# Zero-sum Stochastic Games

Jérôme Renault, TSE Université Toulouse

Stochastic Methods in Game Theory, Singapore 2015

# Zero-sum stochastic games

1. The basic model: finitely many states and actions

1.1 Description

1.2 The $n$-stage game and the $\lambda$-discounted game

1.3 Limit value, the algebraic approach

1.4 The uniform value

2. A few extensions and recent results

2.1 Counterexamples : A simple compact continuous game with no limit value. A hidden stochastic game with no limit value

2.2. 1-player games (Dynamic programming, Gambling Houses and MDP): General results. The compact non expansive case.

2.3 Uniform convergence of $(v_n)_n$ and $(v_\lambda)_\lambda$ are equivalent.

2.4. Repeated Games with incomplete information : Lack of information on one side, repeated games with an informed controller. Lack of information on both sides.

2.5. Some open problems.

# Description

A zero-sum game is a 2-player game where players have opposite interests. Given a real-valued matrix $A = (a_{i,j})$ in $\mathbb{R}^{I \times J}$, we denote the value of $A$ by $\mathrm{val}(A)$ or $\mathrm{val}_{\Delta(I)\times\Delta(J)}(A)$:

$$\mathrm{val}(A) = \max_{x\in\Delta(I)} \min_{y\in\Delta(J)} \sum_{i\in I}\sum_{j\in J} x_i y_j a_{i,j} = \min_{y\in\Delta(J)} \max_{x\in\Delta(I)} \sum_{i\in I}\sum_{j\in J} x_i y_j a_{i,j}$$

A zero-sum stochastic game is a dynamic zero-sum game (with a Markovian structure, played in discrete time).

Basic model (Shapley, 1953): a set of states $K$ with an initial state $k_1$, a set of actions $I$ for player 1, a set of actions $J$ for player 2, a payoff function $g : K \times I \times J \longrightarrow \mathbb{R}$, and a transition $q : K \times I \times J \longrightarrow \Delta(K)$. In the basic model, $K$, $I$ and $J$ are assumed to be non empty *finite* sets.

Progress of the game:

- stage 1: players simultaneously choose $i_1 \in I$ and $j_1 \in J$. $i_1$ and $j_1$ are publicly announced, and P1's stage payoff is $g(k_1, i_1, j_1)$.
- stage $t \geq 2$ : $k_t$ is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, and announced to both players. Players then simultaneously choose $i_t \in I$ et $j_t \in J$. The actions $i_t$ et $j_t$ are announced, and P1's payoff is $g(k_t, i_t, j_t)$.

Notations and vocabulary.

$q(k'|k,i,j)$: probability that the state of stage $t+1$ is $k'$ if the state of stage $t$ is $k$ and $i$ and $j$ are played at that stage.
State $k$ is absorbing if $q(k|k,i,j) = 1$ for all $(i,j)$ in $I \times J$.
A stochastic game is absorbing if it has a unique non absorbing state.

Example 1:

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

A play of the game is a sequence $(k_1, i_1, ...., k_t, i_t, j_t, ...)$ in $(K \times I \times J)^\infty$.
A history of the game is a finite sequence $(k_1, i_1, j_1, ...., k_{t-1}, i_{t-t}, j_{t-1}, k_t)$ in $(K \times I \times J)^{t-1} \times K$ for some positive integer $t$, representing the information available to the players before they play at stage $t$.

A (behavior) strategy of player 1, resp. player 2, associates to every history a mixed action in $\Delta(I)$, resp. $\Delta(J)$, to be played in case this history occurs. A strategy of a player is pure if it associates to each history a Dirac measure, i.e. a pure action in $I$ or $J$.

Denote by $\Sigma$ and $\mathscr{T}$ the sets of strategies of player 1 and 2, respectively. A couple of strategies in $\Sigma \times \mathscr{T}$ induces a probability distribution $I\!P_{k_1,\sigma,\tau}$ over the set of plays, endowed with the product $\sigma$-algebra.

Remark: A mixed strategy of a player is a probability distribution over his set of pure strategies (endowed with the product $\sigma$-algebra). By Kuhn's theorem (Aumann, 1962), one can show that mixed strategies and behavior strategies are equivalent here.

# The $n$-stage game and the $\lambda$-discounted game

• Given a positive integer $n$, the $n$-stage game with initial state $k_1$ is the zero-sum game $\Gamma_n(k_1)$ with strategy spaces $\Sigma$ and $\mathscr{T}$ and payoff function:

$$\forall(\sigma,\tau)\in\Sigma\times\mathscr{T},\ \ \gamma_n^{k_1}(\sigma,\tau)=I\!E_{k_1,\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n}g(k_t,i_t,j_t)\right).$$

It has a value:
$v_n(k_1)=\max_{\sigma\in\Sigma}\min_{\tau\in\mathscr{T}}\gamma_n^{k_1}(\sigma,\tau)=\min_{\tau\in\mathscr{T}}\max_{\sigma\in\Sigma}\gamma_n^{k_1}(\sigma,\tau)$.

• Given a discount rate $\lambda$ in $(0,1]$, the $\lambda$-discounted game with initial state $k_1$ is the zero-sum game $\Gamma_\lambda(k_1)$ with strategy spaces $\Sigma$ and $\mathscr{T}$ and payoff function:

$$\forall(\sigma,\tau)\in\Sigma\times\mathscr{T},\ \ \gamma_\lambda^{k_1}(\sigma,\tau)=I\!E_{k_1,\sigma,\tau}\left(\lambda\sum_{t=1}^{\infty}(1-\lambda)^{t-1}g(k_t,i_t,j_t)\right).$$

It has a value denoted by $v_\lambda(k_1)$.
$\delta=1-\lambda=\frac{1}{1+r}$ is called the discount factor, and $r$ is the interest rate.

Proposition $v_n$ and $v_\lambda$ are characterized by the Shapley equations:

- For $n \geq 0$ and $k$ dans $K$:

$$(n+1)\, v_{n+1}(k) = \mathrm{Val}_{\Delta(I)\times\Delta(J)} \left( g(k,i,j) + \sum_{k' \in K} q(k'|k,i,j)\, n\, v_n(k') \right).$$

And in any $n$-stage game, players have Markov optimal strategies.

- For $\lambda$ in $(0,1]$ and $k$ in $K$:

$$v_\lambda(k) = \mathrm{Val}_{\Delta(I)\times\Delta(J)} \left( \lambda\, g(k,i,j) + (1-\lambda) \sum_{k' \in K} q(k'|k,i,j)\ v_\lambda(k') \right).$$

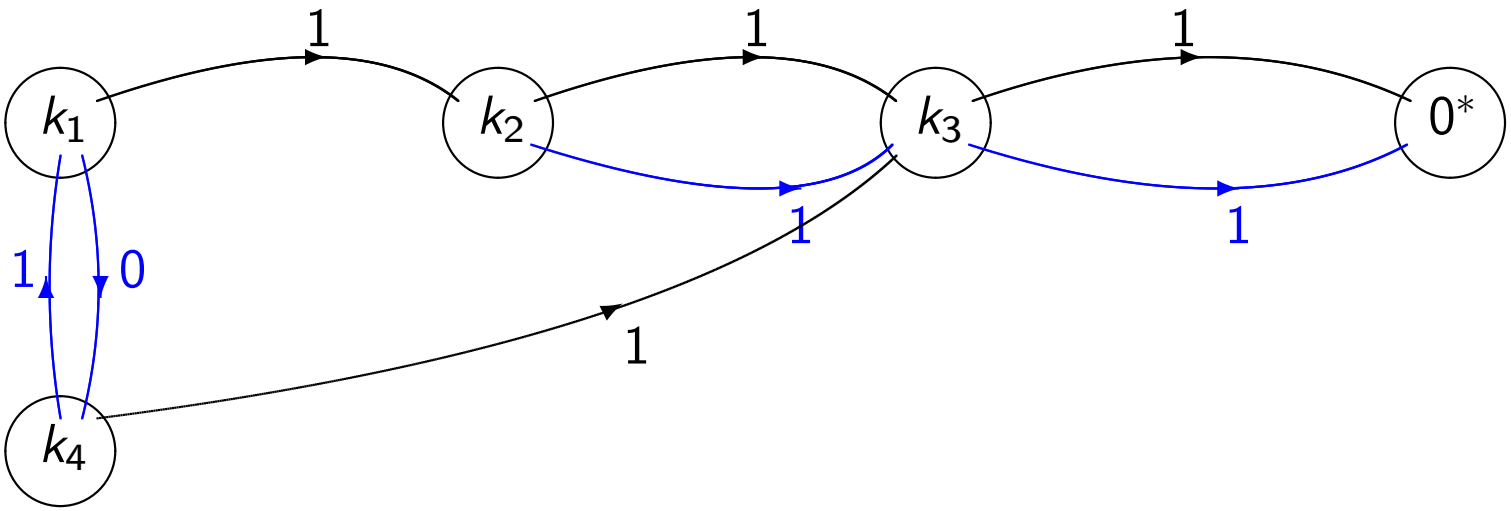
And in any $\lambda$-discounted game, players have stationary optimal strategies.

Example 1: $v_1 = \frac{1}{2}$, $v_{n+1} = \frac{1}{2 - \frac{n}{n+1} v_n}$ for $n \geq 1$, and $v_\lambda = \frac{1}{1+\sqrt{\lambda}}$ for each $\lambda$.

Shapley operator: for $v$ in $\mathbb{R}^K$, define $\Psi(v)$ in $\mathbb{R}^K$ by: for each $k$, $\Psi(v)^k = \mathrm{Val}_{\Delta(I)\times\Delta(J)} \left( g(k,i,j) + \sum_{k' \in K} q(k'|k,i,j)\, v^{k'} \right)$. $\Psi$ is non expansive for $\|.\|_\infty$ on $\mathbb{R}^K$, and the Shapley equations read:

$$n v_n = \Psi((n-1)v_{n-1}) = \Psi^n(0), \text{ and } v_\lambda = \lambda \Psi\left(\frac{1-\lambda}{\lambda} v_\lambda\right).$$

Example 2: A one-player game, with deterministic transitions and actions Black and Blue for Player 1. Payoffs are 1 or 0 in each case.




For $\lambda$ small enough, $v_\lambda(k_1) = \frac{1-\lambda}{2-\lambda}$ and it is optimal in the $\lambda$-discounted game to alternate between states $k_1$ and $k_4$.

For $n \geq 0$, $(2n+3)v_{2n+3} = (2n+4)v_{2n+4} = n+3$ (first alternate between $k_1$ and $k_4$, then go to $k_2$ 3 or 4 stages before the end).

