

Tabular MDP. The environment starts in state s_0 and has horizon $H = 1$. Transition probabilities from state s_0 are given for the two binary actions a_0, a_1 (which send the agent to the other state with complementary probability).

