The limit game approach for stochastic games

Miquel Oliu-Barton

PSL, Université Paris Dauphine, Ceremade

French Symposium on Game Theory

May 27, 2015

Université Paris 7

Stochastic games

Introduced by Shapley 53, two-person zero-sum stochastic games are described by a 5-tuple $\Gamma = (S, I, J, g, q)$ where

- S is a finite set of states
- I and J are finite sets of actions for player 1 and 2 respectively
- $g : S \times I \times J \to \mathbb{R}$ is the stage-payoff function
- $q : S \times I \times J \to \Delta(S)$ is a transition function

The game is played as follows:

At stage $m \geq 1$, knowing the current state s_m

- The players choose actions $(i_m, j_m) \in I \times J$
- A stage-payoff $g_m := g(s_m, i_m, j_m)$ is recorded
- A new state s_{m+1} is chosen according to $q(\cdot \,|s_m, i_m, j_m)$

Example

a_{11}	a_{12}
a_{21}	a_{22}

Example

A

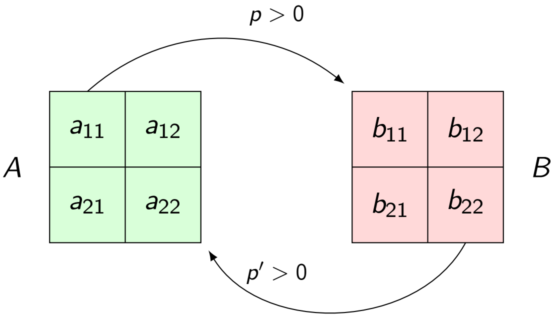
a_{11}	a_{12}
a_{21}	a_{22}

b_{11}	b_{12}
b_{21}	b_{22}

B

Example

A stochastic game with two states and two actions



The discounted game Γ_λ

- The payoff in the λ-discounted game Γ_λ is the geometric mean of the stage payoffs $(g_m)_{m\geq 1}$, where $g_m = g(s_m, i_m, j_m)$
- A stationary strategy is a mapping from S to mixed actions
- For any pair of stationary strategies $(x, y) \in \Delta(I)^S \times \Delta(J)^S$ let $\gamma_\lambda = \gamma_\lambda(\cdot\,, x, y) := \left(\mathbb{E}^s_{x,y}\left[\sum_{m\geq 1} \lambda(1-\lambda)^{m-1} g_m\right]\right)_{s\in S}$ be the vector of expected payoffs. Then

$$\gamma_\lambda = \lambda g + (1-\lambda) Q \gamma_\lambda$$

where $Q = (q(s'|s, x(s), y(s))_{s,s'}$ and $g = (g(s, x(s), y(s))_s$

- **Shapley 53**. The game Γ_λ has a value $v_\lambda \in \mathbb{R}^S$ and the players have optimal stationary strategies

The general evaluation game Γ_θ

- More generally, let $\theta \in \Delta(\mathbb{N}^*)$. The payoff in Γ_θ in given by $\sum_{m \geq 1} \theta_m g_m$
- A Markovian strategy is a strategy that depends, at stage any $m \geq 1$, only of the current stage and state (m, s_m)
- The λ-discounted game Γ_λ corresponds to the evaluation $\theta_m = \lambda(1-\lambda)^{m-1}$ for all $m \geq 1$
- The n-stage game Γ_n corresponds to the evaluation $\theta_m = \mathbb{1}_{m \leq n}$ for all $m \geq 1$
- **Similarly**, the game Γ_θ has a value $v_\theta \in \mathbb{R}^S$ and the players have optimal Markovian strategies

The asymptotic value

Theorem (Mertens and Neyman 81, Mertens, Sorin and Zamir 94)

Let $(\theta^n)_n$ be a sequence of decreasing evaluations (i.e. $\theta^n_m \geq \theta^n_{m+1}$ for all $n, m \geq 1$) such that $\lim_{n\to\infty} \theta^n_1 = 0$. Then $v := \lim_{n\to\infty} v_{\theta^n}$ exists and is independent of the sequence $(\theta^n)_n$.

