# Operator approach to stochastic games with varying stage duration

G.Vigeral (with S. Sorin)

CEREMADE Universite Paris Dauphine

$26$ January 2016,
ADGO II, Santiago de Chile

# Table of contents

# Table of contents

# Zero-sum stochastic game

A zero-sum stochastic game $\Gamma$ is a 5-tuple $(\Omega, I, J, g, \rho)$ where:

- $\Omega$ is the set of states.
- $I$ (resp. $J$) is the action set of Player 1 (resp. Player 2).
- $g : I \times J \times \Omega \to [-1, 1]$ is the payoff function (that Player 1 maximizes and Player 2 minimizes).
- $\rho : I \times J \times \Omega \to \Delta(\Omega)$ is the transition probability.

# How the Game is played

An initial state $\omega_1$ is given, known by each player.
At each stage $k \in \mathbb{N}$:

- the players observe the current state $\omega_k$.
- According to the past history, Player 1 (resp. Player 2) chooses a mixed action $x_k$ in $X = \Delta(I)$ (resp. $y_k$ in $Y = \Delta(J)$). Done independently by each player.
- An action $i_k$ of Player 1 (resp. $j_k$ of Player 2) is drawn according to his mixed strategy $x_k$ (resp. $y_k$).
- This gives the payoff at stage $k$: $g_k = g(i_k, j_k, \omega_k)$.
- A new state $\omega_{k+1}$ is drawn according to $\rho(i_k, j_k, \omega_k)$.

# The $n$-stage game

For any stochastic game $\Gamma$, any finite horizon $n \in \mathbb{N}$, and any starting state $\omega_1$, the $n$-stage game $\Gamma_n$ is the zero-sum game with payoff

$$\mathbb{E}\left\{\sum_{k=1}^{n} g_k\right\},$$

that Player 1 maximizes and Player 2 minimizes.

The value of $\Gamma_n(\omega_1)$ is denoted by $V_n(\omega_1)$.
Normalized value $v_n = \frac{V_n}{n}$.

# The discounted game

For any stochastic game $\Gamma$, any discount factor $\lambda \in ]0,1[$, and any starting state $\omega_1$, the discounted game $\Gamma_\lambda(\omega_1)$ is the zero-sum game with payoff

$$\mathbb{E}\left\{\sum_{k=1}^{+\infty}(1-\lambda)^{k-1}g_k\right\},$$

that Player 1 maximizes and Player 2 minimizes.

The value of $\Gamma_\lambda(\omega_1)$ is denoted by $W_\lambda(\omega_1)$.
Normalized value $w_\lambda = \lambda v_\lambda$.

# Recursive structure

Shapley (1953) proved that the values satisfy a recursive structure:

$$
\begin{aligned}
V_n(\omega) &= \sup_{x\in X}\inf_{y\in Y}\left\{g(x,y,\omega)+E_{\rho(x,y,\omega)}(V_{n-1}(\cdot))\right\} \\
&= \inf_{y\in Y}\sup_{x\in X}\left\{g(x,y,\omega)+E_{\rho(x,y,\omega)}(V_{n-1}(\cdot))\right\} \\
W_\lambda(\omega) &= \sup_{x\in X}\inf_{y\in Y}\left\{g(x,y,\omega)+(1-\lambda)E_{\rho(x,y,\omega)}(W_\lambda(\cdot))\right\} \\
&= \inf_{y\in Y}\sup_{x\in X}\left\{g(x,y,\omega)+(1-\lambda)E_{\rho(x,y,\omega)}(W_\lambda(\cdot))\right\}.
\end{aligned}
$$

# Shapley operator

This can be summarized by:

$$
\begin{aligned}
V_n &= \Psi(V_{n-1}) = \Psi^n(0) \\
W_\lambda &= \Psi((1-\lambda)W_\lambda) \\
w_\lambda &= \lambda\Psi\left(\frac{1-\lambda}{\lambda}w_\lambda\right) = \left(\lambda\Psi\left(\frac{1-\lambda}{\lambda}\cdot\right)\right)^\infty
\end{aligned}
$$

for some operator $\Psi$.

