

# QTA2025 暑期求职笔试训练营-week3

欧岱松

## 1. 聚会握手次数

侯哥和他的室友小 y 正在举办一个聚会, 他邀请了另外 10 对室友。在聚会期间, 侯哥对聚会上的每个人 (不包括他自己) 进行调查, 并询问每个人握了多少次手。假设握手符合以下条件:

- a) 每个人都没有与室友握手
- b) 每个人握手的次数都不同
- c) 两个人最多只会握手一次

请问侯哥的室友小 y 在聚会中握了多少只手?

我们可以知道总共有 22 个人参加游戏, 21 个人被调查, 且每个人握手的次数都不同, 并且不与自己的室友握手。继续分析, 对于聚会中的任何一个人, 最多握 20 次手, 最少 0 次。侯哥调查了 21 个人, 得到了 21 个不同的握手次数。结合上一步的结论, 这意味着这 21 个人的握手次数恰好是集合  $\{0, 1, 2, \dots, 20\}$  的一个排列。注意到不能和室友握手的特性, 和握手次数的特性, 我们考虑不同握手次数的人的配对关系。让我们把握了  $k$  次手的人称为  $P_k$ 。考虑握手最多的人 ( $P_{20}$ ) 和握手最少的人 ( $P_0$ ):

- $P_{20}$  握了 20 次手, 这意味着他/她与除了自己和自己室友之外的所有人都握了手。
- $P_0$  握了 0 次手, 即没有和任何人握手。
- 由于  $P_{20}$  与几乎所有人都握了手, 但他/她唯独没有和自己的室友握手。而  $P_0$  正好是那个没和任何人握手的人。因此,  $P_0$  必定是  $P_{20}$  的室友。

接着考虑次高和次低的人:  $P_{19}$  和  $P_1$ 。通过类似的逻辑可以推断出,  $P_1$  必定是  $P_{19}$  的室友。类似的, 我们可以发现一个配对模式:  $(P_{20}, P_0), (P_{19}, P_1), (P_{18}, P_2), \dots, (P_{11}, P_9)$ 。

接着我们需要考虑小 y 的身份。通过上述配对, 我们找到了 10 对室友, 共 20 个人。这正好是侯哥邀请的“另外 10 对室友”。在被调查的 21 人 (握手次数为  $0, 1, \dots, 20$ ) 中, 只有一个人没有被配对:  $P_{10}$ , 即握了 10 次手的人。这个剩下的人不属于那 10 对客人, 那么他/她必然是侯哥的室友——小 y。所以侯哥的室友小 y 在聚会中握了 10 只手。

值得说明的是, 为什么小 y 在此前不会成为某一个  $P_j$ , 不失一般性, 令小 y 为  $P_{20}$ , 则侯哥必然握手 0 次, 这意味着剩余的 20 个人里面不存在握手 0 次的人, 这与此前握手次数是 0-20 天的全排列矛盾, 故而我们的结论是稳健的。

## 2. 军训转圈圈

汇丰商学院的军训活动中, 教官令 100 名同学横排站开。第一秒, 教官喊口令“向左转”, 然后同学们有的会左转、有的会右转, 这样转完后一些同学会面对面; 下一秒, 这些面对面的同学都会向

后转;再下一秒仍是如此。关于转动最终会不会结束,以下说法正确的是:

- A: 转动一定会结束,最长持续时间 100 秒;
- B: 转动一定会结束,最长持续时间 200 秒;
- C: 转动有可能不会结束,一直持续下去的概率为  $1/100$ ;
- D: 转动有可能不会结束,一直持续下去的概率为  $1/10$ ;

**转动会不会结束?**转动一定会结束。我们给队伍中的每个位置编号,从左到右依次为  $1, 2, 3, \dots, 100$ 。现在,我们定义一个量  $P$  为所有朝右看的同学所在位置编号的总和。当一对“面对面”的同学(即一个朝右的在左,一个朝左的在右)向后转时,我们来观察  $P$  的变化。假设这对同学分别在位置  $i$  和  $i+1$ 。**转身后:**朝右的同学在位置  $i$ ,位置编号  $i$  被计入总和  $P$  中。**转身后:**原来在位置  $i$  的同学转向了左边,而原来在位置  $i+1$  的同学转向了右边。 **$P$  的变化:**在新的总和中,我们去掉了  $i$  但加上了  $i+1$ 。因此,新的总和  $P_{\text{新}} = P_{\text{旧}} - i + (i+1) = P_{\text{旧}} + 1$ 。这意味着,每一次发生转身动作,我们定义的量  $P$  都会至少严格地加 1(若有多位面对面的同学,则会增加更多)。由于学生总数是固定的,  $P$  的值有一个上限(当所有朝右的同学都集中在队伍最右侧时,  $P$  取得最大值)。总的来说,一个严格递增且有上限的整数序列必然会在有限步内停止增加。因此,转身过程必然会结束。

