Online Learning in Games

Rida Laraki and Guillaume Vigeral

IASD Lecture 4, 2024
Games in extensive form

Generalities

- Normal form : each players plays only once and simultaneously.

Generalities

- Normal form : each players plays only once and simultaneously.
- Extensive forms : players play one after another, may play several times.

Generalities

- Normal form : each players plays only once and simultaneously.
- Extensive forms : players play one after another, may play several times.
- Examples : chess, poker, $\cdots$

Generalities

- Normal form : each players plays only once and simultaneously.
- Extensive forms : players play one after another, may play several times.
- Examples : chess, poker, $\cdots$
- Particular case of perfect information : players observe (and remember) everything. Chess is an example, not poker.

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

- a **root** r

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

- a **root** r
- a finite set of **nodes** Z

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

- a **root** r
- a finite set of **nodes** Z
- a fonction **predecessor** ϕ from $Z \setminus \{\theta\}$ to Z, such that for each z there is an integer n with $\phi^n(z) = r$.

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

- a **root** r
- a finite set of **nodes** Z
- a fonction **predecessor** ϕ from $Z \setminus \{\theta\}$ to Z, such that for each z there is an integer n with $\phi^n(z) = r$.
- terminal nodes : $T = Z \setminus Im\phi$ and **positions** $P = Im\phi = Z \setminus T$

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

- a **root** r
- a finite set of **nodes** Z
- a fonction **predecessor** ϕ from $Z \setminus \{\theta\}$ to Z, such that for each z there is an integer n with $\phi^n(z) = r$.
- terminal nodes : $T = Z \setminus Im\phi$ and **positions** $P = Im\phi = Z \setminus T$
- a **partition** $\{Z_i, i \in N\}$ of Z

Description of a game with perfect information

N finite set of players. The game in extensive form Λ is discribed by an finite oriented tree with no cycles, labelled nodes and payoffs for terminal nodes. That is we have :

- a **root** r
- a finite set of **nodes** Z
- a fonction **predecessor** ϕ from $Z \setminus \{\theta\}$ to Z, such that for each z there is an integer n with $\phi^n(z) = r$.
- terminal nodes : $T = Z \setminus Im\phi$ and **positions** $P = Im\phi = Z \setminus T$
- a **partition** $\{Z_i, i \in N\}$ of Z
- a **payoff function** $g : T \to \mathbb{R}^{|N|}$.

Denote by S the successor correspondance defined on P : $z' \in S(z)$ iff $z \in \phi(z')$.

Play of the game

- Let i be the player such that $r \in Z_i$.

Play of the game

- Let i be the player such that $r \in Z_i$.
- Player i chooses a successor z_1 of r.

Play of the game

- Let i be the player such that $r \in Z_i$.

- Player i chooses a successor z_1 of r.

- Inductively, let $j \in N$ such that $z_t \in Z_j$. Player j chooses $z_{t+1} \in S(z_t)$.

Play of the game

- Let i be the player such that $r \in Z_i$.
- Player i chooses a successor z_1 of r.
- Inductively, let $j \in N$ such that $z_t \in Z_j$. Player j chooses $z_{t+1} \in S(z_t)$.
- When $z_t \in T$ the play is over and each player i gets the payoff $g^i(z_t)$.

Strategies

Definition

A (pure) strategy of player i is a function s^i from Q^i to Z such that $s^i(z) \in S(z)$ for all z. Denote S^i the set of strategies of player i.

Strategies

Definition

A (pure) strategy of player i is a function s^i from Q^i to Z such that $s^i(z) \in S(z)$ for all z. Denote S^i the set of strategies of player i.

Normal form

A startegy profile $s = \{s^i, i \in N\}$ naturally induce a play, hence a payoff $g(s)$. The *normal form* Γ of the original extensive form game with perfect information Λ is the normal form game (N, S, g).

Interpretation : before the beggining of the game an arbiter asks each player (independantly) what they would do in each of their controlled node. He can then implement the play and tell each player his payoff.

Definition

A (pure or mixed) Nash equilibria of the game in extensive form is a (pure or mixed) Nash equilibria of its normal form.

Pure equilibria

We admit for now the following result (we will prove something stronger in the next slides)

Theorem

Any game with perfect information has an equilibria in pure strategies

An important corollary is determinacy of games like chess

Theorem (Zermelo, 1912)

Consider a perfect information 2-person zero sum game.

