Stochastic game with partial observation and Borel evaluations

Hugo Gimbert (LABRI), Jérôme Renault (GREMAQ-TSE), Sylvain Sorin (IMJ), Xavier Venel (CES), Wieslaw Zielonka (LIAFA)

French symposium on Games, 27 Mai 2015

Outline

1. A first model: no state, perfect information
2. A second example: state variable, perfect observation
3. Repeated games with signals

Introduction: Stochastic games with perfect information

This talk: 2 player 0-sum stochastic games with infinitely many stages.

Model:

- I and J are non empty sets
- W is a fixed subset of $(I \times J)^\infty$.

How the game is played:

- P1 first chooses i_1 in I,
- P2 chooses j_1 in J,
- P1 chooses i_2 in I,
- etc... .

Who wins: A play is a sequence ω in $(I \times J)^\infty$. P1 wins the game if and only if the induced play ω is in W.

Definition

The game is determined if either P1 or P2 has a winning strategy,

$$\text{i.e. if} \quad \exists i_1 \in I, \forall j_1 \in J, \exists i_2 \in I, \forall j_2 \in J, \ldots\ldots, (i_1, j_1, \ldots, i_n, j_n, \ldots) \in W$$
$$\text{or} \quad \forall i_1 \in I, \exists j_1 \in J, \forall i_2 \in I, \exists j_2 \in J, \ldots\ldots, (i_1, j_1, \ldots, i_n, j_n, \ldots) \notin W$$

Example

$I = J = \{0, 1\}$, $W \subset [0, 1]$ closed. (Gale-Stewart, 1953)

Theorem[Martin, 1975]

If W is a Borel subset of $(I \times J)^\infty$, the game is determined.

Remark

Under the axiom of choice, there exists an undetermined game.
Axiom of Determinacy: all games with I and J countable are determined.

Definition

The game is determined if either P1 or P2 has a winning strategy,

$$\text{i.e. if} \quad \exists i_1 \in I, \forall j_1 \in J, \exists i_2 \in I, \forall j_2 \in J, \ldots\ldots, (i_1, j_1, \ldots, i_n, j_n, \ldots) \in W$$
$$\text{or} \quad \forall i_1 \in I, \exists j_1 \in J, \forall i_2 \in I, \exists j_2 \in J, \ldots\ldots, (i_1, j_1, \ldots, i_n, j_n, \ldots) \notin W$$

Example

$I = J = \{0, 1\}$, $W \subset [0, 1]$ closed. (Gale-Stewart, 1953)

Theorem[Martin, 1975]

If W is a Borel subset of $(I \times J)^\infty$, the game is determined.

Remark

Under the axiom of choice, there exists an undetermined game.
Axiom of Determinacy: all games with I and J countable are determined.

Outline

1. A first model: no state, perfect information
2. A second example: state variable, perfect observation
3. Repeated games with signals

Stochastic games with perfect observation

Model:

- I and J are non empty finite sets
- K is a countable set of states,
- $q : K \times I \times J \longrightarrow \Delta(K)$ is a transition
- k_1 a state in K.

How the game is played:

- k_1 is fixed and announced to P1 and P2
- P1 and P2 simultaneously choose i_1 in I and j_1 in J, k_2 is selected according to $q(k_1, i_1, j_1)$
- i_1, j_1 and k_2 are announced.
- etc.....

"Who wins what": let f be a "good" function from $\Omega = (K \times I \times J)^\infty$ to $\mathbb{R}$.For each infinite history $h \in \Omega$,

$$\text{Player 1 wins } f(h) \text{ and Player 2 wins } -f(h).$$

- A strategy for P1 is a rule specifying what he should play in any situation, formally $\sigma = (\sigma_t)_{t\geq 1}$, with $\sigma_t : (I \times J \times K)^{t-1} \longrightarrow \Delta(I)$.
- Similarly, a strategy for P2, $\tau = (\tau_t)_{t\geq 1}$, with $\tau_t : (I \times J \times K)^{t-1} \longrightarrow \Delta(J)$.
- A profile (σ, τ) induces a proba $I\!P_{k_1,\sigma,\tau}$ on the set of plays $\Omega = (K \times I \times J)^{\infty}$.
- Let f be a bounded measurable function from $(K \times I \times J)^{\infty}$ to $I\!R$
- The payoff if P1 follows σ and P2 follows τ is

$$I\!E_{k_1,\sigma,\tau}(f)$$

Theorem[Martin, 1998]

Given a bounded measurable function $f : \Omega \longrightarrow I\!R$, the stochastic game with objective f has a value, i.e.

