Zero-sum Markov Games with Incomplete Information on One Side

Eilon Solan, Catherine Rainer & Galit Ashkenazi-Golan

Theory and Algorithms in Graph and Stochastic Games, Mons, Belgium

Zero Sum Games

The Model

- Two players: $\{1, 2\}$
- A set of actions of Player 1: $A = \{a_1, ..., a_{m_1}\}$
- A set of actions of Player 2: $B = \{b_1, ..., b_{m_2}\}$
- A payoff function: $u(a_i, b_j) = (u_1(a_i, b_j), -u_1(a_i, b_j))$, and for mixed actions $x \in \Delta(A)$, $y \in \Delta(B)$, $u(x, y) = \Sigma_i \sigma_j x(a_i) y(b_j) u(a_i, b_j)$

Where for a set W, $\Delta(W) = conv(W)$

The Minmax Theorem Von Neumann (1928)

Recall - Player 1 is maximizing and Player 2 is minimizing.

The Minmax Theorem

$$Max_{x \in \Delta(A)} Min_{y \in \Delta(B)} u(x,y) = Min_{y \in \Delta(B)} Max_{x \in \Delta(A)} u(x,y)$$

Aumann and Maschler (1968, 1995)

Bayesian Games with Incomplete Information on One Side

- Two players: $\{1, 2\}$
- Two states: $\{\alpha, \beta\}$
- A zero-sum payoff matrix for each state
- The state is randomized at the beginning of the game (with probability p to state α)
- Only Player 1 is informed of the state
- The game is played infinitely many times
- Observable actions, non observable payoffs

Notations:

$u(p)$ - the one shot value of the game when both players are not informed
$v(p)$ - the value of the infinitely repeated game

