2. Conventions & perturbed dynamics

Jonathan Newton, Kyoto University

Conventions & perturbed dynamics

Conventional behavior amongst members of a population is a regularity such that

1. Everyone conforms to this regularity all of the time.
2. Everyone expects (1).
3. Given (1), everyone wishes to conform.

For example,

- Which side of the road to drive on?
- Is it good manners to hold a door open?
- Appropriate clothes to wear to the office.

Conventions & perturbed dynamics

Conventional behavior amongst members of a population is a regularity such that

1. Everyone conforms to this regularity all of the time.
2. Everyone expects (1).
3. Given (1), everyone wishes to conform.

Given the possibility of multiple conventions (stable equilibria), how can we decide which, if any, will occur in the long run?

Conventions & perturbed dynamics

Conventional behavior amongst members of a population is a regularity such that

1. Most people conform to this regularity most of the time.
2. Most people mostly expect (1) most of the time.
3. Given (1), most people wish to conform most of the time.

That is, we perturb aspects of the process, so that one or more of actions, expectations or preferences are noisy/stochastic/random.

Stochastic dynamics, cycles and trees

Characterization of a popular class of models that can handle a wide variety of behavior.

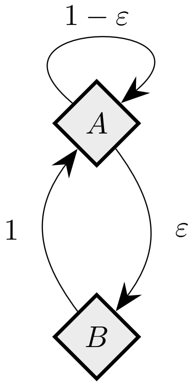
Alice's stochastic choice

Every period $t \in \mathbb{N}$, she can play A or B. She prefers A, so will usually play A. However, after playing A, she will sometimes make a mistake and play B.

- Define a family of Markov chains P^ε on state space $\{A, B\}$.
- If state is A, then play A with probability $1 - \varepsilon$, play B with probability ε.
- If state is B, then play A with probability 1.

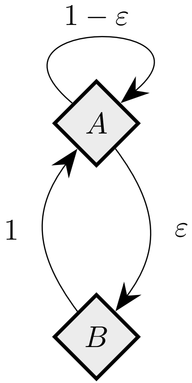
Alice's stochastic choice

- Define a family of Markov chains P^{ε} on state space $\{A, B\}$.
- If state is A, then play A with probability $1-\varepsilon$, play B with probability ε.
- If state is B, then play A with probability 1.



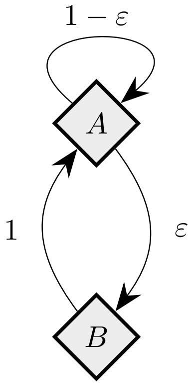
Alice's stochastic choice

- Define a family of Markov chains P^{ε} on state space $\{A, B\}$.
- If state is A, then play A with probability $1 - \varepsilon$, play B with probability ε.
- If state is B, then play A with probability 1.
- Invariant probability distribution μ^{ε} puts weight $\mu^{\varepsilon}(A) = \frac{1}{1+\varepsilon}$, $\mu^{\varepsilon}(B) = \frac{\varepsilon}{1+\varepsilon}$.



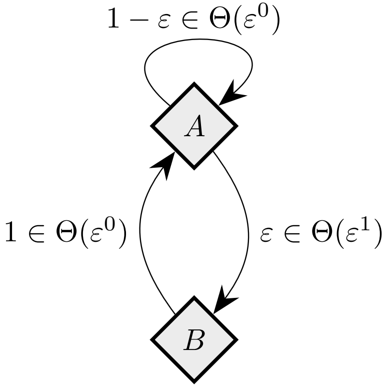
Alice's stochastic choice

- Define a family of Markov chains P^ε on state space $\{A, B\}$.
- If state is A, then play A with probability $1-\varepsilon$, play B with probability ε.
- If state is B, then play A with probability 1.
- Invariant probability distribution μ^ε puts weight $\mu^\varepsilon(A) = \frac{1}{1+\varepsilon}$, $\mu^\varepsilon(B) = \frac{\varepsilon}{1+\varepsilon}$.
- So if we let $\varepsilon \to 0$, then $\mu^\varepsilon(A) \to 1$.



