## Monte Carlo ES (Exploring Starts), for estimating $\pi \approx \pi_*$

 $\pi(s) \in \mathcal{A}(s)$  (arbitrarily), for all  $s \in \mathcal{S}$ 

 $G \leftarrow \gamma G + R_{t+1}$ 

Initialize:

$$Q(s, a) \in \mathbb{R}$$
 (arbitrarily), for all  $s \in \mathcal{S}$ ,  $a \in \mathcal{A}(s)$   
 $Returns(s, a) \leftarrow \text{empty list, for all } s \in \mathcal{S}$ ,  $a \in \mathcal{A}(s)$ 

Loop forever (for each episode):

Loop forever (for each episode):

Choose 
$$S_0 \in \mathcal{S}$$
,  $A_0 \in \mathcal{A}(S_0)$  randomly such that all pairs have probability  $> 0$ 

Concrete an episode from  $S_0$ ,  $A_0$  following  $\pi$ :  $S_0$ ,  $A_0$ ,  $R_1$ ,  $R_2$ ,  $R_3$ ,  $R_4$ ,  $R_4$ ,  $R_5$ 

Append G to  $Returns(S_t, A_t)$ 

 $\pi(S_t) \leftarrow \operatorname{arg\,max}_a Q(S_t, a)$ 

Choose 
$$S_0 \in \mathcal{S}$$
,  $A_0 \in \mathcal{A}(S_0)$  randomly such that  $S_0$  Generate an episode from  $S_0, A_0$ , following  $\pi$ :  $S_0$ 

Generate an episode from  $S_0, A_0$ , following  $\pi: S_0, A_0, R_1, \ldots, S_{T-1}, A_{T-1}, R_T$ 

 $Q(S_t, A_t) \leftarrow \text{average}(Returns(S_t, A_t))$ 

Generate an episode from 
$$S_0, A_0$$
, following  $\pi$ :  $S_0$ 

$$G \leftarrow 0$$

Generate an episode from 
$$S_0, T_0$$
, is nowing  $\pi$ .  $S_0, T_0$   
 $G \leftarrow 0$   
Loop for each step of episode,  $t = T - 1, T - 2, \dots, 0$ :

e. 
$$t = T - 1, T - 2, \dots$$