Discrete Mean Field Games

Josu Doncel Nicolas Gast Bruno Gaujal Jean-Yves Le Boudec

NPCG, April, 2019

Mean Field Interaction and Optimization

In a system with mean field interaction, compare the optimal strategy of the system with the optimal strategy of the mean field limit.

Mean Field Interaction and Optimization

In a system with mean field interaction, compare the optimal strategy of the system with the optimal strategy of the mean field limit.

Mean field interaction results in an aggregate information that may be harmful for efficient optimization.

This tension (information - optimisation) can be seen in centralized optimization as well as in games.

In the general case, this tension may lead to coarse results.

Interesting examples lie out of these general cases.

Summary of the talk

Part 1: Mean Field Markov Decision Processes

We consider a Markovian system composed of N *objects*. Each object lives in a finite set $\mathcal{S} = \{1 \dots S\}$.

The state of the system at time k is $X^N(k) := \left(X_1^N(k) \dots X_N^N(k)\right)$, with transition P

For all $i \in \mathcal{S}$, $M^N(k)$ the empirical distribution of the objects at time k:

$$M^N(k) := \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(k)},$$

At every time k, a centralized controller chooses an action $A^N(k) \in \mathcal{A}$ where $\mathcal{A}$ (compact).

Exchangeability: Objects are exchangeable ($\sigma^{-1} P \sigma = P$) and only observable through their states.

Part 1: Mean Field Markov Decision Processes

We consider a Markovian system composed of N *objects*. Each object lives in a finite set $\mathcal{S} = \{1 \dots S\}$.

The state of the system at time k is $X^N(k) := \left(X_1^N(k) \dots X_N^N(k)\right)$, with transition P

For all $i \in \mathcal{S}$, $M^N(k)$ the empirical distribution of the objects at time k:

$$M^N(k) := \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(k)},$$

At every time k, a centralized controller chooses an action $A^N(k) \in \mathcal{A}$ where $\mathcal{A}$ (compact).

Exchangeability: Objects are exchangeable ($\sigma^{-1} P \sigma = P$) and only observable through their states.

This implies that $M^N()$ *is a controlled Markov chain.*

Drift, Value Function, Policies

For every k, a *policy* π_k is a function $\mathcal{P}(\mathcal{S}) \to \mathcal{A}$. Let $M_\pi^N(k)$ be the state distribution at time k under π. $r(m, a)$ is the reward under distribution m and action a.

The value of a policy π over the horizon $[0; H^N]$ starting from m_0 is

$$V_\pi^N(m) := \mathbb{E}\Big(\sum_{k=0}^{H^N-1} r(M_\pi^N(k), \pi(M_\pi^N(k))) \Big| M_\pi^N(0) = m \Big).$$

The goal of the controller is to find an optimal policy that maximizes the value. We denote by $V_*^N(m)$ the optimal value when starting from m:

$$V_*^N(m) = \sup_\pi V_\pi^N(m)$$

Mean Field Limit

The drift $F^N(m,a)$ is:

$$F^N(m,a) = \mathbb{E}\Big(M^N(k+1) - M^N(k) \mid M^N(k) = m, A^N(k) = a\Big).$$

Assume $\left\|NF^N(m,a) - f(m,a)\right\| \to 0$, unif. in m and a (can be relaxed).

Limit system: Action function $\alpha : [0;T] \to \mathcal{A}$ is piecewise Lipschitz continuous.

$$m(t) = m(0) + \int_0^t f(m(s), \alpha(s))ds = \phi_t(m_0, \alpha) \qquad (1)$$

$$v_\alpha(m_0) := \int_0^T r\left(\phi_s(m_0,\alpha), \alpha(s)\right) ds \qquad (2)$$

The optimal value of the deterministic limit system $v_*(m_0)$:

$$v_*(m_0) = \sup_\alpha v_\alpha(m_0), \qquad (3)$$

where the supremum is taken over all action functions from $[0;T] \to \mathcal{A}$.

