

Stochastic Games

Hugo Gimbert¹

¹CNRS, LaBRI, Bordeaux

GAMES-EPIT Spring School, Carcans-Maubuisson, 2011

Outline

- 1 Stochastic games with perfect information
 - Markov chains and probability measures
 - Simple stochastic games
 - A few results about stochastic games with perfect information
- 2 Stochastic games for economists
 - Games in strategic form
 - Discounted games (Shapley)
 - Undiscounted games (Mertens Neyman)
- 3 Determinacy of Blackwell games

Introduction

Stochastic games: played over time

- ⇒ some events are controlled some are random.
- ⇒ discrete time.
- ⇒ zero-sum.

Three players: Maximizer, Minimizer and Nature.

Problems:

Existence of values.

Description of (ε -)optimal strategies.

Computation of values and strategies.

Probability distributions

Definition

\mathbf{S} a finite or countable set. The set of probability distributions:

$$\mathcal{D}(\mathbf{S}) = \left\{ \delta : \mathbf{S} \rightarrow [0, 1] \mid \sum_s \delta(s) = 1 \right\}.$$

Example

Random choice of an integer:

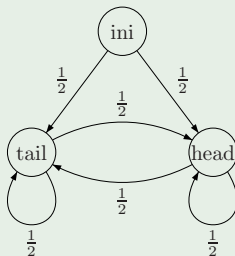
$$\mathbf{S} = \mathbb{N} \text{ and } \delta(n) = \frac{1}{2^{n+1}}$$

Throw a coin forever: $\mathbf{S} = \{H, T\}^\omega$.

Intuitively, all realisations in \mathbf{S}^ω are equiprobable.
How to measure probabilities of such experiments?

Markov chain

Example



Definition (Markov chain)

Let \mathbf{S} be a finite or countable set. A Markov chain on \mathbf{S} is a tuple $\mathcal{M} = (\mathbf{S}, s_0 \in \mathbf{S}, p : \mathbf{S} \rightarrow \mathcal{D}(\mathbf{S}))$. A trajectory is an infinite sequence in S^ω .

How to measure probability of sets of trajectories?

Example: what is the probability that the number of heads minus the number of tails is unbounded along a trajectory?

Measuring probabilities in a Markov chain

Definition (Cylinders and events)

A cylinder is a subset of S^ω of type $s_0 \cdots s_n S^\omega$ with $s_0 \cdots s_n \in S^*$.

The set of *events* (Borel-measurable sets) is the smallest collection of subsets of S^ω which contains cylinders and which is stable by complementation (denoted \neg) and countable union.

Example

- $E = \{(x_n)_{n \in \mathbb{N}} \in \{0, 1\}^\omega \mid x_n = 0 \text{ for infinitely many } n\}$,
- $E = \{\limsup \frac{x_0 + x_1 + \dots + x_n}{n+1} = \frac{1}{2}\}$,
- any singleton, $E = \{3 + \sum_n \frac{x_n}{2^{n+1}} = \pi\}$.

Existence of a probability measure

Theorem

For every Markov chain $\mathcal{M} = (\mathbf{S}, s_0 \in \mathbf{S}, p : \mathbf{S} \rightarrow \mathcal{D}(\mathbf{S}))$ with a finite or countable set of states, there exists a unique probability measure \mathbb{P}_{s_0} which associates to every event $E \subseteq \mathbf{S}^\omega$ a probability $\mathbb{P}_{s_0}(E) \in [0, 1]$, with the following properties:

Unity: $\mathbb{P}_{s_0}(\mathbf{S}^\omega) = 1$.

Cylindres: $\mathbb{P}_{s_0}(t_0 \cdots t_n \mathbf{S}^\omega) = \begin{cases} 0 & \text{if } t_0 \neq s_0 \\ p(t_0, t_1)p(t_1, t_2) \cdots p(t_{n-1}, t_n) & \text{otherwise.} \end{cases}$

Additivity: if $E \cap F = \emptyset$ then $\mathbb{P}_{s_0}(E \cup F) = \mathbb{P}_{s_0}(E) + \mathbb{P}_{s_0}(F)$.

σ -additivity: $\mathbb{P}_{s_0}\left(\bigcup_{n \in \mathbb{N}} E_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}_{s_0}\left(\bigcup_{m \leq n} E_m\right)$

Markov property

In a Markov chain, all knowledge about future states is given by the current state, the past states make no difference.

Markov property: For every prefix $s_0 \cdots s_n \in S^*$, the shifted probability measure:

$$\mathbb{P}_{s_0 \cdots s_n} : E \mapsto \mathbb{P}_{s_0} (s_0 \cdots s_n \cdot E \mid s_0 \cdots s_n S^\omega)$$

only depends on s_n .

Reachability probabilities

Definition

Probability to access a subset $\mathbf{T} \subseteq \mathbf{S}$ from an initial state s_0

$$\text{val}(s_0) = \mathbb{P}_{s_0}(\mathbf{S}^* \mathbf{T} \mathbf{S}^\omega) = \mathbb{P}_{s_0}(\exists n, S_n \in \mathbf{T}) .$$

Proposition

The reachability probabilities are characterized by:

$$\text{val}(s) = \begin{cases} 1 & \text{if } s \in \mathbf{T} \\ 0 & \text{if } \mathbf{T} \text{ is not accessible from } s \\ \sum_{t \in \mathbf{S}} p(s, t) \cdot \text{val}(t) & \text{otherwise.} \end{cases}$$

Remark

Unique solution computable in cubic time.

Third equality holds by the Markov property.

Proof of unicity

$$\text{val}(s) = \begin{cases} 1 & \text{if } s \in \mathbf{T} \\ 0 & \text{if } \mathbf{T} \text{ is not accessible from } s \\ \sum_{t \in \mathbf{S}} p(s, t) \cdot \text{val}(t) & \text{otherwise.} \end{cases}$$

Proof.

Let $(v_1(s))_{s \in \mathbf{S}}$ and $(v_2(s))_{s \in \mathbf{S}}$ two solutions.

Let $U \subseteq \mathbf{S}$ the set of states s which maximize $\phi : s \mapsto (v_1(s) - v_2(s))$.

For every $s \in \mathbf{S}$, $(v_1(s) - v_2(s)) = \sum_{t \in \mathbf{S}} p(s, t) \cdot (v_1(t) - v_2(t))$.

