Two models of stochastic games with stage duration

Ivan Novikov

Université Paris-Dauphine, CEREMADE

Table of contents

Zero-sum stochastic games (1)

A zero-sum stochastic game is a 5-tuple (Ω, I, J, g, P), where:

- Ω is a non-empty set of states;
- I is a non-empty set of actions of player 1;
- J is a non-empty set of actions of player 2;
- $g : I \times J \times \Omega \to \mathbb{R}$ is a payoff function of player 1;
- $P : I \times J \times \Omega \to \Delta(\Omega)$ is a transition probability function.

We assume that I, J, Ω are finite.
$\Delta(\Omega) :=$ the set of probability measures on Ω.

Zero-sum stochastic games (2)

A stochastic game (Ω, I, J, g, P) proceeds in stages as follows. At each stage n:

1. The players observe the current state ω_n;
2. Players choose their mixed actions, $x_n \in \Delta(I)$ and $y_n \in \Delta(J)$;
3. Pure actions $i_n \in I$ and $j_n \in J$ are chosen according to $x_n \in \Delta(I)$ and $y_n \in \Delta(J)$;
4. Player 1 obtains a payoff $g_n = g(i_n, j_n, \omega_n)$, while player 2 obtains the payoff $-g_n$;
5. The new state ω_{n+1} is chosen according to the probability law $P(i_n, j_n, \omega_n)$.

The above description of the game is known to the players.

Strategies and total payoff

- Strategies σ, τ of players consist in choosing at each stage a mixed action;
- The players can take into account the previous actions of players, as well as the current and previous states;
- Let $B = \{b_m\}$ be a non-increasing positive sequence. Total payoff: $E^{\omega}_{\sigma,\tau}\left(\sum_{i=1}^{\infty} b_i g_i\right)$;
- Depends on initial state ω and strategies of the players;
- λ-discounted total payoff: take $b_m = \lambda(1-\lambda)^{m-1}$;
- Payoff in the game with horizon N: take $b_m = 1/N$ if $1 \leq m \leq N$ and $b_m = 0$ if $m > N$.

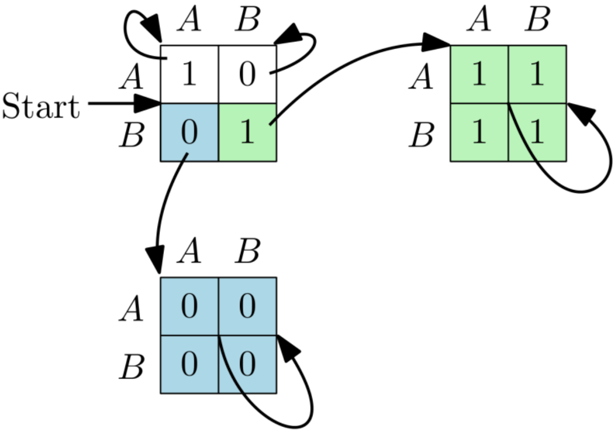
Value

- Value $v_B : \Omega \to \mathbb{R}$:

$$v_B(\omega) = \sup_\sigma \inf_\tau E^\omega_{\sigma,\tau}\left(\sum_{i=1}^\infty b_i g_i\right)$$
$$= \inf_\tau \sup_\sigma E^\omega_{\sigma,\tau}\left(\sum_{i=1}^\infty b_i g_i\right).$$

- The value exists in our finite case.

Example of a game (Big match)



- 3 states, 2 actions for each player.
- Value = 1/2. Player 2's optimal strategy is (1/2, 1/2).
- The figure is from the internet.

Table of contents

Continuous-time Markov games (1)

- Finite state space Ω, action spaces I, J of two players;
- Instantaneous payoff function g.
- Infinitesimal generator of the game $q : I \times J \to \{\text{matrices } |\Omega| \times |\Omega| \text{ satisfying property } *\}$.
- Matrix $A = (a_{ij})$ satisfies property $*$, if $a_{ij} \geq 0$ for all $i \neq j$, $a_{ii} \leq 0$, and $\sum_{j=1}^{|\Omega|} a_{ij} = 0$ for all i.
- If the state at time t is ω' and the players play (i, j) in the interval $[t, t + h]$, then at the time $t + h$ the state is distributed according to $e^{hq(i,j)}(\omega', \cdot)$.

Continuous-time Markov games (2)

- Players choose their (Markov) strategies
  $\sigma : \Omega \times [0, \infty) \to \Delta(I)$, $\tau : \Omega \times [0, \infty) \to \Delta(J)$.
- There are some measurability conditions.
- Initial state is ω_0.
- λ-discounted payoff: $E^{\omega_0}_{\sigma,\tau} \left(\int_0^\infty \lambda e^{-\lambda t} g(i_t, j_t, \omega_t) dt \right)$.
- The value is defined as before.