Convergence to the Mean Field Limit

If r and f are Lipschitz continuous then,

Theorem (convergence of the value)

$$\lim_{N\to\infty} V_*^N\left(M^N(0)\right) = v_*(m_0) \quad \textit{when} \quad M^N(0) \to m_0$$

Theorem (Asymptotically Optimal Policy)

If α_* is an optimal action function for the limiting system and if $\lim_{N\to\infty} M^N(0) = m_0$ almost surely [resp. in probability], then we have:

$$\lim_{N\to\infty} \left| V_{\alpha_*}^N - V_*^N \right| = 0,$$

almost surely [resp. in probability].

Sketch of the proof

We construct 2 auxiliary systems:

- The random flow $\phi_t(m_0, A_\pi^N)$, driven by policy π.
- the Markovian system M_α^N, driven action function α.

Sketch of the proof

We construct 2 auxiliary systems:

- The random flow $\phi_t(m_0, A^N_\pi)$, driven by policy π.
- the Markovian system M^N_α, driven action function α.

Using the first system, $\mathbb{E}v(A^N_{\pi^*}) \leq v(\alpha^*)$
One can show that $V^N_{\pi^*} - \mathbb{E}v(A^N_{\pi^*}) \to 0$ (Kurtz's result)
This implies $\lim V^N_{\pi^*} \leq v(\alpha^*)$.

Using the second system, $V^N_{\alpha^*} \leq V^N_{\pi^*}$.
One can show that $v(\alpha^*) - V^N_{\alpha^*} \to 0$.
This implies $\lim V^N_{\pi^*} \geq v(\alpha^*)$

Convergence rate to the Mean Field Limit

Theorem

for any action function,

$$\sqrt{N}\left(M_\alpha^N(t) - m_\alpha(t)\right) \xrightarrow{\text{Law}} G_t.$$

G_t is Gaussian.

Theorem

There exists $\beta > 0$ s.t.

$$\sqrt{N}\left|V_*^N(m_0) - v_*(m_0)\right| \leq \beta + o(1).$$

Conclusion I

As N grows, policies that do not take into account the state of the system (*i.e.* open loop) are asymptotically as good as adaptive policies (closed loop).

This result does not hold in general for long run average costs, even if the mean field limits under all actions have the same global attactor.

Part 2: Repeated Games with Random Outcomes

Consider the model investigated by Green (1980) and Sabourian (1990). Several similar results exist (Al-Najjar, Smorodinsky, 2001) with variations on the interactions between players.

A repeated game $H^{\infty} = (K, A, X, F, \delta, r)$.

K is the set of players,
A is the set of strategies,
X the set of outcomes,
Random outcome distribution $F(a) \in \mathcal{P}(X)$
$0 < \delta < 1$ is the discount factor,
The stage reward of player k is $r_k(a_k, x)$ (bounded)

The stage expected reward for player k is $R_k(a) = \int_X r_k(a_k, x) dF(a)(x)$.

Total reward for player k: $\mathbb{E} \sum_{t=0}^{\infty} \delta^t r_k(a_k(t), x(t))$

Anonymous Players

Players are anonymous implies the diagram commutes:

$$\begin{tikzcd}
 & \mathcal{P}(A) \arrow[dr, "G"] & \\
A \arrow[ur, "emp"] \arrow[rr, "F"] & & \mathcal{P}(X)
\end{tikzcd}$$

Anonymous Players

Players are anonymous implies the diagram commutes:

$$\begin{tikzcd} & \mathcal{P}(A) \arrow[dr, "G"] & \\ A \arrow[ur, "emp"] \arrow[rr, "F"] & & \mathcal{P}(X) \end{tikzcd}$$

Main assumption for the following result to hold: The function G is continuous if $\mathcal{P}(A)$ is endowed with the weak topology and $\mathcal{P}(X)$ with the total variation norm.

SPE is ϵ-Nash

Theorem (Sabourian, 1990)

For any ε, $\exists N$ s.t. $\forall$ repeated game H^∞ with $K \geq N$ anonymous players, any SPE s is such that $s_t(h_t)$ is an ε-equilibirum of H, for all histories h_t.

