Zero-Sum Stochastic Games with Vanishing Stage Duration and Public Signals

Ivan Novikov

Université Paris-Dauphine, CEREMADE

11/06/2024

Table of contents

Zero-sum stochastic games with perfect observation of the state (1)

A zero-sum stochastic game (with perfect observation of the state) 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 with perfect observation of the state (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 payoff $-g_n$;
5. The new state w_{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.
- λ-discounted total payoff: $E^{\omega}_{\sigma,\tau}\left(\lambda \sum_{i=1}^{\infty}(1-\lambda)^{i-1} g_i\right)$;
- Depends on $\lambda \in (0,1)$, initial state ω, and strategies of the players;
- Value $v_\lambda : \Omega \to \mathbb{R}$:

$$
\begin{aligned}
v_\lambda(\omega) &= \sup_{\sigma} \inf_{\tau} E^{\omega}_{\sigma,\tau}\left(\lambda \sum\nolimits_{i=1}^{\infty}(1-\lambda)^{i-1} g_i\right) \\
&= \inf_{\tau} \sup_{\sigma} E^{\omega}_{\sigma,\tau}\left(\lambda \sum\nolimits_{i=1}^{\infty}(1-\lambda)^{i-1} g_i\right).
\end{aligned}
$$

Limit of λ-discounted game Γ^λ

- $v_\lambda(\omega) = \sup_\sigma \inf_\tau E^\omega_{\sigma,\tau}\left(\lambda \sum_{i=1}^\infty (1-\lambda)^{i-1} g_i\right)$;
- One can ask: what happens if players become more and more patient? I.e., players are willing to wait a lot to obtain a big payoff;
- Mathematically, it means that $\lambda \to 0$;
- Thus, one is interested in the uniform (in ω) limit $\lim_{\lambda \to 0} v_\lambda(\omega)$;
- The limit always exists in the finite framework, but may fail to exits in a more general setting.

Table of contents

Kernel

- Kernel $q : I \times J \times \Omega \to \mathbb{R}^{|\Omega|}$.

$$q(i,j,\omega)(\omega') = \begin{cases} P(i,j,\omega)(\omega') & \text{if } \omega \neq \omega'; \\ P(i,j,\omega)(\omega') - 1 & \text{if } \omega = \omega'. \end{cases}$$

- Recall that $P(i,j,\omega)(\omega')$ is the probability that the next state is ω', if the current state is ω and players' actions are (i,j);
- Hence the closer kernel q is to 0, the more probable it is that the next state coincides with the current one.

Stochastic games with stage duration

- Consider a family of stochastic games G_h, parametrized by $h \in (0, 1]$;
- h represents stage duration;
- Players now play at times $0, h, 2h, \ldots$, instead of playing at times $0, 1, 2, \ldots$;
- State space Ω and action spaces I and J of player 1 and player 2 are independent of h;
- Payoff function g_h of player 1 and kernel q_h depend on h.

Stochastic games with stage duration

- Payoff $g_h = hg$;
- Kernel $q_h = hq$;
- $h = 1$: "Usual" stochastic game;
- When h small, g_h is close to zero (players receive almost nothing each turn), and q_h is close to zero (the next state with a high probability will be the same).

Comparison (1)

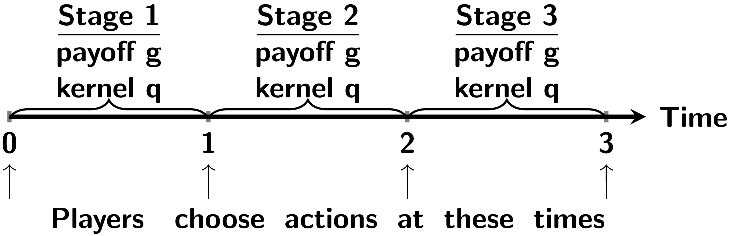


Figure: "Usual" stochastic game: duration of each stage is 1

Comparison (2)

