Finite state approximation for continuous-time Markov games with ergodic payoffs [1]

Ismael Hernández-Noriega
Mathematics Department
CINVESTAV-IPN

COLOQUIO DE SISTEMAS ESTOCÁSTICOS 2011

October 2011

[1] Joint work with T. Prieto-Rumeau and O. Hernández-Lerma.

- Markov games are a class of dynamic stochastic games where the state process evolves in time as a Markov process. If time parameter evolves in an interval, then we have a continuous-time Markov game.
- Continuous-time Markov games have been widely analyzed in the literature. However, the majority of these results only show the existence of Nash equilibria and the corresponding optimal gains without a clear at all way to compute them.
- A finite state approximation approach is proposed because of its computational viability.
- The motivation for studying this approach come from its potential and computable application to many real-world problems, as in the case of queueing systems, telecommunication networks, and population systems with catastrophes.

Markov game models

$$M_n := \{S_n, (A(i), B(i), i \in S_n), q_n(j|i, a, b), r_{k,n}(i, a, b), k = 1, 2\}, n \geq 0.$$

- **State space** $S_0 = \{0, 1, 2, \ldots\}$ and $S_n = \{0, 1, \ldots, n\}$.
- **Admissible control sets** $A(i)$ y $B(i)$ for P_1 and P_2, respectively.
- **Transition rates** $q_n(j|i, a, b) \geq 0$ if $i \neq j$.
- **Reward rates** $r_{k,n}(i, a, b)$ for P_k.

Notation

Notation. If X is a complete and separable metric space, its Borel $\sigma-$algebra is denoted by $\mathcal{B}(X)$, while $\mathcal{P}(X)$ stands for the space of probability measures on $\mathcal{B}(X)$ endowed with the topology of weak convergence.

Strategies

Let $n \geq 0$.

- A **randomized Markov strategy** for P_1 is a family $\pi^1 = \{\pi^1_t, t \geq 0\}$ of stochastic kernels satisfying:
  - (i) for each $t \geq 0$ and $i \in S_n$, $\pi^1_t(\cdot|i)$ is a probability measure on $A(i)$ such that $\pi^1_t(A(i)|i) = 1$, and
  - (ii) for each $D \in \mathcal{B}(A)$ and $i \in S_n$, the function $t \mapsto \pi^1_t(D|i)$ is Borel measurable in $t \geq 0$.
- For each $n \geq 0$, let $\Pi^m_{k,n}$ be the set of all randomized Markov strategies for P_k, $k = 1, 2$.
- A strategy $\pi^1 = (\pi^1_t) \in \Pi^m_{k,n}$ is called **stationary** there is a probability measure $\pi^1(\cdot|i) \in \mathcal{P}(A(i))$ such that $\pi^1_t(\cdot|i) = \pi^1(\cdot|i)$ for all $i \in S_n$ and $t \geq 0$.
- For each $n \geq 0$, let $\Pi^s_{k,n} \subset \Pi^m_{k,n}$ be the set of all stationary strategies for P_k.

Notation

Given a pair $(\pi_t^1, \pi_t^2) \in \Pi_{1,n}^m \times \Pi_{2,n}^m$:

$$q(j|i, \pi_t^1, \pi_t^2) := \int_{B(i)} \int_{A(i)} q(j|i, a, b) \pi_t^1(da|i) \pi_t^2(db|i). \tag{1}$$

$$r(i, \pi_t^1, \pi_t^2) := \int_{B(i)} \int_{A(i)} r(i, a, b) \pi_t^1(da|i) \pi_t^2(db|i). \tag{2}$$

In particular, for stationary strategies $(\pi^1, \pi^2) \in \Pi_{1,n}^s \times \Pi_{2,n}^s$ we write (1) and (2) as $q(j|i, \pi^1, \pi^2)$ and $r(i, \pi^1, \pi^2)$, respectively

Ergodic payoffs

- Pathwise average payoff (PAP)

$$J^{p}_{k,n}(i,\pi^1,\pi^2) := \limsup_{T\to\infty} \frac{1}{T}\int_0^T r_{k,n}(x(t),\pi^1,\pi^2)dt.$$

- Expected average payoff (EAP)

$$J^{e}_{k,n}(i,\pi^1,\pi^2) := \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_i^{\pi^1,\pi^2}[r_{k,n}(x(t),\pi^1,\pi^2)]dt.$$

Nash equilibrium

For each $n \geq 0$, a pair of strategies $(\pi^{*1}, \pi^{*2}) \in \Pi^m_{1,n} \times \Pi^m_{2,n}$ is called a Nash (or noncooperative) equilibrium for the PAP criterion if, for all $i \in S_n$ and $(\pi^1, \pi^2) \in \Pi^m_{1,n} \times \Pi^m_{2,n}$,

$$J^p_{1,n}(i, \pi^{*1}, \pi^{*2}) \geq J^p_{1,n}(i, \pi^1, \pi^{*2}) \qquad P_i^{\pi^1, \pi^{*2}} \; a.s., \tag{3}$$

and

$$J^p_{2,n}(i, \pi^{*1}, \pi^{*2}) \geq J^p_{2,n}(i, \pi^{*1}, \pi^2) \qquad P_i^{\pi^{*1}, \pi^2} \; a.s. \tag{4}$$

We can see that in a Nash equilibrium a player cannot get a higher payoff if he/she changes his/her strategy unilaterally.

