WhiteBox

Symposium on Explainability

LOEWE Exzellente Forschung für Hessens Zukunft | Centre for Cognitive Science | TECHNISCHE UNIVERSITÄT DARMSTADT

Matthias Schultheis

Inferring Interpretable Properties in Sequential Decision Tasks

Reinforcement learning (RL) methods can be used to generate behavior in sequential decision-making tasks based on model specifications with interpretable parameters. To find parameters that explain and reproduce observed behavior, inverse RL (IRL) methods can be applied. Existing methods, however, are limited to specific systems and parameters.

Goal: In this subproject we develop and apply new (I)RL approaches for generating interpretable representations for given behavor of an observed agent. We focus on partial-observable stochastic tasks with the goals to estimate costs and noise quantities in sensorimotor tasks as well as inferring discount parameters in economic decision-making tasks.

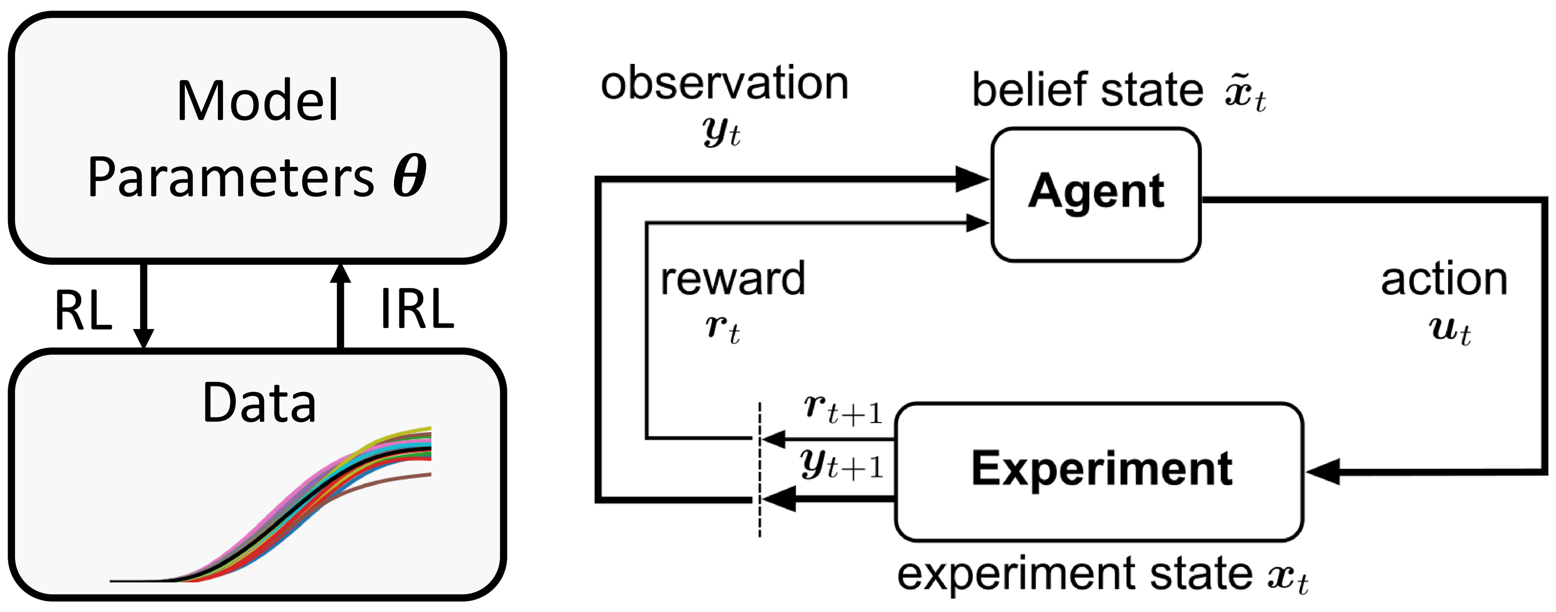
IRL for Stochastic Systems

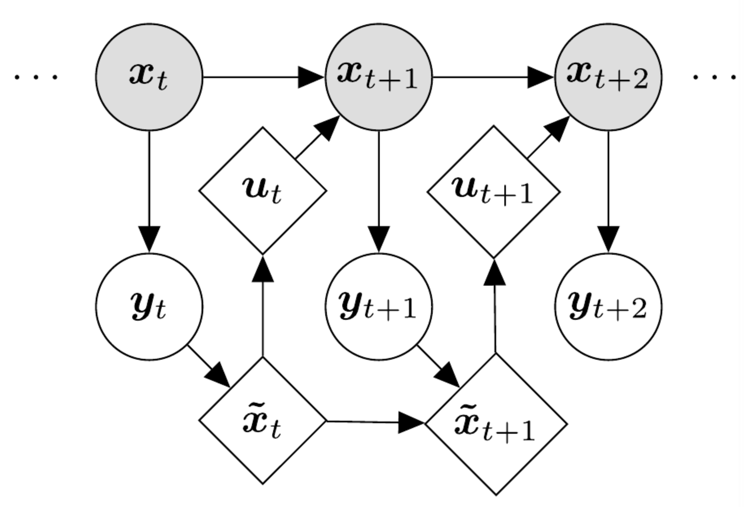
Problem: Previous IRL methods are

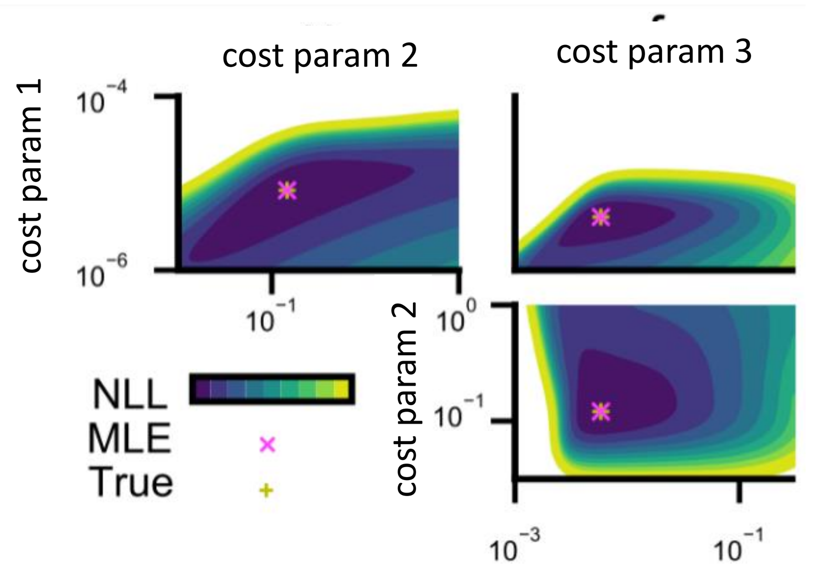
- agnostic of agent‘s suboptimality (softmax decision rule)
- limited to estimation of cost function

