The 40th conference on
The mathematics of operations research
Landelijk Netwerk Mathematische Besliskunde
Lunteren, January 13, 2015

Rational Learning in Repeated Games

Ehud Kalai
Northwestern University

Definition:
An n-person strategic game is a function:

$u: S \to R^n$ with $S = \times_i S^i$.

S^i is the set of *strategies* of player I,

S is the set of *strategy profiles* or *strategy configurations*, and

u^i is the *payoff function* of player i.

Nash equilibrium

Extends the idea of equilibrium from supply and demand to general behavior between interacting players

Has taken over as a major analytical tool in economics

Operations management, political science and computer science are going through similar transformations

Coincides with behavior predicted by survival of the fittest.

Definition: A **Nash equilibrium** is a configuration of individual strategies, each optimal (best response) relative to the others, i.e., no player has an incentive to unilaterally deviate from the configuration.

Simple familiar examples:

Everybody driving on the right side of the road.

Markets, real and on the web.

Complementarities in production:

a. Simultaneous production of software and of hardware
also
b. No production of software with no production of hardware.
But
production of software without production of hardware is not an equilibrium.

Common language, common system of measurements…

Example: be generous or selfish

Aka Prisoners' dilemma

when a $1 donation yields your opponent $3.

		She: generous	She: selfish
He	generous	2, 2	-1, 3
	selfish	3,-1	0, 0

The only Nash equilibrium is non-cooperative:
Both players choose the selfish action.

Example: a shy woman with a bold man,

Aka match pennies

he wants to be with her she wants to be alone.

		She: .5 in	She: .5 out
He	.5 in	1, 0	0, 1
	.5 out	0, 1	1, 0

This game has no "pure strategy" Nash equilibrium, but it has a "mixed strategy" equilibrium: each player chooses one of the two options with equal probability.

According to Nash: if we allow for mixed strategies, every finite game has a Nash equilibrium.

Computer choice game with known types

Each has to choose $\mathcal{PC}$ or $\mathcal{M}$.
He likes $\mathcal{PC}$ she likes $\mathcal{M}$, but they also like to make the same choice.

His payoff: 1 if they choose the same computer (0 otherwise)
+ .2 if he chooses $\mathcal{PC}$ (0 otherwise).

Her payoff: 1 if they choose the same computer (0 otherwise)
+ .2 if she chooses $\mathcal{M}$ (0 otherwise).

This game has
two pure strategy equilibria:
(1) both choose $\mathcal{PC}$ and (2) both choose $\mathcal{M}$;

and one mixes strategy equilibrium:
he randomizes .6 to .4 between $\mathcal{PC}$ and $\mathcal{M}$, and
she randomizes .4 to .6 between $\mathcal{PC}$ and $\mathcal{M}$.

Aka Prisoners' dilemma

Example: be generous or selfish, **repeated play**

- The same two players play the "stage game" in periods 1,2,…, with "perfect monitoring."

		She: generous	She: selfish
He	generous	2, 2	-1, 3
He	selfish	3,-1	0, 0

- At the end of every period, each is told the choice of his opponent and receives a payoff according to the table.
- A strategy is a function *f* from histories of play to period choice, e.g. $f^1\begin{pmatrix} g & g & g \\ g & s & g \end{pmatrix} = (gen.\, wp\ .9, selfish\ wp\ .10)$
- Present value is computed with a discount parameter d.

(f^1, f^2) is an **equilibrium** if each f^i maximizes the expected present value of the total future payoffs.

Example: If d > 1/3, both play tit for tat ; average payoff = 2,2 .

Example of a Bayesian game

Computer choice game with **unknown types**

- Each of two players have to choose $\mathcal{PC}$ or $\mathcal{M}$.
- Each is of one random type: likes $\mathcal{PC}$ or likes $\mathcal{M}$, with prob .50 - .50
- Each player knows his own type, but only the probabilities of the opponent's type.
- Identical individual payoff functions:

1 if you choose the same computer as your opponent (0 otherwise)

+ .2 if you chooses the computer you like (0 otherwise).

In Bayesian Nash equilibrium strategies are type dependent.

For example: choose your favorite computer, i.e., Choose $\mathcal{PC}$, if you like $\mathcal{PC}$; and choose $\mathcal{M}$ if you like $\mathcal{M}$.

Notice that the equilibrium is efficient if they happen to be of the same type, with prob. .5; and inefficient otherwise.

The mechanism design literature deals with fixing such inefficiencies.

Example of a **Bayesian repeated game**

Computer choice game with unknown types played repeatedly

- First, players are assigned types as above, to remain fixed throughout the repeated play.
- Then, the computer choice game is played in periods 1,2,… with perfect monitoring and with discounted sum of payoffs.

Strategy profile $\boldsymbol{f} = (\boldsymbol{f}^1, \boldsymbol{f}^2)$:
a vector of vectors of repeated game strategies.

For Player 1, for example: $\boldsymbol{f}^1 = (\boldsymbol{f}^{1,\mathcal{PC}}, \boldsymbol{f}^{1,\mathcal{M}})$

- $\boldsymbol{f}^{1,\mathcal{PC}}(h)$ describes the computer choice probabilities of his $\mathcal{PC}$- type, after the history of observed past choices h.
- $\boldsymbol{f}^{1,\mathcal{M}}(h)$ is the same, but for his $\mathcal{M}$- type.

Example of a **Baye**… **n repeated game**

Computer choic… …known types played repeatedly

- First, pla… …above, to remain f…
- Then, th… periods 1,2,… with … discounted sum of payoffs.

