Mean field stochastic games

Tembine Hamidou

Supelec, France

AlgoGT, Grenoble, June 20-21, 2011

Mean field stochastic games?

- Games,
- Stochastic games (*Shapley 1953*)
- "Mean field" + "stochastic games"

Mean field-like games

- Continuum of players: von Neumann'44, Nash'51, Wardrop'52, Aumann'64, Selten'70, anonymous sequential games, Jovanovic & Rosenthal'88, Bergins et al. etc
- Differential games in evolutionary equilibrium selection: Tarasyev'1994, Hofbauer, Sorger'02
- Mean field stochastic control & games: Huang, Caines, Malhame'03-, Kotelenez & Kurtz'07-08-, Li & Zhang'08, Yin, Mehta, Meyn, and Shanbhag'10-, Feng et al.'10-,
- Mean field games - coupling HJBF+Kolmogorov equations : Lasry & Lions'06-, Buckdahn, Li, Peng'07-, Gueant'09-, Gomes et al.'09, Dogbe'10-, Achdou et al.'10-, LaChapelle'10- Başar et al.'11, etc
- QSA and mean field interaction:Weibull& Benaïm'03-, Weintraub, Benkard, Van Roy'05-, Sandholm '06-, Adlaska, Johari, Goldsmith'08-, Benaïm & Le Boudec'08-, Gast, Gaujal'09- Bardenave'09-. etc

Static Aggregative Games: some references

- R. Aumann, Markets with a continuum of traders, Econometrica 32 (1964),
- Selten, R. (1970), Preispolitik der Mehrprodktenunternehmung in der statischen Theorie, Springer-Verlag.
- Dubey, P., A. Mas-Colell, and M. Shubik (1980): "Efficiency Properties of Strategic Market Games: An Axiomatic Approach", Journal of Economic Theory 22, 339-362.
- Corchon, L. (1994): "Comparative Statics for Aggregative Games. The Strong Concavity Case", Mathematical Social Sciences 28, 151-165

A common property: **"The payoff function of each player depends its own action and an aggregative term of the other actions".**

$$r_j(a_1, \ldots, a_n) = \tilde{r}_j(a_j, \phi(a)),$$

Example: (–) Cournot oligopoly $\phi(a) = \sum_i a_i$ (–) Interference models in wireless networks, $\phi(a) = \sum_i w_i a_i$

The basic setting

- A class of stochastic games with
  - resource states,
  - types,
  - individual states and,
  - actions per type-state.
- At each stage, there is a random number of interacting players.
- The states and the actions of all the interacting players determine together the instantaneous payoffs and the transitions to the next states.

How the game is played?

- $\mathcal{N} = \{1, 2, \ldots, n\}$ set of players and $\mathcal{S}$ set of resource states. Each player has its own state $x_{j,t}^n \in \mathcal{X}$ (possibly different classes) and can make a decision based on its current state and the resource state $\mathcal{A}(x_j^n, s)$
- Discrete time $\mathbb{T}_n = \{0, \delta_n, 2\delta_n, \ldots\}$.
- At each time step t, a random set $\mathcal{B}^n(t) \in 2^{\mathcal{N}}$ of players interact. The individual states change $L^n(x_{t+1}^n | x_t^n, s_t, a_t^n)$; the resource state changes $(\tilde{q}(s_{t+1} | s_t, a_t^n))$; each player j in $\mathcal{B}^n(t)$ gets a payoff $r_t(.)$,
- The system goes to $t+1$, a new random set $\mathcal{B}^n(t+1)$ is drawn.

Stochastic games with additional individual states

Mean field stochastic games (1)

Mean field stochastic game of interest will have many interacting players

- Large system of players seeking their "interest"
- Main issue: the relation between the state and actions of each player and the resulting mass behavior

$$M_t^n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{\{x_{j,t}^n = x\}}$$

Mean field stochastic games (2)

How to combine mean field and stochastic games? • Input: several decision-makers, • Modelling: microscopic description (states, controls, transitions, effects, payoffs). Question: How to reduce the complexity?