# Limit values - The algebraic approach

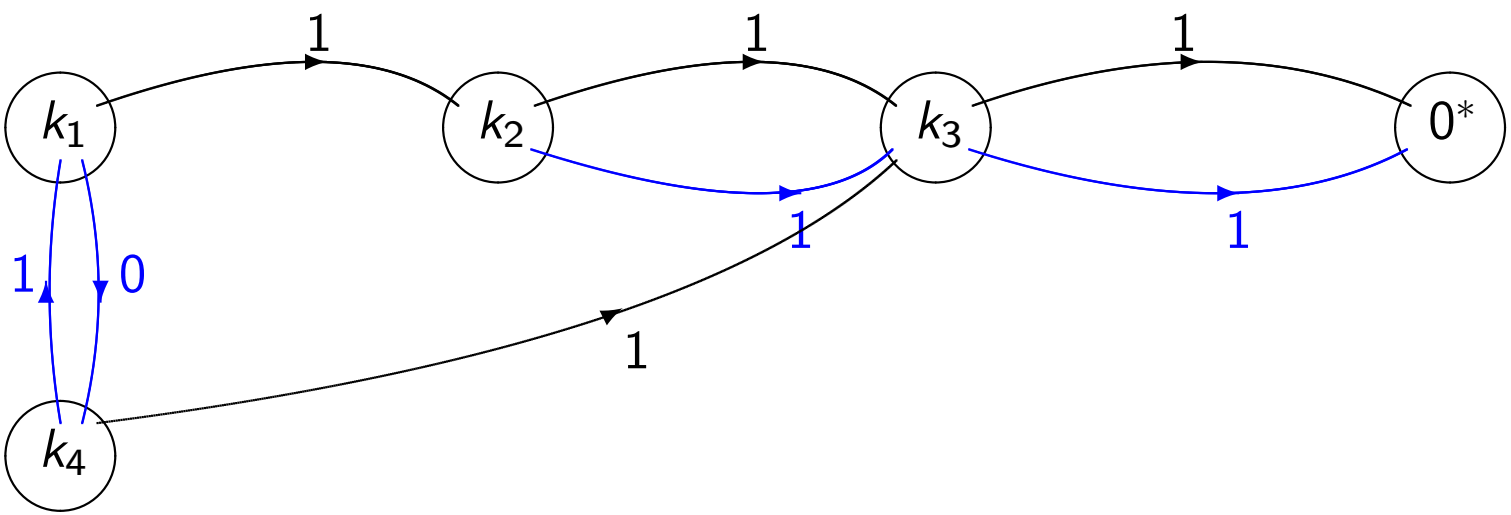
1) The 1-player case: Markov Decision Processes.
For $\lambda > 0$, player 1 has a *pure* stationary optimal strategy in the $\lambda$-discounted game. Such a strategy is defined by a mapping $f : K \to I$. The $\lambda$-discounted payoff satisfies:

$$\gamma_\lambda^k(f) = \lambda g(k, f(k)) + (1-\lambda) \sum_{k' \in K} q(k'|k, f(k)) \gamma_\lambda^{k'}(f).$$

Can be written in matrix form: $(I - (1-\lambda)A)v = \lambda\alpha$, where $v = (\gamma_\lambda^k(f))_k$, $I$ is the identity matrix, $A = (q(k'|k, f(k))_{k,k'}$ is a stochastic matrix, and $\alpha = (g(k, f(k)))_k$. $(I - (1-\lambda)A)$ is invertible, and its inverse has coefficients which are rational fractions of its coefficients. So for each $f$ and $k$, the payoff $\gamma_\lambda^k(f)$ is a rational function of $\lambda$.

Theorem (Blackwell, 1962): In the 1-player case, there exists $\lambda_0 > 0$ and a pure stationary strategy which is optimal in any game with discount $\lambda \leq \lambda_0$. For $\lambda \leq \lambda_0$ and $k$ in $K$, the value $v_\lambda(k)$ is a bounded rational fraction of $\lambda$, hence converges when $\lambda$ goes to 0.

Example 2 again: $f$ is the strategy which alternates forever between $k_1$ and $k_4$. There exists no strategy which is optimal in all $n$-stage games with $n$ sufficiently large.



## 2) Stochastic games: The algebraic approach

For each $\lambda$, players have stationary optimal strategies $x_\lambda$ and $y_\lambda$:
Consider the following set:

$$\begin{aligned} A = \quad & \{(\lambda, x_\lambda, y_\lambda, w_\lambda) \in (0,1] \times (I\!R^I)^K \times (I\!R^J)^K \times I\!R^K, \forall k \in K, \\ & x_\lambda(k), y_\lambda(k) \text{ stationary optimal in } \Gamma_\lambda(k),\ w_\lambda(k) = v_\lambda(k)\}. \end{aligned}$$

$A$ can be written with finitely many polynomial inequalities:

$$\forall i,j,k,\ \sum_i x_\lambda^i(k) = 1, x_\lambda^i(k) \geq 0, \sum_j y_\lambda^j(k) = 1, y_\lambda^j(k) \geq 0,$$

$$\forall j,k,\ \sum_{i \in I} x_\lambda^i(k)(\lambda g(k,i,j) + (1-\lambda)\sum_{k'} q(k'|k,i,j) w_\lambda(k')) \geq w_\lambda(k),$$

$$\forall i,k,\ \sum_{j \in J} y_\lambda^j(k)(\lambda g(k,i,j) + (1-\lambda)\sum_{k'} q(k'|k,i,j) w_\lambda(k')) \leq w_\lambda(k).$$

$A$ is semi-algebraic (can be written a finite union of sets, each of these sets being defined as the conjunction of finitely many weak or strict polynomial inequalities).

The projection of a semi-algebraic set is still semi-algebraic (Tarski-Seidenberg elimination theorem).
So $A^* = \{(\lambda, v_\lambda), \lambda \in (0,1]\}$ is also a semi-algebraic subset of $I\!R \times I\!R^K$.
Implies the existence of a bounded Puiseux series development of $v_\lambda$ in a neighborhood of $\lambda = 0$.

Theorem (Bewley Kohlberg 1976)

There exists $\lambda_0 > 0$, a positive integer $M$, coefficients $r_m \in I\!R^K$ for each $m \geq 0$ such that for all $\lambda \in (0, \lambda_0]$, and all $k$ in $K$:

$$v_\lambda(k) = \sum_{m=0}^{\infty} r_m(k)\, \lambda^{m/M}.$$

Example 1 : $v_\lambda = \frac{1-\sqrt{\lambda}}{1-\lambda} = (1-\sqrt{\lambda})(1+\lambda+\ldots+\lambda^n+\ldots.)$

Corollaries:

1) $v_\lambda$ converges when $\lambda$ goes to 0.
2) $v_\lambda$ has bounded variation at 0, i.e. for any sequence $(\lambda_i)_{i\geq 1}$ of discount factors decreasing to 0, we have $\sum_{i\geq 1} \| v_{\lambda_{i+1}} - v_{\lambda_i} \| < \infty$.
3) $v_n$ also converges, and $\lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$.

Proof of 3): (A. Neyman) We compare $v_n$ with the value $w_n := v_{1/n}$ of the $\frac{1}{n}$ discounted game. Using the Shapley operator, we have for all $n$:

$$v_{n+1} = \frac{1}{n+1}\Psi(nv_n), \text{ and } w_{n+1} = \frac{1}{n+1}\Psi(nw_{n+1}).$$

Since $\Psi$ is non expansive,
$\|w_{n+1} - v_{n+1}\| \leq \frac{n}{n+1}\|w_{n+1} - v_n\| \leq \frac{n}{n+1}(\|w_{n+1} - w_n\| + \|w_n - v_n\|)$.
We obtain:

$$(n+1)\|w_{n+1} - v_{n+1}\| - n\|w_n - v_n\| \leq n\|w_{n+1} - w_n\|.$$

Summing from $n = 1$ to $m$ gives:

$$\|w_{m+1} - v_{m+1}\| \leq \frac{1}{m+1}\sum_{n=1}^{m} n\|w_{n+1} - w_n\|.$$

By the bounded variation property, we have $\sum_{n=1}^{\infty}\|w_{n+1} - w_n\| < \infty$.
It is a simple exercise to show that if $(a_n)_n$ is a sequence of non negative real numbers satisfying $\sum_{n=1}^{\infty} a_n < \infty$, the sequence $(na_n)_n$ Cesaro-converges to 0.
We conclude that $\|w_{m+1} - v_{m+1}\| \xrightarrow[m\to\infty]{} 0$.

Remarks:

- An "elementary" proof of the existence of the limit value has been given by M. Oliu-Barton (2014).

- Computing the limit is not easy.

# The Uniform Value

Fix the initial state. $\lim_n v_n = \lim_\lambda v_\lambda$ exists, so we know the value of the stochastic game when $n$ is large (or $\lambda$ is small) and known to the players.
Question: when $n$ is large or $\lambda$ is small, can players play well without knowing exactly $n$ or $\lambda$ ?

Definition: Given a real $v$, say that:
Player 1 guarantees $v$ if: $\forall \varepsilon > 0$, $\exists \sigma \in \Sigma$, $\exists n_0$, $\forall n \geq n_0$, $\forall \tau \in \mathscr{T}$, $\gamma_n(\sigma, \tau) \geq v - \varepsilon$.
Player 2 guarantees $v$ if: $\forall \varepsilon > 0$, $\exists \tau \in \mathscr{T}$, $\exists n_0$, $\forall n \geq n_0$, $\forall \sigma \in \Sigma$, $\gamma_n(\sigma, \tau) \leq v + \varepsilon$.
If $v$ can be guaranteed by both players, then $v$ is called the uniform value of the stochastic game.

Strong concept: If the uniform value exists, it is unique, equal to $\lim_n v_n = \lim_\lambda v_\lambda$, and we have: $\forall \varepsilon > 0$ $\exists \lambda_0 > 0$, $\sigma, \tau$ s.t. for all $\lambda \leq \lambda_0$, $\sigma', \tau'$ we have: $\gamma_\lambda(\sigma, \tau') \geq v - \varepsilon$ and $\gamma_\lambda(\sigma', \tau) \leq v + \varepsilon$.

# The Uniform Value

Fix the initial state. $\lim_n v_n = \lim_\lambda v_\lambda$ exists, so we know the value of the stochastic game when $n$ is large (or $\lambda$ is small) and known to the players.
Question: when $n$ is large or $\lambda$ is small, can players play well without knowing exactly $n$ or $\lambda$ ?

Definition: Given a real $v$, say that:
Player 1 guarantees $v$ if: $\forall \varepsilon > 0$, $\exists \sigma \in \Sigma$, $\exists n_0$, $\forall n \geq n_0$, $\forall \tau \in \mathscr{T}$, $\gamma_n(\sigma, \tau) \geq v - \varepsilon$.
Player 2 guarantees $v$ if: $\forall \varepsilon > 0$, $\exists \tau \in \mathscr{T}$, $\exists n_0$, $\forall n \geq n_0$, $\forall \sigma \in \Sigma$, $\gamma_n(\sigma, \tau) \leq v + \varepsilon$.
If $v$ can be guaranteed by both players, then $v$ is called the uniform value of the stochastic game.

Strong concept: If the uniform value exists, it is unique, equal to $\lim_n v_n = \lim_\lambda v_\lambda$, and we have: $\forall \varepsilon > 0$ $\exists \lambda_0 > 0$, $\sigma, \tau$ s.t. for all $\lambda \leq \lambda_0$, $\sigma', \tau'$ we have: $\gamma_\lambda(\sigma, \tau') \geq v - \varepsilon$ and $\gamma_\lambda(\sigma', \tau) \leq v + \varepsilon$.

# The Big Match

Absorbing game given by:
$$\begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} 1^* & 0^* \\ 0 & 1 \end{pmatrix} \end{array}$$

$v_n = v_\lambda = 1/2$ for all $n$ and $\lambda$.
Player 2 guarantees 1/2 by playing i.i.d. 1/2L+1/2R.

Prop (Blackwell Ferguson 1968) : the Big Match has a uniform value.

Proof: define the r.v. $L_t = \sum_{s=1}^{t-1} \mathbf{1}_{j_s=L}$, $R_t = \sum_{s=1}^{t-1} \mathbf{1}_{j_s=R}$, and $m_t = R_t - L_t \in \{-(t-1), ..., 0, ..., t-1\}$. $R_1 = L_1 = m_1 = 0$.
Given a fixed parameter $M \geq 1$ define the strategy $\sigma_M$ of player 1 by:
at any stage $t$, $\sigma_M$ plays $T$ with probability $\frac{1}{(m_t+M+1)^2}$.
$\sigma_M$ is well defined.
At any stage $t$ in the non absorbing state, we have $-M \leq m_t \leq t-1$, and $\sigma_M$ plays $T$ with a probability in the interval $[\frac{1}{(M+t)^2}, 1]$.