For example

$f^{1,\mathcal{PC}}(\emptyset) = (.80,.20)$ means that initially he chooses $\mathcal{PC}$ wp .80 and $\mathcal{M}$ wp .20.

$f^{1,\mathcal{PC}}\left(\begin{smallmatrix}\mathcal{MMM}\\ \mathcal{MMM}\end{smallmatrix}\right) = (0,1)$ means that if they both chose $\mathcal{M}$s in the first three periods, in the fourth period he would choose $\mathcal{M}$ for sure.

Strategy profile $f = (f^1, f^2)$:
a vector of vectors of repeated game strategies.

<u>For Player 1, for example</u>: $f^1 = (f^{1,\mathcal{PC}}, f^{1,\mathcal{M}})$

- $f^{1,\mathcal{PC}}(h)$ describes the computer choice probabilities of his $\mathcal{PC}$- type, after the history of observed past choices *h*.
- $f^{1,\mathcal{M}}(h)$ is the same, but for his $\mathcal{M}$- type.

Strategy profile $f = (f^1, f^2)$ is **a Bayesian equilibrium** if for every player and type, his repeated game strategy maximized the type's expected present value of payoff, given the distribution of the opponent's types and their repeated game strategies.

Sketch of Kalai and Lehrer (1993) process of learning: for example:

<u>assume that Pl 1 turns out to be a $\mathcal{PC}$ type and Pl 2 a $\mathcal{M}$ type:</u>

They will play $(f^{1,\mathcal{PC}}, f^{2,\mathcal{M}})$, with

$f^{1,\mathcal{PC}}$ being optimal against the .50 -.50 belief that he is facing $f^{2,\mathcal{PC}}$ vs $f^{2,\mathcal{M}}$;

(similarly for Pl 2).

But maximizing the PV of future payoffs, implies that Pl 1 (and Pl 2) plays optimally relative to **Bayesian updated beliefs**:

Instead of the initial prior belief, $prob(t^2=\mathcal{PC})=.5$, after any history of past plays h, he will use the posterior belief, $prob(t^2=\mathcal{PC} \mid h)$, and optimize against it.

The $prob(t^2=\mathcal{PC} \mid h)$ must converge, but not necessarily to the true probability, which is zero since Pl 2 is a $\mathcal{M}$-type.

The convergence is a direct consequence of the Martingale convergence theorem.

But maximizing the PV of future payoffs, implies that Pl 1 (and Pl 2) plays optimally relative to **Bayesian updated beliefs**:

Instead of the initial prior belief, $prob(t^2=\mathcal{PC})=.5$, after any history of past plays h, he will use the posterior belief, $prob(t^2=\mathcal{PC} \mid h)$, and optimize against it.

The $prob(t^2=\mathcal{PC} \mid h)$ must converge, but not necessarily to the true probability, which is zero since Pl 2 is a $\mathcal{M}$-type.

Nevertheless, Pl 1's predictions of the future play will become accurate, he will predict Pl 2's choices under $f^{2,\mathcal{M}}$, as if he knew that she is the $\mathcal{M}$- type.

But maximizing the PV of future payoffs, implies that Pl 1 (and Pl 2) plays optimally relative to **Bayesian updated beliefs**:
Instead of the initial prior belief, $prob(t^2=\mathcal{PC})=.5$, after any history of past plays h, he will use the posterior belief, *prob(* … optimize against it.

This follows from the merging literature, see Kalai and Lehrer (1994) for sufficient conditions that hold in repeated games

The …
to th…
type.

Nevertheless, Pl 1's predictions of the future play will become accurate, he will predict Pl 2's choices under $f^{2,\mathcal{M}}$, as if he knew that she is the $\mathcal{M}$- type.

It follows that:

Theorem (Kalai and Lehrer 1993b) at a Bayesian eq players converge to play a **subjective equilibrium** of the repeated game: At such an equilibrium they each play optimally relative to his beliefs, which may be false, but not contradictable by the observed data.

See von Hayek (1937) for the idea of subjective equilibrium, and

see also Battigalli (1987) Fudenberg and Levin (1993) on the idea of self-confirming subjective equilibria in different contexts

It follows that:

Theorem (Kalai and Lehrer 1993b) at a Bayesian eq players converge to play a **subjective equilibrium** of the repeated game: At such an equilibrium they each play optimally relative to his beliefs, which may be false, but not contradictable by the observed data.

Theorem (Kalai and Lehrer 1993a) at Bayesian equilibria of two-person games, the play will converge to an approximate Nash equilibrium of the repeated game, as if the types of both players are common knowledge.

The first theorem holds for any number of players.

For the second theorem with more than two players assume subjective independence: knowing his realized type, every player believes that his opponents' types are independent of each other.

Example: Repeated Production

n players producing widgets at time periods 1,2,3,…

At the beginning of each period, a producer decides

1. the type of widgets he will produce, and
2. his selling price

At the end of the period he observes his competitors' choices and collects his period's profit.

Each producer knows only his own (constant) production capabilities and costs.

At a Bayesian equilibrium he maximizes the expected present value of all his future profits.

A producer may act strategically. For example, he may:

Learn: experiments with period choices to test the competitors responses.

Teach: sell widgets at low prices in selected periods to deceive his competitors about his cost and discourage their participation.

Nevertheless, with time:

The producers learn to predict the future choices of their competitors,

and play as if everybody's capabilities and costs are common knowledge.

Thank you!