Theorem 1

Under first and second order Lyapunov conditions, continuity-compactness conditions and irreducibility conditions:

(a) For each $n \geq 0$, there exists a pair of constants $g^*_{1,n}$, $g^*_{2,n}$, a pair of functions $u^*_{1,n}, u^*_{2,n} \in \mathbb{B}(S_n)_w$, and a pair $(\pi^{*1}_n, \pi^{*2}_n) \in \Pi^s_{1,n} \times \Pi^s_{2,n}$ satisfying that, for every $i \in S_n$,

$$g^*_{1,n} = \sup_{\pi^1 \in \Pi^s_1} \left\{ r_{1,n}(i, \pi^1, \pi^{*2}) + \sum_{j \in S} q_n(j|i, \pi^1, \pi^{*2}) u^*_{1,n}(j) \right\}, \qquad (5)$$

$$g^*_{2,n} = \sup_{\pi^2 \in \Pi^s_2} \left\{ r_{2,n}(i, \pi^{*1}, \pi^2) + \sum_{j \in S} q_n(j|i, \pi^{*1}, \pi^2) u^*_{2,n}(j) \right\}. \qquad (6)$$

(b) The pair $(\pi^{*1}, \pi^{*2}) \in \Pi^s_{1,n} \times \Pi^s_{2,n}$ in (a) is a Nash equilibrium for the PAP criterion.

(c) A pair (π^{*1}, π^{*2}) is a Nash equilibrium if and only if (5) and (6) are satisfied.

Definición

Given the original model M_0 and a sequence of game models $\{M_n\}_{n\geq 1}$, we say that $\{M_n\}_{n\geq 1}$ converges to M_0 as $n \to \infty$ if for each fixed $i, j \in S$, the functions $r_{k,n}(i,\cdot,\cdot)$ and $q_n(j|i,\cdot,\cdot)$, for $n \geq i, j$ y $k = 1, 2$, respectively converge to $r_{k,0}(i,\cdot,\cdot)$ and $q_0(j|i,\cdot,\cdot)$ uniformly on $A(i) \times B(i)$.

Theorem 2

We will suppose that $\{M_n\}_{n\geq 1}$ converges to M_0. Under the same assumptions from Theorem 1:

(a) For $k = 1, 2$, the optimal gains $g^*_{k,n}$ of M_n converge to the optimal gains $g^*_{k,0}$ of M_0 as $n \to \infty$, and

(b) If $(\pi_n^{*1}, \pi_n^{*2}) \in \Pi^s_{1,n} \times \Pi^s_{2,n}$ is a Nash equilibrium for the game model M_n, with $n \geq 1$, then any limiting strategy of $\{\pi_n^{*1}, \pi_n^{*2}\}_{n\geq 1}$ is a Nash equilibrium for M_0.

Proof Part (a). For $n \geq 0$ y $k = 1, 2$, let $(\pi^1, \pi^2) \in \Pi^s_{1,n} \times \Pi^s_{2,n}$ fixed.

- Let $\mu_n^{\pi^1,\pi^2}$ be the unique invariant probability measure of the process $\{x_n(t)\}_{t \geq 0}$.
- $$J^e_{k,n}(i, \pi^1, \pi^2) = \sum_{j \in S_n} r_{k,n}(j, \pi^1, \pi^2)\mu_n^{\pi^1,\pi^2}(j) =: g_{k,n}(\pi^1, \pi^2).$$
  $$J^p_{k,n}(i, \pi^1, \pi^2) = g_{k,n}(\pi^1, \pi^2) \quad P_i^{\pi^1,\pi^2} - a.s.$$
- The bias function
  $$u_{k,n}^{\pi^1,\pi^2}(i) := \int_0^\infty [E_i^{\pi^1,\pi^2} r_{k,n}(x(t), \pi^1, \pi^2) - g_{k,n}(\pi^1, \pi^2)]dt.$$
- The Poisson equation
  $$g_{k,n}(\pi^1, \pi^2) = r_{k,n}(i, \pi^1, \pi^2) + \sum_{j \in S_n} q_n(j|i, \pi^1, \pi^2)u_{k,n}^{\pi^1,\pi^2}(j) \quad \forall i \in S_n.$$

