Evolutionary game theory and cognition

Artem Kaznatcheev

McGill University

November 12, 2009

Zero-sum games

- A game between two players (Alice and Bob) is represented by a matrix G of pairs.

Example

$$\begin{bmatrix} (3,1) & (2,3) \\ (-1,2) & (3,-1) \end{bmatrix}$$

Zero-sum games

- A game between two players (Alice and Bob) is represented by a matrix G of pairs.

Example

$$\begin{bmatrix} (3,1) & (2,3) \\ (-1,2) & (3,-1) \end{bmatrix}$$

- If Alice plays strategy i and Bob plays strategy j then $(a,b) := G_{ij}$ is the outcome, where a corresponds to the change in Alice's utility and b to Bob's.

Zero-sum games

Definition

A game G is a zero-sum game if for each $(a, b) := G_{ij}$ we have $a + b = 0$.

Zero-sum games

Definition

A game G is a zero-sum game if for each $(a, b) := G_{ij}$ we have $a + b = 0$.

Example

$$\begin{bmatrix} (1, -1) & (-1, 1) \\ (-1, 1) & (1, -1) \end{bmatrix}$$

Zero-sum games

Definition

A game G is a zero-sum game if for each $(a, b) := G_{ij}$ we have $a + b = 0$.

Example

$$\begin{bmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{bmatrix}$$

- Zero-sum games are the epitome of competition. Any gain for Alice is a loss for Bob, and vice-versa.

Coordination games

Definition

A two-strategy game G is a coordination game if we have

$$G = \begin{bmatrix} (a_1, b_1) & (c_2, d_1) \\ (c_1, d_2) & (a_2, b_2) \end{bmatrix}$$

And $a_1 > c_1$, $a_2 > c_2$, $b_1 > d_1$, $b_2 > d_2$.

Coordination games

Definition

A two-strategy game G is a coordination game if we have

$$G = \begin{bmatrix} (a_1, b_1) & (c_2, d_1) \\ (c_1, d_2) & (a_2, b_2) \end{bmatrix}$$

And $a_1 > c_1$, $a_2 > c_2$, $b_1 > d_1$, $b_2 > d_2$.

Examples

$$\begin{bmatrix} (1,1) & (-1,-1) \\ (-1,-1) & (1,1) \end{bmatrix}, \begin{bmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{bmatrix}, \begin{bmatrix} (4,4) & (0,2) \\ (2,0) & (3,3) \end{bmatrix}$$

Coordination games

Definition

A two-strategy game G is a coordination game if we have

$$G = \begin{bmatrix} (a_1, b_1) & (c_2, d_1) \\ (c_1, d_2) & (a_2, b_2) \end{bmatrix}$$

And $a_1 > c_1$, $a_2 > c_2$, $b_1 > d_1$, $b_2 > d_2$.

Examples

$$\begin{bmatrix} (1,1) & (-1,-1) \\ (-1,-1) & (1,1) \end{bmatrix}, \begin{bmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{bmatrix}, \begin{bmatrix} (4,4) & (0,2) \\ (2,0) & (3,3) \end{bmatrix}$$

- The diagonals are always better for both players, they just have to figure out how to pick the same strategy.
- Captures the idea of win-win, lose-lose situations.

What do these two types of games tell us?

- Zero-sum and coordination games are mutually exclusive: there is no game that is both zero-sum and a coordination game.

What do these two types of games tell us?

- Zero-sum and coordination games are mutually exclusive: there is no game that is both zero-sum and a coordination game.
- Upside: zero-sum and coordination provide a good duality between impossibility of cooperation and obvious cooperation.

What do these two types of games tell us?

- Zero-sum and coordination games are mutually exclusive: there is no game that is both zero-sum and a coordination game.
- Upside: zero-sum and coordination provide a good duality between impossibility of cooperation and obvious cooperation.
- Downside: both types of games are really boring. The most interesting games (from a mathematical and modeling point of view) are neither zero-sum nor coordination.

What do these two types of games tell us?

- Zero-sum and coordination games are mutually exclusive: there is no game that is both zero-sum and a coordination game.
- Upside: zero-sum and coordination provide a good duality between impossibility of cooperation and obvious cooperation.
- Downside: both types of games are really boring. The most interesting games (from a mathematical and modeling point of view) are neither zero-sum nor coordination.
- Being non-zero-sum does not ensure cooperation.

Prisoner's dilemma

$$\begin{bmatrix} (b-c, b-c) & (-c, b) \\ (b, -c) & (0, 0) \end{bmatrix}$$

- b is the benefit of receiving and c is the cost of giving.
- Strategy 1 is called cooperate or C and strategy 2 is called defect or D.

Prisoner's dilemma

$$\begin{bmatrix} (b-c, b-c) & (-c, b) \\ (b, -c) & (0, 0) \end{bmatrix}$$

- b is the benefit of receiving and c is the cost of giving.
- Strategy 1 is called cooperate or C and strategy 2 is called defect or D.
- The rational strategy (or Nash equilibrium) is mutual defection.

