Inference of Strategic Behavior based on Incomplete Observation Data

Antti Kangasrääsiö
Samuel Kaski
first.last@aalto.fi

Problem Setting

We wish to infer the parameters of the task, preferences or limitations of users when they are performing tasks involving strategic behavior in inverse reinforcement learning (IRL) context.

A limitation with existing methods for IRL is that they make very specific assumptions about the type of observation data: trajectories denoted as ξ = (s0, a1, s1, …, aT, sT). We extend this setting to arbitrary noise models σ(ξ).

Background: In IRL, a RL model is used to explain the strategic behavior of a used in a situation similar to a Markov decision process (S, A, T, R, γ). The strategic behavior is assumed to follow an optimal policy for the MDP. Given observations of the user's behavior, we wish to infer the parameters of the MDP, such that the optimal policy matches the observed behavior.

Our Contributions

We demonstrate that IRL inference is possible even when the observation noise model is an arbitrary function σ(ξ). This extends the state of the art which is only able to deal with few special types of observation noise (missing / probabilistic state observations).

We derive the exact Bayesian likelihood for this problem, but demonstrate that it may be very expensive to evaluate

We propose two approximations: a Monte-Carlo estimate and an ABC estimate, which are faster to evaluate

IRL-SD Problem

Let M be a MDP (S, A, T, R, γ) with parameters θ. Let the true parameters be $\theta^* \in \Theta$ and assume agent behaving according to an optimal policy for M_{θ^*}. Assume the agent has taken paths $(\xi_1, \ldots, \xi_N)$ and we observe *summaries* $\Xi_\sigma = (\xi_{1\sigma}, \ldots, \xi_{N\sigma})$, where $\xi_{i\sigma} \sim \sigma(\xi_i)$ and σ is a known summary function. The *inverse reinforcement learning problem from summary data (IRL-SD)* is then:

Given (1) set of summaries Ξ_σ of an agent demonstrating optimal behavior; (2) summary function σ; (3) MDP M with θ unknown; (4) bounded space Θ; and optionally (5) prior $P(\theta)$.
Estimate $\hat{\theta} \in \Theta$ such that simulated behavior from $M_{\hat{\theta}}$ agrees with Ξ_σ, or the posterior $P(\theta|\Xi_\sigma)$.

Exact Likelihood

$$L(\theta|\Xi_\sigma) = \prod_{i=1}^{N} P(\xi_{i\sigma}|\theta) = \prod_{i=1}^{N} \sum_{\xi_i \in \Xi_{\sigma n}} P(\xi_{i\sigma}|\xi_i) P(\xi_i|\theta),$$

$$P(\xi_i|\theta) = P(s_0^i) \prod_{t=0}^{T_i - 1} \pi_\theta^*(s_t^i, a_t^i) P(s_{t+1}^i | s_t^i, a_t^i).$$

Expensive to evaluate

Monte-Carlo Likelihood

$$\hat{L}(\theta|\Xi_\sigma) = \prod_{i=1}^{N} \frac{1}{N_{MC}} \sum_{\xi_n \in \Xi_{MC}} \frac{P(\xi_{i\sigma}|\xi_n) P(\xi_n|\theta)}{P(\xi_n|\theta)}$$

$$= \prod_{i=1}^{N} \frac{1}{N_{MC}} \sum_{\xi_n \in \Xi_{MC}} P(\xi_{i\sigma}|\xi_n).$$

Applicable when we know σ as a distribution

ABC Likelihood

Assume a function for generating summary datasets Ξ_σ^{sim} given MDP M, parameters θ, number of episodes N, and summary function σ: $\texttt{RLSUM}(M_\theta, N, \sigma)$. Also assume a *discrepancy function* δ,

$$\delta(\Xi_\sigma^A, \Xi_\sigma^B) \to [0, \infty),$$

which quantifies the dissimilarity between two observation datasets.

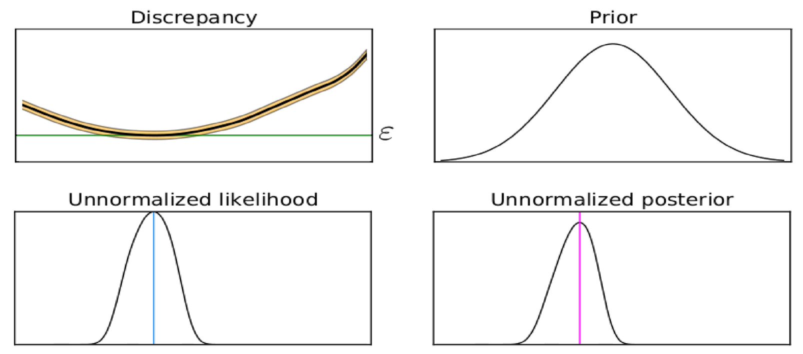
By combining $\texttt{RLSUM}(M_\theta, |\Xi_\sigma|, \sigma)$ with δ, we define

$$d_\theta \sim \delta(\texttt{RLSUM}(M_\theta, |\Xi_\sigma|, \sigma), \Xi_\sigma).$$

