Mean Field Asymptotics of Markov Decision Evolutionary Games and Teams[*]

Hamidou Tembine, Jean-Yves Le Boudec, Rachid El-Azouzi, Eitan Altman [†]

Abstract

We introduce Markov Decision Evolutionary Games with N players, in which each individual in a large population interacts with other randomly selected players. The states and actions of each player in an interaction together determine the instantaneous payoff for all involved players. They also determine the transition probabilities to move to the next state. Each individual wishes to maximize the total expected discounted payoff over an infinite horizon. We provide a rigorous derivation of the asymptotic behavior of this system as the size of the population grows to infinity. We show that under any Markov strategy, the random process consisting of one specific player and the remaining population converges weakly to a jump process driven by the solution of a system of differential equations. We characterize the solutions to the team and to the game problems at the limit of infinite population and use these to construct almost optimal strategies for the case of a finite, but large, number of players. We show that the large population asymptotic of the microscopic model is equivalent to a (macroscopic) Markov decision evolutionary game in which a local interaction is described by a single player against a population profile. We illustrate our model to derive the equations for a dynamic evolutionary Hawk and Dove game with energy level.

1. Introduction

We consider a large population of players in which frequent interactions occur between small numbers of chosen individuals. Each interaction in which a player is involved can be described as one stage of a dynamic game. The state and actions of the players at each stage determine an immediate payoff (also called *fitness* in behavioral ecology) for each player as well as the transition probabilities of a controlled Markov chain associated with each player. Each player wishes to maximize its expected fitness averaged over time.

This model extends the basic evolutionary games by introducing a controlled state that characterizes each player. The stochastic dynamic games at each interaction replace the matrix games, and the objective of maximizing the expected long-term payoff over an infinite time horizon replaces the objective of maximizing the outcome of a matrix game. Instead of a choice of a (possibly mixed) action, a player is now faced with the choice of decision rules (called strategies) that determine what actions should be chosen at a given interaction for given present and past observations.

This model with a finite number of players, called a mean field interaction model, is in general difficult to analyze because of the huge state space required to describe the sate of all players. Then taking the asymptotics as the number of players grows to infinity, the whole behavior of the population is replaced by a deterministic limit that represents the system's state, which is fraction of the population at each individual state that use a given action.

In this paper we study the asymptotic *dynamic* behavior of the system in which the population profile evolves in time. For large N, under mild assumptions (see Section 3), the mean field converges to a deterministic measure that satisfies a non-linear ordinary differential equation for under any stationary strategy. We show that the mean field interaction is asymptotically equivalent to a Markov decision evolutionary game. When the rest of the population uses a fixed strategy u, any given player sees an equivalent game against a collective of players whose state evolves according to the ordinary differential equation (ODE) which we explicitly compute. In addition to providing the exact limiting asymptotic, the ODE approach provides tight approximations for fixed large N. The mean field asymptotic calculations for large N for given choices of strategies allows us to compute the equilibrium of the game in the asymptotic regime.

Related Work. Mean field interaction models have already been used in standard evolutionary games in a completely different context: that of evolutionary game dynamics (such as replicator dynamics) see e.g. [8] and references therein. The paradigm there has been to associate relative growth rate to actions according to the fitness they achieved, then study the asymptotic trajectories of the state

of the system, i.e. the fraction of users that adopt the different actions. *Non-atomic* Markov Decision Evolutionary Games have been applied in [9] to firm idiosyncratic random shocks and in [1] to cellular communications.

Structure. The remainder of this paper is organized as follows. In next section we present the model assumptions and notations. In Section 3 we present some convergence results of the ODE in the random number of interacting players. In Section 4 a resource competition between animals with two types of behaviors and several states is presented. All the sketch of proofs are given in Appendix. Section 5 concludes the paper.

2. Model description

2.1. Markov Decision Evolutionary Process With N Players

We consider the following model, which we call Markov Decision Evolutionary Game with N players.
- There are $N \in \mathbb{N}$ players.
- Each player has its own state. A state has two components: the *type* of the player and the *internal state*. The type is a constant during the game. The state of player j at time t is denoted by $X_j^N(t) = (\theta_j, S_j^N(t))$ where θ_j is the type. The set of possible states $\mathscr{X} = \{1, \ldots, \Theta\} \times \mathscr{S}$ is finite.
- Time is discrete, taking values in $\frac{\mathbb{N}}{N} := \{0, \frac{1}{N}, \frac{2}{N}, \ldots\}$.
- The *global detailed description* of the system at time t is $X^N(t) = (X_1^N(t), \ldots, X_N^N(t))$.

Define $M^N(t)$ to be the current population profile i.e $M_x^N(t) = \frac{1}{N}\sum_{j=1}^{N} 1_{\{X_j^N(t)=x\}}$. At each time t, $M^N(t)$ is in the finite set $\{0, \frac{1}{N}, \frac{2}{N}, \ldots, 1\}^{\sharp\mathscr{X}}$. and $M_{\theta,s}^N(t)$ is the fraction of players who belong to population of type θ (also called subpopulation θ) and have internal state s. Also let $\bar{M}_\theta^N = N\sum_{s\in\mathscr{S}} M_{\theta,s}^N(t)$ be the size of subpopulation θ (independent of t by hypothesis). We do not make any specific hypothesis on the ratios $\frac{\bar{M}_\theta^N}{N}$ as N gets large (it may be constant or not, it may tend to 0 or not).

- *Strategies and local interaction:* At time slot t, an ordered list $\mathscr{B}^N(t)$, of players in $\{1,2,\ldots,N\}$, without repetition, is selected randomly as follows. First we draw a random number of players $K(t)$ such that

$$\mathbb{P}(K(t) = k | M^N(t) = \vec{m}) = J_k^N(\vec{m})$$

where the distribution $J_k^N(\vec{m})$ is given for any N, $\vec{m} \in \{0, \frac{1}{N}, \frac{2}{N}, \ldots, 1\}^{\sharp\mathscr{X}}$. Second, we set $\mathscr{B}^N$ to an ordered list of $K(t)$ players drawn uniformly at random among the $N(N-1)\ldots(N-K(t)+1)$ possible ones. By abuse of notation we write $j \in \mathscr{B}^N(t)$ with the meaning that j appears in the list $\mathscr{B}^N(t)$.

Each player such that $j \in \mathscr{B}^N(t)$ takes part in a one-shot event at time t, as follows. First, the player chooses an

action a in the finite set $\mathscr{A}$ with probability $u_\theta(a|s)$ where (θ, s) is the current player state. The stochastic array u is the strategy profile of the population, and u_θ is the strategy of subpopulation θ. A vector of probability distributions u which depend only on the type of the player and its internal state is called *stationary strategy*.

Second, say that $\mathscr{B}^N(t) = (j_1, \ldots, j_k)$. Given the actions $a_{j_1}, \ldots, a_{j_k}$ drawn by the k players, we draw a new set of internal states $(s'_{j_1}, \ldots, s'_{j_k})$ with probability $L_{\underline{\theta};\underline{s};\underline{a};\underline{s}'}^N(k, \vec{m})$,

$$\text{where } \underline{\theta} = (\theta_{j_1}, \ldots, \theta_{j_k}), \ \underline{s} = (s_{j_1}, \ldots, s_{j_k})$$
$$\underline{a} = (a_{j_1}, \ldots, a_{j_k}), \ \underline{s}' = (s'_{j_1}, \ldots, s'_{j_k})$$

Then the collection of k players makes one synchronized transition, such that

$$S_{j_i}^N(t + \frac{1}{N}) = s'_{j_i} \quad i = 1, \ldots, k$$

Note that $S_j^N(t + \frac{1}{N}) = S_j^N(t)$ if j is not in $\mathscr{B}^N(t)$.

