固定策略 去掉奖励函数 → Markov 链 **MDP** → MRP