The distribution of d_θ corresponds with the ability of θ to satisfy our requirements for solving the IRL-SD problem. Finally we define an approximate likelihood function,

$$\tilde{L}_\varepsilon(\theta|\Xi_\sigma) = P(d_\theta \leq \varepsilon|\theta),$$

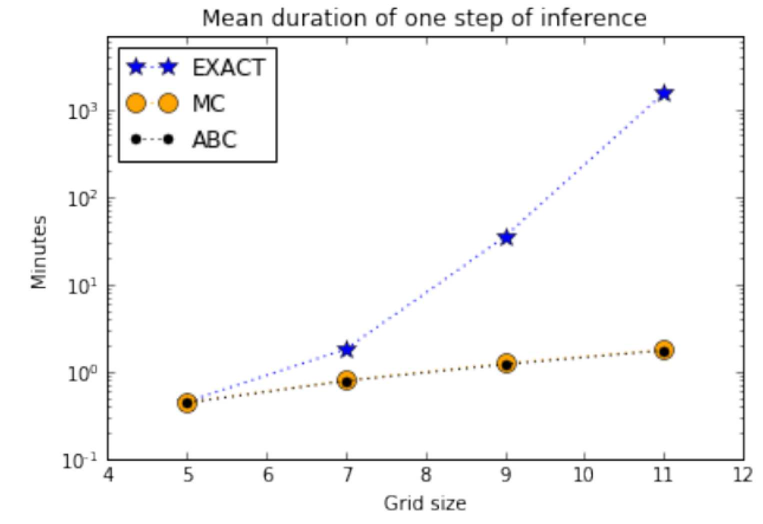
where the approximation threshold $\varepsilon \in [0, \infty)$.



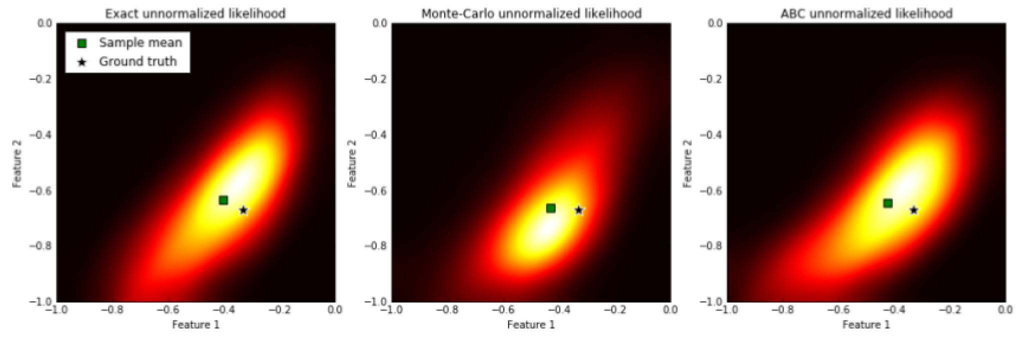
Applicable when we can evaluate σ at will

Experiments (Grid World)

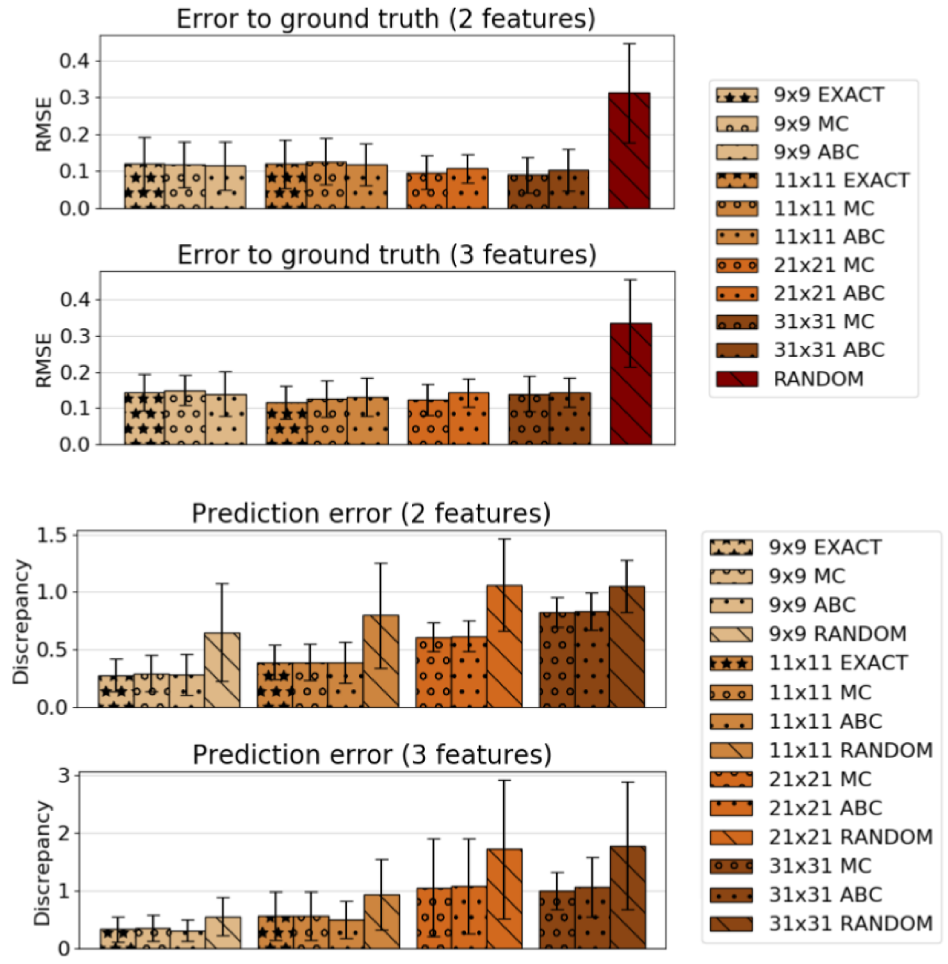
Algorithm runtime (one step) with different grid sizes



