Stable Equilibria in Sampling Best Response Dynamics

Undergraduate Project Report

Arqam Patel

November 2023

1 Acknowledgements

I would like to thank Dr Srinivas Arigapudi and Dr Subhajit Dutta for their guidance for the project.

2 Introduction to game theory

Game theory is a branch of mathematics that studies the strategic interactions among rational agents. In its simplest form, a game is defined by its players, actions, and payoffs. Evolutionary game theory is an extension of classical game theory that seeks to model how the strategic interactions among agents evolve through a series of games over time.

2.1 Players

In a game, the entities making decisions are referred to as players. A player can represent an individual, a group, or even an inanimate object like a computer program. We consider games consisting of a finite set of n players, denoted by $N = \{1, 2, \dots, n\}$.

2.2 Actions

Each player has a set of actions they can take. These are the basic decisions that a player can make, often represented as $A_i = \{s_1, s_2, \dots s_m\}$ for player i. The actions of the players collectively determine the outcome.
In a game involving n players, an *action profile* is a tuple $(a_1, a_2, \dots, a_n)$, where $a_i \in A_i$ represents the action taken by player i. An action profile thus specifies a particular combination of actions chosen by a set of players in the game at a given instance. Since they are indistinguishable in our context, we use the words actions and strategies interchangeably.

2.3 Outcomes

An outcome refers to a specific end-state that results from a particular combination of actions taken by all players in the game. It can be represented as a point in the set of all possible outcomes, typically denoted by O. In deterministic games, which are the subject of our study, the outcome is uniquely determined by the action profile of the players.
The set O can thus be constructed as the Cartesian product of all players' action spaces. For a game involving n players with action spaces $A_1, A_2, \dots, A_n$, the set of all possible outcomes O can be defined as:

$$O = A_1 \times A_2 \times \dots \times A_n$$

Each outcome $o \in O$ is a tuple $(a_1, a_2, \dots, a_n)$, where a_i is the action taken by player i.

2.4 Preferences

Given any two outcomes, a player can always discern which is better (or if the two are identical) in terms of utility. Preferences in game theory are formally represented by a relation over the set of outcomes O, which captures how a player ranks different outcomes. Given any two outcomes $a, b \in O$, the following can be true:

- $a \succ b$ means the player strictly prefers action a over b.
- $a \sim b$ means the player is indifferent between a and b.
- $b \succ a$ means the player strictly prefers action b over a.

We assume these preferences to be transitive, i.e., if $a \succ b$ and $b \succ c$, then $a \succ c$.

2.5 Payoffs

Payoffs represent a measure for the outcomes of a game for each player, allowing us to represent preferences succinctly. Once an outcome is realized, each player receives the payoff, which quantifies the utility received by that player. The payoff functions $u_i : O \to \mathbb{R}$ map outcomes to real numbers, specifying the payoff for each player i under each possible outcome. Formally, the payoff function for player i and outcome o can be written as:

$$u_i(o) = u_i(a_1, a_2, \ldots, a_n)$$

For two player games, we use the notation $v_i(s_j, s_k)$, which represents the payoff to player i for following strategy j upon encountering a player following strategy j.

2.6 Best Response Functions

The best response of player i is the action that maximises the payoff function, given that the other players do not change their actions. The Best Response function $B_i : A_{-i} \to \mathcal{P}(A_i)$ for player i maps each action profile a_{-i} of the other players to a set of player i's best responses. It is formally defined as:

$$B_i(a_{-i}) = \arg\max_{s_j \in A_i} u_i(s_j, a_{-i})$$

3 Problem definition

In game theory, a "dynamic" refers to a model or process that describes how the game evolves over time. Dynamics introduce a temporal element, allowing us to study how players might adjust their strategies in response to the actions of others, how strategies evolve, or how equilibria are reached or disrupted over time.

We consider a population N of n identical players $\{1, 2, \ldots n\}$, each having identical action spaces $A_i = \{s_1, s_2 \ldots s_m\}$, playing a sequence of "games" in discrete time steps. The players are identical, hence the two-player interaction payoffs $v_i(s_j, s_k) = v(s_j, s_k)$ are constant for all players $i \in N$.

3.1 Best Response Dynamic

In a **best response dynamic**, each player decides its action for the next iteration of the game based on the observed action profile of all other players. Since it can observe every other player's action, each agent can compute the proportions of players following specific actions to determine the payoff values and thus decide the action it takes in the next time step.

We define the payoff of player i for an action profile as the expected payoff of a randomly sampled two-player interaction involving i.

$$u_i(a_1, a_2, \ldots, a_n) = \mathbb{E}[v(a_i, a_j)]$$

where j is sampled uniformly randomly from $\{1, 2, \ldots, n\} \setminus \{i\}$.

