

**How to Discount
or
New Games to Play**

Hugo Gimbert and Wiesław Zielonka

LIAFA, Université Denis Diderot, Paris

Stochastic games

Two players : **0** (maximizer) and **1** (minimizer)

S - a finite set of states,

for each state $s \in S$, $A^0(s)$ and $A^1(s)$ are finite set of states of player 0 and player 1,

$p(s' \mid s, a^0, a^1)$ — probability of going to the state s' if the current state is s and player 0 and 1 have chosen $a^0 \in A^0(s)$ and $a^1 \in A^1(s)$ respectively.

$$\sum_{s' \in S} p(s' \mid s, a^0, a^1) = 1$$

$r(s, a^0, a^1)$ - immediate reward (in \mathbb{R} or in \mathbb{N}) if at the state s actions a^0 and a^1 are chosen.

A play :

$$s_0, (a_0^0, a_0^1), s_1, (a_1^0, a_1^1), s_2, (a_2^0, a_2^1), s_3, (a_3^0, a_3^1), \dots$$

$\forall i, (a_i^0, a_i^1) \in A^0(s_i) \times A^1(s_i)$, yields an infinite sequence of immediate rewards:

$$r(s_0, a_0^0, a_0^1), r(s_1, a_1^0, a_1^1), r(s_2, a_2^0, a_2^1), r(s_3, a_3^0, a_3^1), \dots$$

payoff function:

$$u : \mathbb{R}^\omega \rightarrow \mathbb{R}$$

$u(r_0 r_1 r_2 \dots)$ the amount player 0 pays to player 1

Payoffs

Game theory and economics:

mean payoff:

$$u(r_0 r_1 r_2 \dots) = \limsup_n \frac{1}{n} \sum_{0 \leq i < n} r_i$$

discounted ($0 < \lambda < 1$):

$$u_\lambda(r_0 r_1 r_2 \dots) = (1 - \lambda) \sum_{0 \leq i} \lambda^i r_i$$

computer science:

parity ($r_i \in \mathbb{N}$):

$$u(r_0 r_1 r_2 \dots) = \begin{cases} 1 & \text{if } \limsup_n r_n \text{ is even,} \\ 0 & \text{otherwise} \end{cases}$$

Perfect Information Games

$S = S_0 \cup S_1$ — partition of states onto the states of player 0 and player 1.

For each state $s \in S$,

$A(s)$ — the set of actions available at s .

$p(s'|s, a)$ — probability of going to s' if the current state is s and $a \in A(s)$ is executed.

$$\sum_{s' \in S} p(s'|s, a) = 1$$

$r(s, a)$ — immediate reward.

Deterministic games

Perfect information games such that for all $s \in S$ and $a \in A(s)$ there is exactly one state s' chosen with probability 1.

What is the aim of each player?

Player 0 tries to **minimize** the payment expectation,
player 1 tries to **maximize** the payment expectation.

Strategies

In general can depend on all past history.

Positional strategies — depend only on the current state.

Pure strategies — deterministic strategies, choose one available action with probability 1

Pure positional strategies

Fixing strategies σ_0 and σ_1 and an initial state s determines a unique **probability measure** $\mathbb{P}_{s,\sigma_0,\sigma_1}$ over the set of all **plays**.

The **gain** of player 1 is the expectation of the payoff mapping u :

$$\mathbb{E}_{s,\sigma_0,\sigma_1}(u) = \int u(p) \mathbb{P}_{s,\sigma_0,\sigma_1}(dp)$$

Optimal strategies

The strategies $\sigma_0^\#$ and $\sigma_1^\#$ are **optimal** if for all states s and all strategies σ_0 and σ_1 of players Min and Max:

$$\mathbb{E}_{s, \sigma_0^\#, \sigma_1} (u) \leq \mathbb{E}_{s, \sigma_0^\#, \sigma_1^\#} (u) \leq \mathbb{E}_{s, \sigma_0, \sigma_1^\#} (u)$$



Game Value

What Is Known?

