687 2017-09-14

Note Title 9/14/2017

Stationary vs **Non-stationary**

$$P_r(S_{t+1} = s' \mid S_t = s \wedge A_t = A) = P_r(S_{i+1} = s' \mid S_i = s \wedge A_i = A) \quad \forall i$$

Similarly for rewards R. **Policy** may vary (learning!)

What about variation over multiple episodes? Environment may shift in some problems.

Cart-Pole Problem (Inverted Pendulum)

$\gamma = 1$

state $= (\theta, x, \dot{\theta}, \dot{x})$ accelerate (apply force)

$s_0 = (0, 0, 0, 0)$ actions (L, R) $R_t = 1$ always (max time balanced)

episode ends if hit end of track.
or if pole falls ($> 45°$)
or at $t \geq 20$ secs $\Delta t = .02s$

Not Markovian! Why? → This constraints requires we know the time.

↳ state $= (\theta, x, \dot{\theta}, \dot{x}, t)$

A case of a **Finite Horizon MDP**:

$$\exists L : \forall t \geq L : S_t = \tilde{s}$$

L the horizon. Add t to state space (perhaps implicitly).

Can also have Indefinite Horizon - all episodes terminate, but no bound
Infinite Horizon - some episodes may never terminate.

Partial Observability:

- Agent does not know true state - only has observations
- Can be noisy, incomplete.
- Can have Markovian states + non-Markovian observations
- Two approaches:
  - POMDP: state → sensor → agent
  - Function Approximation: state → agent but observations modeled in agent
    ↑ used in this course.

Any system can be modeled either way. Affects whether environment is an MDP (Markovian).

↓

Func. Approx gives simpler model of the environment.