• Mean field approach: asymptotics & macroscopic interaction

$$\begin{tikzcd} M^n[u,m_0](t) \arrow[r, "t \longrightarrow +\infty"] \arrow[d, "n \longrightarrow +\infty"] & \varpi^n[u,m_0] \arrow[d, "n \longrightarrow +\infty"] \\ m[u,m_0](t) \arrow[r, "t \longrightarrow +\infty"] & ? \end{tikzcd}$$

$H0$: **Indistinguishability per type +** $(x_1^n, \ldots, x_n^n)$ **Markovian.**

joint work with Le Boudec, ElAzouzi & Altman, 2008-

Micro to Macro[1]: $\delta_n \longrightarrow 0$

Let $\mathcal{M}_n^d = \{m \mid nm \in \mathbb{N}^d\}$. Suppose that

D0: For every s, the function $w_s(m)$ is C^1 in m.

D1: $\exists\ 0 < \delta_n, \epsilon_n \searrow 0$, and a C^1- function $f : \mathbb{R}^d \times \mathcal{S} \longrightarrow \mathbb{R}^d$ such that $\lim_n \sup_{\|m\|\leq 1} \| \frac{f^n(m,s)}{\delta_n} - f(m,s) \| = 0$, where $x \in \mathcal{X}$ and

$$f_x^n(m,s) = \int_{m' \in \mathcal{M}_n^d} \mathbf{1}_{\|m'-m\|\leq 2}(m'_x - m_x)\mathcal{L}^n(dm'; m, s),$$

D2: $\sup_n \frac{1}{\delta_n} \int_{m' \in \mathcal{M}_n^d} \| m' - m \| \mathcal{L}^n(dm'; m, s) < +\infty$

D3: $\lim_n \frac{1}{\delta_n} \int_{m' \in \mathcal{M}_n^d} \mathbf{1}_{\|m'-m\|>\epsilon_n} \| m' - m \| \mathcal{L}^n(dm'; m, s) = 0$,

D4: $M_0^n = m_0^n$ converges to $m_0 \in \Delta(\mathcal{X})$.

[1] Kurtz, 1970

Deterministic mean field dynamics

ODE

Assume D0-D4. Then, for all $\epsilon > 0$, $T < +\infty$,

$$\lim_{n} \mathbb{P}\left(\sup_{t\in[0,T]} \parallel M^n_{\frac{t}{\delta_n}} - m_t[m_0] \parallel > \epsilon \right) = 0,$$

where $m_t[m_0]$ is the unique solution of the ODE $\dot{m}_t = \tilde{f}(m_t)$ starting from $m_0 \in \Delta(\mathcal{X})$ where $\tilde{f}(m_t) := \sum_{s\in\mathcal{S}} w_s(m_t) f(m_t, s)$.

The result extends to (i) *controlled case*, (ii) diffusion process (*noisy mean field limit*) under weaker conditions. D2-D3 relaxed and second order term controlled, (iii) Mean field with *switching*, (iv) *fractional* mean field. *Limitation of the deterministic mean field approach (see Schnakenberg'79)*

The standard deterministic evolutionary game dynamics based on revision protocols are in the form

$$\dot{m}_t(x) = \sum_{x' \in \mathcal{X}} \mathcal{L}_{x'x}(m_t) m_t(x') - m_t(x) \sum_{x' \in \mathcal{X}} \mathcal{L}_{xx'}(m_t), \quad (1)$$

which is a specific mean field equation.
Examples: replicator, Smith, Brown-von Neumann-Nash, logit, imitation dynamics etc.