Thus U contains all states accessible from U .

If U does not intersect T , ϕ is zero on U .

If U intersects T , ϕ is zero on U .

Thus the maximum of ϕ is 0.

Thus for every state s , $v_1(s) \leq v_2(s)$.

By symmetry, $v_1 = v_2$. □

Outline

- 1 **Stochastic games with perfect information**
 - Markov chains and probability measures
 - **Simple stochastic games**
 - A few results about stochastic games with perfect information
- 2 **Stochastic games for economists**
 - Games in strategic form
 - Discounted games (Shapley)
 - Undiscounted games (Mertens Neyman)
- 3 **Determinacy of Blackwell games**

A stochastic game with finite duration

Example (The coin game)

You throw a coin, maximum three times.

You stop whenever you want.

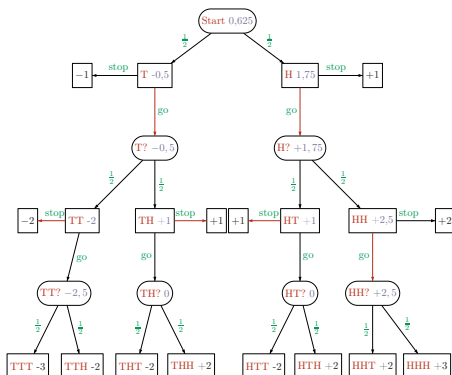
- If more tails than heads, you **lose** as many euros as there are tails.
- If more heads than tails, you **win** as many euros as there are heads.
- In case of equality (one tail and one head), you win one euro.

Do you accept playing this game?

What is your strategy?

What payoff can you expect on average?

Dynamic programming (Bellman, 40's)



Dynamic programming (Bellman): compute optimal choices and values in a bottom-up fashion.

Maximize on choice nodes, average on random nodes.

Optimal strategy: choose any successor with maximal value.

Advertising your research:

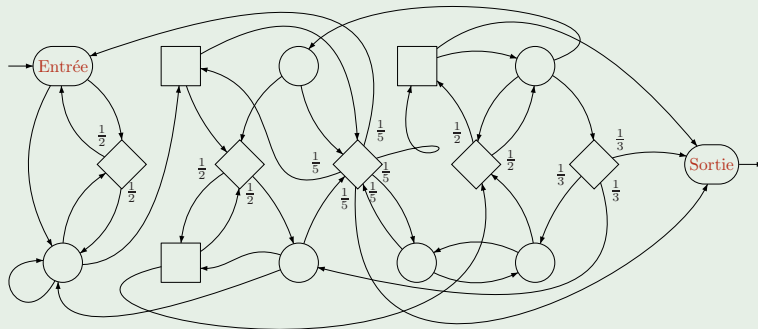
Quoting Bellman, speaking about the 50's:

” An interesting question is, ‘Where did the name, dynamic programming, come from?’ The 1950s were not good years for mathematical research. (...) Hence, I felt I had to do something to shield [the Secretary of Defense] and the Air Force from the fact that I was really doing mathematics inside the RAND Corporation.(...) Let’s take a word that has an absolutely precise meaning, namely *dynamic*, in the classical physical sense. It also has a very interesting property as an adjective, and that is it’s impossible to use the word *dynamic* in a pejorative sense. Try thinking of some combination that will possibly give it a pejorative meaning. It’s impossible. I thought *dynamic programming* was a good name. It was something not even a Congressman could object to. So I used it as an umbrella for my activities.”

General case

- Maximizer (circles) against Minimizer (squares)
- Cycles in the game graph.

Example (A simple stochastic games)



No way to use dynamic programming.

Strategies

Definition (Stochastic game with perfect information)

Simple stochastic game: $\mathbf{G} = (\mathbf{S}, \mathbf{S}_{\text{Max}}, \mathbf{S}_{\text{Min}}, \mathbf{S}_R, E, \mathbf{T}, p : \mathbf{S}_R \rightarrow \mathcal{D}(\mathbf{S}))$ where \mathbf{S} is partitioned between $(\mathbf{S}_{\text{Max}}, \mathbf{S}_{\text{Min}}, \mathbf{S}_R)$.

Deterministic transitions: $E \subseteq (\mathbf{S}_{\text{Max}} \cup \mathbf{S}_{\text{Min}}) \times \mathbf{S}$

Transition probability: $p(t, s)$ from state $s \in \mathbf{S}_R$ to state $t \in \mathbf{S}$.

Strict competition Max wants to reach \mathbf{T} and Min does not want to.

Strategies are used by players to take decisions.

Definition (Strategies)

Strategy for Max:

function $\sigma : \mathbf{S}^* \mathbf{S}_{\text{Max}} \rightarrow \mathcal{D}(\mathbf{S})$ such that $\sigma(s_0, \dots, s_n)(t) > 0 \implies (s_n, t) \in E$.

Probability measure associated to strategies

Definition

When players play with strategies σ and τ from the initial state s then $\mathbb{P}_s^{\sigma, \tau}(\cdot)$ is the probability measure on the Markov chain with states $s\mathbf{S}^*$ and transition probabilities induced by the two strategies:

$$\mathbb{P}_s^{\sigma, \tau}(s_0 \cdots s_n s_{n+1} \mid s_0 \cdots s_n) = \begin{cases} \sigma(s_0 \cdots s_n)(s_{n+1}) & \text{if } s_n \in \mathbf{S}_{\text{Max}} \\ \tau(s_0 \cdots s_n)(s_{n+1}) & \text{if } s_n \in \mathbf{S}_{\text{Min}} \\ p(s_{n+1}, s_n) & \text{if } s_n \in \mathbf{S}_R . \end{cases} \quad (1)$$

Simple stochastic games

Winning condition: event $W \subseteq S^\omega$

The maximizer wants to maximize the probability $\mathbb{P}_s^{\sigma, \tau}(W)$.

The minimizer wants to minimize the probability $\mathbb{P}_s^{\sigma, \tau}(W)$.

Simple stochastic games: target vertices $\mathbf{T} \subseteq \mathbf{S}$.

$$W = S^*TS^\omega = \{\exists n \in \mathbb{N}, S_n \in \mathbf{T}\} .$$

Value of a game

Who should choose his strategy first?