SPE is ϵ-Nash

Theorem (Sabourian, 1990)

For any ε, $\exists N$ s.t. $\forall$ repeated game H^∞ with $K \geq N$ anonymous players, any SPE s is such that $s_t(h_t)$ is an ε-equilibirum of H, for all histories h_t.

Proof sketch: Let s be an SPE, $\tilde{a}_k$ any action, $s' = ((\tilde{a}_k, s^0_{-k}), s^1)$.

SPE is ϵ-Nash

Theorem (Sabourian, 1990)

For any ε, $\exists N$ s.t. $\forall$ repeated game H^∞ with $K \geq N$ anonymous players, any SPE s is such that $s_t(h_t)$ is an ε-equilibirum of H, for all histories h_t.

Proof sketch: Let s be an SPE, $\tilde{a}_k$ any action, $s' = ((\tilde{a}_k, s^0_{-k}), s^1)$.

$$\mathbb{E}r_k^\infty(s) = R_k(s^0) + \int_X V_k(s(1), x) dF(s^0)(x)$$

SPE is ϵ-Nash

Theorem (Sabourian, 1990)

For any ε, $\exists N$ s.t. $\forall$ repeated game H^∞ with $K \geq N$ anonymous players, any SPE s is such that $s_t(h_t)$ is an ε-equilibirum of H, for all histories h_t.

Proof sketch: Let s be an SPE, $\tilde{a}_k$ any action, $s' = ((\tilde{a}_k, s^0_{-k}), s^1)$.

$$\mathbb{E}r_k^\infty(s) = R_k(s^0) + \int_X V_k(s(1), x)dF(s^0)(x)$$

$$\mathbb{E}r_k^\infty(s') = R_k(\tilde{a}_k, s^0_{-k}) + \int_X V_k(s'(1), x)dF(s'^0)(x)$$

SPE is ϵ-Nash

Theorem (Sabourian, 1990)

For any ε, $\exists N$ s.t. $\forall$ repeated game H^∞ with $K \geq N$ anonymous players, any SPE s is such that $s_t(h_t)$ is an ε-equilibirum of H, for all histories h_t.

Proof sketch: Let s be an SPE, $\tilde{a}_k$ any action, $s' = ((\tilde{a}_k, s^0_{-k}), s^1)$.

$$\mathbb{E}r_k^\infty(s) = R_k(s^0) + \int_X V_k(s(1), x) dF(s^0)(x)$$

$$\mathbb{E}r_k^\infty(s') = R_k(\tilde{a}_k, s^0_{-k}) + \int_X V_k(s'(1), x) dF(s'^0)(x)$$

Since $\mathbb{E}r_k^\infty(s) \geq \mathbb{E}r_h^\infty(s')$, then

$$\begin{aligned} R_k(\tilde{a}_k, s^0_{-k}) - R_k(s^0) &\leq \int_X V_k(s(1), x) dF(s^0)(x) - \int_X V_k(s'(1), x) dF(s'^0) \\ &\leq W \int_X |dF(s^0)(x) - dF(s'^0)(x)|. \end{aligned}$$

This is the TV distance between outcome measures. $M(s'^0)$ and $M(s^0)$ are close for the weak norm when N grows, so the TV distance is also small.

Conclusion II

As N grows, All SPE are ϵ-greedy at each time t, under mean field interaction and continuity of the signals.

Mean field games combine the two previous models: Large number of players and Markovian evolution.
The continuity assumption on G does not hold for mean field games under the natural construction, given below.

Part 3: Discrete MFG: Dynamic Population Game

State and action sets $\mathcal{E} = \{1, \dots, E\}$, $\mathcal{A} = \{1, \dots, A\}$
Population distribution $\mathbf{m}(t)$ in $\mathcal{P}(\mathcal{E})$.
Mixed strategy $\pi_i(t)$ in $\mathcal{P}(\mathcal{A})$.

Part 3: Discrete MFG: Dynamic Population Game

State and action sets $\mathcal{E} = \{1, \ldots, E\}$, $\mathcal{A} = \{1, \ldots, A\}$

Population distribution $\mathbf{m}(t)$ in $\mathcal{P}(\mathcal{E})$.

