Mean Field Competitive Binary MDPs and Structured Solutions

Minyi Huang

School of Mathematics and Statistics
Carleton University
Ottawa, Canada

MFG2017, UCLA, Aug 28–Sept 1, 2017

Outline of talk

- The competitive binary MDP model (Huang and Ma'16, 17)
- The stationary MFG equation (with discount)
  - Model features: monotone state dynamics, resetting action, positive externality
  - Check: ergodicity, existence, uniqueness, solution structure, pure strategies, stable/unstable equilibria
  - Exploit positive externality to prove uniqueness
- Extensions

Model: Individual controlled state $x_t^i \in [0,1]$, action $a_t^i \in \{0,1\}$.

Motivation for this class of competitive MDP models –

- Except for LQ cases, general nonlinear MFGs rarely have closed-form solutions
- We introduce this class of MDP models which have simple solution structures – threshold policy (see next page)
  - i) There is a long tradition of looking for threshold policies to various decision problems in the operations research community
  - ii) A threshold policy is also studied in R. Johari's ride sharing problem
- Sometimes simple models tell more; the binary MDP modeling is of interest in its own right

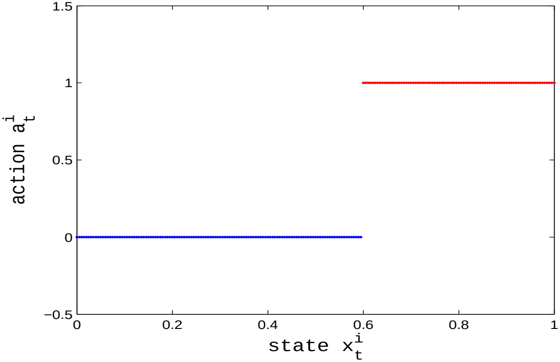


Figure : threshold is 0.6

- Dynamic games of MDPs were initial introduced by L. Shapley (1953) and called stochastic games.
- Literature on large population games with discrete states and/or actions (or discrete time MDP)
    - Weintraub et al. (2005, 2008), Huang (2012), Gomes et al. (2013), Adlakha, Johari, and Weintraub (2015), Kolokoltsov and Bensoussan (2016), Saldi, Basar and Raginsky (2017). Also see anonymous games e.g. Jovanovic and Rosenthal (1988)

The MF MDP model:

- N players (or agents $\mathcal{A}_i$, $1 \leq i \leq N$) with states x_t^i, $t \in \mathbb{Z}_+$, as controlled Markov processes (no coupling)
  - State space: $\mathbf{S} = [0,1]$ Interpret state as <u>unfitness</u>
  - Action space: $\mathbf{A} = \{a_0, a_1\}$. a_0: inaction, a_1: active control
- Agents are **coupled** via the costs

(Huang and Ma, CDC'16, CDC'17)

Examples of binary action space (action or inaction)

- maintenance of equipment;
- network security games;
- wireless medium access control (channel shared by users; transmission or back off; collisions possible)
- (flu) vaccination games, etc.

Binary choice models are widely used in various decision problems

- Schelling, T. (1973), Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities, *J. Conflict Resol.*
- Brock, W. and Durlauf, S. (2001), Discrete choice with social interactions, *Rev. Econ. Studies.*
- Schelling, T. (2006), Micromotives and Macrobehavior
- Babichenko, Y. (2013), Best-reply dynamics in large binary-choice anonymous games, *Games Econ. Behavior*
- Lee, L., Li, J. and Lin, X. (2014), Binary choice models with social network under heterogeneous rational expectations, *Rev. Econ. Statistics*

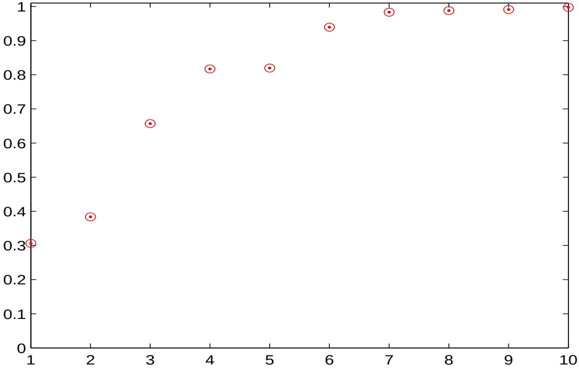
Dynamics:

The controlled transition kernel for x_t^i. For $t \geq 0$ and $x \in \mathbf{S}$,

$$P(x_{t+1}^i \in B | x_t^i = x, a_t^i = a_0) = Q_0(B|x),$$
$$P(x_{t+1}^i = 0 | x_t^i = x, a_t^i = a_1) = 1,$$

- $Q_0(\cdot|x)$: **stochastic kernel** defined for $B \in \mathcal{B}(\mathbf{S})$ (Borel sets).
- $Q_0([x,1]|x) = 1$. Under inaction a_0, state **gets worse**.
- Transition of x_t^i is not affected by other a_t^j, $j \neq i$.