Simulation of population games

- large population of $n = 8000$ players,
- random selection of single player per time slot.

$$
\left(\begin{array}{cccc}
 & R & P & S \\
R & (0,0) & (-1,1) & (1,-1) \\
P & (1,-1) & (0,0) & (-1,1) \\
S & (-1,1) & (1,-1) & (0,0)
\end{array}\right)
$$

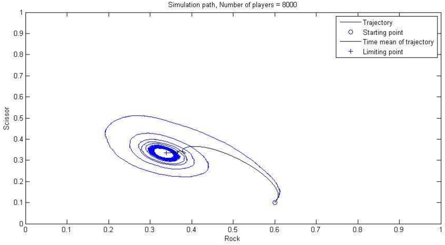
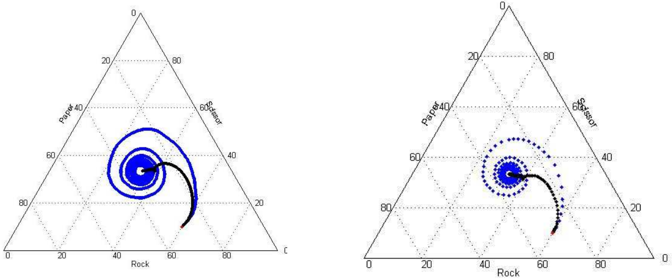
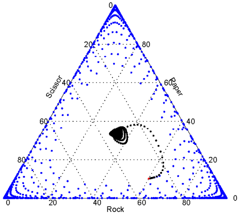
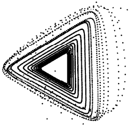


Illustration (2)



RSP: 8000 players





Micro to Macro:$\delta_n \longrightarrow \delta > 0$.

[H1] Assume that the marginal of L^n converges to q
We assume H1 holds and the limiting kernel is continuous.
• The mean field (limit) is in discrete time,
• The evolution of m_t is given by the Chapman-Kolmogorov equation.
• each generic player is facing an evolving macroscopic object.
The risk-sensitive criterion is considered.

Risk-sensitive MFSG

The previous approaches on mean field are risk-neutral type. Not all behavior, however, can be captured by risk-neutral objective functions.

The certainty-equivalent expectation $e(R)$ is defined by $g(e(R)) = \mathbb{E}(g(R))$. When g is exponential

$$e(R) = g^{-1}\left(\mathbb{E}(g(R))\right) = \mu \log\left(\mathbb{E}\left(e^{\frac{R}{\mu}}\right)\right), \ \mu \neq 0$$

- Howard & Matheson 1972: Risk-sensitive Markov decision processes. Management Science, 18, 356-369.
- Jacobson 1973: Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. IEEE Trans. Automat. Contr., 18, 124-131.

Intuitive view of the criterion

$$R_{\mu_*} := \frac{1}{\mu_*} * sign(\mu_*) \log \mathbb{E} \left(e^{\mu_* [g_{T+1}(x_{T+1}) + \sum_{t'=t}^{T} r_{t'}(x_{t'}, u_{t'}, M^n_{t'})]} \right),$$

Taylor expansion at μ_* close to zero leads

$$R_{\mu_*} = \mathbb{E}(R) + \frac{\mu_*}{2} var(R) + o(\mu_*)$$

Takes into consideration not only the expectation but also the variance!

- If $\mu_* \longrightarrow 0$ risk-neutral.
- If $\mu_* > 0$ risk-seeking
- If $\mu_* < 0$ risk-averse.

One-shot risk-sensitive games

- payoff is $\mu_\theta \log \left(\mathbb{E}_{m_\theta, m_{-\theta}} e^{\frac{1}{\mu_\theta} r_\theta(a)} \right)$,
- Risk-sensitive mean field equilibrium (static) $\tilde{r}_{\mu,\theta} = e^{\frac{1}{\mu_\theta} r_\theta}$ $\forall \theta, \ \langle (m_\theta - m_\theta^*), \tilde{r}_{\mu,\theta}(m^*) \rangle \leq 0, \ \forall m_\theta \in \Delta(\mathcal{A}(\theta))$
- Existence (VI), Uniqueness (Stable population games).