It can easily be shown that this form of interaction has following properties: (1) X^N is Markov and (2) players can be observed only through their state.

The model is entirely specified by the probability distributions J^N, the Markov transition kernels L^N and the strategy profile u. In this paper, we assume that J^N and L^N are fixed for all N, but u can be changed and does not depend on N (though it would be trivial to extend our results to strategies that depend on N, but this appears to be unnecessary complication). We are interested in large N.

It follows from our assumptions that

1. $M^N(t)$ is Markov.

2. for any fixed $j \in \{1, \ldots, N\}$, $(X_j^N(t), M^N(t))$ is Markov. This means that the evolution of one specific player $X_j^N(t)$ depends on the other players only through the occupancy measure $M^N(t)$.

2.2. Payoffs

We consider two types of instantaneous payoff and one discounted payoff:
- *Instant Gain:* This is the random gain $G_j^N(t)$ obtained by one player whenever it is involved in an event at time t. We assume that it depends on this player's state just before the event and just after the event, the chosen action, and on the states and actions of all players involved in this event. Formally, if player $j \in \mathscr{B}^N(t)$

$$G_j^N(t) = g^N(x_j, a_j, x'_j, x_{\mathscr{B}^N(t)\setminus j}, a_{\mathscr{B}^N(t)\setminus j}, x'_{\mathscr{B}^N(t)\setminus j})$$

where $x_j = X_j^N(t)$, a_j is the action chosen by player j, $x'_j = X_j^N(t + \frac{1}{N})$, $x_{\mathscr{B}^N(t)\setminus j}$ [resp. $x'_{\mathscr{B}^N(t)\setminus j}$] is the list of states at

time t [resp. at time $t + \frac{1}{N}$] of players other than j involved in the event, $a_{\mathscr{B}^N(t)\setminus j}$ is the list of their actions and $g()$ is some non random function defined on the set of appropriate lists. Whenever j is not in $\mathscr{B}^N(t)$, $G_j^N(t) = 0$. We assume that $G_j^N(t)$ is bounded, i.e. there is a non random number C_0 such that, with probability 1: for all j,t: $|G_j^N(t)| \leq C_0$

• *Expected Instant Payoff:* It is defined as the expected instant gain of player j, given the state x of j and the population profile $\vec{m}$. By our indistinguishability assumption, it does not depend on the identity of a player, so we can write it as

$$r^N(u,x,\vec{m}) := \mathbb{E}\left(G_j^N(t) \,\big|\, X_j^N(t) = x, M^N(t) = \vec{m}\right)$$

Note that this conditional expectation contains the case when j is not in $\mathscr{B}^N(t)$, i.e. when $G_j^N(t) = 0$.

• *Discounted Long-Term Payoff:* It is defined as the expected discounted long term payoff of one player, given the initial state of this player and the population: $\bar{r}^N(u; x, \vec{m}) :=$

$$\mathbb{E}\left(\sum_{\substack{t=0 \text{ step } 1/N}}^{\infty} e^{-\beta t} G_j^N(t) | X_j(0) = x, M^N(0) = \vec{m}\right)$$

where β is a positive parameter (existence follows from the boundedness of G_j^N). The fact that it does not depend on the identity j of the player, but only on its initial state x and the initial population profile $\vec{m}$, follows from the indistinguishability assumption.

We defined the Discounted Long-Term Payoff in terms of the instant gain, as this is the most natural definition. The following proposition shows that the alternative definition, by means of the expected instant payoff, is equivalent.

Proposition 2.2.1. *For all player state x and population profile $\vec{m}$*

$$\bar{r}^N(u;x,\vec{m}) = \mathbb{E}\left(\sum_{\substack{t=0 \text{ step } 1/N}}^{\infty} e^{-\beta t} r^N(u, X_j^N(t), \vec{M}^N(t))\right.$$
$$\left. | X_j(0) = x, M^N(0) = \vec{m}\right)$$

2.3. Focus on One Single Player

We are interested in the following special case (here we make the dependency on the strategy explicit). There are two types of players, i.e. $\Theta = 2$. There is exactly one player (the player of interest) with type 1. All other players have type 2. In this case we use the notation $R^N(u_1, u_2; s, \vec{m})$ for the discounted long-term payoff obtained by the player in type 0, when her strategy is u_1 and all other players's strategy is u_2, given that this player's initial internal state is s and the initial type 2 subpopulation profile is $\vec{m}$. Note that

$$R^N(u_1, u_2; s, \vec{m}) = \bar{r}^N(u_1, u_2; (1,s), \vec{m}')$$

with $m'_{1,s'} = \frac{1}{N} 1_{s=s'}$ and $m'_{2,s'} = m_{2,s'}$ for all $s' \in \mathscr{S}$.

Markov Decision Evolutionary Game

Player j may choose a strategy u^j. We look for a (Nash) equilibrium u such that if all players use u then no player has an incentive to deviate from u. For any finite N one can map this into a standard Markov game. This is true for both the case where the number of players is known and in the case it is unknown when taking a decision. Therefore we know that a stationary equilibrium exists in the discounted case. A stationary equilibrium is solution of the fixed point equation:

$$\forall j, \; u_{j,\theta} \in \arg\max_{v_{j,\theta}} R^N(v_{j,\theta}, u_{-j}; s, m)$$

By assuming symmetry per type we can show that a stationary equilibrium exists which is a solution of the fixed point equation

$$\forall \theta, u_\theta \in \arg\max_{v_\theta} R^N(v_\theta, u; s, m)$$

Markov Decision Evolutionary Team

We wish to find a stationary u that maximizes R^N averaged over all players.

$$u = (u_1, \ldots, u_\Theta) \in \arg\max_v R^N(v; s, m)$$

3. Main Results

3.1. Scaling Assumptions

We are interested in the large N regime and obtain that, for any fixed j, (X_j^N, M^N) converges weakly to a simple process. This requires the weak convergence of $M^N(0)$ to some $\vec{m}_0$.

We assume that the parameters of the model and the payoff per time unit converge as $N \to \infty$, i.e.

$$\begin{cases} J_k^N(\vec{m}) \to J_k(\vec{m}) \\ L^N_{\theta;s;a;s'}(k,\vec{m}) \to \quad L_{\theta;s;a;s'}(k,\vec{m}) \\ r^N(u,x,\vec{m}) \to r(u,x,\vec{m}) \end{cases} \quad (1)$$

Our main scaling assumption is

H1 $\sum_k k^2 J_k(\vec{m}) < \infty$ for all $\vec{m} \in \Delta$. This ensures that the second moment of the number of players involved in an event per time slot is bounded.

Note that H1 excludes the case where the number of players involved in an event per time slot scales like N (i.e. synchronous transitions of all players at the same time). There may be large N asymptotic results for such cases [10] but the limit is not given by an ODE. In contrast, H1 is automatically true if the number of players involved in an event

per time slot is upper bounded by a non random constant. We also need some technical assumptions, which are usually true and can be verified by inspection.

H2 $\sum_k J_k(\vec{m}) > 0$ for all $\vec{m} \in \Delta$ (Δ is the simplex $\{\vec{m} : m_{\theta,s} \geq 0, \sum_{\theta,s} m_{\theta,s} = 1\}$). This ensures that the mean number of players involved in an event per time slot, $\sum_{k \geq 0} k J_k(\vec{m})$ is non zero.

Define the *drift* of $M^N(t)$ as

$$\vec{f}^N(u,\vec{m}) = \mathbb{E}\left(M^N(t + \frac{1}{N}) - M^N(t)|M^N(t) = \vec{m}\right)$$

Note that we make explicit the dependency on the strategy u but not on J and L, assumed to be fixed.

It follows from our hypotheses that

$$\lim_{N \to \infty} N f^N(u,\vec{m}) := f(u,\vec{m}) \qquad (2)$$

exists.