**最长持续时间是多久?**为了找到最长的持续时间,我们需要构想一个使得转身次数最多的初始排列。我们可以将朝右的同学看作一个“粒子”(‘>’),朝左的同学看作另一个“粒子”(‘<’)。转身的过程  $> <$  变成  $< >$ ,可以看作是两种粒子相遇后互相穿过对方。整个过程将在所有朝左的同学都移动到所有朝右的同学的左边时结束。最终的稳定状态是  $<<<\dots>>>$ 。持续时间由需要“旅行”最远距离的那个同学决定。**最坏情况:**假设在位置 1 的同学朝右,而其余 99 名同学都朝左 ( $> <<<\dots<$ )。为了到达队伍最右边的最终位置 100,这位同学需要“穿过”所有 99 名朝左的同学。每一次相遇并转身,他相当于向右前进了一个位置。因此,他需要 99 秒才能走完全程。同理,如果在位置 100 的同学朝左,其余 99 名都朝右 ( $>\dots>>> <$ ),同样需要 99 秒。最长的旅行距离是 99 个位置,这意味着最长的持续时间是 99 秒。

我们的分析表明,转动一定会结束,最长持续时间为 99 秒,加上初始转动的一秒,故而正确的选项是 A。

### 3. 联合正态分布的条件协方差

若  $x, y \sim N(0, 1)$  i.i.d, 计算  $\text{Cov}(x, y | x + y > 0)$

0

根据定义,条件协方差为:

$$\text{Cov}(x, y | x + y > 0) = E[xy | x + y > 0] - E[x | x + y > 0]E[y | x + y > 0]$$

为了简化问题,我们引入新的变量:

$$u = x + y$$

$$v = x - y$$

由于  $x$  和  $y$  是独立同分布的正态分布,  $u$  和  $v$  也是正态分布,据此我们可以计算相应的期望、方差、协方差。

- $E[u] = E[x] + E[y] = 0$

- $E[v] = E[x] - E[y] = 0$
- $Var(u) = Var(x) + Var(y) = 2$
- $Var(v) = Var(x) + Var(y) = 2$
- $Cov(u, v) = E[x^2] - E[y^2] = 0$

因为  $u, v$  是联合正态分布且协方差为 0, 所以  $u, v$  相互独立。我们得到:

- $u \sim N(0, 2)$
- $v \sim N(0, 2)$

我们将  $x, y$  用  $u, v$  表示, 得到:

- $x = \frac{u+v}{2}, \quad y = \frac{u-v}{2}$
- $E[x|u > 0] = E\left[\frac{u+v}{2} \middle| u > 0\right] = \frac{1}{2}(E[u|u > 0] + E[v|u > 0])$

由于  $u, v$  相互独立, 条件  $u > 0$  不影响  $v$  的分布, 所以  $E[v|u > 0] = E[v] = 0$ 。因此:

$$E[x|u > 0] = \frac{1}{2}E[u|u > 0]$$

同理,  $E[y|u > 0] = \frac{1}{2}E[u|u > 0]$ 。

对于一个变量  $Z \sim N(0, \sigma^2)$ , 其条件期望  $E[Z|Z > 0] = \sigma\sqrt{\frac{2}{\pi}}$ , 证明见附录 2。在我们的问题中,  $u \sim N(0, 2)$ , 所以有:

$$E[u|u > 0] = \sqrt{2} \cdot \sqrt{\frac{2}{\pi}} = \frac{2}{\sqrt{\pi}}$$

因此,

$$E[x|x+y > 0] = E[y|x+y > 0] = \frac{1}{2} \cdot \frac{2}{\sqrt{\pi}} = \frac{1}{\sqrt{\pi}}$$