$$
\begin{aligned}
\Psi(f)(\omega) &= \sup_{x\in X}\inf_{y\in Y}\left\{g(x,y,\omega) + E_{\rho(x,y,\omega)}(f(\cdot))\right\} \\
&= \inf_{y\in Y}\sup_{x\in X}\left\{g(x,y,\omega) + E_{\rho(x,y,\omega)}(f(\cdot))\right\}.
\end{aligned}
$$

$\Psi$ is nonexpansive for the infinite norm

$$
\|\Psi(f) - \Psi(f')\|_\infty \leq \|f - f'\|_\infty
$$

.

# Framework

This was proven by Shapley in the finite case but true in a very wide framework.
For example

- if $\Omega$ finite, $X$ and $Y$ compact, $g$ and $\rho$ continuous.
- $\Omega$, $X$ and $Y$ are compact metric, $g$ and $\rho$ continuous.

See Maitra Partasarathy, Nowak, Mertens Sorin Zamir for more general frameworks.

# Table of contents

# Definition

- Definition due to Neyman (2013).
- Instead of playing at time $1, 2, \cdots, n, \cdots$, players play at times $t_1, t_2, \cdots, t_n, \cdots$
- The intensity of both payoff and transition at time $t_k$ is $h_k = t_{k+1} - t_k$
- That is $g_h = hg$ and $\rho_h = (1-h)Id + h\rho$.
- Shapley operator of "exact game" with duration $h$ : $\Psi_h = (1-h)Id + h\Psi$

# Some natural questions

1. What happens, for a fixed horizon $t$ or discount factor $\lambda$, when the duration $h_i$ of each stage vanishes ? Does the value converge, to which limit ?
2. What happens, for a fixed sequence of stage duration $h_i$, when the horizon goes to infinity or the discount factor goes to 0. Does the normalized value converge, to which limit ?
3. What happens when both $\lambda$ (or $\frac{1}{n}$) and $h_i$ go to 0 ?
4. What can be said of optimal strategies in games with varying duration ?

Neyman answers questions 1 3 4 for finite discounted games. Here we use the operator approach to give a general answer to 1 2 3.

# Game with finite horizon and varying duration

- Finite horizon $t$, finite sequence of stage duration $h_1, \cdots, h_n$ with $\sum h_i = t$.
  The value $V$ of such a game satisfies $V = z_n$ with

$$z_{i+1} = \Psi_{h_i}(z_i) = (1 - h_i)z_i + h_i\Psi(z_i)$$

- $\frac{z_{i+1} - z_i}{h_i} = -(Id - \Psi)(z_i)$
- Eulerian scheme associated to $f' = -(Id - \Psi)(f)$.
- One can use general results associated to such schemes, for any non expansive operator defined on a Banach space.

# Eulerian schemes in Banach spaces

For general nonexpansive $\Psi$:

## Proposition (Miyadera-Oharu '70, Crandall-Liggett '71)

$$\|f_{nh}(z_0) - \Psi_h^n(z_0)\| \leq \|z_0 - \Psi(z_0)\| \, h\sqrt{n}.$$

## Proposition (V. '10)

*If* $z_{i+1} = (1-h_i)z_i + h_i\Psi(z_i)$, *then*

$$\|f_t(z_0) - x_n\| \leq \|z_0 - \Psi(z_0)\| \sqrt{\sum_{i=1}^{n} h_i^2}.$$

*with* $t = \sum_{i=1}^{n} h_i$.

# Result with $t$ fixed

- Let $h = \max h_i$ and $t = \sum h_i$, then

$$\|V - f(t)\| \leq K\sqrt{ht}.$$

- Hence as the mesh $h$ goes to 0, the value of the game goes to $f(t)$.
- $f(t)$ can be interpreted as the value of a game played in continuous time (Neyman '13).