H3 We assume that the convergence in Equation (2) is uniform in $\vec{m}$ and the limit is Lipschitz-continuous in $\vec{m}$. This is in particular true if one can write, for every strategy u, $f^N(u,\vec{m}) = \frac{1}{N}\phi_u(\frac{1}{N},\vec{m})$, with ϕ_u defined on $[0,\varepsilon] \times \Delta$ where $\varepsilon > 0$ and Φ_u is continuously differentiable.

H4 $\mathbb{P}(X_j^N(t + 1/N) = y|X_j^N(t) = x, M^N(t) = m, M^N(t + 1/N) = m')$ converges uniformly in $\vec{m},\vec{m}'$ and the limit is Lipschitz-continuous in $\vec{m},\vec{m}'$. This is in particular true if one can write, for every strategy u, as $\xi_{u,x;y}(1/N,m,m')$. with ξ defined on $[0,1] \times \Delta \times \Delta$ and $\xi_{u,x;y}$ is continuously differentiable.

Our model satisfies the assumptions in [2], therefore we have the following result:

Theorem 3.1.1 ([2])**.** *Assume that $\lim_{N \to \infty} M^N(0) = \vec{m}_0$ in probability. For any stationary strategy u, and any time t, the random process $M^N(t) = \frac{1}{N}\sum_{j=1}^N \delta_{X_j^N(t)}$ converges in distribution to the (non-random) solution of the ODE*

$$\dot{\vec{m}}(t) = f(u,\vec{m}(t)) \qquad (3)$$

with initial condition $\vec{m}_0$.

3.2. Convergence results

We focus on one player, without loss of generality we can call her player 1, and consider the process (X_1^N, M^N). For any finite N, X_1^N and M^N are not independent, however in the limit we have the following:

Theorem 3.2.1. *Assume that $\lim_{N \to \infty} M^N(0) = \vec{m}_0$ and $\lim_{N \to \infty} X_1^N(0) = x_0 = (\theta_1, s_0)$ in probability. The discrete time process $(X_1^N(t), M^N(t))$ defined for $t \in \frac{\mathbb{N}}{N}$, converges weakly to the continuous time jump and drift process $(X_1(t), \vec{m}(t))$, where $\vec{m}(t)$ is solution of the ODE Equation (3) with initial condition $\vec{m}_0$ and $X_1(t)$ is a continuous time, non homogeneous jump process, with initial state x_0. The rate of transition of $X_1(t)$ from state $x_1 = (\theta_1, s_1)$ to state $x_1' = (\theta_1, s_1')$ is*

$$A(x_1, x_1'; \vec{m}(t), u) = \sum_{k \geq 1} J_k(\vec{m}) A_k(s_1, s_1'; \vec{m}(t), u)$$

with $A_k(s_1, s_1'; \vec{m}(t), u) =$

$$\sum_{\underline{\theta}; \underline{s}; \underline{a}; \underline{s}'} L_{\theta_1, \underline{\theta}; s, \underline{s}; \underline{a}; s', \underline{s}'}(k, \vec{m}(t)) \prod_{j=1}^k u_{\theta_j}(a_j|s_j) \prod_{j=2}^k m_{\theta_j, s_j}(t)$$

where $\underline{\theta} = (\theta_2, ..., \theta_k), \underline{s} = (s_2, ..., s_k)$
$\underline{a} = (a_1, ..., a_k), \underline{s}' = (s_2', ..., s_k')$

Note that, contrary to results based on propagation of chaos, we do not assume that the distribution of player states at time 0 is exchangeable. In contrast, we will use Theorem 3.2.1 precisely in the case where player 1 is different from other players. Theorem 3.2.1 motivates the following definition.

Definition 3.3. *To a game as defined in Section 2.1 we associate a "Macroscopic Markov Decision Evolutionary Game", defined as follows. There is one player, (player 1), with state $X_1(t)$ and a population profile $\vec{m}(t)$. The initial condition of the game is $X_1(0) = x$, $\vec{m}(0) = \vec{m}_0$. The population profile is solution to the ODE (3) and $X_1(t)$ evolves as a jump process as in Theorem 3.2.1.*

Further, let $\bar{r}(u; x, \vec{m})$ be the discounted long-term payoff of player 1 in this game, given that $X_1(0) = x$ and $\vec{m}(0) = \vec{m}_0$, i.e. $\bar{r}(u; x, \vec{m}) =$

$$\mathbb{E}\left(\int_0^\infty e^{-\beta t} r(u, X_1(t), m(t))|X_1(0) = x, \vec{m}(0) = \vec{m}_0\right)$$

We also consider, as in Section 2.3, the case with $\Theta = 2$ types and define by analogy $R(u_1, u_2; s, \vec{m})$ as the discounted long-term payoff when player 1 starts in state s and the population profile starts in state $\vec{m}$, with player 1 using strategy u_1 and other players strategy u_2.

In order to exploit the convergence in distribution of the process focused on one player, we need that the payoff be continuous in the topology of this convergence. This is stated in the next theorem.

Theorem 3.3.1. *Let $E = \mathscr{S} \times \Delta$ and $D_E[0,\infty)$ the set of cadlag functions from $[0,\infty)$ to $\mathbb{R}$, equipped with Skorohod's topology. The mapping*

$$D_E[0,\infty) \rightarrow \mathbb{R}$$
$$(s,m) \mapsto \int_0^\infty e^{-\beta t} r(u,s(t),m(t))\, dt$$

is continuous.

Using Theorem 3.2.1 and Theorem 3.3.1 we obtain the following, which is the main result of this paper. It says that when N goes to infinity, the Markov Decision Evolutionary Game with $N(t)$ of players becomes equivalent to the associated Macroscopic Markov decision evolutionary game. This reduces any multi-player problem into an effective one-player problem.

Theorem 3.3.2 (Asymptotically equivalent game)**.** *When N goes to infinity we have (a) the discrete time process X_1^N converges in distribution to the continuous time process X_1 (b) $\vec{r}^N(u;x,\vec{m}) \rightarrow \vec{r}(u;x,\vec{m})$ and (c) $R^N(u_1,u_2;s,\vec{m}) \rightarrow R(u_1,u_2;s,\vec{m})$*

3.4. Case with Global Attractor

Assume that, for some strategy u, the ODE (3) has a global attractor $\vec{m}^*$ (this may or may not hold, depending on the ODE). If in addition the model with N players is irreducible, with stationary probability distribution ϖ^N for M^N, then $\lim_{N\longrightarrow\infty} \varpi^N = \delta_{\vec{m}^*}$ where $\delta_{\vec{m}^*}$ is the Dirac mass at $\vec{m}^*$ (follows from [2]). i.e. the large time distribution of $M^N(t)$ converges, as $N \rightarrow \infty$, to the attractor $\vec{m}^*$.

Also, $(X_j^N(t), M^N(t))$ converges to a continuous time, homogeneous Markov jump process with time-independent transition matrix:

$$A(x_1,x_1';u) = \sum_{k \geq 1} J_k(\vec{m}) A_k(s_1,s_1';\vec{m}^*,u)$$

Assume that the transition matrix $A(x_1,x_1';u)$ is also irreducible and let $\pi()$ be its unique stationary probability. Also let π^N be the first marginal of the stationary probability of (X_1^N, M^N). It is natural in this case to replace the definition of the long term payoffs $R^N(u_1,u_2;s,\vec{m})$ and $R^N(u_1,u_2;s,\vec{m})$ by their stationary counterparts

$$R_{st}^N(u_1,u_2) := \sum_s \pi^N(s) R^N(u_1,u_2;s,\vec{m}^*)$$
$$R_{st}(u_1,u_2) := \sum_s \pi(s) R(u_1,u_2;s,\vec{m}^*)$$

3.5. Single player per type selected per time slot

Consider the special case where at each time slot, only one player per type between the N is randomly selected and has a chance to change its action, i.e. $\sharp \mathscr{B}^N = 1$ w.p 1.