Prisoner's dilemma

$$\begin{bmatrix} (b-c, b-c) & (-c, b) \\ (b, -c) & (0, 0) \end{bmatrix}$$

- b is the benefit of receiving and c is the cost of giving.
- Strategy 1 is called cooperate or C and strategy 2 is called defect or D.
- The rational strategy (or Nash equilibrium) is mutual defection.
- The best for the players taken together (or Pareto optimum) is mutual cooperation.

Nash equilibrium

Definition

A strategy pair (p, q) is a Nash equilibrium of a game G if for all other strategies r we have:

$$fst(G(p, q)) \geq fst(G(r, q))$$

and

$$snd(G(p, q)) \geq snd(G(p, r))$$

Nash equilibrium

Definition

A strategy pair (p, q) is a Nash equilibrium of a game G if for all other strategies r we have:

$$fst(G(p, q)) \geq fst(G(r, q))$$

and

$$snd(G(p, q)) \geq snd(G(p, r))$$

- Informally: neither Alice nor Bob can improve their payoff by unilateral change of strategy.

Nash equilibrium

Definition

A strategy pair (p, q) is a Nash equilibrium of a game G if for all other strategies r we have:

$$fst(G(p,q)) \geq fst(G(r,q))$$

and

$$snd(G(p,q)) \geq snd(G(p,r))$$

- Informally: neither Alice nor Bob can improve their payoff by unilateral change of strategy.
- If we only allow pure strategies then replace $G(i,j)$ by G_{ij}
- If we allow mixed strategies, then every game has at least one Nash equilibrium

Pareto optimum

Definition

A strategy pair (p, q) is a Pareto optimum of a game G is there is no other strategy pair (p', q') such that $G(p', q') > G(p, q)$

Pareto optimum

Definition

A strategy pair (p, q) is a Pareto optimum of a game G is there is no other strategy pair (p', q') such that $G(p', q') > G(p, q)$

- Informally: there is no other strategy such that both Alice and Bob get a better payoff.
- Every game has at least one Pareto optimum

Pareto optimum

Definition

A strategy pair (p, q) is a Pareto optimum of a game G is there is no other strategy pair (p', q') such that $G(p', q') > G(p, q)$

- Informally: there is no other strategy such that both Alice and Bob get a better payoff.
- Every game has at least one Pareto optimum

Example

$$\begin{bmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{bmatrix}, \begin{bmatrix} (2,-3) & (-1,1) \\ (0,0) & (-2,2) \end{bmatrix}$$

Cognitive demands of rationality

- Alice needs to be aware of her own utility function

Cognitive demands of rationality

- Alice needs to be aware of her own utility function
- To check if she is currently in Nash equilibrium (at least from her perspective) Alice needs to be able to simulate the game in her mind (thus she must understand the interaction)

Cognitive demands of rationality

- Alice needs to be aware of her own utility function
- To check if she is currently in Nash equilibrium (at least from her perspective) Alice needs to be able to simulate the game in her mind (thus she must understand the interaction)
- To find a Nash equilibrium Alice needs to be able to simulate the game and she must be able to place yourself in Bob's shoes.

Cognitive demands of rationality

- Alice needs to be aware of her own utility function
- To check if she is currently in Nash equilibrium (at least from her perspective) Alice needs to be able to simulate the game in her mind (thus she must understand the interaction)
- To find a Nash equilibrium Alice needs to be able to simulate the game and she must be able to place yourself in Bob's shoes.
- Do we even expect humans to be able to do all of this?

Cognitive demands of rationality

- Alice needs to be aware of her own utility function
- To check if she is currently in Nash equilibrium (at least from her perspective) Alice needs to be able to simulate the game in her mind (thus she must understand the interaction)
- To find a Nash equilibrium Alice needs to be able to simulate the game and she must be able to place yourself in Bob's shoes.
- Do we even expect humans to be able to do all of this?
- Let's bound rationality and see what happens!

Evolutionary game theory

- Strategy is a genetic trait and immutable by the agent.
- All cognition is stripped away

Evolutionary game theory

- Strategy is a genetic trait and immutable by the agent.
- All cognition is stripped away
- Game payoffs change the fitness of the agent.
- Agents reproductive rate increases with higher fitness.

Evolutionary game theory

- Strategy is a genetic trait and immutable by the agent.
- All cognition is stripped away
- Game payoffs change the fitness of the agent.
- Agents reproductive rate increases with higher fitness.
- Simplest model of biological evolution.
- Also applicable outside of biology.

Evolutionary game theory

- Strategy is a genetic trait and immutable by the agent.
- All cognition is stripped away
- Game payoffs change the fitness of the agent.
- Agents reproductive rate increases with higher fitness.
- Simplest model of biological evolution.
- Also applicable outside of biology.
- What happens to rationality?