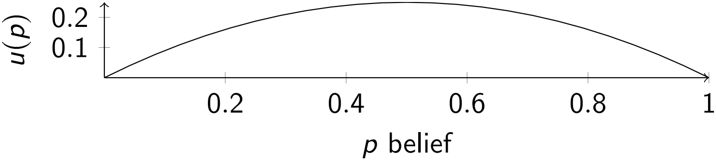
Aumann and Maschler - First Example

A Case Where Information Should Not Be Revealed

State α	L	R
T	1	0
B	0	0

State β	L	R
T	0	0
B	0	1

Result	L	R
T	p	0
B	0	1-p



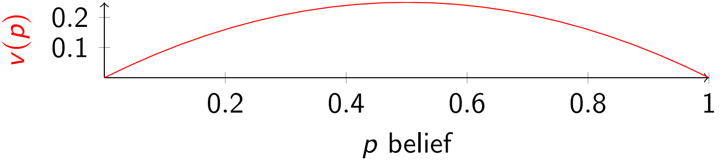
Aumann and Maschler - First Example

A Case Where Information Should Not Be Revealed

State α	L	R
T	1	0
B	0	0

State β	L	R
T	0	0
B	0	1

Result	L	R
T	p	0
B	0	1-p



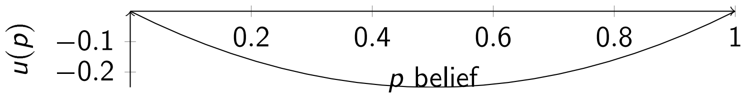
Aumann and Maschler - Second Example

A Case Where Information Should Be Revealed

State α	L	R
T	-1	0
B	0	0

State β	L	R
T	0	0
B	0	-1

Result	L	R
T	-p	0
B	0	-1+p



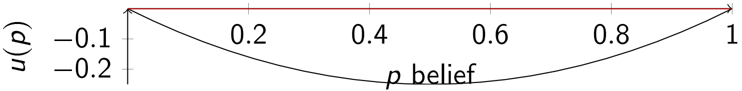
Aumann and Maschler - Second Example

A Case Where Information Should Be Revealed

State α	L	R
T	-1	0
B	0	0

State β	L	R
T	0	0
B	0	-1

Result	L	R
T	-p	0
B	0	-1+p



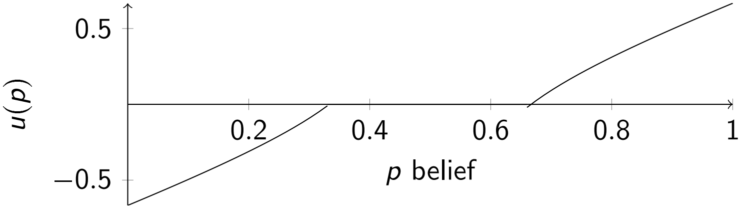
Aumann and Maschler - Third Example

A case where some information is revealed

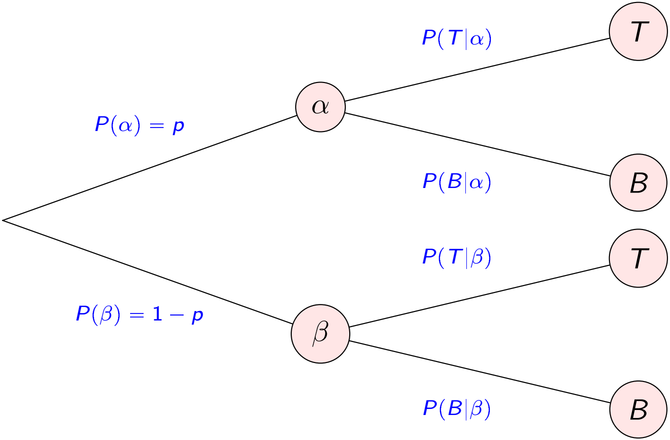
State α	L	R
T	1	0
B	0	2

State β	L	R
T	-2	0
B	0	-1

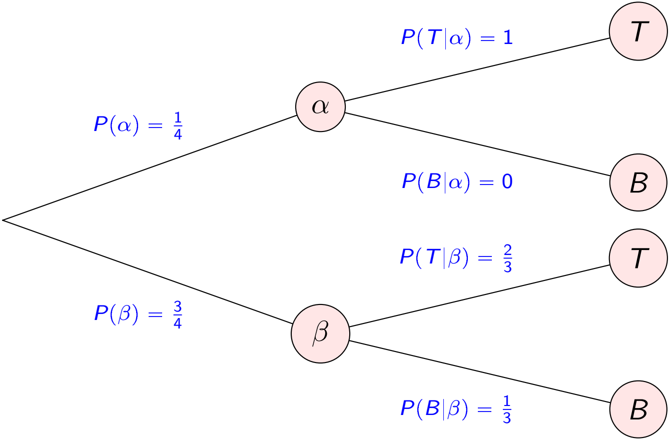
Result	L	R
T	3p-2	0
B	0	3p-1



Aumann and Maschler - Third Example



Aumann and Maschler - Third Example



$$P(\alpha|T) = \frac{\frac{1}{4}}{\frac{1}{4}+\frac{3}{4}\frac{2}{3}} = \frac{1}{3}$$

$$P(\alpha|B) = 0$$

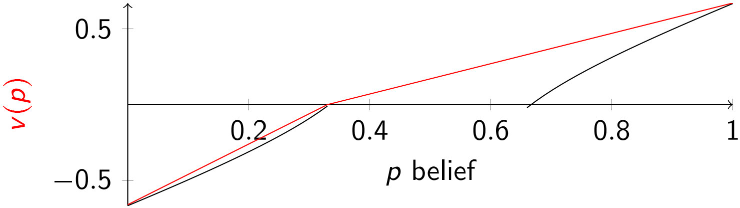
Aumann and Maschler - Third Example

A case where some information is revealed

State α	L	R
T	1	0
B	0	2

State β	L	R
T	-2	0
B	0	-1

Result	L	R
T	3p-2	0
B	0	3p-1



Aumann and Maschler (1968, 1995)

The result

- Characterization of the value function - concavification of the value function of the one shot game when no player knows the state
- The optimal strategy of the informed player involves a single state of revelation of information

The Model

Markov Games with Incomplete Information on One Side

- Two players: $\{1, 2\}$
- Two states: $\{\alpha, \beta\}$
- A zero-sum payoff matrix for each state
- The game switches between states α and β in a Markovian way.
- Observable actions, non observable payoffs
- Only Player 1 is informed of the transitions
- Player 2 does not know the state, denote by p the probability with which he believes that the state is α (p changes along the game)
- Time duration between stages goes to zero:
  - Discount factor: $e^{-r/n}$
  - Switching probability at each state: from α to β probability $1 - e^{-\lambda_1/n}$ and from β to α with probability $1 - e^{-\lambda_2/n}$. This talk: $\lambda_2 = 0$, a single transition.

Cardaliaguet, Rainer, Rosenberg and Vieille (2016) and Gensbittel (unpublished)

The model

- Markov games: the state changes over time (only as a function of current state)
- Incomplete information on one side
- Time duration between stages goes to zero

The result

- Characterization of the limit value function
- Characterization of the limit optimal strategy of the informed player (this strategy involves repeated revelation of information)
- The characterization through a system of differential equations (does not find the value or the optimal strategy explicitly)

Our Goal

To provide a finite-stage algorithm to compute the limit value of the Markov game, as duration between stages goes to zero (and so do the switching probabilities).