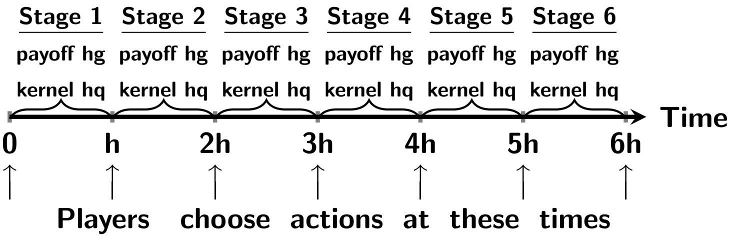


Figure: Stochastic game with stage duration h: stage payoff and kernel are proportional to h

Discounted games with stage duration

- For a game with stage duration h, the total payoff is (depending on the discount factor λ, initial state ω, and strategies σ, τ of players)

$$E^{\omega}_{\sigma,\tau}\left(\lambda \sum_{k=1}^{\infty}(1-\lambda h)^{k-1}(g_k)_h\right);$$

- Why such a choice? Easy explanation:
- The total payoff is λ-discounted game with stage duration 1 is $E^{\omega}_{\sigma,\tau}\left(\lambda \sum_{k=1}^{\infty}(1-\lambda)^{k-1}g_k\right)$. The total payoff of λ-discounted game with stage duration h is $E^{\omega}_{\sigma,\tau}\left(\sum_{k=1}^{\infty}\lambda h(1-\lambda h)^{k-1}g_k\right)$;
- So, it may be seen as a game with discount factor λh. I.e., the discount factor is proportional to h, just as the payoff g and the kernel q.

Real meaning behind the total payoff of the game with stage duration h

- Total payoff: $E^{\omega}_{\sigma,\tau}\left(\lambda\sum_{k=1}^{\infty}(1-\lambda h)^{k-1}(hg_k)\right)$;
- When h is small, the total payoff of the λ-discounted stochastic game with stage duration h is close to the total payoff of the analogous λ-discounted continuous-time game;
- In a continuous-time game, players can choose actions at any time, and at each time t they receive instantaneous payoff g_t. The total payoff is (depending on the discount factor λ) $\int_0^{\infty}\lambda e^{-\lambda t}g_t dt$.

Papers about games with stage duration

- “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).

Discounted games with stage duration (main properties)

- We denote by $v_{h,\lambda}$ the value of the game with total payoff $E^{\omega}_{\sigma,\tau}\left(\lambda \sum_{k=1}^{\infty}(1-\lambda h)^{k-1}(g_k)_h\right)$;
- Main question: What happens with $v_{h,\lambda}$ when $h \to 0$?

Proposition (A. Neyman)

$\lim_{h\to 0} v_{h,\lambda}$ *exists and is a unique solution of a functional equation.*

Proposition (S. Sorin, G. Vigeral)

$\lim_{\lambda\to 0}\lim_{h\to 0} v_{h,\lambda}$ *exists if and only if* $\lim_{\lambda\to 0} v_{1,\lambda}$ *exists, and in the case of existence we have* $\lim_{\lambda\to 0}\lim_{h\to 0} v_{h,\lambda} = \lim_{\lambda\to 0} v_{1,\lambda}$.

- $\lim_{\lambda\to 0} v_{1,\lambda}$ should be considered as the limit value of the discrete-time stochastic game, whereas $\lim_{\lambda\to 0}\lim_{h\to 0} v_{h,\lambda}$ should be considered as the limit value of analogous continuous-time game.

Table of contents

Stochastic Games with Public Signals (1)

- Now players cannot perfectly obseserve the current state;
- Players know the initial probability distribution on the states and some information about the current state.

Stochastic Games with Public Signals (2)

A *zero-sum stochastic game with public signals* is a 7-tuple $(A, \Omega, f, I, J, g, P)$, where:

- A is a non-empty set of signals;
- Ω is a non-empty set of states;
- $f : \Omega \to A$ is a partition of Ω;
- 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 stage payoff function of player 1;
- $P : I \times J \times \Omega \to \Delta(\Omega)$ is the transition probability function.

We assume that I, J, Ω, A are finite.