The asymptotic value

Theorem (Mertens and Neyman 81, Mertens, Sorin and Zamir 94)

Let $(\theta^n)_n$ be a sequence of decreasing evaluations (i.e. $\theta^n_m \geq \theta^n_{m+1}$ for all $n, m \geq 1$) such that $\lim_{n\to\infty} \theta^n_1 = 0$. Then $v := \lim_{n\to\infty} v_{\theta^n}$ exists and is independent of the sequence $(\theta^n)_n$.

Main open problems

1. Characterize the asymptotic value $v \in \mathbb{R}^S$
   - Constant payoff conjecture (Sorin, Venel, Vigeral 2010)
2. Describe asymptotically optimal strategies
   - General evaluation of the stage payoffs

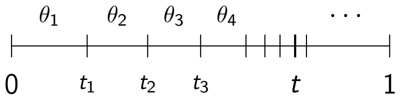
Key observation. The discounted game plays a central role

Fraction $t \in [0,1]$

- For any $t \in [0,1]$ and $\theta \in \Delta(\mathbb{N}^*)$ let

$$\varphi(t,\theta) = \min\left\{ n \geq 1 \,\middle|\, \sum_{m=1}^{n} \theta_m \geq t \right\}$$

- Thus, at stage $\varphi(t,\theta)$ a fraction t of the game has been played



- For any $m \geq 1$, let $t_m = \sum_{k=1}^{m} \theta_k$ and $\theta_m^+ = \frac{\theta_m}{1-t_m}$
- **Remark**. The sequence $(\theta_m^+)_m$ is constant iff θ follows a geometric distribution (i.e. in the λ-discounted case)

Families of stationary strategies

- A family of strategies $(x_\lambda)_{\lambda>0}$ is asymptotically optimal if

$$\liminf_{\lambda\to 0} \gamma_\lambda(s, x_\lambda, j) \geq v(s), \quad \text{for all } j \in J^S \text{ and } s \in S$$

- A family of strategies $(x_\lambda)_{\lambda>0}$ is regular if for all $s \in S$ and $i \in I$ there exists $c(s,i) \geq 0$ and $e(s,i) \geq 0$ such that

$$x^i_\lambda(s) \sim_{\lambda\to 0} c(s,i)\lambda^{e(s,i)}$$

Remark. For regular $(x_\lambda)_\lambda$ and $(y_\lambda)_\lambda$, $\lim_{\lambda\to 0}\gamma_\lambda(s_0, x_\lambda, y_\lambda)$ exists

Families of stationary strategies

- A family of strategies $(x_\lambda)_{\lambda>0}$ is asymptotically optimal if

$$\liminf_{\lambda\to 0} \gamma_\lambda(s, x_\lambda, j) \geq v(s), \quad \text{for all } j \in J^S \text{ and } s \in S$$

- A family of strategies $(x_\lambda)_{\lambda>0}$ is regular if for all $s \in S$ and $i \in I$ there exists $c(s,i) \geq 0$ and $e(s,i) \geq 0$ such that

$$x^i_\lambda(s) \sim_{\lambda\to 0} c(s,i)\lambda^{e(s,i)}$$

Remark. For regular $(x_\lambda)_\lambda$ and $(y_\lambda)_\lambda$, $\lim_{\lambda\to 0}\gamma_\lambda(s_0, x_\lambda, y_\lambda)$ exists

Proposition (Bewley and Kohlberg 76, Solan and Vieille 03, O-B 14)

The players have asymptotically optimal regular strategies.

The limit game $\mathcal{L} = (\mathcal{L}^s)_s$

The limit game $\mathcal{L}^s$ is a one-shot game $(\mathbf{X}, \mathbf{Y}, \mathbf{G}^s)$ played as follows

The limit game $\mathcal{L} = (\mathcal{L}^s)_s$

The limit game $\mathcal{L}^s$ is a one-shot game $(\mathbf{X}, \mathbf{Y}, \mathbf{G}^s)$ played as follows

- The players choose a coefficient-exponent pair for each state and action $\mathbf{x} := (c(s', i), e(s', i))_{s', i}$ $\quad \mathbf{y} := (c(s', j), e(s', j))_{s', j}$

The limit game $\mathcal{L} = (\mathcal{L}^s)_s$

The limit game $\mathcal{L}^s$ is a one-shot game $(\mathbf{X}, \mathbf{Y}, \mathbf{G}^s)$ played as follows