1. (Shapley 53) Discounted stochastic games have a value, both players have positional optimal strategies.
2. (Bowley, Kohlberg 76, Mertens, Neyman 81) Mean-payoff stochastic games have a value and

$$\lim_{\lambda \nearrow 1} \text{valOfDiscounted}_\lambda = \text{valOfMeanPayoff}$$

3. (de Alfaro, Majumdar 01) Stochastic parity games have a value.
4. Perfect information : both players have optimal pure positional strategies.

Relationship between parity and mean-payoff games?

Yes, for deterministic games. Apparently non for stochastic games.

Discounted Parity Games?

Luca de Alfaro, Thomas A. Henzinger and Rupak Majumdar, Discounting the Future in Systems Theory, ICALP 2003

Parity Games Revisited

priorities:	0	1	2	3	4	5	6
winning player:	0	1	0	1	0	1	0

Binary Priority Games

Choose the maximal priority visited infinitely often, the player associated with this priority becomes the winner.

priorities:	0	1	2	3	4	5	6
winning player:	1	1	0	0	0	1	1

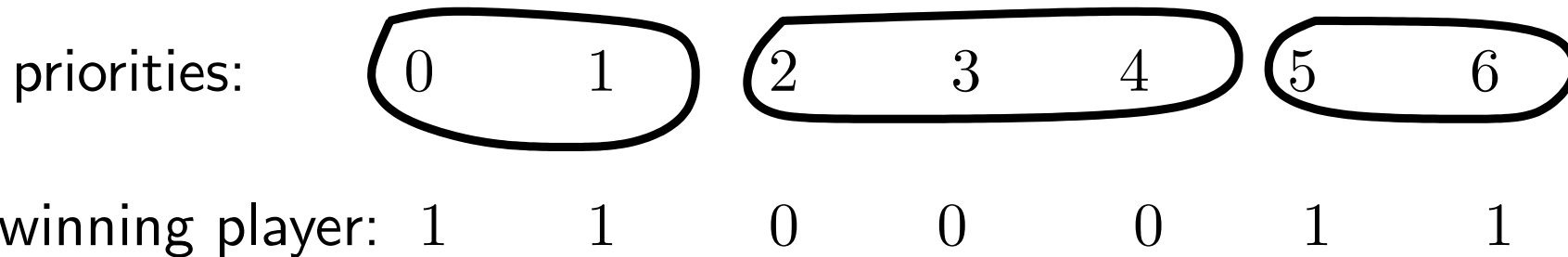
Binary Priority Games = Parity Games

Binary Priority Game

priorities:	0	1	2	3	4	5	6
winning player:	1	1	0	0	0	1	1

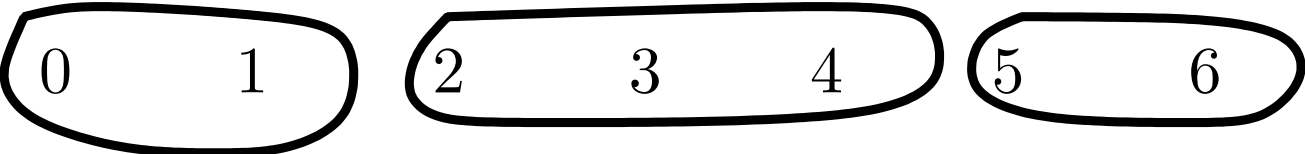
Binary Priority Games = Parity Games

Binary Priority Game



Binary Priority Games = Parity Games

Binary Priority Game to Parity Game

priorities: 

winning player: 1 1 0 0 0 1 1

new priorities: 1 2 3

new winning player: 1 0 1

From Winning Player to Binary Payoff Mapping

Binary Priority Game

priorities: 0 1 2 3 4 5 6

~~winning player:~~

1 1 0 0 0 1 1

payoff for
player 1

From Binary Payoffs to any Payoffs

Simple Priority Game

priorities:	0	1	2	3	4	5	6
payoff for player 1	25	-3	0	6	1	6	-4

Solving Simple Priority Deterministic Games

priorities:

