Decoding Rewards in Competitive Games: Inverse Game Theory with Entropy Regularization

Junyi Liao[1], Zihan Zhu[2], Ethan X. Fang[1],
Zhuoran Yang[3], Vahid Tarokh[1]

[1] Department of Electrical and Computer Engineering, Duke University
[2] Department of Statistics and Data Science, University of Pennsylvania
[3] Department of Statistics and Data Science, Yale University

ICML 2025

Motivation

- Many real-world scenarios involve strategic interactions: markets, cyber-security, multi-agent system, etc.
- Can we infer underlying preferences (rewards) from observed agent behavior?
- We observe equilibrium behavior *only* (no rewards).
- Our goal: infer reward functions assuming agents play a Quantal Response Equilibrium (QRE).

Problem & Key Challenges

Setting:

- Two-player zero-sum Markov games.
- Observe agents' strategies (in the form of empirical policies).
- Goal: Recover the underlying reward function that explains observed behavior.

Main Challenges:

- **Non-identifiability:** There may exists multiple reward functions leading to the same QRE.
- **Offline Setting:**
  - **Noisy observation:** Empirical policies deviate from exact QRE;
  - **Partial coverage:** observed strategies may fail to comprehensively cover the state-action space.
- **High-dimensional features:** Overfitting and instability when the feature space is large.

Our Contributions

- **General Framework:** We study inverse reinforcement learning in entropy-regularized Markov games under the QRE assumption.
- **Identifiability Theory:** We characterize when the reward function is uniquely or partially recoverable.
- **Efficient Estimation:** We propose an algorithm that constructs confidence sets for reward functions and verify its performance empirically.
- **Statistical Guarantees:** We establish finite-sample convergence bounds for our algorithm under mild assumptions.

Key Ideas

Goal: Recover the reward function from observed strategies in a zero-sum Markov game under entropy regularization, which is formulated as

$$\max_{\mu_h} \min_{\nu_h} \mu_h(\cdot|s)^\top Q_h(s,\cdot,\cdot)\nu_h(\cdot|s) + \frac{1}{\eta}\mathcal{H}(\mu_h(\cdot|s)) - \frac{1}{\eta}\mathcal{H}(\nu_h(\cdot|s))$$

This problem is *concave* in μ_h and *convex* in ν_h.

Key Assumptions:

- Agents follow the Quantal Response Equilibrium (QRE);
- Rewards and transitions have **linear** structures:

$$r_h(s,a,b) = \phi(s,a,b)^\top \omega_h, \quad \mathbb{P}_h(\cdot|s,a,b) = \phi(s,a,b)^\top \pi_h(\cdot).$$

The QRE satisfies a group of softmax constraints (obtained by KKT conditions for the optimization problem). Under the linear assumption, one can transform nonlinear constraints to linear constraints.

Methodology

Both identification and estimation crucially rely on the linear structure of the QRE constraints to understand and estimate rewards.

1. Identification (Theoretical Task)

- Assume the exact QRE (μ^*, ν^*) is known.
- The reward parameter θ satisfies linear constraints

$$\begin{bmatrix} A(\nu^*) \\ B(\mu^*) \end{bmatrix} \theta = \begin{bmatrix} c(\mu^*) \\ d(\nu^*) \end{bmatrix}$$

- Analyze when θ is uniquely or partially identifiable.

2. Estimation (Statistical Task)

- Estimate QRE $(\hat{\mu}, \hat{\nu})$ from data.
- Construct a confidence set using relaxed constraints (least squares):

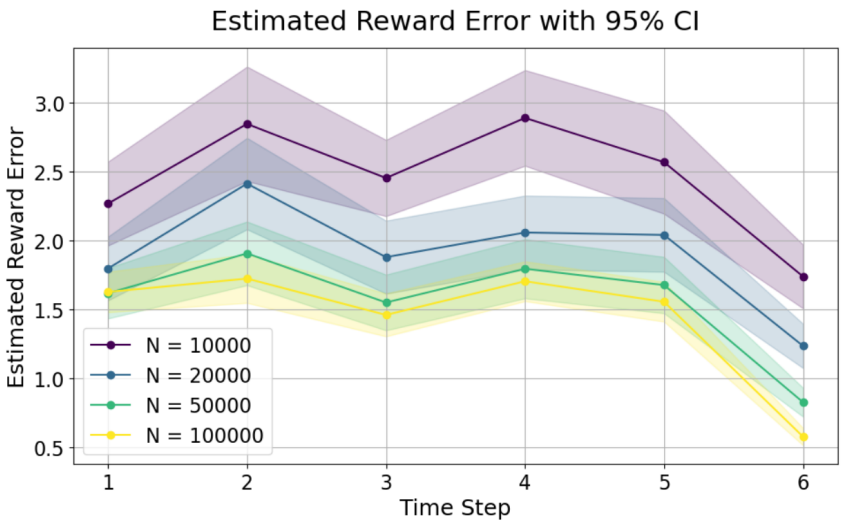
$$\left\| \begin{bmatrix} A(\hat{\nu}) \\ B(\hat{\mu}) \end{bmatrix} \theta - \begin{bmatrix} c(\hat{\mu}) \\ d(\hat{\nu}) \end{bmatrix} \right\|^2 \leq \kappa$$

Theory Highlights

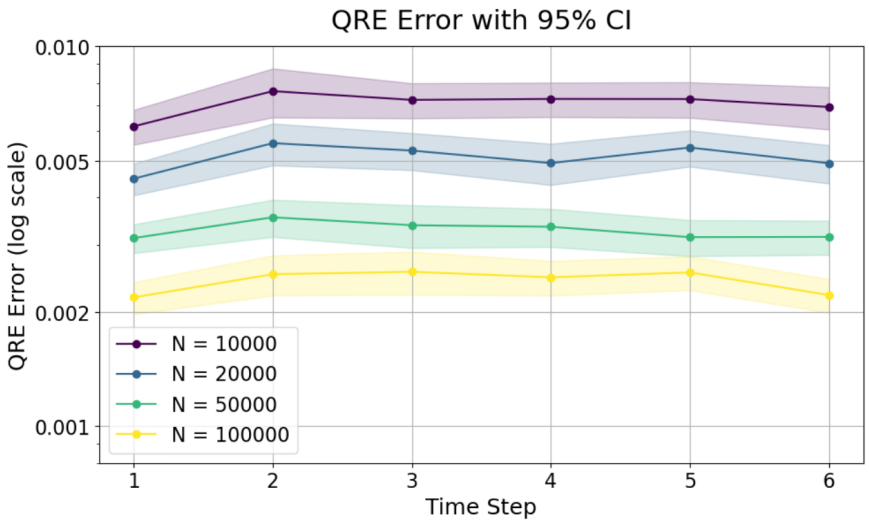
- We establish the identifiability conditions via rank conditions on matrices comprised of features ϕ (See Proposition 3.5).
- The confidence set covers all feasible parameters with high probability, even in the partially identifiable case.
- Convergence rate: $\mathcal{O}(N^{-1/2})$ in sample size N (See Theorem 3.9).

Experiments

- Synthetic Markov games with known ground-truth rewards.
- Evaluate
  1. the error between ground-truth rewards and estimated rewards;
  2. the error between the corresponding QREs.
- Empirical convergence matches theoretical guarantees.



(a) The reconstruction error of reward functions



(b) The error of estimated QRE

Takeaways

- Introduced a framework for inverse reinforcement learning in entropy-regularized Markov games.
- Provided identifiability conditions and estimation guarantees.
- Future: general-sum games, nonlinear setting.

Thanks!