- The players choose a coefficient-exponent pair for each state and action $\mathbf{x} := (c(s',i), e(s',i))_{s',i}$ $\quad \mathbf{y} := (c(s',j), e(s',j))_{s',j}$
- The payoff is given by $\mathbf{G}^s(\mathbf{x},\mathbf{y}) := \lim_{\lambda\to 0} \gamma_\lambda(s, \mathbf{x}_\lambda, \mathbf{y}_\lambda)$, where $\mathbf{x}_\lambda$ and $\mathbf{y}_\lambda$ are defined for each $\lambda > 0$, (s', i) and (s', j) as

$$\mathbf{x}_\lambda^i(s') \;\propto\; c(s',i)\lambda^{e(s',i)} \quad \text{and} \quad \mathbf{y}_\lambda^j(s') \;\propto\; c(s',j)\lambda^{e(s',j)}$$

The limit game $\mathcal{L} = (\mathcal{L}^s)_s$

The limit game $\mathcal{L}^s$ is a one-shot game $(\mathbf{X}, \mathbf{Y}, \mathbf{G}^s)$ played as follows

- The players choose a coefficient-exponent pair for each state and action $\mathbf{x} := (c(s', i), e(s', i))_{s', i}$ $\quad \mathbf{y} := (c(s', j), e(s', j))_{s', j}$
- The payoff is given by $\mathbf{G}^s(\mathbf{x}, \mathbf{y}) := \lim_{\lambda \to 0} \gamma_\lambda(s, \mathbf{x}_\lambda, \mathbf{y}_\lambda)$, where $\mathbf{x}_\lambda$ and $\mathbf{y}_\lambda$ are defined for each $\lambda > 0$, (s', i) and (s', j) as

$$\mathbf{x}^i_\lambda(s') \ \propto \ c(s', i)\lambda^{e(s', i)} \ \text{ and } \ \mathbf{y}^j_\lambda(s') \ \propto \ c(s', j)\lambda^{e(s', j)}$$

Lemma

The limit game $\mathcal{L}$ has value v and the players have optimal strategies. If $\mathbf{x}$ and $\mathbf{y}$ are optimal then for each $s \in S$

$$v(s) = \min_{j \in J^S} \mathbf{G}^s(\mathbf{x}, j) = \mathbf{G}^s(\mathbf{x}, \mathbf{y}) = \max_{i \in I^S} \mathbf{G}^s(i, \mathbf{y})$$

Some properties

Lemma

There exists an optimal strategy for player 1 in $\mathcal{L}$ such that

- *For all (s,i), either $c(s,i)=0$ or*

  $$c(s,i)>0 \text{ and } e(s,i)\in\left\{0,\frac{1}{N},\ldots,\frac{N-1}{N},1\right\}$$

  where $N \leq |S||I|^{\sqrt{|S||I|}}$.
- *For all s, $\min_{i\in I} e(s,i)=0$ and $\sum_{i\in I} c(s,i)\mathbb{1}_{\{e(s,i)=0\}}=1$.*

On the contrary, one cannot bound $(c(s,i))_{s,i}$ in terms of S, I, J.

Asymptotically optimal strategies in Γ_θ

- Let $\theta \in \Delta(\mathbb{N}^*)$ be any evaluation and let $\mathbf{x} = (c(s,i), e(s,i))_{s,i}$ be a strategy for player 1 in the limit game
- Define a Markovian strategy $\mathbf{x}_\theta$ as follows:
  At stage m, chose $i \in I$ according to $\mathbf{x}_{\theta_m^+}(s_m) \in \Delta(I)$, i.e.

$$\mathbf{x}_\theta(m, s_m)^i \;\propto\; c(s_m, i)\left(\frac{\theta_m}{1 - t_m}\right)^{e(s_m,i)}$$

Remark. The strategy x_θ uses the optimal pairs (c, e) at every stage, but wit different weights.

Conjecture

Let $(\theta^n)_n$ be a sequence of decreasing evaluations such that $\lim_{n\to\infty} \theta_1^n = 0$. If $\mathbf{x}$ is optimal for player 1 in $\mathcal{L}$, then

$$\liminf_{n\to\infty} \gamma_{\theta^n}(s, \mathbf{x}_{\theta^n}, y_n) \geq v(s), \quad \forall (y_n)_n \; \forall s$$

A stochastic game $\Gamma = (S, I, J, g, q)$ is absorbing if there exists $s \in S$ such that $q(s' \mid s', i, j) = 1$ for all (i, j) and $s' \neq s$.