Mixed strategy $\pi_i(t)$ in $\mathcal{P}(\mathcal{A})$.

Rate matrices $Q_{ija}(\mathbf{m}(t))$ is a rate at which a player in state i moves to state j when choosing action a, when the population distribution is $\mathbf{m}(t)$.

Population Evolution $\mathbf{m}^\pi(t) \in \mathcal{P}(\mathcal{E})$ the population distribution at t, under π.

Its evolution is similar to Mean field MDP (with explicit rate):

$$m_j^\pi(t) = m_j(0) + \int_0^t \left(\sum_{i \in \mathcal{E}} \sum_{a \in \mathcal{A}} m_i^\pi(u) Q_{ija}(\mathbf{m}^\pi(u)) \pi_{i,a}(u) \right) du.$$

Part 3: Discrete MFG: Dynamic Population Game

State and action sets $\mathcal{E} = \{1, \ldots, E\}$, $\mathcal{A} = \{1, \ldots, A\}$

Population distribution $\mathbf{m}(t)$ in $\mathcal{P}(\mathcal{E})$.

Mixed strategy $\pi_i(t)$ in $\mathcal{P}(\mathcal{A})$.

Rate matrices $Q_{ija}(\mathbf{m}(t))$ is a rate at which a player in state i moves to state j when choosing action a, when the population distribution is $\mathbf{m}(t)$.

Population Evolution $\mathbf{m}^{\pi}(t) \in \mathcal{P}(\mathcal{E})$ the population distribution at t, under π.

Its evolution is similar to Mean field MDP (with explicit rate):

$$m_j^{\pi}(t) = m_j(0) + \int_0^t \left(\sum_{i \in \mathcal{E}} \sum_{a \in \mathcal{A}} m_i^{\pi}(u) Q_{ija}(\mathbf{m}^{\pi}(u)) \pi_{i,a}(u) \right) du.$$

Explicit interactions

$Q_{ija}(\mathbf{m}^{\pi}(t))$ depends explicitly on the population distribution. Other mean field models, such as Gomes 2010, only consider the special case where $Q_{ija}(\mathbf{m}^{\pi}(t)) = Q_{ija}$. This makes the population dynamics linear.

Cost function

Player 0 chooses her own strategy $\pi^0 : \mathbb{R} \times \mathcal{E} \to \mathcal{P}(\mathcal{A})$.
$\mathbf{x}^{\pi^0,\pi}(t) \in \mathcal{P}(\mathcal{E})$ is the probability distribution of Player 0 state when she uses strategy π^0 against a population who plays strategy π.
Player 0 has an instantaneous cost $c_{i,a}(\mathbf{m}^{\pi}(t))$ in state i when using action a.
The total discounted cost of Player 0 is

$$V(\pi^0, \pi) = \int_0^{\infty} \left(\sum_{i \in \mathcal{E}} \sum_{a \in \mathcal{A}} x_i^{\pi^0,\pi}(t) c_{i,a}(\mathbf{m}^{\pi}(t)) \pi^0_{i,a}(t) e^{-\beta t} \right) \, dt$$

$\beta > 0$ is a discount factor.

Best Response

The best response of Player 0 to π is $BR(\pi)$, the strategies that minimize her discounted cost:

$$BR(\pi) := \arg\min_{\pi^0 \in S} V(\pi^0, \pi).$$

Definition (Mean Field Equilibrium (MFE))

A strategy π is a mean field equilibrium if it is a fixed point for the best-response correspondance:

$$\pi^{MFE} \in BR(\pi^{MFE}).$$

Existence of a MFE

- The rate functions $Q_{ija}(\mathbf{m})$ are Lipschitz-continuous in $\mathbf{m}$.
- The cost functions $c_{i,a}(\mathbf{m})$ are continuous in $\mathbf{m}$.

Theorem

Any discrete mean field game $\mathcal{G}$ whose rate and cost satisfy the assumptions above admits a mean field equilibrium.

