Operator approach to stochastic games with varying stage duration

G.Vigeral (with S. Sorin)

CEREMADE Universite Paris Dauphine

4 December 2015,
Stochastic methods in Game theory
Table of contents

1. Zero-sum stochastic games

2. Exact games with varying stage duration
   - Finite horizon
   - Discounted evaluation

3. Discretization of a continuous timed game

4. Conclusion and remarks
Table of contents

1. Zero-sum stochastic games

2. Exact games with varying stage duration
   - Finite horizon
   - Discounted evaluation

3. Discretization of a continuous timed game

4. Conclusion and remarks

G.Vigeral (with S. Sorin)  Operator approach
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).
- $g : I \times J \times \Omega \rightarrow [-1, 1]$ is the payoff function (that Player 1 maximizes and Player 2 minimizes).
- $\rho : I \times J \times \Omega \rightarrow \Delta(\Omega)$ is the transition probability.
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

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

that Player 1 maximizes and Player 2 minimizes.

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

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

that Player 1 maximizes and Player 2 minimizes.

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

G.Vigeral (with S. Sorin) 
Operator approach
Shapley (1953) proved that the values satisfy a recursive structure:

\[
V_n(\omega) = \sup_{x \in X} \inf_{y \in Y} \left\{ g(x, y, \omega) + E_{\rho(x, y, \omega)}(V_{n-1}(\cdot)) \right\}
\]

\[
= \inf_{y \in Y} \sup_{x \in X} \left\{ g(x, y, \omega) + E_{\rho(x, y, \omega)}(V_{n-1}(\cdot)) \right\}
\]

\[
W_\lambda(\omega) = \sup_{x \in X} \inf_{y \in Y} \left\{ g(x, y, \omega) + (1 - \lambda) E_{\rho(x, y, \omega)}(W_\lambda(\cdot)) \right\}
\]

\[
= \inf_{y \in Y} \sup_{x \in X} \left\{ g(x, y, \omega) + (1 - \lambda) E_{\rho(x, y, \omega)}(W_\lambda(\cdot)) \right\}.
\]
This can be summarized by:

\[
V_n = \Psi(V_{n-1}) = \Psi^n(0)
\]

\[
W_{\lambda} = \Psi((1 - \lambda)W_{\lambda})
\]

\[
w_{\lambda} = \lambda \Psi\left(\frac{1 - \lambda}{\lambda} w_{\lambda}\right) = \left(\lambda \Psi\left(\frac{1 - \lambda}{\lambda}\right)\right)^{\infty}
\]

for some operator \(\Psi\).

\[
\Psi(f)(\omega) = \sup_{x \in X} \inf_{y \in Y} \left\{ g(x, y, \omega) + E_{\rho(x,y,\omega)}(f(\cdot)) \right\}
\]

\[
= \inf_{y \in Y} \sup_{x \in X} \left\{ g(x, y, \omega) + E_{\rho(x,y,\omega)}(f(\cdot)) \right\}.
\]

\(\Psi\) is nonexpansive for the infinite norm

\[
\|\Psi(f) - \Psi(f')\|_{\infty} \leq \|f - f'\|_{\infty}
\]

.
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

1. Zero-sum stochastic games

2. Exact games with varying stage duration
   - Finite horizon
   - Discounted evaluation

3. Discretization of a continuous timed game

4. Conclusion and remarks
Definition due to Neyman (2013).

Instead of playing at time 1, 2, ⋯, n, ⋯, players play at times $t_1$, $t_2$, ⋯, $t_n$, ⋯

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 $\rho_h = (1 - h)I + h\rho$.

Shapley operator of "exact game" with duration $h$:

$\Psi_h = (1 - h)I + h\Psi$
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.
Exact games with varying stage duration

Finite horizon

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)**

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

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

15 G.Vigeral (with S. Sorin) Operator approach
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 $h_i$,
  $$\left\| \frac{V - f(t)}{t} \right\| \leq \frac{K}{\sqrt{t}}.$$

- 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 $v_n$. 
Discount factor $\lambda = \text{weight on the payoff on } [0, 1] \text{ compared to } [0, +\infty]$.

- Infinite sequence of stage durations $h_1, \cdots, h_n, \cdots$.
- When $h$ is constant, normalized value $w^h = \lambda \psi_h \left( \frac{1-\lambda h}{\lambda} \right)$.
- In general $w$ is

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

with

$$D^h (f) = \lambda \psi_h \left( \frac{1-\lambda h}{\lambda} f \right).$$
Result with $\lambda$ fixed and vanishing duration

- For a uniform duration $h$, $w^h_\lambda = w_\mu$ with $\mu = \frac{\lambda}{1 + 1 - \lambda h}$.
- 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>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Zero-sum stochastic games</td>
</tr>
<tr>
<td>2</td>
<td>Exact games with varying stage duration</td>
</tr>
<tr>
<td></td>
<td>- Finite horizon</td>
</tr>
<tr>
<td></td>
<td>- Discounted evaluation</td>
</tr>
<tr>
<td>3</td>
<td>Discretization of a continuous timed game</td>
</tr>
<tr>
<td>4</td>
<td>Conclusion and remarks</td>
</tr>
</tbody>
</table>
Discretization of a continuous timed game

Model

- **Finite** state space.
- $P^t(i,j)$ is a continuous time homogeneous Markov chain on $\Omega$, indexed by $\mathbb{R}^+$, with generator $Q(i,j)$:
  \[ \dot{P}^t(i,j) = P^t(i,j)Q(i,j). \]
- $\overline{G}^h$ is the discretization with mesh $h$ of the game in continuous time $\overline{G}$ where the state variable follows $P^t$ 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 $P^t$. 
Results

- Shapley operator is

\[
\bar{\Psi}_h(f) = \text{val}_{X \times Y} \{ g^h + P^h \circ f \}
\]

where \( g^h(\omega_0, x, y) \) stands for \( \mathbb{E}[\int_0^h g(\omega_t; x, y) dt] \) and

\[
P^h(x, y) = \int_{I \times J} P^h(i, j) x(di) y(dj).
\]

- \( \| \bar{\Psi}_h(f) - \Psi_h(f) \| = (1 + \| f \|) O(h^2) \)

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

1. Zero-sum stochastic games

2. Exact games with varying stage duration
   - Finite horizon
   - Discounted evaluation

3. Discretization of a continuous timed game

4. Conclusion and remarks
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

- 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).
- What happens when all the weight goes to infinity (analogous to $t$ goes to infinity or $\lambda$ to 0).
- What if the state is not observed?
**A stupid game**

- Only one player
- He observes (and remember) his moves but not the state.
- Starting state $\omega$.
- Clearly $w_\lambda$ tends to 1 as $\lambda$ goes to 0.
A (not so) stupid game with varying stage duration.

As long as Player 1 plays change the probability of being in the first state satisfies
\[ p_{t+1} = (1 - h)p_t + h(1 - p_t) = (1 - 2h)p_t + h. \]

Hence \( h \leq \frac{1}{2} \), \( w^h_\lambda \) tends to \( \frac{1}{2} \) as \( \lambda \) goes to 0. In fact \( w^h_\lambda \leq \frac{1}{2} \) for any \( h \leq \frac{1}{2} \).
Thank you for your attention

Thank you !