MARKOV DECISION PROCESSES

MARKOV DECISION PROCESSES (MDP)

DEFINITION

$\mathcal{S}$ set of states, S_t random variable for state at time t

$\mathcal{A}$ set of actions, A_t random variable for the action at time t

R_t random variable for the reward at time t

$$p(s', r \mid s, a) \doteq \Pr(S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a)$$

$$p(s, a, s') \doteq \Pr(S_{t+1} = s' \mid S_t = s, A_t = a)$$

$$d_0(s) = \Pr(S_0 = s)$$

$$r(s, a, s') \doteq \mathbb{E}[R_{t+1} \mid S_t = s, A_t = a, S_{t+1} = s']$$

$$r(s, a) \doteq \mathbb{E}[R_{t+1} \mid S_t = s, A_t = a]$$

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+1+k}$$

MARKOV DECISION PROCESSES (MDP)

DEFINITION - POLICY

An agent selects an action from a distribution defined in the *policy* π

$$\forall t,\ \pi(a \mid s) \doteq \Pr(A_t = a \mid S_t = s)$$

EXERCISES

$$\Pr(A_0 = a) = ?$$

EXERCISES

$$
\begin{aligned}
\Pr(A_0 = a) &= \sum_{s_0} \Pr(S_0 = s_0, A_0 = a) \\
&= \sum_{s_0} \Pr(A_0 = a \mid S_0 = s_0) \Pr(S_0 = s_0) \\
&= \sum_{s_0} d_0(s) \pi(a \mid s)
\end{aligned}
$$

EXERCISES

$$\Pr(S_3 = s, A_1 = a) = ?$$

EXERCISES

$$
\begin{aligned}
\Pr(S_3 = s, A_1 = a) &= \sum_{s_0,a_0,s_1,s_2,a_2} \Pr(S_3 = s, A_2 = a_2, S_2 = s_2, A_1 = a, S_1 = s_1, A_0 = a_0, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} \Pr(S_3 = s \,|\, A_2 = a_2, S_2 = s_2, \ldots, S_0 = s_0) \Pr(A_2 = a_2, S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} \Pr(S_3 = s \,|\, A_2 = a_2, S_2 = s_2) \Pr(A_2 = a_2, S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} \Pr(S_3 = s \,|\, A_2 = a_2, S_2 = s_2) \Pr(A_2 = a_2 \,|\, S_2 = s_2, \ldots, S_0 = s_0) \Pr(S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} \Pr(S_3 = s \,|\, A_2 = a_2, S_2 = s_2) \Pr(A_2 = a_2 \,|\, S_2 = s_2) \Pr(S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s) \Pr(A_2 = a_2 \,|\, S_2 = s_2) \Pr(S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s) \pi(a_2 \,|\, s_2) \Pr(S_2 = s_2, \ldots, S_0 = s_0)
\end{aligned}
$$

EXERCISES

$$
\begin{aligned}
\Pr(S_3 = s, A_1 = a) &= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) \Pr(S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) \Pr(S_2 = s_2 \mid A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \Pr(A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) \Pr(S_2 = s_2 \mid A_1 = a, S_1 = s_1) \Pr(A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) p(s_1, a_1, s_2) \Pr(A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) p(s_1, a_1, s_2) \Pr(A_1 = a \mid S_1 = s_1, \ldots, S_0 = s_0) \Pr(S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) p(s_1, a_1, s_2) \Pr(A_1 = a \mid S_1 = s_1) \Pr(S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) p(s_1, a_1, s_2)\pi(a, s_1) \Pr(S_1 = s_1, \ldots, S_0 = s_0)
\end{aligned}
$$

EXERCISES

$$
\begin{aligned}
\Pr(S_3 = s, A_1 = a) &= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) \Pr(S_2 = s_2, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) \Pr(S_2 = s_2 \mid A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \Pr(A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2) \Pr(S_2 = s_2 \mid A_1 = a, S_1 = s_1) \Pr(A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2) \Pr(A_1 = a, S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2) \Pr(A_1 = a \mid S_1 = s_1, \ldots, S_0 = s_0) \Pr(S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2) \Pr(A_1 = a \mid S_1 = s_1) \Pr(S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2)\pi(a, s_1) \Pr(S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0} d_0(s_0) \sum_{a_0} \pi(a_0 \mid s_0) \sum_{s_1} p(s_0, a_0, s_1)\pi(a \mid s_1) \sum_{s_2} p(s_1, a, s_2) \sum_{a_2} \pi(a_2 \mid s_2)p(s_2, a_2, s)
\end{aligned}
$$