接下来我们计算  $E[xy|u > 0]$ 。

$$xy = \frac{u^2 - v^2}{4}$$

$$E[xy|u > 0] = \frac{1}{4}(E[u^2|u > 0] - E[v^2|u > 0])$$

由于  $u, v$  独立,  $E[v^2|u > 0] = E[v^2] = Var(v) = 2$ 。

由于  $u \sim N(0, 2)$  的分布关于 0 对称,  $E[u^2|u > 0] = E[u^2] = Var(u) = 2$ 。

所以,

$$E[xy|u > 0] = \frac{1}{4}(2 - 2) = 0$$

将所有计算结果代入协方差公式:

$$\begin{aligned} Cov(x, y|x+y > 0) &= E[xy|x+y > 0] - E[x|x+y > 0]E[y|x+y > 0] \\ &= -\frac{1}{\pi} \end{aligned}$$

## 4. Chameleons Jump

There are 5 chameleons and they are initially sitting on red, blue, green, yellow and purple flowers respectively. For every 5 seconds, a random chameleon will jump to another random flower with at least one chameleon already there, and turn into that color. For how much time do we expect that they all change into the same color?

我们可以将其建模为有吸收态的马尔可夫过程。我们的目标是计算从初始状态  $s_1$  到达吸收态  $s_7$  的期望步数。我们定义  $E_i$  为从状态  $s_i$  出发，首次到达吸收态  $s_7$  所需要的平均步数。

根据定义，我们有  $E_7 = 0$ ，对于任何非吸收态  $s_i$ ，其期望吸收时间  $E_i$  满足以下递推关系：

$$E_i = 1 + \sum_{j=1}^7 P_{ij} E_j$$

其中， $P_{ij}$  是从状态  $s_i$  一步转移到状态  $s_j$  的概率。下面我们对每个状态建立方程。

状态  $s_1 = \{1, 1, 1, 1, 1\}$

- 任意一次跳跃必然导致两种颜色合并，使状态转移到  $s_2 = \{2, 1, 1, 1\}$ 。
- 转移概率:  $P_{12} = 1$ 。
- 方程 (1):  $E_1 = 1 + P_{12}E_2 = 1 + E_2$ 。

状态  $s_2 = \{2, 1, 1, 1\}$  (4 种颜色)

- 选中 2 只同色的变色龙之一（概率  $\frac{2}{5}$ ），它跳到另外 3 种颜色之一的花上。状态不变，仍为  $\{2, 1, 1, 1\}$  ( $s_2$ )。所以  $P_{22} = \frac{2}{5}$ 。
- 选中 3 只“单身”变色龙之一（概率  $\frac{3}{5}$ ）。它会跳到另外 3 种颜色的花上。
  - \* 有  $\frac{1}{3}$  的概率跳到含 2 只变色龙的花上，形成  $\{3, 1, 1\}$  ( $s_3$ )。故  $P_{23} = \frac{3}{5} \times \frac{1}{3} = \frac{1}{5}$ 。
  - \* 有  $\frac{2}{3}$  的概率跳到另外 2 种单身颜色的花上，形成  $\{2, 2, 1\}$  ( $s_4$ )。故  $P_{24} = \frac{3}{5} \times \frac{2}{3} = \frac{2}{5}$ 。
- 方程 (2):  $E_2 = 1 + \frac{2}{5}E_2 + \frac{1}{5}E_3 + \frac{2}{5}E_4 \implies \frac{3}{5}E_2 = 1 + \frac{1}{5}E_3 + \frac{2}{5}E_4$ 。

状态  $s_3 = \{3, 1, 1\}$  (3 种颜色)

- 选中 3 只同色之一（概率  $\frac{3}{5}$ ），跳到 2 种单身颜色之一的花上，形成  $\{2, 2, 1\}$  ( $s_4$ )。故  $P_{34} = \frac{3}{5}$ 。
- 选中 2 只单身之一（概率  $\frac{2}{5}$ ）。
  - \* 有  $\frac{1}{2}$  概率跳到含 3 只变色龙的花上，形成  $\{4, 1\}$  ( $s_5$ )。故  $P_{35} = \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$ 。
  - \* 有  $\frac{1}{2}$  概率跳到另一单身颜色的花上，形成  $\{3, 2\}$  ( $s_6$ )。故  $P_{36} = \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$ 。
- 方程 (3):  $E_3 = 1 + \frac{3}{5}E_4 + \frac{1}{5}E_5 + \frac{1}{5}E_6$ 。

状态  $s_4 = \{2, 2, 1\}$  (3 种颜色)