Thus H1 and H2 are automatically satisfied. The resulting ODE (see [3]) becomes

$$\frac{d}{dt} m_x(t) = \sum_{x'} m_{x'} L_{x',x}(\vec{m},u,\Theta) - m_x \sum_{x'} L_{x,x'}(\vec{m},u,\Theta)$$

The term $\sum_{x'} m_{x'} L_{x',x}(\vec{m},u,\Theta)$ is the *incoming flow* in to x and the *outgoing flow* from x is $m_x \sum_{x'} L_{x,x'}(\vec{m},u,\Theta)$.

We then obtain a large class of *state-dependent* evolutionary game dynamics. Note that in general the trajectories of the mean dynamics need not to converge. In the case of single player selected in each time slot of $1/N$ and linear transition in m, the time averages under the replicator dynamics converge its interior rest points or the boundaries of the simplex.

3.6. Equilibrium and optimality

Let $\mathscr{U}_s$ be the set of strategies. Consider the optimal control problems

$$(OPT_N) \begin{cases} \text{Maximize } R^N(u,u;s,\vec{m}_0) \\ \text{s.t } u \in \mathscr{U}_s \end{cases}$$

$$(OPT_\infty) \begin{cases} \text{Maximize } R(u,u;s,\vec{m}_0) \\ \text{s.t } u \in \mathscr{U}_s \end{cases}$$

The strategy u is an $\varepsilon-$optimal strategy for the N-optimal control problem if

$$R^N(u,u;s,\vec{m}_0) \geq -\varepsilon + \sup_v R^N(v,v;s,\vec{m}_0).$$

Also consider the fixed-point problems

$$(FIX_N) \begin{cases} \text{find } u \in \mathscr{U}_s \text{ such that} \\ u \in \arg\max_{v \in \mathscr{U}_s}\{R^N(v,u;s,\vec{m}_0)\} \end{cases}$$

$$(FIX_\infty) \begin{cases} \text{find } u \in \mathscr{U}_s \text{ such that} \\ u \in \arg\max_{v \in \mathscr{U}_s}\{R^N(v,u;s,\vec{m}_0)\} \end{cases}$$

A solution to (FIX_N) or (FIX_∞) is a (Nash) equilibrium. We say that u is an $\varepsilon-$equilibrium for the game with N [resp. $N \rightarrow \infty$] players if $R^N(u,u;s,\vec{m}_0) \geq \sup_v R^N(v,u;s,\vec{m}_0) - \varepsilon$ [resp. $R(u,u;s,\vec{m}_0) \geq \sup_v R(v,u;s,\vec{m}_0) - \varepsilon$].

Note that the definition of equilibrium and optimal strategy may depend on the initial conditions. If, for any $u \in \mathscr{U}_s$, the hypotheses in Section 3.4 hold, then we may relax this dependency.

Theorem 3.6.1 (Finite N)**.** *For every discount factor $\beta > 0$ the optimal control problem (OPT_N) (resp. the fixed-point problem (FIX_N)) has at least one $0-$optimal strategy (resp. $0-$equilibrium). In particular, there a ε_N-optimal strategy (resp. ε_N-equilibrium) with $\varepsilon_N \longrightarrow 0$.*

Theorem 3.6.2 (Infinite N). *Optimal strategies (resp. equilibrium strategies) exist in the limiting regime when $N \to \infty$ under uniform convergence and continuity of $R^N \to R$. Moreover, if $\{U^N\}$ is a sequence of ε_N-optimal strategies (resp. ε_N-equilibrium strategies) in the finite regime with $\varepsilon_N \longrightarrow \varepsilon$, then, any limit of subsequence $U^{\phi(N)} \longrightarrow U$ is an $\varepsilon-$ optimal strategy (resp. $\varepsilon-$equilibrium) for game with infinite N.*

4. Illustrating example

We present in this section an example of a dynamic version of the Hawk and Dove problem where each individual has three energy levels. We derive the mean field limit for the case where all users follow a given policy and where possibly one player deviates. We then further simplify the model to only two energy states per player. In that case we are able to fully identify and compute the equilibrium in the limiting MDEG. Interestingly, we show that the ODE converges to a fixed point which depends on the initial condition.

Consider an homogenous population of N animals. An animal plays the role of a *player*. Occasionally two animals find themselves in competition on the same piece of food. Each animal has three states $x = 0, 1, 2$ which represents its energy level. An animal can adopt an aggressive behavior (Hawk) or a peaceful one (Dove, passive attitude). At the state $x = 0$ there is no action. We describe the fitness of an animal (some arbitrary player) associated with the possible outcomes of the meeting as a function of the decisions taken by each one of the two animals. The fitnesses represent the following:

- An encounter Hawk-Dove or Dove-Hawk results in zero fitness to the Dove and in $\bar{v}$ of value for the Hawk that gets all the food without fight. The state of the Hawk (the winner) is incremented $a = 1_{\{x'_H = \min(x_H + 1, 2)\}}$ and the state of the Dove is $b = 1_{\{x'_D = \max(x_D - 1, 0)\}}$.

- An encounter Dove-Dove results in a peaceful, equal-sharing of the food which translates to a fitness of $\frac{\bar{v}}{2}$ to each animal and the state of each animal change with the sum of the two distributions $\frac{1}{2}a + \frac{1}{2}b$

- An encounter Hawk-Hawk results in a fight in which with $p = 1/2$ chances, one (resp. the other) animal obtains the food but also in which there is a positive probability for each one of the animals to be wounded $1/2$. Then the fitness of the animal 1 is $\frac{1}{2}(\bar{v} - c) + \frac{1}{2}(-c) = \frac{1}{2}\bar{v} - c$, where the $-c$ term represents the expected loss of fitness due to being injured.

$i\backslash j$	(g_i^N, g_j^N)	$X_i^N(t+\frac{1}{N}), X_j^N(t+\frac{1}{N})$
$D-D$	$(\frac{\bar{v}}{2}, \frac{\bar{v}}{2})$	$\frac{1}{2}\delta_{\min(x_1-1,0),\max(x_2+1,2)}$
		$+\frac{1}{2}\delta_{\max(x_1+1,2),\min(x_2-1,0)}$
$D-H$	$(0,v)$	$(\min(x_1-1,0),\max(x_2+1,2))$
$H-H$	$\frac{1}{2}v - c$	$\frac{1}{2}\delta_{\min(x_1-1,0),\max(x_2+1,2)}$
		$+\frac{1}{2}\delta_{\max(x_1+1,2),\min(x_2-1,0)}$

The vector of frequencies of states at time t is given by $M_x^N(t) = \frac{1}{N}\sum_{j=1}^N 1_{\{X_j^N(t)=x\}}$ for $x = 0, 1, 2$ and the action set is $A_x = \{H, D\}$ in each state $x \neq 0$, $A_0 = \{\}$.

The assumptions in Section 3 are satisfied (pairwise interaction, $\sharp\mathscr{B}^N(t) = 2$) and the occupancy measure $M^N(t)$ converges to $m(t)$.

