# Q&A on Intro to RL & MDPs

Christopher Mutschler

# Overview

*Exercise 3.2* Is the MDP framework adequate to usefully represent *all* goal-directed learning tasks? Can you think of any clear exceptions? □

# IID

- **What does ”i.i.d.” mean? (textually)**
  - Independent and indentically distributed

- Why is iid not guaranteed in RL?

  - Independently distributed: Actions have consequences. With the selected actions and the current state, we are in we influence the sample that we see next. Samples within a trajectory are correlated. (aka choices are made according to the trajectory)

  - Identically distributed:
    1. We only see a (very small) subset of the data which is only a rough approximation of the *real* data space
    2. Our behavior policy changes, and we always maximize the target:

$$\mathbb{E}_{\tau \sim \pi}\left[\sum_t \gamma^t R_t\right]$$

this is not stationary!