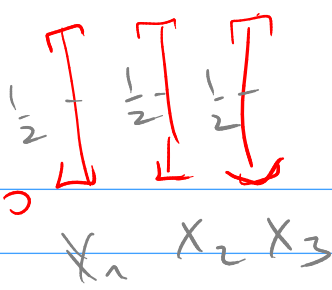
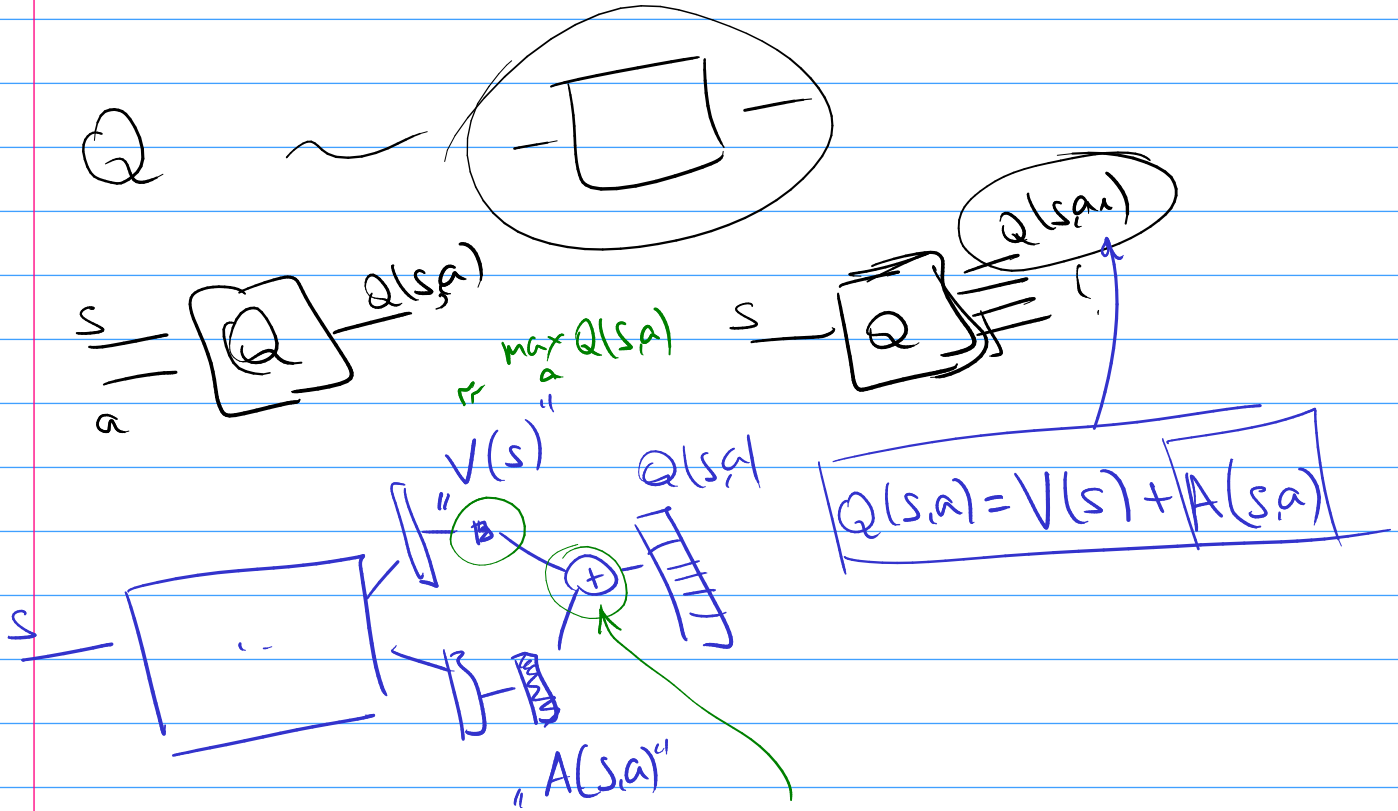


Winner's  
curse



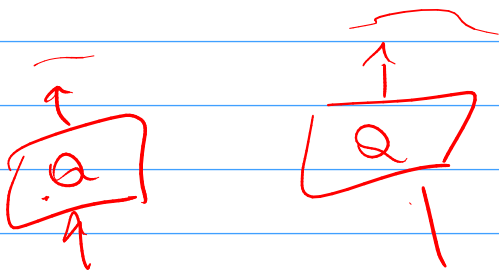
$$\max_i E[X_i] = \frac{1}{2}$$

$$E[\max X_i]$$

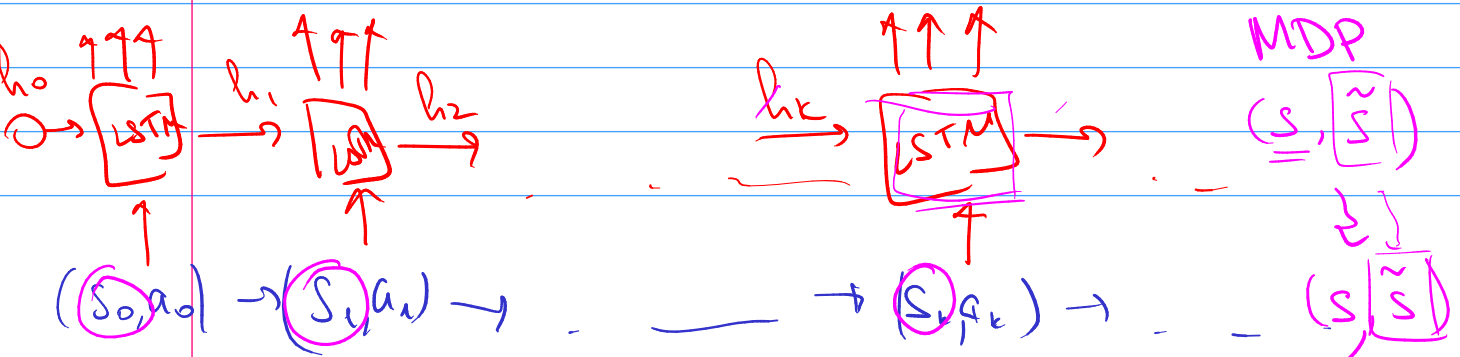


$$Q(s,a) = V(s) + (A(s,a) - \max_{a'} A(s,a'))$$

$$Q(s,a) = V(s) + (A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a'))$$



$$(s_0, a_0) \rightarrow (s_1, a_1) \rightarrow (s_2, a_2) \rightarrow \dots$$



policy improvement theorem

$$\pi^* = \text{MCTS}(\pi, V)$$