An famous example: the big match

1^*	0^*
0	1

Proposition (O.-B.)

Let Γ be an absorbing game. Let x be optimal for player 1 in $\mathcal{L}$ and let $(\theta^n)_n$ be a sequence of decreasing evaluations such that $\lim_{n\to\infty} \theta_1^n = 0$. Then $(\mathsf{x}_{\theta^n})_n$ is asymptotically optimal

Occupation measures for regular strategies

- Let $(\mathbf{x}, \mathbf{y})$ be a fixed pair of regular strategies (i.e. described by pairs coefficients-exponents).
- Let $(\mathbf{x}_\lambda, \mathbf{y}_\lambda)_\lambda$ be the corresponding regular stationary strategies
- Let $Q_\lambda = (q(s'|s, \mathbf{x}_\lambda, \mathbf{y}_\lambda))_{s,s'}$ and $g_\lambda = (g(s, \mathbf{x}_\lambda, \mathbf{y}_\lambda))_s$
- Let $g_0 = \lim_{\lambda \to 0} g_\lambda$. Then

Occupation measures for regular strategies

- Let $(\mathbf{x}, \mathbf{y})$ be a fixed pair of regular strategies (i.e. described by pairs coefficients-exponents).
- Let $(\mathbf{x}_\lambda, \mathbf{y}_\lambda)_\lambda$ be the corresponding regular stationary strategies
- Let $Q_\lambda = (q(s'|s, \mathbf{x}_\lambda, \mathbf{y}_\lambda))_{s,s'}$ and $g_\lambda = (g(s, \mathbf{x}_\lambda, \mathbf{y}_\lambda))_s$
- Let $g_0 = \lim_{\lambda \to 0} g_\lambda$. Then

$$\begin{aligned}\mathbf{G}(\mathbf{x}, \mathbf{y}) &= \lim_{\lambda \to 0} \gamma_\lambda(\,\cdot\,, \mathbf{x}_\lambda, \mathbf{y}_\lambda) \\ &= \lim_{\lambda \to 0} \left(\lambda(\mathrm{Id} - (1-\lambda)Q_\lambda)^{-1} g_0\right)\end{aligned}$$

- Recall that for each (s, s')

$$\lambda(\mathrm{Id} - (1-\lambda)Q_\lambda)^{-1}(s, s') = \mathbb{E}^s_{\mathbf{x},\mathbf{y}}\left[\sum_{m \geq 1} \lambda(1-\lambda)^{m-1} \mathbb{1}_{\{s_m = s'\}}\right]$$

- Let $\Pi_\infty := \lambda(\mathrm{Id} - (1-\lambda)Q_\lambda)^{-1}$ be this occupation measure

Current position

- Similarly, for any $t \in \mathbb{R}_+$, define the occupation measure

$$\Pi_t := \lim_{\lambda \to 0} \sum_{m=1}^{\lfloor \frac{t}{\lambda} \rfloor} \lambda (1-\lambda)^{m-1} Q_\lambda^{m-1}$$

Current position

- Similarly, for any $t \in \mathbb{R}_+$, define the occupation measure

$$\Pi_t := \lim_{\lambda \to 0} \sum_{m=1}^{\lfloor \frac{t}{\lambda} \rfloor} \lambda(1-\lambda)^{m-1} Q_\lambda^{m-1}$$

- Let $P_t = \frac{d}{dt}\Pi_t \simeq \lim_{\lambda \to 0} Q_\lambda^{\lfloor \frac{t}{\lambda} \rfloor}$. For any $s \in S$, $P_t(s) \in \Delta(S)$ is interpreted as the current position at time t

Current position

- Similarly, for any $t \in \mathbb{R}_+$, define the occupation measure

$$\Pi_t := \lim_{\lambda \to 0} \sum_{m=1}^{\lfloor \frac{t}{\lambda} \rfloor} \lambda (1-\lambda)^{m-1} Q_\lambda^{m-1}$$

- Let $P_t = \frac{d}{dt}\Pi_t \simeq \lim_{\lambda \to 0} Q_\lambda^{\lfloor \frac{t}{\lambda} \rfloor}$. For any $s \in S$, $P_t(s) \in \Delta(S)$ is interpreted as the current position at time t

Proposition (Jaffuel and O.-B.)