Alice's stochastic choice

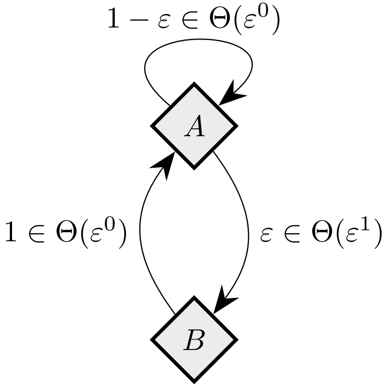
- Order of magnitude of probabilities can be given in terms of powers of ε.



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 :$$
$$\forall \varepsilon < \bar{\varepsilon},\ d_1\varepsilon^r \leq f(\varepsilon) \leq d_2\varepsilon^r$$

Alice's stochastic choice

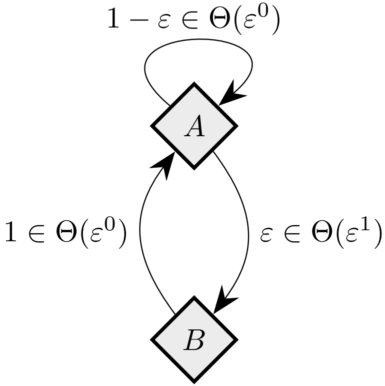
- Order of magnitude of probabilities can be given in terms of powers of ε.
- Here we only have ε^0 and ε^1, but in general can have ε^r, where larger r corresponds to lower probability events.



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 :$$
$$\forall \varepsilon < \bar{\varepsilon},\ d_1 \varepsilon^r \leq f(\varepsilon) \leq d_2 \varepsilon^r$$

Alice’s stochastic choice

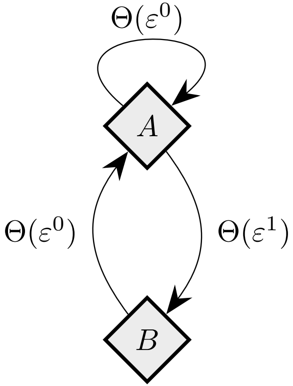
- Order of magnitude of probabilities can be given in terms of powers of ε.
- Here we only have ε^0 and ε^1, but in general can have ε^r, where larger r corresponds to lower probability events.
- For example, r might be higher for transitions that require more mistakes, larger mistakes, greater coordination of mistakes across players.



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 :$$
$$\forall \varepsilon < \bar{\varepsilon},\ d_1\varepsilon^r \leq f(\varepsilon) \leq d_2\varepsilon^r$$

Alice's stochastic choice

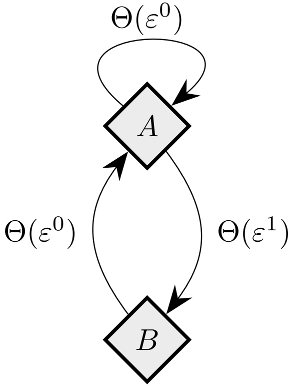
- If we don't know the exact transition probabilities, but know their orders of magnitude (r values),...



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 :$$
$$\forall \varepsilon < \bar{\varepsilon},\ d_1 \varepsilon^r \leq f(\varepsilon) \leq d_2 \varepsilon^r$$

Alice's stochastic choice

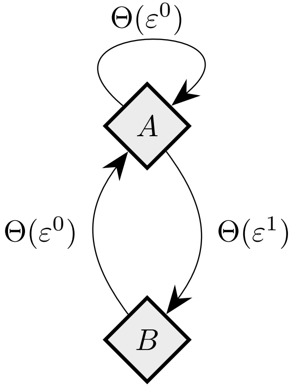
- If we don't know the exact transition probabilities, but know their orders of magnitude (r values),...
- ...can still show that if $\varepsilon \to 0$, then $\mu^{\varepsilon}(A) \to 1$.