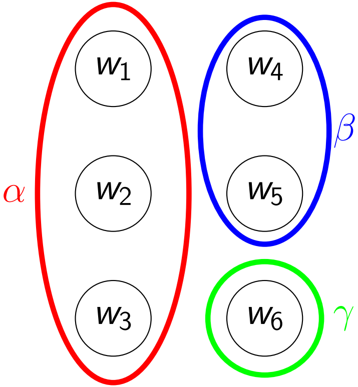
Stochastic Games with Public Signals (3)

The game $(A, \Omega, f, I, J, g, P)$ proceeds in stages as follows. At each stage n:

1. The current state is ω_n. Players do not observe it, but they observe the signal $\alpha_n = f(\omega_n) \in A$ and the actions of each other at the previous stage;
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 payoff $-g_n$;
5. The new state w_{n+1} is chosen according to the probability law $P(i_n, j_n, \omega_n)$. The new signal is $\alpha_{n+1} = f(\omega_{n+1})$.

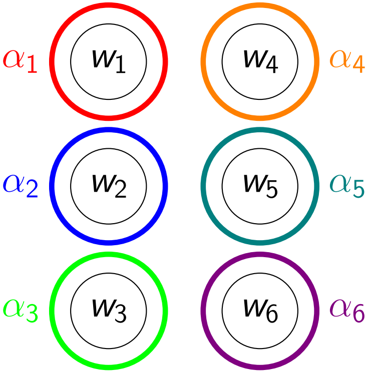
The above description of the game is known to the players. Players do not observe the payoff.

An example of the partition function f (1)



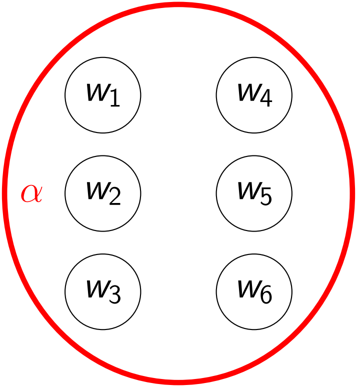
There are 3 public signals, and $f(w_1) = f(w_2) = f(w_3) = \alpha$, $f(w_4) = f(w_5) = \beta$, $f(w_6) = \gamma$.

Examples of the partition function f (2)



The perfect observation of the state, i.e. there are 6 public signals $\alpha_1, \ldots, \alpha_6$; and $f(w_i) := \alpha_i$.

Examples of the partition function f (3)



The state-blind case. There is only one signal α, and $f(w_i) := \alpha$

Stage duration

- We still can consider games with stage duration h in this new setting;
- Payoff $g_h = hg$;
- Kernel $q_h = hq$;
- State space Ω, signal set A, partition function f, and action spaces I and J of player 1 and player 2 are independent of h;
- The total payoff is still $E^{\omega}_{\sigma,\tau}\left(\lambda \sum_{k=1}^{\infty}(1-\lambda h)^{k-1}(g_k)_h\right)$;
- $v_{h,\lambda}$ is the value of the game with such a total payoff.

An example (stage duration 1)

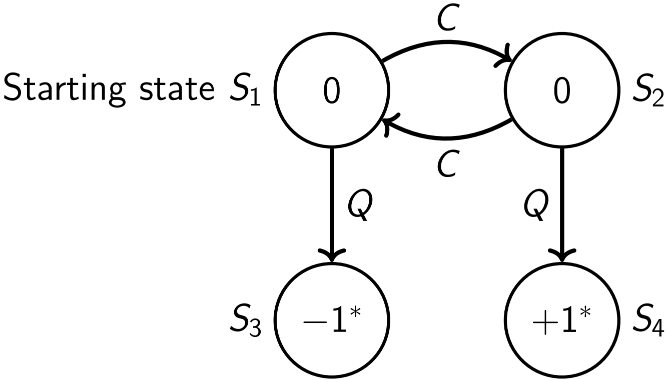


Figure: 1-player game in which each stage has duration 1

- Perfect observation of the state: Play C and later Q.
- State-blind case: the same!