The stochastic deterioration is similar to hazard rate modelling in the maintenance literature (Bensoussan and Sehti, 2007; Grall et al. 2002)



x_t^i under inaction. It is getting worse and worse.

The cost of $\mathcal{A}_i$:

$$J_i = E\sum_{t=0}^{T} \rho^t c(x_t^i, x_t^{(N)}, a_t^i), \quad 1 \le i \le N.$$

- $0 < T \le \infty$; $\rho \in (0,1)$: discount factor.
- Population average state: $x_t^{(N)} = \frac{1}{N}\sum_{i=1}^N x_t^i$.
- The one stage cost:

$$c(x_t^i, x_t^{(N)}, a_t^i) = R(x_t^i, x_t^{(N)}) + \gamma 1_{\{a_t^i = a_1\}}$$

- $R \ge 0$: unfitness-related cost; $\quad \gamma > 0$: effort cost

The MFG equation system (**MFG Eq**$_{[0,T]}$):

$$\begin{cases} V(t,x) = \min \Big[\rho \int_0^1 V(t+1,y) Q_0(dy|x) + R(x,z_t), \\ \qquad\qquad\qquad\qquad \rho V(t+1,0) + R(x,z_t) + \gamma \Big], \quad 0 \le t < T \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad V(T,x) = R(x,z_T), \\ z_t = E x_t^i, \quad 0 \le t \le T \qquad \text{consistency condition in MFG} \end{cases}$$

- Notation: $b_{s,t} = (b_s, b_{s+1}, \dots, b_t)$. Find a solution $(\hat{z}_{0,T}, \hat{a}^i_{0,T})$ such that $\{x^i_t, 0 \le t \le T\}$ is generated by $\{\hat{a}^i_t(x), 0 \le t \le T\}$ as best response.
- **Fact**: Given $z_{0,T}$, the best response is a **threshold policy**

Theorem: i) Under technical conditions, (**MFG Eq**$_{[0,T]}$) has a solution. ii) The resulting set of decentralized strategies for the N players is an ϵ-Nash equilibrium.

Below we will focus on its stationary version.

Action space $\{0,1\}$. Threshold policy:

$$a_t^i = \begin{cases} 1 & x_t^i \geq r \\ 0 & x_t^i < r \end{cases}$$

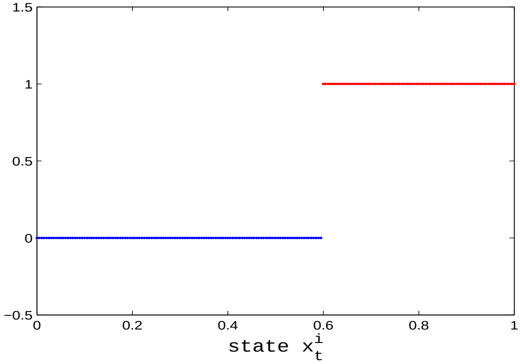


Figure : $r = 0.6$

Stationary equation system for MFG

$$V(x) = \min \left[\rho \int_0^1 V(y) Q_0(dy|x) + R(x,z), \quad \rho V(0) + R(x,z) + \gamma\right]$$

$$z = \int_0^1 x\pi(dx) \quad \text{for probability measure } \pi.$$

$(V, \hat{z}, \hat{\pi}, \hat{a}^i)$ is called a stationary equilibrium with discount if

i) the feedback policy $\hat{a}^i$ is the best response with respect to $\hat{z}$ (in the associated infinite horizon control problem);

ii) The distribut'n of $\{x_t^i, t \geq 0\}$ under $\hat{a}^i$ converges to the **stationary distribution** $\hat{\pi}$;

iii) $(\hat{z}, \hat{\pi})$ satisfies the second equation.

