# MDPREFINED $\rightarrow$ MRP

ASHWIN RAO (STANFORD CME 241)

We start with a MDPRefined given by transitions function $\mathcal{P}^a_{s,s'}$ denoting the probability of transition from state $s$ to state $s'$ upon taking action $a$ and given by rewards function $\mathcal{R}^a_{s's'}$ denoting the reward obtained when transitioning from state $s$ to state $s'$ upon taking action $a$.

First we construct an MDP from $(\mathcal{P}, \mathcal{R})$. The reward function of this MDP will be:

$$\sum_{s'} \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'}$$

Next we construct an MRPRefined from $(\mathcal{P}, \mathcal{R})$ and a policy $\pi(a|s)$. The transitions function of this MRPRefined will be:

$$\sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'}$$

The reward function of this MRPRefined will be:

$$\frac{\sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'}}{\sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'}}$$

Next we construct MRP1 from the above MDP. The reward function of this MRP1 will be:

$$\sum_a \pi(a|s) \cdot (\sum_{s'} \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'}) = \sum_a \sum_{s'} \pi(a|s) \cdot \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'} = \sum_{s'} \sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'}$$

Next we construct MRP2 from the above MRPRefined. The reward function of this MRP2 will be:

$$\sum_{s'} (\sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'}) (\frac{\sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'}}{\sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'}}) = \sum_{s'} \sum_a \pi(a|s) \cdot \mathcal{P}^a_{s,s'} \cdot \mathcal{R}^a_{s,s'}$$

Hence, MRP1 and MRP2 are the same MRP.