Are mean-field games the limits of finite stochastic games?

Josu Doncel, Nicolas Gast, Bruno Gaujal

Are mean-field games the limits of finite stochastic games? *

Josu Doncel
Inria
Univ. Grenoble Alpes
josu.doncel@inria.fr

Nicolas Gast
Inria
Univ. Grenoble Alpes
nicolas.gast@inria.fr

Bruno Gaujal
Inria
Univ. Grenoble Alpes
bruno.gaujal@inria.fr

1. INTRODUCTION

Mean-field games model the rational behavior of an infinite number of indistinguishable players in interaction [5]. An important assumption of mean-field games is that, as the number of player is infinite, the decisions of an individual player do not affect the dynamics of the mass. Each player plays against the mass. A mean-field equilibrium corresponds to the case when the optimal decisions of a player coincide with the decisions of the mass. This leads to a simpler computation of the equilibrium.

It has been shown in [3, 4, 7] that for some games with a finite number of players, the Nash equilibria converge to mean-field equilibria as the number of players tends to infinity. In fact, many authors argue that mean-field games are a good approximation of symmetric stochastic games with a large number of players, the rationale behind this being that the impact of one player becomes negligible when the number of players goes to infinity.

In this paper, we question this assertion. We show that, in general, this convergence does not hold. In fact, the "tit for tat" principle allows one to define many equilibria in repeated or stochastic games with N players. However, in mean-field games, the deviation of a single player is not visible by the population and therefore the "tit for tat" principle cannot be applied. The conclusion is that, even if N-player games have many equilibria with a good social cost, this may not be the case for the limit game.

2. STOCHASTIC GAMES WITH IDENTICAL PLAYERS

We consider a class of stochastic games, as introduced by Shapley [6] for zero sum games and generalized by Fink [2] to N players. Players are identical and anonymous, in the sense that the dynamics, the costs and the action policies only depend on the *population distribution*, as detailed below.

The finite state space of each player is $\mathcal{S} = \{1, \dots, S\}$ and its finite action set is $\mathcal{A} = \{1, \dots, A\}$. The players' states at time t are denoted by $\mathbf{X}(t) = (X_1(t), \dots, X_n(t), \dots, X_N(t))$, with $X_n(t) \in \mathcal{S}$. The state of the players evolve in continuous time: each player n takes actions (denoted $A_n(t) \in \mathcal{A}$) at instants distributed according to a Poisson process, independently of the others. The superposition of all these N activation processes forms a Poisson process whose intensity is proportional to N, the number of players. If we only observe the system at these activation instants, and using uniformization, the system can be seen as a discrete time model where players take actions at discrete times $\mathcal{T}_N = \{i/N\}_{i\in\mathbb{N}}$.

The population distribution at time t is denoted by $\mathbf{M}(t) \in \mathcal{P}(\mathcal{S})$, where $\mathcal{P}(\mathcal{S})$ is the set of probability measures on $\mathcal{S}$. As the set $\mathcal{S}$ is finite, $\mathbf{M}(t)$ is a vector with $|\mathcal{S}|$ components and for all $s \in \mathcal{S}$, $M_s(t)$ being the fraction of players in state s at time t:

$$M_s(t) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{X_n(t)=s\}}.$$

We assume that players interact according to a mean field model, namely, the Markovian evolution of $\mathbf{X}(t)$ at time $t \in \mathcal{T}_N$ can be written as

$$\mathbb{P}\left(X_n(t+\frac{1}{N}) = j \middle| X_n(t) = i, \mathbf{M}(t) = \mathbf{m}, A_n(t) = a\right) = \frac{1}{N}P_{ij}(a, \mathbf{m}). \quad (1)$$

We consider that each player chooses its own *stationary mixed strategies*. Such a strategy is a measurable function $\pi : \mathcal{S}\times\mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{A})$, that associates to each state $i \in \mathcal{S}$ and each population distribution, a probability measure $\pi_i(\mathbf{m})$ on the set of possible actions – $\mathcal{P}(\mathcal{S})$ and $\mathcal{P}(\mathcal{A})$ are the sets of probability measures over $\mathcal{S}$ and $\mathcal{A}$ (as $\mathcal{A}$ is finite, $\mathcal{P}(\mathcal{A})$ is the simplex). We denote by $\pi_{i,a}(\mathbf{m})$ the probability that, under $\mathbf{m}$, a player in state i takes action a,