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 : \\ \forall \varepsilon < \bar{\varepsilon},\ d_1\varepsilon^r \le f(\varepsilon) \le d_2\varepsilon^r$$

Alice's stochastic choice

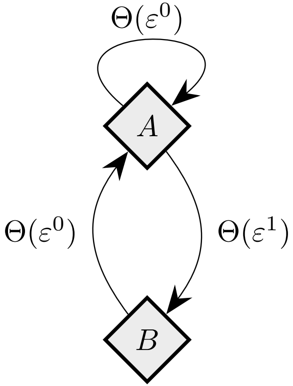
- If we don't know the exact transition probabilities, but know their orders of magnitude (r values),...
- ...can still show that if $\varepsilon \to 0$, then $\mu^{\varepsilon}(A) \to 1$.
- That is, r values are a useful statistic when we consider small perturbations.



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 :$$
$$\forall \varepsilon < \bar{\varepsilon},\ d_1 \varepsilon^r \leq f(\varepsilon) \leq d_2 \varepsilon^r$$

Alice's stochastic choice

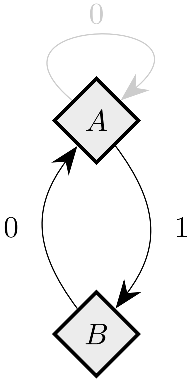
- If we don't know the exact transition probabilities, but know their orders of magnitude (r values),...
- ...can still show that if $\varepsilon \to 0$, then $\mu^{\varepsilon}(A) \to 1$.
- That is, r values are a useful statistic when we consider small perturbations.
- Assume small perturbations, i.e. small ε, from now on.



$$f(\varepsilon) \in \Theta(\varepsilon^r) \Leftrightarrow \exists d_1, d_2, \bar{\varepsilon} > 0 :$$
$$\forall \varepsilon < \bar{\varepsilon},\ d_1 \varepsilon^r \leq f(\varepsilon) \leq d_2 \varepsilon^r$$

Alice's stochastic choice

- Let's write the r values on the edges of our graph.
- Don't forget that larger values mean (much!) smaller probabilities.
- Interested in flow of probability *between* states.



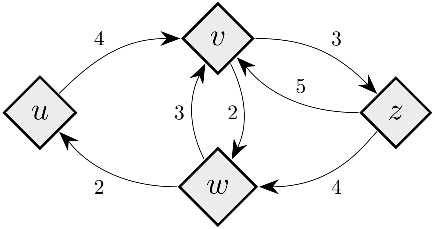
Evolution & Social Science, Brisbane 2023

Multiple players

- For situations with multiple players, can define an appropriate state space.
- For example, the set of all possible strategy profiles, or the set of all Nash equilibria.

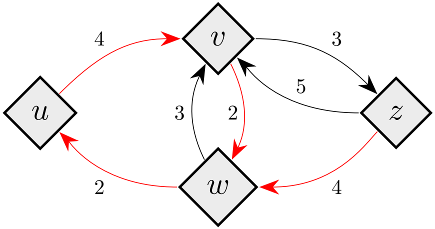
Multiple players

- Let's consider a situation with multiple states and solve it.
- Here we have four states, u, v, w, z.
- r values are given on edges. E.g. Transition from w to u has an order of magnitude of ε^2.



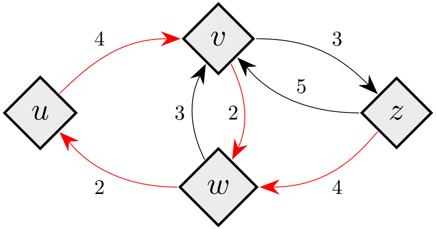
Cycles in behavior

- Select the most likely transitions from each state (shown in red).



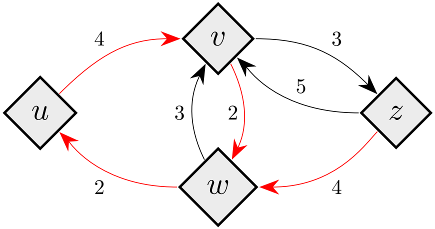
Cycles in behavior

- Select the most likely transitions from each state (shown in red).
- There will be a cycle in red edges.