Remark: dynamic equilibrium?
$\forall \theta, \ \sum_{t=0}^{T} \langle (m_{\theta,t} - m_{\theta,t}^*), \tilde{r}_{\mu,\theta}(m_t^*) \rangle \ \leq 0, \ \forall m_{\theta,t}$
Next we introduce state-dependency.

Long-term

$\mathcal{F}_t^n := \sigma(x_{t'}^n, a_{t'}^n,\ t' \leq t)$. Given a history $h_t = (x_0^n, a_0^n, \ldots, x_t^n, a_t^n)$ in $\mathcal{F}_t^n$, the state profile x_{t+1}^n evolves according to the transition probability $L^n(x'; x, u) = \mathbb{P}\left(x_{t+1}^n = x' \mid h_t\right)$. The mean field evolves according to the kernel

$$\mathcal{L}^n(m'; m, u) := \mathbb{P}(M_{t+1}^n = m' \mid \tilde{h}_t)$$

$$F_{\infty,\mu}(\sigma, x, m) = \mu \log \mathbb{E}\left(e^{\frac{1}{\mu}\sum_{t\geq 0} r_t(x_t, a_t, m_t)}\right) = g^{-1}(\mathbb{E}(g(R_\infty))) \quad (2$$

Let $v_{j,\mu}(x, m) = \sup_\sigma F_{\infty,\mu}(\sigma, x, m)$ and $q_{x\sigma x'}(m)$ be the marginal of the limiting of L^n for a generic player j.

The risk-sensitive mean field DP principle:

$$\left\{\begin{array}{c} g(v^*_{j,\mu}(x_t,m_t)) = \max_{u\in\mathcal{A}(x_t)}\left[e^{\frac{1}{\mu}r_t(x_t,u,m_t)}\sum_{x'} q_{x_t u x'}(m_t) g(v^*_{j,\mu}(x',m_{t+1}))\right] \\ m_{t+1}(x') = \sum_{\bar{x}\in\mathcal{X}} m_t(\bar{x})\mathcal{L}_{\bar{x},x'}(u^*_t,m_t) \\ u^*_t \in \arg\max_u e^{\frac{1}{\mu}r_t(x_t,u,m_t)}\sum_{x'} q_{xux'}(m_t) g(v^*_{j,\mu}(x',m_{t+1})). \end{array}\right.$$

Risk-sensitive mean field equilibrium

A pair $(u_t^*, m_t^*)_{t\geq 0}$ is a risk-sensitive mean field equilibrium if $\{u_t^*\}_{t\geq 0}$ is a risk-sensitive best-response to the mean field trajectory $\{m_t^*\}_{t\geq 0}$ and for any time t, u_t^* generates the mean field m_t^*.

Assumption H2. The mapping $r_t(.)$ is positive. The infinite sum in $F_{\infty,\mu}$ is finite. In the continuous case, the payoff and the transition probabilities are continuous in (a, m).

Irreducible case

Assume $H0 - H2$ holds. Assume a stationary strategy π satisfies:

- $\forall x,\ g(v^*_{j,\mu}(x, m^*)) = e^{\frac{1}{\mu} r(x,\pi(x),m^*)} \sum_{x'} q_{x\pi(x)x'}(m^*) g(v^*_{j,\mu}(x', m^*))$,
- The strategy π generates a Markov decision process with unique positive recurrent class,
- $m^*(x') = \sum_{\bar{x}\in\mathcal{X}} m^*(\bar{x})\mathcal{L}_{\bar{x},\pi(\bar{x}),x'}(m^*)$

Then, π is a risk-sensitive compatible best-response to the mean field (among all the general strategies).