EXERCISES

$$
\begin{aligned}
\Pr(S_3 = s, A_1 = a) &= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2)\pi(a, s_1) \Pr(S_1 = s_1, \ldots, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2)\pi(a, s_1) \Pr(S_1 = s_1 \mid A_0 = a_0, S_0 = s_0) \Pr(A_0 = a_0, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2)\pi(a, s_1)p(s_0, a_0, s_1) \Pr(A_0 = a_0, S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2)\pi(a, s_1)p(s_0, a_0, s_1) \Pr(A_0 = a_0 \mid S_0 = s_0) \Pr(S_0 = s_0) \\
&= \sum_{s_0,a_0,s_1,s_2,a_2} p(s_2, a_2, s)\pi(a_2 \mid s_2)p(s_1, a_1, s_2)\pi(a, s_1)p(s_0, a_0, s_1)\pi(a_0 \mid s_0)d_0(s_0) \\
&= \sum_{s_0} d_0(s_0) \sum_{a_0} \pi(a_0 \mid s_0) \sum_{s_1} p(s_0, a_0, s_1)\pi(a \mid s_1) \sum_{s_2} p(s_1, a, s_2) \sum_{a_2} \pi(a_2 \mid s_2)p(s_2, a_2, s)
\end{aligned}
$$

EXERCISES

$$\Pr(S_5 = s' \,|\, A_2 = a, S_4 = s) = ?$$

EXERCISES

$$
\begin{aligned}
\Pr(S_5 = s' \,|\, A_2 = a, S_4 = s) = \; & \Pr(S_5 = s' \,|\, S_4 = s) \\
= \; & \sum_{a_4} \Pr(S_5 = s', A_4 = a_4 \,|\, S_4 = s) \\
= \; & \sum_{a_4} \Pr(S_5 = s' \,|\, A_4 = a_4, S_4 = s) \Pr(A_4 = a_4 \,|\, S_4 = s) \\
= \; & \sum_{a_4} \pi(s, a_4) p(s, a_4, s')
\end{aligned}
$$

EXERCISES

$$\mathbb{E}[R_5 \mid S_3 = s, A_4 = a] = ?$$

EXERCISES

$$
\begin{aligned}
\mathbb{E}[R_5 \mid S_3 = s, A_4 = a] &= \sum_{a_3, s_4, s_5} \mathbb{E}[R_5 \mid S_5 = s_5, S_4 = s_4, A_3 = a_3, S_3 = s, A_4 = a] \Pr(S_5 = s_5, S_4 = s_4, A_3 = a_3 \mid S_3 = s_3, A_4 = a) \\
&= \sum_{a_3, s_4, s_5} \mathbb{E}[R_5 \mid S_5 = s_5, S_4 = s_4, A_4 = a] \Pr(S_5 = s_5, S_4 = s_4, A_3 = a_3 \mid S_3 = s_3, A_4 = a) \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) \Pr(S_5 = s_5, S_4 = s_4, A_3 = a_3 \mid S_3 = s_3, A_4 = a) \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) \Pr(S_5 = s_5 \mid S_4 = s_4, A_3 = a_3, S_3 = s_3, A_4 = a) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s, A_4 = a) \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) \Pr(S_5 = s_5 \mid S_4 = s_4, A_4 = a) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s, A_4 = a) \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s, A_4 = a)
\end{aligned}
$$

EXERCISES

$$
\begin{aligned}
\mathbb{E}[R_5 \mid S_3 = s, A_4 = a] &= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s, A_4 = a) \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\Pr(A_4 = a, S_4 = s_4, A_3 = a_3 \mid S_3 = s)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\Pr(A_4 = a \mid S_4 = s_4, A_3 = a_3, S_3 = s) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\Pr(A_4 = a \mid S_4 = s_4) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\pi(a, s_4) \Pr(S_4 = s_4, A_3 = a_3 \mid S_3 = s)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\pi(a, s_4) \Pr(S_4 = s_4 \mid A_3 = a_3, S_3 = s) \Pr(A_3 = a_3 \mid S_3 = s)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \sum_{a_3} \pi(s, a_3) \sum_{s_4} p(s, a_3, s_4) \sum_{a_4} \pi(a_4 \mid s_4) \sum_{s_5} p(s_4, a_4, s_5) r(s_4, a_4, s_5) \\
&= \sum_{a_3} \pi(s, a_3) \sum_{s_4} p(s, a_3, s_4) \sum_{a_4} \pi(a_4 \mid s_4) r(s_4, a_4)
\end{aligned}
$$