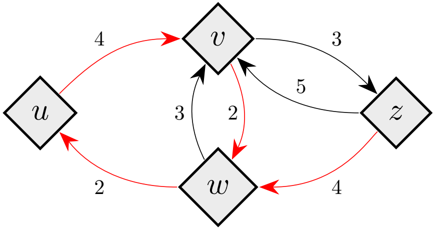
Cycles in behavior

- Select the most likely transitions from each state (shown in red).
- There will be a cycle in red edges.
- E.g. $u \to v \to w \to u$ is such a cycle.



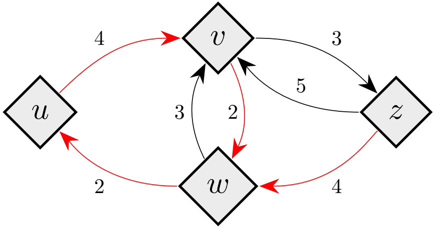
Cycles in behavior

- Select the most likely transitions from each state (shown in red).
- There will be a cycle in red edges.
- E.g. $u \to v \to w \to u$ is such a cycle.
- Within this cycle, most time will be spent at u.



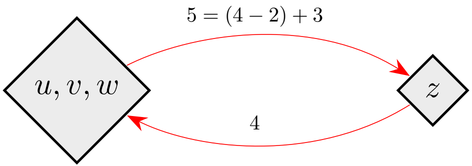
Cycles in behavior

- Select the most likely transitions from each state (shown in red).
- There will be a cycle in red edges.
- E.g. $u \to v \to w \to u$ is such a cycle.
- Within this cycle, most time will be spent at u.
- *Technical*. Time spent at v approximately $\frac{\varepsilon^{-2}}{\varepsilon^{-4}} = \varepsilon^2$ of the time spent in $\{u, v, w\}$.



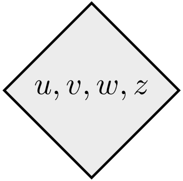
Aggregating cycles

- Merge the cycle into a single state.
- Incoming edges have same weight as before.
- Outgoing edges adjusted by the time spent within the cycle at the relevant states.
- Here we see that more time will be spent at $\{u, v, w\}$ than at z.
- *Technical*. Time spent at z approximately $\frac{\varepsilon^{-4}}{\varepsilon^{-5}} = \varepsilon^{1}$ of the time spent in $\{u, v, w, z\}$.



Aggregating cycles

- Repeat until you end up with one state.
- This cyclic decomposition has taught us about behavior at different levels.
  - Within $\{u, v, w, z\}$, the chain spends almost all of the time in $\{u, v, w\}$.
  - Within $\{u, v, w\}$, the chain spends almost all of the time at u.
  - u is the unique stochastically stable state.

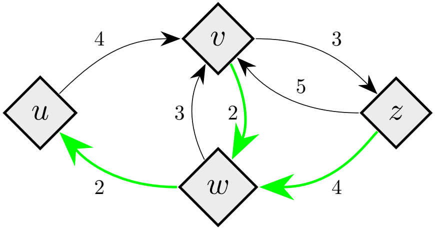


Disaggregating cycles

Chiu-Liu-Edmond's algorithm

Evolution & Social Science, Brisbane 2023

- Reversing the process, selecting "small" edges at each step, we construct a tree (no cycles, path from every vertex to a root vertex).
- The process we just followed is Chiu-Liu Edmond's algorithm for finding a tree with minimal sum of edge weights.

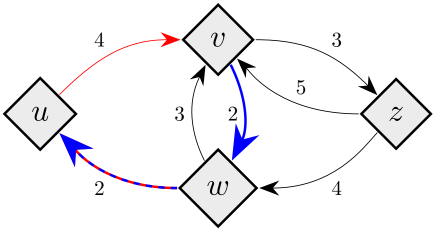


Theorem (tree characterization)

Stochastically stable states correspond to the roots of minimal spanning trees.

