 B 是那个被选中的盒子被主持人选中，那么 $P(A|B)=\frac{P(A)P(B|A)}{P(B)}=\frac{\frac{1}{4}\times\frac{1}{2}}{\frac{1}{4}\times\frac{1}{3}+\frac{2}{4}\times\frac{1}{2}}=\frac{3}{8}$。

在每个盒子初始有奖品概率不同的情况下，这个方法依然适用。

随机变量

随机变量是定义在样本空间上确定的实值函数。

函数 $X:\Omega\to R$ 被称为一个随机变量。

期望

对于一个随机变量，定义其期望为

$$E[X] = \sum_{\omega} P(\omega) X(\omega) = \sum_{x} x P(X = x)$$

这里 $X = x$ 表示一个事件，等价于集合 $\{\omega | \omega \in \Omega, X(\omega) = x\}$。

严谨地讲，我们还需要讨论 $E[X]$ 是否收敛，但 OI 比赛中一般不用关心。

期望 例1

你每秒钟有 50% 概率会往前走且只走一米，50% 概率不动，问走 1 米的期望秒数。

期望 例1

我们可以认为样本空间是由所有可能的操作序列构成的集合，且随机变量 X 返回操作序列的长度。

$1 \times (\frac{1}{2})^1 + 2 \times (\frac{1}{2})^2 + 3 \times (\frac{1}{2})^3 + \cdots = 2$。

随机变量的独立性

对于两个随机变量 X_1, X_2 和实数 $x_1 \in X_1(\Omega), x_2 \in X_2(\Omega)$，如果有

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$$

那么称 X_1, X_2 互相独立。

随机变量的独立性与乘积的期望

如果随机事件 X_1, X_2 互相独立，那么我们有

$$
\begin{aligned}
E[X_1 X_2] &= \sum_x x P(X_1 X_2 = x) \\
&= \sum_x x \sum_{x_1} P(X_1 = x_1, X_2 = \frac{x}{x\ 1}) \\
&= \sum_{x_1} x_1 P(X_1 = x_1) \sum_{x_2} x_2 P(X_2 = x_2) \\
&= E[X_1] E[X_2]
\end{aligned}
$$

期望的线性性质

无论随机变量 X_1, X_2 是否独立，总有

$$E[\alpha X_1 + \beta X_2] = \alpha E[X_1] + \beta E[X_2]$$

因此，我们在计算期望时可以将“大”的期望分解为若干“小”的期望的和。

期望线性性质 例1

有 n 盏灯，第 i 盏灯有 p_i 的概率会亮起。

求期望有多少对相邻的灯同时亮起？

期望线性性质 例1

在这个问题中，样本空间 Ω 是由所有长度为 n 的 01 串（所有灯的亮灭状态）构成的集合。

记随机变量 X 返回相邻的灯同时亮起的对数。

记 X_i $(1 \leq i < n)$ 返回第 i 盏灯和第 $i+1$ 盏灯是否同时亮起。

不难验证对于所有 $\omega \in \Omega$，有 $X(\omega) = \sum_i X_i(\omega)$。

因此 $E[X] = \sum_{i=1}^{n-1} E[X_i] = \sum_{i=1}^{n-1} p_i p_{i+1}$。

期望的分解 例2

随意打乱一个 $1 \dots n$ 的排列 P，求生成的笛卡尔树中，所有结点的子树大小之和的期望值。

$n \le 10^6$

期望的分解 例2

先假设取较小的点为根。

可以利用类似于前面的方法，将题目中的随机变量分解为 $\sum_{i,j}[i \text{ 是} j \text{ 的祖先}]$。

不难发现位置为 i 的数是位置为 j 的数的祖先，当前仅当 P_i 不大于 i 到 j 之间的所有数（包括 P_i, P_j）。

概率就等于 $\frac{1}{|i-j|+1}$。

枚举区间长度，容易计算得到答案。

期望的线性性质 例3

给出一个联通的无向图，并指定起点和若干个终点。

你从起点出发，每次从目前所在点的所有邻边中等概率随机选一条走过去，走到任意一个终点后就会立刻停下来。

问停止在每个终点的概率。

$n \leq 100$

期望的线性性质 例3

我们可以删去终点的所有出边。记从 u 走向 v 的概率为 $G_{u,v}$。

设 E_i 表示经过 i 的期望次数，$p_{t,i}$ 表示走 t 步后到达 i 的概率。那么，有期望的线性性质有 $E_i = \sum_t p_{t,i}$。

因此，$E_v = \sum_t p_{t,v} = p_{0,v} + \sum_t \sum_u p_{t,u} G_{u,v} = p_{0,v} + \sum_u E_u G_{u,v}$。

$p_{0,u}$ 当且仅当 u 是起点时为 1。

那么我们就可以列出方程组解出所有 E_i。因为所有终点最多经过一次，故期望经过次数就等于在这个终点停下的概率。

期望的线性性质

如果随机变量 $X:\Omega\to\{0,1,2,3,...\}$，有

$$
\begin{aligned}
\sum_{n=0}^{\infty} P(X>n) &= \sum_{n=0}^{\infty}\sum_{j=n+1}^{\infty} \mathrm{P}(X=j) \\
&= \sum_{j=1}^{\infty}\sum_{n=0}^{j-1} \mathrm{P}(X=j) \\
&= \sum_{j=1}^{\infty} j\mathrm{P}(X=j) \\
&= \sum_{j=0}^{\infty} j\mathrm{P}(X=j) \\
&= \mathbf{E}[X]
\end{aligned}
$$

