Stochastic Methods in Game Theory — Jérôme Renault
Institute of Mathematical Sciences, Singapour — November 2015

Zero-sum Stochastic Games

Theses notes, to be partially covered during the tutorial, concern the theory of zero-sum stochastic games. They focus on long-term games played in discrete time. The first part contains the fundamental basic results of the theory. The second part presents a few extensions, recent results and open problems and is obviously biased towards my own tastes, research interests and knowledge.

Outline:

1. The basic model: Stochastic games with finitely many states and actions.

1.1 Description, examples.
1.2 The n-stage game and the λ-discounted game
1.3 Limit value, the algebraic approach
1.4 The Big Match. 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.

3. References

1 The basic model

1.1 Description

Zero-sum games are 2-player games where the sum of the payoffs of the players is 0, they are games of pure competition between the players. Zero-sum stochastic games are dynamic zero-sum games played in discrete time. The basic model is due to Shapley (1953), and is given by 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 $g : K \times I \times J \longrightarrow I\!R$, and a transition mapping q from $K \times I \times J$ to the simplex[1] $\Delta(K)$ of probability distributions over K. In the basic model, K, I and J are assumed to be non empty *finite* sets.

The progress of the game is the following:

- The initial state is k_1, known to the players. At stage 1, player 1 and player 2 simultaneously choose $i_1 \in I$ and $j_1 \in J$. Then P1's payoff is $g(k_1, i_1, j_1)$ and P2 's payoff is $-g(k_1, i_1, j_1)$, the actions i_1 et j_1 are publicly announced, and the play proceeds to stage 2.

- A stage $t \geq 2$, the state k_t is selected according to the distribution $q(k_{t-1}, i_{t-1}, j_{t-1})$, and is announced to both players. Player 1 and player 2 then simultaneously choose $i_t \in I$ et $j_t \in J$. P1's payoff is $g(k_t, i_t, j_t)$ and P2's payoff is $-g(k_t, i_t, j_t)$, the actions i_t et j_t are announced, and the play proceeds to stage $t + 1$.

Notations and vocabulary. We denote by $q(k'|k, i, j)$ the 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. A state k is absorbing if $q(k|k, i, j) = 1$ for all (i, j) in $I \times J$ (when k is reached, the play stays there forever). A stochastic game is absorbing if it has a unique non absorbing state.

A play is a sequence $(k_1, i_1, j_1, k_2, i_2, j_2, ..., k_t, i_t, j_t, ...)$ taking values in $K \times I \times J$. 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, or simply a 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 said to be pure if it associates to each history a Dirac measure, that is an element of I for player 1 and an element of J for player 2.

A strategy is said to be Markov if for any stage t, the mixed action prescribed at stage t only depends on the current state k_t (and not on past states or past actions). A stationary strategy is a Markov strategy such that the mixed action prescribed after any history only depends on the current state (and not on the stage number).

[1]When S is a finite set, we denote by $\Delta(S)$ the set of probability distributions over S. More generally, we will later use the notation $\Delta(S)$ for the set of Borel probability measures on a compact metric set S.

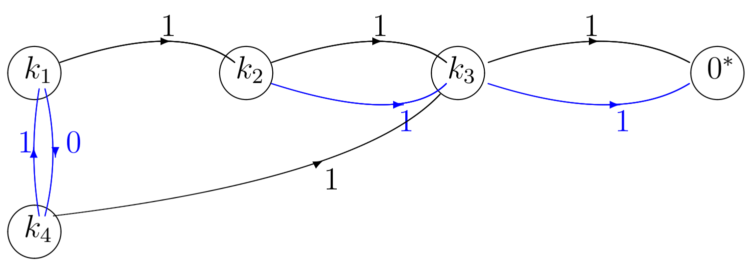
In all the examples an absorbing state will be denoted with a *. For instance, 3* represents an absorbing state where the payoff to player 1 is 3, whatever the actions played.

Example 1:

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

There is a unique non absorbing state which is the initial state. Actions are T and B for player 1, L and R for player 2. If at the first stage the action profile played is (T, L) then the stage payoff is 0 and the play goes to the next stage without changing state. If at the first stage the action profile played is (T, R) or (B, L), the play reaches an absorbing state where at each subsequent stage, whatever the actions played the payoff of player 1 will be 1. If at the first stage the action profile played is (B, R), the play reaches an absorbing state where at each subsequent stage, whatever the actions played the payoff of player 1 will be 0.

Example 2: A one-player game (J is a singleton), with deterministic transitions and actions Black and Blue for Player 1. The payoffs are either 1 or 0 in each case.



Example 3: The "Big Match" :

$$\begin{pmatrix} 1^* & 0^* \\ 0 & 1 \end{pmatrix}$$

We denote by Σ and $\mathcal{T}$ the sets of strategies of player 1 and 2, respectively. A couple of strategies in $\Sigma \times \mathcal{T}$ naturally[2] induces a probability distribution $\mathbf{P}_{k_1,\sigma,\tau}$ over the set of plays $\Omega = (K \times I \times J)^\infty$, endowed with the product σ-algebra. We will denote the expectation with respect to $\mathbf{P}_{k_1,\sigma,\tau}$ by $\mathbb{E}_{k_1,\sigma,\tau}$.

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

[2]just as tossing a coin at every stage induces a probability distribution over sequences of Heads and Tails.

any behavior strategy σ of player 1 there exists a mixed strategy σ' of this player such that, for any pure (or mixed, or behavior) strategy τ of player 2, (σ, τ) and (σ', τ) induce the same probabilities over plays. And vice-versa by exchanging the words "mixed" and "behavior" in the last sentence. Idem by exchanging the roles of player 1 and player 2 above.

1.2 The n-stage game and the λ-discounted game

Definition 1.1. *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 Σ and $\mathcal{T}$, and payoff function:*

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

Because only finitely many stages matter here, $\Gamma_n(k_1)$ can be equivalently seen as a finite zero-sum game played with mixed strategies. Hence it has a value denoted by $v_n(k_1) = \max_{\sigma \in \Sigma} \min_{\tau \in \mathcal{T}} \gamma_n^{k_1}(\sigma, \tau) = \min_{\tau \in \mathcal{T}} \max_{\sigma \in \Sigma} \gamma_n^{k_1}(\sigma, \tau)$. For convenience we write $v_0(k) = 0$ for each k.

Definition 1.2. *Given a discount rate λ in $(0, 1]$, the λ-discounted game with initial state k_1 is the zero-sum game $\Gamma_\lambda(k_1)$ with strategy spaces Σ and $\mathcal{T}$, and payoff function:*

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

By a variant of Sion theorem, it has a value denoted by $v_\lambda(k_1)$. In the economic literature $\delta = 1 - \lambda = \frac{1}{1+r}$ is called the discount factor, r being called the interest rate.

Proposition 1.3. *v_n and v_λ are characterized by the following Shapley equations.*

1) 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.

2) For λ 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 λ-discounted game, players have stationary optimal strategies.

Proof: The proof is standard. For 1), fix n and k and denote by v the value of the matrix game $\left(g(k,i,j)+\sum_{k'\in K} q(k'|k,i,j)\; n\; v_n(k')\right)_{i,j}$. In the game with $n+1$ stages and initial state k, player 1 can play at stage 1 an optimal strategy in this matrix game, then from stage 2 on an optimal strategy in the remaining n-stage stochastic game. By doing so, player 1 guarantees v in $\Gamma_{n+1}(k)$, so $v_{n+1}(k) \geq v$. Proceeding similarly with player 2 gives $v_{n+1}(k) = v$.

The proof of 2) is similar. Notice that by the contracting fixed point theorem, for fixed λ the vector $(v_\lambda(k))_{k\in K}$ is uniquely characterized by the Shapley equations.

It is easy to compute v_n and v_λ in the previous examples (in absorbing games, we simply write v_n and v_λ for the values of the stochastic game where the initial state is the non absorbing state)

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 λ.

Example 2: For λ small enough, $v_\lambda(k_1) = \frac{1-\lambda}{2-\lambda}$ and it is optimal in the λ-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).

Example 3 (The Big Match): $v_n = v_\lambda = 1/2$ for all n and λ.

The Shapley operator is defined as the mapping which associates to each v in $I\!R^K$ the vector $\Psi(v)$ in $I\!R^K$ such that for each k,

$$\Psi(v)^k = \mathrm{Val}_{\Delta(I)\times\Delta(J)}\left(g(k,i,j) + \sum_{k'\in K} q(k'|k,i,j)\; v^{k'}\right).$$

Ψ is non expansive for the sup-norm $\|v\| = \sup_{k\in K}|v^k|$ on $I\!R^K$, and the Shapley equations can be rephrased as:

$$\forall n \geq 1,\; nv_n = \Psi((n-1)v_{n-1}) = \Psi^n(0),$$

$$\forall \lambda \in (0,1], v_\lambda = \lambda\Psi\left(\frac{1-\lambda}{\lambda}v_\lambda\right).$$

1.3 Limit values - The algebraic approach

We are interested here in the limit values when the players become more and more patient, i.e. in the existence of the limits of v_n, when n goes to infinity, and of v_λ, when λ goes to 0.

