Ex 5.5

first-visit:

$$v_s = 10$$

every-visit:

$$v_s = \frac{1}{10}(1+2+3+4+5+6+7+8+9+10) = 5.5$$

Ex 5.6

$V(s)$:

$$V(s) = \frac{\Sigma_{t \subset \mathcal{T}(s)} \rho_{t:T(t)-1} G_t}{\Sigma_{t \subset \mathcal{T}(s)} \rho_{t:T(t)-1}}$$

$Q(s,a)$:

$$Q(s,a) = \frac{\Sigma_{t \subset \mathcal{T}(s,a)} \rho_{t+1:T(t)-1} G_t}{\Sigma_{t \subset \mathcal{T}(s,a)} \rho_{t+1 t:T(t)-1}}$$

○ πの対応場所

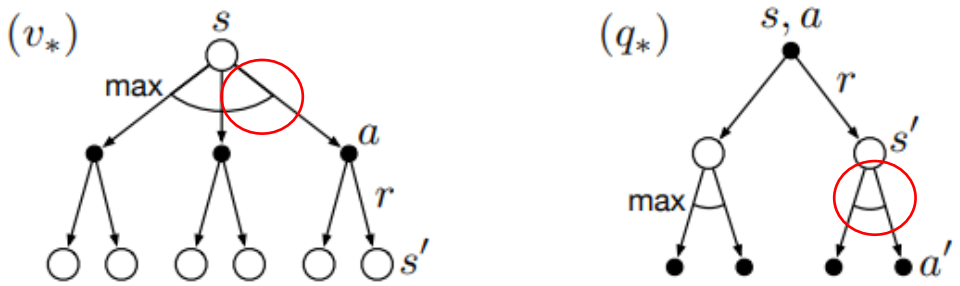

Figure 3.4: Backup diagrams for v_* and q_*

Ex 5.8

$$ans = \mathbb{E}_b\left[\left(\frac{1}{T}\sum_{k=1}^{T}\prod_{t=0}^{k-1}\frac{\pi(A_t|S_t)}{b(A_t|S_t)}G_0\right)^2\right]$$

$$= 0.5(action\ p)\ \cdot 0.1(transition\ p)\cdot\left(\frac{1}{0.5}\right)^2(square)$$

$$+\frac{1}{2}[0.5\cdot 0.9\cdot 0.5\cdot 0.1\cdot 2^{2\cdot 2}+0.5\cdot 0.1\cdot 2^{2\cdot 1}]$$

$$+\frac{1}{3}[0.5^2\cdot 0.9^2\cdot 0.5\cdot 0.1\cdot 2^{2\cdot 3}+0.5\cdot 0.9\cdot 0.5\cdot 0.1\cdot 2^{2\cdot 2}+0.5\cdot 0.1\cdot 2^{2\cdot 1}]$$

$$+\cdots$$

$$= 0.2\sum_{k=1}^{\infty}\frac{1}{k}\sum_{t=0}^{k-1}0.9^t\cdot 2^t$$

$$= 0.2\sum_{k=1}^{\infty}\frac{1}{k}\sum_{t=0}^{k-1}1.8^t$$

$$\geq 0.2\sum_{k=1}^{\infty}\frac{1}{k}\sum_{t=0}^{k-1}1^t$$

$$= 0.2\sum_{k=1}^{\infty}1$$

$$= \infty\ (infinity)$$