Equilibrium in Misspecified Markov Decision Processes

Ignacio Esponda & Demian Pouzo

TE 2021

May 30, 2025

- MDP(Q)

$$v(x) = \max_{a \in \Gamma(x)} \int [r(x, a, x') + \beta v(x')] \, Q(x, a, \mathrm{d}x')$$

where Q is the true transition kernel

- Consider an agent is uncertain about the true Q

- They introduce SMDP$(Q, \mathbb{Q})$
  1. $MDP(Q)$ with the true Q
  2. a nonempty family of transition kernels $\mathbb{Q} := \{Q_\theta : \theta \in \Theta\}$

- Each period: observe x, choose a, and then update belief $\mu \in \mathscr{D}(\Theta)$

- Research question: how to describe the agent's steady-state behavior?

- Their answer: define 'Berk-Nash equilibrium' (some $m \in \mathscr{D}(\mathsf{G})$)

My intuition

1. Nobody knows the true relationship between Y and X
2. For simplification, people study $Y = \beta X + \varepsilon$
3. Question: what's the best β?
4. Answer: OLS is the best linear!

Definition

A distribution over state-action pairs $m \in \mathscr{D}(\mathrm{G})$ is a *Berk-Nash equilibrium* of the SMDP$(Q, \mathbb{Q})$ if the following conditions hold:

1. There exists a belief $\mu \in \mathscr{D}(\Theta)$ such that
   1.1 ***Optimality***. (best on average)
   For all $(x, a) \in \mathrm{G}$ such that $m(x, a) > 0$, a is optimal given x in the MDP$(\bar{Q}_\mu)$, where
   $$\bar{Q}_\mu = \int Q_\theta \mu(\mathrm{d}\theta)$$
   1.2 *Belief Restriction*.
2. *Stationarity*.

Definition

A distribution over state-action pairs $m \in \mathscr{D}(\mathsf{G})$ is a *Berk-Nash equilibrium* of the SMDP$(Q, \mathbb{Q})$ if the following conditions hold:

1. There exists a belief $\mu \in \mathscr{D}(\Theta)$ such that
   1.1 *Optimality*.
   1.2 ***Belief Restriction***. (closest to true)

$$\mu \in \mathscr{D}(\operatorname*{argmin}_{\theta \in \Theta} K_Q(m, \theta))$$

   where

$$K_Q(m, \theta) := \sum_{(x,a) \in \mathsf{G}} \mathbb{E}^Q_{(x,a)} \left[\log \left(\frac{Q(x, a, x')}{Q_\theta(x, a, x')} \right) \right] m(x, a)$$

   is weighted Kullback-Leibler divergence
2. *Stationarity*.

Definition

A distribution over state-action pairs $m \in \mathscr{D}(\mathsf{G})$ is a *Berk-Nash equilibrium* of the SMDP$(Q, \mathbb{Q})$ if the following conditions hold:

1. There exists a belief $\mu \in \mathscr{D}(\Theta)$ such that
   1.1 *Optimality*.
   1.2 *Belief Restriction*.
2. ***Stationarity***. (m_{X} is stationary if choosing the optimal action)
   For all $x' \in \mathsf{X}$,

$$
\begin{aligned}
m_{\mathsf{X}}(x') &= \sum_{(x,a)\in\mathsf{G}} Q(x, a, x')m(x, a) \\
&= \sum_{(x,a)\in\mathsf{G}} Q(x, a, x')m_{\mathsf{A}|\mathsf{X}}(a|x)m_{\mathsf{X}}(x)
\end{aligned}
$$

Existence

Theorem 1

If the following regularity conditions hold

1. *The parameter space* Θ *is a compact subset of an Euclidean space*
2. *The map* $\theta \mapsto Q_\theta(x, a, x')$ *is continuous for any* $(x, a, x') \in \mathsf{G} \times \mathsf{X}$
3. *There is a dense set* $\hat{\Theta} \subseteq \Theta$ *such that, for all* $\theta \in \hat{\Theta}$ *and* $(x, a, x') \in \mathsf{G} \times \mathsf{X}$,

$$Q(x, a, x') > 0 \quad \textit{implies} \quad Q_\theta(x, a, x') > 0$$

then there exists a Berk-Nash equilibrium for SMDP$(Q, \mathbb{Q})$

Identification

Proposition 2

Let m be a Berk-Nash equilibrium of the SMDP$(Q, \mathbb{Q})$. If the following conditions hold

1. $Q \in \mathbb{Q}$
2. *for any* $\theta, \theta' \in \operatorname{argmin}_{\theta \in \Theta} K_Q(m, \theta)$ *and* $(x, a) \in \mathsf{G}$

$$Q_\theta(x, a, \cdot) = Q_{\theta'}(x, a, \cdot)$$

then for all (x, a) in the support of m, a is optimal given x in the MDP(Q)

average = true → best on average = best on true

Bayesian Learning

Question: Under which condition the agent's steady-state behavior can be represented by a Berk-Nash equilibrium?

- The Bayesian agent's problem

$$v(x, \mu) = \max_{a \in \Gamma(x)} \int \{r(x, a, x') + \beta v(x', \mu')\} \bar{Q}_\mu(x, a, \mathrm{d}x')$$

- policy $\sigma \colon \mathsf{X} \times \mathscr{D}(\Theta) \to \mathscr{D}(\mathsf{A})$

- optimal policy σ: for any $(x, \mu, a) \in \mathsf{X} \times \mathscr{D}(\Theta) \times \mathsf{A}$

$$\sigma(x, \mu, a) > 0 \quad \text{implies} \quad a \text{ is a maximizer given} (x, \mu)$$

The convergence of time average

Notion of steady state: time average converges

- Let SMDP$(Q, \mathbb{Q})$ be regular and let σ be an optimal policy
- Let $m_t(h)$ be the frequency of state-action pairs up to time t
- Let $\mathbb{P}^{\sigma}$ be the probability distribution over histories induced by σ
- Suppose that there exists a positive $\mathbb{P}^{\sigma}$-measure set $\mathscr{H}$ such that $m_t(h) \to m$ for all histories $h \in \mathscr{H}$
  (So it does not imply uniqueness!!!)

Limiting Distribution is Berk-Nash Equilibrium

If one of the following two conditions holds, then m is a Berk-Nash equilibrium of SMDP$(Q, \mathbb{Q})$

1. iid → stationary
   r and all Q_θ do not depend on current state x, and
   for any $\theta, \theta' \in \mathrm{argmin}_{\theta \in \Theta} K_Q(m, \theta)$

$$Q_\theta(x, a, \cdot) = Q_{\theta'}(x, a, \cdot), \quad m - a.e.$$

2. unique closest → 100% believe → equilibrium
   for any $\theta, \theta' \in \mathrm{argmin}_{\theta \in \Theta} K_Q(m, \theta)$ and $(x, a) \in$ G

$$Q_\theta(x, a, \cdot) = Q_{\theta'}(x, a, \cdot)$$