An example (vanishing stage duration)

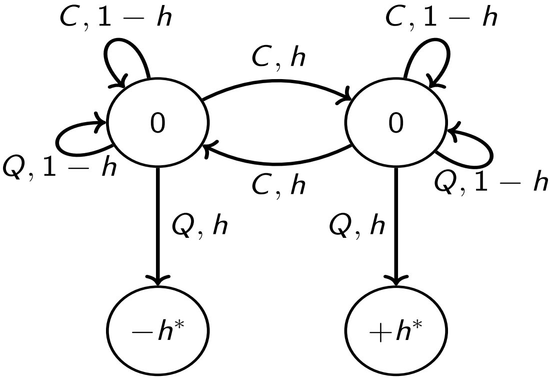


Figure: 1-player game with stage duration h

- Perfect observation of the state: Player will end up in the state S_4. Thus $\lim_{h\to 0} v_{h,\lambda} = \frac{1}{(1+\lambda)^2}$.
- State-blind case: We can prove that the player will play C forever. Thus $\lim_{h\to 0} v_{h,\lambda} = 0$.

First result

Theorem

In the state-blind case, the uniform limit $\lim_{h\to 0} v_{h,\lambda}$ *exists and is a unique viscosity solution of a partial differential equation.*

- The proof is similar to the proof of a similar result in the paper of Sylvain Sorin (2018) "Limit Value of Dynamic Zero-Sum Games with Vanishing Stage Duration".

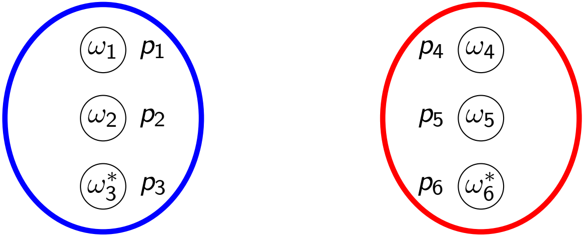
Limit value in games with public signals

- We consider $\lim_{\lambda \to 0} v_{h,\lambda}$;
- Even in finite setting, $\lim_{\lambda \to 0} v_{h,\lambda}$ may not exist;
- First example of inexistence is in the paper of Bruno Ziliotto (2016) "Zero-sum repeated games: Counterexamples to the existence of the asymptotic value and the conjecture maxmin $= \lim v_n$";
- A similar counterexample is in the paper of Bruno Ziliotto and Jérôme Renault (2020) "Hidden stochastic games and limit equilibrium payoffs";
- We now consider a game which is equivalent to the game from the latter paper.

Second result (1)

Theorem

There is a stochastic game G with public signals in which the uniform limit $\lim_{\lambda\to 0}\lim_{h\to 0} v_{h,\lambda}$ exists, but the pointwise limit $\lim_{\lambda\to 0} v_{1,\lambda}$ does not exist.



Signal MINUS

Payoff -1

Player 1's actions: T, B, Q

Player 2's actions: L, R

Signal PLUS

Payoff $+1$

Player 1's actions: T, M, B

Player 2's actions: L, M, R, Q

Second result (2)

The transition matrices for non-absorbing states:

State ω_1:

	L	R
T	ω_1	ω_2
B	ω_2	ω_1
Q	ω_5	ω_5

State ω_2:

	L	R
T	$\frac{1}{2}\omega_1 + \frac{1}{2}\omega_2$	ω_2
B	ω_2	$\frac{1}{2}\omega_1 + \frac{1}{2}\omega_2$
Q	ω_3^*	ω_3^*

State ω_4:

	L	M	R	Q
T	ω_4	ω_5	ω_5	ω_2
M	ω_5	ω_4	ω_5	ω_2
B	ω_5	ω_5	ω_4	ω_2

State ω_5:

	L	M	R	Q
T	$\frac{2}{3}\omega_4 + \frac{1}{3}\omega_5$	ω_5	ω_5	ω_6^*
M	ω_5	$\frac{2}{3}\omega_4 + \frac{1}{3}\omega_5$	ω_5	ω_6^*
B	ω_5	ω_5	$\frac{2}{3}\omega_4 + \frac{1}{3}\omega_5$	ω_6^*

Informal proof (1)

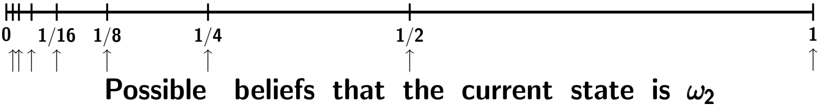


Figure: Discrete case (i.e. stage duration is $h = 1$). Possible beliefs of player 1 that the current state is ω_2 if player 2 plays optimally. As λ becomes smaller, player 1 can wait longer and longer to achieve higher probabilities.