- 选中唯一的单身（概率  $\frac{1}{5}$ ），跳到另外两种含 2 只变色龙的花上，形成  $\{3, 2\}$  ( $s_6$ )。故  $P_{46} = \frac{1}{5}$ 。

– 选中 4 只非单身之一（概率  $\frac{4}{5}$ ）。

\* 有  $\frac{1}{2}$  概率跳到另一组含 2 只变色龙的花上，形成  $\{3, 1, 1\}$  ( $s_3$ )。故  $P_{43} = \frac{4}{5} \times \frac{1}{2} = \frac{2}{5}$ 。

\* 有  $\frac{1}{2}$  概率跳到含 1 只变色龙的花上，状态不变，仍为  $\{2, 2, 1\}$  ( $s_4$ )。故  $P_{44} = \frac{4}{5} \times \frac{1}{2} = \frac{2}{5}$ 。

• 方程 (4):  $E_4 = 1 + \frac{1}{5}E_6 + \frac{2}{5}E_3 + \frac{2}{5}E_4 \implies \frac{3}{5}E_4 = 1 + \frac{2}{5}E_3 + \frac{1}{5}E_6$ 。

状态  $s_5 = \{2, 2, 1\}$  (3 种颜色)

• – 选中唯一的单身（概率  $\frac{1}{5}$ ），跳到另外两种含 2 只变色龙的花上，形成  $\{3, 2\}$  ( $s_6$ )。故  $P_{46} = \frac{1}{5}$ 。

– 选中 4 只非单身之一（概率  $\frac{4}{5}$ ）。

\* 有  $\frac{1}{2}$  概率跳到另一组含 2 只变色龙的花上，形成  $\{3, 1, 1\}$  ( $s_3$ )。故  $P_{43} = \frac{4}{5} \times \frac{1}{2} = \frac{2}{5}$ 。

\* 有  $\frac{1}{2}$  概率跳到含 1 只变色龙的花上，状态不变，仍为  $\{2, 2, 1\}$  ( $s_4$ )。故  $P_{44} = \frac{4}{5} \times \frac{1}{2} = \frac{2}{5}$ 。

• 方程 (4):  $E_4 = 1 + \frac{1}{5}E_6 + \frac{2}{5}E_3 + \frac{2}{5}E_4 \implies \frac{3}{5}E_4 = 1 + \frac{2}{5}E_3 + \frac{1}{5}E_6$ 。

状态  $s_5 = \{4, 1\}$  (2 种颜色)

• – 选中唯一的单身（概率  $\frac{1}{5}$ ），跳跃后形成  $\{5\}$  ( $s_7$ ，吸收态)。故  $P_{57} = \frac{1}{5}$ 。

– 选中 4 只同色之一（概率  $\frac{4}{5}$ ），跳跃后形成  $\{3, 2\}$  ( $s_6$ )。故  $P_{56} = \frac{4}{5}$ 。

• 方程 (5):  $E_5 = 1 + \frac{1}{5}E_7 + \frac{4}{5}E_6 = 1 + \frac{1}{5}(0) + \frac{4}{5}E_6 = 1 + \frac{4}{5}E_6$ 。

状态  $s_6 = \{3, 2\}$  (2 种颜色)

• – 选中 3 只同色之一（概率  $\frac{3}{5}$ ），跳跃后状态不变，仍为  $\{3, 2\}$  ( $s_6$ )。故  $P_{66} = \frac{3}{5}$ 。

– 选中 2 只同色之一（概率  $\frac{2}{5}$ ），跳跃后形成  $\{4, 1\}$  ( $s_5$ )。故  $P_{65} = \frac{2}{5}$ 。

• 方程 (6):  $E_6 = 1 + \frac{3}{5}E_6 + \frac{2}{5}E_5 \implies \frac{2}{5}E_6 = 1 + \frac{2}{5}E_5$ 。

求解方程组

我们得到了一个包含 6 个未知数 ( $E_1$  到  $E_6$ ) 的线性方程组。我们从最简单的方程开始，自底向上代入求解。

1. 解  $E_5, E_6$ : 将方程 (5)  $E_5 = 1 + \frac{4}{5}E_6$  代入方程 (6)  $\frac{2}{5}E_6 = 1 + \frac{2}{5}E_5$  后，解得:  $E_6 = \frac{35}{2} = 17.5$   
将  $E_6$  的值代入方程 (5)，解得:  $E_5 = 15$