4.1. ODE and Stationary strategies

Consider the following fixed parameters $\mu_1 = L_{0,1}$, $\mu_2 = L_{0,2}$. The population profile is denoted by $\vec{m} = (m_0, m_1, m_2)$ and the stationary strategy is described by the parameters v_1, v_2 where $v_1 := u(H|1)$, $v_2 = u(H|2)$

$\dot{m}_2 = m_0 L_{0,2} + m_1 L_{1,2}(u, m) - m_2 L_{2,1}(u, m)$
$\dot{m}_1 = m_0 L_{0,1} + m_2 L_{2,1}(u, m) - m_1 L_{1,2}(u, m) - m_1 L_{1,0}(u, m))$
$\dot{m}_0 = m_1 L_{10}(u, m) - (\mu_1 + \mu_2)m_0$

where $L_{12}(u, m) =$

$$m_0 + v_1\left(\frac{v_1 m_1}{2} + (1-v_1)m_1 + \frac{v_2 m_2}{2} + (1-v_2)m_2\right)$$
$$+(1-v_1)\left(\frac{(1-v_1)m_1}{2} + \frac{(1-v_2)m_2}{2}\right)$$
$$L_{2,1}(u,m) = v_2\left(\frac{v_1 m_1}{2} + \frac{v_2 m_2}{2}\right)$$
$$+(1-v_2)\left(\frac{(1-v_1)m_1}{2} + v_2 m_2 + \frac{(1-v_2)m_2}{2}\right)$$
$$L_{10}(u,m) := v_1\left(\frac{v_1 m_1}{2} + \frac{v_2 m_2}{2}\right)$$
$$+\ (1-v_1)\left(v_1 m_1 + \frac{(1-v_1)m_1}{2} + v_2 m_2 + \frac{(1-v_2)m_2}{2}\right),$$

For $\mathscr{B}^N = \{j_1, j_2\}$, $x'_j, x_i \in \{0,1,2\}$,

$$\frac{d}{dt}m_x = \sum_{x_1,x_2,x'_2} m_{x_1} m_{x_2} L_{x_1,x_2;x,x'_2}(u, \vec{m})$$
$$+ \sum_{x_1,x_2,x'_1} m_{x_1} m_{x_2} L_{x_1,x_2;x'_1,x}(u, \vec{m})$$
$$- m_x \sum_{x_2,x'_1,x'_2} m_{x_2} L_{x,x_2;x'_1,x'_2}(u, \vec{m})$$
$$- m_x \sum_{x_1,x'_1,x'_2} m_{x_1} L_{x_1,x;x'_1,x'_2}(u, \vec{m})$$

4.2. Computation of $R(u_1, u_2; s, \vec{m})$.

We want to compute the value

$$V(u_1, u_2, x) := \mathbb{E}_x \int_0^\infty e^{-\beta t} r(u_1, u_2, x(t), m(t))\, dt$$

$$s.t. \; \dot{m}(t) = f(u_2, m(t)), m(0) = m_0, \; x(0) = x.$$

$$
\begin{aligned}
V(u_1, u_2, x) &= \mathbb{E}_x \int_0^\Delta e^{-\beta t} r(u_1, u_2, x(t), m(t)) \, dt \\
&\quad + \mathbb{E}_x \int_\Delta^\infty e^{-\beta t} r(u_1, u_2, x(t), m(t)) \, dt \\
&= \mathbb{E}_x \int_0^\Delta e^{-\beta t} r(u_1, u_2, x(t), m(t)) \, dt \\
&\quad + \mathbb{E}_x e^{-\beta \Delta} V(u_1, u_2, x(\Delta))
\end{aligned}
$$

This implies that

$$
\begin{aligned}
0 &= \mathbb{E}_x \frac{1}{\Delta} \int_0^\Delta e^{-\beta t} r(u_1, u_2, x(t), m(t)) \, dt \\
&\quad + \frac{e^{-\beta \Delta} - 1}{\Delta} \mathbb{E}_x V(u_1, u_2, x(\Delta)) \\
&\quad + \frac{\mathbb{E}_x V(u_1, u_2, x(\Delta)) - V(u_1, u_2, x)}{\Delta}
\end{aligned}
\tag{4}
$$

Using Ito's formula and Lebesgue integration properties, we obtain that: $\frac{\mathbb{E}_x V(u_1, u_2, x(\Delta)) - V(u_1, u_2, x)}{\Delta}$ goes to $\sum_{x'} D_{m_{x'}} V(u_1, u_2, x') \frac{d}{dt} m_{x'}$, where $D_{m_{x'}} V$ is the derivative of V in a weak sense, $\frac{e^{-\beta \Delta} - 1}{\Delta} \longrightarrow -\beta$, and the term

$$\mathbb{E}_x \frac{1}{\Delta} \int_0^\Delta e^{-\beta t} r(u_1, u_2, x(t), m(t)) \, dt \longrightarrow r(u_1, u_2, x, m_0)$$

when Δ goes to zero. Thus, we obtain

$$\beta V(u_1, u_2, x) = r(u_{1,x}, u_{2,x}, x, m) + \sum_{x'} (D_{m_{x'}} V(u_1, u_2, x')) f_{x'}(u_2, m) \tag{5}$$

where $u_{i,x} = u_i(H|x)$.

The global optimality is then given by the Hamilton-Jacobi-Bellman equation obtained by maximizing the right-hand side of the equation (5) over the action set.

$$\beta \Psi(x) = \max_{u_{1,x}, u_{2,x}} \left\{ r(u_{1,x}, u_{2,x}, x, m) + \sum_{x'} (D_{m_{x'}} \Psi(u_1, u_2, x')) f_{x'}(u_2, m) \right\}$$

and optimality conditions of the best response to u_2 is given by

$$\beta \Phi(u_2, x) = \max_{a \in \{H, D\}} \left\{ r(a, u_{2,x}, x, m) + \sum_{x'} (D_{m_{x'}} \Phi(u_2, x')) f_{x'}(u_2, m) \right\}$$

Theses equations are in general difficult to solve and the solutions are not necessarily regular (e.g. viscosity solutions). Numerical approaches based on multi-grid techniques of Hamilton-Jacobi-Bellman-Issacs equations can be found [5].

4.3. The case of two energy levels

In order to derive closed form expressions for solutions of our ODE, we consider two states, i.e., each animal has two states $x = 1, 2$ which represents its energy levels. Thus, the ODE can be expressed as follows:

$$\dot{m}_2(t) = (1 - m_2(t)) L_{1,2}(u, m) - m_2(t) L_{2,1}(u, m) \tag{6}$$

which can be rewritten as

$$\dot{m}_2(t) = a_1 + a_2 m_2(t) + a_3 (m_2(t))^2 \tag{7}$$

with $a_1 = 1$, $a_2 = \frac{u_2}{2} - 2 < 0$, $a_3 = \frac{1 - u_2}{2} > 0$.

Let $m[u, m_0](t)$ be the solution of the ODE given u and a initial distribution $m(0) = m_0$. We distinguish two cases:

Case 1 $u_2 = 1$ (fully aggressive when it is possible): the ODE becomes $\dot{m}_2(t) = 1 - \frac{3}{2} m_2(t)$ and the solution has the form

$$m_2[1, m_0](t) = \frac{2}{3} [1 - c_1 e^{-\frac{3}{2} t}] \tag{8}$$

with $c_1 = 1 - \frac{3}{2} m_0$ and $m_1[u, m_0](t) = 1 - m_2[u, m_0](t)$

Case 2 $u_2 \neq 1$, (less aggressive in state 2)

$$m_2[u, m_0](t) = \gamma_-(u) + \frac{\gamma_+(u) - \gamma_-(u)}{1 - c_2 e^{(\gamma_+(u) - \gamma_-(u)) a_2 t}} \tag{9}$$

where
$$c_2 = 1 + \frac{\gamma_+(u) - \gamma_-(u)}{m_2(0) - \gamma_-(u)},$$

$$\gamma_-(u) = \frac{2 - u_2/2 - (2 + u_2^2/4)^{\frac{1}{2}}}{1 - u_2} < 1,$$

$$\gamma_+(u) = \frac{2 - u_2/2 + (2 + u_2^2/4)^{\frac{1}{2}}}{1 - u_2} > 1$$

Note that in both cases there is a unique strategy-dependent global attractor.

