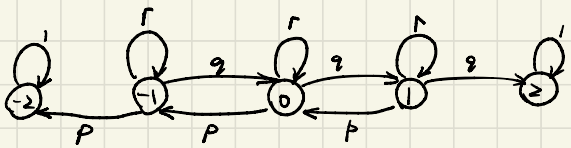
HW7

支一轩 新雅一智1 2020010038

1. a) 状态空间 $S=\{-2,-1,0,1,2\}$

状态转移矩阵



	-2	-1	0	1	2
-2	1	0	0	0	0
-1	p	r	q	0	0
0	0	p	r	q	0
1	0	0	p	r	q
2	0	0	0	0	1

b) 已知1分的情况下，再赢2局的结果

① → ① → ③ 只有这一种可能

$P=r\cdot q$

2. a) $V(A)=-1+\gamma[P_{AB}V(B)+P_{AC}V(C)]=0.1V(B)+0.1V(C)-1$

$V(B)=-1+\gamma[P_{BB}V(B)+P_{BD}V(D)]=0.1V(B)+0.1V(D)-1$

$V(C)=-1+\gamma[P_{CA}V(A)+P_{CD}V(D)]=0.1V(A)+0.1V(D)-1$

$V(D)=0+\gamma[P_{DD}V(D)]=0.2V(D)$

有 $\begin{cases} V(A)=-1.2233 \\ V(B)=-10/9\approx-1.1111 \\ V(C)=-1.1223 \\ V(D)=0 \end{cases}$

	A	B	C	D
A	0	0.5	0.5	0
B	0	0.5	0	0.5
C	0.5	0	0	0.5
D	0	0	0	1

b) 1° 价值迭代

i 初始化 $V(s)$ 均为0

ii 反复执行 $V(s)\leftarrow R(s)+\gamma\sum_{s'}P_{ss'}V(s')$

iii 当 $\Delta V(s)<$ threshold，可认为已收敛，停止迭代

2° 策略迭代

i 初始化任一策略 π

ii 反复执行 $V_\pi(s)=R(s)+\gamma\sum_{s'}P_{ss'}V_\pi(s')$（计算当前策略下 $V(s)$）

$\pi(s)\leftarrow\arg\max_a[R(s,a)+\gamma\sum_{s'}P_{ss'}V_\pi(s')]$（用新状态价值函数更新策略）

iii 同上至收敛.

3. $q_{TV}[4,left] = R_b + V_{TV}[3]$

$q_{TV}[7,right] = R_t + V_{TV}[0]$

下用 V_i 计 $V_{TV}[i]$

$V_0 = 0$

$V_1 = R_t + \frac{1}{4}(V_0 + V_1 + V_2 + V_4)$

$V_2 = R_t + \frac{1}{4}(V_1 + 2V_2 + V_5)$

$V_3 = R_t + \frac{1}{4}(V_0 + V_3 + V_4 + V_6)$

$V_4 = R_t + \frac{1}{4}(V_1 + V_3 + V_5 + V_7)$

$V_5 = R_t + \frac{1}{4}(V_2 + V_4 + V_5 + V_8)$

$V_6 = R_t + \frac{1}{4}(V_3 + 2V_6 + V_7)$

$V_7 = R_t + \frac{1}{4}(V_4 + V_6 + V_7 + V_8)$

$V_8 = 0$

解得 $V_{TV}[3] = V_3 = -7$

$V_{TV}[0] = V_0 = 0$

故有 $q_{TV}[4,left] = R_b + V_{TV}[3] = -8$

$q_{TV}[7,right] = R_t + V_{TV}[0] = -1$