$V_\pi(s) = E_\pi[G_t | S_t = s] = E_\pi[R_{t+1} + \gamma R_{t+2} + \dots | S_t = s]$

(1) 二 Markov 性： $P(S_{t+1} | S_t = s)$ 与 t 无关

$$= \sum_{a \in A} \pi(a|s) E_\pi[R_{t+1} + \gamma V_\pi(S_{t+1}) | S_t = s]$$

$$= \sum_{a \in A} \pi(a|s) \sum_{i=0}^{\infty} \gamma^i E_\pi[R_{t+i+1} | S_t = s]$$

$$= \sum_{a \in A} \pi(a|s) \sum_{i=0}^{\infty} \gamma^i \sum_{s' \in S} P[S_{t+i+1} = s' | S_t = s] \cdot E_\pi[R_{t+i+1} | S_{t+i+1} = s']$$

① 证明 $E_\pi[R_t | S_t]$ 与 t 无关

② 证明 $E_\pi[R_{t+i+1} | S_t = s]$ 与 t 无关

$E_\pi[R_{t+i+1} | S_{t+j+1}]$ 与 t 无关 $(j \le i)$

则 $R[t+i+1 | S_{t+j}] = \sum_{s' \in S} P[S_{t+j+1} = s' | S_{t+j}] \cdot E_\pi[R_{t+j+1} | S_{t+j+1} = s']$

$P[S_{t+j+1} = s' | S_{t+j}]$ 无关 $\Rightarrow$ 与 t 无关

$E_\pi[R_{t+j+1} | S_{t+j+1} = s']$ 无关

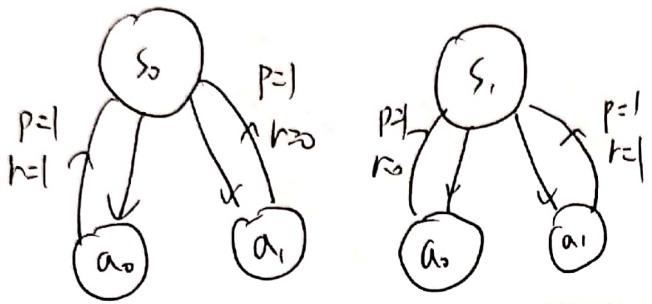
∴ 证明，$\forall j \ge 0$，

有 $R[t+i+1 | S_t]$ 与 t 无关

综上，$V_\pi(s) = E_\pi[G_t | S_t = s]$ 与 t 无关，

等价于 $V_\pi(s) = E_\pi[G_0 | S_0 = s]$

(2)



π_0: $\pi_0(a_0|s_0)=1$
$\pi_0(a_1|s_0)=0$
$\pi_0(a_0|s_1)=0$
$\pi_0(a_1|s_1)=0$

π_1: $\pi_1(a_0|s_0)=0$
$\pi_1(a_1|s_0)=1$
$\pi_1(a_0|s_1)=0$
$\pi_1(a_1|s_1)=1$

$V_{\pi_0}(s_0)=1+\gamma+\gamma^2+\cdots$

$=\lim_{n\to\infty} 1\cdot\frac{1-\gamma^n}{1-\gamma}=\frac{1}{1-\gamma}$

$V_{\pi_0}(s_1)=0$

$V_{\pi_1}(s_0)=\infty$

$V_{\pi_1}(s_1)=1+\gamma+\gamma^2+\cdots$

$=\lim_{n\to\infty} 1\cdot\frac{1-\gamma^n}{1-\gamma}=\frac{1}{1-\gamma}$

(3)

证明:

$V_*(s)$

$=\max_a\left(R_s^a+\gamma\sum_{s'\in S}P_{ss'}^a V_*(s')\right)$

$V_{\pi_0}(s_0)>V_{\pi_1}(s_0)$

$V_{\pi_0}(s_1)<V_{\pi_1}(s_1)$

相悖于

(3) ① 先证 $\max_{a}(\forall s\in S,\ \mathbb{E}_{a\sim\pi'(s)}[Q_\pi(s,a)]\geq \mathbb{E}_{a\sim\pi(s)}[Q_\pi(s,a)]) \Rightarrow (\forall s\in S, V_{\pi'}(s)\geq V_\pi(s))$

$$V_\pi(s)=\mathbb{E}_{a\sim\pi(s)}[Q_\pi(s,a)]\leq \mathbb{E}_{a\sim\pi'(s)}[Q_\pi(s,a)]$$

$$=\mathbb{E}_{a\sim\pi'(s)}[r(s,a)+\gamma V_\pi(s_1)\mid s_0=s]$$

$$=\mathbb{E}_{a\sim\pi'(s)}[r(s,a)+\gamma \mathbb{E}_{a_1\sim\pi(s_1)}[Q_\pi(s,a_1)]\mid s_0=s]$$

$$\leq \mathbb{E}_{a\sim\pi'(s)}[r(s,a)+\gamma \mathbb{E}_{a_1\in\pi'(s_1)}[Q_\pi(s_1,a_1)]\mid s_0=s]$$

$$=\mathbb{E}_{\substack{a\sim\pi'(s)\\ a_1\sim\pi'(s)}}[r(s,a)+\gamma r(s_1,a_1)+\gamma^2 V_\pi(s_2)\mid s_0=s]$$

$\Leftarrow$ 迭代得到，对 $\forall\pi'$!

$$V_\pi(s)\leq \mathbb{E}_{a_t\sim\pi'(s_t)}\left[\sum_{t=0}^{+\infty}\gamma^t\mathbb{E}[r(s_t,a_t)]+\gamma^{t+1}V_\pi(s_{t+1})\mid s_0=s\right]$$

$$=V_{\pi'}(s)$$

② 再证 π^* 为 optimal：

π^*：对 $\forall(s,a)\in S\times A$，有 $a\in\operatorname{argmax}_{a'\in A} Q_\pi(s,a')$

证明：若 $\exists(s_0,a_0)$ 使 $a_0\notin\operatorname{argmax}_{a'\in A} Q_\pi(s_0,a')$，则不是最优

则可构造 π'：

$$\begin{cases}\pi'(s)=\pi(s) & s\neq s_0\\ \pi'(s_0)\in\operatorname{argmax}_{a'\in A} Q_\pi(s_0,a') & \end{cases}$$

则对 $\forall s$，$\mathbb{E}_{a\in\pi'(s)}[Q_\pi(s,a)]\geq \mathbb{E}_{a\in\pi(s)}[Q_\pi(s,a)]$，

则有 $\forall s\in S$，$V_{\pi'}(s)\geq V_\pi(s)$

不存在比 π^* 更优的策略

同理，对于一个非最优策略 π，我们总能构造 π' 使 $V_{\pi'}(s)>V_\pi(s)$

最优有：$\forall s\in S$，$\pi^*(s)=\operatorname{argmax}_{a\in A} Q^*(s,a)$ 为最优