Note: This solution definition is different from that in **anonymous games** which only deal with a fixed point equation for a measure (describe the continuum population) **without considering ergodicity**

What we want to study?

- Q1 – Existence (based on ergodicity)
- Q2 – Threshold policy exists?
- Q3 – Uniqueness
- Q4 – Under what condition can a pure equilibrium strategy fail to exist?
- Q5 – How does the model parameter affect the solution? Related to comparative statistics
- Q6 – Does the equilibrium have stability, or can this fail? Can the model show oscillatory behaviors (as observed in vaccination situations)?

Assumptions:

(A1) $\{x_0^i, i \geq 1\}$ are independent random variables taking values in $\mathbf{S}$.

(A2) $R(x,z)$ is a continuous function on $\mathbf{S} \times \mathbf{S}$. For each fixed z, $R(\cdot, z)$ is strictly increasing.

(A3) i) $Q_0(\cdot|x)$ satisfies $Q_0([x,1]|x) = 1$ for any x, and is strictly stochastically increasing; ii) $Q_0(\cdot|x)$ has a positive density for all $x < 1$.

(A4) $R(x, \cdot)$ is increasing for fixed x. (**i.e. positive externality**)

(A5) $\gamma > \beta \max_z \int_0^1 [R(y,z) - R(0,z)] Q_0(dy|0)$.

Remarks:

- Montonicity in (A2): cost increases when state is poorer.
- (A3)-i) means dominance of distributions
- (A5) Effort cost should not be too low; this prevents zero action threshold; this condition can be refined if more information is known on R (such as sub-modularity).

Theorem (existence on Q1, Q2) Assume (A1)-(A5). Then

$$V(x) = \min \left[\rho \int_0^1 V(y) Q_0(dy|x) + R(x,z), \quad \rho V(0) + R(x,z) + \gamma\right]$$

$$z = \int_0^1 x\pi(dx)$$

has **a solution** (V, z, π, a^i) for which the best response a^i is a threshold policy.

Note:

- a^i: $x \in [0,1] \mapsto \{a_0, a_1\}$, is implicitly specified by the first equation
- On the right hand side, if the first term is smaller, then $a^i(x) = a_0$ (otherwise, a_1)

Theorem (uniqueness on Q3)

In the previous theorem, if we further assume $R(x, z) = R_1(x)R_2(z)$ and $R_2 > 0$ is strictly increasing on **S**, then

$$V(x) = \min \left[\rho \int_0^1 V(y)Q_0(dy|x) + R(x, z), \quad \rho V(0) + R(x, z) + \gamma\right]$$

$$z = \int_0^1 x\pi(dx)$$

has **a unique solution** (V, z, π, a^i) .

Remark: Monotonicity of R_2 indicates positive externalities (crucial in uniqueness analysis), i.e, a person benefits from the improving behaviour of the population.

How to prove the existence theorem?

- Show ergodicity of the controlled Makov chain $\{x_t^i\}$ under an interior threshold policy (so that π in the equation makes sense); the case of threshold ≥ 1 is handled separately
- Further use Brouwer's fixed point theorem to show existence

How to prove uniqueness?

- Use two comparison theorems

Theorem (ergodicity, Q1) : For threshold $\theta \in (0,1)$, $\{x_t^{i,\theta}, t \geq 0\}$ is uniformly ergodic with stationary probability distribution π_θ, i.e.,

$$\sup_{x \in \mathbf{S}} \|P^t(x,\cdot) - \pi_\theta\| \leq K r^t$$

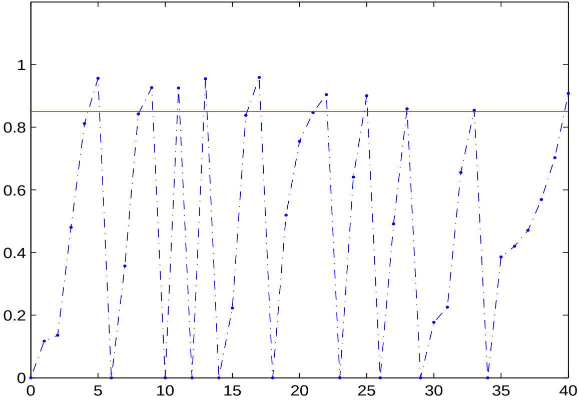
for some constants $K > 0$ and $r \in (0,1)$, where $\|\cdot\|$ is the total variation norm of signed measures.