Previous existence proofs based on (strict) convexity of c_{ia}.
The best-response function $\pi \mapsto BR(\pi)$ is not hemi-continuous ($BR(\pi)$ may not be convex).

Existence of a MFE

- The rate functions $Q_{ija}(\mathbf{m})$ are Lipschitz-continuous in $\mathbf{m}$.
- The cost functions $c_{i,a}(\mathbf{m})$ are continuous in $\mathbf{m}$.

Theorem

Any discrete mean field game $\mathcal{G}$ whose rate and cost satisfy the assumptions above admits a mean field equilibrium.

Previous existence proofs based on (strict) convexity of c_{ia}.
The best-response function $\pi \mapsto BR(\pi)$ is not hemi-continuous ($BR(\pi)$ may not be convex).
Instead, we consider a fixed point equation in $\mathbf{m}$ by defining $\phi(\mathbf{m}) = \{x^{\pi^0} \forall \pi^0 \in arg \min_{\pi} V(\pi, \mathbf{m})\}$.
$V(\pi, m)$ is continuous for the weak topology implies ϕ is semi-upper continuous and compact. This implies that ϕ has a fixed point.

Tightness of the assumptions

- If Q is not Lipschitz-continuous in $\mathbf{m}$, then the evolution of the population is not well defined: it may have several solutions or none.
- There exist games with non-continuous cost functions c that do not admit mean field equilibrium. Example:

$$\mathcal{G} = \left(\mathcal{E} = \{1,2\}, \mathcal{A} = \{a,b\}, Q_a = 0, Q_b = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, m(0) = (1,0) \right.$$

$$\left. c_a(m_1,m_2) = 0, c_b(m_1,m_2) = \begin{cases} -1 & \text{if } m_2 \leq 1/2 \\ 1 & \text{otherwise} \end{cases} \right).$$

Tightness of the assumptions

- If Q is not Lipschitz-continuous in $\mathbf{m}$, then the evolution of the population is not well defined: it may have several solutions or none.
- There exist games with non-continuous cost functions c that do not admit mean field equilibrium. Example:

$$\mathcal{G} = \left(\mathcal{E} = \{1,2\}, \mathcal{A} = \{a,b\}, Q_a = 0, Q_b = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, m(0) = (1,0) \right.$$

$$\left. c_a(m_1,m_2) = 0, c_b(m_1,m_2) = \begin{cases} -1 & \text{if } m_2 \leq 1/2 \\ 1 & \text{otherwise} \end{cases} \right).$$

$m_2(t)$ is non-decreasing. Let $\tau = \sup\{t : m_2(t) \leq 1/2\}$
The best-response of Player 0 to any $m(\cdot)$ is policy $\pi^{(\tau)}$:

Play "b" until τ and "a" after τ.

This cannot be a MFE: for any T, under policy $\pi^{(T)}$, $m_2(t) = 1 - e^{-\min(t,T)}$, so that $\tau = \ln 2$ or $+\infty$.

$\pi^{(\ln 2)}$ is the best response to $\pi^{(\infty)}$ and vice-versa.

Markov Games with N Exchangeable Players

To a MFG $\mathcal{G}$ we associate a stochastic N-player game $\mathcal{G}^N$ (the construction can also be done from the N player game to the mean field limit).

Markov Games with N Exchangeable Players

To a MFG $\mathcal{G}$ we associate a stochastic N-player game $\mathcal{G}^N$ (the construction can also be done from the N player game to the mean field limit).
In $\mathcal{G}^N$, Each player becomes active according to a Poisson process with rate 1, independently of the others.

Markov Games with N Exchangeable Players

To a MFG $\mathcal{G}$ we associate a stochastic N-player game $\mathcal{G}^N$ (the construction can also be done from the N player game to the mean field limit).
In $\mathcal{G}^N$, Each player becomes active according to a Poisson process with rate 1, independently of the others.
This can be extended to a set of active players $R(t)$ whose size is bounded in L_1 and L_2.