- *If the only possible payoffs are 1 (P1 wins) and -1 (P2 wins) then one player has a pure winning strategy.*

Pure equilibria

We admit for now the following result (we will prove something stronger in the next slides)

Theorem

Any game with perfect information has an equilibria in pure strategies

An important corollary is determinacy of games like chess

Theorem (Zermelo, 1912)

Consider a perfect information 2-person zero sum game.

- *If the only possible payoffs are 1 (P1 wins) and* -1 *(P2 wins) then one player has a pure winning strategy.*
- *If the only possible payoffs are 1 (P1 wins),* -1 *(P2 wins) and 0 (draw) then one player has a pure strategy that garantee at least a draw.*

Subgame perfection

Definition

For each position (non terminal node) $p \in P$, the subgame Λ_p is the game starting from p intead of r. In particular $Lambda_r$ is the original game Λ.
A profile σ is a subgame perfect equilibria if for every p, the profile $\sigma[p]$ induced by σ in the subgame Λ_p is a Nash equilibria of this subgame.

Subgame perfection

Definition

For each position (non terminal node) $p \in P$, the subgame Λ_p is the game starting from p intead of r. In particular $Lambda_r$ is the original game Λ.
A profile σ is a subgame perfect equilibria if for every p, the profile $\sigma[p]$ induced by σ in the subgame Λ_p is a Nash equilibria of this subgame.

It is important to undestrand why there may be equilibria that are not subgame perfect : existence of **not credible threats**.

Subgame perfection

Definition

For each position (non terminal node) $p \in P$, the subgame Λ_p is the game starting from p intead of r. In particular $Lambda_r$ is the original game Λ.
A profile σ is a subgame perfect equilibria if for every p, the profile $\sigma[p]$ induced by σ in the subgame Λ_p is a Nash equilibria of this subgame.

It is important to undestrand why there may be equilibria that are not subgame perfect : existence of **not credible threats**.

Theorem

Every game with perfect information admits a subgame perfect pure equilibrium.

Subgame perfection

Definition

For each position (non terminal node) $p \in P$, the subgame Λ_p is the game starting from p intead of r. In particular $Lambda_r$ is the original game Λ.
A profile σ is a subgame perfect equilibria if for every p, the profile $\sigma[p]$ induced by σ in the subgame Λ_p is a Nash equilibria of this subgame.

It is important to undestrand why there may be equilibria that are not subgame perfect : existence of **not credible threats**.

Theorem

Every game with perfect information admits a subgame perfect pure equilibrium.

Proof : backward induction, see example on blackboard.

Nature

- Motivation : model the randomness in some games with perfect information like backgammon or monopoly.

- Add a player called nature without payoff. For each node controlled by nature the successor is chosen randomly with a predetermined lottery.

- All previous results hold except determinacy (since there is randomness of course you cannot garantee to win with probability 1 in backgammon).

Example

Extensive, normal form, pure Nash and subgame perfect equilibria of this game.

- A bank (=player 1) can invest with a heavy or low risk. In the first case there is an important gain with proba $1/2$, a moderate loss with proba $1/3$, and default with proba $1/6$. In the second case there is a moderate gain with proba 1.

Example

Extensive, normal form, pure Nash and subgame perfect equilibria of this game.

- A bank (=player 1) can invest with a heavy or low risk. In the first case there is an important gain with proba $1/2$, a moderate loss with proba $1/3$, and default with proba $1/6$. In the second case there is a moderate gain with proba 1.

- In case of default the state (=player 2) decides to save the bank or not.

Example

Extensive, normal form, pure Nash and subgame perfect equilibria of this game.

- A bank (=player 1) can invest with a heavy or low risk. In the first case there is an important gain with proba $1/2$, a moderate loss with proba $1/3$, and default with proba $1/6$. In the second case there is a moderate gain with proba 1.

- In case of default the state (=player 2) decides to save the bank or not.

- Payoffs are : $(3, 2)$ for moderate gain, $(10, 4)$ for an important gain, $(-1, -1)$ for a moderate loss, $(-4, -10)$ for a default that is saved, et $(-40, -16)$ for a default that is not.

What do you think the state should do in practice ? Why ? Does it depend on additional assumptions ?

Model

- Generalization of the previous games to frameworks when players do not observe everything.