We will show that $\sigma_M$ guarantees $\frac{M}{2(M+1)}$, close to 1/2 for $M$ large.

$$\text{Prop}: \forall T \geq 1, \forall M \geq 0, \forall \tau \in \mathscr{T},\ I\!E_{\sigma_M,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq \frac{M}{2(M+1)} - \frac{M}{2T}.$$

Proof of the Prop: Fix $T$. Assume w.l.o.g. that player 2 plays a fixed deterministic sequence $y = (j_1, \ldots j_t, \ldots) \in \{L, R\}^\infty$.
Define the random time of absorption:

$$t^* = \inf\{s \in \{1, \ldots, T\}, i_s = T\} \in \{1, \ldots, T+1\}.$$

Recall that $R_t = m_t + L_t = t - 1 - L_t$, so $R_t = \frac{1}{2}(m_t + t - 1)$. For $t \leq t^*$, we have $m_t \geq -M$, so $R_{t^*} \geq \frac{1}{2}(t^* - M - 1)$.
Define also $X_t$ as the random variable of the current limit value: $X_t = 1/2$ if $t \leq t^* - 1$, $X_t = 1$ if $t \geq t^*$ and $j_{t^*} = L$, and $X_t = 0$ if $t \geq t^*$ and $j_{t^*} = R$.

$$\begin{aligned} I\!E_{\sigma_M,y}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) &= I\!E_{\sigma_M,y}\ \frac{1}{T}(R_{t^*} + (T - t^* + 1)\mathbf{1}_{j_{t^*}=L}) \\ &\geq I\!E_{\sigma_M,y}\ \frac{1}{T}(\frac{1}{2}(t^* - M - 1) + (T - t^* + 1)\mathbf{1}_{j_{t^*}=L}) \\ &= -\frac{M}{2T} + I\!E_{\sigma_M,y}\ \frac{1}{T}(\frac{1}{2}(t^* - 1) + (T - t^* + 1)\mathbf{1}_{j_{t^*}=L}) \end{aligned}$$

So

$$\mathbb{E}_{\sigma_M,y}\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq -\frac{M}{2T} + \mathbb{E}_{\sigma_M,y}\left(\frac{1}{T}\sum_{t=1}^{T} X_t\right).$$

It is finally enough to prove:

Lemma: $\forall t \in \{1,...,T\}, \forall y \in \{L,R\}^\infty, \forall M \geq 1$, $\mathbb{E}_{\sigma_M,y}(X_t) \geq \frac{M}{2(M+1)}$.

Proof of the lemma. by induction on $t$.
For $t=1$,
$\mathbb{E}_{\sigma_M,y}(X_1) = \frac{1}{2}(1-\frac{1}{(M+1)^2}) + \frac{1}{(M+1)^2}\mathbf{1}_{j_1=L} \geq \frac{1}{2}(1-\frac{1}{(M+1)^2}) \geq \frac{M}{2(M+1)}$.

Assume the lemma true for some $t$. Consider $y=(j_1,...)$ in $\{L,R\}^\infty$, and write $y=(j_1,y_+)$ with $y_+=(j_2,j_3,...)\in\{L,R\}^\infty$.

If $j_1=L$, $\mathbb{E}_{\sigma_M,y}(X_{t+1}) = \frac{1}{(M+1)^2}1 + (1-\frac{1}{(M+1)^2})\mathbb{E}_{\sigma_{M-1},y_+}(X_t)$. By assumption, $\mathbb{E}_{\sigma_{M-1},y_+}(X_t) \geq \frac{M-1}{2M}$, so $\mathbb{E}_{\sigma_M,y}(X_{t+1}) \geq \frac{M}{2(M+1)}$.

Otherwise $j_1=R$, and $\mathbb{E}_{\sigma_M,y}(X_{t+1}) = (1-\frac{1}{(M+1)^2})\mathbb{E}_{\sigma_{M+1},y_+}(X_t)$
$\geq (1-\frac{1}{(M+1)^2})\frac{M+1}{2(M+2)} = \frac{M}{2(M+1)}$. *end of proof for the Big Match.*

Theorem (Mertens Neyman 1981):
Every zero-sum stochastic game with finitely many states and actions has a uniform value.

Proof: in the written notes. uses the BV property of $(v_\lambda)$.

The theorem extends to more general models where states and actions can be infinite, provided:
1) stage payoffs are bounded
2) for each state $k$ and discount $\lambda$ the corresponding discounted game has a value $v_\lambda(k)$
3) one can find $(\lambda_i)_i$ decreasing to 0 s.t. $\frac{\lambda_{i+1}}{\lambda_i} \longrightarrow 1$ and $\sum_i \|v_{\lambda_{i+1}} - v_{\lambda_i}\|_\infty < \infty$
4) states and payoffs (not necessarily actions) are observed by the players.

Remark: In the variant of the Big Match where Player 1 can not observe the actions played by player 2, the $n$-stage and $\delta$-discounted values are still the same, but the uniform value no longer exists.

Theorem (Mertens Neyman 1981):
Every zero-sum stochastic game with finitely many states and actions has a uniform value.

Proof: in the written notes. uses the BV property of $(v_\lambda)$.

The theorem extends to more general models where states and actions can be infinite, provided:
1) stage payoffs are bounded
2) for each state $k$ and discount $\lambda$ the corresponding discounted game has a value $v_\lambda(k)$
3) one can find $(\lambda_i)_i$ decreasing to 0 s.t. $\frac{\lambda_{i+1}}{\lambda_i} \longrightarrow 1$ and $\sum_i \|v_{\lambda_{i+1}} - v_{\lambda_i}\|_\infty < \infty$
4) states and payoffs (not necessarily actions) are observed by the players.

Remark: In the variant of the Big Match where Player 1 can not observe the actions played by player 2, the $n$-stage and $\delta$-discounted values are still the same, but the uniform value no longer exists.

Theorem (Mertens Neyman 1981):
Every zero-sum stochastic game with finitely many states and actions has a uniform value.

Proof: in the written notes. uses the BV property of $(v_\lambda)$.

The theorem extends to more general models where states and actions can be infinite, provided:
1) stage payoffs are bounded
2) for each state $k$ and discount $\lambda$ the corresponding discounted game has a value $v_\lambda(k)$
3) one can find $(\lambda_i)_i$ decreasing to 0 s.t. $\frac{\lambda_{i+1}}{\lambda_i} \longrightarrow 1$ and $\sum_i \|v_{\lambda_{i+1}} - v_{\lambda_i}\|_\infty < \infty$
4) states and payoffs (not necessarily actions) are observed by the players.

Remark: In the variant of the Big Match where Player 1 can not observe the actions played by player 2, the $n$-stage and $\delta$-discounted values are still the same, but the uniform value no longer exists.

Theorem (Mertens Neyman 1981):
Every zero-sum stochastic game with finitely many states and actions has a uniform value.

Proof: in the written notes. uses the BV property of $(v_\lambda)$.

The theorem extends to more general models where states and actions can be infinite, provided:
1) stage payoffs are bounded
2) for each state $k$ and discount $\lambda$ the corresponding discounted game has a value $v_\lambda(k)$
3) one can find $(\lambda_i)_i$ decreasing to 0 s.t. $\frac{\lambda_{i+1}}{\lambda_i} \longrightarrow 1$ and $\sum_i \|v_{\lambda_{i+1}} - v_{\lambda_i}\|_\infty < \infty$
4) states and payoffs (not necessarily actions) are observed by the players.

Remark: In the variant of the Big Match where Player 1 can not observe the actions played by player 2, the $n$-stage and $\delta$-discounted values are still the same, but the uniform value no longer exists.

# 2. A few extensions and recent results

2.1 Counterexamples : A simple compact continuous game with no limit value. A hidden stochastic game with no limit value

2.2 1-player games (Dynamic programming, Gambling Houses and MDP): General results. The compact non expansive case.

2.3 Uniform convergence of $(v_n)_n$ and $(v_\lambda)_\lambda$ are equivalent.

2.4 Repeated Games with incomplete information : Lack of information on one side, repeated games with an informed controller. Lack of information on both sides.

2.5 Some open problems.

# A simple stochastic game with no limit value

- Finite set of states, compact action sets, continuous transitions.

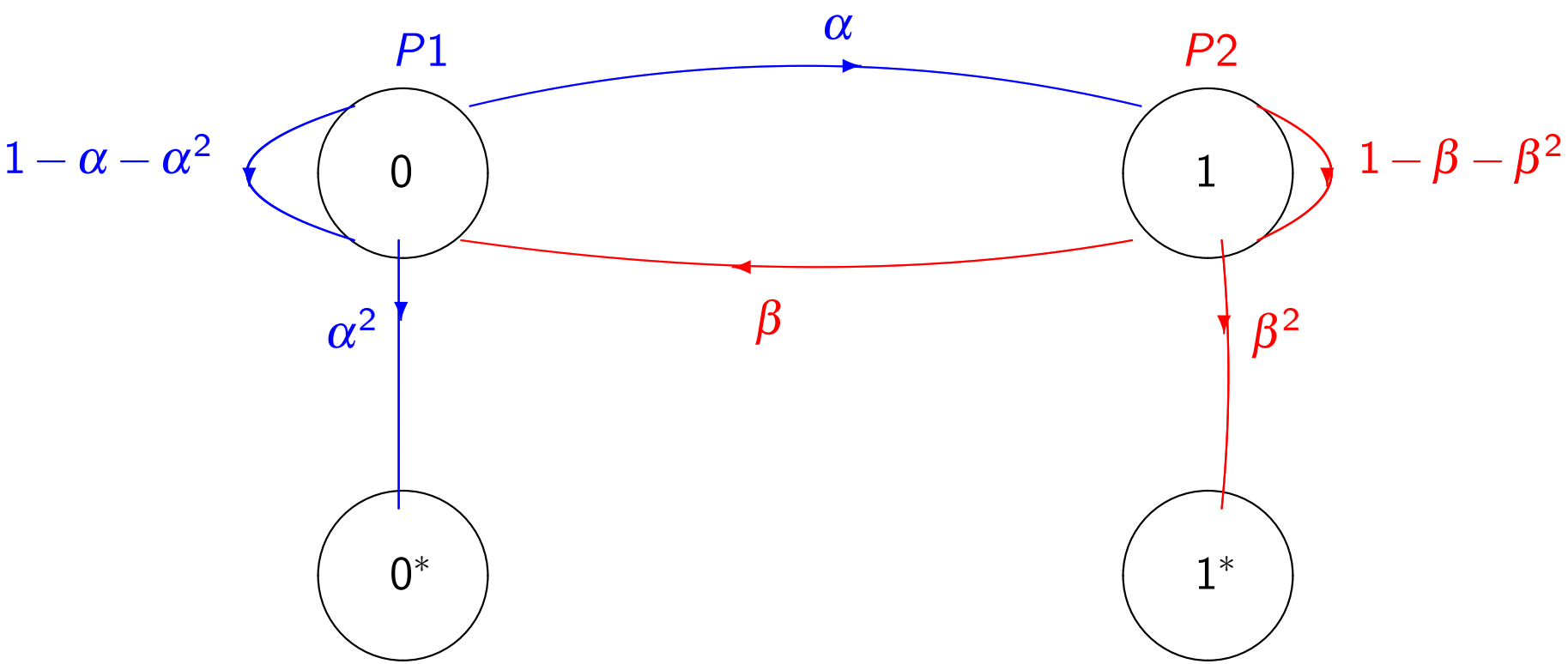
First counterexample: Vigeral (2013), with non semi-algebraic transitions (Bolte, Gaubert, Vigeral 2015)

Following counterexample : polynomial transitions but non semi-algebraic action sets.

Slight variant of a counter-example of Ziliotto (2014), also mentioned in Sorin Vigeral (2015).