- $$\begin{aligned} &g_{k,0}(\pi^1,\pi^2) - g_{k,n}(\pi^1,\pi^2) \\ &\quad = \sum_{i\in S_n} (r_{k,0}(i,\pi^1,\pi^2) - r_{k,n}(i,\pi^1,\pi^2))\mu_n^{\pi^1,\pi^2}(i) \\ &\quad + \sum_{i\in S_n}\left[\sum_{j\in S_0} q_0(j|i,\pi^1,\pi^2)u_{k,0}^{\pi^1,\pi^2}(j) - \sum_{j\in S_n} q_n(j|i,\pi^1,\pi^2)u_{k,n}^{\pi^1,\pi^2}(j)\right] \\ &\quad \times \mu_n^{\pi^1,\pi^2}(i) \end{aligned}$$

- Given $\epsilon > 0$, there exists $N > 0$ such that

$$\sup_{(\pi^1,\pi^2)\in\Pi_{1,n}^s\times\Pi_{2,n}^s} |g_{k,0}(\pi^1,\pi^2) - g_{k,n}(\pi^1,\pi^2)| < \epsilon \quad \forall\, n \geq N.$$

Part (b) follows from Part (a) and the continuity condition on reward functions $g_{k,n}$.

A two-player population system with catastrophes

- **Applications:** Infectious diseases, epidemics, queues, birth-and-death processes with downward jumps.
- **Population size** $i \in S_0 = \{0, 1, 2, \ldots\}$
- **Birth** rate $\lambda > 0$. **Death** rate $\mu > 0$.
- The **immigration set** $A(i)$ for player 1.
- An immigration occurs at a rate $a \in A(i)$.
- The **catastrophe set** $B(i)$ for player 2.
- A catastrophe occurs at a rate $h(i, b) \geq 0$, with $b \in B(i)$.
- The **probability distribution of the perished individuals in the catastrophe** $\{\rho_i(j)\}_{1 \leq j \leq i}$.
- The **immigration set at state** $\mathbf{i = 0}$ is $A(0) = [a_1, a_2]$ with $0 < a_1 < a_2$, and the corresponding transition rates are

$$q_0(1|0, a, b) := -q_0(0|0, a, b) := a \quad \text{for all } a \in A(0).$$

- The **immigration** and **catastrophe sets at state** $\mathbf{i \geq 1}$ are $A(i) = [0, a_2]$ and $B(i) = [b_1, b_2]$ with $0 < b_1 < b_2$, and the corresponding transition rates are

$$q_0(j|i,a,b) := \begin{cases} 0 & \text{if } j > i+1, \\ \lambda i + a & \text{if } j = i+1, \\ -(\mu+\lambda)i - a - h(i,b) & \text{if } j = i, \\ \mu i + h(i,b)\rho_i(1) & \text{if } j = i-1, \\ h(i,b)\rho_i(i-j) & \text{if } 0 \leq j < i-1. \end{cases}$$

- The **net reward for player 1**:

$$r_1(i,a,b) = p_1 i - c_1(i,a,b),$$

  - $p_1 > 0$ is a fixed reward fee per individual in the population,
  - $c_1(i,a,b)$ is the cost for controlling the immigration and the control taken by player 2.

- The **net reward for player 2**:

$$r_2(i, a, b) = p_2 j - c_2(i, a, b),$$

  - j is the number of perished individuals,
  - $p_2 > 0$ is a fixed reward fee,
  - $c_2(i, a, b)$ is the cost for controlling the catastrophe and the control taken by player 1.

Proposition 1

Under suitable assumptions, there exists a Nash equilibrium for the above population system.

For each $n \geq 1$ and $k = 1, 2$, consider the game model M_n with

$$r_{k,n}(i, a, b) := r_{k,0}(i, a, b) \text{ for } i \in S_n, \text{ and } (a, b) \in A(i) \times B(i),$$

and

$$q_n(j|i, a, b) := q_0(j|i, a, b) \text{ for } 0 \leq j < n,$$

$$q_n(n|i, a, b) := \sum_{j \geq n} q_0(j|i, a, b).$$

Truncated game model

Under suitable conditions, the above truncated game model M_n converges the original game model M_0.

1.- A convergence order for $|g^*_{k,n} - g^*_{k,0}|$, $k = 1, 2$.

2.- Computable algorithm.

3.- Numerical approximation.

References

1. Guo, X. P., Hernández-Lerma, O., *Zero-sum games for continuous-time Markov processes in Polish spaces: discounted payoffs*, Adv. Appl. Prob. 39, (2007), pp. 645-668.
2. Guo, X. P., Hernández-Lerma, O., *New optimality conditions for average-payoff continuous-time Markov games in polish spaces*, Sci. China Math., 54, (2011), doi: 10.1007/s11425-011-4186-9
3. Hernández-Noriega, I., Hernández-Lerma, O., *Existence of Nash equilibria for continuous-time Markov games with ergodic payoffs*, (2011), Submitted for publication.
4. Prieto-Rumeau, T., Hernández-Lerma, O., *Bias and overtaking equilibria for zero-sum continuous-time Markov games*, Math. Methods Oper. Res., 61, (2005), pp. 437-454
5. Prieto-Rumeau, T., Lorenzo, J. M., *Approximating ergodic average reward continuous-time controlled Markov chains*, IEEE Trans. Automat. Control 55, (2010), pp. 201-207.