清華大學
TSINGHUA UNIVERSITY

1.(1) $E_\pi[G_t|S_t=s] = \cancel{\sum_{a\in A}\pi(a|s)(R_s^a + r\sum_{s'\in S} P_{ss'}^a V_\pi(s'))}$

$\cancel{= E_\pi[G_t|S_t]}$ $E_\pi[R_{t+1} + r G_{t+1} | S_t=s]$

$= \sum_{a\in A}\pi(a|s) E_\pi[G_t|S_t=s, A_t=a]$

$= R_s + r E_\pi[G_{t+1}|S_t=s]$

$= R_s + r\sum_{s'\in S} P_{ss'} V_\pi(s')$

$= \sum_{a\in A}\pi(a|s)R_s^a + \pi(a|s) r \sum_{s'\in S} P_{ss'}^a$ $E_\pi[G_{t+1}|S_t=s], A_t=a$

$E_0[G_t|S_t=s]$

$= \sum_{a\in A}\pi(a|s)\ E_\pi[G_0|S_0=s, A_0=a]$

只需证 $\forall s,a\in S,A$, $E_\pi[G_t|S_t=s,A_t=a]$

$= E_\pi[G_0|S_0=s,A_0=a]$

$E_\pi[\cancel{G_0|S_0=s}] = \sum_{a\in A}\pi(a|s)R_s^a + \pi(a|s)\, r\sum_{s'\in S}P_{ss'}^a E_\pi[G_1|S_1=s', A_1=a]$

$G_t = R_{t+1} + \cdots = \sum_{k=0}^{\infty} r^k R_{t+k+1}$, 只需证 对 $\forall s'\in S$, $E_\pi[G_{t+1}|S_{t+1}=s', A_{t+1}=a]$

$= E_\pi[G_1|S_1=s', A_1=a]$

$G_0 = R_1 + \cdots = \sum_{k=0}^{\infty} r^k R_{k+1}$

同理展开两式，可发现只需证

对 $\forall s'\in S$, $E_\pi[G_{t+1}|S_{t+1}=s', A_t=a]$

$= E_\pi[G_1|S_1=s', A_1=a]$

$E_\pi[G_t|S_t=s, A_t=a]$

$= E_\pi[R_{t+1}|S_t=s,A_t=a] + rE_\pi[G_{t+1}|S_t=s,A_t=a]$

$= R_s^a + r\sum_{s'\in S}\sum_{a'\in A}\pi(a'|s')P_{ss'}^{a'} E_\pi[G_{t+1}|S_{t+1}=s', A_{t+1}=a']$

$E_\pi[G_1|S_t=s, A_t=a] = R_s^a + r\sum_{s'\in S}\sum_{a'\in A}\pi(a'|s')P_{ss'}^{a'} E_\pi[G_1|S_1=s',A_1=a']$

只需证对 $\forall s',a'\in S,A$, $E_\pi[G_{t+1}|S_{t+1}=s',A_{t+1}=a']$

$= E_\pi[G_1|S_1=s',A_1=a']$

递推下去，只需证

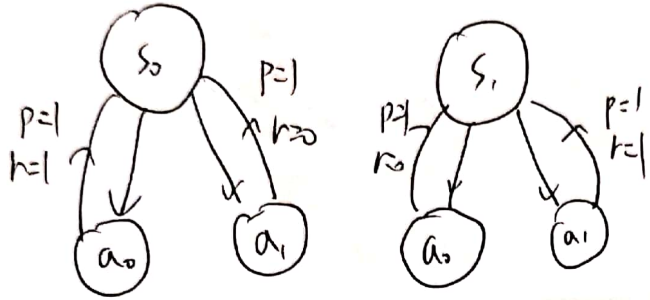
$\forall s',a'\in S,A$, $E_\pi[G_{2t}|S_{2t}=s', A_{2t}=a']$

$= E_\pi[G_t|S_t=s',A_t=a']$

$E_\pi[G_t|S_t=s]$ 按定义(1) $\forall t'\geq t$, $E_\pi[G_{t'}|S_{t'}=s',A_{t'}=a']$

$= E_\pi[G_0|S_0=s]$ ← 显然成立 $= q_\pi(s',a')$

(2)



π_0: $\pi_0(a_0|s_0)=1$

$\pi_0(a_1|s_0)=0$

$\pi_0(a_0|s_1)=1$

$\pi_0(a_1|s_1)=0$

π_1: $\pi_1(a_0|s_0)=0$

$\pi_1(a_1|s_0)=1$

$\pi_1(a_0|s_1)=0$

$\pi_1(a_1|s_1)=1$

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

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

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

$V_{\pi_1}(s_0)=0$

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

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

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

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

非偏序

(3)

证明:

$V_*(s)$

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

(3) 证明：

设 π_1, π_2，

对 $s \in S_1$，$\pi_1(s) < \pi_2(s)$

$s \notin S_1$，$\pi_1(s) \geq \pi_2(s)$

则：令 $\pi_3 = \begin{cases} s \in S_1: \pi_3(s) = \pi_2(s) \\ s \notin S_1: \pi_3(s) = \pi_1(s) \end{cases}$

$$V_{\pi_1}(s) = \sum_{a \in A} \pi_1(a|s)\left(R_s^a + r \sum_{s' \in S} P_{ss'}^a V_{\pi_1}(s)\right)$$

$$V_{\pi_2}(s) = \sum_{a \in A} \pi_2(a|s)\left(R_s^a + r \sum_{s' \in S} P_{ss'}^a V_{\pi_2}(s)\right)$$

对 $s \in S_1$：

$$V_{\pi_3}(s) = \sum_{a \in A} \pi_2(a|s)\left(R_s^a + r \sum_{\substack{s' \in S \\ s' \in S_1}} P_{ss'}^a V_{\pi_2}(s) + r \sum_{\substack{s' \notin S_1 \\ s' \notin S_1}} P_{ss'}^a V_{\pi_1}(s)\right)$$

对 $s \notin S_1$

$$V_{\pi_3}(s) = \sum_{a \in A} \pi_1(a|s)\left(R_s^a + r \sum_{\substack{s' \in S \\ s' \in S_1}} P_{ss'}^a V_{\pi_2}(s) + r \sum_{\substack{s' \in S \\ s' \notin S_1}} P_{ss'} V_{\pi_1}(s)\right)$$

对 $\forall s \in S$

$V_{\pi_3}(s) \geq V_{\pi_1}(s)$

$V_{\pi_3}(s) \geq V_{\pi_2}(s)$

4. 对全部 (π_i, π_j) 进行此步骤，可得一个最优策略 π_*，

对 $\forall \pi_i \in \pi$，

$\pi_* \geq \pi_i$

得证