The elementary self-contained proof here follows Laraki Renault (2015), where a close counterexample can be found.

4 states: $K = \{k_0, k_1, 0^*, 1^*\}$.
In state $k_1$ player 2 chooses $\beta$ in $J = [0, 1/2]$.
In state $k_0$, Player 1 chooses $\alpha \in I = \{0\} \cup \{\frac{1}{2^{2n}}, n \geq 1\}$



Write $x_\lambda = v_\lambda(k_0)$, $y_\lambda = v_\lambda(k_1)$.
Shapley equations:

$$\begin{aligned} x_\lambda &= \max_{\alpha \in I}(1-\lambda)((1-\alpha-\alpha^2)x_\lambda + \alpha y_\lambda), \\ y_\lambda &= \min_{\beta \in J}\left(\lambda + (1-\lambda)((1-\beta-\beta^2)y_\lambda + \beta x_\lambda + \beta^2)\right). \end{aligned}$$

Can be rewriten:

$$\lambda x_\lambda = (1-\lambda)\max_{\alpha \in I}\left(\alpha(y_\lambda - x_\lambda) - \alpha^2 x_\lambda\right) \quad (1)$$

$$\lambda y_\lambda = \lambda + (1-\lambda)\min_{\beta \in J}(\beta(x_\lambda - y_\lambda) + \beta^2(1-y_\lambda)) \quad (2)$$

Since $x_\lambda > 0$, eq. (1) gives that $y_\lambda > x_\lambda$.

Lemma 0: For $\lambda \leq 1/5$, $\beta_\lambda = \frac{y_\lambda - x_\lambda}{2(1-y_\lambda)}$ is optimal for player 2 and

$$4\lambda(1-y_\lambda)^2 = (1-\lambda)(y_\lambda - x_\lambda)^2. \quad (3)$$

Consequence: $y_\lambda - x_\lambda \longrightarrow_{\lambda \to 0} 0$.
Let $\lambda_n$ be a vanishing sequence of discount factors.

Lemma 1: if $y_{\lambda_n}$ and $x_{\lambda_n}$ converge to $v$ in $[0,1]$, then $v \leq 1/2$, $y_{\lambda_n} - x_{\lambda_n} \sim 2\sqrt{\lambda_n}(1-v)$ and $\beta_{\lambda_n} \sim \sqrt{\lambda_n}$.

Proof: Consider an optimal strategy $\alpha_\lambda$ for P1. Using (3), we get $x_\lambda(\lambda + \alpha_\lambda^2) = \lambda x_\lambda \alpha_\lambda^2 + 2\alpha_\lambda\sqrt{\lambda}\sqrt{1-\lambda}(1-y_\lambda) \geq 2\alpha_\lambda\sqrt{\lambda}x_\lambda$. Dividing by $\alpha_\lambda\sqrt{\lambda}$ and passing to the limit gives $v \leq 1/2$.

Lemma 2: If for each $n$, $\sqrt{\lambda_n} \in I$, then $y_{\lambda_n}$ and $x_{\lambda_n}$ converge to $1/2$.

Proof: By considering a converging subsequence we can assume that $y_{\lambda_n}$ and $x_{\lambda_n}$ converge to some $v$ in $[0,1]$. By the previous lemma, $v \leq 1/2$, and we have to show that $v \geq 1/2$. For each $\lambda$ in the subsequence player 1 can choose to play $\alpha = \sqrt{\lambda}$, so:

$$\lambda x_\lambda \geq (1-\lambda)\sqrt{\lambda}(y_\lambda - x_\lambda) - (1-\lambda)\lambda x_\lambda.$$

Dividing by $\lambda$ and passing to the limit, we obtain $v \geq 1/2$.

Lemma 3: If for each $n$, the open interval $(\frac{1}{2}\sqrt{\lambda_n}, 2\sqrt{\lambda_n})$ does not intersect $I$, then $\limsup_n y_{\lambda_n} \leq 4/9$.

Proof: Suppose that (up to a subsequence) $x_{\lambda_n}$ and $y_{\lambda_n}$ converges to some $v \geq 4/9$. It is enough to show that $v = 4/9$. We know that $v \leq 1/2$ by lemma 1. Consider again the maximization problem of player 1, and denote by $\alpha^*(\lambda) = \frac{y_\lambda - x_\lambda}{2x_\lambda} > 0$ the argmax of the *unconstrained* problem.
Since $\alpha^*(\lambda) \sim \sqrt{\lambda}\frac{1-v}{v}$, $\frac{1}{2}\sqrt{\lambda} \leq \alpha^*(\lambda) \geq 2\sqrt{\lambda}$ for $\lambda$ small in the sequence. By assumption $(\frac{1}{2}\sqrt{\lambda}, 2\sqrt{\lambda})$ contains no point in $I$ and the objective function of player 1 is increasing from 0 to $\alpha^*(\lambda)$ and decreasing after $\alpha^*(\lambda)$. 2 possible cases:
If $\alpha_\lambda \leq \frac{1}{2}\sqrt{\lambda}$ we have $\lambda x_\lambda \leq \frac{1}{2}(1-\lambda)\sqrt{\lambda}(y_\lambda - x_\lambda) - \frac{1}{4}(1-\lambda)\lambda x_\lambda$.
Dividing by $\lambda$ and passing to the limit gives: $v \leq \frac{4}{9}$.
Otherwise, $\alpha_\lambda > 2\sqrt{\lambda}$ and we have
$\lambda x_\lambda \leq 2(1-\lambda)\sqrt{\lambda}(y_\lambda - x_\lambda) - 4(1-\lambda)\lambda x_\lambda$. Again, $v \leq \frac{4}{9}$.

Considering the sequences $\lambda_n = \frac{1}{2^{2n}}$ and $\lambda_n = \frac{1}{2^{2n+1}}$ is finally enough to conclude that there is no limit value.

# A hidden stochastic game with no limit value

Hidden stochastic games: at the beginning of each period, players observe past actions and a public signal (but no longer the current state). Stochastic games with public information.

Model given by: a set of states $K$ a set of actions $I$ for player 1, a set of actions $J$ for player 2, a set of signals $S$, a payoff function $g : K \times I \times J \longrightarrow \mathbb{R}$, and a transition $q : K \times I \times J \longrightarrow \Delta(K \times S)$. Here, $K$, $I$, $J$ and $S$ are finite.

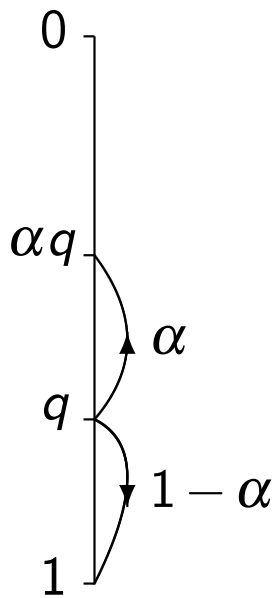
Bruno Ziliotto (2013) constructed a hidden stochastic game with no limit value. ($\liminf v_\delta = 1/2, \limsup v_\delta \geq 5/9$). Disproves 2 conjectures of Mertens/ MSZ.

Here: improvement (R. Ziliotto 2015) .
Theorem: For each $\varepsilon > 0$, there exists a zero-sum HSG with payoffs in $[0,1]$ for P1, 6 states, 2 actions for each player, 6 signals, s.t.:

$$\limsup v_\lambda \geq 1 - \varepsilon \text{ and } \liminf v_\lambda \leq -\varepsilon.$$

Construction done in 4 progressive steps: a Markov chain on $[0,1]$, a Markov Decision Process, a stochastic game with infinite state space, and a final HSG.

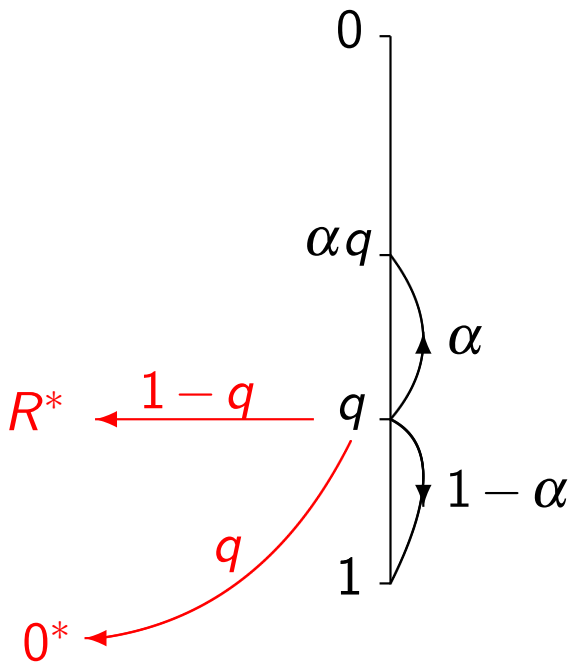
Step 1: a Markov chain on $[0,1]$, with a parameter $\alpha \in (0,1/4)$.
Initial state $q_0 = 1$.




Define for $a$ in $\mathbb{N}$, $T_a = \inf\{t \geq 1, q_t \leq \alpha^a\}$.

$\mathbb{E}(T_{a+1}) = \frac{1}{\alpha}(1 + \mathbb{E}(T_a))$ (grows exponentially with $a$)

## Step 2: a Markov Decision Process on $[0,1]$




Recall $T_a = \inf\{t \geq 1, q_t \leq \alpha^a\}$.
Payoff of the $a$-strategy in the MDP with parameter $\alpha$, reward $R$ and discount factor $\delta$: $R\, s_{\alpha,\delta}(a)$, with

$$s_{\alpha,\delta}(a) = \frac{(1-\alpha^a)(1-\alpha\delta)}{1-\alpha+(1-\delta)\alpha^{-a}\delta^{-a-1}}.$$

(optimal strategies do not depend on $R$)

For $R = 1$, the value is:

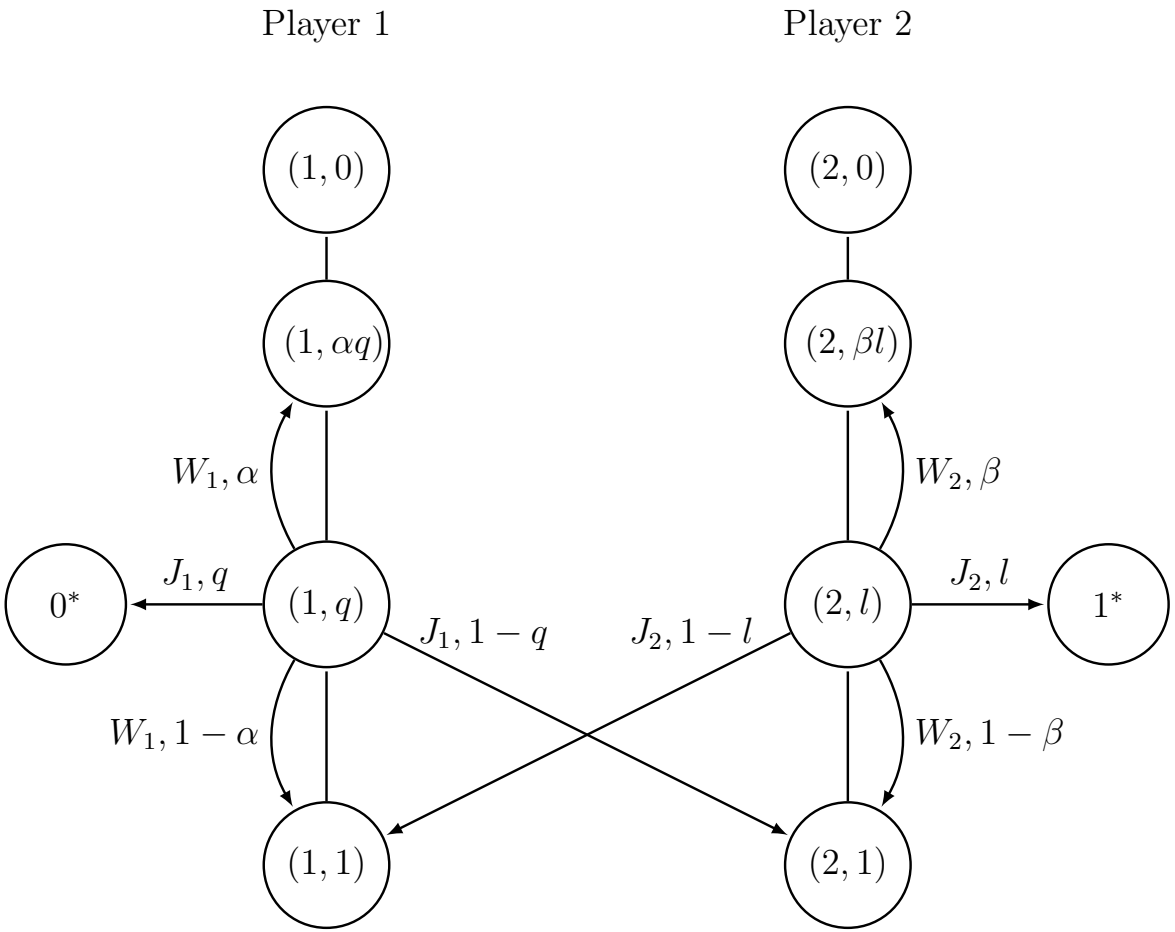
$$v_{\alpha,\delta} = \mathrm{Max}_{a \in IN} s_{\alpha,\delta}(a) = \mathrm{Max}_{a \in IN} \frac{(1-\alpha^a)(1-\alpha\delta)}{1-\alpha+(1-\delta)\alpha^{-a}\delta^{-a-1}} \xrightarrow[\delta \to 1]{} 1.$$