Inferred likelihood densities (example)



Inference quality (error to ground truth, prediction error) with different grid sizes and dimensionality of reward function



Inference

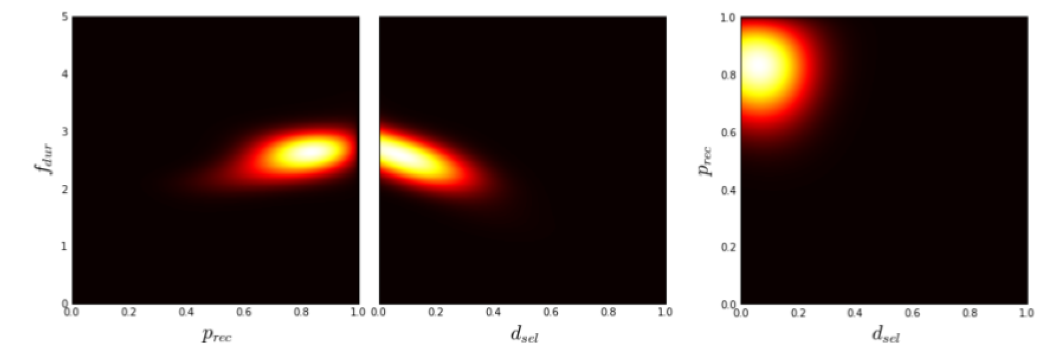
GP surrogate fit using Bayesian optimization

Algorithm 1 Likelihood Estimation for IRL-SD

Input: M, Ξ_σ, Θ, H, N_{opt}
Output: Likelihood estimate $\bar{L}(\theta)$
$D \leftarrow \varnothing$
for $i = 1$ **to** N_{opt} **do**
  $\theta_i \leftarrow \arg\max_\theta Acq(\theta|D, H)$
  $\pi_{\theta_i}^* \leftarrow \texttt{RL}(M_{\theta_i})$
  if Exact **then**
    $d_\theta \leftarrow \log L(\theta_i|\Xi_\sigma)$
  else
    $\Xi_{MC} \leftarrow \{\texttt{SIM}(M_{\theta_i}, \pi_{\theta_i}^*)\}_{n=1 \ldots N_{MC}}$
    if Monte-Carlo **then**
      $d_\theta \leftarrow \log \hat{L}(\theta_i|\Xi_\sigma)$
    else if ABC **then**
      $\Xi_\sigma^{sim} \leftarrow \{\sigma(\Xi_{MC,n})\}_{n=1 \ldots N_{MC}}$
      $d_\theta \leftarrow \delta(\Xi_\sigma^{sim}, \Xi_\sigma)$
    end if
  end if
  $D \leftarrow \{D, (\theta_i, d_\theta)\}$
end for
if ABC **then**
  $\varepsilon \leftarrow \min_\theta G_\mu(\theta|D, H)$
  $\bar{L}(\theta) \leftarrow \Phi(\varepsilon|G_\mu(\theta|D, H), G_s(\theta|D, H))$
else
  $\log \bar{L}(\theta) \leftarrow G(\theta|D, H)$
end if

Full Posterior Inference

Model: Visual search in drop-down menus



Summary

Regarding partial observability in IRL, there now exists formulations for three different situations:

(1) Agent has partial observability of the environment state → POMDP model

(2) External observer has partial observability on *state* level → traditional IRL methods can be extended

<u>**New:**</u> (3) External observer has partial observability on *path* level → presented methods for IRL-SD can be used