$$\sup_\sigma \inf_\tau I\!E_{k_1,\sigma,\tau}(f) = \inf_\tau \sup_\sigma I\!E_{k_1,\sigma,\tau}(f).$$

- A strategy for P1 is a rule specifying what he should play in any situation, formally $\sigma = (\sigma_t)_{t\geq 1}$, with $\sigma_t : (I \times J \times K)^{t-1} \longrightarrow \Delta(I)$.
- Similarly, a strategy for P2, $\tau = (\tau_t)_{t\geq 1}$, with $\tau_t : (I \times J \times K)^{t-1} \longrightarrow \Delta(J)$.
- A profile (σ, τ) induces a proba $I\!P_{k_1,\sigma,\tau}$ on the set of plays $\Omega = (K \times I \times J)^{\infty}$.
- Let f be a bounded measurable function from $(K \times I \times J)^{\infty}$ to $I\!R$
- The payoff if P1 follows σ and P2 follows τ is

$$I\!E_{k_1,\sigma,\tau}(f)$$

Theorem[Martin, 1998]

Given a bounded measurable function $f : \Omega \longrightarrow I\!R$, the stochastic game with objective f has a value, i.e.

$$\sup_\sigma \inf_\tau I\!E_{k_1,\sigma,\tau}(f) = \inf_\tau \sup_\sigma I\!E_{k_1,\sigma,\tau}(f).$$

Example

Example of objective functions: $g : K \times I \times J \longrightarrow I\!R$ is a fixed (bounded) stage payoff function, and
$f(k_1, i_1, j_1, ..., k_t, i_t, j_t, ...) =$

$$\begin{array}{cc} \sup_t g_t & (sup - game) \\ \limsup_t g_t & (limsup - game) \\ \limsup_T \frac{1}{T} \sum_{t=1}^{T} g_t & (limpsup - meangame) \\ \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g_t, \text{ with } \lambda \in (0,1] & (discounted\ game) \end{array}$$

(with $g_t = g(k_t, i_t, j_t)$ the induced payoff of stage t)

Remark

Existence of the value for discounted games: OK by compactness and continuity

Outline

1. A first model: no state, perfect information
2. A second example: state variable, perfect observation
3. Repeated games with signals

Model:

- I and J are non empty finite sets of actions.
- K is a countable set of states,
- C and D are non empty finite sets of signals.
- $q : K \times I \times J \longrightarrow \Delta(K \times C \times D)$ is a transition
- $\pi \in \Delta(K \times C \times D)$ an initial distribution.

How the game is played:

- (k_1, c_1, d_1) is fixed. P1 learns c_1 and P2 learns c_2.
- P1 and P2 simultaneously choose i_1 in I and j_1 in J, (k_2, c_2, d_2) is selected according to $q(k_1, i_1, j_1)$
- c_2 is announced to P1, d_2 is announced to $P2$.
- etc.....

What can we say?

- In general nothing.

What can we say?

- In general nothing.

An encoding of "pick the largest integer"

Example

Both players in the dark,the initial state is k_2, P1 chooses the line, P2 chooses the column.

$$\begin{pmatrix} \circlearrowleft & 0^* \\ \circlearrowleft & 0^* \end{pmatrix} \qquad \begin{pmatrix} \circlearrowleft & \rightarrow \\ \leftarrow & \circlearrowleft \end{pmatrix} \qquad \begin{pmatrix} \circlearrowleft & \circlearrowleft \\ 1^* & 1^* \end{pmatrix}$$

k_1 $\qquad$ k_2 $\qquad$ k_3

An $\circlearrowleft$ means that the play remains in the same state. Payoff for P1 is : 1 in k_1 and 0 in k_2 and k_3.

No limsup value

$$\sup_\sigma \inf_\tau I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) = 0 < 1 = \inf_\tau \sup_\sigma I\!E_{k_2,\sigma,\tau}(\limsup_t g_t)$$

An encoding of "pick the largest integer"

Example

Both players in the dark,the initial state is k_2, P1 chooses the line, P2 chooses the column.

$$\begin{pmatrix} \circlearrowleft & 0^* \\ \circlearrowleft & 0^* \end{pmatrix} \qquad \begin{pmatrix} \circlearrowleft & \rightarrow \\ \leftarrow & \circlearrowleft \end{pmatrix} \qquad \begin{pmatrix} \circlearrowleft & \circlearrowleft \\ 1^* & 1^* \end{pmatrix}$$

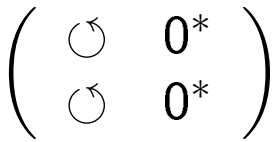
k_1 $\qquad$ k_2 $\qquad$ k_3

An $\circlearrowleft$ means that the play remains in the same state. Payoff for P1 is : 1 in k_1 and 0 in k_2 and k_3.

