A Data-Driven State Aggregation Approach for Dynamic Discrete Choice Models

Sinong Geng[1], Houssam Nassif*[2], Carlos A. Manzanares[3]

Computer Science Department, Princeton University[1], Meta[2], Amazon[3], *Work done at Amazon

Abstract

We consider dynamic discrete choice model (DDM) estimation, which estimates parameters of agent reward functions (also known as structural parameters) using agent behavioural data.
In this work, we present a novel algorithm that provides a data-driven method for selecting and aggregating states, which lowers the computational and sample complexity of estimation.

Problem Definition

Agents make decisions under a Markov Decision Problem (MDP), $M = (\mathscr{S}, \mathscr{A}, r, \gamma, P)$, solving:

$$V^\theta(s) := \max_\pi \sum_{t=0}^{\infty} \gamma^t \mathbb{E}[r(S_t, A_t; \theta) + \mathscr{H}(\pi(S_t, \cdot)) \mid S_0 = s], \quad (1)$$

where $\mathscr{H}(\pi(s,\cdot)) := -\int_{\mathscr{A}} \log(\pi(s,a))\pi(s,a)\,da$ represents information entropy.

The goal of DDM estimation is to estimate θ using agent decision-making behaviours.

Background

NF-MLE [1] maximizes

$$L(\mathbb{D};\theta) := \frac{1}{N}\sum_{i=1}^{N}\left(Q^\theta(s_i,a_i) - \log\left(\sum_{a'\in\mathscr{A}} \exp(Q^\theta(s_i,a'))\right)\right), \quad (2)$$

where $Q^\theta(s,a)$ satisfies the following Bellman equation

$$Q^\theta(s,a) := r(s,a;\theta) + \gamma\mathbb{E}\left[\log\left(\sum_{a'\in\mathscr{A}} \exp(Q^\theta(s',a'))\right) \Big| s,a\right]. \quad (3)$$

In each iteration of NF-MLE, with a candidate θ, the algorithm solves for $Q^\theta(s,a)$ by fixed-point iteration via the Bellman equation (3).

State aggregation chooses a subset of states: $\Pi(\cdot) : \mathscr{S} \to \tilde{\mathscr{S}}$ with $\tilde{\mathscr{S}} := \{\tilde{s}_1, \tilde{s}_2, \cdots \tilde{s}_{n_s}\}$.

State aggregation inevitably comes with asymptotic error. With $\tilde{\theta}^\Pi := \arg\max_\theta \mathbb{E}[\tilde{L}(\mathbb{D};\theta,\Pi)]$:

$$\varepsilon_{asy}(\Pi) := \left\|\tilde{\theta}^\Pi - \theta^*\right\|^2.$$

SAmQ

We propose state aggregation minimizing Q error (SAmQ).

Q error

$$\varepsilon_Q(\Pi) := \max_{(s,a)\in\mathscr{S}\times\mathscr{A}} \left|Q^{\theta^*}(s,a) - Q^{\theta^*}(\Pi(s),a)\right|. \quad (4)$$

Minimize the Q error by clustering Consider a clustering problem with a distance function defined as

$$d(s,s') := \max_{a\in\mathscr{A}} \left|Q^{\theta^*}(s,a) - Q^{\theta^*}(s',a)\right|. \quad (5)$$

Procedure

Step 1 Estimate Q^{θ^*} using IRL [2].

Step 2 Aggregate states by clustering.

Step 3 Estimate structural parameters using NF-MLE with aggregated states.

Algorithm SAmQ

```
Input Dataset: 𝕏, n_s.
Output θ̂
Q̂ ← DeepPQR(𝔻)
Π̂ ← Clustering(𝔻, Q̂, n_s)
θ̂ ← NF-MLE(𝔻, Π̂)
Return θ̂
```

Theory

Theorem 1 *Under some assumptions*

$$\varepsilon_{asy}(\Pi) \le \frac{4}{C_H(1-\gamma)}\varepsilon_Q(\Pi).$$

Theorem 2 *Non-asymptotic error bounds are provided demonstarting the trade-off between variance and bias, with*

$$BiasBound := \frac{4}{C_H(1-\gamma)}\left(\frac{R_{\max}+1}{1-\gamma}\frac{4}{n_s^{\frac{1}{n_a}}-1} + \varepsilon_P\right),$$

$$VarianceBound := \frac{4(R_{max}+1)}{(1-\gamma)C_H}\sqrt{\frac{\log(\frac{4|\Theta|}{\delta})}{2N}} + \frac{R_{max}+1}{(1-\gamma)^2 C_H}\sqrt{\frac{\log(\frac{8n_s n_a|\Theta|}{\delta})}{2N}}\frac{4}{C_{uni}-\sqrt{\frac{\log(\frac{4n_s n_a|\Theta|}{\delta})}{2N}}}.$$