Optimal choice for $a \in IR_+$ would be $a^* = a^*(\alpha,\delta)$ s.t. $\alpha^{a^*} \simeq \sqrt{\frac{1-\delta}{1-\alpha}}$.
Define $\Delta_1(\alpha) = \{\delta, a^* \in IN\} = \{1-(1-\alpha)\alpha^{2a}, a \in IN\}$,
and $\Delta_2(\alpha) = \{\delta, a^* \in IN + [1/4, 3/4]\}$.

Proposition:
For $\delta \in \Delta_1(\alpha)$, $v_{\alpha,\delta} = 1 - \frac{2}{\sqrt{1-\alpha}}\sqrt{1-\delta} + o(\sqrt{1-\delta})$.
For $\delta \in \Delta_2(\alpha)$, $v_{\alpha,\delta} \leq 1 - \frac{1}{\sqrt{\alpha^{1/2}(1-\alpha)}}\sqrt{1-\delta} + o(\sqrt{1-\delta})$.

## Step 3: a stochastic game $\Gamma_{\alpha,\beta}$ with perfect information

States: $X = \{(1,q), q \in [0,1]\} \cup \{(2,l), l \in [0,1]\} \cup 0^* \cup 1^*$, start at $(2,1)$. Sum of payoffs is 1, P1 has payoff 0 in the left part, payoff 1 in the right part.



Proposition: The stochastic game restricted to pure stationary strategies has an equilibrium in *dominant strategies*, and value:

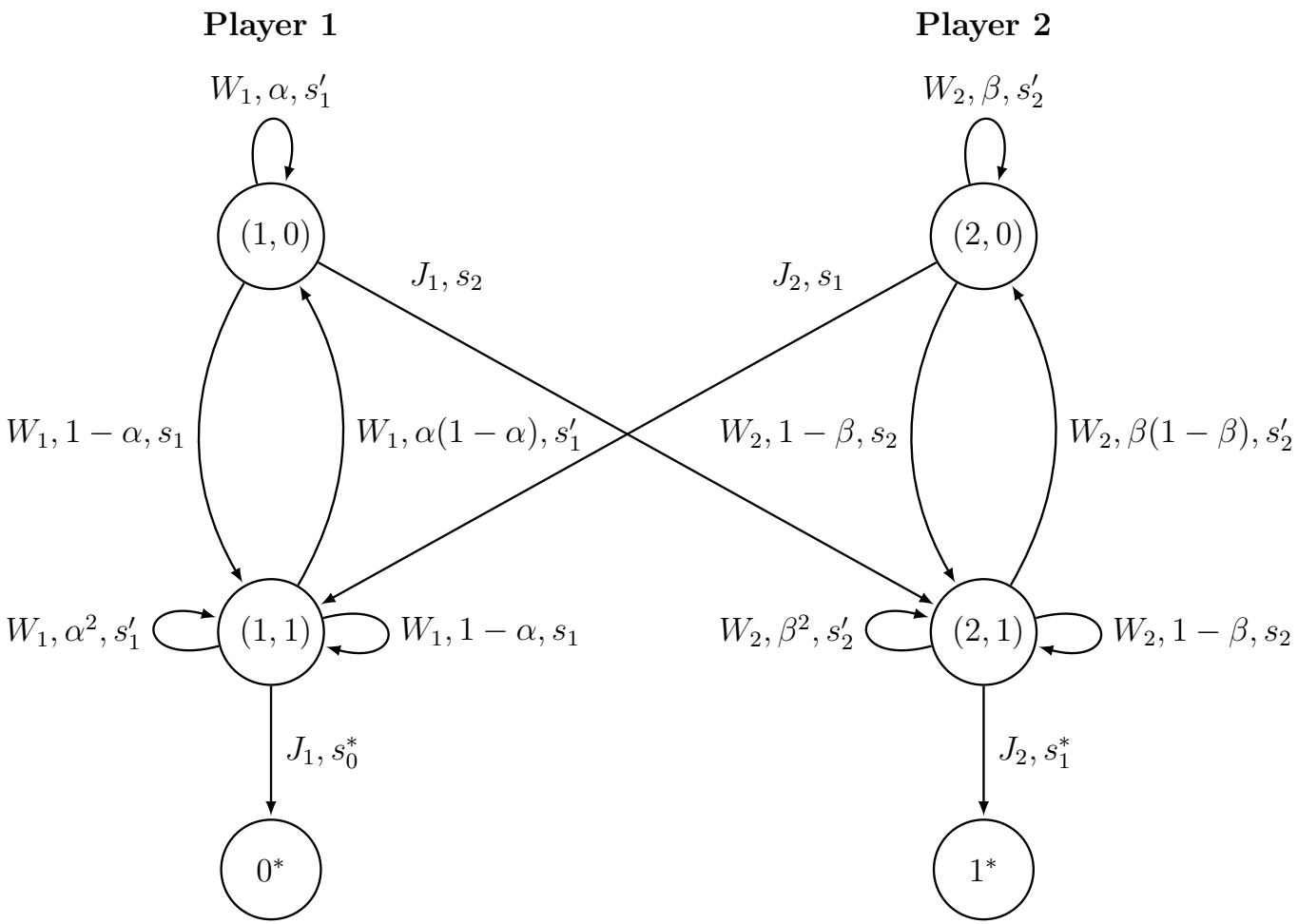
$$v_{\alpha,\beta,\delta} = \frac{1 - v_{\beta,\delta}}{1 - v_{\alpha,\delta} v_{\beta,\delta}}.$$

Proposition: Fix $\varepsilon > 0$. For $n$ large enough, fixing $\alpha = 1/n$ and $\beta = 1/(n+1)$ yields:

$$\limsup_{\delta \to 1} v_{\alpha,\beta,\delta} \geq 1 - \varepsilon, \text{ and } \liminf_{\delta \to 1} v_{\alpha,\beta,\delta} \leq \varepsilon.$$

## Step 4: a constant-sum hidden stochastic game $\Gamma^*_{\alpha,\beta}$

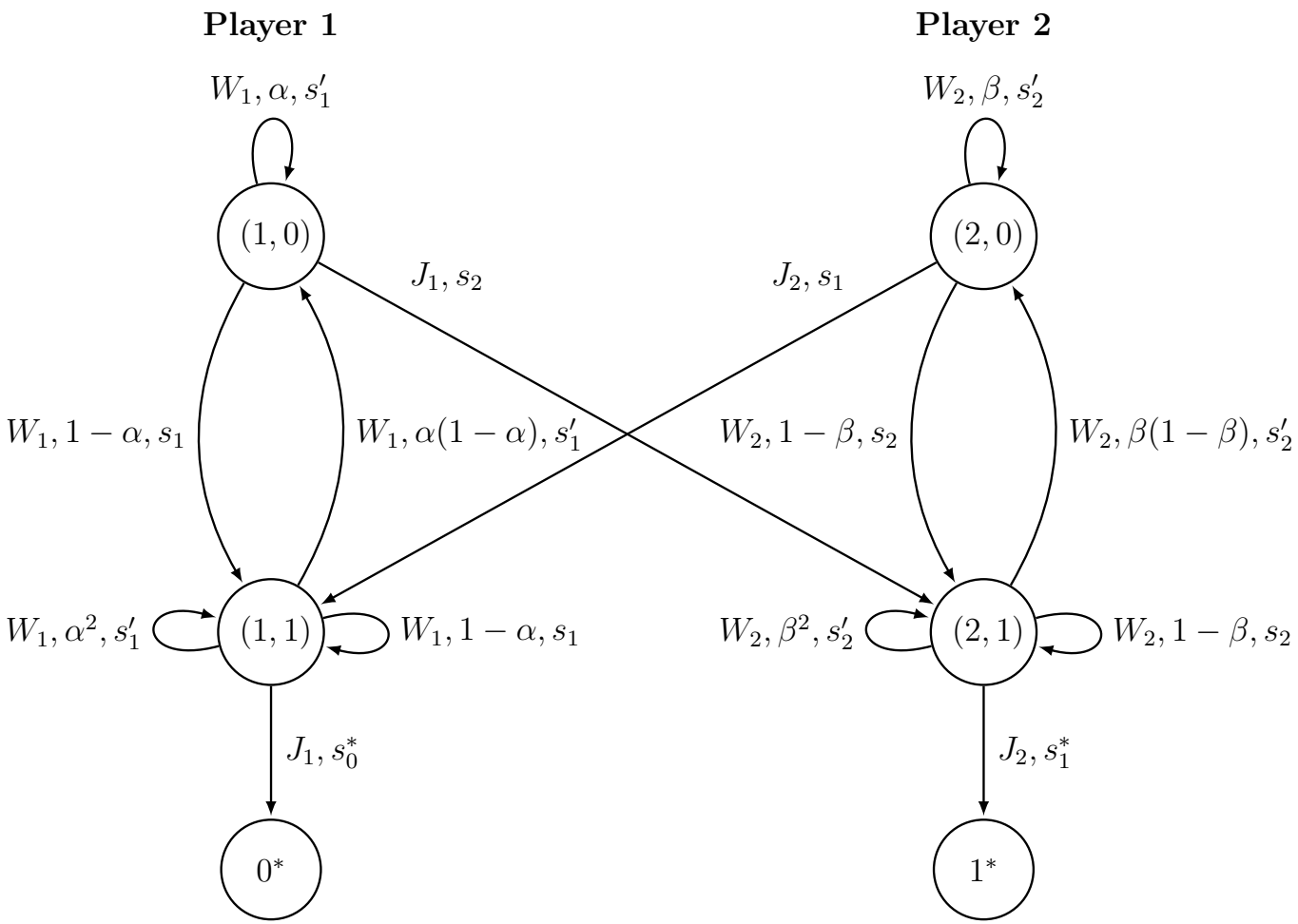
with 6 states and 6 signals, initial state: (2,1).