Idea of proof:

- For each x, m and t, let $w_t(x, m)$ be

$$g(w_t(x_t, m_t)) = \mathbb{E}_\pi \left[g(v^*_{j,\mu}(x_t, m_t)) \mid x_0 = x, m_0 = m \right].$$

- By H2,

$$g(w_{t+1}(x, m)) \leq g(w_t(x, m)), \ \forall \ x.$$

  i.e $t \longrightarrow w_t$ is monotone decreasing in time.
- the ergodic Markov theorem gives the independence of payoff in x :

$$\lim_{t \longrightarrow \infty} \ w_t(x, m) = w^*(m), \ \forall x.$$

- $v^*_{j,\mu}(x, m) - w^*(m)$ satisfies the comparison property. i.e $v^*_{j,\mu}(x, m) - w^*(m) \geq v^*_{j,\mu}(m)$ from which we deduce $w^*(m) \leq 0$
- we deduce that $w^* = 0$

- $e^{\frac{1}{\mu}\sum_{t=0}^{T} r(x_t,a_t,m_t)} \longrightarrow e^{\frac{1}{\mu}\sum_{t=0}^{+\infty} r(x_t,a_t,m_t)}$
- $e^{\frac{1}{\mu}\sum_{t=0}^{T} r(x_t,a_t,m_t)} g(v^*_{j,\mu}(x_{T+1}, m_{T+1})) \longrightarrow e^{\frac{1}{\mu}\sum_{t=0}^{+\infty} r(x_t,a_t,m_t)} g(w^*(m))$
- $v^*_{j,\mu}(x, m) = F_{j,\mu}(\pi, x, m)$

A basic example (Cavazos-Cadena et al. 2000)

Suppose $\mu > 0$, $S = \{\}$.

- $\mathcal{X} = \{0, 1\}$.
- $\mathcal{A}_j(0) = \{0\}$, $\mathcal{A}_j(1) = [0, 1]$.
- $r(0, a_0, m) = 0$; $r(1, a_1, m) = \frac{\mu}{2} - a_1 \frac{\mu}{2} + \tilde{\epsilon}(1 - a_1)\tilde{h}(m)$.

Transition probabilities

- state 0: $q_{000}(m) = 1, q_{001} = 0.$
- state 1 : $q_{1a_11}(m) = a_1 = 1 - q_{1a_10}(m).$

we present the case where $\tilde{\epsilon} = 0$. • For $a_1 < 1$, the more the player investment is high the more he/she has chance to stay in state 1 but has a higher cost (of investment).
• Now assume that $a_1 = 0$. **Putting no effort is good for today but not the future.**
For $a_1 = 1$, the payoff is zero and the state will move to 1 the next slot.

Absorption

For every strategy σ, the state 0 is an absorbing state.

One has, $\mathbb{P}\left(\sum_{t\geq 0} r(x_t, a_t, m_t) = 0 \mid x_0 = 0\right) = 1$. Thus, the expectation

$$\mathbb{E}_{\sigma,x,m}\left[g\left(\sum_{t\geq 0} r(x_t, a_t, m_t)\right) \mid x_0 = 0\right] = g(0).$$

and payoff under σ is zero if the starting state is 0. Hence, $v_{\mu}^{*}(0, m) = 0$, for any m.

• the analysis reduces to the events until absorption i.e the exit time from state 1.

• A pure stationary strategy consists to specify the action to be played in state 1 (because $\mathcal{A}(0) = \{0\}$). let π defined by $\pi(0) = 0, \ \pi(1) = a_1$

Suboptimality

Suppose $\tilde{\epsilon} = 0$,

- The payoff is monotone in a_1.
- There is an optimal payoff. The optimal payoff is $\mu \log 2$.
- There is no stationary strategy that is best response to mean field.

Hints for proof

We first compute the payoff $F_{\infty,\mu}(\pi, x, m)$.

• the state 0 is absorbing under the strategy π with $a_1 = 1$ and the payoff is zero.so $F_\mu(1, m) = 0$.