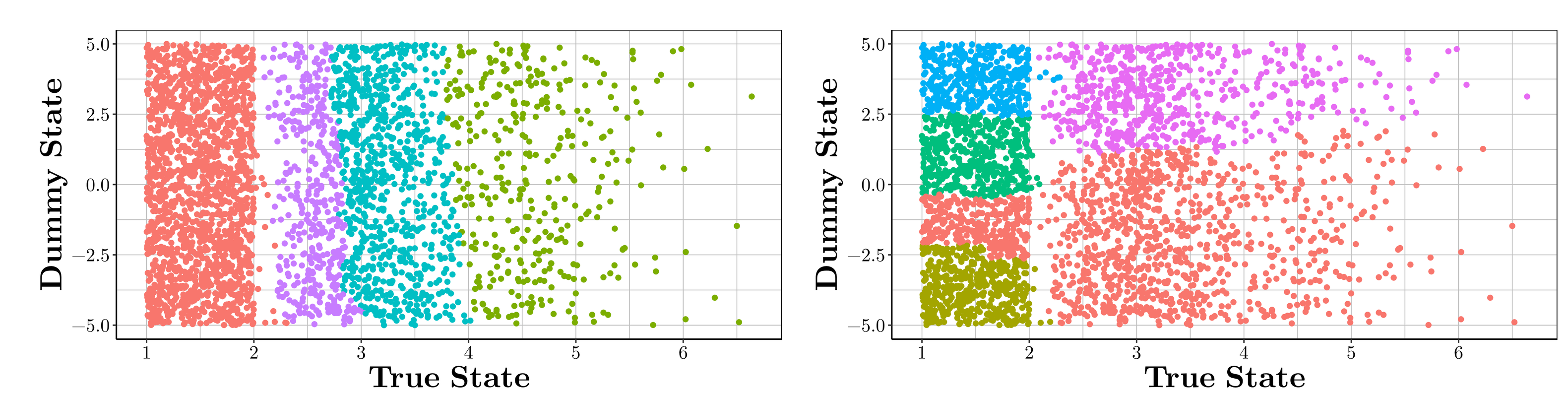
Experiments

Table: Considered methods

Methods	Category	State Aggregation Scheme
SAmQ	Proposed method	SAmQ
NF-MLE	DDM	No aggregation
PQR	IRL	No aggregation
NF-MLE-SA	DDM	By state values
PQR-SA	IRL	By state values
PQR-SAmQ	IRL	SAmQ

Table: MSE for structural parameter estimation

Methods	Number of aggregated states n_s				
	5	10	50	100	1000
SAmQ	0.046 ± 0.045	0.014 ± 0.013	0.002 ± 0.001	0.001 ± 0.000	0.004 ± 0.001
NF-MLE-SA	5.254 ± 2.860	1.569 ± 1.218	0.012 ± 0.003	0.003 ± 0.003	0.008 ± 0.002
PQR-SA	0.334 ± 0.001	0.355 ± 0.019	0.355 ± 0.036	0.332 ± 0.003	0.337 ± 0.005
PQR-SAmQ	1.557 ± 0.173	0.383 ± 0.129	0.354 ± 0.018	0.377 ± 0.023	0.335 ± 0.004
NF-MLE			0.199 ± 0.020		
PQR			1.276 ± 0.094		



(a) SAmQ (b) Ad-hoc state aggregation

Figure: Aggregated states for a simple example with 2-dimensional states. A good aggregation ignores the dummy state, and aggregates by column.

References

[1] John Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, pages 999–1033, 1987.

[2] Sinong Geng, Houssam Nassif, Carlos Manzanares, Max Reppen, and Ronnie Sircar. Deep PQR: Solving inverse reinforcement learning using anchor actions. In *International Conference on Machine Learning*, pages 3431–3441, 2020.