Contribution: New probabilistic method [1, 2]

- for **non-linear partially-observable** stochastic systems
- incorporates **explicit stochasticity** model via partial/noisy observations
- can infer a **wide range of parameters** (e.g., dynamics function, noises, costs)
- Principle: Maximize approximate likelihood determined by assumed-density filtering







Continuous Time POMDPS

Problem:

No framework for **modelling & solving** partially observable Markov decision problems (POMDPs) with continuous time and discrete state-action space

Contribution:

Formalization of continuous time POMDP model [3]

- Hamilton-Jacobi-Bellman (HJB)-type equation characterizes **optimal solution**
- Solution via collocation or advantage updating

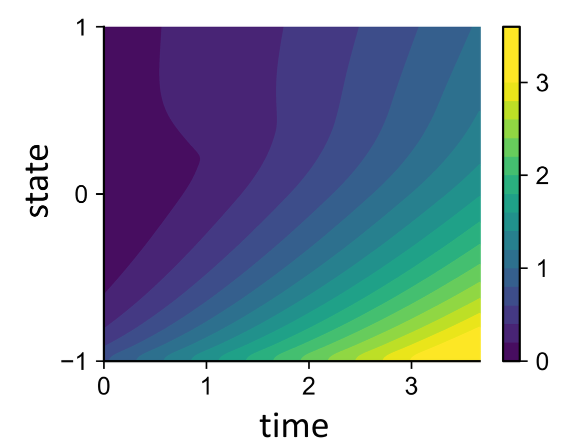
RL with Non-Exponential Discounting

Problem:

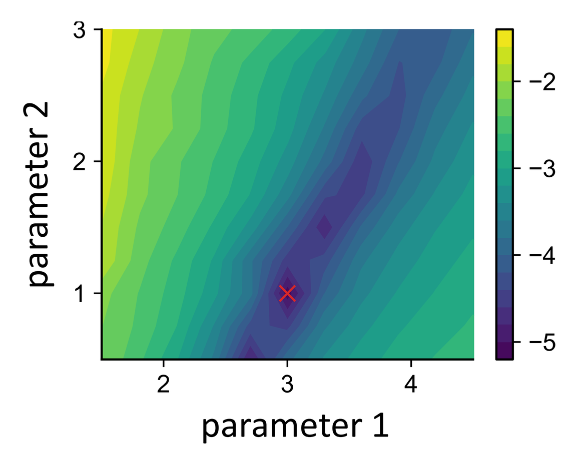
- Most forward and inverse RL methods assume **exponential discount function** to discount future rewards: $R = \sum_{i=0}^{\infty} \boldsymbol{\gamma}^{\boldsymbol{t}} r(s_t, a_t)$
- Human decision behavior in economic tasks and systems with **non-exponential end time** distribution are not well modelled with exponential discounting.