It is always interesting to study first the 1-player case.

1.3.1 1-player case: Markov Decision Process

We assume here that player 2 does not exist, that is J is a singleton. For any $\lambda > 0$, player 1 has an optimal stationary strategy in the λ-discounted game. Moreover since the matrix games appearing in 2) of the Shapley equations only have one column, this stationary optimal strategy can be taken to be pure. So we just have to consider strategies given by a mapping $f : K \longrightarrow I$, with the interpretation that player 1 plays $f(k)$ whenever the current state is k.

The λ-discounted payoff when f is played and the initial state is k satisfies:

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

Consider the vector $v = (\gamma_\lambda^k(f))_k$. The above equations can be written in matrix form: $(I - (1-\lambda)A)v = \lambda\alpha$, where I is the identity matrix, $A = (q(k'|k, f(k))_{k,k'}$ is a stochastic matrix independent of λ, and $\alpha = (g(k, f(k)))_k$ is a fixed vector. $(I-(1-\lambda)A)$ being invertible, we know that its inverse has coefficients which are rational fractions of its coefficients. Consequently, we obtain that:

For a given pure stationary strategy f, the payoff $\gamma_\lambda^k(f)$ is a rational fraction of λ.

Now we have finitely many such strategies to consider, and a given f is optimal in the λ-discounted game with initial state k if and only if: $\gamma_\lambda^k(f) \geq \gamma_\lambda^k(f')$ for all f'. Because a non-zero polynomial only has finitely many roots, we obtain that for λ small enough, the same pure optimal strategy f has to be optimal in any discounted game. And clearly f can be taken to be optimal whatever the initial state is.

Theorem 1.4. *(Blackwell, 1962) In the 1-player case, there exists $\lambda_0 > 0$ and a pure stationary optimal strategy f which is optimal in any λ-discounted game with $\lambda \leq \lambda_0$. For $\lambda \leq \lambda_0$ and k in K, the value $v_\lambda(k)$ is a bounded rational fraction of λ, hence converges when λ goes to 0.*

In example 2, 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.

1.3.2 Stochastic games: The algebraic approach

Back to the 2-player case, we know that in each discounted game the players have stationary optimal strategies. The following approach[3] is due to Bewley and Kohlberg (1976). 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:

[3] M. Oliu-Barton (2014) provided a proof of the convergence of v_λ using elementary tools.

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

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

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

In particular, the set A is semi-algebraic[4]. One can show that the projection of a semi-algebraic set (keeping a smaller number of coordinates) is still semi-algebraic (Tarski-Seidenberg elimination theorem), so $A^*=\{(\lambda,v_\lambda),\lambda\in(0,1]\}$ is also a semi-algebraic subset of $I\!R\times I\!R^K$. This implies the existence of a bounded Puiseux series development of v_λ in a neighborhood of $\lambda=0$.

Theorem 1.5. *(Bewley Kohlberg) 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:*

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

So when λ is close to 0, for each k $v_\lambda(k)$ is a power series of $\lambda^{1/M}$.

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

Corollary 1.6.

1) v_λ converges when λ goes to 0.

2) v_λ has bounded variation at 0, i.e. for any sequence $(\lambda_i)_{i\geq 1}$ of discount factors decreasing to 0, we have $\sum_{i\geq 1}\|v_{\lambda_{i+1}}-v_{\lambda_i}\|_\infty<\infty$.

3) v_n also converges, and $\lim_{n\to\infty}v_n=\lim_{\lambda\to 0}v_\lambda$.

Proof: 1) is clear by the Puiseux series development.

2) also comes from this development. Fix k in K. When λ is small enough, $v_\lambda(k)=f_k(\lambda^{1/M})$ where f_k is a power series with positive radius of convergence, hence

$$\frac{\partial v_\lambda(k)}{\partial\lambda}(\lambda)=\frac{1}{M}f_k'(\lambda^{1/M})\lambda^{1/M-1}.$$

so that there exists a bound C such that for λ small enough, $|\frac{\partial v_\lambda(k)}{\partial\lambda}(\lambda)|\leq C\lambda^{1/M-1}$. Now, if $0<\lambda_2<\lambda_1$, $|v_{\lambda_1}(k)-v_{\lambda_2}(k)|\leq\int_{\lambda_2}^{\lambda_1}C\lambda^{1/M-1}d\lambda=CM(\lambda_1^{1/M}-\lambda_2^{1/M})$, and the result follows.

[4] A subset of an Euclidean space is semi-algebraic if it 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.

3) The idea[5] is to 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 Ψ 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:

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

And summing these inequalities from $n = 1$ to m gives:

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

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. By the bounded variation property, we have $\sum_{n=1}^{\infty} \|w_{n+1} - w_n\| < \infty$. We conclude that $\|w_{m+1} - v_{m+1}\| \xrightarrow[m\to\infty]{} 0$.

Bewley and Kohlberg also provided an example where v_n is equivalent to $\frac{\ln n}{n}$ when n goes to infinity.

1.4 Uniform value

We fix here the initial state k_1, and omit the dependance on the initial state for a while. We know that $\lim_n v_n = \lim_\lambda v_\lambda$ exists, so we approximately know the value of the stochastic game when n is large and known to the players (and when λ is small and known to the players). But this does not tell us if the players can play approximately well when they do not know exactly how large is n or how small is λ. Do the players have nearly optimal strategies that are robust with respect to the time horizon or the discount factor ? This property is captured by the notion of *uniform value*, which might be considered as the nectar of stochastic games.

Definition 1.7. *Let v be a real number.*

Player 1 can uniformly guarantee v in the stochastic game if: $\forall \varepsilon > 0$, $\exists \sigma \in \Sigma$, $\exists n_0$, $\forall n \geq n_0$, $\forall \tau \in \mathcal{T}$, $\gamma_n(\sigma,\tau) \geq v - \varepsilon$.

Player 2 can uniformly guarantee v in the stochastic game if: $\forall \varepsilon > 0$, $\exists \tau \in \mathcal{T}$, $\exists n_0$, $\forall n \geq n_0$, $\forall \sigma \in \Sigma$, $\gamma_n(\sigma,\tau) \leq v + \varepsilon$.

If v can be uniformly guaranteed by both players, then v is called the uniform value of the stochastic game.

[5] The following proof is, I believe, due to A. Neyman.

It is easily shown that the uniform value, whenever it exists, is unique. The largest quantity uniformly guaranteed by Player 1, resp. smallest quantity uniformly guaranteed by Player 2, can be denoted by:

$$\underline{v} = \sup_{\sigma} \liminf_{n} \left(\inf_{\tau} \gamma_n(\sigma, \tau) \right), \ \overline{v} = \inf_{\tau} \limsup_{n} \left(\sup_{\sigma} \gamma_n(\sigma, \tau) \right).$$

Plainly, $\underline{v} \leq \lim_n v_n \leq \overline{v}$. The uniform value exists if and only if $\underline{v} = \overline{v}$. Whenever it exists it is equal to $\lim_n v_n = \lim_\lambda v_\lambda$, and for each $\varepsilon > 0$ there exists $\lambda_0 > 0$, σ and τ such that for all $\lambda \leq \lambda_0$, σ' and τ' we have: $\gamma_\lambda(\sigma, \tau') \geq v - \varepsilon$ and $\gamma_\lambda(\sigma', \tau) \leq v + \varepsilon$.

1.4.1 The Big Match

The Big Match is the absorbing stochastic game described by:

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

It was introduced by Gillette in 1957. We have seen that $\lim v_n = \lim v_\lambda = 1/2$ here. It is easy to see that player 2 can uniformly guarantee 1/2 by playing at each stage the mixed action 1/2 L +1/2 R independently of everything. It is less easy to see what can be uniformly guaranteed by player 1, and one can show that no stationary or Markov strategy of Player 1 can uniformly guarantee a positive number here. However, Blackwell and Ferguson (1968) proved that the uniform value of the Big Match exists.

Proposition 1.8. *The Big Match has a uniform value*

Proof: All we have to do is prove that Player 1 can uniformly guarantee 1/2. First define the following random variables, for all positive integer t: g_t is the payoff of player 1 at stage t, $i_t \in \{T, B\}$ is the action played by player 1 at stage t, $j_t \in \{L, R\}$ is the action played by player 2 at stage t, $L_t = \sum_{s=1}^{t-1} \mathbf{1}_{j_s=L}$ is the number of stages in 1,...,$t-1$ where player 2 has played L, $R_t = \sum_{s=1}^{t-1} \mathbf{1}_{j_s=R} = t - 1 - L_t$ is the number of stages in 1,...,$t-1$ where player 2 has played 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 (a positive integer) let us define the following strategy σ_M of player 1: at any stage t, σ_M plays T with probability $\frac{1}{(m_t+M+1)^2}$, and B with the remaining probability.