Tree surgery

- Tree surgery, where a minimal tree is conjectured to be rooted at some vertex v, before it is shown that an even lower cost tree can be constructed, usually by adding a path from v to some u and deleting surplus edges.
- Immediate corollary (radius-coradius theorem) is that if this is possible from any vertex to u, then u is stochastically stable.



- Take any tree rooted at v *(not illustrated)*.
- Add edges $v \to w \to u$. Adds $2 + 2 = 4$.
- Remove other edges exiting w, u.
  Removes at least $2 + 4 = 6$.
- So obtain tree rooted at u with lower sum of edge weights.
- Therefore v is not stochastically stable.

Next...

- The tree characterization and associated surgery is used directly in almost all existing applications of stochastic stability in economics.
- However, sometimes when we have behavioral rules and a target strategy profile with special properties, we can bypass direct calculations.

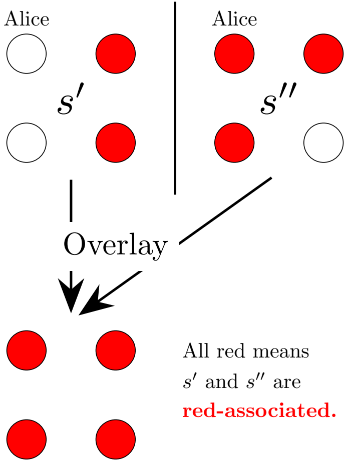
Asymmetry

The regularity from which classic results on risk dominance arise.

Asymmetry

Definition

- Alice is a member of a set of players.
- Each player has two strategies, red and transparent.
- Strategy profiles s' and s'' are red-associated if every player plays red at s' or s''.
- Alice is asymmetric towards red if, for red-associated s' and s'', her probability of choosing red given current profile s' is at least as great (to order of mag.) as her probability of choosing transparent given s''.



Asymmetry towards red implies

$$Prob(\text{Alice plays red} \mid s') \geq Prob(\text{Alice plays transparent} \mid s'')$$

Probability to order of magnitude, i.e. r value.

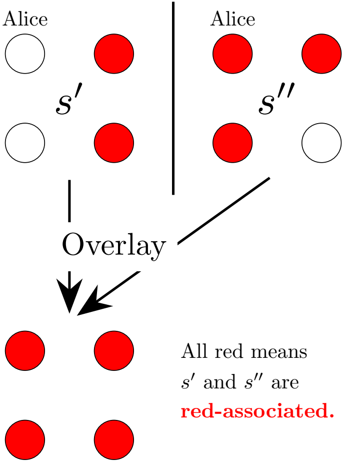
Asymmetry

Main theorem

Consider a process in which, every period some subset of players independently update their strategies given the current strategy profile. Assume that each player updates with positive probability.

Theorem (Asymmetry implies stability)

If every player is asymmetric towards red*, then* $s^* = (red, \ldots, red)$ *is stochastically stable.*



All red means s' and s'' are **red-associated.**

Asymmetry towards red implies

$$Prob(\text{Alice plays red} \mid s') \geq Prob(\text{Alice plays transparent} \mid s'')$$

Probability to order of magnitude, i.e. r value.

Evolution & Social Science, Brisbane 2023

Coordination on networks

- Finite set of players N.
- Each $i \in N$ plays strategy $s_i \in \{A, B\}$.
- Payoff of player i at profile s is

$$\Pi_i(s_i, s_{-i}) = \sum_{j \in N \setminus i} w_{ij}\, \pi(s_i, s_j).$$

- $w_{ij} \geq 0$ is the influence of player j on player i.

	A	B
A	$\pi(A, A)$	$\pi(A, B)$
B	$\pi(B, A)$	$\pi(B, B)$

Coordination game with A a risk dominant strategy.

$$\pi(A, A) - \pi(B, A) \geq \pi(B, B) - \pi(A, B) > 0$$

Behavioral rules

Based on payoff differences

- Payoff loss for i when changing strategy from s,

$$\Delta_i(s) = \Pi(s_i, s_{-i}) - \Pi(s_i', s_{-i}), \qquad s_i' \neq s_i.$$

Behavioral rules

Based on payoff differences

- Payoff loss for i when changing strategy from s,

$$\Delta_i(s) = \Pi(s_i, s_{-i}) - \Pi(s_i', s_{-i}), \qquad s_i' \neq s_i.$$

- Probability of player i changing strategy has order of magnitude $\varepsilon^{f_i(\Delta_i)}$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ is non-decreasing.

