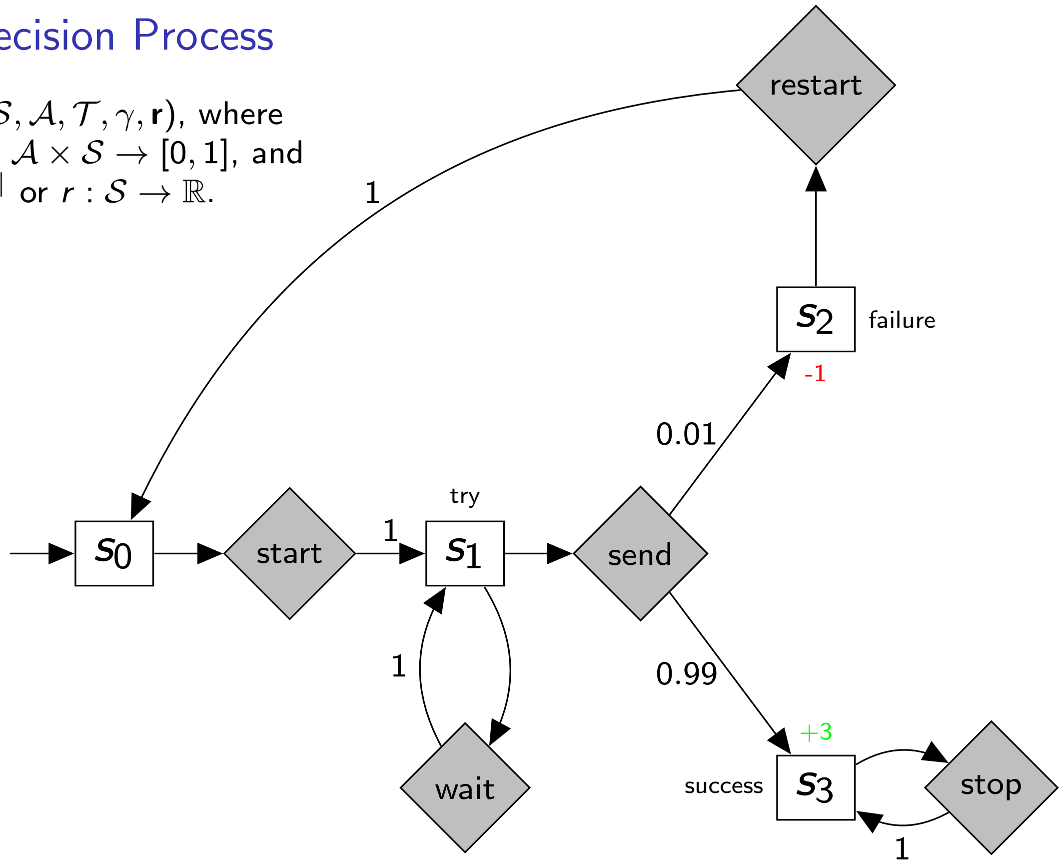
Markov Decision Process

$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \gamma, \mathbf{r})$, where
$\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$, and
$\mathbf{r} \in \mathbb{R}^{|\mathcal{S}|}$ or $r : \mathcal{S} \to \mathbb{R}$.



Inverse Reinforcement Learning

Given:

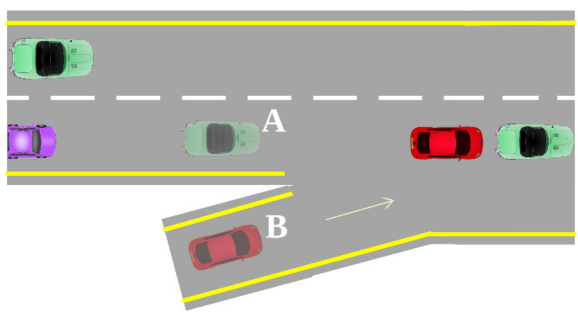
- $\mathcal{M} \setminus \{\mathbf{r}\}$,
- $\mathcal{D} = \{\zeta_i\}_{i=1}^{N}$, where $\zeta_i = \{(s_{i,1}, a_{i,1}), \ldots, (s_{i,T}, a_{i,T})\}$,
- features $\mathbf{X} \in \mathbb{R}^{|\mathcal{S}| \times d}$,

find $\mathbf{r}$. Motivation:

- Reward functions can be difficult to construct in practice
- Rewards are more generalisable than policies

Applications

- Autonomous vehicle control
  - Helicopter tricks
  - Robot movement among people
- Predicting behaviour
  - Taxi destinations
  - Pedestrian movement
  - Energy efficient driving



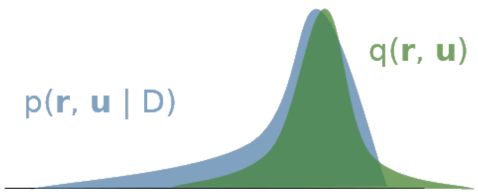
Variational Inference

How to calculate the posterior probability distribution?

$$p(\mathbf{r}, \mathbf{u} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathbf{r})p(\mathbf{r} \mid \mathbf{u})p(\mathbf{u})}{p(\mathcal{D})}$$

Solution: approximate $p(\mathbf{r}, \mathbf{u} \mid \mathcal{D})$ with $q(\mathbf{r}, \mathbf{u}) = q(\mathbf{r} \mid \mathbf{u})q(\mathbf{u})$, where

- $q(\mathbf{r} \mid \mathbf{u}) = p(\mathbf{r} \mid \mathbf{u})$,
- $q(\mathbf{u}) = \mathcal{N}(\mathbf{u}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.



Benefits

- More precise reward predictions
- Variance estimates guide further data collection
- Bayesian treatment guards against overfitting

Theoretical Results

Proposition

MDP value functions $V(s) : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}$ (for $s \in \mathcal{S}$) are Lebesgue measurable.

Proposition

If the initial values of the MDP value function satisfy the following bound, then the bound remains satisfied throughout value iteration:

$$|V_{\mathbf{r}}(s)| \leq \frac{\|\mathbf{r}\|_\infty + \log|\mathcal{A}|}{1-\gamma}.$$

Theorem

Whenever the derivative exists,

$$\frac{\partial}{\partial t} \iint V_{\mathbf{r}}(s)q(\mathbf{r} \mid \mathbf{u})q(\mathbf{u})\, d\mathbf{r}\, d\mathbf{u} = \iint \frac{\partial}{\partial t}[V_{\mathbf{r}}(s)q(\mathbf{r} \mid \mathbf{u})q(\mathbf{u})]\, d\mathbf{r}\, d\mathbf{u},$$

where t is any scalar part of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, or $\boldsymbol{\lambda}$.