Proof. Show Doeblin's condition by checking

$$\inf_{x \in \mathbf{S}} P(x_4 = 0 | x_0 = x) \geq \eta > 0.$$

and aperiodicity. Here 4 is the **magic number**.

The regenerative process:



x_t^i under threshold $\theta = 0.85$

Numerics: Recall the MFG solution equation system

$$V(x) = \min \left[\rho \int_0^1 V(y) Q_0(dy|x) + R(x,z), \quad \rho V(0) + R(x,z) + \gamma \right]$$

$$z = \int_0^1 x \pi(dx)$$

Algorithm:

- Step 1. Initialize $z^{(0)}$
- Step 2. Given $z^{(k)}$, solve $V^{(k)}(x)$ and associated threshold policy $a^{(k)}$ from the DP Eqn
- Step 3. Use $a^{(k)}$ to find the mean field $\bar{z}^{(k+1)}$ via $\pi^{(k)}$.
- Step 4. $z^{(k+1)} = 0.98 z^{(k)} + 0.02 \bar{z}^{(k+1)}$ (cautious update!)
- Step 5. Go back to step 2 (until accuracy attained).

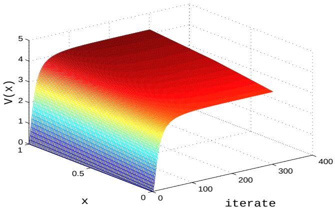


Figure : $V(x)$ defined on $[0, 1]$

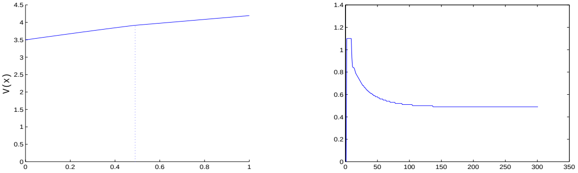


Figure : Left: $V(x)$ (threshold 0.49) Right: search of the threshold 0.49

The comparison theorems use very intuitive ideas. Used to answer **Q3**.

- **First comparison theorem** (with $R(x, z) = R_1(x)R_2(z)$): If $z \downarrow$ (i.e., decreasing population threat), best response's threshold $\theta \uparrow$
- **Second comparison theorem**: If the threshold $\theta \in (0, 1) \uparrow$, the resulting regenerative process $x_t^{i\,\theta}$'s long term mean $\uparrow$ (since more chance to "grow"; lawn, say, under biweekly cut)

Method to prove the second comparison theorem: Take the threshold $\theta \in (0,1)$. Set $x_0^{i,\theta} = 0$, and the Markov chain $x_t^{i,\theta}$ evolves under inaction until τ_θ:

$$\tau_\theta = \inf\{t | x_t^{i,\theta} \geq \theta\}.$$

Reset $x_{\tau_\theta+1}^{i,\theta} = 0$, and the Markov chain restarts at $\tau_\theta + 1$.
Denote $S_k = \sum_{t=0}^{k} x_t^{i,\theta}$. Let S_k' be defined with $\theta' \in (\theta, 1)$. Then

$$\lambda_\theta := \lim_{k\to\infty} \frac{1}{k} \sum_{t=0}^{k-1} x_t^{i,\theta} = \frac{ES_{\tau_\theta}}{1 + E\tau_\theta}, \quad \text{similarly define } \lambda_{\theta'}.$$

The second comparison theorem: For $0 < \theta < \theta' < 1$, then

i) $\frac{ES_{\tau_\theta}}{1+E\tau_\theta} \leq \frac{ES'_{\tau_{\theta'}}}{1+E\tau_{\theta'}}$

ii) So $\lambda_\theta \leq \lambda_{\theta'}$

A useful lemma for proving the second comparison theorem.
Let $0 < r < r' < 1$. Consider a Markov process $\{Y_t, t \geq 0\}$ with state space $[0,1]$, and transition kernel $Q_Y(\cdot|y)$ which satisfies $Q_Y([y,1]|y) = 1$ for any $y \in [0,1]$, and is stochastically monotone. Suppose $Y_0 \equiv y_0 < r$. Define the stopping time

$$\tau = \inf\{t | Y_t \geq r\}$$

Lemma If $E\tau < \infty$, then $E\sum_{t=0}^{\tau} Y_t < \infty$ and

$$\frac{E\sum_{t=0}^{\tau} Y_t}{1 + E\tau} = \frac{EY_0 + EY_1 + \sum_{k=1}^{\infty} E(Y_{k+1}1_{\{Y_k<r\}})}{2 + \sum_{k=1}^{\infty} P(Y_k < r)}$$

Facts:

- Apply the lemma to the regenerative process $x_t^{i,\theta}$ for one cycle.
- Show that increasing $r = \theta$ to $r' = \theta'$ increases the ratio in the lemma.