Some intuition for σ_M can be given. Assume we are still in the non absorbing state at stage t. If player 2 has played R often at past stages, player 1 is doing well and has received good payoffs, m_t is large and σ_M plays the risky action T with small probability. On the other hand if Player 2 is playing L often, player 1 has received low payoffs but Player 2 is taking high risks; m_t is small and σ_M plays the risky action T with high probability.

Notice that σ_M is well defined. If $m_t = -M$ then σ_M plays T with probability 1 at stage t and then the game is over. So the event $m_t \leq -M-1$ has probability 0 as long as the play is in the non absorbing state. At any stage t in the non absorbing state, we have $-M \leq m_t \leq t-1$, and σ_M plays T with a probabilty in the interval $[\frac{1}{(M+t)^2}, 1]$.

We will show that σ_M uniformly guarantees $\frac{M}{2(M+1)}$, which is close to $1/2$ for M large. More precisely we will prove that:

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

To conclude the proof of proposition 1.8, we now prove (1). Notice that we can restrict attention to strategies of player 2 which are pure, and (because there is a unique relevant history of moves of player 1) independent of the history. That is, we can assume w.l.o.g. that player 2 plays a fixed deterministic sequence $y = (j_1, ...j_t, ...) \in \{L, R\}^{\infty}$.

T being fixed until the end of the proof, we define the random variable t^* as the time of absorption:

$$t^* = \inf\{s \in \{1, ..., T\}, i_s = T\},\ \text{with the convention } t^* = T+1 \text{ if } \forall s \in \{1, ..., T\}, i_s = B$$

Recall that $R_t = m_t + L_t = t - 1 - L_t$, so that $R_t = \frac{1}{2}(m_t + t - 1)$. 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 following fictitious payoff of player 1: $X_t = 1/2$ 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 $j_{t^*} = R$. X_t is the random variable of the limit value of the current state.

A simple computation shows:

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

To prove (1), it is thus enough to show the following lemma.

Lemma 1.9. *For all t in $\{1,...,T\}$, y in $\{L,R\}^\infty$ and $M \geq 1$, $I\!E_{\sigma_M,y}\left(X_t\right) \geq \frac{M}{2(M+1)}$.*

Proof of the lemma. The proof is by induction on t. For $t=1$, $I\!E_{\sigma_M,y}\left(X_1\right) = \frac{1}{2}(1-\frac{1}{(M+1)^2}) + \frac{1}{(M+1)^2}\mathbf{1}_{j_1=L} \geq \frac{1}{2}(1-\frac{1}{(M+1)^2}) \geq \frac{M}{2(M+1)}$.

Assume the lemma is true for $t \in \{1,...,T-1\}$. 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 $j_1 = L$, $I\!E_{\sigma_M,y}\left(X_{t+1}\right) = \frac{1}{(M+1)^2}1 + (1-\frac{1}{(M+1)^2})I\!E_{\sigma_{M-1},y_+}(X_t)$. By the induction hypothesis, $I\!E_{\sigma_{M-1},y_+}(X_t) \geq \frac{M-1}{2M}$, so $I\!E_{\sigma_M,y}\left(X_{t+1}\right) \geq \frac{M}{2(M+1)}$. Otherwise $j_1 = R$, and $I\!E_{\sigma_M,y}\left(X_{t+1}\right) = (1-\frac{1}{(M+1)^2})I\!E_{\sigma_{M+1},y_+}(X_t)$ $\geq (1-\frac{1}{(M+1)^2})\frac{M+1}{2(M+2)} = \frac{M}{2(M+1)}$. The lemma is proved, concluding the proof of proposition 1.8.

Remark: It is crucial here that player 1 observes at the end of every stage the action played by player 2. In the variant of the Big Match where Player 1 can not observe at all the actions played by player 2, the n-stage and δ-discounted values are still the same, but one can easily show that the uniform value does not exist anymore.

1.4.2 The existence result

The following theorem is due to J-F. Mertens and A. Neyman (1981).

Theorem 1.10. *(Mertens Neyman 1981)*
Every zero-sum stochastic game with finitely many states and actions has a uniform value.

The rest of this section is devoted to the proof of theorem 1.10. Without loss of generality we assume that all payoffs are in $[0,1]$, and fix $\varepsilon \in (0,1)$ in the sequel.

We know by the algebraic approach that there exists $C > 0$, $M \geq 1$, $\lambda_0 > 0$ such that for all $0 < \lambda_1 < \lambda_2 \leq \lambda_0$:

$$\|v_{\lambda_1} - v_{\lambda_2}\| \leq \int_{\lambda_1}^{\lambda_2} \psi(s)ds \ \text{ with } \ \psi(s) = \frac{C}{s^{1-1/M}}.$$

All is needed about ψ is that it is non negative and integrable: $\int_0^1 \psi(s)ds < \infty$.

Definition 1.11. *Define the mapping D from $(0,\lambda_0]$ to $I\!R$ by:*

$$D(y) = \frac{12}{\varepsilon}\int_y^{\lambda_0} \frac{\psi(s)}{s}ds + \frac{1}{\sqrt{y}}.$$

The proof of the next lemma is left to the reader.

Lemma 1.12.
a) D is positive, decreasing, $D(y) \xrightarrow[y\to 0]{} +\infty$ and $\int_0^{\lambda_0} D(y)dy < \infty$.
b) $D(y(1-\varepsilon/6)) - D(y) \xrightarrow[y\to 0]{} +\infty$ and $D(y) - D(y(1+\varepsilon/6)) \xrightarrow[y\to 0]{} +\infty$.
c) $yD(y) \xrightarrow[y\to 0]{} 0$.

Definition 1.13. *Define the mapping φ from $[0, \lambda_0]$ to $I\!R$ by:*

$$\varphi(\lambda) = \int_0^{\lambda} D(y)dy - \lambda D(\lambda).$$

Note that φ is increasing and $\varphi(0) = \lim_{\lambda \to 0} \varphi(\lambda) = 0$.

We fix the initial state k_1 and denote the limit value $\lim_{\lambda} v_{\lambda}(k_1)$ by $v(k_1)$. We now define a nice strategy σ for player 1 in the stochastic game with initial state k_1. While playing at some stage $t+1$, player 1 knows the current state k_{t+1} and the previous payoff g_t, he will update a fictitious discount factor λ_{t+1} and play at stage $t+1$ a stationary optimal strategy in the stochastic game with discount factor λ_{t+1} and initial state k_{t+1}. The definition of the sequence of random discount factors $(\lambda_t)_t$ below, joint with the introduction of an auxiliary sequence $(d_t)_t$, will end the definition of σ.

One first chooses $\lambda_1 > 0$ such that: (i) $v_{\lambda_1}(k_1) \geq v(k_1) - \varepsilon$, (ii) $\varphi(\lambda_1) < \varepsilon$, and (iii) $\forall y \in (0, \lambda_1],\ D(y(1 - \varepsilon/6)) - D(y) > 6$ and $D(y) - D(y(1 + \varepsilon/6)) > 6$. Put $d_1 = D(\lambda_1)$, and by induction define for each $t \geq 1$:

$$d_{t+1} = \max\{d_1, d_t + g_t - v_{\lambda_t}(k_{t+1}) + 4\varepsilon\}, \text{ and } \lambda_{t+1} = D^{-1}(d_{t+1}).$$

We have $\lambda_{t+1} \leq \lambda_1$ for each t. Notice that if the current payoff g_t is high, then λ_{t+1} will have a tendency to decrease : player 1 plays in a more patient way. On the contrary if g_t is small then λ_{t+1} will have a tendency to increase : player 1 plays more for the short-run payoffs. σ being defined, we now fix an arbitrary strategy τ of player 2. We simply write $\mathbf{P}$ for $\mathbf{P}_{k_1,\sigma,\tau}$ and $I\!E$ for $I\!E_{k_1,\sigma,\tau}$.

By construction, the following properties hold on every play. The proofs of a), b) and d) are left to the reader.

Lemma 1.14. *For all $t \geq 1$,*

a) $|d_{t+1} - d_t| \leq 6$,

b) $|\lambda_{t+1} - \lambda_t| \leq \frac{\varepsilon \lambda_t}{6}$,

c) $|v_{\lambda_t}(k_{t+1}) - v_{\lambda_{t+1}}(k_{t+1})| \leq \varepsilon \lambda_t$.

d) $d_{t+1} - d_t \leq g_t - v_{\lambda_t}(k_{t+1}) + 4\varepsilon + \mathbf{1}_{\lambda_{t+1} = \lambda_1}$.