Markov Games with N Exchangeable Players

To a MFG $\mathcal{G}$ we associate a stochastic N-player game $\mathcal{G}^N$ (the construction can also be done from the N player game to the mean field limit).
In $\mathcal{G}^N$, Each player becomes active according to a Poisson process with rate 1, independently of the others.
This can be extended to a set of active players $R(t)$ whose size is bounded in L_1 and L_2.
State $\mathbf{X}(t) = (X_1(t) \dots X_N(t)) \in \mathcal{E}^N$ is a continuous time Markov chain.

Markov Games with N Exchangeable Players

To a MFG $\mathcal{G}$ we associate a stochastic N-player game $\mathcal{G}^N$ (the construction can also be done from the N player game to the mean field limit).
In $\mathcal{G}^N$, Each player becomes active according to a Poisson process with rate 1, independently of the others.
This can be extended to a set of active players $R(t)$ whose size is bounded in L_1 and L_2.
State $\mathbf{X}(t) = (X_1(t) \dots X_N(t)) \in \mathcal{E}^N$ is a continuous time Markov chain.
If we assume

- Mean field interaction between the players.
- time uniformization (with intensity N): $(t_n)_{n \in \mathbb{Z}^+}$ is a Poisson process with rate N.

$$\mathbb{P}\left(X_k(t_{n+1}) = j | X_k(t_n) = i, k = R(t_n), \mathbf{M}(t_n) = \mathbf{m}, A_k(t_n) = a\right) = \frac{Q_{ija}(\mathbf{m})}{N}$$

Costs

Given a strategy π^k used by player k and a strategy π used by all the others, $V(\pi^k, \pi)$ is the expected discounted cost of player n:

$$V^N(\pi^k, \pi) = \mathbb{E}\left[\sum_{t \in \mathcal{T}_N} e^{-\beta t} c_{X_k(t), A_k(t)}(\mathbf{M}^\pi(t)) \,\middle|\, \begin{array}{l} A_k \text{ chosen w.r.t. } \pi^k \\ A_{-k} \text{ chosen w.r.t. } \pi \end{array}\right].$$

Equilibrium of the N player game

Definition (Equilibrium of the N player game)

For a given set of strategies S, a strategy $\pi \in S$ is called a symmetric equilibrium in S if for any strategy $\pi^k \in S$:

$$V^N(\pi, \pi) \leq V^N(\pi^k, \pi).$$

When all other players apply the strategy π, then π minimizes the objective of the kth player over all possible strategies in S.

Subsets of Admissible Strategies

In a full information setting, $A_k(t)$ is a (possibly random) function of all $X(t')$, and all $A(t')$, $t' < t$. Such a strategy is hard to analyze. Therefore, in the following, we will consider two natural subclasses for the set of admissible strategies:

- (Markov) – A strategy π is Markovian (and stationary) if $A_k(t)$ is a (possibly random) function of $\mathbf{M}(t)$ and $\mathbf{X}(t)$:

$$\mathbb{P}\left(A_k(t) = a \mid \mathcal{F}_t\right) = \pi_{a,X_k(t)}(\mathbf{M}(t)).$$

- (Local) – A strategy π is a local if the strategy only depends on the player's internal state and on time.

$$\mathbb{P}\left(A_k(t) = a \mid \mathcal{F}_t\right) = \pi_{a,X_k(t)}(t).$$

  Under local strategies, the actions may depend on time, hence may track the law of the population $\mathbf{M}(t)$ (but not $\mathbf{M}(t)$ itself). Also notice that a local strategy is not necessarily Markovian because of its dependence on time.

Nash Equilibria Limits

Theorem

(i) *Let π be an equilibrium of $\mathcal{G}$. There exists N_0 s.t. $\forall N \geq N_0$, π is a local ε-equilibrium of the N-player game. (Cecchin, Fisher 2017).*

(ii) *If $(\pi^N)_N$ is a sequence of local equilibria for the N player game, there is a sub-sequence that converges weakly to a mean field equilibrium of $\mathcal{G}$.*

Proof: Based on Tembine, 2009 or Kolokotsov (value of local strategies converges uniformly).