Contribution: Methods [4] for **stochastic systems** to

- determine the optimal behavior for **non-exponential discount** functions
- **learn** the discount function from data



value function (time dependent)



likelihood and inferred discount parameters

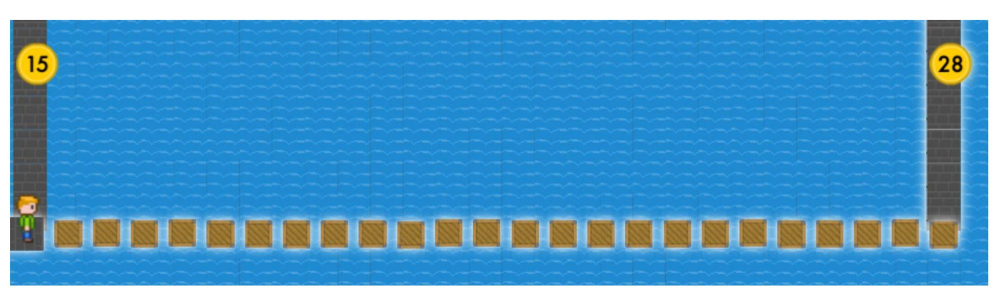
Normative Models of Human Discounting

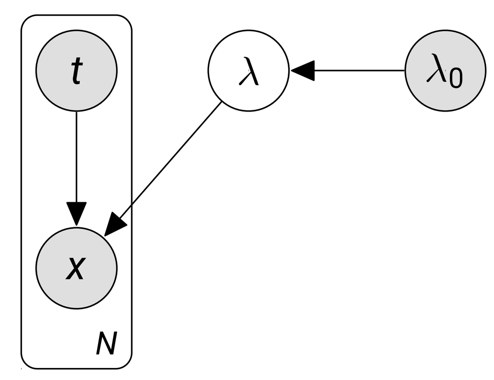
Problem:

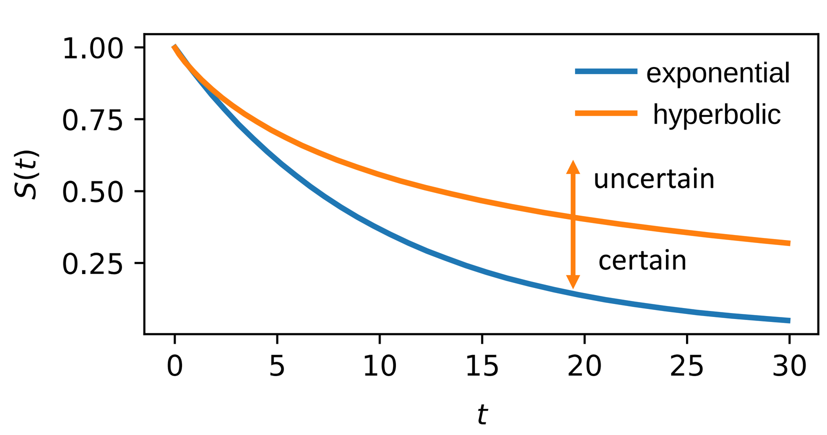
Many **theories for explaining non-exponential** discounting but **no experimental validation**

Contribution:

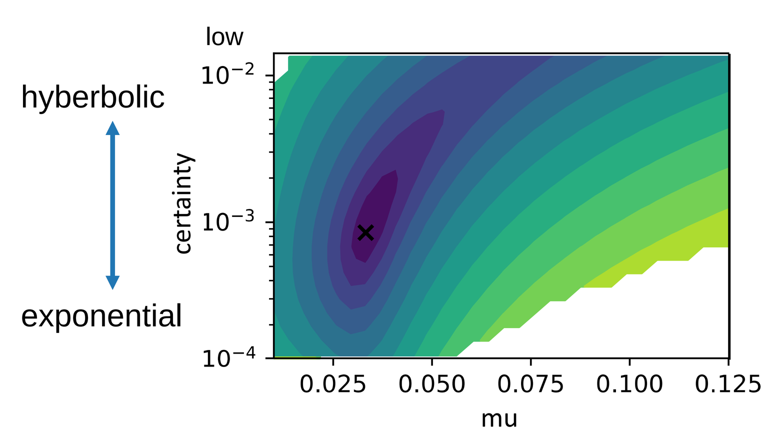
- Probabilisic model predicting behavior to become „more exponential“ with **increasing certainty**
- **Experiment** with uncertainty over risk to validate model (ongoing)







discount function



log likelihood

[1] Schultheis, Straub, Rothkopf (2021). Inverse optimal control adapted to the noise characteristics of the human sensorimotor system. NeurIPS
[2] Straub, Schultheis, Koeppl, Rothkopf (2023). Probabilistic invers optimal control with local linearization for non-linear partially observable systems. *(subm. to ICML)*
[3] Alt, Schultheis, Koeppl (2020). POMDPs in continuous time and discrete spaces. NeurIPS
[4] Schultheis, Rothkopf, Koeppl (2022). Reinforcement Learning with Non-Exponential Discounting. NeurIPS

TU Darmstadt / 27.03.2023
LOEWE-Schwerpunkt WhiteBox