Stochastic games with stage duration (discounted case)

- We want to approximate a continuous-time game by stochastic games.
- Consider a family of stochastic games G_h, parametrized by $h \in (0, 1]$.
- h represents stage duration.
- Players can play only at times $0, h, 2h, \ldots$
- State can change only at times $h, 2h, \ldots$
- State space Ω and action spaces I and J of player 1 and player 2 are independent of h.

Stochastic games with stage duration (discounted case 2)

- Payoff function g_h of player 1 and transition probability P_h depend on h.

- First model: $P_h(i,j) = e^{hq(i,j)}$.
- Payoff at n-th stage is $\int_{(n-1)h}^{nh} \lambda e^{-\lambda t} g_t dt$.
- Total payoff is $\int_0^{\infty} \lambda e^{-\lambda t} g_t dt$.
- The value is $v^1_{\lambda,h}$.

- Second model: $P_h(i,j) = Id + hq(i,j)$.
- Payoff at n-th stage is $\lambda h(1-\lambda h)^{n-1} g_n$.
- Total payoff is $\lambda h \sum_{k=1}^{\infty} (1-\lambda h)^{k-1} g_k$.
- The value is $v^2_{\lambda,h}$.

- If $h \to 0$ then both $v^1_{\lambda,h}$ and $v^2_{\lambda,h}$ approach the λ-discounted value of the continuous-time game.

Papers about games with stage duration

First model:

- “Limit Value of Dynamic Zero-Sum Games with Vanishing Stage Duration” by Sylvain Sorin (2018);
- “Markov Games with Frequent Actions and Incomplete Information—The Limit Case” by Pierre Cardaliaguet, Catherine Rainer, Dinah Rosenberg, Nicolas Vieille (2016);
- “Continuous-time limit of dynamic games with incomplete information and a more informed player” by Fabien Gensbittel (2016).

Second model:

- “Stochastic games with short-stage duration” by Abraham Neyman (2013);
- “Operator approach to values of stochastic games with varying stage duration” by Sylvain Sorin and Guillaume Vigeral (2016).

Stochastic games with stage duration (general case 1)

- We want to consider more general payoffs.
- Let $k : [0, +\infty) \to \mathbb{R}$ be a nonincreasing continuous positive function with $\int_0^\infty k(t)dt = 1$.

- First model: $P_h(i,j) = e^{hq(i,j)}$.
- Payoff at n-th stage is $\int_{(n-1)h}^{nh} k(t)g_t dt$.
- Total payoff is $\int_0^\infty k(t)g_t dt$.
- The value is $v^1_{k,h}$.
- The limit $\lim_{h\to 0} v^1_{k,h}$ was studied in "Limit Value of Dynamic Zero-Sum Games with Vanishing Stage Duration" by Sylvain Sorin (2018);

Stochastic games with stage duration (general case 2)

- Second model: $P_h(i,j) = Id + hq(i,j)$.
- Payoff at n-th stage is $h \cdot k((n-1)h) \cdot g_n$.
- Total payoff is $\sum_{k=1}^{\infty} hk((n-1)h)g_k = hg(0)g_1 + hg(h)g_2 + \dots$
- The value is $v^2_{k,h}$.
- The limit $\lim_{h\to 0} v^2_{k,h}$ in the discounted case was studied in [1, 2].
- The limit $\lim_{h\to 0} v^2_{k,h}$ in the case of repeated N times game $\left(k(t) = \begin{cases} 1/N, & \text{if } 0 \le t \le N, \\ 0, & \text{otherwise.} \end{cases} \right)$ was studied in [2].
- Our goal is to study this limit in the general case.

[1] “Stochastic games with short-stage duration” by Abraham Neyman (2013);
[2] “Operator approach to values of stochastic games with varying stage duration” by Sylvain Sorin and Guillaume Vigeral (2016).

The limit (1)

Proposition (S. Sorin, 2018)

The limit $\lim_{h\to 0} v^1_{k,h}$ *exists and is a unique viscosity solution of*

$$0 = \frac{d}{dt} v(t,\omega) + \mathtt{Val}_{I\times J}[k(t)g(i,j,\omega) + \langle q(i,j)(\omega,\cdot)\,, v(t,\cdot)\rangle],$$

where

- $\langle f(\cdot), g(\cdot)\rangle = \sum_{x\in X} f(x)g(x)$;
- $\mathtt{Val}_{I\times J}(G)$ *is a value of the one-shot game* G *with action spaces* I, J.

We want to prove an analogous result for $\lim_{h\to 0} v^2_{k,h}$.

The limit (2)

Proposition (I.N.)