Proof of c):

$$\begin{aligned}
|v_{\lambda_t}(k_{t+1}) - v_{\lambda_{t+1}}(k_{t+1})| &\leq \|v_{\lambda_t} - v_{\lambda_{t+1}}\| \\
&\leq \left| \int_{\lambda_t}^{\lambda_{t+1}} \psi(s) ds \right| \\
&\leq \max\{\lambda_t, \lambda_{t+1}\} \left| \int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(s)}{s} ds \right| \\
&\leq 2\lambda_t \left| \int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(s)}{s} ds \right|.
\end{aligned}$$

Now,

$$\int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(s)}{s} ds = \frac{\varepsilon}{12}\left(D(\lambda_t) - \frac{1}{\sqrt{\lambda_t}}\right) - \frac{\varepsilon}{12}\left(D(\lambda_{t+1}) - \frac{1}{\sqrt{\lambda_{t+1}}}\right) = \frac{\varepsilon}{12}\left((d_t - d_{t+1}) + (\frac{1}{\sqrt{\lambda_{t+1}}} - \frac{1}{\sqrt{\lambda_t}})\right).$$

If $\lambda_t \leq \lambda_{t+1}$, $0 \leq \int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(s)}{s} ds \leq \frac{\varepsilon}{2}$ by point a) of lemma 1.14. So $\left|\int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(s)}{s} ds\right| \leq \frac{\varepsilon}{2}$, and this also holds if $\lambda_t \geq \lambda_{t+1}$. We obtain $|v_{\lambda_t}(k_{t+1}) - v_{\lambda_{t+1}}(k_{t+1})| \leq \varepsilon \lambda_t$.

Definition 1.15. *Define the random variable*

$$Z_t = v_{\lambda_t}(k_t) - \varphi(\lambda_t).$$

When λ_t is close to 0, Z_t is close to $v(k_t)$.

Proposition 1.16. $(Z_t)_t$ *is a sub-martingale, and for all* $t \geq 1$*:*

$$\mathbb{E}(Z_t) \geq 2\varepsilon \mathbb{E}(\sum_{s=1}^{t-1} \lambda_s) + Z_1.$$

Proposition 1.16 is the key to Mertens and Neyman's proof. Assume for the moment the proposition and let us see how the proof of the theorem follows.

We have for each $t \geq 1$, $\mathbb{E}(Z_t) \geq Z_1$, so $\mathbb{E}(v_{\lambda_t}(k_t)) \geq v_{\lambda_1}(k_1) - \varphi(\lambda_1) + \mathbb{E}(\varphi(\lambda_t)) \geq v_{\lambda_1}(k_1) - \varphi(\lambda_1)$, so

$$\mathbb{E}(v_{\lambda_t}(k_t)) \geq v_{\lambda_1}(k_1) - \varepsilon. \tag{2}$$

Since $Z_{t+1} \leq 1$, we have by proposition 1.16 that $2\varepsilon \mathbb{E}(\sum_{s=1}^t \lambda_s) \leq 1 - Z_1 \leq 1 + \varepsilon$, so $\mathbb{E}(\sum_{s=1}^t \lambda_s) \leq \frac{1}{\varepsilon}$. We obtain $\mathbb{E}(\sum_{s=1}^t \lambda_1 \mathbf{1}_{\lambda_1 = \lambda_s}) \leq \frac{1}{\varepsilon}$, and

$$\mathbb{E}(\sum_{s=1}^{t} \mathbf{1}_{\lambda_1 = \lambda_s}) \leq \frac{1}{\lambda_1 \varepsilon}. \tag{3}$$

Using d) and c) of lemma 1.14, we have:

$$g_t \geq v_{\lambda_{t+1}}(k_{t+1}) - \varepsilon\lambda_t + (d_{t+1} - d_t) - 4\varepsilon - \mathbf{1}_{\lambda_{t+1}=\lambda_1}.$$

So for each T,

$$I\!\!E\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq I\!\!E\left(\frac{1}{T}v_{\lambda_{t+1}}(k_{t+1})\right) - \varepsilon I\!\!E\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_t\right) + I\!\!E\left(\frac{1}{T}(d_{T+1} - d_1)\right) - 4\varepsilon - \frac{1}{T}I\!\!E\left(\sum_{t=1}^{T}\mathbf{1}_{\lambda_{t+1}=\lambda_1}\right)$$

Unsing the inequalities (2) and (3), we obtain

$$I\!\!E\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq v_{\lambda_1}(k_1) - \varepsilon - \varepsilon - \frac{d_1}{T} - 4\varepsilon - \frac{1}{\varepsilon\lambda_1 T}.$$

And for T large enough, we have:

$$I\!\!E\left(\frac{1}{T}\sum_{t=1}^{T} g_t\right) \geq v(k_1) - 8\varepsilon,$$

independently of the strategy τ of player 2. This shows that player 1 uniformly guarantees $v(k_1)$ in the stochastic game with initial state k_1. By symmetry, player 2 can do as well and theorem 1.10 is proved.

We finally come back to the proof of the key proposition.

Proof of proposition 1.16: Fix $t \geq 1$, and define $C_1 = \varphi(\lambda_t) - \varphi(\lambda_{t+1})$, $C_2 = v_{\lambda_{t+1}}(k_{t+1}) - v_{\lambda_t}(k_{t+1})$ and $C_3 = \lambda_t(g_t - v_{\lambda_t}(k_{t+1}))$. A simple computation shows that:

$$Z_{t+1} - Z_t - (C_1 + C_2 - C_3) = \lambda_t g_t + (1 - \lambda_t)v_{\lambda_t}(k_{t+1}) - v_{\lambda_t}(k_t).$$

Denote by $\mathcal{H}_t$ the σ-algebra generated by histories in $(K \times I \times J)^{t-1} \times K$ (before players play at stage t), by definition of σ one has:

$$I\!\!E(\lambda_t g_t + (1 - \lambda_t)v_{\lambda_t}(k_{t+1})|\mathcal{H}_t) \geq v_{\lambda_t}(k_t).$$

Consequently, we obtain:

$$I\!\!E(Z_{t+1} - Z_t|\mathcal{H}_t) \geq I\!\!E(C_1 + C_2 - C_3|\mathcal{H}_t). \tag{4}$$

We have $|C_2| \leq \varepsilon\lambda_t$ by point c) of lemma 1.14. By definition of d_{t+1}, we have $d_{t+1} - d_t \geq g_t - v_{\lambda_t}(k_{t+1}) + 4\varepsilon$, hence $C_3 \leq \lambda_t(d_{t+1} - d_t) - 4\varepsilon\lambda_t$. We now prove

$$C_1 \geq \lambda_t(d_{t+1} - d_t) - \varepsilon\lambda_t \tag{5}$$

If $\lambda_{t+1} < \lambda_t$, then $d_{t+1} > d_t$ and $C_1 = \varphi(\lambda_t) - \varphi(\lambda_{t+1}) \geq \lambda_t(d_{t+1} - d_t) - (\lambda_t - \lambda_{t+1})(d_{t+1} - d_t)$ $\geq \lambda_t(d_{t+1} - d_t) - \varepsilon\lambda_t$ by a) and b) of lemma 1.14. If $\lambda_{t+1} > \lambda_t$, then $d_{t+1} < d_t$ and $\varphi(\lambda_{t+1}) - \varphi(\lambda_t) \leq \lambda_{t+1}(d_t - d_{t+1}) = \lambda_t(d_t - d_{t+1}) + (\lambda_{t+1} - \lambda_t)(d_t - d_{t+1}) \leq \lambda_t(d_t - d_{t+1}) + \varepsilon\lambda_t$. And (5) is proved.

Back to inequality (4), we obtain:

$$I\!\!E(Z_{t+1} - Z_t|\mathcal{H}_t) \geq I\!\!E\left(\lambda_t(d_{t+1} - d_t) - \varepsilon\lambda_t - \varepsilon\lambda_t - \lambda_t(d_{t+1} - d_t) + 4\varepsilon\lambda_t|\mathcal{H}_t\right) = 2\varepsilon I\!\!E\left(\lambda_t|\mathcal{H}_t\right)$$

which proves that $(Z_t)_t$ is a sub-martingale and for all $t \geq 0$, $I\!\!E(Z_{t+1}) \geq 2\varepsilon I\!\!E(\sum_{s=1}^{t} \lambda_s) + Z_1$. This ends the proof of proposition 1.16.

Remark: Mertens-Neyman 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 λ 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.