Definition

The inferior and superior values of a state $s \in \mathbf{S}$ are:

$$\text{val}_*(s) = \sup_{\sigma} \inf_{\tau} \mathbb{P}_s^{\sigma, \tau}(W) \quad (\text{Max chooses first}) ,$$

$$\text{val}^*(s) = \inf_{\tau} \sup_{\sigma} \mathbb{P}_s^{\sigma, \tau}(W) \quad (\text{Min chooses first}) .$$

$$\text{val}_* \leq \text{val}^* .$$

Theorem (Martin, 81)

Whenever W is Borel measurable, these two values are equal. This defines $\text{val}(s)$, the value of the game (from state s).

True in a more general framework.

Simple stochastic games

Theorem

- *A simple stochastic game has a value.*
- *This value is realized by simple strategies called positional strategies.*
- *This value is computable.*

Proposition

The inferior $\text{val}_* : \mathbf{S} \rightarrow [0, 1]$ and superior values $\text{val}^* : \mathbf{S} \rightarrow [0, 1]$ are solution of the system:

$$v(s) = \begin{cases} 1 & \text{if } s \in \mathbf{T} , \\ \max_{(s,t) \in E} v(t) & \text{if } s \in \mathbf{S}_{\text{Max}} , \\ \min_{(s,t) \in E} v(t) & \text{if } s \in \mathbf{S}_{\text{Min}} , \\ \sum_{t \in \mathbf{S}} p(t, s) \cdot v(t) & \text{if } s \in \mathbf{S}_R . \end{cases}$$

Remark

In general, several solutions.

Markov decision processes

Markov decision processes: $\mathbf{S}_{\text{Min}} = \emptyset$ thus the minimizer has no control on the play, and

$$\text{val}(s) = \sup_{\sigma} \mathbb{P}_s^{\sigma} (\exists n, S_n \in \mathbf{T}) .$$

The vector $(\text{val}(s))_{s \in \mathbf{S}}$ is the optimal solution in $[0, 1]^{\mathbf{S}}$ of the LP:

$$\text{minimize } \sum_{s \in \mathbf{S}} v(s)$$

$$\forall s \in \mathbf{S}_{\text{Max}}, \forall (s, t) \in E,$$

$$v(s) \geq v(t)$$

$$\forall s \in \mathbf{S}_R,$$

$$v(s) = \sum_{t \in \mathbf{S}} p(s, t) \cdot v(t)$$

$$\forall s \in \mathbf{T},$$

$$v(s) = 1$$

Proof.

The set U of vertices s where the difference $\text{val}^*(s) - v(s)$ is maximal is closed by moves of player Max as well as probabilistic transitions. □

Proposition

In a Markov decision process, values are computable in polynomial time.

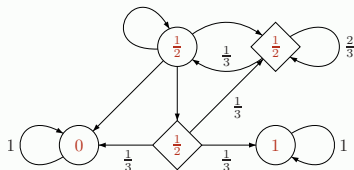
Question: once the values are computed, is it easy to compute an optimal strategy?

Canonical optimal strategy: from state s , choose randomly between all successors t such that $\text{val}(s) = \text{val}(t)$.

Proof.

$\mathbb{E}_s^\sigma[\text{val}(S_n)]$ is constant equal to $\text{val}(s)$. Almost-surely, either a target state or a state with value 0 is eventually reached. \square

Example



Proposition

In a Markov decision process, values are computable in polynomial time.

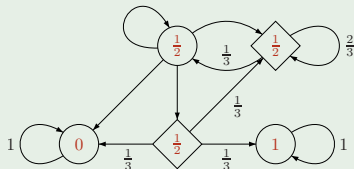
Question: once the values are computed, is it easy to compute an optimal strategy?

Canonical optimal strategy: from state s , choose randomly between all successors t such that $\text{val}(s) = \text{val}(t)$.

Proof.

$\mathbb{E}_s^\sigma[\text{val}(S_n)]$ is constant equal to $\text{val}(s)$. Almost-surely, either a target state or a state with value 0 is eventually reached. \square

Example



Strategy improvement algorithm

Algorithm

1 Start with an arbitrary positional strategy $\sigma_0 : \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$.

2 Repeat

- For each $s \in \mathbf{S}$, compute

$$\text{val}(\sigma_n)(s) = \mathbb{P}_s^{\sigma_n}(\exists n \in \mathbb{N}, S_n \in \mathbf{T}) .$$

- For each state s , switch $\sigma_n(s)$ to a successor vertex $\sigma_{n+1}(t)$ maximizing $\text{val}(\sigma_n)(t)$.

3 until no switch is required.

Correctness: for every s , $\text{val}(\sigma_0)(s) \leq \text{val}(\sigma_1)(s) \leq \text{val}(\sigma_0)(s) \dots$ with strict inequality on one state at least. Finite number of positional strategy.

Remark

Pros: easy implementation, no need to use linear programming.

Cons: exponential in the worst case (Condon et Melekopoglou, 1994).

(Open?) problem: find an algorithm to solve Markov decision processes which does not rely on linear programming.

Two-player case

Definition (Positional strategy)

Strategy $\sigma : \mathbf{S}^* \mathbf{S}_{\text{Max}} \rightarrow \mathcal{D}(\mathbf{S})$ is

pure if $\sigma : \mathbf{S}^* \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$,

stationary if $\sigma : \mathbf{S}_{\text{Max}} \rightarrow \mathcal{D}(\mathbf{S})$,

positional if it is both pure and stationary $\sigma : \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$.

Theorem (Existence of optimal positional strategies)

In every simple stochastic game, there exists positional strategies

$\sigma^\# : \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$ such that for every strategies σ, τ ,

$$\mathbb{P}_s^{\sigma, \tau^\#}(\exists n, S_n \in \mathbf{T}) \leq \mathbb{P}_s^{\sigma^\#, \tau^\#}(\exists n, S_n \in \mathbf{T}) \leq \mathbb{P}_s^{\sigma^\#, \tau}(\exists n, S_n \in \mathbf{T}) .$$

The common value is $\text{val}(s) := \text{val}_(s) = \text{val}^*(s)$.*

*These strategies are **optimal**.*

Proof.