$$\lim_{t \longrightarrow \infty} m_2[u, m_0](t) = \begin{cases} \gamma_-(u) & \text{if } u_2 \neq 1 \\ 2/3 & \text{if } u_2 = 1 \end{cases}$$

The expected instant payoff of a player using the stationary strategy v when the population profile is $m[u, m_0](t)$, is given by

$$r(v, u, 2, m[u, m_0](t)) = v[\bar{v} - c m_2 u_2] + (1 - v) r(v, u, 1, m[u, m_0](t))$$

$$r(v, u, 1, m[u, m_0](t)) = \frac{1}{2} (1 - m_2[u, m_0](t) u_2) \bar{v}$$

where $m_2[u, m_0](t)$ is given by (8) (resp. (9)) for $u_2 = 1$ (resp. $u_2 \neq 1$). Now, we can compute explicitly the best response against u for a given initial m_0. Let

$$\beta_2(u, 2, m_0, t) = r(H, u, 2, m[u, m_0](t)) - r(D, u, 2, m[u, m_0](t)).$$

The best response, $BR(x, u, m[u, m_0](t))$, against u at t is

$$BR(x, u, m[u, m_0](t)) = \begin{cases} \text{play Hawk if } \beta_2(u, x, m_0, t) > 0 \\ \text{play Dove if } \beta_2(u, x, m_0, t) < 0 \end{cases}$$

This implies that it is better to play Hawk for $\frac{\bar{v}}{2c} > \frac{\gamma}{1+\gamma}$ where $\gamma = \max(2/3, m_0)$. Since the solution of the ODE is strictly monotone in time for each stationary strategy, there is at most one time for which β_2 is zero. It is easy to see that if $\frac{\bar{v}}{2c} > \frac{2}{3}$ then the strategy which to play Hawk in state 2 and Dove in state 1 is an equilibrium.
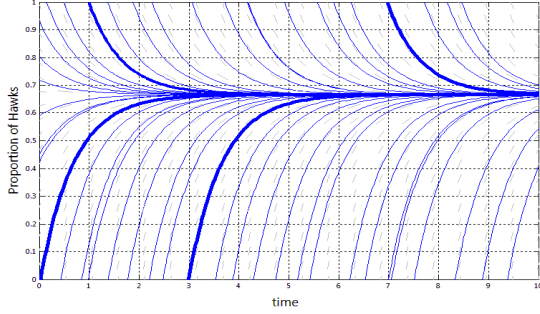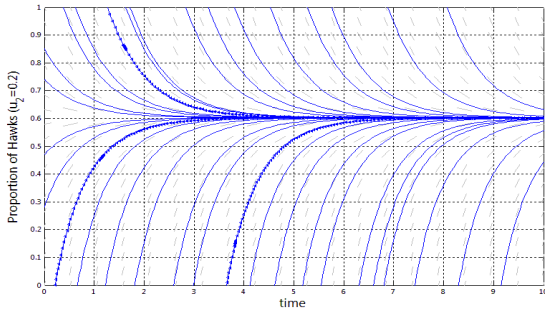

Figure 1. Global attractor for $u_2 = 1$

Figure 2. Global attractor for $u_2 = 0.2$

5. Concluding remarks

The goal of this paper has been to develop mean field asymptotic of interactions with large number of players using stochastic games. Due to the curse of the size of the population, the applicability of atomic stochastic games has been severely limited. As an alternative, we proposed a method for Markov decision evolutionary games where players make decisions only based on their own state and the global system state. We have showed under mild assumptions convergence results, where asymptotics were taken in the number of players. The population state profile satisfies a system of non-linear ordinary differential equations. We have considered very simple class of strategies that are functions only of player's own state and the population profile. We applied to Hawk-Dove interaction with several energy level and formulated the ODEs. We show that the best response depends on the initial conditions.

Appendix

Sketch of proof of Proposition 2.2.1

Let τ^N be the first time after $t = 0$ that $X_j^N(t)$ hits in some given state. We show that

$$\bar{r}^N = \frac{1}{N} \mathbb{E} \sum_{\substack{s=0 \\ \text{step } 1/N}}^{\tau^N} e^{-\beta t} r^N \left(X_j^N(s), M^N(s) \right) \qquad (10)$$

Define for $t \in \mathbb{N}/N$:

$$Z_t^N = \sum_{\substack{s=0 \\ \text{step } 1/N}}^{t} e^{-\beta s} \left(G^N(s) - r^N \left(X_j^N(s), M^N(s) \right) \right)$$

we have, for $0 \le s \le t$:

$$\begin{aligned} Q &:= \mathbb{E} \left(Z_t^N - Z_s^N | \mathscr{F}_s^N \right) \\ &= \sum_{\substack{u'=0 \\ \text{step } 1/N}}^{t} e^{-\beta u'} \mathbb{E} \left(G^N(u') - r^N \left(X_j^N(u'), M^N(u') \right) | \mathscr{F}_s^N \right) \end{aligned}$$

which can be written as

$$\sum_{\substack{u'=0 \\ \text{step } 1/N}}^{t} e^{-\beta u'} \mathbb{E} \left(\mathbb{E} \left(G^N(u') - r^N \left(X_j^N(u'), M^N(u') \right) | \mathscr{F}_{u'}^N \right) | \mathscr{F}_s^N \right)$$

$$= 0$$

thus Z_t^N is an $\mathscr{F}_t^N$− martingale. Now τ^N is a stopping time with respect to the filtration $\mathscr{F}_t^N$ thus, by Doob's stopping time theorem: $\mathbb{E}Z_{t \wedge \tau^N}^N = \mathbb{E}Z_{0 \wedge \tau^N}^N = 0$ Further, $Z_{t \wedge \tau^N}^N \le K|\tau^N|$ for some constant K. Since τ^N is almost surely finite and has a finite expectation, we can apply dominated convergence (with $t \to \infty$) and obtain $\mathbb{E}Z_{\tau^N}^N = 0$.

Sketch of Proof of Theorem 3.2.1

To prove the weak convergence of Z^N, we check the following steps: Without loss of generality, we took the set of states as $\mathscr{S} = \{0, 1, 2, \ldots, \sharp \mathscr{S}\}$ X_j^N has a jump r with probability

$$q_{i,i+r}^N(M^N(k)) = \frac{1}{N} L_{i,i+r}^N(M^N(k), u))$$

and M^N is the continuous process with drift f^N.

- We introduce of $\tilde{X}_j^N$ by scaling with step size $\frac{1}{N}$. Then, $Z^N = (X^N, M^N)$ is approximate in some sense by a discrete time process $\tilde{Z}^N = (\tilde{X}^N, \tilde{m}^N)$ where $\tilde{m}^N(k) = m(\lfloor Nt \rfloor)$ m solution of the ODE with $\tilde{X}_j^N$ is the discrete time jump process with transition matrix

$$q_{i,i+r}^N(\tilde{m}^N(k)) = \frac{1}{N} L_{i,i+r}^N(m(\frac{k}{N}), u)).$$

We show that $d(X_j^N, \tilde{X}_j^N) \longrightarrow 0$ for any compact of time intervals.

-
$$\tilde{Z}^N = (\tilde{X}^N, \tilde{m}^N) \Longrightarrow (\tilde{X}, \tilde{m})$$

$M^N([Nt]) \longrightarrow m(t)$. We derive the weak convergence of Z^N to (X, m) where m is deterministic and X is random.

Approximation by a discrete time process

The following lemma follows from the lemma 1 and 3 in Benaim and Weibull (2003,2008), in which we incorporate behaviorial strategies.

Lemma 5.0.1. *For every $t > 0$ there exists a constant c such that for every $\varepsilon > 0$ and N large enough one has*

$$P(\sup_{0 \le \tau \le T} ||M^N(\tau) - m(\tau)|| > \varepsilon \mid M^N(0) = m_0, u) \le 2(\sharp S) e^{-\varepsilon^2 CN}$$

for all $m_0 \in \Delta_d$, all every stationary strategy u.

