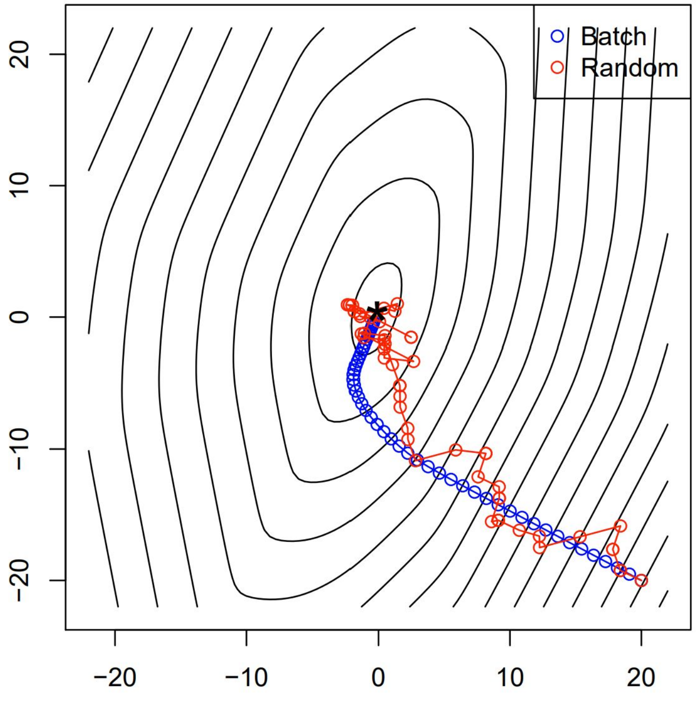
AIL 722: Reinforcement Learning

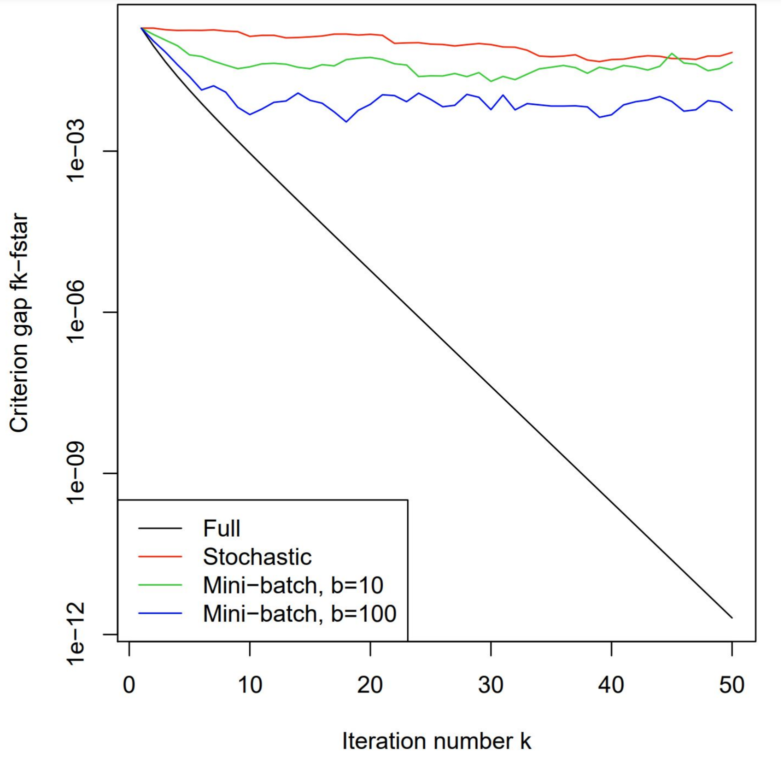
Lecture 34: Baselines

Raunak Bhattacharyya

ScAI YARDI SCHOOL OF ARTIFICIAL INTELLIGENCE
INDIAN INSTITUTE OF TECHNOLOGY DELHI

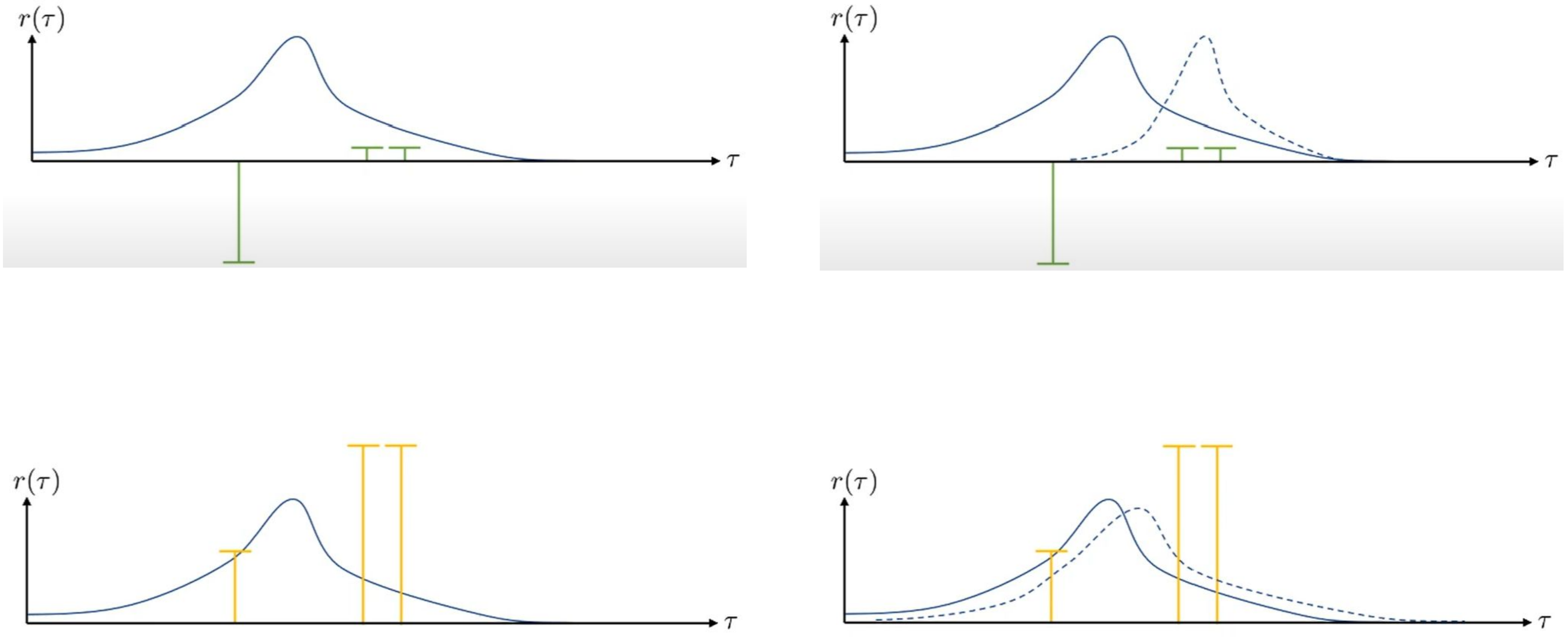
Why Does Variance Matter?





Convex Optimisation, Ryan Tibshirani

Why Does Variance Matter in Policy Gradients



Adapted from Sergey Levine

Aside: Control Variates

$$\text{Var}(X - Y) = E[(X - Y)^2] - (E[X - Y])^2$$

$$= E[X^2 - 2XY + Y^2] - (E[X] - E[Y])^2$$

$$= E[X^2] - 2E[XY] + E[Y^2] - (E[X])^2 - (E[Y])^2 + 2E[X]E[Y]$$

$$= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 - (2E[XY] - 2E[X]E[Y])$$

$$= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

A route to reduce the variance of our gradient estimator

Baseline

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim p_\theta(\tau)} \left[\nabla_\theta \log p_\theta(\tau) \, r(\tau) \right]$$

$$\text{Let } Y = \nabla_\theta \log p_\theta(\tau) \cdot \Big(r(\tau) - b \Big)$$

$$\mathbb{E}\Big[Y \Big] = \mathbb{E}\Big[\nabla_\theta \log p_\theta(\tau) \cdot r(\tau) \Big] - b \cdot \mathbb{E}\Big[\nabla_\theta \log p_\theta(\tau) \Big]$$

$$E[\nabla_\theta \log p_\theta(\tau)] = \int p_\theta(\tau) \nabla_\theta \log p_\theta(\tau) \, d\tau$$

$$= \int \nabla_\theta p_\theta(\tau) \, d\tau$$

Remember this?

$$= \nabla_\theta \int p_\theta(\tau) \, d\tau \qquad = \nabla_\theta 1$$

$$E\Big[Y \Big] = E\Big[\nabla_\theta \log p_\theta(\tau) \cdot r(\tau) \Big]$$

Our revised estimator Y does not introduce bias