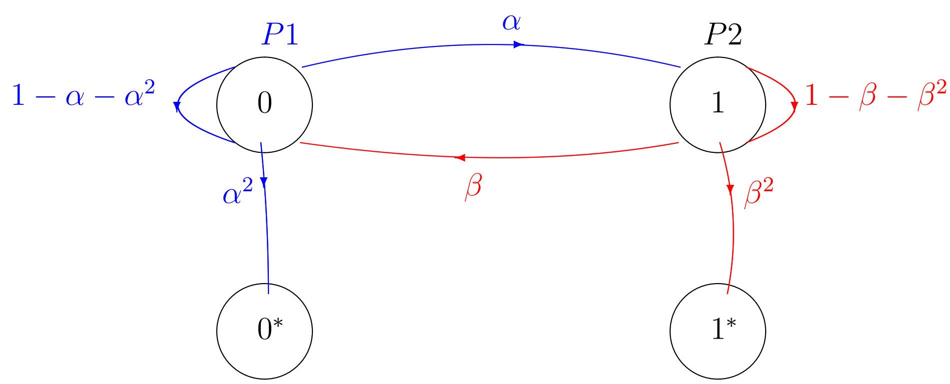
2 A few extensions and recent results

We want to go beyond the "simple" case of finitely many states and actions. Before presenting positive results, we start with recent counterexamples.

2.1 Counterexamples

2.1.1 A simple stochastic game with compact action sets and no limit value

It was long believed that stochastic games with compact state space, continuous transitions and payoff functions have a limit value. The first counter-example is due to G. Vigeral (2013), who was also studying with S. Gaubert and J. Bolte the case of semi-algebraic transitions. The elementary example below is a very slight variation on a example by B. Ziliotto (2013), mentioned in Sorin Vigeral (2015). (A variant, where each player controls his own state variable, is in Laraki Renault 2015).



There are 4 states: $K = \{k_0, k_1, 0^*, 1^*\}$. States 0^* and 1^* are absorbing, and the payoff in state k_0, resp. k_1, is 0, resp. 1. In state k_0, Player 1 chooses α in some fixed set $I \subset [0, 1/2]$, and the next state is k_1 with probability α, 0^* with probability α^2 and k_0 with the remaining probability $1-\alpha-\alpha^2$. Similarly, in state k_1 player 2 chooses β in J, and the next state is k_0 with probability β, 1^* with probability β^2 and k_1 with the remaining probability. To obtain divergence of the values, we introduce a dissymmetry between players and assume that:

$$I = \{\frac{1}{2^{2n}}, n \geq 1\} \cup \{0\}, \text{ and } J = [0, 1/2].$$

During the lecture we will prove:

Theorem 2.1.
$\liminf_{\lambda\to 0} v_\lambda(k_0) = \liminf_{\lambda\to 0} v_\lambda(k_1) = 4/9,$ *and* $\limsup_{\lambda\to 0} v_\lambda(k_0) = \limsup_{\lambda\to 0} v_\lambda(k_1) = 1/2.$

2.1.2 A hidden stochastic game with no limit value

Hidden stochastic games are a generalization of the basic model by assuming that at the beginning of each stage, the players observe past actions and a public signal (but no longer the current state). They are also called Stochastic Games with Public Information.

Hence a hidden stochastic game is 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 $g : K \times I \times J \longrightarrow I\!R$, and a transition $q : K \times I \times J \longrightarrow \Delta(K \times S)$. Here, K, I, J and S are assumed non empty and finite.

Bruno Ziliotto (2013) constructed a hidden stochastic game with no limit value (where $\liminf v_\lambda = 1/2, \limsup v_\lambda \geq 5/9$). This disproved 2 important conjectures by J-F. Mertens: 1) the existence of the limit value in any general repeated game with finitely many states, actions and signals, and 2) the equality between the largest quantity guaranteed by player 1 and the limit value for games where player 1 always has more information than player 2.

One can even slightly improve on B. Ziliotto's construction and we will show (Renault Ziliotto 2015):

Theorem 2.2. *For each $\varepsilon > 0$, there exists a zero-sum hidden stochastic game with P1's payoffs in $[0,1]$, 6 states, 2 actions for each player and 6 signals such that:*

$$\liminf_{\lambda \to 0} v_\lambda \leq \varepsilon \text{ and } \limsup_{\lambda \to 0} v_\lambda \geq 1 - \varepsilon.$$

2.2 1-Player games

While looking for positive results, it is interesting to start with the one-player case, where the existence of the limit and uniform values is fairly understood.

2.2.1 General results: the long-term value

We consider a general dynamic programming problem with *bounded payoffs*: $\Gamma(z_0) = (Z, F, r, z_0)$ given by 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, and for each state z in Z, $F(z)$ is a non empty subset of Z. (An equivalent MDP variant of the model exists with an explicit set of actions A, and transitions given by a function from $Z \times A$ to Z.)

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...

The set of admissible plays at z_0 is defined as: $S(z_0) = \{s = (z_1, ..., z_t, ...) \in Z^\infty, \forall t \geq 1, z_t \in F(z_{t-1})\}$.

For $n \geq 1$, the value of the n-stage problem with initial state z is defined as:

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

For $\lambda \in (0,1]$, the value of the λ-discounted problem with initial state z is defined as:

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

More generally, define an evaluation $\theta = (\theta_t)_{t \geq 1}$ as a probability on positive integers. The θ-payoff of a play $s = (z_t)_{t \geq 1}$ is $\gamma_\theta(s) = \sum_{t=1}^{\infty} \theta_t r(z_t)$, and the θ-value of $\Gamma(z)$ is

$$v_\theta(z) = \sup_{s \in S(z)} \gamma_\theta(s).$$

The set of all evaluations is denoted by Θ. The total variation of an evaluation θ is defined as: $TV(\theta) = \sum_{t=1}^{\infty} |\theta_{t+1} - \theta_t|$. Given an evaluation $\theta = \sum_{t \geq 1} \theta_t \delta_t$ (here δ_t is the Dirac measure on stage t) and some non negative integer m, we 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 the convergence of $(v_n)_n$, when $n \to \infty$, of $(v_\lambda)_\lambda$, when $\lambda \to 0$, or more generally of $(v_{\theta^k})_k$, 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.*

We now only consider uniform convergence of the value functions. Denote by $\mathcal{V}$ the set of functions from Z to $[0,1]$, endowed with the supremum metric $d_\infty(v, v') = \sup_{z \in Z} |v(z) - v(z')|$. Saying that a sequence $(v^k)_{k \geq 1}$ of functions from Z to $[0,1]$ uniformly converges is the same as saying that the sequence (v^k) converges in the metric space $\mathcal{V}$. Notice that in a metric space, convergence of a sequence (v^k) happens if and only if:

1) the sequence $(v^k)_k$ has at most one limit[6], and

2) the set $\{v^k, k \geq 1\}$ is totally bounded[7].

The above equivalence holds for any sequence in a metric space. But here we consider the special case of value functions of a dynamic programming problem, with long term limits. It will turn out that 1) is automatically satisfied.

Definition 2.3. *Define for all z in Z,*

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

[6]A limit point of $(v^k)_k$ being defined as a limit of a converging subsequence of $(v^k)_k$.

[7]For each $\varepsilon > 0$, the set can be covered by finitely many balls of radius ε. Equivalently, the completion of the set is compact. Equivalently, from any sequence in the set one can extract a Cauchy subsequence.

The following results apply in particular to the sequences $(v_n)_n$ and $(v_\lambda)_\lambda$

Theorem 2.4. *(R., 2014)*
Consider a sequence of evaluations $(\theta^k)_k$ such that $TV(\theta^k) \to_{k\to\infty} 0$.
Any limit point of $(v_{\theta^k})_k$ is v^*.

Corollary 2.5. *Consider a sequence of evaluations $(\theta^k)_k$ such that $TV(\theta^k) \to_{k\to\infty} 0$.*
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) Assume that Z is endowed with a distance d such that: a) (Z,d) is a totally bounded metric space, and b) 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) Assume that Z is endowed with a distance d such that: a) (Z,d) is a precompact metric space, b) r is uniformly continuous, and c) F is non expansive, i.e. $\forall z \in Z, \forall z' \in Z, \forall z_1 \in F(z), \exists z_1' \in F(z')$ s.t. $d(z_1, z_1') \leq d(z,z')$. Same conclusions as corollary 3).

The above results can be extended to the case of stochastic dynamic programming, (i.e. when $F(z)$ is a set of probability distributions on Z for each z). In this case it is often convenient to define the value functions v_n, v_λ, v_θ directly by their Bellman equations.

Notice that life is much simpler in the particular case where the problem is leavable, i.e. when $z \in F(z)$ for each z. Then without any assumption, $(v_n)_n$ is non decreasing and *pointwise* converge to v^*, where: $v^* = \inf\{v : Z \to [0,1], \textit{excessive}^{8}, v \geq r\}$.