Definition of $\tau^\#$: in state $s \in \mathbf{S}_{\text{Min}}$ choose a successor $\tau^\#(s)$ with minimal value:

$$\text{val}_*(\tau^\#(s)) = \min_{(s,t) \in E} \text{val}_*(t) .$$

Definition of $\sigma^\#$: in state $s \in \mathbf{S}_{\text{Max}}$ choose with uniform probability between all successors of maximal value.

$$(\sigma^\#(s)(t) > 0) \iff \text{val}^*(t) = \max_{(s,t') \in E} \text{val}^*(t') .$$

Let $v(s) = \mathbb{P}_s^{\sigma^\#, \tau^\#} (\exists n, S_n \in \mathbf{T})$. We show that for every σ, τ ,

$$\mathbb{P}_s^{\sigma, \tau^\#} (\exists n, S_n \in \mathbf{T}) \leq v_*(s) \leq v^*(s) \leq \mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T}) .$$



Proof.

Start with proving:

$$\mathbb{P}_s^{\sigma, \tau^\#} (\exists n, S_n \in \mathbf{T}) \leq \text{val}_*(s) .$$

For $n \in \mathbb{N}$ let:

$$\mathbb{E}_s^{\sigma, \tau^\#} [\text{val}_*(S_n)] = \sum_{s \in \mathbf{S}} \text{val}_*(s) \cdot \mathbb{P}_s^{\sigma, \tau^\#} (S_n = s) .$$

Be $s_n \in \mathbf{S}_{\text{Max}}$ or $s_n \in \mathbf{S}_{\text{Min}}$ or $s_n \in \mathbf{S}_R$,

$$\mathbb{E}_s^{\sigma, \tau^\#} [\text{val}_*(S_{n+1}) \mid S_n = s_n] \leq \text{val}_*(s_n) .$$

Consequently:

$$\begin{aligned} \mathbb{P}_s^{\sigma, \tau^\#} (\exists n, S_n \in \mathbf{T}) &\leq \mathbb{E}_s^{\sigma, \tau^\#} \left[\liminf_n \text{val}_*(S_n) \right] \\ &\leq \liminf_n \mathbb{E}_s^{\sigma, \tau^\#} [\text{val}_*(S_n)] \\ &\leq \text{val}_*(s) . \end{aligned}$$

Proof.

Now we show:

$$\mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T}) \geq \text{val}^*(s) .$$

First,

$$\mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T} \mid \forall n, \text{val}^*(S_n) > 0) = 1 . \quad (2)$$

For $s \in \mathbf{S}$, let $f(s) = \inf_\tau \mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T} \mid \forall n, \text{val}^*(S_n) > 0)$.

Let $U = \{s \in \mathbf{S} \mid \text{val}^*(s) > 0 \text{ and } f(s) \text{ minimal}\}$.

Then U is stable by $\sigma^\#$ and probabilistic transitions.

As a consequence, U contains a target vertex and f has value 1 on U .

Then, for every $s_n \in \mathbf{S}_{\text{Max}}$ or $s_n \in \mathbf{S}_{\text{Min}}$ or $s_n \in \mathbf{S}_R$,

$$\mathbb{E}_s^{\sigma^\#, \tau} [\text{val}^*(S_{n+1}) \mid S_n = s_n] \geq \text{val}^*(s_n) . \quad (3)$$

Consequently:

$$\begin{aligned} \mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T}) &\geq \mathbb{P}_s^{\sigma^\#, \tau} (\forall n, \text{val}^*(S_n) > 0) \\ &\geq \mathbb{E}_s^{\sigma^\#, \tau} \left[\limsup_n \text{val}^*(S_n) \right] \end{aligned}$$

Proof.

Finally, we have shown that for every σ, τ ,

$$\mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T}) \geq v^*(s) .$$

The strategy $\sigma^\# : \mathbf{S}_{\text{Max}} \rightarrow \mathcal{D}(\mathbf{S})$ is stationary but not pure. To complete the proof, one has to turn $\sigma^\#$ to a positional strategy $\sigma^{\#\#} : \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$ which guarantees the same payoff, an easy exercise... □

Remark

*This theorem was proved in a more general setting by Gillette (1957) using a wrong Tauberian theorem, then corrected by Liggett et Lippman (1969) using a wrong argument, and eventually proved by Bewley and Kohlberg (1978)...
⇒ double check proofs of game theoretists.*

Proof.

Finally, we have shown that for every σ, τ ,

$$\mathbb{P}_s^{\sigma^\#, \tau} (\exists n, S_n \in \mathbf{T}) \geq v^*(s) .$$

The strategy $\sigma^\# : \mathbf{S}_{\text{Max}} \rightarrow \mathcal{D}(\mathbf{S})$ is stationary but not pure. To complete the proof, one has to turn $\sigma^\#$ to a positional strategy $\sigma^{\#\#} : \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$ which guarantees the same payoff, an easy exercise... □

Remark

*This theorem was proved in a more general setting by Gillette (1957) using a wrong Tauberian theorem, then corrected by Liggett et Lippman (1969) using a wrong argument, and eventually proved by Bewley and Kohlberg (1978)...
⇒ double check proofs of game theoretists.*

Strategy enumeration

Proposition (Computing best response)

For every positional strategy $\sigma : \mathbf{S}_{\text{Max}} \rightarrow \mathbf{S}$, one can compute

$$\text{val}(\sigma)(s) = \inf_{\tau} \mathbb{P}_s^{\sigma, \tau} (\exists n, S_n \in \mathbf{T})$$

in polynomial time.

Algorithm

Enumerate all positional strategies and choose the best one.

Other known algorithms have the same complexity in the worst case.

Proposition (Condon)

Deciding whether $\text{val}(s) \geq \frac{1}{2}$ is in $NP \cap co-NP$.

Is this decision problem in P?

Beyond reachability games

Every Borel-measurable event $W \subseteq \mathbf{S}^\omega$ defines a class of stochastic games.

Example

- Parity games: states labeled by priorities $c : \mathbf{S} \rightarrow \{0, \dots, d\}$ and

$$W = \{ \limsup c(s_n) \text{ is even} \}$$

- Positive average games: states labeled by rewards $r : \mathbf{S} \rightarrow \mathbb{R}$ and

$$W = \{ \limsup \frac{r(s_0) + \dots + r(s_n)}{n+1} \geq 0 \}$$

- Exotic games: $W =$ "the play is ultimately periodic".

A link with parity games

Proposition (Jurdzinski)

Deciding whether the winner in a parity game is in $UP \cap co-UP$.

Is this decision problem in P?