Evolutionary stable strategy

Definition

A strategy s is an evolutionary stable strategy for a game G if for all other strategies r we have (a) $fst(G(s,s)) > fst(G(r,s))$, or (b) $fst(G(s,s)) = fst(G(r,s))$ and $fst(G(s,r)) > fst(G(r,r))$.

Evolutionary stable strategy

Definition

A strategy s is an evolutionary stable strategy for a game G if for all other strategies r we have (a) $fst(G(s,s)) > fst(G(r,s))$, or (b) $fst(G(s,s)) = fst(G(r,s))$ and $fst(G(s,r)) > fst(G(r,r))$.

- Consider a population all with strategy s, a mutant with strategy r cannot invade the population only if one of the following conditions holds:

Evolutionary stable strategy

Definition

A strategy s is an evolutionary stable strategy for a game G if for all other strategies r we have (a) $fst(G(s,s)) > fst(G(r,s))$, or (b) $fst(G(s,s)) = fst(G(r,s))$ and $fst(G(s,r)) > fst(G(r,r))$.

- Consider a population all with strategy s, a mutant with strategy r cannot invade the population only if one of the following conditions holds:
  - r has a higher fitness than s in a population of all s.

Evolutionary stable strategy

Definition

A strategy s is an evolutionary stable strategy for a game G if for all other strategies r we have (a) $fst(G(s,s)) > fst(G(r,s))$, or (b) $fst(G(s,s)) = fst(G(r,s))$ and $fst(G(s,r)) > fst(G(r,r))$.

- Consider a population all with strategy s, a mutant with strategy r cannot invade the population only if one of the following conditions holds:
  - r has a higher fitness than s in a population of all s.
  - r has the same fitness when interacting with s and the same or greater fitness when interacting with other r.

Evolutionary stable strategy

Definition

A strategy s is an evolutionary stable strategy for a game G if for all other strategies r we have (a) $fst(G(s,s)) > fst(G(r,s))$, or (b) $fst(G(s,s)) = fst(G(r,s))$ and $fst(G(s,r)) > fst(G(r,r))$.

- Consider a population all with strategy s, a mutant with strategy r cannot invade the population only if one of the following conditions holds:
  - r has a higher fitness than s in a population of all s.
  - r has the same fitness when interacting with s and the same or greater fitness when interacting with other r.
- Compare this to the Nash equilibrium conditions.

Evolutionary stable strategy

Definition

A strategy s is an evolutionary stable strategy for a game G if for all other strategies r we have (a) $fst(G(s,s)) > fst(G(r,s))$, or (b) $fst(G(s,s)) = fst(G(r,s))$ and $fst(G(s,r)) > fst(G(r,r))$.

- Consider a population all with strategy s, a mutant with strategy r cannot invade the population only if one of the following conditions holds:
  - r has a higher fitness than s in a population of all s.
  - r has the same fitness when interacting with s and the same or greater fitness when interacting with other r.
- Compare this to the Nash equilibrium conditions.
- The conditions are almost identical: we can think of the evolutionary process as a rational process (entity?)!.

Wait a second: what about cooperation?

- The ESS predicts mutual defection in the Prisoner's dilemma, but we observe cooperation through out nature.

Wait a second: what about cooperation?

- The ESS predicts mutual defection in the Prisoner's dilemma, but we observe cooperation through out nature.
- The assumptions of the ESS:

Wait a second: what about cooperation?

- The ESS predicts mutual defection in the Prisoner's dilemma, but we observe cooperation through out nature.
- The assumptions of the ESS:
  - Random interactions (inviscid population)

Wait a second: what about cooperation?

- The ESS predicts mutual defection in the Prisoner's dilemma, but we observe cooperation through out nature.
- The assumptions of the ESS:
  - Random interactions (inviscid population)
  - No repeated interactions

Wait a second: what about cooperation?

- The ESS predicts mutual defection in the Prisoner's dilemma, but we observe cooperation through out nature.
- The assumptions of the ESS:
  - Random interactions (inviscid population)
  - No repeated interactions
  - Zero cognition in individual agents

Wait a second: what about cooperation?

- The ESS predicts mutual defection in the Prisoner's dilemma, but we observe cooperation through out nature.
- The assumptions of the ESS:
  - Random interactions (inviscid population)
  - No repeated interactions
  - Zero cognition in individual agents
- Various augmentations of the model create fascinating results, among them: cooperation.

Cognitively relevant augmentations

- Kin selection: the ability to recognize your children, siblings and parents

Cognitively relevant augmentations

- Kin selection: the ability to recognize your children, siblings and parents
- Direct reciprocity: the ability to remember previous interactions

Cognitively relevant augmentations

- Kin selection: the ability to recognize your children, siblings and parents
- Direct reciprocity: the ability to remember previous interactions
- Indirect reciprocity: the ability to track social constructs like reputation