0 1 2 3 4 5 6

payoff for
player 1

25 -3 0 6 1 6 -4

1 0 0 0 0 0 0

binary
priority
games

1 0 0 1 0 1 0

1 0 0 1 1 1 0

1 0 1 1 1 1 0

1 1 1 1 1 1 0

From Simple Priority Games to Mean Payoff Priority Games (1)

Immediate reward mapping:

for $s \in S$ and $(a^0, a^1) \in A^0(s) \times A^1(s)$:

$$r(s, a^0, a^1) \in \mathbb{R} \times \mathbb{N}$$

$$r(s, a^0, a^1) = (\text{reward}(s, a^0, a^1), \text{priority}(s, a^0, a^1))$$

From Simple Priority Games to Mean Payoff Priority Games (2)

Simple Priority Games:

for all “transitions” (s, a^0, a^1) :

the same priority \implies the same reward

$(r_0, p_0), (r_1, p_1), \dots$ - the sequence of immediate rewards,

let p maximal priority visited infinitely often

the subsequence with $p_i = p$ is constant – the most trivial and dull sequence

Allow different rewards for the same priority

Priority Mean Payoff Games

$(r_0, p_0), (r_1, p_1), \dots$ - the sequence of immediate rewards,

Let p the maximal priority visited infinitely often in the sequence above.

$i_0 < i_1 < i_2 < i_3 < \dots$ the infinite sequence of all indices such that $p = p_{i_0} = p_{i_1} = p_{i_2} = \dots$

The corresponding subsequence of immediate rewards $r_{i_0} r_{i_1} r_{i_2} r_{i_3} \dots$ is not constant.

Player 1 wins

$$\limsup_n \frac{1}{n} \sum_{l=0}^{n-1} r_{i_l}$$

Multi-discounted Games

$$r(s, a^0, a^1) = (\text{reward}(s, a^0, a^1), \lambda(s, a^0, a^1))$$

$$\text{reward}(s, a^0, a^1) \in \mathbb{R}$$

$$\lambda(s, a^0, a^1) \in [0; 1)$$

$(r_0, \lambda_0), (r_1, \lambda_1), (r_2, \lambda_2) \dots$ the sequence of immediate rewards.

Player 1 receives:

$$\sum_{i=0}^{\infty} (1 - \lambda_i) \lambda_0 \cdots \lambda_{i-1} r_i$$

From Multi-discounted Games to Mean Payoff Priority Games

Suppose that $\Lambda = \{\lambda_0, \dots, \lambda_k\}$ a finite set of discount parameters (variables) and

$$\lambda(s, a^0, a^1) \in \Lambda$$

$$u_\Lambda = \sum_{i=0}^{\infty} (1 - \lambda_i) \lambda_0 \cdots \lambda_{i-1} r_i$$

is a function of variables Λ .

Theorem 1. *We consider perfect information games with the same state and action space.*

The first game $\Gamma(\lambda_0, \dots, \lambda_k)$ is a multi-discount game with discount factors $\lambda_0, \dots, \lambda_k$ and the reward mapping $r = (\text{reward}, \lambda)$.

The second game is a priority mean payoff game with the reward mapping $r' = (\text{reward}, \text{priority})$ and such that

$$\lambda(s, a^0, a^1) = \lambda_i \quad \text{iff} \quad \text{priority}(s, a^0, a^1) = i$$

Then

$$\lim_{\lambda_k \rightarrow 1} \dots \lim_{\lambda_0 \rightarrow 1} \text{val}(\lambda_k, \dots, \lambda_0) = \text{val}_{mean}$$

where $\text{val}(\lambda_k, \dots, \lambda_0)$ is the value of the multi-discount game and val_{mean} is the value of the priority mean payoff game. Both players have optimal pure positional strategies for priority mean payoff games with a finite state space.