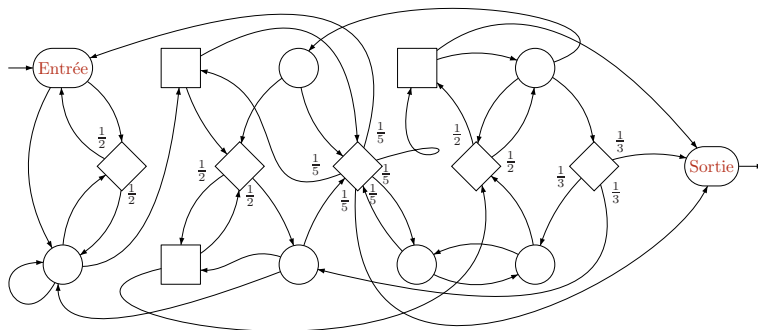
This question is **easier** to answer for parity games than for simple stochastic games.

Reduction of a parity game to a simple stochastic game:

- a state with a high good priority d goes with high probability p_d to a target state.
- a state with a high bad priority d goes with high probability to p_d an absorbing sink state.

If the probabilities p_d are well chosen, the parity game has positive value if and only if the simple stochastic game has positive value.

Yesterday



No player: Markov chain, linear equations.

One player: Markov Decision Problem, linear programming.

Two players: Simple Stochastic Games, strategy enumeration.

Outline

- 1 Stochastic games with perfect information
 - Markov chains and probability measures
 - Simple stochastic games
 - A few results about stochastic games with perfect information
- 2 Stochastic games for economists
 - Games in strategic form
 - Discounted games (Shapley)
 - Undiscounted games (Mertens Neyman)
- 3 Determinacy of Blackwell games

Pure strategies for games with perfect information

Theorem (Corollary of the second Martin theorem)

For every stochastic game with perfect-information and a Borel-measurable objective $W \subseteq \mathbf{S}^\omega$ and pure strategies are sufficient:

$$\text{val}_*(s) = \sup_{\sigma} \inf_{\tau} \mathbb{P}_s^{\sigma, \tau}(W) = \inf_{\tau} \sup_{\sigma} \mathbb{P}_s^{\sigma, \tau}(W) = \text{val}^*(s) ,$$

Remark

Existence of the value and ε -optimal strategies: for every $\varepsilon > 0$ there exists σ_ε such that for every τ ,

$$\mathbb{P}_s^{\sigma_\varepsilon, \tau}(W) \geq \text{val}(s) - \varepsilon .$$

What about . . .

- *the existence of (0-)optimal strategies?*
- *the memory needed by optimal strategies?*
- *the computability of the values?*

A transfer theorem

Theorem (G., Zielonka, 07)

Let $W \subseteq S^\omega$ be a winning condition such that in every **Markov decision process** equipped with W or $S^\omega \setminus W$, the maximizer has an **optimal positional strategy**.

Then in every stochastic game with **perfect information** and winning condition W , both players have **optimal positional strategies**.

- transfers positionality from one-player games to two-player games.
- applies to positive average games, parity games, reachability games and many others.

A sufficient criterion for positionality

Definition

A winning condition $W \subseteq S^\omega$ is **submixing** if for every sequence w_0, w_1, w_2, \dots of finite non-empty words of S^* ,

$$\begin{cases} w_0 w_2 w_4 \dots \notin W \\ w_1 w_3 w_5 \dots \notin W \end{cases} \implies w_0 w_1 w_2 w_3 w_4 w_5 \dots \notin W .$$

Mixing two losing plays gives a losing play.

Definition

W is **tail** if for every $u \in S^\omega$, $u \in W$ iff some suffix of u is in W .

Theorem (G. 07)

If W is **both tail and submixing** then in every Markov decision process equipped with W , the maximizer has a **positional optimal** strategy.

Applies to positive average games, parity games, reachability games and many others.

About tail games

Theorem (G., Horn, 10)

In every stochastic game with perfect information and a tail winning condition, both players have (0-)optimal strategies.

Proof.

For ε sufficiently small, an ε -optimal strategy can be transformed into an optimal strategy. □

Theorem (G.,Horn, 10)

Let W be a tail winning condition.

*Suppose **there exists an algorithm to decide in every W -game whether player Max has a strategy to win with probability 1 (almost-surely).***

*Then **values** of W -games are **computable**.*

*Moreover **no more memory is required** for playing **optimally** than for winning with probability 1.*

Outline

- 1 Stochastic games with perfect information
 - Markov chains and probability measures
 - Simple stochastic games
 - A few results about stochastic games with perfect information
- 2 Stochastic games for economists
 - Games in strategic form
 - Discounted games (Shapley)
 - Undiscounted games (Mertens Neyman)
- 3 Determinacy of Blackwell games

Games in strategic form

- Matrix games, one-shot games, popular in economics.
- Emile Borel and John von Neumann.

Definition

Set of actions I for player 1 and J for player 2. Matrix of payoffs $(a_{i,j})_{i \in I, j \in J}$.

How to play:

- simultaneously, player 1 chooses action $i \in I$ and player 2 chooses action $j \in J$
- then player 1 gives $a_{i,j}$ to player 2.

Example

	L	R
T	2	-1
B	0	1

Mixed strategies

Example

	<i>L</i>	<i>R</i>
<i>T</i>	2	-1
<i>B</i>	0	1

- 1 What maximal payoff can player 1 expect?
- 2 Playing always the same action is a bad idea.
- 3 Player 1 plays *T* with probability p_T and *B* with probability $p_B = (1 - p_T)$.

$$\frac{p_T \cdot T + p_B \cdot B}{2p_T \quad -p_T + (1 - p_T)}$$

- 4 If $p_T = \frac{1}{4}$ then player 2 is indifferent between *L* and *R*.
 Similarly, if $p_L = \frac{1}{2}$ then player 1 is indifferent between *T* and *B*.
- 5 Then strategies $x = (\frac{1}{4}, \frac{3}{4})$ and $y(\frac{1}{2}, \frac{1}{2})$ form an **equilibrium** with payoff $\frac{1}{2}$.

Borel-von Neumann Theorem

Theorem

For every matrix $A \in \mathbb{R}^{I \times J}$

$$\sup_{x \in \mathcal{D}(I)} \inf_{y \in \mathcal{D}(J)} x^t \cdot A \cdot y = \inf_{y \in \mathcal{D}(J)} \sup_{x \in \mathcal{D}(I)} x^t \cdot A \cdot y .$$

This common value is called the **value** of the matrix game.