类比条件概率

如果事件 A 一定发生，那么考虑随机变量 X，我们有

$$P(X=x|A)=\frac{P(X=x,A)}{P(A)}$$

对于随机变量 X,Y，记 $X|Y=y$ 为样本空间 $Y=y$ 上的一个新随机变量，那么 $E[X|Y=y]$ 就是一个定值。

如果不明确 Y 的取值，则 $E[X|Y]$ 是一个新的随机变量，其期望 $E[E[X|Y]]$ 表示在 Y 的各种特定输出 y 下，$E[X|Y=y]$ 的期望值。

全期望公式

$$
\begin{aligned}
E[E[X|Y]] &= \sum_y E[X|Y=y]P(Y=y) \\
&= \sum_y P(Y=y) \sum_x xP(X=x|Y=y) \\
&= \sum_x x \sum_y P(X=x|Y=y)P(Y=y) \\
&= \sum_x xP(X=x) \\
&= E[X]
\end{aligned}
$$

全期望公式

$$E[E[X|Y]] = E[X]$$

一个感性的例子是，如果你要统计全年段所有学生的分数平均值，只要对所有班级的平均分求个加权平均数。

全期望公式 例1

给出一个有向图和起点终点。你需要求出从起点出发，每个从所有出边中等概率随机选一条走，最终到达终点的期望步数。

$n \leq 100$

全期望公式 例1

我们认为样本空间是由许多路径组成的集合。随机变量 X 返回路径的长度。

设 u 的出边数为 d_u，对于所有能从 u 直接到达的 v，从 u 开始的路径的下一个点是 v 的概率是 $\frac{1}{d_u}$。

记事件"路径第一个点为 u" 为 E_u。

我们有 $E[X|E_u] = E[E[(X|E_u)|(\text{第二个点是}v)]] = \sum_v \frac{1}{d_u}(1 + E[X|E_v])$

因此可以列出方程组解出所有 $E[X|E_i]$。

一点习题

习题1

你每秒钟有 50% 概率会往前走且只走一米，50% 概率不动，问 n 秒后走出 m 米的概率。

习题1

设 $dp_{i,j}$ 表示 i 秒后走出 j 米的概率。

$$dp_{i,j} = 0.5 \times dp_{i-1,j-1} + 0.5 \times dp_{i-1,j}$$

习题2

你站在一个有数字的矩阵的左下角，你要走到右上角去，每次只能向右或向上走。

你每次有 50% 的概率向上走，50% 概率向右走。

若处于上边界或右边界，则只能往一个方向。

求路径上数字和的期望值。

习题2

记 $v_{i,j}$ 表示在位置 (i,j) 上的值。

设 $f_{i,j}$ 表示经过 (i,j) 的概率。

设 $g_{i,j}$ 表示走到 (i,j) 时路径上数字和的期望值乘 $f_{i,j}$。

不难得到转移方程。

习题2

我们还有另一种方法。

$\mathbf{ans} = \sum_{i,j} v_{i,j} f_{i,j}$。

一种理解方式是，路径上数字和的期望值就等于每个数字的大小乘上它被经过的概率。

习题3

你每秒钟有 50% 概率会往前走且只走 1 米，50% 概率会往回退 1 米（若处于起点则必定前进），求第一次到达 10 米的期望秒数。

习题3

设 f_i 表示已经走了 i 米，要到达 10 米还要走的期望步数。

那么有

- $f_{10} = 0$
- $f_i = 1 + \frac{1}{2}f_{i-1} + \frac{1}{2}f_{i+1}\ (1 < i < 10)$。
- $f_0 = f_1 + 1$。

我们可以视 f_0 为未知数，将其他所有 f_i 用 f_0 和常数表示出来，并通过 $f_{10} = 0$ 解出 f_0。

习题4

你现在正参与一个抽奖活动，初始中奖概率为 2%。
当你连续 50 次没有中奖时，则下一次中奖概率提升到 4%；若依旧没有中奖，则下一中奖概率提升到 6%；每次没抽到概率都会累加以此类推，直到中奖。
假设你一直抽到了中奖为止，求每次抽奖的期望中奖概率。

习题4

我们容易构造一个带点权的图，那么在这个图上从源点开始随机游走直到走回源点，每次走到的点的期望点权值就是答案。

考虑利用全概率公式枚举所有可能的路径：

$E[X] = \sum_p P(p)E[X|P]$

$E[X|P]$ 就等于路径平均值。

习题5

如果你特别有钱，无限地抽了下去呢？

习题5

一种感性的处理方法是，一个点被访问的概率与走一圈它被访问的期望次数成正比。

下面是另一种意识流：

可以考虑所有长度为 n 的路径，当 $n \to \infty$，路径平均值的期望应该等于答案。

我们把所有路径补长到下一次中奖。

考虑如果将所有点的点权减去答案，路径平均期望值应该等于 0。

于是期望路径和等于 0。

所以得到了上面的东西。

Bonus

今天早上 T2

> n 种颜色的球，每种颜色 a_i 个。每次你会等概率取出两个球，将前一个球的颜色变成后一个球的颜色，然后把它们放回去。问期望操作几次使所有球颜色相同。
> $n \leq 2000, a_i \leq 100000$

Thank you!