Model

- Generalization of the previous games to frameworks when players do not observe everything.
- They may not know what other players do (e.g. in voting afternoon voters do not know what happened in the morning)

Model

- Generalization of the previous games to frameworks when players do not observe everything.
- They may not know what other players do (e.g. in voting afternoon voters do not know what happened in the morning)
- They may not know what other players know : private information (e.g. poker).

Model

- Generalization of the previous games to frameworks when players do not observe everything.
- They may not know what other players do (e.g. in voting afternoon voters do not know what happened in the morning)
- They may not know what other players know : private information (e.g. poker).
- We model it by uncertainty of the players on the current node on the tree.

Model

- Generalization of the previous games to frameworks when players do not observe everything.
- They may not know what other players do (e.g. in voting afternoon voters do not know what happened in the morning)
- They may not know what other players know : private information (e.g. poker).
- We model it by uncertainty of the players on the current node on the tree.
- In addition to previous structure, each set of nodes P^i of positions controlled by Player i is partitionned in a number of information sets P^i_k.

Model

- Generalization of the previous games to frameworks when players do not observe everything.
- They may not know what other players do (e.g. in voting afternoon voters do not know what happened in the morning)
- They may not know what other players know : private information (e.g. poker).
- We model it by uncertainty of the players on the current node on the tree.
- In addition to previous structure, each set of nodes P^i of positions controlled by Player i is partitionned in a number of information sets P^i_k.
- For each P^i_k there is an action set A^i_k and a function f^i_k.

Model

- Generalization of the previous games to frameworks when players do not observe everything.
- They may not know what other players do (e.g. in voting afternoon voters do not know what happened in the morning)
- They may not know what other players know : private information (e.g. poker).
- We model it by uncertainty of the players on the current node on the tree.
- In addition to previous structure, each set of nodes P^i of positions controlled by Player i is partitionned in a number of information sets P^i_k.
- For each P^i_k there is an action set A^i_k and a function f^i_k.
- $f^i_k : P^i_k \times A^i_k \to Z$ such that, for any $p \in P^i_k$, $f^i_k(p, \cdot)$ is a bijection between A^i_k and $S(p)$.

Play of the game

Similar to the perfect information game except that

- If the current node is $p_t \in Q^i_k$ then player i (and only him) is informed that it is his turn.

Play of the game

Similar to the perfect information game except that

- If the current node is $p_t \in Q_k^i$ then player i (and only him) is informed that it is his turn.
- He is informed that p_t is in Q_k^i (i.e. the value of k) but not of the precise node p_t.

Play of the game

Similar to the perfect information game except that

- If the current node is $p_t \in Q^i_k$ then player i (and only him) is informed that it is his turn.
- He is informed that p_t is in Q^i_k (i.e. the value of k) but not of the precise node p_t.
- He has to choose an action $a^i_t \in A^i_k$ (not observed by other players).

Play of the game

Similar to the perfect information game except that

- If the current node is $p_t \in Q^i_k$ then player i (and only him) is informed that it is his turn.
- He is informed that p_t is in Q^i_k (i.e. the value of k) but not of the precise node p_t.
- He has to choose an action $a^i_t \in A^i_k$ (not observed by other players).
- The next node is $p_{t+1} = f^i_k(p_t, a^i_t)$.

Play of the game

Similar to the perfect information game except that

- If the current node is $p_t \in Q^i_k$ then player i (and only him) is informed that it is his turn.
- He is informed that p_t is in Q^i_k (i.e. the value of k) but not of the precise node p_t.
- He has to choose an action $a^i_t \in A^i_k$ (not observed by other players).
- The next node is $p_{t+1} = f^i_k(p_t, a^i_t)$.
- Nature works exactly as in the perfect information case.

See examples on the blackboard.

Strategies

- A pure strategy of player i is an element of $\Pi_k A^i_k$.

Strategies

- A pure strategy of player i is an element of $\Pi_k A^i_k$.
- A mixed strategy of player i is an element of $\Delta\left(\Pi_k A^i_k\right)$.

Strategies

- A pure strategy of player i is an element of $\Pi_k A^i_k$.
- A mixed strategy of player i is an element of $\Delta\left(\Pi_k A^i_k\right)$.
- A behavior strategy of player i is an element of $\Pi_k \Delta(A^i_k))$.

Strategies

- A pure strategy of player i is an element of $\Pi_k A^i_k$.
- A mixed strategy of player i is an element of $\Delta\left(\Pi_k A^i_k\right)$.
- A behavior strategy of player i is an element of $\Pi_k \Delta(A^i_k))$.