There exists **optimal strategies** x^* and y^* which realize the maxima and minima.

Remark (Proof history)

- First ad-hoc proof by von Neumann, topology and functional calculus.
- Second proof Brouwer fixpoint theorem.
- Third proof by Kakutani fixpoint theorem.
- Fourth proof, elementary, by J. Ville "Traité du calcul des probabilités et ses applications", "applications aux jeux de hasard". Proof chosen by von Neumann for his famous book "Games and economic behaviour".

Theorem (Borel, von Neumann)

For every matrix $A \in \mathbb{R}^{I \times J}$

$$\sup_{x \in \mathcal{D}(I)} \inf_{y \in \mathcal{D}(J)} x^t \cdot A \cdot y = \inf_{y \in \mathcal{D}(J)} \sup_{x \in \mathcal{D}(I)} x^t \cdot A \cdot y .$$

This common value is called the **value** of the matrix game.

There exists **optimal strategies** x^* and y^* which realize the sup and inf.

Proof.

It is sufficient to prove:

$$\text{either } \exists y \in \mathcal{D}(J), A \cdot y \leq 0 \text{ or } \exists x \in \mathcal{D}(I), x^t \cdot A > 0.$$

Let $C \subseteq \mathbb{R}^I$ be the cone generated by columns $C_j = (a_{i,j})_{i \in I}$ and elementary columns $C_j = (1_{i=j})_{i \in I}$.

- If $(0)_{i \in I}$ is in the cone C then one obtains y .
- Otherwise there exists a separating hyperplane (Hahn-Banach separation theorem) between C and $(0)_{i \in I}$, which gives x .

Computation of the value

Proposition

The value v of a matrix game $(A_{i,j})_{I \times J}$ is the optimal solution to:

maximize v under constraints

$$x_i \geq 0 \quad (i \in I)$$

$$\sum_{i \in I} x_i = 1$$

$$\sum_{i \in I} a_{ij} x_i \geq v \quad (j \in J)$$

Linear programming came after the result of Borel and von Neumann. Von Neumann developed the theory of duality in linear programming.

Non-zero sum case

Example

	<i>L</i>	<i>R</i>
<i>T</i>	2, 1	0, 0
<i>B</i>	0, 0	1, 2

Equilibrium $x = \frac{2}{3} \cdot T + \frac{1}{3} \cdot B$ and $y = \frac{1}{3} \cdot L + \frac{2}{3} \cdot R$.

Definition (Nash equilibrium)

Non zero-sum matrix game specified by two $I \times J$ matrices A_1, A_2 . Then $(x_*, y_*) \in \mathcal{D}(I) \times \mathcal{D}(J)$ is a **Nash-equilibrium** if $\forall x \in \mathcal{D}(I), \forall y \in \mathcal{D}(J)$:

$$\begin{aligned}x^t \cdot A_1 \cdot y_* &\leq x_*^t \cdot A_1 \cdot y_* \\x_*^t \cdot A_2 \cdot y_* &\leq x_*^t \cdot A_2 \cdot y_* .\end{aligned}$$

No player has an incentive to deviate.
In general, arbitrary number of players.

Theorem (Nash)

Every non-zero sum matrix game has a Nash equilibrium.

Proof.

Kakutani's theorem.

Theorem (Kakutani)

*Let S be a non-empty, compact and convex subset of some Euclidean space E . Let $\phi : S \rightarrow 2^S$ be a set-valued function on S with a **closed graph** and the property that $\phi(x)$ is **non-empty** and **convex** for all $x \in S$. Then ϕ **has a fixed point** ($x \in \phi(x)$).*

Apply Kakutani's theorem to:

$$\phi(x, y) \rightarrow \{(x', y') \mid x'^t \cdot A \cdot y = \max_{x''} x''^t \cdot A \cdot y \text{ and } x^t \cdot A \cdot y' = \min_{y''} x^t \cdot A \cdot y''\} .$$

Any fixpoint is an equilibrium. □

Follows previous proof of von Neumann.

Quoted from an interview of John Nash (after translation).

”

–After your license, you worked for the RAND Corporation (. . .).

–Yes, for three summers. It was sponsored by the Air Force and was one of the indirect modalities that the government was using to finance research: instead of giving the money directly to scientists, the government was giving money to the army, who in turn was giving money to the scientists.

”

Outline

- 1 Stochastic games with perfect information
 - Markov chains and probability measures
 - Simple stochastic games
 - A few results about stochastic games with perfect information
- 2 Stochastic games for economists
 - Games in strategic form
 - Discounted games (Shapley)
 - Undiscounted games (Mertens Neyman)
- 3 Determinacy of Blackwell games

Shapley: the birth of stochastic games

Definition (Stochastic games)

Two players 1 and 2 with sets of actions I and J .

Finitely many states \mathbf{S} .

With each state s and actions i, j are associated:

- **transition probabilities** to other states $p(t \mid s, i, j)$.
The probability $\mu(s, i, j) = 1 - \sum_t p(t \mid s, i, j)$ is strictly positive, this is the probability that the game stops.
- the **reward** $a(s, i, j) \in \mathbb{R}$ taken from player 2 by player 1.

The discounted Big Match

Example (The discounted Big Match)

	A_2	B_2
A_1	1	0
B_1	0^*	1^*

Player 1 tries to predict the choice of player 2, and earns one point each time he is right.

It goes on and on as long as player 1 predicts A.

If someday player 1 predicts B , all the future choices of both players are bound to be the same forever: if player 1 was right (player 2 played B_2), he will be right forever otherwise (player 2 played A_2) he will be wrong forever.

Let $0 \leq \lambda < 1$ a fixed **continuation probability**.

States $\mathbf{S} = \{0^*, s, 1^*\}$ and **actions** $I = \{A_1, B_1\}$ and $J = \{A_2, B_2\}$.

Rewards: $a(s, \cdot, A_1) = 0$, $a(s, \cdot, A_2) = 1$, $a(0^*, \cdot, \cdot) = 0$ and $a(1^*, \cdot, \cdot) = 1$.

Non-zero transition probabilities: $p(s, A_1, \cdot)(s) = \lambda$, $p(s, A_2, B_1)(1^*) = \lambda$, $p(s, A_2, B_2)(0^*) = \lambda$, $p(0^*, \cdot, \cdot)(0^*) = \lambda$ and $p(1^*, \cdot, \cdot)(1^*) = \lambda$.