Remark: in the basic model of stochastic games, one can similarly define the θ-value of any evaluation θ. The existence of the uniform value (Mertens-Neyman 1981) implies that v_{θ^k} converges to the same limit as (v_n) and (v_λ) as soon as : for each k, θ^k is non increasing, and θ_1^k goes to 0 when $k \to \infty$. Assuming only that $TV(\theta^k) \to_{k\to\infty} 0$ is not enough to obtain such convergence (Ziliotto 2015).

2.2.2 The uniform convergence of $(v_n)_n$ and $(v_\lambda)_\lambda$ are equivalent.

The results of the previous subsection show in particular that if (v_n) and (v_λ) uniformly converge, they have the same limit. For these two particular sequences of evaluations, we have a stronger result.

Theorem 2.6. *(Lehrer-Sorin 1992) In a 1-player game, (v_n) converges uniformly if and only if (v_λ) converges uniformly. In case of convergence, the limit is the same.*

[8] v excessive means that $v(z) \geq v(z')$ if $z' \in F(z)$, i.e. that v is non increasing on any trajectory.

2.2.3 The compact non expansive case and the uniform value

We have stronger results if the state space is assumed to be compact, payoffs are continuous and transitions are non expansive. We consider here a stochastic dynamic programing problem (also called 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 $[0,1]$.

Here $\Delta_f(X)$ is the set of probabilities with finite support over X. We assume that transitions have finite support for simplicity, however many results concerning the limit value and its characterization can go through without this assumption. When we will study the uniform value, this assumption will be useful to define strategies avoiding measurability issues.

Here a player chooses u_1 in $F(x_0)$, then x_1 is selected according to u_1 and yields the payoff $r(x_1)$, then the player chooses u_2 in $F(x_1)$, etc... We define as usual the n- stage value function: $v_n(x_0) = \sup_{\sigma \in S(x_0)} I\!E_\sigma \left(\frac{1}{n}\sum_{t=1}^n r(x_t)\right)$, where $S(x_0) = \{\sigma = (u_1, ..., u_t, ...) \in Z^\infty, u_1 \in F(x_0), \forall t \geq 1, u_{t+1} \in F(u_t)\}$. We define similarly the λ-discounted value $v_\lambda(z_0)$, and more generally for any evaluation θ we have the θ-value $v_\theta(z_0)$.

We assume here that X is a compact metric space with metric denoted by d. The set $\Delta(X)$ of Borel probability measures over X is also a compact metric space (for the weak-* topology), and we will use the Kantorovich-Rubinstein metric[9]: for u and u' 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 assimilate an element x in X with the corresponding Dirac measure in $\Delta(X)$. The Graph of Γ can be viewed as a subset of $\Delta(X) \times \Delta(X)$, and we denote by $\overline{\text{conv}}\text{Graph}(\Gamma)$ its closed convex hull in $\Delta(X) \times \Delta(X)$. We define the set of *invariant measures* as:

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

We will assume that r is continuous, and extend r to a continuous affine function defined on $\Delta(X)$: for u in $\Delta(X)$, $r(u)$ is the expectation of r with respect to u. We will also assume non expansive transitions.

$$\forall x \in X, \forall x' \in X, \forall u \in \Gamma(x), \exists u' \in \Gamma(x'),\ s.t.\ d_{KR}(u,u') \leq d(x,x').$$

[9] In the second expression, $\Pi(u,u')$ denotes the set of probabilities on $X \times X$ with first marginal u and second marginal u'.

This assumption is always satisfied when X is finite[10], or when X is a simplex and $\Gamma(x)$ is the set of splittings at x, i.e. the set of Borel probabilities on X with mean x.

One can apply here a variant of property 4) of corollary 2.5 to prove uniform convergence of (v_n) and (v_λ), but we can obtain a stronger result with a better characterization of the limit value and the existence of the uniform value.

Theorem 2.7. *(R-Venel 2013) Assume the state space is compact, payoffs are continuous and transitions are non expansive. Then (v_n) and (v_λ) uniformly converge to v^*, where for each initial state x,*

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

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

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

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).

The theorem also extends to general sequences of evaluations with vanishing total variation.

For partially observable Markov decision processes (POMDP) with finite set of states, actions and signals, the existence of the uniform value was first proved by Rosenberg, Solan and Vieille (2002). The present theorem can not be applied as is in this case, because transitions are not non expansive with respect to the KR-metric. However, an alternative metric introduced in (Renault Venel 2013) can be used to apply the theorem to this class of games.

Recently, Venel and Ziliotto (2015) proved for these models the existence of the uniform value in pure strategies, i.e. without the assumption that F has convex values.

2.3 The CV of $(v_n)_n$ and $(v_\lambda)_\lambda$ are equivalent.

The equivalence between the uniform convergence of $(v_n)_n$ and $(v_\lambda)_\lambda$, which holds in general in 1-player games, has been recently proved (Ziliotto 2015) to extend to a large class of stochastic games.

It applies in particular to the following setup. Assume the set of states K and the set of actions I and J are compact metric spaces, that the transition $q : K \times I \times J \longrightarrow \Delta(K)$ and the payoff $g : K \times I \times J \longrightarrow I\!\!R$ are jointly continuous. Together with an initial state k , (K, I, J, q, g) define a stochastic game. Then one can show that for each n and each λ the value of the n-stage game $v_n(k)$ and $v_\lambda(k)$ exist and satisfy the Shapley equations: $\forall n \geq 0, \forall \lambda \in (0, 1]$, $\forall k \in K$,

[10] if $d(x, x') = 2$ for all x, x', then $d_{KR}(z, z') = \|z - z'\|_1$ for all z, z' in $\Delta(X)$.

$$\begin{aligned}(n+1)\, v_{n+1}(k) &= \sup_{x\in\Delta(I)} \inf_{y\in\Delta(J)} \left(g(k,x,y)) + n\, I\!\!E_{q(k,x,y)}(v_n) \right), \\ &= \inf_{y\in\Delta(J)} \sup_{x\in\Delta(I)} \left(g(k,x,y)) + n\, I\!\!E_{q(k,x,y)}(v_n) \right).\end{aligned}$$

$$\begin{aligned}v_\lambda(k) &= \sup_{x\in\Delta(I)} \inf_{y\in\Delta(J)} \left(\lambda\, g(k,x,y) + (1-\lambda)\, I\!\!E_{q(k,x,y)}(v_\lambda) \right), \\ &= \inf_{y\in\Delta(J)} \sup_{x\in\Delta(I)} \left(\lambda\, g(k,x,y) + (1-\lambda)\, I\!\!E_{q(k,x,y)}(v_\lambda) \right),\end{aligned}$$

Theorem 2.8. *(Ziliotto, 2015) In a compact continuous stochastic game, (v_n) converges uniformly if and only if (v_λ) converges uniformly. In case of convergence, the limit is the same.*

B. 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 $I\!\!R$, its image $\Psi(f)$ by the Shapley operator, defined by:

$$\forall k \in K, \Psi(f)(k) = \sup_{x\in\Delta(I)} \inf_{y\in\Delta(J)} \left(\lambda\, g(k,x,y) + (1-\lambda)\, I\!\!E_{q(k,x,y)}(f) \right),$$

is also Borel measurable.

2.4 Repeated Games with incomplete information

2.4.1 Lack of information on one side: the cav u theorem

Repeated games with lack of information on one side were introduced by Aumann and Maschler in the 1960's. In the basic model presented here, players repeat at every stage the same matrix game, which is only partially known to player 2.

Formally, we have a finite family $(G^k)_{k\in K}$ of payoff matrices in $I\!\!R^{I\times J}$, and an initial belief $p \in \Delta(K)$ for player 2. All these quantity are known by the players. The game is played as follows: 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, and at the end of every stage the actions played are publicly observed.

As usual, we define the value $v_n(p)$ of the n-stage game with average payoffs $I\!\!E(\frac{1}{n}\sum_{t=1}^n G^k(i_t, j_t))$.

In the infinitely repeated game with initial belief p, for each N player 2 can play an optimal strategy of the N-stage game with belief p, independently on consecutive blocks of N stages. This easily implies that $(v_n)_n$ CV and that P2 can guarantee $\lim_n v_n$ in the infinitely repeated game. 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 *Completely revealing* or *Non revealing* guarantees 0.

Proposition 2.9. *Recursive formula: for all $n \geq 1$ and p in $\Delta(K)$,*

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

where $x = (x^k(i))_{i \in I, k \in K}$, with x^k the mixed action used at stage 1 by player 1 if the state is k, $g(p, x) = \min_j \sum_{k,i,j} p^k G^k(x^k(i), j)$ is the expected payoff of stage 1 if player 2 plays a best reply against x, $x(p)(i) = \sum_{k \in K} p^k x^k(i)$ is the probability that action i is played at stage 1, and $\hat{p}(x, i)$ is the conditional probability on K given p, x, i.