At time $t \in \mathcal{T}_N$, the player n suffers an instantaneous cost $c_{X_n(t),A_n(t)}(\mathbf{M}(t))$, function of her state $X_n(t)$, the action that she takes $A_n(t)$ and the population distribution $\mathbf{M}(t)$. The objective of player n is to choose a strategy π^n from some set of admissible strategies Π, in order to minimize her expected discounted payoff, knowing the strategies of the others. The discount factor is denoted by β. Given a strategy $\pi^n \in \Pi$ used by player n and a strategy $\pi \in \Pi$ used by all the others, we denote by $V(\pi^n, \pi)$ the expected discounted payoff of player n:

$$V^N(\pi^n, \pi) = \mathbb{E}\left[\sum_{t\in\mathcal{T}_N} e^{-\beta t} c_{X_n(t),A_n(t)}(\mathbf{M}(t)) \middle| \begin{array}{l} A_n \text{ has d.b. } \pi^n \\ A_{n'} \text{ has d.b. } \pi \ (n' \neq n) \end{array}\right].$$

An equilibrium for this game is a strategy π such that a player has no another admissible strategy that leads to a higher payoff. This notion depends naturally on the set of

*This work is partially supported by the EU project QUANTICOL, 600708.

admissible strategies.

Definition 1 (Nash Equilibrium). *For a given set of strategies Π, a strategy $\pi \in \Pi$ is called a symmetric Nash equilibrium in Π for the N player game if, for any strategy $\pi^n \in \Pi$,*

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

The existence of a stationary Nash equilibria is proven in [2] when $c_{ia}(\mathbf{m})$ and $P_{ij}(a, \mathbf{m})$ are continuous in $\mathbf{m}$.

To make the connection with the mean field game model presented below, let us remark that when all players use a stationary strategy π, the evolution of the population distribution satisfies

$$\frac{1}{N}\mathbb{E}\left(M_j(t+\frac{1}{N}) - M_j(t)\middle|\mathbf{M}(t)\right) = \sum_{i\in\mathcal{S}}\sum_{a\in\mathcal{A}} M_i(t)Q_{ij}(a, \mathbf{M}(t))\pi_{i,a}(\mathbf{M}(t)), \quad (2)$$

where the *rate* matrix $Q(a, \mathbf{M}(t)) := P(a, \mathbf{M}(t)) - Id$.

3. MEAN-FIELD GAME MODEL

We consider an infinite homogeneous population of players. We denote by $\mathbf{m}^\pi(t) \in \mathcal{P}(\mathcal{S})$ the population distribution at time t. As the state space is finite, $\mathbf{m}^\pi(t)$ is a vector whose ith component, $m_i^\pi(t)$, is the proportion of players in state i at time t. We assume that the initial condition $\mathbf{m}^\pi(0) = \mathbf{m}_0$ is fixed. For $t \geq 0$, the population distribution evolves over time according to the following differential equation: for $j \in \mathcal{S}$

$$\dot{m}_j^\pi(t) = \sum_{i\in\mathcal{S}}\sum_{a\in\mathcal{A}} m_i^\pi(t)Q_{ij}(a, \mathbf{m}^\pi(t))\pi_{i,a}(\mathbf{m}^\pi(t)). \quad (3)$$

The rationale behind this differential equation is Equation (2): the players in state i that take action $a \in \mathcal{A}$ move to state j at rate $Q_{ij}(a, \mathbf{m}^\pi(t))$.

We now concentrate on a particular player, that we call Player 0. Player 0 chooses her own strategy π^0. We denote by $\mathbf{x}(t)$ the probability distribution of Player 0 when she applies strategy π^0 and the population uses strategy π. For a given state $i \in \mathcal{S}$, $x_i(t)$ denotes the probability for Player 0 to be in state i at time t. The distribution $\mathbf{x}$ evolves over time according to the following equation: for $j \in \mathcal{S}$

$$\dot{x}_j(t) = \sum_{i\in\mathcal{S}}\sum_{a\in\mathcal{A}} x_i(t)Q_{ij}(a, \mathbf{m}^\pi(t))\pi_{i,a}^n(t). \quad (4)$$

If Player 0 is in state i and takes an action a, it suffers from the instantaneous cost $c_{i,a}(\mathbf{m}^\pi(t))$, introduced in Section 2. Given a population strategy π and the strategy of Player 0 π^n, we define the discounted cost of Player 0 as

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

where β is the discount factor. Player 0 chooses the strategy that minimizes her expected cost, which depends as well on the strategy π. When Player 0 does so, we say it does the best-response to the mass strategy π.