Markov Equilibria May Not Converge to MFE

Let us consider a matching game version of the prisoner's dilemma. The state space: $\mathcal{S} = \{A, B\}$ and $\mathcal{A} = \mathcal{S}$. Population distribution is $\mathbf{m} = (m_A, m_B)$.

Cost of a player:

$$C(i, i, \mathbf{m}) = \begin{cases} m_A + 3m_B & \text{if } i = A \\ 2m_B & \text{if } i = B \end{cases}$$

This is the expected cost of a player matched with another player at random and using the cost matrix:

	A	B
A	$1, 1$	$3, 0$
B	$0, 3$	$2, 2$

Lemma

Always playing B is the unique mean-field equilibrium.

Non-convergence of Markov strategies (II)

Let us define the following stationary strategy for N players:

$$\pi^N(\mathbf{M}) = \begin{cases} B & \text{if } M_A < 1 \\ A & \text{if } M_A = 1. \end{cases}$$

"play A as long as my opponent plays A. Play B forever as soon as my opponent plays B."

Non-convergence of Markov strategies (II)

Let us define the following stationary strategy for N players:

$$\pi^N(\mathbf{M}) = \begin{cases} B & \text{if } M_A < 1 \\ A & \text{if } M_A = 1. \end{cases}$$

"play A as long as my opponent plays A. Play B forever as soon as my opponent plays B."

Lemma

For $\beta < 1$ and any N, π^N is a sub-game perfect equilibrium of the N-player stochastic game.

Non-convergence of Markov strategies (III)

Assume all players, except player 0, use strategy π^N and let us compute the best response of player 0 to π^N.

If player 0 plays A, its cost is $\frac{1}{N}\sum_{i=0}^{\infty} e^{-\beta i/N} = 1/\beta + O(1/N)$.

If player 0 chooses action B, then its cost is $\int_0^\infty 2M_B e^{-\beta t} dt + O(1/N) = \int_0^\infty \frac{2e^{-\beta t}}{1+(N-1)e^{-Nt}} dt + O(1/N) \geq 1/\beta + O(1/N)$.

Non-convergence of Markov strategies (III)

Assume all players, except player 0, use strategy π^N and let us compute the best response of player 0 to π^N.

If player 0 plays A, its cost is $\frac{1}{N}\sum_{i=0}^{\infty} e^{-\beta i/N} = 1/\beta + O(1/N)$.

If player 0 chooses action B, then its cost is $\int_0^{\infty} 2M_B e^{-\beta t} dt + O(1/N) =$
$\int_0^{\infty} \frac{2e^{-\beta t}}{1+(N-1)e^{-Nt}} dt + O(1/N) \geq 1/\beta + O(1/N)$.

Therefore, when $\beta < 1$, player 0 has no incentive to deviate from π^N: it is a sub-game perfect equilibrium.

Non-convergence of Markov strategies (III)

Assume all players, except player 0, use strategy π^N and let us compute the best response of player 0 to π^N.

If player 0 plays A, its cost is $\frac{1}{N}\sum_{i=0}^{\infty} e^{-\beta i/N} = 1/\beta + O(1/N)$.

If player 0 chooses action B, then its cost is $\int_0^\infty 2M_B e^{-\beta t} dt + O(1/N) = \int_0^\infty \frac{2e^{-\beta t}}{1+(N-1)e^{-Nt}} dt + O(1/N) \geq 1/\beta + O(1/N)$.

Therefore, when $\beta < 1$, player 0 has no incentive to deviate from π^N: it is a sub-game perfect equilibrium.

Similar example over a finite horizon.