For any profile of (pure, mixed, behavior) strategy one can infer a probability on terminal nodes and hence an (expected) payoff for each player.

Strategies

- A pure strategy of player i is an element of $\Pi_k A^i_k$.
- A mixed strategy of player i is an element of $\Delta\left(\Pi_k A^i_k\right)$.
- A behavior strategy of player i is an element of $\Pi_k \Delta(A^i_k))$.

For any profile of (pure, mixed, behavior) strategy one can infer a probability on terminal nodes and hence an (expected) payoff for each player.
Basically in a mixed strategy the player choose a global lottery, in a behavior strategy he chooses a family of local lotteries. These are not the same things (consider the dimension of the sets for example)!

Strategies

- A pure strategy of player i is an element of $\Pi_k A_k^i$.
- A mixed strategy of player i is an element of $\Delta\left(\Pi_k A_k^i\right)$.
- A behavior strategy of player i is an element of $\Pi_k \Delta(A_k^i))$.

For any profile of (pure, mixed, behavior) strategy one can infer a probability on terminal nodes and hence an (expected) payoff for each player.
Basically in a mixed strategy the player choose a global lottery, in a behavior strategy he chooses a family of local lotteries. These are not the same things (consider the dimension of the sets for example)!
Example of drunken driver on blackboard.

Kuhn's result

Behavioral strategies are more intuitive and less numerous. On the other hand, Nash existence theorem holds for mixed strategy. Fortunately, mixed and behavior strategies are "equivalent" in a wide setting.

Kuhn's result

Behavioral strategies are more intuitive and less numerous. On the other hand, Nash existence theorem holds for mixed strategy. Fortunately, mixed and behavior strategies are “equivalent” in a wide setting.
We say that a game is with perfect memory if players never forget anything (it certainly does not mean they know everything !)

Kuhn's result

Behavioral strategies are more intuitive and less numerous. On the other hand, Nash existence theorem holds for mixed strategy. Fortunately, mixed and behavior strategies are "equivalent" in a wide setting.
We say that a game is with perfect memory if players never forget anything (it certainly does not mean they know everything !)

Theorem (Kuhn, 1953)

Assume the game is with perfect memory. Then if s^i is a mixed strategy, there exists an equivalent behavior strategy σ^i : whatever the other players do the distribution on terminal nodes will be the same if i play s^i or σ^i. Similarly, for every behavior strategy σ^i there exists an (in fact several in most cases) equivalent mixed strategy s^i.

Kuhn's result

Behavioral strategies are more intuitive and less numerous. On the other hand, Nash existence theorem holds for mixed strategy. Fortunately, mixed and behavior strategies are "equivalent" in a wide setting.
We say that a game is with perfect memory if players never forget anything (it certainly does not mean they know everything !)

Theorem (Kuhn, 1953)

Assume the game is with perfect memory. Then if s^i is a mixed strategy, there exists an equivalent behavior strategy σ^i : whatever the other players do the distribution on terminal nodes will be the same if i play s^i or σ^i. Similarly, for every behavior strategy σ^i there exists an (in fact several in most cases) equivalent mixed strategy s^i.

An immediate corollary of Kuhn and Nash theorems is

Corollary

Any game in extensive form has an equilibrium in behavior strategies.

Model

A one-shot game $\Gamma = (N, A, g)$ is repeated over and over. In each stage t players play simultaneously, get a payoff g_t and then observe the actions of everyone.

Model

A one-shot game $\Gamma = (N, A, g)$ is repeated over and over. In each stage t players play simultaneously, get a payoff g_t and then observe the actions of everyone.

One can model it has an extensive game with imperfect information. Players play say behavior strategies (perfect memory so it is not important) and try to maximize the mean of their stage payoffs.

Model

A one-shot game $\Gamma = (N, A, g)$ is repeated over and over. In each stage t players play simultaneously, get a payoff g_t and then observe the actions of everyone.

One can model it has an extensive game with imperfect information. Players play say behavior strategies (perfect memory so it is not important) and try to maximize the mean of their stage payoffs.

Two main way of aggregating stage payoffs :

- In the game with horizon $T \in \mathbb{N}$ player i maximizes the expectation of $\frac{1}{T}\sum_{t=1}^{T} g_t^i$.