Example (to answer Q4): Assume $R(x, z) = x(c + z)$ where $c > 0$. $Q_0(\cdot|x)$ has the same distribution as $x + (1 - x)\xi$ where ξ has uniform distribution on $[0, 1]$. Consider all $\gamma > 0$. Denote $c^* = \frac{E\xi}{2} = \frac{1}{4}$.

Then

- If $\gamma \in (0, \frac{\rho c}{2}] \cup (\frac{\rho(c+c^*)}{2}, \infty)$, there exists a unique pure strategy as a stationary equilibrium with discount
- If $\gamma \in (\frac{\rho c}{2}, \frac{\rho(c+c^*)}{2}]$, there exists **no pure strategy** as a stationary equilibrium with discount

This result can be extended to

- the general model for $R(x, z)$
- the long run average cost case

Q5: Check the dependence of equilibrium solution (z, θ) on γ.

This is actually a question on comparative statistics formalized by J. R. Hicks (1939) and P. A. Samuelson (1947)

- It studies how a change of the model parameters affects the equilibrium
- Important in the economic literature, game theory and optimization; however, most literature considered static models
- Example of dynamic models: D. Acemoglu, M. K. Jensen (2015) J. Political Econ. (on large dynamic economies)

Theorem (monotone comparative statistic, Q5). Suppose

- both γ_1 and γ_2 (effort cost) satisfy (A5), and $\gamma_1 < \gamma_2$;
- $(\gamma_i, z_i, \theta_i)$ is solved from the MFG equation system.

Then

$$\theta_1 < \theta_2, \quad z_1 < z_2.$$

Interpretation:
Effort being more expensive → less willing to act → worse average state of the population

Q6. Instability of the equilibrium solution can occur

- Take $R(x,z) = x(c + R_0(z))$ where $R_0(z)$ has a quick leap from a near 0 value to nearly 1 around some value z_c
- Such R_0 is not purely artificial; similar phenomena appear in vaccination models where group risk has sharp decrease toward 0 when the vaccination coverage exceeds a certain critical level.

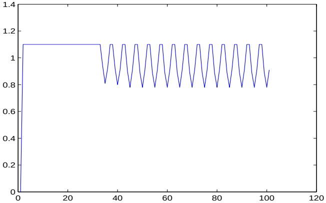


Figure : Iteration of threshold in **best response** w.r.t. mean field; swing back and forth

Game with long-run average costs: Define

$$J^i(x_0^1, \dots, x_0^N, \pi^1, \dots, \pi^N) = \limsup_{T \to \infty} \frac{1}{T} E \sum_{t=0}^{T-1} c(x_t^i, x_t^{(N)}, a_t^i),$$

where $c(x_t^i, x_t^{(N)}, a_t^i) = R(x_t^i, x_t^{(N)}) + \gamma 1_{\{a_t^i = a_1\}}$
The solution equation system of the MFG:

$$\begin{cases} \lambda + h(x) = \min\{\int_0^1 h(y) Q_0(dy|x) + R(x,z), \quad R(x,z) + \gamma\}, \\ z = \int_0^1 x \nu(dx), \end{cases}$$

where ν is the limiting distribution of the closed-loop state X_t^i.
Idea:

- Consider discount factor ρ and denote value function $V_\rho(x)$.
- Define $h_\rho(x) = V_\rho(x) - V_\rho(0)$. Try to get a subsequence of $\{h_\rho\}$ converging to h when $\rho \uparrow 1$ under an additional concavity condition on $R(\cdot, z)$ for each fixed z.

Further possible extensions:

- Unbounded state space (such as half line)
- Pairwise coupling in the cost
- Continuous time modeling (example, compound Poisson process with impulse control; CDC'17, with Zhou); a Levy process with one-sided jumps (such as a subordinator) also fits the monotone state modeling

Thank you!