**Value** $v^*_{\alpha,\beta,\delta} = v_{\alpha,\beta,\delta}$.

## Step 4: a constant-sum hidden stochastic game $\Gamma^*_{\alpha,\beta}$

with 6 states and 6 signals, initial state: (2,1).




**Value** $v^*_{\alpha,\beta,\delta} = v_{\alpha,\beta,\delta}$.

# The 1 Player case

Consider a general dynamic programming problem with bounded payoffs: $\Gamma(z_0) = (Z, F, r, z_0)$ with a non empty set of states $Z$, an initial state $z_0$, a transition correspondence $F$ from $Z$ to $Z$ with non empty values, and a reward mapping $r$ from $Z$ to $[0,1]$. Here $Z$ can be any set.

A player chooses $z_1$ in $F(z_0)$, has a payoff of $r(z_1)$, then he chooses $z_2$ in $F(z_1)$, etc...
Set of admissible plays at $z_0$:
$S(z_0) = \{s = (z_1, ..., z_t, ...) \in Z^\infty, \forall t \geq 1, z_t \in F(z_{t-1})\}$.

Value of the $n$-stage problem with initial state $z$:

$$v_n(z) = \sup_{s \in S(z)} \gamma_n(s), \text{ where } \gamma_n(s) = \frac{1}{n} \sum_{t=1}^{n} r(z_t).$$

Value of the $\lambda$-discounted problem with initial state $z$:

$$v_\lambda(z) = \sup_{s \in S(z)} \gamma_\lambda(s), \text{ where } \gamma_\lambda(s) = \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} r(z_t).$$

More generally: let $\Theta$ be the set of probabilities on positive integers. Given an evaluation $\theta = \sum_{t\geq 1} \theta_t \delta_t$ in $\Theta$, the $\theta$-value of $\Gamma(z)$ is

$$v_\theta(z) = \sup_{s\in S(z)} \gamma_\theta(s), \text{ where } \gamma_\theta(s) = \sum_{t=1}^{\infty} \theta_t r(z_t).$$

The total variation of $\theta$ is : $TV(\theta) = \sum_{t=1}^{\infty} |\theta_{t+1} - \theta_t|$.
For $m \geq 0$, write $v_{m,\theta}$ for the value function associated to the shifted evaluation $\theta \oplus m = \sum_{t=1}^{\infty} \theta_t \delta_{m+t}$.

What can be said in general about $\lim_{n\to\infty} v_n$, $\lim_{\lambda\to 0} v_\lambda$, or more generally of $\lim_{k\to\infty} v_{\theta^k}$, when $(\theta^k)_k$ is a sequence of evaluations such that $TV(\theta^k) \to_{k\to\infty} 0$?
Many things, if we focus on *uniform convergence.*

More generally: let $\Theta$ be the set of probabilities on positive integers. Given an evaluation $\theta = \sum_{t\geq 1} \theta_t \delta_t$ in $\Theta$, the $\theta$-value of $\Gamma(z)$ is

$$v_\theta(z) = \sup_{s\in S(z)} \gamma_\theta(s), \text{ where } \gamma_\theta(s) = \sum_{t=1}^{\infty} \theta_t r(z_t).$$

The total variation of $\theta$ is : $TV(\theta) = \sum_{t=1}^{\infty} |\theta_{t+1} - \theta_t|$.
For $m \geq 0$, write $v_{m,\theta}$ for the value function associated to the shifted evaluation $\theta \oplus m = \sum_{t=1}^{\infty} \theta_t \delta_{m+t}$.

What can be said in general about $\lim_{n\to\infty} v_n$, $\lim_{\lambda\to 0} v_\lambda$, or more generally of $\lim_{k\to\infty} v_{\theta^k}$, when $(\theta^k)_k$ is a sequence of evaluations such that $TV(\theta^k) \to_{k\to\infty} 0$?
Many things, if we focus on *uniform convergence.*

Denote by $\mathscr{V}$ the metric space of functions from $Z$ to $[0,1]$, with the sup metric $d_\infty(v,v') = \sup_{z\in Z}|v(z)-v(z')|$. In general, a sequence in $\mathscr{V}$ converges iff 1) $(v^k)_k$ has at most one limit point, and 2) the set $\{v^k, k \geq 1\}$ is totally bounded.

$$\text{Define } \forall z \in Z,\ v^*(z) = \inf_{\theta\in\Theta} \sup_{m\geq 0} v_{m,\theta}(z).$$

Consider a sequence of evaluations $(\theta^k)_k$ such that $TV(\theta^k) \to_{k\to\infty} 0$.

Theorem (R. 2014) : Any limit point of $(v_{\theta^k})_k$ is $v^*$.

Corollaries:

1) If $(v_{\theta^k})_k$ converges, the limit is $v^*$.

2) $$(v_{\theta^k})_k \text{ converges} \iff \text{the set } \{v_{\theta^k}, k\geq 1\} \text{ is totally bounded,}$$
$$\iff \text{the set } \{v_{\theta^k}, k\geq 1\}\cup\{v^*\} \text{ is compact.}$$

3) If $(Z,d)$ is a totally bounded metric space, and if the family $(v_\theta)_{\theta\in\Theta}$ is uniformly equicontinuous, then there is general uniform convergence of the value functions to $v^*$, i.e.
$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Theta\ s.t.\ TV(\theta) \leq \alpha,\ \|v_\theta - v^*\| \leq \varepsilon.$

4) If $(Z,d)$ is a precompact metric space, $r$ is uniformly continuous, and $F$ is non expansive, same conclusions as 3).

Denote by $\mathscr{V}$ the metric space of functions from $Z$ to $[0,1]$, with the sup metric $d_\infty(v,v') = \sup_{z\in Z}|v(z)-v(z')|$. In general, a sequence in $\mathscr{V}$ converges iff 1) $(v^k)_k$ has at most one limit point, and 2) the set $\{v^k, k\geq 1\}$ is totally bounded.

$$\text{Define } \forall z \in Z,\ v^*(z) = \inf_{\theta\in\Theta}\sup_{m\geq 0} v_{m,\theta}(z).$$

Consider a sequence of evaluations $(\theta^k)_k$ such that $TV(\theta^k) \to_{k\to\infty} 0$.
Theorem (R. 2014) : Any limit point of $(v_{\theta^k})_k$ is $v^*$.
Corollaries:
1) If $(v_{\theta^k})_k$ converges, the limit is $v^*$.

2) $$(v_{\theta^k})_k \text{ converges} \iff \text{the set } \{v_{\theta^k}, k\geq 1\} \text{ is totally bounded,}$$
$$\iff \text{the set } \{v_{\theta^k}, k\geq 1\}\cup\{v^*\} \text{ is compact.}$$

3) If $(Z,d)$ is a totally bounded metric space, and if the family $(v_\theta)_{\theta\in\Theta}$ is uniformly equicontinuous, then there is general uniform convergence of the value functions to $v^*$, i.e.
$\forall\varepsilon>0, \exists\alpha>0, \forall\theta\in\Theta\ s.t.\ TV(\theta)\leq\alpha,\ \|v_\theta - v^*\|\leq\varepsilon$.
4) If $(Z,d)$ is a precompact metric space, $r$ is uniformly continuous, and $F$ is non expansive, same conclusions as 3).

Results extend to the case of stochastic dynamic programming.

Particular easy case : if the problem is leavable ($z \in F(z)$ for each $z$).
Then $(v_n)$ and $(v_\lambda)$ *pointwise* converge to $v^*$, where:
$v^* = \inf\{v : Z \to [0,1], \textit{excessive}, v \geq r\}$. ( $v$ excessive means that $v(z) \geq v(z')$ if $z' \in F(z)$, i.e. that $v$ is non increasing on any trajectory.)

Theorem (Lehrer-Sorin 1992 ): In a 1-player game, $(v_n)$ converges uniformly if and only if $(v_\lambda)$ converges uniformly. In case of convergence, the limit is the same.

Results extend to the case of stochastic dynamic programming.

Particular easy case : if the problem is leavable ($z \in F(z)$ for each $z$). Then $(v_n)$ and $(v_\lambda)$ *pointwise* converge to $v^*$, where: $v^* = \inf\{v : Z \to [0,1], \textit{excessive}, v \geq r\}$. ( $v$ excessive means that $v(z) \geq v(z')$ if $z' \in F(z)$, i.e. that $v$ is non increasing on any trajectory.)

Theorem (Lehrer-Sorin 1992 ): In a 1-player game, $(v_n)$ converges uniformly if and only if $(v_\lambda)$ converges uniformly. In case of convergence, the limit is the same.

# 1 player: The compact non expansive case.

Consider a stochastic dynamic programing problem (or Gambling House) $\Gamma = (X, F, r, x_0)$ given by a non empty set of states $X$, an initial state $x_0$, a transition multifunction $F$ from $X$ to $Z := \Delta_f(X)$ with non empty values, and a reward mapping $r$ from $X$ to $[0, 1]$.

Assume $X$ is compact metric. Then $\Delta(X)$ is also compact metric space for the Kantorovich-Rubinstein metric: for $z$ and $z'$ in $\Delta(X)$,

$$\begin{aligned} d_{KR}(u, u') &= \sup_{f: X \to I\!R, 1-Lip} \left| \int_{x \in X} f(x) du(x) - \int_{x \in X} f(x) du'(x) \right| \\ &= \min_{\pi \in \Pi(u,u')} \int_{(x,x') \in X \times X} d(x, x') d\pi(x, x'). \end{aligned}$$

$X$ is now viewed as a subset of $\Delta(X)$, and we define the set of *invariant measures* as:

$$R = \{u \in \Delta(X), (u, u) \in \overline{\text{conv}}\text{Graph}(\Gamma)\}$$

Theorem (R-Venel 2013) Assume the state space is compact metric, payoffs are continuous and transitions are non expansive. Then $(v_n)$ and $(v_\lambda)$ uniformly converge to $v^*$, where for each initial state $x$,

$$v^*(x) = \inf\{w(x), w : \Delta(X) \to [0,1] \text{ affine } C^0 \text{ } s.t.$$

$$(1)\ \forall x' \in X, w(x') \geq \sup_{u \in F(x')} w(u)$$

$$(2)\ \forall u \in R, w(u) \geq r(u)\}.$$

Moreover, the uniform value exists if $F$ has convex values (or if one allows the player to play a behavior strategy, i.e. to select randomly an element $u$ in $F(x)$ while at state $x$).

- Extends to sequences of evaluations with vanishing total variation.
- Case of POMDP with finite sets of actions, states and signals: the uniform value exists in behavior strategies (Rosenberg Solan Vieille 2002)
- Recently, Venel and Ziliotto (2015) showed the existence of the uniform value in pure strategies for such models.

Theorem (R-Venel 2013) Assume the state space is compact metric, payoffs are continuous and transitions are non expansive. Then $(v_n)$ and $(v_\lambda)$ uniformly converge to $v^*$, where for each initial state $x$,

$$v^*(x) = \inf\{w(x), w : \Delta(X) \to [0,1] \text{ affine } C^0 \text{ s.t.}$$

$$(1)\ \forall x' \in X, w(x') \geq \sup_{u \in F(x')} w(u)$$

$$(2)\ \forall u \in R, w(u) \geq r(u)\}.$$

Moreover, the uniform value exists if $F$ has convex values (or if one allows the player to play a behavior strategy, i.e. to select randomly an element $u$ in $F(x)$ while at state $x$).

- Extends to sequences of evaluations with vanishing total variation.
- Case of POMDP with finite sets of actions, states and signals: the uniform value exists in behavior strategies (Rosenberg Solan Vieille 2002)
- Recently, Venel and Ziliotto (2015) showed the existence of the uniform value in pure strategies for such models.

