$$J(\pi; f) = \mathbb{E}_{\epsilon_{0:\infty}}\Big[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_{t+1} = f(s_t, a_t) + \epsilon_t, a_t = \pi(s_t)\Big]$$

RL objective:

Expectation is over transition noise, i.e. aleatoric uncertainty

Posterior over dynamics models:

$p(f \mid \mathscr{D})$

How should we use this?

$$J(\pi; f) = \mathbb{E}_{???}\Big[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_{t+1} = f(s_t, a_t) + \epsilon_t, a_t = \pi(s_t) \Big]$$

Sources of Uncertainty

Decision-making Under Uncertainty

37

Stochastic dynamics

Return = discounted sum of rewards

Deterministic policy

What is the expectation over?

Sources of Uncertainty

Decision-making Under Uncertainty

RL objective:

$$J(\pi; f) = \mathbb{E}_{\epsilon_{0:\infty}}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_{t+1} = f(s_t, a_t) + \epsilon_t, a_t = \pi(s_t)\right]$$

Expectation is over transition noise, i.e. aleatoric uncertainty

Posterior over dynamics models:

$$p(f \mid \mathcal{D})$$

How should we use this?

Model Averaging

$$\pi^{Greedy} = \arg\max_{\pi} \mathbb{E}_{p(f|\mathcal{D})}\Big[J(\pi; f)\Big]$$