Model

A one-shot game $\Gamma = (N, A, g)$ is repeated over and over. In each stage t players play simultaneously, get a payoff g_t and then observe the actions of everyone.

One can model it has an extensive game with imperfect information. Players play say behavior strategies (perfect memory so it is not important) and try to maximize the mean of their stage payoffs.

Two main way of aggregating stage payoffs :

- In the game with horizon $T \in \mathbb{N}$ player i maximizes the expectation of $\frac{1}{T}\sum_{t=1}^{T} g_t^i$.
- In the λ-discounted game ($\lambda \in]0,1]$) player i maximizes the expectation of $\lambda \sum_{t=1}^{+\infty}(1-\lambda)^{t-1} g_t^i$.

Model

A one-shot game $\Gamma = (N, A, g)$ is repeated over and over. In each stage t players play simultaneously, get a payoff g_t and then observe the actions of everyone.
One can model it has an extensive game with imperfect information. Players play say behavior strategies (perfect memory so it is not important) and try to maximize the mean of their stage payoffs.
Two main way of aggregating stage payoffs :

- In the game with horizon $T \in \mathbb{N}$ player i maximizes the expectation of $\frac{1}{T}\sum_{t=1}^{T} g_t^i$.
- In the λ-discounted game ($\lambda \in]0,1]$) player i maximizes the expectation of $\lambda \sum_{t=1}^{+\infty}(1-\lambda)^{t-1} g_t^i$.

Benefit of repetition

An obvious benefit is that one can “convexify” the set of equilibrium of the one shot game. For example in war of sexes players can alternate between their prefered occupation

Benefit of repetition

An obvious benefit is that one can “convexify” the set of equilibrium of the one shot game. For example in war of sexes players can alternate between their prefered occupation

Less obvious benefit is that repetition favorizes cooperation.

Benefit of repetition

An obvious benefit is that one can "convexify" the set of equilibrium of the one shot game. For example in war of sexes players can alternate between their prefered occupation

Less obvious benefit is that repetition favorizes cooperation.

Threats of a later punishment can incentivize players to cooperate and achieve a "social optimum" that is unattainable in the one shot game.

Benefit of repetition

An obvious benefit is that one can "convexify" the set of equilibrium of the one shot game. For example in war of sexes players can alternate between their prefered occupation
Less obvious benefit is that repetition favorizes cooperation.
Threats of a later punishment can incentivize players to cooperate and achieve a "social optimum" that is unattainable in the one shot game.
Huge literature on these "folk theorems", we will illustrate this on the prisoner dilemma.

Repeated prisonner dilemma : discounted case

	C	D
C	3,3	0,4
D	4,0	1,1

Denote E_λ the set of equilibrium payoffs in the λ-discounted game and E_T the set of equilibrium payoffs in the game with finite horizon T.

Repeated prisonner dilemma : discounted case

	C	D
C	3,3	0,4
D	4,0	1,1

Denote E_λ the set of equilibrium payoffs in the λ-discounted game and E_T the set of equilibrium payoffs in the game with finite horizon T.

Proposition

$(3,3) \in E_\lambda$ *if* $\lambda \leq 2/3$ *(players are patient enough).*

Repeated prisonner dilemma : discounted case

	C	D
C	3,3	0,4
D	4,0	1,1

Denote E_λ the set of equilibrium payoffs in the λ-discounted game and E_T the set of equilibrium payoffs in the game with finite horizon T.

Proposition

$(3,3) \in E_\lambda$ *if* $\lambda \leq 2/3$ *(players are patient enough).*

Démonstration.

Consider the profile in which each player plays the following strategy : play C as long as everyone has always played C, and play D forever if someone ever played D in the past. This is an equilibria for $\lambda \leq 2/3$ $\square$

Repeated prisonner dilemma : finite horizon case

	C	D
C	3,3	0,4
D	4,0	1,1

Very different result

Proposition

$E_T = \{(1,1)\}$ *for every* T.

Repeated prisonner dilemma : finite horizon case

	C	D
C	3,3	0,4
D	4,0	1,1

Very different result

Proposition

$E_T = \{(1,1)\}$ *for every* T.

Démonstration.

In the last stage player will play D on the equilibrium path. Hence, playing C in stage $T-1$ makes no sense (deviating to D will increase the payoff in stage $T-1$ and not decrease the payoff in stage T). Induction gives that players will always play D in any equilibrium. $\square$