2. 解  $E_3, E_4$ : 将  $E_5 = 15, E_6 = 17.5$  代入关于  $E_3, E_4$  的方程组，解得:  $E_4 = \frac{125}{6} E_3 = 20$

3. 解  $E_2$ : 将  $E_3 = 20, E_4 = 125/6$  代入关于  $E_2$  的方程，解得:  $E_2 = \frac{200}{9}$

4. 解  $E_1$ : 代入方程 (1)  $E_1 = 1 + E_2$ ，解得:  $E_1 = 1 + \frac{200}{9} = \frac{209}{9}$

最终结果

我们计算出，从初始状态  $s_1$  到达吸收态  $s_7$  的期望步数为  $\frac{209}{9}$  步。

由于每个时间步长（一次跳跃）为 5 秒，所以总的期望时间为：

$$\text{总期望时间} = \frac{209}{9} \text{ 步} \times 5 \text{ 秒/步} = \frac{1045}{9} \text{ 秒} \approx 116.11 \text{ 秒}$$

## 5. 圆环上随机游走

1. 圆环上有  $n$  个点, 一只蚂蚁从圆环的某点出发随机游走 (即每一步有概率  $\frac{1}{2}$  顺时针移动到相邻点, 有概率  $\frac{1}{2}$  逆时针移动到相邻点), 求蚂蚁走遍圆环上所有点位所需移动步数的期望?
2. 如果蚂蚁走遍圆环上所有点后立即停下, 那么蚂蚁所停位置在圆环上的概率分布是怎样的?

### 1) 问题一: 走遍所有点位所需移动步数的期望

我们将其分解为  $n - 1$  个阶段, 计算每新发现一个点所需的期望步数, 然后将它们相加。  
具体来说, 我们做如下分段:

- 阶段 1: 从已访问 1 个点 (起始点) 到访问第 2 个点。
- 阶段 2: 从已访问 2 个点到访问第 3 个点。
- ...
- 阶段  $k$ : 从已访问  $k$  个点到访问第  $k + 1$  个点。
- ...
- 阶段  $n-1$ : 从已访问  $n - 1$  个点到访问第  $n$  个点。

总的期望步数  $E$  就是所有这些阶段期望步数之和:  $E = \sum_{k=1}^{n-1} T_k$ , 其中  $T_k$  是完成阶段  $k$  所需的期望步数。

### 计算总期望步数 $E$

由附录的证明可以知道,

$$T_k = k \text{ 步}$$

接着, 现在我们将所有阶段的期望步数加起来:

$$\begin{aligned} E &= T_1 + T_2 + T_3 + \cdots + T_{n-1} \\ &= 1 + 2 + 3 + \cdots + (n - 1) \\ &= \frac{n(n - 1)}{2} \end{aligned}$$

所以, 蚂蚁走遍圆环上所有  $n$  个点位所需移动步数的期望为  $\frac{n(n-1)}{2}$ 。

### 2) 蚂蚁停止位置的概率分布

当蚂蚁走遍所有点后, 它会停在最后一个被访问的点。我们需要分析这个点的位置概率分布, 考虑到蚂蚁的起始点是第一个被访问的点, 所以它永远不可能是最后一个被访问的点。这意味着, 最终停止的位置只可能是除了起始点之外的  $n - 1$  个点之一。又由于圆环具有对称性, 蚂蚁每一步向左和向右的概率都是  $\frac{1}{2}$ , 所以每个点被访问的概率是相等的。所以, 所有  $n - 1$  个非起始点成为最后一个被访问的点的概率必须是均等的。故而, 由于每个点的概率都相等, 所以每个点的概率就是  $\frac{1}{n-1}$ 。如果蚂蚁走遍所有点后停下, 其停止位置均匀分布在除起始点以外的所有  $n - 1$  个点上。每个点被作为停止位置的概率都是  $\frac{1}{n-1}$ 。

## 6. 猜测棋子位置

1. 一个棋子沿着数轴始终以固定速度向一个方向移动。假设起点已知, 移动的方向未知, 移动速度  $v$  为整数。你可以在每个时间点猜一次棋子位置, 设计一个策略在可数的次数内猜出棋子位置
2. 假设棋子的起点未知, 你会如何设计策略?