# UCV of $(v_n)$ and $(v_\lambda)$ are equivalent.

Consider a compact continuous stochastic game: $K$, $I$ and $J$ are compact metric spaces, the transition $q : K \times I \times J \longrightarrow \Delta(K)$ and the payoff $g : K \times I \times J \longrightarrow \mathbb{R}$ are jointly continuous. One can show that for each $n$ and $\lambda$, $v_n$ and $v_\lambda$ exists and satisfy the Shapley equations.

Theorem (Ziliotto, 2015) In a compact continuous stochastic game, $(v_n)$ converges uniformly if and only if $(v_\lambda)$ converges uniformly. In case of convergence, the limit is the same.

Ziliotto also showed that this result extends to the general case of a stochastic game where:

- $K$, $I$ and $J$ are Borel subsets of Polish spaces, $q$ and $g$ are Borel measurable and $g$ is bounded.
- For each $n \geq 1$ and each $\lambda \in (0,1]$, the corresponding stochastic game has a value which is measurable with respect to the initial state, and such that the above Shapley equations holds.
- For each Borel measurable bounded function $f$ from $K$ to $\mathbb{R}$, its image $\Psi(f)$ by the Shapley operator, defined by: $\forall k \in K, \Psi(f)(k) = \sup_{x \in \Delta(I)} \inf_{y \in \Delta(J)} \left(\lambda\, g(k,x,y) + (1-\lambda)\, \mathbb{E}_{q(k,x,y)}(f)\right)$, is also Borel.

# UCV of $(v_n)$ and $(v_\lambda)$ are equivalent.

Consider a compact continuous stochastic game: $K$, $I$ and $J$ are compact metric spaces, the transition $q : K \times I \times J \longrightarrow \Delta(K)$ and the payoff $g : K \times I \times J \longrightarrow \mathbb{R}$ are jointly continuous. One can show that for each $n$ and $\lambda$, $v_n$ and $v_\lambda$ exists and satisfy the Shapley equations.

Theorem (Ziliotto, 2015) In a compact continuous stochastic game, $(v_n)$ converges uniformly if and only if $(v_\lambda)$ converges uniformly. In case of convergence, the limit is the same.

Ziliotto also showed that this result extends to the general case of a stochastic game where:

- $K$, $I$ and $J$ are Borel subsets of Polish spaces, $q$ and $g$ are Borel measurable and $g$ is bounded.
- For each $n \geq 1$ and each $\lambda \in (0,1]$, the corresponding stochastic game has a value which is measurable with respect to the initial state, and such that the above Shapley equations holds.
- For each Borel measurable bounded function $f$ from $K$ to $\mathbb{R}$, its image $\Psi(f)$ by the Shapley operator, defined by: $\forall k \in K, \Psi(f)(k) = \sup_{x \in \Delta(I)} \inf_{y \in \Delta(J)} \left(\lambda\, g(k,x,y) + (1-\lambda)\, \mathbb{E}_{q(k,x,y)}(f)\right)$, is also Borel.

# UCV of $(v_n)$ and $(v_\lambda)$ are equivalent.

Consider a compact continuous stochastic game: $K$, $I$ and $J$ are compact metric spaces, the transition $q : K \times I \times J \longrightarrow \Delta(K)$ and the payoff $g : K \times I \times J \longrightarrow \mathbb{R}$ are jointly continuous. One can show that for each $n$ and $\lambda$, $v_n$ and $v_\lambda$ exists and satisfy the Shapley equations.

Theorem (Ziliotto, 2015) In a compact continuous stochastic game, $(v_n)$ converges uniformly if and only if $(v_\lambda)$ converges uniformly. In case of convergence, the limit is the same.

Ziliotto also showed that this result extends to the general case of a stochastic game where:

- $K$, $I$ and $J$ are Borel subsets of Polish spaces, $q$ and $g$ are Borel measurable and $g$ is bounded.
- For each $n \geq 1$ and each $\lambda \in (0,1]$, the corresponding stochastic game has a value which is measurable with respect to the initial state, and such that the above Shapley equations holds.
- For each Borel measurable bounded function $f$ from $K$ to $\mathbb{R}$, its image $\Psi(f)$ by the Shapley operator, defined by: $\forall k \in K, \Psi(f)(k) = \sup_{x \in \Delta(I)} \inf_{y \in \Delta(J)} \left(\lambda\, g(k,x,y) + (1-\lambda)\, \mathbb{E}_{q(k,x,y)}(f)\right)$, is also Borel.

# Lack of information on one side: the cav $u$ theorem

Repeated Games with lack of information on one side (Aumann Maschler 1966): a finite family $(G^k)_{k\in K}$ of payoff matrices in $\mathbb{R}^{I\times J}$, and $p\in\Delta(K)$ define a zero-sum repeated game where: first, some $k$ is selected according to $p$, $k$ remains fixed and is told to player 1 only, then $G^k$ is repeated over and over.

As usual, define the value $v_n(p)$ of the $n$-stage game with average payoffs $\mathbb{E}(\frac{1}{n}\sum_{t=1}^{n} G^k(i_t,j_t))$.
Easy to see that $(v_n)_n$ CV and that P2 can uniformly guarantee $\lim_n v_n$.
Can P1 guarantee $\lim_n v_n$ as well ?

Example: 2 states $K=\{a,b\}$, and $p=(1/2,1/2)$.
$G^a=\begin{pmatrix}4&0&2\\4&0&-2\end{pmatrix}$ and $G^b=\begin{pmatrix}0&4&-2\\0&4&2\end{pmatrix}$.

What should do player 1 ? Playing $CR$ or $NR$ guarantees 0.

Recursive formula:

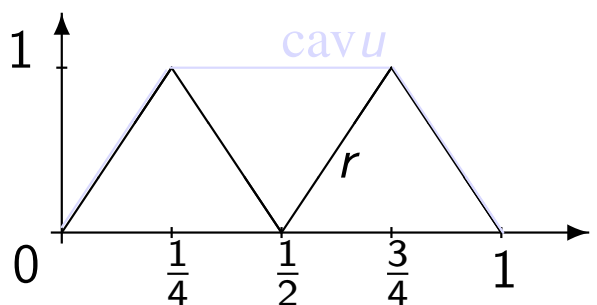
$$v_n(p) = \sup_{x \in \Delta(I)^K} \left( \frac{1}{n} g(p,x) + \frac{n-1}{n} \sum_{i \in I} x(p)(i) v_{n-1}(\hat{p}(x,i)) \right).$$

where $p \in \Delta(K)$, $g(p,x) = \min_j(\sum_k p^k G^k(x^k, j))$ and $\hat{p}(x,i)$ is the conditional belief on $\Delta(K)$ given $p$, $x$, $i$.

Can be written as a leavable gambling house $F : X \rightrightarrows \Delta_f(X)$, where $X$ is the simplex $\Delta(K)$ and $F(x) = \{\mu \in \Delta_f(X), m(\mu) = x\}$.

Well known here (Aumann Maschler 1966): define $u(q) = \mathrm{Val}(\sum_k q^k G^k)$ for each $q$ in $\Delta(K)$, then $(v_n)$ CV to

$$v^* = \mathrm{cav}\, u = \inf\{v : \Delta(K) \to [0,1], v \text{ concave } v \geq u\}$$



Recursive formula:

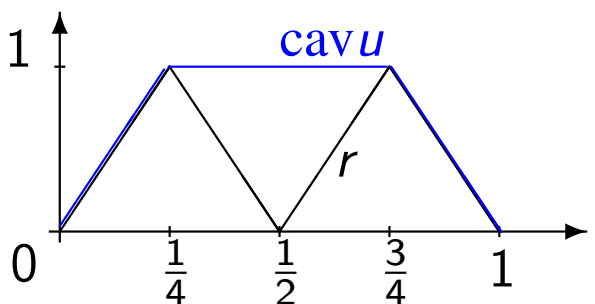
$$v_n(p) = \sup_{x \in \Delta(I)^K} \left( \frac{1}{n} g(p,x) + \frac{n-1}{n} \sum_{i \in I} x(p)(i) v_{n-1}(\hat{p}(x,i)) \right).$$

where $p \in \Delta(K)$, $g(p,x) = \min_j(\sum_k p^k G^k(x^k, j))$ and $\hat{p}(x,i)$ is the conditional belief on $\Delta(K)$ given $p$, $x$, $i$.

Can be written as a leavable gambling house $F : X \rightrightarrows \Delta_f(X)$, where $X$ is the simplex $\Delta(K)$ and $F(x) = \{\mu \in \Delta_f(X), m(\mu) = x\}$.

Well known here (Aumann Maschler 1966): define $u(q) = \mathrm{Val}(\sum_k q^k G^k)$ for each $q$ in $\Delta(K)$, then $(v_n)$ CV to

$$v^* = \mathrm{cav}\, u = \inf\{v : \Delta(K) \to [0,1], v \text{ concave } v \geq u\}$$



# The *cavu* theorem with non observable actions

2 player repeated games with lack of information on one side and signals: $K$ states, $(G^k)_k$ payoffs, actions $I$ and $J$ and now signals $C$ and $D$ and a signalling function $l : K \times I \times J \longrightarrow \Delta(C \times D)$.
Define

$$NR(p) = \{x = (x^k)_{k\in K} \in \Delta(I)^K, \forall k \in K, \forall k' \in K \text{ s.t. } p^k p^{k'} > 0, \forall j \in J, \sum_{i\in I} x_i^k l_2(k,i,j) = \sum_{i\in I} x_i^{k'} l_2(k',i,j)\}.$$

If player 1 plays a strategy $x$ in $NR(p)$, the *a posteriori* of player 2 will remain a.s. constant: player 2 can deduce no information on the selected state $k$. The value of the non revealing game becomes:

$$u(p) = \max_{x\in NR(p)} \min_{y\in\Delta(J)} \sum_{k\in K} p^k G^k(x^k, y) = \min_{y\in\Delta(J)} \max_{x\in NR(p)} \sum_{k\in K} p^k G^k(x^k, y),$$

whith $u(p) = -\infty$ if $NR(p) = \emptyset$.

Theorem (Aumann Maschler): The repeated game with initial probability $p$ has a uniform value given by $cavu(p)$.

# The *cavu* theorem with non observable actions

2 player repeated games with lack of information on one side and signals: $K$ states, $(G^k)_k$ payoffs, actions $I$ and $J$ and now signals $C$ and $D$ and a signalling function $l : K \times I \times J \longrightarrow \Delta(C \times D)$.
Define

$$NR(p) = \{x = (x^k)_{k\in K} \in \Delta(I)^K, \forall k \in K, \forall k' \in K \text{ s.t. } p^k p^{k'} > 0,\ \forall j \in J, \sum_{i\in I} x_i^k l_2(k,i,j) = \sum_{i\in I} x_i^{k'} l_2(k',i,j)\}.$$

If player 1 plays a strategy $x$ in $NR(p)$, the *a posteriori* of player 2 will remain a.s. constant: player 2 can deduce no information on the selected state $k$. The value of the non revealing game becomes:

$$u(p) = \max_{x\in NR(p)} \min_{y\in\Delta(J)} \sum_{k\in K} p^k G^k(x^k, y) = \min_{y\in\Delta(J)} \max_{x\in NR(p)} \sum_{k\in K} p^k G^k(x^k, y),$$

whith $u(p) = -\infty$ if $NR(p) = \emptyset$.
Theorem (Aumann Maschler): The repeated game with initial probability $p$ has a uniform value given by $\text{cav}u(p)$.

# The value of repeated games with an informed controller

General zero-sum repeated game: Markov Dynamic Games (MDG).

- Five non empty and finite sets

a set of states: $K$,
sets of actions: $I$ for player 1, and $J$ for player 2,
sets of signals: $C$ for player 1, and $D$ for player 2.