# Asymptotic results

- For any $h_i$,
$$\|\frac{V - f(t)}{t}\| \leq \frac{K}{\sqrt{t}}.$$
- All the repeated games with varying stage duration have the same (normalized) asymptotic behavior.
- Same asymptotic behavior for the normalized value in continuous time $\frac{f(t)}{t}$ and for the normalized value of the original game $v_n$.

# Game with discount factor and varying duration

- Discount factor $\lambda$ = weight on the payoff on $[0,1]$ compared to $[0,+\infty]$.
- Infinite sequence of stage durations $h_1,\cdots,h_n,\cdots$.
- When $h$ is constant, normalized value $w_\lambda^h=\lambda\Psi_h\left(\frac{1-\lambda h}{\lambda}\right)$.
- In general $w$ is

$$\left(\prod_{i=1}^{+\infty} D_\lambda^{h_i}\right)(0)$$

with

$$D_\lambda^h(f)=\lambda\Psi_h\left(\frac{1-\lambda h}{\lambda}f\right).$$

# Result with $\lambda$ fixed and vanishing duration

- For a uniform duration $h$, $w^h_\lambda = w_\mu$ with $\mu = \frac{\lambda}{1+\lambda-\lambda h}$.
- For any $\lambda$ and $h_i \leq h$, the value $w$ of the $\lambda-$discounted game with stage durations $h_i$ satisfies

$$\|w - \hat{w}_\lambda\| \leq Kh$$

  with $\hat{w}_\lambda := w_{\frac{\lambda}{1+\lambda}}$.
- Hence as the mesh $h$ goes to 0, the value of the game goes to $w_{\frac{\lambda}{1+\lambda}}$. Already known when the game is finite (Neyman 2013).
- $\hat{w}_\lambda$ can be interpreted as the value of a game played in continuous time (Neyman '13).

# Asymptotic results

- Assumption: there exists nondecreasing $k:]0,1] \to \mathbb{R}^+$ and $\ell : [0,+\infty] \to \mathbb{R}^+$ with $k(\lambda) = o(\sqrt{\lambda})$ as $\lambda$ goes to 0 and

$$\|D^1_\lambda(z) - D^1_\mu(z)\| \leq k(|\lambda - \mu|)\ell(\|z\|)$$

  for all $(\lambda,\mu) \in ]0,1]^2$ and $z \in Z$.
- Always true for Shapley operators of games with bounded payoff.
- Then for any $\lambda$ and $h_i$, the value $w$ of the $\lambda-$discounted game with stage durations $h_i$ satisfies

$$\|w - w_\lambda\| \leq K\lambda.$$

- All the repeated games with varying stage duration have the same (normalized) asymptotic behavior as $\lambda$ goes to 0.
- Same asymptotic behavior for the normalized value in continuous time $\hat{w}_\lambda$ and for the normalized value of the original game $w_\lambda$.

# Table of contents

# Model

- *Finite* state space.
- $\mathsf{P}^t(i,j)$ is a continuous time homogeneous Markov chain on $\Omega$, indexed by $\mathbb{R}^+$, with generator $Q(i,j)$:

$$\dot{\mathsf{P}}^t(i,j) = \mathsf{P}^t(i,j)Q(i,j).$$

- $\overline{G}^h$ is the discretization with mesh $h$ of the game in continuous time $\overline{G}$ where the state variable follows $\mathsf{P}^t$ and is controlled by both players (Zachrisson '64, Tanaka Wakuta '77, Guo Hernadez-Lerma '03, Neyman '12)
- Players act at time $s = kh$ by choosing actions $(i_s, j_s)$ (at random according to some $x_s$, resp. $y_s$), knowing the current state.
- Between time $s$ and $s+h$, state $\omega_t$ evolves with conditional law $\mathsf{P}^t$

# Results

- Shapley operator is

$$\overline{\Psi}_h(f) = \underset{X\times Y}{\mathtt{val}}\{g^h + \mathsf{P}^h \circ f\}$$

  where $g^h(\omega_0, x, y)$ stands for $\mathsf{E}[\int_0^h g(\omega_t; x, y)dt]$ and $\mathsf{P}^h(x,y) = \int_{I\times J} \mathsf{P}^h(i,j)x(di)y(dj)$.
- $\|\overline{\Psi}_h(f) - \Psi_h(f)\| = (1 + \|f\|)O(h^2)$
  where $\Psi$ is the Shapley operator of the (discrete time) stochastic game with payoff $g$ and transition $Id + Q$.
- Hence all the results of previous section involving small $h$ still hold.