Since C is independent of N, and $(e^{-\varepsilon^2 C})^N$ is summable, we can use the dominated convergence theorem: for all $\varepsilon > 0$,

$$\sum_N \mathbb{P}\left(\sup_{0 \le \tau \le T} \| M^N(\tau) - m(\tau) \|_\infty > \varepsilon \mid M^N(0) = m_0, u\right) < \infty,$$

By Borel-Cantelli's lemma, for every fixed $t < \infty$, the random variable $v^{N,t} := \sup_{0 \le \tau \le t} \| M^N(\tau) - m(\tau) \|_\infty$ converges almost completely towards 0. This $v^{N,t}$ implies that converges almost surely to 0.

We introduce of $\tilde{X}_j^N$ by scaling with step size $\frac{1}{N}$. Then, $Z^N = (X^N, M^N)$ is approximate in some sense by a discrete time process $\tilde{Z}^N = (\tilde{X}^N, \tilde{m}^N)$ where $\tilde{m}^N(k) = m(\lfloor Nt \rfloor)$ m solution of the ODE where $\tilde{X}_j^N$ is the discrete time jump process with transition matrix

$$q_{i,i+r}^N(\tilde{m}^N(k)) = \frac{1}{N} L_{i,i+r}(m(\frac{k}{N}), u)).$$

Using the lemma 5.0.1 and uniform Lipschitz continuity of of L^N, we obtain that

$$\sup_{i,j} \sup_{0 \le \tau \le t} \| q_{i,j}^N(M^N(\tau)) - q_{i,j}(m(\tau)) \|$$

$$\le K(\varepsilon_N + \sup_{0 \le \tau \le t} \|M^N(\tau) - m(\tau)\|).$$

Hence, we can write $\|M^N(\tau) - m(\tau)\| \le K(\varepsilon_N + \frac{1}{N^2})$ over set of event $\Omega_\varepsilon = \{\|M^N(\tau) - m(\tau)\| \le \varepsilon\}$ and $P(\Omega_\varepsilon) \ge 1 - 2(\sharp S) e^{-\varepsilon^2 CN} \to 1$. Thus,

$$P(X^N_{j,|[0,t]} = \tilde{X}^N_{j,|[0,t]} | k \text{ transitions}) \ge \mathbb{E}(\varepsilon^{Bin(\frac{1}{N}, Nt)})$$

$$\mathbb{E}(\varepsilon^{Bin(\frac{1}{N}, Nt)}) = (1 - \frac{1}{N} + \frac{1}{N}\varepsilon)^{Nt}$$

$$P(X^N_{j,|[0,t]} = \tilde{X}^N_{j,|[0,t]} | k \text{ transitions}) \ge e^\varepsilon$$

and this holds for any ε arbitrary small. We define $d(X,Y) = \sum_{k=0} \frac{1}{2^k} d(X_k, Y_k)$ where $d(X_k, Y_k) = 1_{X_k \ne Y_k}$. Then, $d(X^N_{j,|[0,t]}, \tilde{X}^N_{j,|[0,t]}) \longrightarrow 0$ when N goes to infinity.

Convergence of the discrete time process To prove the weak convergence of $(\tilde{X}_j^N, \tilde{M}^N)$, we check the following steps:

- the discrete time empirical measures $\tilde{M}^N$ are tight (follows from Sznitman for finite states) and converges to a martingale problem. The limit $\tilde{m}$ is deterministic measure and is solution of ODE which has the unique solution m (given m_0, u). Thus, $\tilde{m} = m$.

- Conditionally to $\tilde{M}^N$, $\tilde{X}_j^N$ converges to a martingale problem. The jump and drift process $\tilde{X}$ with time dependent transition is given by the limit of the marginal of $A^N(. | \tilde{M}^N, m_0, x_0, u)$. We derive the weak convergence of $(\tilde{X}_j^N, \tilde{M}^N)$ to $(\tilde{X}, \tilde{m})$ where $\tilde{m}$ is deterministic and $\tilde{X}$ is random. For this we use the Theorem 17.25 and its discrete time approximation in Theorem 17.28 pages 344-347 in Kallenberg.

Sketch of Proof of Theorem 3.3.1

Since Skorohod's topology is induced by a metric, it is sufficient to show that whenever $(X_j^N, m^N) \to (x, m)$ in Skorohod's topology, we have:

$$\lim_{N \longrightarrow \infty} \int_0^\infty e^{-\beta t} r^N(v, X_j^N(t), m^N(t)) dt$$

$$= \int_0^\infty e^{-\beta t} r(v, x(t), m(t)) dt$$

By [4], page 117, there is some sequence of increasing bijections $\lambda_n : [0, \infty) \to [0, \infty)$ s.t.

$$\frac{\lambda_n(t) - \lambda_n(s)}{t - s} \to 1 \text{ uniformly in } t \text{ and } s$$

and $\| y_n(t) - y(\lambda_n(t)) \| \to 0$ uniformly in t

over compact subsets of $[0, \infty)$. Fix $\varepsilon > 0$, arbitrary and consider

$$h^N := |\int_0^\infty e^{-\beta t} r^N(X^N(t), v, m^N(t)) dt$$
$$- \int_0^\infty e^{-\beta t} r(x(t), v, m(t)) dt|$$
$$\leq \int_0^\infty e^{-\beta t} |r^N(x^N(t), v, m^N(t)) - r(x(t), v, m(t))| dt$$

First let $K = \sup_{x \in \mathscr{S}, v, m \in \Delta} |r(x, v, m)| < \infty$ by hypothesis, and pick some time T large enough such that $e^{-\beta T} K / \beta \leq \varepsilon/3$. Thus

$$h^N \leq \varepsilon/3 + \int_0^T e^{-\beta t} |r(x^N(t), v, m^N(t)) - r(x(t), v, m(t))| dt \tag{11}$$

Second, we use the distance on E defined by

$$d((x, m), (x', m')) = \| m - m' \| + 1_{x \neq x'} \tag{12}$$

Let $K' = \sup_{x \in \mathscr{S}, v, m \in \Delta_d} \frac{|r(x, v, m) - r(x', v, m')|}{\| m - m' \|} < \infty$

by hypothesis. It is easy to see that for all $x, x' \in \mathscr{S}$ and $m, m' \in \Delta_d$:

$$\| r(x, v, m) - r(x', v, m') \| \leq K' d((x, m), (x', m')) \tag{13}$$

Thus, by Equation (11):

$$h^N \leq \varepsilon/3 + K' \int_0^T e^{-\beta t} d((x^N(t), m^N(t)), (x(t), m(t))) dt \tag{14}$$

By [4], page 117, there is some sequence of increasing bijections $\lambda^N : [0, \infty) \to [0, \infty)$ s.t.

$$\frac{\lambda^N(t) - \lambda^N(s)}{t - s} \to 1 \text{ uniformly in } t \text{ and } s$$

and $d((x^N(t), m^N(t), (x^N(\lambda^N(t)), m^N(\lambda^N(t)))) \to 0$

uniformly in t over compact subsets of $[0, \infty)$. Thus there is some $N_0 \in \mathbb{N}$ such that for $N \geq N_0$ and $t \in [0, T]$:

$$d((x^N(t), m^N(t), (x^N(\lambda^N(t)), m^N(\lambda^N(t)))) \leq \frac{\varepsilon \beta e^{\beta T}}{3K'} \tag{15}$$

Thus, by the triangular inequality for d: $h^N \leq$

$$\leq \frac{\varepsilon}{3} + K' \int_0^T e^{-\beta t} d((x^N(t) m^N(t)), (x(\lambda^N(t)), m(\lambda^N(t)))) dt$$

$$+ K' \int_0^T e^{-\beta t} d((x(\lambda^N(t)), m(\lambda^N(t))), (x(t), m(t)) dt$$

$$\leq \frac{2\varepsilon}{3} + K' \int_0^T e^{-\beta t} d((x(\lambda^N(t)), m(\lambda^N(t))), (x(t), m(t))) dt \tag{16}$$