### 1) 起点已知, 方向、速度未知

我们已知起点  $x_0$ , 未知速度  $v$ , 它是一个非零整数, 即  $v \in \mathbb{Z} \setminus \{0\} = \{1, -1, 2, -2, 3, -3, \dots\}$ 。方向包含在  $v$  的正负号中。在时间  $t$  棋子的位置是  $p(t) = x_0 + v \cdot t$ 。我们需要设计一个在每个时间点  $t$  的猜测序列  $g(t)$ , 保证不论真实的  $v$  是多少, 总会在某个时间点  $t_0$  猜中, 即  $g(t_0) = p(t_0)$ 。

所有可能的整数速度  $v$  构成的集合可以构造与自然数集的一一映射, 故而是可数无穷的。在每一个时间切面上猜测一种可能的模式, 从而构造一个可数无穷的猜测位置集合, 枚举某一种初速度可能的位置。

#### 1. 枚举所有可能性:

我们将所有可能的速度  $v$  排列成一个序列, 比如:

$$v_1 = 1, \quad v_2 = -1, \quad v_3 = 2, \quad v_4 = -2, \quad v_5 = 3, \quad v_6 = -3, \quad \dots$$

这个序列覆盖了所有非零整数。任何一个特定的速度  $v_{actual}$  都在这个序列的某个位置上。

#### 2. 设计猜测序列:

我们的策略是, 在第  $k$  个时间点, 我们专门用来测试第  $k$  种速度的可能性是否为真。

- 在时间  $t = 1$  时, 我们猜测速度是  $v_1 = 1$ 。那么我们猜的位置是  $g(1) = x_0 + v_1 \cdot 1 = x_0 + 1$ 。
- 在时间  $t = 2$  时, 我们猜测速度是  $v_2 = -1$ 。那么我们猜的位置是  $g(2) = x_0 + v_2 \cdot 2 = x_0 - 2$ 。
- 总的来说: 在时间点  $t = k$  时, 我们猜测棋子的速度是  $v_k$ , 并猜在相应的位置上:  $g(k) = x_0 + v_k \cdot k$ 。

**策略有效性证明** 假设棋子实际的速度是  $v_{actual}$ 。因为我们的速度序列  $v_1, v_2, v_3, \dots$  包含了所有非零整数, 所以  $v_{actual}$  必然在这个序列的某个位置, 比如说第  $m$  个位置, 即  $v_{actual} = v_m$ 。

现在我们在时间  $t = m$  时会发生什么:

- 棋子的实际位置是:  $p(m) = x_0 + v_{actual} \cdot m = x_0 + v_m \cdot m$ 。
- 我们的猜测位置是: 根据策略, 在  $t = m$  时, 我们猜测的速度是  $v_m$ , 所以我们猜的位置是  $g(m) = x_0 + v_m \cdot m$ 。

可以看到,  $p(m) = g(m)$ 。我们在第  $m$  步时准确地猜中了棋子的位置。由于任何可能的速度  $v$  都有一个对应的序号  $m$ , 所以我们保证总能在可数的步数内猜中。

### 2) 起点未知, 方向、速度未知

现在, 未知量增加了一个: 起点  $x_0$  也是一个未知的整数, 但此时, 可以看作两个无穷可数集合的并  $(x_0, v)$ , 仍然满足可列假设此时, 我们仍然需要将所有可能性排成一个序列, 然后依次测试。

### 1. 枚举所有可能性:

现在的每一种可能性是一个整数对  $(x_0, v)$ 。所有这种整数对的集合  $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$  也是可数无穷的。我们可以用“对角线”的方式将所有这些可能性排成一个序列，例如：

$$\begin{aligned} H_1 &= (x_0 = 0, v = 1) \\ H_2 &= (x_0 = 1, v = 1) \\ H_3 &= (x_0 = 0, v = -1) \\ H_4 &= (x_0 = -1, v = 1) \\ &\vdots \end{aligned}$$

故而我们有了一个不遗漏、不重复的方法，能将所有可能的轨迹假设  $H_k = (x_{0,k}, v_k)$  一一列出。

### 2. 设计猜测序列:

策略与第一部分完全类似。在第  $k$  个时间点，我们专门用来测试第  $k$  个轨迹假设  $H_k$  是否为真。

- **通用策略:** 在时间点  $t = k$  时，我们猜测棋子的轨迹是  $H_k = (x_{0,k}, v_k)$ ，并猜在相应的位置上： $g(k) = x_{0,k} + v_k \cdot k$ 。

