Stationary vs Non-stationary

\[ P_i \left( S_{t+1} = s' \mid S_t = s, A_t = A \right) = P_r \left( S_{i+1} = s' \mid S_i = s, A_i = A \right) \quad \forall i \]

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

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

*Cart-Pole Problem* (Inverted Pendulum)

\[
\text{angle} \quad \text{state} = (\theta, x, \dot{\theta}, \dot{x}) \quad \text{accelerate (force)}
\]

\[
\text{force} \quad s_0 = (0, 0, 0, 0) \quad \text{actions (L, R)} \quad R_f = 1 \quad \text{always}
\]

Episode ends if hit end of track.

- if pole falls ($> 45^\circ$)
- if at \( t \approx 20 \text{secs} \)

\[
\Delta t = 0.02s
\]

Not Markovian! Why?

This constraint requires we know the time.

A case of a Finite Horizon MDP:

\[
F.H.: H_t \geq L: S_{L+1}^r
\]

At the horizon, add \( t \) to state space (perhaps implicitly).
Can also have **Infinite 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 $\rightarrow$ sensor $\rightarrow$ agent
  - Function Approximation: state $\rightarrow$ agent but observations modeled in agent

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

Func. Approx gives simpler model of the environment.