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

Definition 2 (Mean-Field Equilibrium). *A strategy is a (symmetric) mean-field equilibrium if it is a fixed point for the best-response function, that is,*

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

Applying the Kakutani fixed point theorem for infinite dimension spaces over a modified version of the correspondence BR, one can show the existence of a mean-field equilibrium for these mean-field games under very mild continuity assumptions. The proof is available in [1].

Theorem 1 (Existence of equilibrium, [1]). *Assume that $Q_{ij}(a, \mathbf{m})$ and $c_{ia}(\mathbf{m})$ are continuous in $\mathbf{m}$. Then, there always exists a mean-field equilibrium.*

4. (NON-)CONVERGENCE

4.1 Mean-field games as a natural limit

Using classical arguments from mean field theory, one can show that when the strategy of the players is given, the population distribution for the N player game described in Equation (2) converges to the population distribution of the mean field limit (3). More precisely, in an N player-game, the influence of a single player on the mass $\mathbf{m}$ is of order $1/N$. In fact, it is shown in [7] that if the cost functions $c_{i,a}(\mathbf{m})$, the transition kernels $Q_{ij}(a, \mathbf{m})$ and the policy $\pi_i(\mathbf{m})$ are continuous in $\mathbf{m}$ then the dynamics of the population converges to the solution of the differential equation (3) and the evolution of one player converges to solution of (4).

Furthermore, it is shown in [1] that under such continuity conditions
(i) if π be a mean-field equilibrium, then there exists N_0 such that for all $N \geq N_0$, π is a ε-equilibrium of the N player game; and
(ii) if $(\pi^N)_N$ is a sequence of Lipschitz-continuous strategies such that π^N is a equilibrium for the N player game, then, any sub-sequence of the sequence $(\pi^N)_N$ has a sub-sequence that converges weakly to a mean-field equilibrium. This is a more or less direct consequence of Theorem 3.3.2 in [7].

4.2 Counter-example in the general case

The positive result presented in the previous section uses an important unnatural assumption: the strategy of a player should be continuous in $\mathbf{m}$. In the definition of a Nash equilibrium, there is no *a priori* reason to restrict to strategies that are continuous in $\mathbf{m}$, even when the cost and transition kernel are continuous in $\mathbf{m}$. In this section, we show that removing this assumption leads to a very different result.

To construct such an example, let is consider a symmetrical version of the prisoner's dilemma. The state space of a player is $\mathcal{S} = \{C, D\}$ (that stand for Cooperate and Defect, respectively) and the action set is $\mathcal{A} = \mathcal{S}$. Let m_C (resp. m_D) be the proportion of players in state C (resp. D). The population distribution is $\mathbf{m} = (m_C, m_D)$. The instantaneous cost of a player depends on her state i and on the mass $\mathbf{m}$:

$$c_{i,a}(\mathbf{m}) = \begin{cases} m_C + 3m_D & \text{if } i = C \\ 2m_D & \text{if } i = D \end{cases}$$

This cost corresponds to the expected cost of a player that is matched with another player at random and suffers a cost according to the following cost matrix:

	C	D
C	$1,1$	$3,0$
D	$0,3$	$2,2$

(5)

The state of a player coincides with her current action. When a player chooses action a, the next state becomes a with probability one. This gives the following transition matrices:

$$P(C,\mathbf{m}) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad P(D,\mathbf{m}) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

A player that has no impact on the actions of the other players should play D: indeed, as indicated by the cost matrix (5), its cost is smaller when she plays D. This leads to the following result:

Lemma 1. *There exists a unique mean-field equilibrium π^∞ that consists in always playing D.*

Proof. We consider that Player 0 has state vector x and the mean-field is m. Taking into account that $x_C(t) + x_D(t) = m_C(t) + m_D(t) = 1$, her expected cost is

$$\int_0^\infty [x_C(t)m_C(t) + 3x_C(t)m_D(t) + 2x_D(t)m_D(t)]e^{-\beta t}dt$$
$$= \int_0^\infty [x_C(t) + 2m_D(t)]e^{-\beta t}dt.$$

This cost is minimal when $x_C(t)$ is minimal. This shows that the best response of player 0 is to always play D. $\square$

A similar proof shows that π^∞ is also a Nash equilibrium for the corresponding stochastic game with N players. However it is not the only one. Let us define the following stationary strategy:

$$\pi^N(m) = \begin{cases} D & \text{if } m_C < 1 \\ C & \text{if } m_C = 1. \end{cases}$$

This strategy can be rephrased as "play C as long as everyone else is playing C. Play D as soon as another player deviates and play D". This strategy is called a reward and punishment strategy: as long as the others are cooperating, I reward them by also cooperating myself. If one person defects, playing D will punish everyone, including the defector, who therefore has no incentive to defect.