Synchronous Case (N players)

- Each player n has an state $X_k(t) \in \mathcal{E}$ ($\mathbf{X}(t) = (X_0(t), \ldots, X_{N-1}(t))$ and chooses actions in $\mathcal{A}$.
- Main difference with the asynchronous model: $\forall t \in \mathbb{Z}^+$, all players choose an action $A_k(t) \in \mathcal{A}$ simultaneously.
- A player in state i who chooses action a goes to state j with probability $P_{ija}(\mathbf{X}(t))$.
- Given $\mathbf{X}(t)$, the evolution of all players are independent.
- The players are exchangeable : for any permutation σ,
$P_{ija}(X_0(t), \ldots, X_{N-1}(t)) = P_{ija}(X_{\sigma(0)}(t), \ldots, X_{\sigma(N-1)}(t))$.
Exchangeability implies that the dependence in $\mathbf{X}(t)$ can be replaced by a dependence on the population distribution $\mathbf{M}(t)$:

$$\mathbb{P}\left(\mathbf{X}(t+1) = \mathbf{j} | \mathbf{X}(t) = \mathbf{i}, \mathbf{A}(t) = \mathbf{a}\right) = \prod_{n=1}^{N} P_{i_n j_n a_n}(\mathbf{M}(t)),$$

Corresponding Mean Field Game

$$m_j^{\pi}(t+1) = \sum_{i \in \mathcal{E}} \sum_{a \in \mathcal{A}} m_i^{\pi}(t) P_{i,j,a}(\mathbf{m}^{\pi}(t)) \pi_{i,a}(\mathbf{m}(t)). \tag{6}$$

We denote by π^0 the strategy of player 0. The probability that Player 0 is in state $j \in \mathcal{E}$ evolves over time according to the following equation:

$$x_j(t+1) = \sum_{i \in \mathcal{E}} \sum_{a \in \mathcal{A}} x_i(t) P_{i,j,a}(\mathbf{m}^{\pi}(t)) \pi_{i,a}^0(\mathbf{m}(t)). \tag{7}$$

In this case, the cost of Player 0, becomes

$$V(\pi^0, \pi) = (1-\delta) \sum_{t=0}^{\infty} \sum_{i \in \mathcal{E}} \sum_{a \in \mathcal{A}} \delta^t x_i(t) c_{i,a}(\mathbf{m}^{\pi}(t)) \pi_{i,a}^0(\mathbf{m}(t)).$$

Player 0 chooses the strategy that minimizes her expected cost. When Player 0 does so, we say it uses the best-response to the mass strategy π.

$$BR(\pi) = \arg\min_{\pi^0} V(\pi^0, \pi).$$

An Important Special Case: Repeated Games

The state of a player is her current action ($\mathbf{X}(t) = \mathbf{A}(t)$) and the evolution of the state becomes: Under state $x = a$ and selecting action b, the next state becomes b with probability one.
The folk theorem holds for anonymous players with deterministic outcomes: any acheivable cost $V \leq V^*$ (cost a the static NE) is the cost of a Nash equilibrium, if the discount factor β is large enough. Sabourian results do not apply here:

$$F(a) = \delta_{M(a)} \text{ is not continuous.}$$

The Folk Theorem does not hold at the mean field limit: An example similar to the one in continuous time shows that not all equilibria pass at the mean field limit. Actually, for any $V < V^*$, the equilibria whose cost is V are based on the "tit for tat" principle. We claim that none of these equilibria survive at the mean field limit.

Conclusion III: Free Riders in Mean Field Games

When the number of players is infinite, the deviation of a single player is not visible by the population. The equilibria based on punishments of one player do not exist at the mean-field limit. In other words, mean field games cannot fight against (informed) free riders.

Conclusion III: Free Riders in Mean Field Games

When the number of players is infinite, the deviation of a single player is not visible by the population. The equilibria based on punishments of one player do not exist at the mean-field limit. In other words, mean field games cannot fight against (informed) free riders.

This is all the more damaging because these equilibria have very good social costs: mean-field games may fail to capture the best equilibria.

Conclusion III: Free Riders in Mean Field Games

When the number of players is infinite, the deviation of a single player is not visible by the population. The equilibria based on punishments of one player do not exist at the mean-field limit. In other words, mean field games cannot fight against (informed) free riders.

This is all the more damaging because these equilibria have very good social costs: mean-field games may fail to capture the best equilibria.

On the brighter side, Markov strategies may not be realistic (population distribution may not be observable). A result similar to Sabourian should hold.