On can show that the problem of player 1 is similar to facing a leavable stochastic dynamic programming problem given $F : X \rightrightarrows \Delta_f(X)$, where X is the simplex $\Delta(K)$ and $F(x) = \{\mu \in \Delta_f(X), mean(\mu) = x\}$.

Define for each p in $\Delta(k)$ the value of the "non revealing game at p" as the value of the average matrix game $\sum_k p^k G^k$:

$$u(p) = \mathrm{Val}(\sum_k p^k G^k).$$

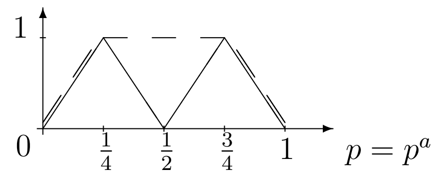
The following result is the basis of the theory of repeated games with incomplete information:

Theorem 2.10. *(Aumann Maschler 1966):*

(v_n) uniformly converges to

$$\mathrm{cav} u = \inf\{v : \Delta(K) \to I\!\!R, v \text{ concave } v \geq u\},$$

and the repeated game with incomplete information has a uniform value.



Example: u and cav$u$.

2.4.2 The cav$u$ theorem with non observable actions

The previous model can be extended to the case where at the end of each stage, each player receives a private signal depending on the selected state and the actions played. In a 2 player repeated game with lack of information on one side and signals, we still have a finite set of states K, payoff matrices $(G^k)_k$, finite action sets I and J and now finite signal sets C and D together with a signaling function $l : K \times I \times J \longrightarrow \Delta(C \times D)$. If the state is k, at the end of a stage where i and j have been played, a couple (c, d) is selected according to $l(k, i, j)$, player 1 learns c whereas player 2 learns d.

Again, it is not difficult to show that $\lim_n v_n$ exists and can be guaranteed by player 2. The problem of player 1 is now equivalent to a non leavable stochastic dynamic problem, and the signalling function will only play a role through its second marginal l_2 on the set D of signals of player 2. Define the set of non revealing strategies of player 1 at p as:

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

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

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

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

Theorem 2.11. *(Aumann Maschler 1967): The repeated game with initial probability p has a uniform value given by* cav$u(p)$.

2.4.3 The value of repeated games with an informed controller

We now consider the general model of zero-sum dynamic game with finitely many states, actions and signals (Mertens Sorin Zamir 1994 Core DP, 2015 Cambridge U. Press). A Markov Dynamic Game (MDG) is given by:

• 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 $\Delta(K \times C \times D)$.

The progress of the game is as follows:

At stage 1: (k_1, c_1, d_1) is selected according to π, 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 stage 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 $g(k_t, i_t, j_t)$.

As usual, a pair of behavioral strategies (σ, τ) induces a probability over plays. What about the existence of $\lim_n v_n$ and $\lim_\lambda v_\lambda$? 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 of player 2 on the initial state k_1. This belief, deduced from π and the initial signal of player 2, is denoted by p. We write $X = \Delta(K)$ the set of possible such beliefs.

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 2.12. *(R. 2012, R-Venel 2013): Under HX and HY, the repeated game has a uniform value. And in the game where the initial belief of player 2 is p, the limit value is:*

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

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

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

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

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

Remarks:

• extends to the case of evaluations with vanishing total variation.

• the existence of the uniform value has been extended to the case where Player 1 controls the transitions and has more information on the state than Player 2 (Gensbittel, Oliu-Barton, Venel 2014).

2.4.4 Lack of information on both sides

We now consider the more symmetric model where both players have partial information on the matrix game to be repeated. K (resp. L) is the finite set of private states for P1 (resp. $P2$), there is a family $(G^{k,l})_{(k,l)\in K\times L}$ of payoff matrices in $I\!R^{I\times J}$ and initial probabilities $p \in \Delta(K)$ and q in $\Delta(L)$. This defines 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, and the actions played are publicly observed at the end of each stage.

Definition 2.13. *The non revealing value function u is defined by:*

$$\forall p \in \Delta(K), \forall q \in \Delta(L),\ u(p,q) = \mathrm{Val}_{\Delta(I)\times\Delta(J)}(\sum_{k,l} p^k q^l G^{k,l}).$$

Given a continuous function v on $\Delta(K)\times\Delta(L)$, we denote by $\mathrm{cav_I}\, v$ the concavification of v with respect to the first variable, the second variable being fixed: for each q, $\mathrm{cav_I}\, v(.,q) = \mathrm{cav}(.,q)$. Similarly $\mathrm{vex_{II}}\, v$ denotes the convexification of v with respect to the second variable.

Theorem 2.14.
(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.

(Mertens-Zamir 1971): (v_n) and (v_λ) uniformly converge to the unique continuous function v on $\Delta(K)\times\Delta(L)$ such that:

$$\begin{cases} v &= \mathrm{vex_{II}}\,\max\{u,v\} \\ v &= \mathrm{cav_I}\,\min\{u,v\} \end{cases}$$

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.

2.5 Some open problems

2.5.1 Computing the value.

a) In the basic model.
b) In repeated game with incomplete info on one side, where the state follows an exogeneous Markov chain observed by player 1 only (R. 2006). $K = \{a, b\}$, $p = (1/2, 1/2)$, $M = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}$, $G^a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $G^b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

One can show that the uniform value exists in such model (it is a particular case of theorem 2.12).
If $\alpha = 1$, the value is $1/4$ (Aumann Maschler cav$u$ theorem).
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): $\frac{1}{v} = u_0 + u_0 u_1 + u_0 u_1 u_2 + ...$, 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 $\alpha = 0.9$?

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

2.5.3 Find nice conditions for compact non expansive stochastic games to have a limit value.

(such as semi-algebraicity in Bolte Gaubert Vigeral 2015, acyclicity in Laraki Renault...)

2.5.4 Finite MDG: Find other value functions which will always converge. Continuous-time games (à la Neyman) ?

2.5.5 How large is the set of information structures ?

K is a fixed finite set of parameters.

An information structure is defined as an element u in $Z := \Delta_f(K \times \mathbb{N} \times \mathbb{N})$. 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 $g : K \times \mathbb{N} \times \mathbb{N} \longrightarrow [-1, 1]$
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 $\mathtt{val}(u, g)$ 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 $\mathbb{N}$, player 2 chooses j in $\mathbb{N}$, and player 1's payoff is $g(k, i, j)$.

Define (Gensbittel R. work in progress):

$$d^*(u, v) = \sup_g |\mathtt{val}(u, g) - \mathtt{val}(v, g)|.$$

Let Z^* be the quotient space of Z. (Z^*, d^*) is a metric space, is it totally bounded ?

2.5.6 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).

This existence question is 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 i receives a payoff $u^i(S)$, depending on the set S 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:

To conclude, let me stress again that many important and interesting works do not appear at all in these short notes. Here are a few examples, without even mentioning differential games:

- 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 ε-optimal stationary strategies (Thuijsman Vrieze 1991, 1992, Flesch Thuijsman Vrieze 1998...)
- continuous-time stochastic games (Neyman 2012), continuous-time approachs (Cardaliaguet *et al.* 2012...)
- continuous-time limits where the duration of a stage goes to 0 (Neyman 2013, Cardaliaguet *et al.* 2015...)
- 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 (Rosenberg Vieille 2002)...
- ......

Acknowledgement:
Thanks to Bruno Ziliotto for comments and proofreading of these notes.

References

[1] Aumann R.J. and M. Maschler. (1995, reedition from 1966-67-68) Repeated games with incomplete information, with the collaboration of R. Stearns. M.I.T. Press.

[2] Blackwell D. (1962). Discrete dynamic programming, *Annals of Mathematical Statistics*, 33, 719-726.

[3] Blackwell D. and T. Ferguson (1968). The Big Match, *Annals of Mathematical Statistics*, 33, 882-886.

[4] Benedetti R. and J. Risler (1990). *Real algebraic and semi-algebraic sets*, Hermann.

[5] Bewley T. and Kohlberg E. (1976a). The asymptotic theory of stochastic games, *Mathematics of Operation Research,* 1, 197-208.

[6] Bewley T. and Kohlberg E. (1976b). The asymptotic solution of a recursion equation occuring in stochastic games, *Mathematics of Operation Research,* 1, 321-336.

[7] Bressaud X. and Quas, A. (2013). Dynamical analysis of a repeated game with incomplete information, submitted paper.

[8] Cardaliaguet P., Laraki R and S. Sorin.(2012), a continuous time approach for the asymptotic value in two-person zero-sum games, *SIAM Journal of Control and Optimization,* 50, 1573-1596.