As a by-product the algorithm provides a strategy for Player 1, which is approximately optimal for small duration between stages.

Model Presented Here

We present here the algorithm for a special case: the single transition game, when the game begins at state α and moves to the absorbing state β.

In this case, if Player 1 reveals no information, the belief of Player 2 drifts towards 0.

The options of Player 1

Suppose the value of the game at belief p_0 is $v(p_0)$, and we are now at belief $p > p_0$

No revelation of information (sliding)

- Let φ denote the value with no revelation
- For one stage, the belief of Player 2 begins at p and ends at $pe^{-\frac{\lambda_1}{n}}$
- φ satisfies:

$$\varphi(p) = (1 - e^{-\frac{r}{n}})u(p) + e^{-\frac{r}{n}}\varphi(pe^{-\frac{\lambda_1}{n}})$$

- φ is a solution of a differential equation, specifically $\varphi'(p) = \frac{r(u(p)-\varphi(p))}{\lambda_1 p}$

The options of Player 1

Suppose the value of the game at belief p_0 is $v(p_0)$, and we are now at belief $p > p_0$

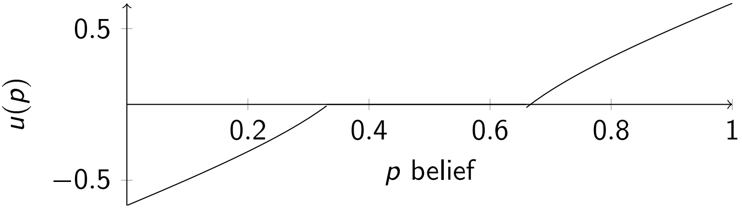
Jumping from p to p_0

- Let ψ denote the value when jumping
- The payoff in first stage $u(p)$, with weight $1 - e^{\frac{r}{n}}$
- Belief becomes $pe^{-\frac{\lambda_1}{n}}$
- Player 1 makes Player 2 have a belief of p with probability $\frac{pe^{-\frac{\lambda_1}{n}} - p_0}{p - p_0}$, and with the complement probability have a belief of p_0
- The resulting ψ is a convex combination of $u(p)$ and $v(p_0)$, a linear function with a known slope that depends on p

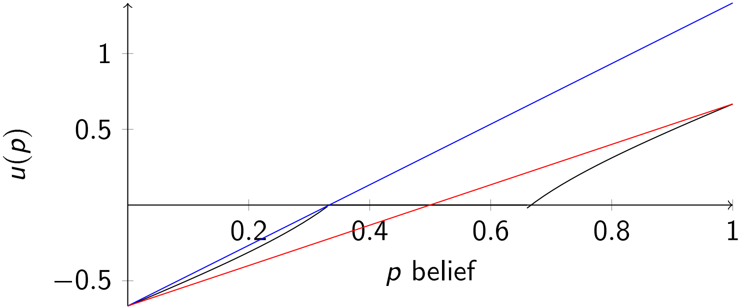
Splitting the belief

Turns out to be unprofitable when the smaller is belief is not close to current one

The options of Player 1



The options of Player 1



The Algorithm

- Begin at $p = 0$ and work your way towards $p = 1$. Suppose now at p_0 (meaning $v(p_0)$ is known).
- Find the p from which a jump gives the highest slope (that slope is the derivative of ψ)
- Compare to the derivative of φ, the derivative when sliding
- Choose sliding or jumping according to the derivative that is the higher one for p's near p_0 (and slightly larger)
- The slide or jump goes on until $p = 1$ or until the now smaller derivative becomes the higher one
- Repeat until $p = 1$

The proof has two main parts:

- The algorithm ends after a finite number of iterations.
- The resulting v is the function answering the characterization of Cardaliaguet, Rainer, Rosenberg and Vieille.

Two remarks

- Why no splitting is important?
- The general Markov case:
  - The belief drifts towards the stationary distribution (p^*) of the Markov process.
  - An initial step is added:finding the value at p^*. Then the algorithm is employed twice : once towards 0 and then towards 1.

Open questions

- Optimal strategy of Player 2.
- Three states - even with two moves only:
  - Value of the game - an expression or an algorithm to find it.
  - Optimal strategies for both players.
- Two states - incomplete information for both sides.

THANK YOU