# Table of contents

# Conclusion

- We recover and generalize some results of Neyman '13, using only properties of nonexpansive operators.
- Only assumptions are : a) $\Psi$ is well defined and 1-Lipschitz b) the current state is observed.
- Same asymptotic structure of original game, games with varying duration, and game in continuous time.
- Counterexamples of convergence of values with observations of states (V., Ziliotto, Sorin V.) are thus also oscillating with varying duration.

# Open questions

- 1) What if the state is not observed ?
- 2) What happens with a general weight on the payoff (not finite horizon or constant discount factor) ? When $h$ goes to 0, results by Neyman (finite games) and Sorin (using viscosity techniques). But what happens when all the weight goes to infinity ? (analogous to $t$ goes to infinity or $\lambda$ to 0).

# Open problems

$X$ Banach space, $\Psi : X \to X$ 1-Lipschitz.

$$z_{i+1} = \alpha_i z_i + \beta_i \Psi(\gamma_i z_i)$$

with $\alpha_i + \beta_i \gamma_i \leq 1$.
Asymptotic behavior of $z_n$ ?

- With no geometric asumptions on $X$
- no fixed point of $\Psi$.
- with as few assuptions as possible on $\Psi$, and hopefully none.
- Not looking for convergence results, but for comparison between two sequences.

# What I know

Particular case : $\alpha_i = 1 - \beta_i,\ \gamma_i = 1$

$$\begin{aligned} z_{i+1} &= (1-\beta_i)z_i + \beta_i \Psi(z_i) \\ \hat{z}_{i+1} &= (1-\hat{\beta}_i)\hat{z}_i + \hat{\beta}_i \Psi(\hat{z}_i) \end{aligned}$$

Then

$$\|z_n - \hat{z}_m\| \leq \|z_0 - \hat{z}_0\| + C\sqrt{(\sigma_n - \hat{\sigma}_m)^2 + \tau_n + \hat{\tau}_m}$$

where $\sigma_n = \sum_{i=1}^n \beta_i$ and $\tau_n = \sum_{i=1}^n \beta_i^2$
In particular if $\sigma_n = \hat{\sigma}_m$ then

$$\|z_n - \hat{z}_m\| = O(\sqrt{\sigma_n})$$

# What I know (II)

Particular case : $\alpha_i = 0$, $\gamma_i = \frac{1-\beta_i}{\beta_i}$

$$z_{i+1} = \beta_i \Psi(\frac{1-\beta_i}{\beta_i} z_i)$$

With a very mild assumption on $\Psi$, if $\beta_i$ converges slowly to 0, $z_n$ is asymptotically close to the fixed point of $\beta_n \Psi(\frac{1-\beta_n}{\beta_n} \cdot)$

# And you ?

What do you know ?

# Open problems

Up to some renormalization,

$$z_{n+1} = (1-h_n)(1-\lambda_n h_n)z_n + h_n\lambda_n \Psi\left(\frac{1-\lambda_n h_n}{\lambda_n} z_n\right)$$

with $(\lambda_n, h_n) \in [0,1]^2$.
$\lambda_n$ : local discount factor, $h_n$ : local stage length.
Comparison between sequence $z$ and $\hat{z}$ associated to $(h,\lambda)$ and $(\hat{h},\hat{\lambda})$ ? We know when:

- $\lambda_n = \hat{\lambda}_n = \lambda$ fixed
- $\lambda_n = \frac{1}{\sum_1^n h_i}$ and $\hat{\lambda}_n = \frac{1}{\sum_1^n \hat{h}_i}$

General formula for the difference $\|z_n - \hat{z}_m\|$ ?

# Thank you for your attention

Muchas gracias !