This strategy is an equilibrium for the N player game:

Lemma 2. *For $\beta < 1$ and N large, π^N is a Nash equilibrium of the N-player stochastic game.*

Proof. Assume that all players, except player 0, play the strategy π and let us compute the best response of player 0.

If at time 0, $m_C < 1$, then the best response of player 0 is to play D. On the other hand, if $m_C = 1$ and if player 0 becomes active, then if she uses π, she will suffer a cost $\frac{1}{N}\sum_{i=0}^\infty e^{-\beta i/N} = \int \exp(-\beta t)dt + O(1/N) = 1/\beta + \mathrm{O(1/N)}$. If player 0 deviates from π at this time step and chooses action D, all players will also deviate after the next time step. This implies that $m_D(t) = 1 - \exp(-t)$ and that the player 0 will suffer a cost equal to $\int_0^\infty (x_C(t) + 2 - 2e^{-t})e^{-\beta t}dt + O(1/N) \geq 2/(\beta(\beta+1)) + O(1/N)$ when N is large. This shows that when $\beta < 1$, player 0 has no incentive to deviate from the strategy π so that, π is a Nash equilibrium. $\square$

The Nash-equilibrium π^N leads to a totally different situation from the mean-field equilibrium. In the first case, when all players start in C, they always play C whereas in the second case they will all play D. This shows that for this game, there exists a sequence of Nash-equilibria that do not converge to a mean-field equilibrium (although the cost and the transition kernel are both Lipschitz-continuous). π^N is a social optimum while π^∞ has the worst possible social cost.

In mean-field games, a single player have a negligible impact in the population distribution so that her individual actions are not visible by the population. This is the reason why the population can not punish a single player that deviates from the desired pattern. In other words, in the limit game the "tit for tat" principle cannot be applied.

5. CONCLUSION

In this paper, we consider the convergence of finite stochastic games to mean-field games when the number of players tends to infinity. When we restrict to strategies that are continuous in $\mathbf{m}$, every Nash equilibria converge to a mean-field equilibrium. However, when removing this restriction, convergence to mean-field games does not hold in general.

When the number of players is finite, it is possible to define many equilibria by using the "tit for tat" principle. Indeed, players can cooperate to punish a player that does not follow a desired pattern. These equilibria are linked to the Folk theorem for repeated games, that states that, when a one-shot game is repeated infinitely often, for each achievable cost that is not worse than the Nash equilibrium of the one-shot game, there exists a "tit for tat" equilibrium that achieves this payoff. When the number of players is infinite, the deviation of a single player is not visible by the population and, therefore, the equilibria that are based on the "tit for tat" principle do not scale at the mean-field limit. This is all the more damaging because these equilibria have very good social costs: mean-field games fail to describe the best equilibria.

References

[1] J. Doncel, N. Gast, and B. Gaujal. Mean-Field Games with Explicit Interactions. 2016. URL: `https://hal.inria.fr/hal-01277098`.

[2] A. M. Fink. Equilibrium in a stochastic n-person game. *Journal of science of the hiroshima university, series ai (mathematics)*, 28(1):89–93, 1964.

[3] D. A. Gomes, J. Mohr, and R. R. Souza. Discrete time, finite state space mean field games. *Journal de mathématiques pures et appliquées*, 93(3):308–328, 2010.

[4] O. Guéant, J.-M. Lasry, and P.-L. Lions. Mean field games and applications. In, *Paris-princeton lectures on mathematical finance 2010*, pp. 205–266. Springer, 2011.

[5] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[6] L. S. Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

[7] H. Tembine, J.-Y. L. Boudec, R. El-Azouzi, and E. Altman. Mean field asymptotics of markov decision evolutionary games and teams. In *Gamenets*, 2009, pp. 140–150.