Characterization of values

Proposition (Shapley)

The operator $\Phi : \mathbb{R}^S \rightarrow \mathbb{R}^S$ defined by:

$$\phi(v)(s) = \text{val} \left(\left[a(s, i, j) + \sum_t p(t | s, i, j) \cdot v(t) \right]_{i,j} \right),$$

has a unique fixpoint.

This fixpoint is the value of the discounted stochastic game.

Players have optimal stationary strategies.

The discounted Big Match

Example

The Shapley operator is:

$$\phi(v)(s) = \text{val} \left(\left[\begin{array}{cc} 0 + \lambda \cdot v & 1 + \lambda \cdot v \\ \frac{1}{1-\lambda} = \lambda + \lambda^2 + \dots & 0 \end{array} \right] \right)$$

Player 1 plays top with probability $0 < x < 1$.

Indifference of player 2:

$$\lambda \cdot v \cdot x + \frac{1}{1-\lambda} (1-x) = (1 + \lambda \cdot v) \cdot x$$

$$\implies x = \lambda$$

$$\implies v = \lambda + \lambda^2 \cdot v \implies v = \frac{\lambda}{1-\lambda^2}.$$

Discounted value:

$$v_\lambda = (1-\lambda) \cdot v = \frac{\lambda}{1+\lambda} \xrightarrow{\lambda \rightarrow 1} \frac{1}{2}.$$

Proof of Shapley theorem

$$\phi(v)(s) = \text{val} \left(\left[a(s, i, j) + \sum_t p(t | s, i, j) \cdot v(t) \right]_{i,j} \right),$$

Proof.

The operator is contracting with factor $\lambda = \max_{s,i,j} \sum_t p(t | s, i, j) < 1$.

Unique fixpoint v_λ .

Stationary optimal strategy: in state s , play optimally in the matrix game:

$$\left[a(s, i, j) + \sum_t p(t | s, i, j) \cdot v(t) \right]_{i,j}.$$

By induction, for every $n \in \mathbb{N}$, this ensures payoff at least $v_\lambda - \lambda^n \cdot K$ over the first n rounds. □

Tarski's theorems

Theorem (Tarski)

To any formula $\Phi(X_1, \dots, X_m)$ in the vocabulary $\{0, 1, +, \cdot, <\}$ one can effectively associate two objects:

- i) a quantifier free formula $\bar{\Phi}(X_1, \dots, X_m)$ in the same vocabulary,
- ii) a proof of the equivalence $\Phi \leftrightarrow \bar{\Phi}$ that uses only the axioms for real closed fields (ordered fields with no ordered proper algebraic extension, i.e. the intermediate value property for polynomials).

Corollary (Tarski)

First order theory over $(\mathbb{R}, +, \cdot, \geq)$ is decidable.

Semi-algebraic sets (defined by polynomial inequalities) are closed under projection.

Calculability of Shapley games

Proposition

Being a fixpoint of:

$$\phi(v)(s) = \text{val} \left(\left[a(s, i, j) + \sum_t p(t | s, i, j) \cdot v(t) \right]_{i,j} \right),$$

is expressible by FOT with depth-one existential quantification.

Proof.

Quantify over the existence of stationary strategies $\sigma : \mathbf{S} \rightarrow \mathcal{D}(I)$ and $\tau : \mathbf{S} \rightarrow \mathcal{D}(J)$ that ensure and defend payoff $v(s)$ from state s . □

Corollary

It is decidable in exponential time whether the value of a discounted game is greater than 0.

Bewley-Kohlberg algebraic approach to Shapley games

Proposition

Replace the stopping probabilities by a discount factor $0 \leq \lambda < 1$. Being a fixpoint of:

$$\phi_\lambda(v)(s) = \text{val} \left(\left[(1 - \lambda)a(s, i, j) + \lambda \sum_t p(t | s, i, j) \cdot v(t) \right]_{i,j} \right),$$

is expressible by FOT with depth-one existential quantification. Let $v_\lambda : \mathbf{S} \rightarrow \mathbb{R}$ be the unique fixpoint of ϕ_λ .

Corollary

The mapping:

$$\lambda \rightarrow v_\lambda$$

is semi-algebraic.

Puiseux series

Definition

A Puiseux series is a series with negative and fractional exponents:

$$\sum_{k=k_0}^{+\infty} c_k T^{\frac{k}{n}}, k_0 \in \mathbb{Z}.$$

Theorem (Newton-Puiseux theorem)

The space of Puiseux series is a real closed field.

Corollary

A semi-algebraic function has an expansion near 0 in Puiseux series.

Bewley-Kohlberg algebraic approach to Shapley games

Corollary

The mapping:

$$\text{val} : \lambda \rightarrow v_\lambda$$

is semi-algebraic.

Corollary

$\lim_{\lambda \rightarrow 1} v_\lambda$ *exists.*

Proof.

The mapping val has an expansion near 0 in Puiseux series over the variable $(1 - \lambda)$. Because v_λ is uniformly bounded by $\max_{s,i,j} r(s,i,j)$, the coefficients of negative powers are zero. Hence the limit is the constant term of the series. □

Outline

- 1 Stochastic games with perfect information
 - Markov chains and probability measures
 - Simple stochastic games
 - A few results about stochastic games with perfect information
- 2 Stochastic games for economists
 - Games in strategic form
 - Discounted games (Shapley)
 - Undiscounted games (Mertens Neyman)
- 3 Determinacy of Blackwell games

The Big Match

	A_2	B_2
A_1	1	0
B_1	0^*	1^*

Player 1 tries to predict the choice of player 2, and earns one point each time he is right.

It goes on and on as long as player 1 predicts A .

If someday player 1 predicts B , all the future choices of both players are bound to be the same forever: if player 1 was right (player 2 played B_2), he will be right forever otherwise (player 2 played A_2) he will be wrong forever.

The Big Match (Gillette)

	A_2	B_2
A_1	1	0
B_1	0^*	1^*

Discounted payoff.

One can check that $v_\lambda = \frac{1}{2}$. A stationary optimal strategy for player 1 consists in playing B with probability $x_\lambda = \frac{\lambda}{1+\lambda}$.

Average payoff.

If player 2 plays each action with equal probability, he guarantees himself a loss of at most $\frac{1}{2}$ **on average**.

Can player 1 guarantee payoff at least $\frac{1}{2}$, on average?