EXERCISES

$$
\begin{aligned}
\mathbb{E}[R_5 \mid S_3 = s, A_4 = a] &= \sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \frac{\pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\Pr(A_4 = a \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \Pr(A_4 = a, A_3 = a_3', S_4 = s_4' \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \Pr(A_4 = a \mid A_3 = a_3', S_4 = s_4', S_3 = s) \Pr(A_3 = a_3', S_4 = s_4' \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \Pr(A_4 = a \mid S_4 = s_4') \Pr(A_3 = a_3', S_4 = s_4' \mid S_3 = s)}
\end{aligned}
$$

EXERCISES

$$
\begin{aligned}
\mathbb{E}[R_5 \mid S_3 = s, A_4 = a] &= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \Pr(A_4 = a \mid S_4 = s_4') \Pr(A_3 = a_3', S_4 = s_4' \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \pi(a \mid s_4) \Pr(A_3 = a_3', S_4 = s_4' \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \pi(a \mid s_4') \Pr(S_4 = s_4' \mid A_3 = a_3', S_3 = s) \Pr(A_3 = a_3' \mid S_3 = s)} \\
&= \frac{\sum_{a_3, s_4, s_5} r(s_4, a, s_5) p(s_4, a, s_5) \pi(a, s_4) p(s_3, a_3, s_4) \pi(a_3 \mid s_3)}{\sum_{a_3', s_4'} \pi(a \mid s_4') p(s, a_3', s_4') \pi(a_3' \mid s)}
\end{aligned}
$$

EXERCISES

$$\Pr(S_3 = s' \mid R_4 = r, S_2 = s) = ?$$

EXERCISES

$$\Pr(S_3 = s' \mid R_4 = r, S_2 = s) = \frac{\Pr(R_4 = r, S_3 = s' \mid S_2 = s)}{\Pr(R_4 = r \mid S_2 = s)}$$

$$= \frac{\sum_{a_2} \pi(a_2 \mid s) p(s, a_2, s') \sum_{a_3} \pi(a_3 \mid s') \sum_{s_4} p(s_4, r \mid s', a_3)}{\sum_{a_2'} \pi(a_2' \mid s) \sum_{s_3'} p(s, a_2', s_3') \sum_{a_3'} \pi(a_3' \mid s_3') \sum_{s_4'} p(s_4', r \mid s_3', a_3')}$$

EXERCISES

$R_1 = 2, R_2 = -4, R_3 = 0, R_4 = 16, \ \forall t \geq 5, R_t = 2$

Using $\gamma = 0.5$, what are $G_0, G_1, G_2, G_3, G_4, G_5$?

EXERCISES

$R_1 = 2, R_2 = -4, R_3 = 0, R_4 = 16, \ \forall t \geq 5, R_t = 2$

Using $\gamma = 0.5$, what are G_0, G_1, G_2, G_3, G_4?

$G_t = R_{t+1} + \gamma G_{t+1}$

$$G_4 = \sum_{k=0}^{\infty} \gamma^k R_{t+1+k} = \sum_{k=0}^{\infty} \gamma^k 2 = 2 \sum_{k=0}^{\infty} \gamma^k = 2 \frac{1}{1-\gamma} = \frac{2}{1-0.5} = 4$$

EXERCISES

$R_1 = 2, R_2 = -4, R_3 = 0, R_4 = -18, \ \forall t \geq 5, R_t = 2$

Using $\gamma = 0.5$, what are G_0, G_1, G_2, G_3, G_4?

$G_t = R_{t+1} + \gamma G_{t+1}$

$G_4 = 4$

$G_3 = R_4 + \gamma G_4 = -18 + 0.5(4) = -16$

$G_2 = R_3 + \gamma G_3 = 0 + 0.5(-16) = -8$

$G_1 = R_2 + \gamma G_2 = -4 + 0.5(-8) = -8$

$G_0 = R_1 + \gamma G_1 = 2 + 0.5(-8) = -2$

NEXT CLASS

WHAT YOU SHOULD DO

1. Watch the material for week 3: Value functions and Bellman Equations.
2. Quiz due Friday night: Value Functions and Bellman Equations 1

Friday: policies and value functions