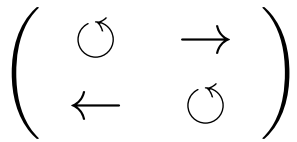
No limsup value

$$\sup_\sigma \inf_\tau I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) = 0 < 1 = \inf_\tau \sup_\sigma I\!E_{k_2,\sigma,\tau}(\limsup_t g_t)$$

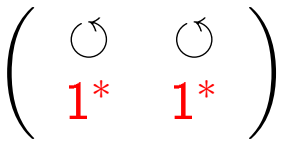
P1 guarantees nothing more than 0



k_1



k_2



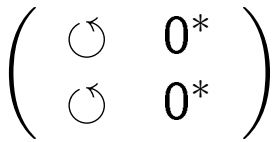
k_3

- Fix a strategy σ of P1
  - α_n is the proba that P1 plays B for the first time at stage n.
  - α^* the proba that he always play T.

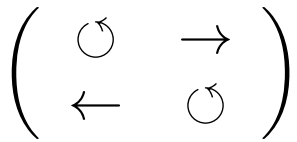
$$\alpha^* + \sum_{n=1}^{\infty} \alpha_n = 1.$$

- For $\varepsilon > 0$, let N be s.t. $\sum_{n=N+1}^{\infty} \alpha_n \leq \varepsilon$: after N, P1 do not play B anymore (almost).
- Define the strategy τ of P2 by:
  - play L at stages 1,...,N
  - then play R forever.
- P1 wins only if he plays B for the first time after N, so with proba at most ε. Hence $I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) \leq \varepsilon$.

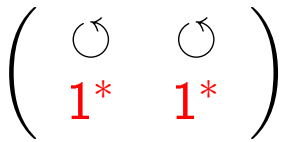
P1 guarantees nothing more than 0



k_1



k_2



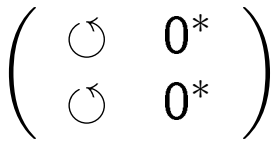
k_3

- Fix a strategy σ of P1
  - α_n is the proba that P1 plays B for the first time at stage n.
  - α^* the proba that he always play T.

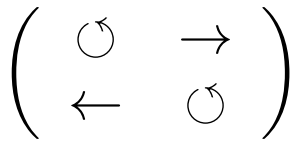
$$\alpha^* + \sum_{n=1}^{\infty} \alpha_n = 1.$$

- For $\varepsilon > 0$, let N be s.t. $\sum_{n=N+1}^{\infty} \alpha_n \leq \varepsilon$: after N, $P1$ do not play B anymore (almost).
- Define the strategy τ of P2 by:
  - play L at stages 1,...,N
  - then play R forever.
- P1 wins only if he plays B for the first time after N, so with proba at most ε. Hence $I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) \leq \varepsilon$.

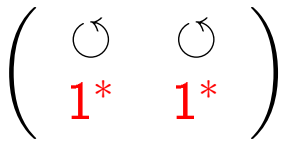
P1 guarantees nothing more than 0



k_1



k_2



k_3

- Fix a strategy σ of P1
  - α_n is the proba that P1 plays B for the first time at stage n.
  - α^* the proba that he always play T.

$$\alpha^* + \sum_{n=1}^{\infty} \alpha_n = 1.$$

- For $\varepsilon > 0$, let N be s.t. $\sum_{n=N+1}^{\infty} \alpha_n \leq \varepsilon$: after N, $P1$ do not play B anymore (almost).
- Define the strategy τ of P2 by:
  - play L at stages 1,...,N
  - then play R forever.
- P1 wins only if he plays B for the first time after N, so with proba at most ε. Hence $I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) \leq \varepsilon$.

And what if the players observe the states ?

Example

The Big Match where P1 does not observe P2's actions. P2 may or may not observe P1's actions.

$$\begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} 1* & 0* \\ 0 & 1 \end{pmatrix} \end{array}$$

No limsup value

$$\sup_\sigma \inf_\tau I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) = 0 < 1/2 = \inf_\tau \sup_\sigma I\!E_{k_2,\sigma,\tau}(\limsup_t g_t)$$

And what if the players observe the states ?