The limit $\lim_{h\to 0} v^2_{k,h}$ *exists and is a unique viscosity solution of*

$$0 = \frac{d}{dt}v(t,\omega) + \mathtt{Val}_{I\times J}[k(t)g(i,j,\omega) + \langle q(i,j)(\omega,\cdot)\,, v(t,\cdot)\rangle].$$

Sketch of the proof

- An idea: prove that $\|v^1_{k,h} - v^2_{k,h}\|_\infty \to 0$ as $h \to 0$.
- $v^1_{k,h} =?$ and $v^2_{k,h} =?$
- Define for $n \in \mathbb{N}^*$

$$\psi^h_n : \; C(\Omega, \mathbb{R}) \to C(\Omega, \mathbb{R}),$$
$$f(\omega) \mapsto \mathtt{Val}_{I\times J}[k((n-1)h)hg(i,j,\omega) + \langle (Id + hq(i,j))(\omega,\cdot)\, , f(\cdot)\rangle];$$
$$\overline{\psi}^h_n : \; C(\Omega, \mathbb{R}) \to C(\Omega, \mathbb{R}),$$
$$f(\omega) \mapsto \mathtt{Val}_{I\times J}\left[\int_{(n-1)h}^{nh} k(t)g(i,j,\omega)dt + \langle \exp\{hq(i,j)\}(\omega,\cdot)\, , f(\cdot)\rangle\right].$$

- We can prove $v^1_{k,h} = \prod_{i=1}^\infty \overline{\psi}^h_i(0)$, $v^2_{k,h} = \prod_{i=1}^\infty \psi^h_i(0)$, where we denote $\prod_{i=1}^\infty S_i(z) := \lim_{i\to\infty}(S_1 \circ S_2 \circ \cdots \circ S_i(z))$.
- Afterwards we make some manipulations with Shapley operators to prove our result.

State-Blind Stochastic Games

- Now players cannot observe the current state.
- Players know the initial probability distribution on the states, and they observe the actions of each other.
- Payoffs are <u>not</u> observed.

The limit (3)

Proposition (S.Sorin, 2018)

The limit $\lim_{h\to 0} v^1_{k,h}$ *exists and is a unique viscosity solution of*

$$0 = \frac{d}{dt} v(t,p) + \mathtt{Val}_{I\times J}[k(t) g(i,j,p) + \langle p * q(i,j), \nabla v(t,p)\rangle],$$

where

- $(p * q(i,j))(\omega) = \sum_{\omega' \in \Omega} p(\omega') \cdot q(i,j,\omega')(\omega)$;
- $\langle f(\cdot), g(\cdot)\rangle = \sum_{x\in X} f(x) g(x)$.

We want to prove the same result for $\lim_{h\to 0} v^2_{k,h}$.

The limit (4)

Proposition (I.N.)

The limit $\lim_{h \to 0} v_{k,h}^2$ *exists and is a unique viscosity solution of*

$$0 = \frac{d}{dt} v(t,p) + \mathtt{Val}_{I \times J}[k(t) g(i,j,p) + \langle p * q(i,j), \nabla v(t,p) \rangle].$$

Sketch of the proof

- An idea: prove that $\|v^1_{k,h} - v^2_{k,h}\|_\infty \to 0$ as $h \to 0$;
- A problem: it does not work here;
- Thus, we want to follow the proof from [1];
- We consider the family $\{v^2_{k,h}\}_{h\in(0,1]}$. It can be proven that it is equilipschitz-continuous and equibounded;
- Hence by the Arzelà–Ascoli theorem the limit $\lim_{h\to 0} v^2_{k,h}$ has at least one accumulation point;
- Afterwards we write the Shapley equation to prove that each accumulation point is a viscosity solution of the above differential equation;
- It can be proven that this differential equation has a unique solution.

[1] "Limit Value of Dynamic Zero-Sum Games with Vanishing Stage Duration" by Sylvain Sorin (2018).

Generalization: varying stage duration

- Now we allow different stage durations for different stages;
- $T \in \mathbb{R}_+$, and there is a sequence $\{h_i\}_{i \in \mathbb{N}}$ with $\sum_{i=1}^{\infty} h_i = T$;
- Players act in times $0, h_1, h_1 + h_2, \ldots$;
- We denote $t_1 = 0, t_n = \sum_{i=1}^{n-1} h_n$;
- First model: $P_h(i,j) = e^{h_n q(i,j)}$;
- Payoff at n-th stage is $\int_{t_n}^{t_{n+1}} k(t) g_t dt$;
- Second model: $P_h(i,j) = 1 + h_n q(i,j)$;
- Payoff at n-th stage is $h_n k(t_n) g_n$;
- The analogues of the above propositions hold in this more general model. We suppose now that $\sup h_i \to 0$.

This is all.

Thank you!