مسئله‌ی ۱.

حل.

الف) حرکت بالا $\rightarrow V = -50 + \gamma + \gamma^2 + \cdots + \gamma^{100}$

$$= -50 + \frac{1-\gamma^{101}}{1-\gamma}$$

حرکت پایین $\rightarrow V = 50 - (\gamma + \gamma^2 + \cdots + \gamma^{100})$

$$= 50 - \left(\frac{1-\gamma^{101}}{1-\gamma}\right)$$

ب) $$-50 + \frac{1-\gamma^{101}}{1-\gamma} > 50 - \frac{1-\gamma^{101}}{1-\gamma}$$

$$\Rightarrow \frac{1-\overset{0}{\cancel{\gamma^{101}}}}{1-\gamma} > 50 \Rightarrow 50\gamma > 49 \Rightarrow$$

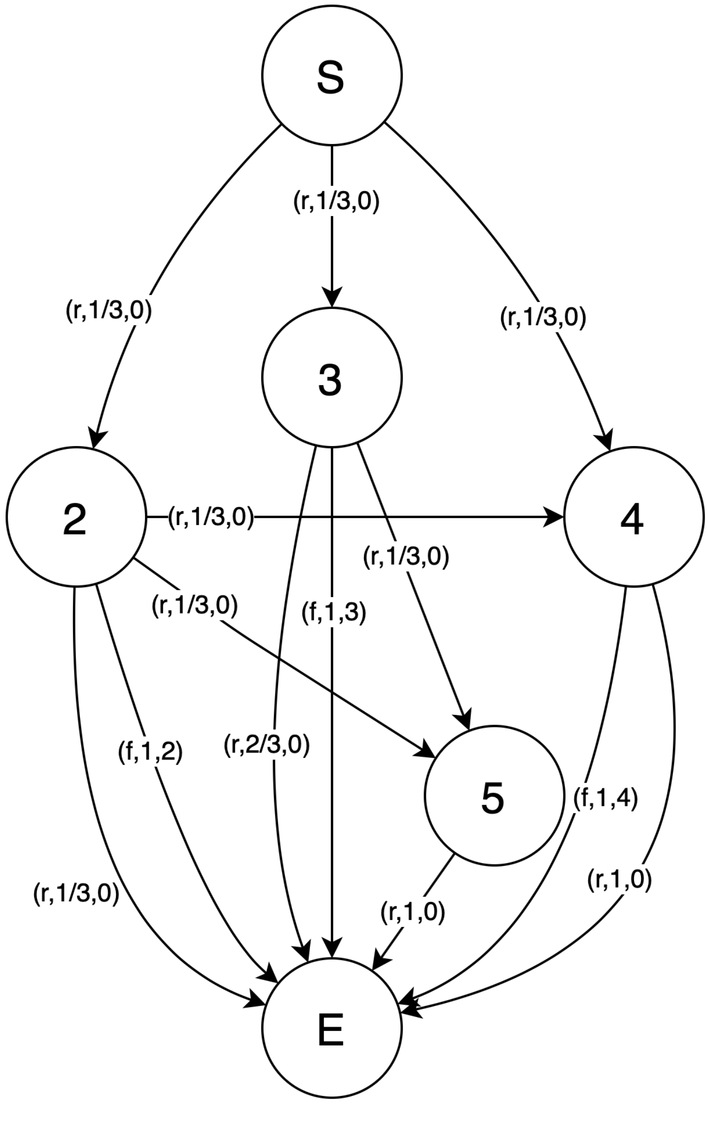
$$\boxed{1 > \gamma > 0.98}$$

▷

مسئله‌ی ۲.

حل.

۱. در این سوال فرض شده است که در ابتدای بازی حتما باید کارت کشیده شود. S استیت شروع و E استیت پایانی است. r اکشن کارت کشیدن و f اکشن پایان بازی است. بر روی هر فلش در FSM زیر ابتدا اکشن، سپس احتمال گذار و در نهایت پاداش دریافتی مشخص شده است.



۲. مقادیر نهایی استیت‌ها و سیاست بهینه را در شکل زیر مشاهده می‌کنید.

$$v_5 = \max \begin{cases} \text{action}=r \rightarrow 1\times(0+0)=0 \\ \text{action}=f \rightarrow 1\times(0+5)=\boxed{5} \end{cases} \quad \text{best action} = f$$

$$v_4 = \max \begin{cases} \text{action}=r \rightarrow 1\times(0+0)=0 \\ \text{action}=f \rightarrow 1\times(0+4)=\boxed{4} \end{cases} \quad \text{best action} = f$$

$$v_3 = \max \begin{cases} \text{action}=r \rightarrow \frac{1}{3}v_5 + \frac{2}{3}\times 0 = \frac{5}{3} \\ \text{action}=f \rightarrow 1\times 3=\boxed{3} \end{cases} \quad \text{best action} = f$$

$$v_2 = \max \begin{cases} \text{action}=r \rightarrow \frac{1}{3}v_4 + \frac{1}{3}v_5 + \frac{1}{3}\times 0 = \boxed{3} \\ \text{action}=f \rightarrow 1\times 2 = 2 \end{cases} \quad \text{best action} = r$$