#### 策略有效性证明

假设棋子实际的轨迹是由整数对  $(x_{0,actual}, v_{actual})$  决定的。因为我们的假设序列  $H_1, H_2, H_3, \dots$  包含了所有可能的整数对，所以这个实际轨迹必然在序列的某个位置，比如说第  $m$  个位置，即  $H_m = (x_{0,m}, v_m) = (x_{0,actual}, v_{actual})$ 。

在时间  $t = m$  时：

- 棋子的实际位置是： $p(m) = x_{0,actual} + v_{actual} \cdot m$ 。
- 我们的猜测位置是：根据策略，在  $t = m$  时，我们测试的是第  $m$  个假设  $H_m$ ，所以我们猜的位置是  $g(m) = x_{0,m} + v_m \cdot m$ 。

由于  $x_{0,m} = x_{0,actual}$  且  $v_m = v_{actual}$ ，所以  $p(m) = g(m)$ 。我们同样在第  $m$  步猜中了位置。此策略保证有效。

## 7. 量化基金的超额回撤

某些公司在笔试中也会以问答题的形式考察候选人对行业整体的理解，例如以下的真题：描述一下你了解的中国量化私募过去五年比较大的行业性超额回撤（指大部分量化基金的超额收益同时发生较大回撤，不考虑大盘波动）。当时发生了什么，是什么原因，在你看来能否预警或者控制？

### 事件一：2021 年 Q4 “因子拥挤” 回撤

#### 核心现象

- **背景:** 2021 上半年量化规模（AUM）急速扩张。
- **表现:** Q4 起，主流策略（中证 500/1000 增强）出现持续近 3 个月的显著超额回撤。
- **性质:** 行业首次面临范围广、时间长的“Alpha 危机”。



### 核心原因

- **因子拥挤 (Factor Crowding)**: 最主要原因。过多资金涌入相似的短周期价量因子，导致因子失效和交易踩踏。
- **市场风格切换 (Style Shift)**: 市场由中小盘成长切换至大盘价值，量化模型主要 Alpha 源失效。

### 预警与控制

- **预警信号**: 因子拥挤度指标、股指期货高升水（基差）、Alpha 收益高度相关性。
- **控制手段**:
  - **管理人**: 控制策略容量与 AUM；研发并配置低相关的新 Alpha 因子。
  - **投资者**: 分散投资于不同类型和管理人。

## 事件二：2024 年初“杠杆与流动性”回撤

### 核心现象

短几周内，小盘股策略和市场中性策略遭遇“暴力”式历史级别超额回撤。对冲端（股指期货）的巨额亏损吞噬了股票端的 Alpha 收益。使用杠杆的 DMA 产品爆仓，加剧市场下跌。

### 核心原因

- **小盘股流动性危机**: 在悲观预期和“雪球”敲入压力下，小盘股出现恐慌性抛售。
- **基差崩塌**: 恐慌性对冲需求导致小盘股指期货深度贴水，对冲仓位巨亏。
- **杠杆产品正反馈**: DMA 产品被动平仓卖出小盘股，加剧了“下跌-基差恶化-再平仓”的负反馈循环。

### 预警与控制

- **预警信号**: 小盘股交易拥挤度、场外杠杆产品（DMA）规模激增、期货高贴水模式的脆弱性。
- **控制手段**:
  - **管理人**: 严格限制杠杆；在模型中加入流动性与基差风险考量。
  - **监管层**: 事后限制融券、引导资金入市；事前可加强对场外杠杆的规范。

## 附录

### 附录一：关于 $T_k = k$ 的证明

在本小节中，我们对之前的结论进行相关证明。

### 1. 建立一维随机游走模型

当蚂蚁已经访问了  $k$  个连续的点后，它的位置必然在这  $k$  个点的两个端点之一。为了访问一个新点（第  $k+1$  个点），它必须最终走出这个已访问的区间。

我们可以将这个过程的建模为一个在有限直线上进行的一维随机游走问题，具体来说：

- 假设已访问的  $k$  个点在数轴上被标记为  $1, 2, 3, \dots, k$ 。
- 那么未访问的新点就是  $0$  和  $k+1$ 。这两个点是我们的“吸收壁”。
- 蚂蚁的“游戏”是在点  $\{1, 2, \dots, k\}$  上进行的。当它到达  $0$  或  $k+1$  时，“游戏”结束。
- 根据对称性，不妨假设蚂蚁当前位于端点  $1$ 。