Example

The Big Match where P1 does not observe P2's actions. P2 may or may not observe P1's actions.

$$\begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} 1* & 0* \\ 0 & 1 \end{pmatrix} \end{array}$$

No limsup value

$$\sup_\sigma \inf_\tau I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) = 0 < 1/2 = \inf_\tau \sup_\sigma I\!E_{k_2,\sigma,\tau}(\limsup_t g_t)$$

We need additional assumptions:

-
-

We need additional assumptions:

- Either on the stage payoff (g) or the global evaluation (f).
- Or on the information structure.

Theorem 1

A repeated game with signals has a sup-value, i.e.

$$\sup_\sigma \inf_\tau I\!E_{\pi,\sigma,\tau}\left(\sup_t g_t\right) = \inf_\tau \sup_\sigma I\!E_{\pi,\sigma,\tau}\left(\sup_t g_t\right).$$

Definition

The game is recursive if there exists K^0 and K^* a partition of K such that

- the stage payoff is 0 on active states, K^0: $\forall k \in K^0$, $g(k) = 0$.
- states in K^* are absorbing: $\forall k \in K^*$, the payoff is constant and $\sum_{s \in S} q(k,i,j)(k,s) = 1$ for all $(i,j) \in I \times J$.

Theorem 2

Consider a recursive stochastic game with non negative payoffs.

- It has a sup-value v and a limsup-value w, and they coincide: $v = w$.
- Moreover

$$\forall \varepsilon > 0, \exists \sigma,\ \exists T_0,\ \forall T \geq T_0,\ \forall \tau,\ I\!E_{\pi,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq v - \varepsilon$$

$$\exists \tau,\ \forall \varepsilon > 0,\ \exists T_0,\ \forall T \geq T_0,\ \forall \sigma,\ I\!E_{\pi,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \leq v + \varepsilon,$$

Definition

The game is recursive if there exists K^0 and K^* a partition of K such that

- the stage payoff is 0 on active states, K^0: $\forall k \in K^0$, $g(k) = 0$.
- states in K^* are absorbing: $\forall k \in K^*$, the payoff is constant and $\sum_{s \in S} q(k,i,j)(k,s) = 1$ for all $(i,j) \in I \times J$.

Theorem 2

Consider a recursive stochastic game with non negative payoffs.

- It has a sup-value v and a limsup-value w, and they coincide: $v = w$.
- Moreover

$$\forall \varepsilon > 0, \exists \sigma,\ \exists T_0,\ \forall T \geq T_0,\ \forall \tau,\ I\!E_{\pi,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq v - \varepsilon$$

$$\exists \tau,\ \forall \varepsilon > 0,\ \exists T_0,\ \forall T \geq T_0,\ \forall \sigma,\ I\!E_{\pi,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \leq v + \varepsilon,$$

"pick the largest integer" again?

Example 3: An example of recursive game with no limsup value (Shmaya, quoted from Rosenberg & al.). Both players in the dark, the initial state is k_2 and is known:

$$\left(\begin{array}{cc} 0 & -2^* \\ 0 & -2^* \end{array}\right)\left(\begin{array}{cc} 0 & 1/2(-1^*)+1/2(k_3) \\ 1/2(1^*)+1/2(k_1) & 0^* \end{array}\right)\left(\begin{array}{cc} 0 & 0 \\ 2^* & 2^* \end{array}\right)$$

k_1 k_2 k_3

For example, if the state is k_2, if player 1 plays T and Player 2 plays R then with proba $1/2$ the payoff is -1 forever, and with probability $1/2$ the play goes to k_3.

$$\sup_\sigma \inf_\tau I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) = -1/2 < 1/2 = \inf_\tau \sup_\sigma I\!E_{k_2,\sigma,\tau}(\limsup_t g_t).$$

"pick the largest integer" again?

Example 3: An example of recursive game with no limsup value (Shmaya, quoted from Rosenberg & al.). Both players in the dark, the initial state is k_2 and is known:

$$\left(\begin{array}{cc} 0 & -2^* \\ 0 & -2^* \end{array}\right)\left(\begin{array}{cc} 0 & 1/2(-1^*)+1/2(k_3) \\ 1/2(1^*)+1/2(k_1) & 0^* \end{array}\right)\left(\begin{array}{cc} 0 & 0 \\ 2^* & 2^* \end{array}\right)$$

k_1 k_2 k_3

For example, if the state is k_2, if player 1 plays T and Player 2 plays R then with proba $1/2$ the payoff is -1 forever, and with probability $1/2$ the play goes to k_3.

$$\sup_\sigma \inf_\tau I\!E_{k_2,\sigma,\tau}(\limsup_t g_t) = -1/2 < 1/2 = \inf_\tau \sup_\sigma I\!E_{k_2,\sigma,\tau}(\limsup_t g_t).$$

Definition

A symmetric signaling stochastic game is a repeated game with signals such that

- there exists a set S with $C = D = I \times J \times S$ satisfying

$$\forall (x,i,j) \in X \times I \times J, \sum_{s,x'} q(x,i,j)(x',(i,j,s),(i,j,s)) = 1.$$

- π is also symmetric.