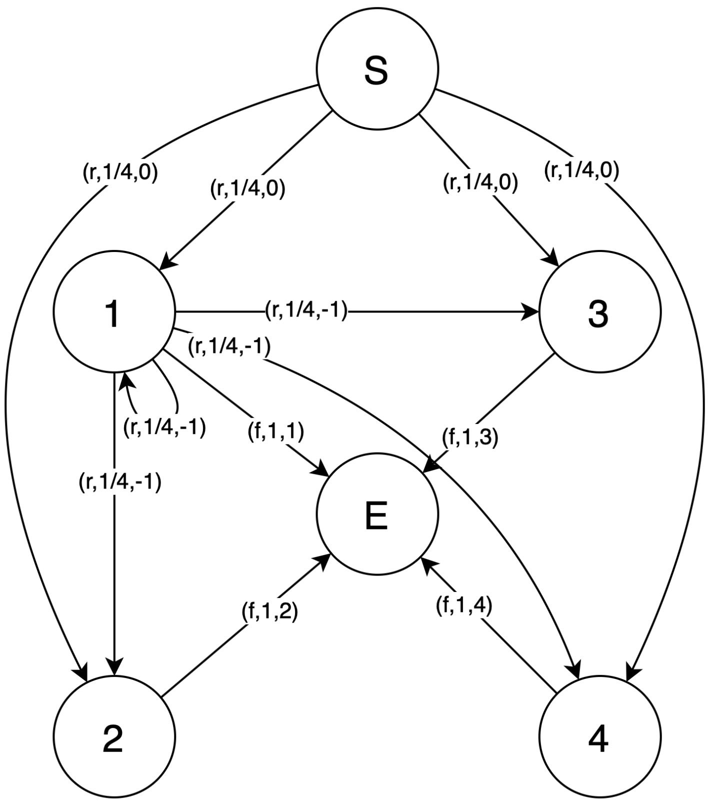
۳. با توجه به این‌که طول طولانی‌ترین مسیر در FSM بالا برابر با ۳ است، به هر ترتیبی که مقدار حالت‌ها را محاسبه کنیم، در سه iteration همگرا می‌شوند. اگر با ترتیب بالا مقادیر را محاسبه کنیم در یک iteration همگرا می‌شوند.

▷

مسئله‌ی ۳.

حل.

۱. با توجه به این‌که می‌توان برای این سوال یک FSM رسم کرد و توابع گذار و پاداش تعریف کرد، این مسئله را می‌توان به صورت یک مسئله‌ی MDP تعریف کرد. شکل زیر FSM این مسئله را نشان می‌دهد. S استیت شروع و E استیت پایانی است. r اکشن تولید عدد و f اکشن پایان بازی است. بر روی هر فلش در FSM زیر ابتدا اکشن، سپس احتمال گذار و در نهایت پاداش دریافتی مشخص شده است. برای سادگی شکل اکشن r تنها برای استیت اول مشخص شده است ولی برای باقی استیت‌ها دقیقا مشابه استیت اول خواهد بود.



۲. جدول زیر مقادیر استیت‌ها را به روش Value Iteration با سه مرحله نشان می‌دهد. برای هر استیت مقدار value برای هر اکشن و مقدار به روز رسانی شده نوشته شده است.

Iteration	State	Action r	Action f	Updated Value
1	1	-1	1	1
	2	-1	2	2
	3	-1	3	3
	4	-1	4	4
2	1	1.25	1	1.25
	2	1.25	2	2
	3	1.25	3	3
	4	1.25	4	4
3	1	1.3	1	1.3
	2	1.3	2	2
	3	1.3	3	3
	4	1.3	4	4

۳. جدول زیر مقادیر استیت‌ها و سیاست بهینه را به روش Policy Iteration با سه مرحله نشان می‌دهد.

Iteration 1:

$\boxed{V_1^* = 1}$ $\boxed{V_2^* = 2}$

$$\left.\begin{aligned} V_3^* &= -1 + \tfrac{1}{4} \times 0.9 (V_1^* + V_2^* + V_3^* + V_4^*) \\ V_4^* &= -1 + \tfrac{1}{4} \times 0.9 (V_1^* + V_2^* + V_3^* + V_4^*) \end{aligned}\right\} \Rightarrow V_3^* = V_4^*$$

$$\Rightarrow V_3^* = -1 + \tfrac{1}{4} \times 0.9 (3 + 2V_3^*) \Rightarrow \boxed{V_3^* = 0.59}$$

$$\boxed{V_4^* = 0.59}$$

$\boxed{Action_1 \rightarrow \ell}$ $\boxed{Action_2 \rightarrow \ell}$ $\boxed{Action_3 \rightarrow \ell}$ $\boxed{Action_4 \rightarrow \ell}$

Iteration 2:

$\boxed{V_1^* = 1}$ $\boxed{V_2^* = 2}$ $\boxed{V_3^* = 3}$ $\boxed{V_4^* = 4}$

$\boxed{Action_1 \rightarrow r}$ $\boxed{Action_2 \rightarrow \ell}$ $\boxed{Action_3 \rightarrow \ell}$ $\boxed{Action_4 = \ell}$

Iteration 3:

$\boxed{V_2^* = 2}$ $\boxed{V_3^* = 3}$ $\boxed{V_4^* = 4}$

$$V_1^* = -1 + \tfrac{1}{4} \times 0.9 (V_1^* + V_2^* + V_3^* + V_4^*) \Rightarrow \boxed{V_1^* = 1.32}$$

$\boxed{Action_1 \rightarrow r}$ $\boxed{Action_2 \rightarrow \ell}$ $\boxed{Action_3 \rightarrow \ell}$ $\boxed{Action_4 \rightarrow \ell}$

▷