Evolution & Social Science, Brisbane 2023

Behavioral rules

Based on payoff differences

- Payoff loss for i when changing strategy from s,

$$\Delta_i(s) = \Pi(s_i, s_{-i}) - \Pi(s_i', s_{-i}), \qquad s_i' \neq s_i.$$

- Probability of player i changing strategy has order of magnitude $\varepsilon^{f_i(\Delta_i)}$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ is non-decreasing.
- That is, probability of changing strategy increases (decreases) in the payoff gain (loss) from doing so.

Behavioral rules

Based on payoff differences

- Payoff loss for i when changing strategy from s,

$$\Delta_i(s) = \Pi(s_i, s_{-i}) - \Pi(s_i', s_{-i}), \qquad s_i' \neq s_i.$$

- Probability of player i changing strategy has order of magnitude $\varepsilon^{f_i(\Delta_i)}$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ is non-decreasing.
- That is, probability of changing strategy increases (decreases) in the payoff gain (loss) from doing so.
- This includes rules such as

Behavioral rules

Based on payoff differences

- Payoff loss for i when changing strategy from s,

$$\Delta_i(s) = \Pi(s_i, s_{-i}) - \Pi(s_i', s_{-i}), \qquad s_i' \neq s_i.$$

- Probability of player i changing strategy has order of magnitude $\varepsilon^{f_i(\Delta_i)}$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ is non-decreasing.
- That is, probability of changing strategy increases (decreases) in the payoff gain (loss) from doing so.
- This includes rules such as
  - Best response with uniform mistakes. $f_i(x) = 0$ if $x \leq 0$, $f_i(x) = 1$ if $x > 0$.

Behavioral rules

Based on payoff differences

- Payoff loss for i when changing strategy from s,

$$\Delta_i(s) = \Pi(s_i, s_{-i}) - \Pi(s_i', s_{-i}), \qquad s_i' \neq s_i.$$

- Probability of player i changing strategy has order of magnitude $\varepsilon^{f_i(\Delta_i)}$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ is non-decreasing.
- That is, probability of changing strategy increases (decreases) in the payoff gain (loss) from doing so.
- This includes rules such as
  - Best response with uniform mistakes. $f_i(x) = 0$ if $x \leq 0$, $f_i(x) = 1$ if $x > 0$.
  - Logit choice. $f_i(x) = 0$ if $x \leq 0$, $f_i(x) = x$ if $x > 0$.

Evolution & Social Science, Brisbane 2023

Coordination on networks

P-D based rules and RD implies asymmetry

Proposition

If player i follows a payoff-difference based rule and A is risk dominant, then i is asymmetric towards A.

Corollary

If each player follows a payoff-difference based rule and A is risk dominant, then $(A, \ldots, A)$ is stochastically stable.

Note that the Corollary applies even if players follow different payoff-difference based rules.

	A	B
A	$\pi(A, A)$	$\pi(A, B)$
B	$\pi(B, A)$	$\pi(B, B)$

Coordination game with A a risk dominant strategy.

$$\pi(A, A) - \pi(B, A) \geq \pi(B, B) - \pi(A, B) > 0$$

Evolution & Social Science, Brisbane 2023

Before we go, let's note

- We have described asymmetry at the level of the individual. It can also be defined for coalitions and for the process as a whole.
- Asymmetry can be used for many behavioral rules.
- For example, imitative rules, coalitional rules, altruistic choice, k-level thinking, strange preferences.
- Can sometimes prove asymmetry for subsets of players, then take these players' actions for granted in solving the remainder of the problem.

newton@kier.kyoto-u.ac.jp

For references, see reading list.