- an initial distribution $\pi \in \Delta(K \times C \times D)$,
- a payoff function $g$ from $K \times I \times J$ to $[0,1]$,
- and a transition $q$ from $K \times I \times J$ to $\Delta(K \times C \times D)$.

At stage 1: $(k_1, c_1, d_1)$ is selected according to $\pi$, player 1 learns $c_1$ and player 2 learns $d_1$. Then simultaneously player 1 chooses $i_1$ in $I$ and player 2 chooses $j_1$ in $J$. The payoff for player 1 is $g(k_1, i_1, j_1)$.

At any stage $t \geq 2$: $(k_t, c_t, d_t)$ is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, player 1 learns $c_t$ and player 2 learns $d_t$. Simultaneously, player 1 chooses $i_t$ in $I$ and player 2 chooses $j_t$ in $J$. The stage payoff for player 1 is $g(k_t, i_t, j_t)$.

A pair of behavioral strategies $(\sigma, \tau)$ induces a probability over plays.
Existence of $\lim_n v_n$ and $\lim v_\lambda$ ? of the uniform value ?

**Hypothesis HX**: Player 1 is informed, in the sense that he can always deduce the state and player 2's signal from his own signal.

Under $HX$, player 1 can always compute the initial belief $p$ of player 2 on the initial state $k_1$.

**Hypothesis HY**: Player 1 controls the transition, in the sense that the marginal of the transition $q$ on $K \times D$ does not depend on player 2's action.

A pair of behavioral strategies $(\sigma, \tau)$ induces a probability over plays.
Existence of $\lim_n v_n$ and $\lim v_\lambda$ ? of the uniform value ?

**Hypothesis HX**: Player 1 is informed, in the sense that he can always deduce the state and player 2's signal from his own signal.

Under $HX$, player 1 can always compute the initial belief $p$ of player 2 on the initial state $k_1$.

**Hypothesis HY**: Player 1 controls the transition, in the sense that the marginal of the transition $q$ on $K \times D$ does not depend on player 2's action.

Theorem (R- 2012): Under HX and HY, the repeated game has a uniform value.

Theorem (R-V 2013): Under the same hyp, the limit value is:

$$\forall p \in X,\ v^*(p) = \inf\{w(p), w : \Delta(X) \to [0,1] \text{ affine } C^0 \text{ s.t.}$$

$$(1)\ \forall p' \in X, w(p') \geq \sup_{a \in \Delta(I)^K} w(q(p',a))$$

$$(2)\ \forall (u,y) \in RR, w(u) \geq y\}.$$

Where $RR = \{(u,y) \in \Delta(X) \times [0,1], \text{there exists } a : X \to \Delta(I)^K \text{ measurable s.t.}$

$$\int_{p \in X} q(p,a(p))du(p) = u \text{ and } \int_{p \in X} \min_{j \in J} g(p,a(p),j)du(p) = y \quad \}.$$

Remarks:

- extends to the case of sequence of evaluations with vanishing TV.
- existence of the uniform value extended to the case where Player 1 controls the transitions and has more info on the state than P2 (Gensbittel, Oliu-Barton, Venel 2014).

# Lack of information on both sides

$K$ (resp. $L$) is the finite set of private states for P1 (resp. $P2$).
A finite family $(G^{k,l})_{(k,l)\in K\times L}$ of payoff matrices in $I\!R^{I\times J}$, $p \in \Delta(K)$ and $q$ in $\Delta(L)$ define a zero-sum repeated game where: first, $(k,l)$ is selected according to $p \otimes q$, $k$ is told to player 1 and $l$ is told to P2. Then $G^{k,l}$ is repeated over and over.

Define the non revealing value function as:
$u(p,q) = \mathrm{Val}_{\Delta(I)\times\Delta(J)}(\sum_{k,l} p^k q^l G^{k,l})$.

Theorem (Aumann Maschler Stearns 1967): The greatest quantity which can be guaranteed by player 1 is $\mathrm{cav_I}\,\mathrm{vex_{II}}\,u(p,q)$, and the smallest quantity which can be guaranteed by player 2 is $\mathrm{vex_{II}}\,\mathrm{cav_I}\,u(p,q)$. The uniform value may fail to exist.

# Lack of information on both sides

$K$ (resp. $L$) is the finite set of private states for P1 (resp. $P2$).
A finite family $(G^{k,l})_{(k,l)\in K\times L}$ of payoff matrices in $I\!R^{I\times J}$, $p \in \Delta(K)$ and $q$ in $\Delta(L)$ define a zero-sum repeated game where: first, $(k,l)$ is selected according to $p \otimes q$, $k$ is told to player 1 and $l$ is told to P2. Then $G^{k,l}$ is repeated over and over.

Define the non revealing value function as:
$u(p,q) = \mathrm{Val}_{\Delta(I)\times\Delta(J)}(\sum_{k,l} p^k q^l G^{k,l})$.

Theorem (Aumann Maschler Stearns 1967): The greatest quantity which can be guaranteed by player 1 is $\mathrm{cav_I}\,\mathrm{vex_{II}}\,u(p,q)$, and the smallest quantity which can be guaranteed by player 2 is $\mathrm{vex_{II}}\,\mathrm{cav_I}\,u(p,q)$. The uniform value may fail to exist.

Theorem (Mertens-Zamir 1971): $(v_n)$ and $(v_\lambda)$ uniformly converges to the unique continuous function $v$ such that:

$$\begin{cases} v & = & \mathrm{vex}_{\mathrm{II}} \max\{u, v\} \\ v & = & \mathrm{cav}_{\mathrm{I}} \min\{u, v\} \end{cases}$$

Extends to:
1) the case of signals independent of the states,
2) the case of correlated initial information,
3) the case where states are not fixed but follows independent Markov chains (Gensbittel R, 2015).

- Oliu-Barton (2015) showed that the associated *Splitting Game* defined on $\Delta(K) \times \Delta(L)$ has a uniform value.
- Extension: Laraki R. 2015, to be presented this week.

Theorem (Mertens-Zamir 1971): $(v_n)$ and $(v_\lambda)$ uniformly converges to the unique continuous function $v$ such that:

$$\begin{cases} v & = & \mathrm{vex}_{\mathrm{II}} \max\{u, v\} \\ v & = & \mathrm{cav}_{\mathrm{I}} \min\{u, v\} \end{cases}$$

Extends to:
1) the case of signals independent of the states,
2) the case of correlated initial information,
3) the case where states are not fixed but follows independent Markov chains (Gensbittel R, 2015).

• Oliu-Barton (2015) showed that the associated *Splitting Game* defined on $\Delta(K) \times \Delta(L)$ has a uniform value.
• Extension: Laraki R. 2015, to be presented this week.

# Some open problems

1. Computing the value.

In the basic model.

In repeated game with incomplete info on one side. $K = \{a, b\}$, $p = (1/2, 1/2)$, $M = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}$, $G^a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $G^b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

If $\alpha = 1$, the value is $1/4$ (Aumann Maschler).

If $\alpha \in [1/2, 2/3]$, the value is $\frac{\alpha}{4\alpha - 1}$ (Hörner *et al.* 2010, Marino 2005 for $\alpha = 2/3$).

For $\alpha \in [2/3, .73]$ (Bressaud Quas 2013): $\frac{1}{v} = u_0 + u_0 u_1 + u_0 u_1 u_2 + \ldots$, where $(u_n)$ is defined by $u_0 = 1$ and $u_{n+1} = max\{\psi(u_n), 1 - \psi(u_n)\}$ with $\psi(u) = 3\alpha - 1 - \frac{2\alpha - 1}{u}$.

What is the value for $\alpha = 0.9$ ?

2. Existence of the limit value in repeated games with lack of information on both sides and general state-dependent signaling ?

3. Find conditions for compact non expansive stochastic games to have a limit value (Bolte Gaubert Vigeral: semi-algebraicity, Laraki Renault: acyclicty...)

4. Finite MDG: Find other value functions which will always converge ? Continous-time games (Neyman) ?

5. How large is the set of information structures ? $K$ is a fixed finite set of parameters.
An information structure is an element $u$ in $Z := \Delta_f(K \times I\!N \times I\!N)$.
Interpretation: $u$ is publicly known, $(k, c, d)$ is selected according to $u$, player 1 learns $c$ and player 2 learns $d$.

How to evaluate an information structure ?

A payoff structure is a mapping $g : K \times I\!N \times I\!N \longrightarrow [-1, 1]$
s.t. for some $L$: $g(k, i, j) = -1$ if $i > L$ and $j \leq L$, and $g(k, i, j) = +1$ if $i \leq L$ and $j > L$.

Given $u$ and $g$, denote by $\mathrm{val}(u, g)$ the value of the zero-sum game where:

- $(k, c, d)$ is selected to $u$, player 1 learns $c$ and player 2 learns $d$.
- Then simultaneously player 1 chooses $i$ in $I\!N$, player 2 chooses $j$ in $I\!N$, and player 1's payoff is $g(k, i, j)$.

Define (Gensbittel R.): $d^*(u, v) = \sup_g |\mathrm{val}(u, g) - \mathrm{val}(v, g)|$.
Let $Z^*$ be the quotient space of $Z$. $(Z^*, d^*)$ is a metric space, is it totally bounded ?

## An equivalent problem for $|K| = 2$:

Given $\varepsilon > 0$, does there exist $p$ such that any bimatrix $(A, B) = (a_{i,j}, b_{i,j})$ in $([-1,1]^2)^{I \times J}$ of arbitrary dimension can be approximated by a $p \times p$ bimatrix in the following sense ?

one can find $x_1,...,x_p$ in $\Delta(I)$, $y_1,...,y_p$ in $\Delta(J)$ such that

1) for each $i$ in $I$, there exists $x^*(i)$ a convex combination of $x_1,...,x_p$ satisfying $\forall j = 1,...,p$, $iAy_j \leq x^*(i)Ay_j + \varepsilon$ and $iBy_j \leq x^*(i)By_j + \varepsilon$, and

2) for each $j$ in $J$, there exists $y^*(j)$ a convex combination of $y_1,...,y_p$ satisfying $\forall i = 1,...p$, $x_iAj \geq x_iAy^*(j) - \varepsilon$ and $x_iBj \geq x_iBy^*(j) - \varepsilon$.

6. Basic Model, non zero-sum case. Existence of a uniform equilibrium payoff ? i.e. of $x$ in $\mathbb{R}^N$ such that $\forall \varepsilon > 0, \exists \sigma = (\sigma^i)_{i \in N}, \exists n_0$ satisfying;

$$\forall n \geq n_0, \forall i \in N, \forall \tau^i, \gamma_n^i(\tau^i, \sigma^{-i}) \leq x^i + \varepsilon \text{ and } \gamma_n^i(\sigma) \geq x^i - \varepsilon.$$

Positive for 2 players (Vieille 00), for 3 players absorbing games (Solan 99).
Even unknown in the case of $n$-player quitting games, with $n \geq 4$: at each stage, each player decides to stop or continue. Whenever at least one player stops, the game is absorbed and each player receives a payoff depending on the set of stopping players.

Warning: for non zero-sum stochastic games, the set of uniform equilibrium payoffs and the limit set of discounted equilibrium payoffs may be disjoint (Sorin, 1986).

Many other interesting things:

- stochastic games with Borel payoff functions (Martin 1975, 1998, Gimbert et al. 2014)
- discounted stochastic games with general state spaces (Nowak 2003, Solan 1998...)
- limiting average value and $\varepsilon$-optimal stationary strategies (Thuijsman Vrieze 1991, 1992, Flesch Thuijsman Vrieze 1998...)
- continuous-time stochastic games (Neyman 2012...)
- maxmin and minmax of stochastic games with unobserved actions (Coulomb 2003, Rosenberg Solan Vieille 2003)
- Big Match with lack to information on one side (Sorin 1984, 1985), stochastic games with incomplete information...
- .....