Theorem 3

Let Γ be a symmetric signaling repeated game, then for every Borelian evaluation f, the game Γ has a value i.e.

$$\sup_\sigma \inf_\tau I\!E_{\pi,\sigma,\tau}(f) = \inf_\tau \sup_\sigma I\!E_{\pi,\sigma,\tau}(f).$$

Definition

A symmetric signaling stochastic game is a repeated game with signals such that

- there exists a set S with $C = D = I \times J \times S$ satisfying

$$\forall (x,i,j) \in X \times I \times J, \sum_{s,x'} q(x,i,j)(x',(i,j,s),(i,j,s)) = 1.$$

- π is also symmetric.

Theorem 3

Let Γ be a symmetric signaling repeated game, then for every Borelian evaluation f, the game Γ has a value i.e.

$$\sup_\sigma \inf_\tau I\!E_{\pi,\sigma,\tau}(f) = \inf_\tau \sup_\sigma I\!E_{\pi,\sigma,\tau}(f).$$

Outline of the proof: applying Martin's theorem to an auxilliary stochastic game

We define an auxiliary stochastic game on joined beliefs: $\widetilde{\Gamma}(p_1)$ where

- the set of states is $Z = \Delta(K) \times S$.
- the sets of actions are I and J,
- the transition function is $\widetilde{q} : Z \times I \times J \to \Delta_f(Z)$:

$$\widetilde{q}((p,s),i,j) = \sum_{s' \in S} q(p,i,j)(s') \delta_{q(p,i,j)(.|s'),s'}$$

- the initial probability is $\delta_{(p,s)} \in \Delta_f(Z)$.

Outline of the proof

Given on observed history $h_o \in (S \times I \times J)^{+\infty}$, we would like to define

$$\widetilde{f}(h_o) = \mathbb{E}(f(h)|h_o)$$

- Problem : we need to define the conditional expectation
  - h_o may have probability 0.
  - Several probalities (not countable) depending on the profile of strategy (σ, τ).
- Solution: build by hand regular conditional probability which does not depend on (σ, τ)
  - for finite histories and finite observed histories, the conditional distribution can be explicitely computed and does not depend of $\mathbb{P}_{\pi,\sigma,\tau}$.
  - Extend to infinite history and infinite observed hisitories.

Outline of the proof

Given on observed history $h_o \in (S \times I \times J)^{+\infty}$, we would like to define

$$\widetilde{f}(h_o) = \mathbb{E}(f(h)|h_o)$$

- Problem : we need to define the conditional expectation
  - h_o may have probability 0.
  - Several probalities (not countable) depending on the profile of strategy (σ, τ).
- Solution: build by hand regular conditional probability which does not depend on (σ, τ)
  - for finite histories and finite observed histories, the conditional distribution can be explicitely computed and does not depend of $\mathbb{P}_{\pi,\sigma,\tau}$.
  - Extend to infinite history and infinite observed hisitories.

Outline of the proof

Given on observed history $h_o \in (S \times I \times J)^{+\infty}$, we would like to define

$$\widetilde{f}(h_o) = I\!E(f(h)|h_o)$$

- Problem : we need to define the conditional expectation
  - h_o may have probability 0.
  - Several probalities (not countable) depending on the profile of strategy (σ, τ).
- Solution: build by hand regular conditional probability which does not depend on (σ, τ)
  - for finite histories and finite observed histories, the conditional distribution can be explicitely computed and does not depend of $I\!P_{\pi,\sigma,\tau}$.
  - Extend to infinite history and infinite observed hisitories.

Conclusions :

- No value for repeated game with signals in general.
- Existence of the value for some Borelian evaluations independently of the signaling structure.
- Existence of the value for symmetric game under any Borelian evaluation.

Further research:

- Characterize the set of information structures where the value always exists ? the limsup exists?
- In the special case of the limsup payoff, Maitra and Sudderth proved the existence of the value whith an operator approach.
  - Can we adopt their approach?
  - Study the decidability of computing the value.

Thanks