Third, let D be the set of discontinuity points of (x, m). Since (x, m) is cadlag, D is enumerable, thus it is negligible for the Lebesgue measure and

$$\int_0^T e^{-\beta t} d((x(\lambda^N(t)), a, m(\lambda^N(t))), (x(t), a, m(t)) dt$$

$$= \int_0^T e^{-\beta t} d((x(\lambda^N(t)), m(\lambda^N(t))), (x(t), m(t)) 1_{t \notin D} dt$$

Now $\lim_{N \to \infty} \lambda^N(t) = t$ and thus for $t \notin D$

$$\lim_{N \to \infty} d((x(\lambda^N(t)), m(\lambda^N(t))), (x(t), m(t)) = 0$$

and thus by dominated convergence

$$\lim_{N \to \infty} \int_0^T e^{-\beta t} d((x(\lambda^N(t)), m(\lambda^N(t))), (x(t), m(t))) dt = 0 \tag{17}$$

and for N large enough the second term in the right-hand side of Equation (16) can be made smaller than $\varepsilon/3$. Finally, for N large enough, $h^N \leq \varepsilon$. This completes the proof.

Sketch of Proof of Theorem 3.3.2

Define the discounted stochastic evolutionary game with random number of interacting players in each local interaction in which each player in x with the mixed action $u(.|x)$ receives $r(u, x, m(t))$ where $m(t)$ is the population profile at t, which evolves under the dynamical system (3) and the between states follows the transition kernel L. Then, a strategy of a player is the same as in the microscopic case and the discounted payoffs

$$R(u_1, u_2, s_0, m_0) = \int_0^\infty e^{-\beta t} r(s(t), u_1, m[u_2](t)) dt$$

is the limit of $R^N(u_1, u_2, s_0, m_0)$ when N goes to infinity, where $m[u_2]$ is the solution of the ODE $\dot{m} = f(u_2, m), m(0) = m_0$. It follows that the asymptotic regime of the microscopic game and the Markov decision evolutionary game (macroscopic game) are equivalent.

Sketch of Proof of Theorem 3.6.1

We show that for every discount factor $\beta > 0$ the optimal control problem (OPT_N) (resp. the fixed-point problem (FIX_N)) has at least one 0−optimal strategy. It follows from the existence of equilibria in stationary strategies for finite stochastic games with discounted payoff: The set of pure strategies is a compact space in the product topology (Tykhonov theorem). Thus, the set of behavioral strategies Σ_j is a compact space and also convex as the set of probabilities on the pure strategies. For every player j and every strategy profile σ the marginal of the

payoffs and constraints functions are continuous for any $\beta > 0 : \alpha_j \longmapsto R_j^N(\alpha_j, \sigma_{-j}, s, m_0)$. Moreover, the stationary strategies is convex, compact and upper and lower hemi-continuous (as a correspondence). Define

$$\gamma_j(s, m_0, \sigma) = \arg \max_{\alpha_j \in \mathscr{U}_s} R_j^N(\alpha_j, \sigma_{-j}, s, m_0).$$

Then, $\gamma_j(m_0, \sigma) \subseteq \Sigma_j$ is a non-empty, convex and compact set and the product correspondence

$$\gamma : \sigma \longmapsto (\gamma_1(s, m_0, \sigma), \ldots, \gamma_N(s, m_0, \sigma))$$

is upper hemi-continuous (its graph is closed). We now use the Glicksberg generalization of Kakutani fixed point theorem, and there is a stationary strategy profile σ^* such that

$$\sigma^* \in \gamma(s, m_0, \sigma^*).$$

Moreover, if the game has symmetric payoffs and strategies for each type, there is a symmetric per type stationary equilibrium. This completes the proof.

Sketch of Proof of Theorem 3.6.2

Let $(U^N)_N$ be a sequence of solution of (FIX_N) i.e equilibrium in the system with N players. Choose a subsequence N_k such that U^{N_k} converges to some point u when k goes to infinity. We can write $R^{N_k}(U^{N_k}, U^{N_k}) - R(U, U) = R^{N_k}(U^{N_k}, U^{N_k}) - R^{N_k}(U, U) + R^{N_k}(U, U) - R(U, U)$. Since $R^N(., .)$ is continuous and converges uniformly to $R(., .)$, R^{N_k} converges uniformly to R, the second term $R^{N_k}(U, U) - R(U, U) \longrightarrow 0$ when $N_k \longrightarrow \infty$ and the first term $R^{N_k}(U^{N_k}, U^{N_k}) - R^{N_k}(U, U)$ can be rewritten as $R^{N_k}(U^{N_k}, U^{N_k}) - R^{N_k}(U, U) = R^{N_k}(U^{N_k}, U^{N_k}) - R(U^{N_k}, U^{N_k}) + R(U^{N_k}, U^{N_k}) - R(U, U) + R(U, U) - R^{N_k}(U, U)$. Each term goes to zero by continuity of R, convergence of U^{N_k} to U, and by uniform convergence of R^N to R. Let U^N be a ε_N-equilibrium. Then, $R^N(U^N, U^N) \geq R^N(v, U^N) - \varepsilon_N$, $\forall v$. Then any limit U of a subsequence of U^N satisfies $R(U, U) \geq R(v, U) - \varepsilon$, $\forall v$. Similarly, if

$$R^N(U^N, U^N) \geq R^N(v, v) - \varepsilon_N, \ \forall v$$

then any omega-limit U of the sequence of U^N satisfies $R(U, U) \geq R(v, v) - \varepsilon$, $\forall v$ i.e U is an ε-optimal strategy. In particular if $(U^N)_N$ is a sequence of ε_N-equilibria (resp. optimal strategies) with $\varepsilon_N \longrightarrow 0$ when N goes to infinity then any accumulation point U of $(U^N)_N$ is a 0-equilibrium (resp. 0-optimal strategy).

References

[1] E. Altman, Y. Hayel, H. Tembine, R. El-Azouzi, "Markov decision Evolutionary Games with Time Average Expected Fitness Criterion", In proc. of Valuetools, October, 2008.

[2] Benaim, M. and Le Boudec, J. Y. , *A Class Of Mean Field Interaction Models for Computer and Communication Systems*, Performance Evaluation, 2008.

[3] Benaim, M. and Weibull J. W. (2003). *Deterministic Approximation of Stochastic Evolution in Games*, Econometrica 71, 873-903

[4] Stewart N. Ethier and Thomas G. Kurtz. Markov Processes, Characterization and Convergence. Wiley, 2005.

[5] Kushner, Harold J., and Paul G. Dupuis, Numerical Methods for Stochastic Control Problems in Continuous Time, New York: Springer-Verlag, 1992.

[6] Kurtz T. G., *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes*, Journal of Applied Probability, Vol. 7, No. 1 (Apr., 1970), pp. 49-58.

[7] Kurtz T. G., *Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes*, Journal of Applied Probability, Vol. 8, No. 2 (Jun., 1971), pp. 344-356.

[8] Tanabe Y., *The propagation of chaos for interacting individuals in a large population*, Mathematical Social Sciences, 2006,51,2,pp.125-152.

[9] G. Y. Weintraub, L. Benkard, B. Van Roy, *Oblivious Equilibrium: A mean field Approximation for Large-Scale Dynamic Games*, Advances in Neural Information Processing Systems, Vol 18, 2006.

[10] Le Boudec J.Y., McDonald D. , and Mundinger J., A Generic Mean Field Convergence Result for Systems of Interacting Objects. In QEST 2007 pages 3–18.