[9] Cardaliaguet P., Rainer C., Rosenberg D. and N. Vieille (2015), Markov games with frequent actions and incomplete information – the limit case, to appear in *Mathematics of Operation Research*.

[10] Coulomb J. M. (2003). Stochastic games without perfect monitoring, *International Journal of Game Theory,* 32, 723-96.

[11] Bolte J., Gaubert S. and G. Vigeral (2015). Definable zero-sum stochastic games, *Mathematics of Operation Research,* 40, 171-191.

[12] Flesch J., Thuijsman F. and O. Vrieze(1998) . Simplifying Optimal Strategies in Stochastic Games, *SIAM Journal of Control and Optimization,* 36, 1331-1347.

[13] Gensbittel F., Oliu-Barton M. and X. Venel (2014). Existence of the uniform value in repeated games with a more informed controller, *Journal of Dynamics and Games* , 1, 411-445.

[14] Gensbittel F. and J. Renault J. (2015) The value of Markov chain games with lack of information on both sides, *Mathematics of Operations Research*, 40, 820-841.

[15] Gimbert H., Renault J., Sorin S., Venel. X and W. Zielonka (2014): On the values of repeated games with signals, to appear in *Annals of Applied Probability.*

[16] Hörner J., D. Rosenberg, E. Solan and N. Vieille (2010) *On Markov Games with Incomplete Information on One Side*, *Operations Research*, 58, 1107-115.

[17] Laraki R. (2010). Explicit formulas for repeated games with absorbing states, *International Journal of Game Theory* , Special Issue in Honor of Michael Maschler, 39, 53-70.

[18] Laraki R. and J.Renault (2015). Acyclic Gambling Games.

[19] Lehrer E. and S. Sorin.(1992) A Uniform tauberian theorem in dynamic programming, *Mathematics of Operations Research*, 17, 303-307.

[20] Maitra A. and T. Parthasarathy (1970). On stochastic games, *Journal of Optimization Theory and Applications,* 5, 289-300.

[21] Martin D.A. (1975). Borel determinacy, *Annals of Mathematics*, 102, 363-371.

[22] Martin D.A. (1998). The determinacy of Blackwell games, *Journal of Symbolic Logic*, f63(4), 1565-1581.

[23] Mertens J.-F. Repeated games. *Proceedings of the International Congress of Mathematicians, Berkeley 1986*, 1528–1577. *American Mathematical Society*, 1987.

[24] Mertens J.-F. and Neyman A. (1981). Stochastic games, *International Journal of Game Theory,* 10, 53-66.

[25] Mertens J.-F. and S. Zamir. The value of two-person zero-sum repeated games with lack of information on both sides. *International Journal of Game Theory*, 1, 39–64, 1971.

[26] Mertens J.-F. and S. Zamir. A duality theorem on a pair of simultaneous functional equations. *Journal of Mathematical Analysis and Applications*, 60, 550–558, 1977.

[27] Neyman A. (2012). Continuous-time Stochastic games, working paper.

[28] Neyman A. (2013), Stochastic Games with Short-Stage Duration. *Dynamic Games and Applications*, 236-278.

[29] Nowak A.S. (2003). Zero-sum stochastic games with Borel state spaces. *Stochastic games and applications*, NATO Science Series, Neyman and S. Sorin eds, 77-91

[30] Nowak A.S. (1985). Universally measurable strategies in zero-sum stochastic games, *Annals of Probability,* 13, 269-287.

[31] Oliu-Barton, M. (2014). The asymptotic value in finite stochastic games, *Mathematics of Operation Research* , 39, 712-721.

[32] Oliu-Barton, M. (2014). The Splitting Game: Value and Optimal Strategies, submitted.

[33] Renault J. (2006) The value of Markov chain games with lack of information on one side, *Mathematics of Operations Research*, 31, 490-512

[34] Renault J. (2011) Uniform value in Dynamic Programming, *Journal of the European Mathematical Society*, 13, 309–330.

[35] Renault J. (2012) The value of repeated games with an informed controller, *Mathematics of Operations Research*, 37, 154-179.

[36] Renault J. (2014) General limit value in Dynamic Programming, *Journal of Dynamics and Games*, 1, 471-484.

[37] Renault J. and B. Ziliotto (2015). Hidden Stochastic Games and Limit Equilibrium Payoffs, arXiv:1407.3028

[38] Rosenberg D., E. Solan and N. Vieille. (2002). Blackwell Optimality in Markov Decision Processes with Partial Observation, *Annals of Statistics*, 30, 1178-1193.

[39] Rosenberg D., E. Solan and N. Vieille. (2002). The MaxMin value of stochastic games with imperfect monitoring, *International Journal of Game Theory,* 32, 133-150.

[40] Rosenberg D. and N. Vieille. (2002). Stochastic games with a single controller and incomplete information, *SIAM Journal of Control and Optimization*43, 86-110.

[41] Shapley L. S. (1953) Stochastic games, *Proceedings of the National Academy of Sciences of the U. S. A.,* 39, 1095-1100.

[42] Shmaya E. (2011).The determinacy of infinite games with eventual perfect monitoring. *Proceedings of the American Mathematical Society*. 139: 3665-3678.

[43] Solan E. (1998). Discounted stochastic games, *Mathematics of Operation Research,* 23, 1010-1021.

[44] Solan E. (1999). Three-Player absorbing games, *Mathematics of Operation Research,* 24, 669-698.

[45] Solan E. and N. Vieille (2001). Quitting games, *Mathematics of Operation Research,* 26, 265-285.

[46] Solan E. and N. Vieille (2002). Uniform value in recursive games *The Annals of Applied Probability,* 12, 1185-1201.

[47] Solan E. and N. Vieille (2003). Quitting games – an example, *International Journal of Game Theory,* 31, 365-381.

[48] S. Sorin. (1984) Big match with lack of information on one side (Part I). *International Journal of Game Theory*, 13, 201–255, 1984.

[49] S. Sorin. (1985) Big match with lack of information on one side (Part II). *International Journal of Game Theory*, 14, 173-204

[50] Sorin S. (1986). Asymptotic properties of a non zero-sum stochastic game, *International Journal of Game Theory,* 15, 101-107.

[51] Sorin S. and G. Vigeral (2015). Reversibility and Oscillations in zero-sum discounted stochastic games, *Journal of Dynamics and Games,* 2, 103-115.

[52] Takahashi M. (1964). Equilibrium points of stochastic non-cooperative n-person game, *J. Sci. Hiroshima Univ.* 28, 95-99.

[53] F. Thuijsman and O. J. Vrieze [1991], Easy initial states in stochastic games, in Stochastic Games and Related Topics, T. E. S. Raghavan, T. S. Ferguson, O. J. Vrieze, and T. Parthasarathy, eds., Kluwer, Dordrecht, the Netherlands, 85–100.

[54] Thuijsman F. and Vrieze O (1992): Note on recursive games , in *Game Theory and Economic Applications*, Lecture Notes in Economics and Mathematical Systems 389, Dutta, Mookherjee, Parthasarathy, Ragahavn and Tijs (eds.), Springer-Verlag, 133-145.

[55] Venel X. and B. Ziliotto (2015) Pathwise uniform value in gambling houses and Partially Observable Markov Decision Processes. arXiv:1505.07495

[56] Vigeral, G. (2013). A Zero-Sum Stochastic Game with Compact Action Sets and no Asymptotic Value, *Dynamic Games and Applications,* 3, 172-186.

[57] Vieille N. (2000a). Two-player stochastic games I: a reduction, *Israël Journal of Mathematics,* 119, 55-91.

[58] Vieille N. (2000b). Two-player stochastic games II: the case of recursive games, *Israël Journal of Mathematics,* 119, 93-126.

[59] Ziliotto B. (2013) Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin=limvn. To appear in Annals of Probability.

[60] Ziliotto B. (2015) General limit value in stochastic game. To appear in *International Journal of Game Theory*.

[61] Ziliotto B. (2015) A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games. To appear in *Mathematics of Operations Research*.

Books and reviews:

[62] Laraki R. and S. Sorin (2014). Advances in zero-sum dynamic games, *Handbook of Game Theory IV*, 27-93. ed. by P. Young and S. Zamir.

[63] Mertens J.-F., S. Sorin and S. Zamir (2015). *Repeated games*, Cambridge University Press.

[64] Neyman A. and S. Sorin (2003). *Stochastic games and applications*, NATO Science Series, Kluwer Academic Publishers.

[65] Sorin S. (2002). *A First course on zero-sum repeated games*, Springer-Verlag.

[66] Vieille N. (2002). Stochastic games: Recent results, *Handbook of Game Theory III*, ed. by R. Aumann and S. Hart, 1833–1850.