• Now assume that $a_1 < 1$. Then,

$$\mathbb{E}\left(e^{\frac{1}{\mu} r(1,a_1,m)}\right) = a_1 e^{(1-a_1)/2} \leq a_1 e^{(1-a_1)/2} < 1, \;\; a_1 \in (0,1)$$

Let τ_π be the exit time from state 1 starting from 1.

$$\tau_\pi = \inf\{t,\ x_t = 0,\ a_1 < 1\}$$

τ_π is a random variable and

$$\mathbb{P}(\tau_\pi = l\ |x_0 = 1) = a_1^{l-1}(1 - a_1),\ \ l \geq 1.$$

We now use the identity
$g(F_{\infty,\mu}(\pi, 1, m)) = \mathbb{E}_\pi\left(g(\sum_{t=0}^{\tau_\pi} r_t) \mid x_0 = 1\right)$
$= \mathbb{E}_\pi\left(g(\tau_\pi r(.)) \mid x_0 = 1\right) = \sum_{l\geq 1} e^{l\frac{1}{\mu}r(.)} a_1^{l-1}(1 - a_1).$

• $F_{\mu}(\pi, 1, m) = \mu \log\left(\frac{(1-a_1)e^{r(.)/\mu}}{1-a_1 e^{r(.)/\mu}}\right)$. monotone in a_1.
• The limit when $a_1 \longrightarrow 1$ goes to $\mu \log 2$.
• comparison rule: $v'(1) = \mu \log 2 \geq 0$ and $v'(0) = 0$. It is clear that $v'(1) \geq F_{\infty,\mu}(\pi, 1, m)$.
Since g is increasing, $g(v'(1)) \geq g(F_{\infty,\mu}(\pi, 1, m))$. For any $a_1 < 1$, one has clearly

$$g(v'(0)) \leq e^{\frac{r(0,m)}{\mu}} \left(q_{000} g(v'(0)) + q_{001} g(v'(1))\right),$$

because $q_{000} = 1$, $q_{001} = 0$.

Similarly, for $a_1 = 1$ one has

$$g(v'(1)) = e^{\frac{r(1,a_1,m)}{\mu}} \left(q_{1a_11}g(v'(1)) + q_{1a_10}g(v'(0))\right),$$

because $r(1,1,m) = 0$ and for $a_1 < 1$ we want to show that

$$g(v'(1)) > e^{\frac{r(1,a_1,m)}{\mu}} \left(q_{1a_11}g(v'(1)) + q_{1a_10}g(v'(0))\right),$$

which is $g(v'(1)) > e^{(1-a_1)/2}(a_1 g(v'(1)) + (1-a_1))$ i.e
$f(a_1) = 1 + a_1 - 2e^{\frac{-(1-a_1)}{2}} < 0, \ 0 \leq a_1 < 1$ and $f(1) = 0$.
Since $g(v'(1)) = 2$, f is strictly increasing and the maximum is 0. Hence we can apply the "comparison principle" which gives $v'(1) \geq \sup_{\sigma} F_{\infty,\mu}(\sigma, x, m)$. In the other hand the limit of the payoff under π when a_1 goes to 1 one gets v'.
This completes the proof.

No $0-$optimality

It is important to notice that the strategy π provides an $\epsilon-$*best response to the risk-sensitive mean field* for arbitrary small $\epsilon > 0$. However, there is no $0-$best response to mean field.

The result extends to small $\tilde{\epsilon} \neq 0$.

Concluding remarks

- The risk-sensitive criterion captures the subpopulation sensitivity to the risk
- Extension: general state space, action space, non-unichain, non H0-2?
- Switching regime, fractional mean field games, non-mean field games etc

Ongoing works

- H. Tembine, J. Y. Le Boudec, R. El Azouzi, E. Altman, Mean field asymptotic of markov decision evolutionary games and teams. in the Proc. of GameNets, May 2009.
- H. Tembine, Large-scale games in large-scale systems, in the Proc. of Valuetools, 2011.
- H. Tembine, Risk-sensitive mean field stochastic games, Workshop mean field games and related topics, Rome, May 12-13, 2011.

Thank you !