2. 甲乙两人进行比赛，设每局比赛甲胜的概率是 p，乙胜的概率是 q，平局的概率是 r，有 p + q + r = 1。设每局比赛后，胜者记‘+1’，负者记‘-1’，平局不记分。当两人中一人积到 2 分时，比赛结束。用 X_n 表示比至第 n 局结束，甲获得的分数，则序列 $\{X_1, X_2, \cdots\}$为一个马尔可夫过程。

1） 请给出状态空间和状态转移矩阵。

2） 问在甲积 1 分的情况下，恰好再赛两局可以结束比赛的概率是多少？

(1) 状态空间 $S=\{-2, -1, 0, 1, 2\}$

状态转移矩阵 $P=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & r & p & 0 & 0 \\ 0 & q & r & p & 0 \\ 0 & 0 & q & r & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

(2) 记 $p_{s,k}$ 为第k局结束时甲分数为S的概率

则 $v_k=[p_{-2,k} \quad p_{-1,k} \quad p_{-0,k} \quad p_{1,k} \quad p_{2,k}]$

由题意得 $X_0=1 \qquad v_0=[0 \quad 0 \quad 0 \quad 1 \quad 0]$

$v_1=v_0P=[0 \quad 0 \quad q \quad r \quad p]$

$v_2=v_1P=[0 \quad q^2 \quad 2qr \quad pq+r^2 \quad pr+p]$

两局结束的概率为 $p_{2,2}=pr+p$

其中恰两局概率为 pr

对应的状态路径为 $1 \longrightarrow 1 \longrightarrow 2$

3. 考虑下方一个 3 × 3 网格图

0	1	2
3	4	5
6	7	8

非终止状态集合 S = {1, 2, …, 7}。每个状态有四种可能的动作{up、down、left、right}，对于每次转移 $R_t = -1$，每个动作会导致状态转移，但当动作会导致智能体移出网格时，状态保持不变。γ = 1，若 π 是等概率随机策略，那么行动价值 $q_\pi(4, \text{left})$、$q_\pi(7, \text{right})$是多少?

3. $v_\pi(0)=v_\pi(8)=0$

$v_\pi(1)=v_\pi(3)=v_\pi(5)=v_\pi(7)$

$v_\pi(2)=v_\pi(6)$

$v_\pi(1)=0.25(-1+v_\pi(1))+0.25(-1+v_\pi(4))+0.25(-1+v_\pi(0))+0.25(-1+v_\pi(2))$

$=-1+0.25v_\pi(1)+0.25v_\pi(2)+0.25v_\pi(4)$

$v_\pi(2)=0.25(-1+v_\pi(2))+0.25(-1+v_\pi(5))+0.25(-1+v_\pi(1))+0.25(-1+v_\pi(2))$

$=-1+0.5v_\pi(1)+0.5v_\pi(2)$

$v_\pi(4)=0.25(-1+v_\pi(1))+0.25(-1+v_\pi(3))+0.25(-1+v_\pi(5))+0.25(-1+v_\pi(7))$

$=-1+v_\pi(1)$

解得 $v_\pi(1) = -7$ $v_\pi(2) = -9$ $v_\pi(4) = -8$

$q_\pi(4, left) = r_4^{left} + v_\pi(3) = -1-7 = -8$

$q_\pi(7, right) = r_7^{right} + v_\pi(8) = -1$