- If the current signal is LEFT, then the smaller is the discount factor λ, the smaller is player 1 can make his belief that the current state is ω_2;
- Analogously, if the current signal is RIGHT, then the smaller is λ, the smaller is player 2 can make his belief that the current state is ω_5;
- Because of that, there is an oscillation when $\lambda \to 0$.

Informal proof (2)

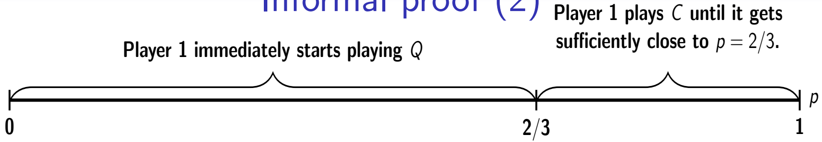
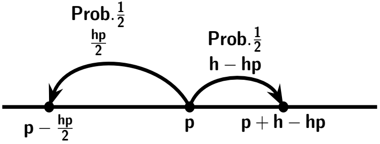
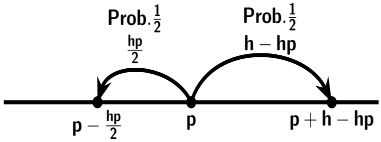


Figure: Continuous case (i.e. $h \approx 0$) with small λ. With prob. $p < 2/3$ that the current state is ω_2, player 1 should immediately start playing Q. Otherwise, his belief $\widetilde{p}$ will start to increase until it becomes $\widetilde{p} = 2/3$, which is bad for player 1. With prob. $p \geq 2/3$ that the current state is ω_2, player 1 can very quickly decrease his belief $\widetilde{p}$ until it becomes $\widetilde{p} \approx 2/3$, which is good for him.



(a) $p > 2/3$ and player 1 plays not Q. $E(\widetilde{p} - p) = \frac{1}{2}(h - hp) + \frac{1}{2} \cdot \frac{-hp}{2} = \frac{h}{4}(2 - 3p) < 0$, thus if λ is small, then player 1 prefers do not play Q until $\widetilde{p}$ is close to $2/3$.



(b) $p < 2/3$ and player 1 plays not Q. $E(\widetilde{p} - p) = \frac{1}{2}(h - hp) + \frac{1}{2} \cdot \frac{-hp}{2} = \frac{h}{4}(2 - 3p) > 0$, thus player 1 prefers to play Q until the state changes.

Informal proof (3)

- Thus very there is a threshold $p = 2/3$ which player 1 cannot cross;
- So, the state is going to get absorbed with prob. $2/3$;
- Similarly, there is a threshold $p = 3/4$ which player 2 cannot cross;
- So, the state is going to get absorbed with prob. $3/4$;
- Thus there is no oscillation as $\lambda \to 0$.

Theorem

There is a stochastic game G with public signals in which the uniform limit $\lim_{\lambda\to 0}\lim_{h\to 0} v_{h,\lambda}$ exists, but the pointwise limit $\lim_{\lambda\to 0} v_{1,\lambda}$ does not exist.

Open question: For the considered above game G, can we say that

1. For any fixed $h \in (0,1]$, the limit $\lim\limits_{\lambda\to 0} v_{h,\lambda}$ does not exist?
2. We have $\left|\limsup\limits_{\lambda\to 0} v_{h,\lambda}(p) - \liminf\limits_{\lambda\to 0} v_{h,\lambda}(p)\right| \to 0$ as $h \to 0$, uniformly in p?

Generalization: varying stage duration

- Now we allow different stage durations for different stages;
- There is a sequence $\{h_i\}_{i\in\mathbb{N}}$;
- Players act in times $h_1, h_1 + h_2, h_1 + h_2 + h_3, \ldots$;
- i-th stage payoff is $h_i g$ and i-th stage kernel is $h_i q$;
- Total payoff is now

$$\lambda \sum_{i=1}^{\infty} \left(\prod_{j=1}^{i-1} (1 - \lambda h_j) \right) h_i g_i.$$

- The analogues of the above theorems hold in this more general model. We suppose now that $\sup h_i \to 0$.

This is all.

Thank you!