Problem: if player 1 plays always top, he loses. He should play bottom from times to times. But if the probability is fixed in advance, his adversary will simply be patient.

The big match: the solution

	A_2	B_2
A_1	1	0
B_1	0^*	1^*

Theorem (Blackwell and Ferguson 68)

For every integer K , player 1 has a strategy which guarantees the payoff $\frac{K}{2(K+1)}$ on average.

Proof.

After having seen the sequence of rewards $a_0 a_1 a_2 \dots a_n$, player 1 plays T with probability

$$\frac{1}{1 + \max(1, M + \sum_{i=0}^n (a_i - \frac{1}{2}))} ,$$

with M large enough. □

Mertens-Neyman theorem

Definition (Uniform value)

A stochastic game has uniform value v_∞ if for every $\varepsilon > 0$ there exists $\sigma^\sharp, \tau^\sharp$ and $N \in \mathbb{N}$ such that:

- 1 when player 1 plays σ^\sharp , the average payoff after from the N -th step is greater than $v_\infty - \varepsilon$,
- 2 when player 2 plays τ^\sharp , the average payoff after more than N steps is less than $v_\infty + \varepsilon$.

Theorem (Mertens and Neyman)

A stochastic game with finite state and action space has a uniform value v_∞ . This value is the limit of discounted values v_λ when λ converges to 1.

$$\sup_{\sigma} \inf_{\tau} \mathbb{E}^{\sigma, \tau} \left[\liminf_n \frac{1}{n+1} \sum_{i=0}^n a_i \right] = \inf_{\tau} \sup_{\sigma} \mathbb{E}^{\sigma, \tau} \left[\limsup_n \frac{1}{n+1} \sum_{i=0}^n a_i \right].$$

Back to Shapley's seminal work

Theorem

A stochastic game with finite state and action space has a uniform value v_∞ . This value is the limit of discounted values v_λ when λ converges to 1.

Question: in the original paper of Shapley, there was no discount factor, but stopping probabilities which depend on the current state s , while λ is fixed. What can be said when λ depends on the current state?

Partial answer: when the stopping probabilities converge to 0 simultaneously or after another, it is possible to recover parity games at the limit (de Alfaró and Majumdar) as well as mixtures of parity and mean-payoff games (Gimbert and Zielonka). To be continued. . .

Blackwell games

Definition

A, B be two alphabets and $f : (AB)^\omega \rightarrow \mathbb{R}$ bounded and Borel-measurable.
At step n , players 1 and 2 choose simultaneously actions a_n and b_n .
After play $a_0 b_0 a_1 b_1 \dots$, player 2 gives $f(a_0 b_0 a_1 b_1 \dots)$ to player 1.

Definitions

Strategy for player 1: $\sigma : (AB)^* \rightarrow \mathcal{D}(A)$

Strategy for player 2: $\tau : (AB)^* \rightarrow \mathcal{D}(B)$

Expected payoff: $\mathbb{E}^{\sigma, \tau} [f]$ the expected value of f in the Markov chain induced by σ and τ .

Example

Discounted games

Undiscounted games

Big Match

Martin's second determinacy theorem

Theorem (Martin's second determinacy theorem 98)

For every bounded Borel-measurable $f : (AB)^\omega \rightarrow \mathbb{R}$,

$$\sup_{\sigma} \inf_{\tau} \mathbb{E}^{\sigma, \tau} [f] = \inf_{\tau} \sup_{\sigma} \mathbb{E}^{\sigma, \tau} [f] .$$

Corollary (Weaker form of Mertens-Neyman theorem (81))

Stochastic games with undiscounted payoff have a value:

$$\sup_{\sigma} \inf_{\tau} \mathbb{E}^{\sigma, \tau} \left[\liminf_n \frac{\sum_{i=0}^n a_i}{n+1} \right] = \inf_{\tau} \sup_{\sigma} \mathbb{E}^{\sigma, \tau} \left[\limsup_n \frac{\sum_{i=0}^n a_i}{n+1} \right] .$$

Sketch of proof

Proof.

Reduction to a Gale-Stewart game. Assume $f : (AB)^\omega \rightarrow (0, 1]$.

Actions of player 1: valuation $h : A \times B \rightarrow [0, 1]$.

Actions of player 2: pairs of actions $A \times B$.

- Player 1 chooses a valuation $h_0 : A \times B \rightarrow [0, 1]$,
- player 2 chooses a move $a_0 b_0$,
- player 1 chooses $h_1 : A \times B \rightarrow [0, 1]$,
- 2 chooses $a_1 b_1$ and so on...

Let v_0 be arbitrary and $v_{n+1} = h_n(a_n b_n)$.

Constraints:

- $\text{val}[h_{n+1}(a, b)]_{a, b \in A \times B} \geq v_{n+1}$.
- $v_{n+1} > 0$.



Proof.

Actions of player 1: valuation $h : A \times B \rightarrow [0, 1]$.

Actions of player 2: pairs of actions $A \times B$.

Let $v_0 = v$ and $v_{n+1} = h_n(a_n b_n)$.

Constraints:

- $\text{val}[h_{n+1}(a, b)]_{a, b \in A \times B} \geq v_{n+1}$.
- $v_{n+1} > 0$.

Rule: A play is won by player 1 if and only if:

$$\limsup_n v_n \leq f(a_0 b_0 a_1 b_1 a_2 \cdots) .$$

Proposition

Either player 1 or player 2 wins this Gale-Stewart game G_v .

If player 1 wins G_v then player 1 has a strategy σ^\sharp to ensure $\mathbb{E}^{\sigma^\sharp, \tau} [f] \geq v$ in the Blackwell game.

If player 2 wins G_v then for every $\varepsilon > 0$, player 2 has a strategy τ^\sharp to ensure $\mathbb{E}^{\sigma, \tau^\sharp} [f] \leq v + \varepsilon$ in the Blackwell game.

Conclusion

Summary:

- various games, various algorithmic problems
- mathematical programming: linear algebra, linear programming, FOT.
- general determinacy results: Martin's theorems.
- uniform determinacy results: Shapley, Bewley-Kohlberg, Mertens-Neyman.

Research directions:

- solving Markov decision processes with a simple algorithm,
- solving Simple Stochastic Games in polynomial time,
- Mertens-Neyman theorems for multiple discount factors,
- explore games with partial observation and probabilistic automata. . .