### 2. 建立期望步数的递推关系

我们要求的是从点  $1$  出发，首次到达点  $0$  或点  $k+1$  所需的期望步数。

- 设  $E_i$  为从点  $i$  ( $i \in \{1, 2, \dots, k\}$ ) 出发，首次到达  $0$  或  $k+1$  所需的期望步数。我们的目标是计算  $E_1$ 。
- 对于任何内部点  $i$  ( $1 < i < k$ )，从该点走一步，有  $\frac{1}{2}$  的概率到达  $i-1$ ，有  $\frac{1}{2}$  的概率到达  $i+1$ 。因此，其期望步数满足以下关系：

$$E_i = 1 + \frac{1}{2}E_{i-1} + \frac{1}{2}E_{i+1}$$

- **边界条件：**当蚂蚁到达  $0$  或  $k+1$  时，它已经访问了新点，“游戏”结束。因此，从这些点出发到达目标所需的步数为  $0$ 。

$$E_0 = 0 \quad \text{且} \quad E_{k+1} = 0$$

### 3. 求解递推方程

对于  $E_i$  的通用递推关系，我们可以改写为  $E_{i+1} - E_i = E_i - E_{i-1} - 2$ 。这是一个二阶常系数非齐次差分方程，其通解形式为一个二次多项式。

设解的形式为  $E_i = Ai^2 + Bi + C$ 。代入递推关系可以解得  $A = -1$ 。所以解的一般形式是：

$$E_i = -i^2 + Bi + C$$

现在我们用边界条件  $E_0 = 0$  和  $E_{k+1} = 0$  来确定系数  $B$  和  $C$ ，其中  $B = K + 1$  和  $C = 0$  是待定系数。因此，我们得到了期望步数的精确通解公式：

$$E_i = -i^2 + (k+1)i$$

### 4. 计算最终结果

我们的蚂蚁从端点  $1$  出发，所以我们要求解的是  $E_1$ ：

$$E_1 = 1 \cdot (k+1-1) = k$$

类似的，我们也可以计算从另一个端点  $k$  出发的情况，即求解  $E_k$ ：

$$E_k = k \cdot (k+1-k) = k$$

## 结论

无论蚂蚁位于已发现的  $k$  个连续点的哪个端点（位置 1 或位置  $k$ ），它成功访问到一个新点（即另一个端点外侧的节点）的期望步数恰好为  $k$  步，证毕。

## 附录二：正态分布性质的证明

为了证明对于一个服从均值为 0、方差为  $\sigma^2$  的正态分布的随机变量  $Z$ （即  $Z \sim N(0, \sigma^2)$ ），其在  $Z > 0$  条件下的期望值为  $\sigma\sqrt{\frac{2}{\pi}}$ ，我们可以从条件期望的根本定义出发。根据定义，条件期望  $E[Z|Z > 0]$  等于在  $Z > 0$  区间上  $z$  与其概率密度函数  $f(z)$  乘积的积分，再除以该事件发生的总概率  $P(Z > 0)$ 。

首先，我们确定分母  $P(Z > 0)$  的值。根据对称性， $P(Z > 0) = \frac{1}{2}$ 。

接着，我们计算分子，也就是积分  $\int_0^\infty z f(z) dz$ 。将  $Z$  的概率密度函数  $f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$  代入，我们需要计算的积分是  $\int_0^\infty z \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} dz$ 。

为求解此积分，我们可以采用换元法，令  $u = \frac{z^2}{2\sigma^2}$ 。通过对  $u$  进行微分，我们得到  $du = \frac{z}{\sigma^2} dz$ ，这意味着表达式中的  $z dz$  可以被替换为  $\sigma^2 du$ 。积分的下限在  $z = 0$  时为  $u = 0$ ，上限在  $z \rightarrow \infty$  时  $u \rightarrow \infty$ ，因此积分限保持不变。代入后，原积分化为  $\frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty e^{-u} (\sigma^2 du)$ ，简化后得到  $\frac{\sigma}{\sqrt{2\pi}} \int_0^\infty e^{-u} du$ 。由于标准积分  $\int_0^\infty e^{-u} du$  的计算结果为 1，因此整个分子的值就是  $\frac{\sigma}{\sqrt{2\pi}}$ 。

最后，我们将分子和分母的结果合并，便可得到最终的条件期望：

$$E[Z|Z > 0] = \sigma\sqrt{\frac{2}{\pi}}$$

证毕。