There exists disjoint sets $R_1, \ldots, R_L \subset S$ and matrices $\mu \in \mathbb{R}^{S \times L}$, $A \in \mathbb{R}^{L \times L}$ and $M \in \mathbb{R}^{L \times S}$ such that $P_t = \mu e^{At} M$, for all $t \geq 0$

Remark. Our proof is constructive: by considering for some $\delta < \frac{1}{N}$

$$P_t^{e+} := \lim_{\lambda \to 0} Q_\lambda^{\lfloor \frac{t}{\lambda} \rfloor + \lfloor \frac{1}{\lambda^{e+\delta}} \rfloor}, \quad \forall e \in \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}$$

we provide an algorithm which gives $R_1, \ldots, R_L$, μ, A and M.

Remark. $\lim_{\lambda\to 0}\sum_{m=1}^{\lfloor \frac{t}{\lambda}\rfloor + M}\lambda(1-\lambda)^{m-1} = 1-e^{-t}$, $\forall M\in\mathbb{N}$. Thus, the "current position" at time t corresponds to many stages (think of a periodic Markov chain). A time-change provides the position at the fraction $t\in[0,1]$ of the game, i.e. $p_t := P_{-\ln(1-t)}$

Remark. $\lim_{\lambda\to 0}\sum_{m=1}^{\lfloor \frac{t}{\lambda}\rfloor+M}\lambda(1-\lambda)^{m-1}=1-e^{-t}$, $\forall M\in\mathbb{N}$. Thus, the "current position" at time t corresponds to many stages (think of a periodic Markov chain). A time-change provides the position at the fraction $t\in[0,1]$ of the game, i.e. $p_t := P_{-\ln(1-t)}$

Definition. For any initial state $s\in S$, the payoff at time t is defined as $\langle p_t(s), g\rangle$. It follows that

$$\mathbf{G}^s(\mathbf{x},\mathbf{y})=\int_0^1\langle p_t(s), g\rangle dt=\int_0^\infty e^{-t}\langle P_t(s), g\rangle dt$$

Conjecture

Let $(\mathbf{x},\mathbf{y})$ be optimal in the limit game $\mathcal{L}$. Then the current payoff $\langle p_t(s), g\rangle$ is independent of $t\in[0,1]$

Remark. $\lim_{\lambda\to 0}\sum_{m=1}^{\lfloor \frac{t}{\lambda}\rfloor + M}\lambda(1-\lambda)^{m-1} = 1-e^{-t}$, $\forall M\in\mathbb{N}$. Thus, the "current position" at time t corresponds to many stages (think of a periodic Markov chain). A time-change provides the position at the fraction $t\in[0,1]$ of the game, i.e. $p_t := P_{-\ln(1-t)}$

Definition. For any initial state $s\in S$, the payoff at time t is defined as $\langle p_t(s), g\rangle$. It follows that

$$\mathbf{G}^s(\mathbf{x},\mathbf{y}) = \int_0^1 \langle p_t(s), g\rangle dt = \int_0^\infty e^{-t}\langle P_t(s), g\rangle dt$$

Conjecture

Let $(\mathbf{x},\mathbf{y})$ be optimal in the limit game $\mathcal{L}$. Then the current payoff $\langle p_t(s), g\rangle$ is independent of $t\in[0,1]$

Proposition (Sorin and Vigeral, O.-B.)

Let Γ be an absorbing game. Then $\langle p_t(s), g\rangle$ is independent of t

Conclusion

- The limit game captures (and simplifies) the relevant information of the stochastic games Γ_θ, as the weights of all stages are close to 0. Optimal strategies in $\mathcal{L}$ generate ε-optimal strategies in any sufficiently "long" game

Conclusion

- The limit game captures (and simplifies) the relevant information of the stochastic games Γ_θ, as the weights of all stages are close to 0. Optimal strategies in $\mathcal{L}$ generate ε-optimal strategies in any sufficiently "long" game

- The explicit information about the occupation times and current position during the play turned to be helpful in solving some problems for absorbing games. The generalization to general stochastic games is a working project

Moltes gràcies !

Merci pour votre attention