Homework6

2

(1)

解:

设甲的分数为 X。比赛结束时，甲的分数可能为 $+2$（甲胜）或 -2（乙胜）。

因此，状态空间 $S = \{-2, -1, 0, 1, 2\}$。其中 -2 和 2 是终止状态。

状态转移矩阵 P 如下所示，行和列的顺序对应状态 $\{-2, -1, 0, 1, 2\}$：

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & r & p & 0 & 0 \\ 0 & q & r & p & 0 \\ 0 & 0 & q & r & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(2)

解:

甲当前积 1 分 (状态为 1)。比赛在恰好两局后结束，意味着第一局比赛没有结束，第二局比赛结束了。

第一局后比赛未结束：则甲不能获胜概率 p，否则并非再赛两局结束；若甲平局：概率 r，其后甲胜一局恰好结束比赛；若甲输：概率 q，两者比分变为 0:0, 不可能在接下来第二局结束比赛。

因此，在甲积 1 分的情况下，恰好再赛两局可以结束比赛的唯一方式是：第一局平局，第二局甲胜。其概率为

$$P(\text{结束于第二局}|X_0 = 1) = P(X_1 = 1|X_0 = 1) \times P(X_2 = 2|X_1 = 1) = r \times p$$

4

(1)

解:

给定 $R_A = -1, R_B = -1, R_C = -1, R_D = 0$ 和 $\gamma = 0.5$。

转移概率：

- 对 A：$P_{AB} = 0.5, P_{AC} = 0.5$
- 对 B：$P_{BB} = 0.5, P_{BD} = 0.5$
- 对 C：$P_{CA} = 0.5, P_{CD} = 0.5$
- 对 D：$P_{DD} = 1.0$

由状态价值的定义:

$$
\begin{aligned}
V(A) &= -1 + 0.5 \times (0.5V(B) + 0.5V(C)) \\
V(B) &= -1 + 0.5 \times (0.5V(B) + 0.5V(D)) \\
V(C) &= -1 + 0.5 \times (0.5V(A) + 0.5V(D)) \\
V(D) &= 0 + 0.5 \times (1.0V(D)) \implies V(D) = 0
\end{aligned}
$$

将 $V(D) = 0$ 代入：

联立解得:

$$
\begin{cases}
V(A) = -\frac{76}{45} \\
V(B) = -\frac{4}{3} \\
V(C) = -\frac{64}{45} \\
V(D) = 0
\end{cases}
$$

(2)

解:

当马尔可夫回报过程的状态空间非常大时，直接求解线性方程组（如第一部分所示）变得不切实际。在这种情况下，可以采用迭代算法来近似计算状态价值。

其中一种求解思路价值迭代 (Value Iteration) 如下所示：

- **核心思想**：该方法通过一系列的迭代来逼近真实的状态价值。在每一次迭代中，它使用前一次迭代得到的价值函数来计算当前所有状态的更新价值，这个更新过程基于状态价值的线性方程组。持续进行迭代，直到价值函数的变化足够小，即价值函数收敛。

- **初始化阶段**：为所有非终止状态 $s \in S$ 任意赋予一个初始价值 $V_0(s)$。对于题目中的终止状态 D，其价值是已知的，即 $V_k(D) = 0$ 对于所有迭代次数 k 都成立。通常可以将所有非终止状态的初始价值设为 0。

- **迭代更新规则**：对于每一次迭代 $k = 0, 1, 2, \ldots$，对每一个状态 s (在本题中为 A, B, C)，同步更新其价值 $V_{k+1}(s)$。更新规则基于即时回报 R_s 和后继状态的期望折现价值 $\gamma \sum_{s'} P_{ss'} V_k(s')$：

$$V_{k+1}(s) = R_s + \gamma \sum_{s'} P_{ss'} V_k(s')$$

  具体到本题中的状态 A, B, C（其中 $R_A = R_B = R_C = -1$, $R_D = 0$, $\gamma = 0.5$）：

$$\begin{aligned}
&V_{k+1}(A) = -1 + 0.5\left(0.5V_k(B) + 0.5V_k(C)\right) \\
&V_{k+1}(B) = -1 + 0.5\left(0.5V_k(B) + 0.5V_k(D)\right) = -1 + 0.25V_k(B) \\
&V_{k+1}(C) = -1 + 0.5\left(0.5V_k(A) + 0.5V_k(D)\right) = -1 + 0.25V_k(A) \\
&V_{k+1}(D) = 0
\end{aligned}$$

- **终止条件**：迭代过程持续进行，直到某次迭代后所有状态的价值变化量都小于一个预设的阈值 $\epsilon > 0$。即，当满足以下条件时停止迭代：

$$\max_{s \in S} |V_{k+1}(s) - V_k(s)| < \epsilon$$

  此时，$V_{k+1}(s)$ 就作为各状态价值的近似解。