We observe that the payoffs are uniquely determined by the distribution - sampling probabilities - and thus the proportions of the various actions in the population of interest. Hence, only the proportions of the various strategies are relevant for computing strategy payoffs. Considering player i adopts strategy a_i and proportions ($p_{i,h}$ of all players except i adopt s_h, we define a new payoff function mapping a tuple of action and the various proportions to payoffs:

$$U_i(a_i, p_{i,1}, p_{i,2} \ldots p_{i,m}) = \sum_{h=1}^{m} p_{i,h} \cdot v(a_i, s_h)$$

Equivalently, $\mathbf{U}_i = \mathbf{V}\mathbf{p}_i$ where, $\mathbf{p}_i$ is the action profile vector $(m \times 1)$ specifying the proportion $p_{i,h}$ of each strategy s_h in the population of players excluding i, $\mathbf{U}_i$ is the payoff vector $(m \times 1)$ specifying the payoff $U_{i,h}$ of each action s_h given the action profile, and $\mathbf{V}$ is the two player payoff matrix $(m \times m)$, $v_{j,h} = v(s_j, s_h)$

We can also define a reparametrized version of the best response function in this case:

$$B'_i(p_{i,1}, p_{i,2} \dots p_{i,m}) = \arg\max_{s_j \in A_i} U_i(s_j, p_{i,1}, p_{i,2} \dots p_{i,m})$$

3.2 Sampling Best Response Dynamic

This dynamic attempts to replicate the information uncertainty that real life agents operate under. For each player i, we draw a random uniform sample (without replacement) $T = \{t_1, t_2, \dots t_{k_i}\}$ of size $k_i \geq 2$ from the set of players except itself $\{1, 2, \dots n\} \setminus i$. We then use the sample proportions as the estimates for the proportions of the population following each strategy:

$$\hat{p}_{i,h} = \frac{1}{k_i} \sum_{g=1}^{k_i} \mathbb{I}(a_{t_g} = s_h)$$

We then adopt a similar approach to the best response dynamic described above, substituting the true proportions with the sample estimates.

In the simplest case, **k-sampling best response dynamic**, we assume that the sampling parameter k_i is the same for each player.

However, the subject of our study is the $\mathbf{k_1}$ $\mathbf{k_2}$ **sampling best response dynamic**, where we have two cohorts in the population, with uniform sampling parameters k_1 and k_2 respectively. This is the simplest case of a population with non-homogeneous sampling parameters, and thus provides a starting point for studying the behaviour of such systems.

3.3 Asymptotically Stable Equilibrium

Let $x_{j,t}$ be the proportion of players in the population adopting strategy s_j at time (or time step) t. A state is said to be an asymptotically stable equilibrium if $\exists x_1, x_2, ...x_m$ with $\sum x_j = 1$ such that:

$$\lim_{t \to \infty} x_{j,t} = x_j \forall j$$

4 Literature review

Ritzberger and Weibull (1995) study the evolution of an infinite population of players in continuous time under various dynamics. Their main result postulates that pure equilibria (where a single strategy is adopted by the whole population) constitute the only stable equilibria under a large variety of games.
Sandholm (2001) studies the k-sampling best response dynamic in coordination games, where players sampled from the population get opportunities to update their strategies by sampling the behavior of k opponents and playing a best response to that sample. The paper shows that if the game has a *p-dominant* strategy played initially by a fraction of the population, and the population size is large enough, play converges to the an equilibrium of the p-dominant strategy with high probability. Specifically, if the proportion of players adopting s_i exceeds $\frac{1}{k}$, this proportion, after a large number of games, converges in probability to 1. This holds for any $k \geq 2$. The proof involves analyzing a stochastic process (Markov chain) for the number of players not currently playing the dominant strategy.

5 Methodology

5.1 Finite population, discrete time space case

Here, we attempt to model the $\mathbf{k_1}$ $\mathbf{k_2}$ **sampling best response dynamic** for coordiation games. We consider a population of n players, where αn players have sampling parameter k_1, while $(1-\alpha)n$ players have parameter k_2. For all players, we consider identical action spaces of two strategies $A_i = \{s_1, s_2\}$. We consider coordination games, where players have a positive payoff for facing a player with the same strategy, and zero for facing a player with a different strategy. Thus, $v(s_1, s_2) = v(s_2, s_1) = 0$. $v(s_1, s_1)$ and $v(s_2, s_2)$ are assumed to be some positive constants a and b respectively. The payoff matrix is thus:

$$V_i = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

Since we have only two possible actions, the action profile of players except i, a_{-i} can be parametrized by the proportion p_i of players except i adopting action s_1 (we will simply have $(1 - p_i)$ proportion of players with action s_2).

$$U_i(a_i, p_i) = p_i \cdot v(a_i, s_1) + (1 - p_i) \cdot v(a_i, s_2)$$

We substitute the true proportion p_i with the sample proportion $\hat{p}_i$. We can thus define the *approximate* best response function in this case using $\hat{p}_i$:

$$B_i'(\hat{p}_i) = \begin{cases} \{s_1\} & \text{if } \hat{p}_i \cdot a > (1 - \hat{p}_i) \cdot b \\ \{s_1, s_2\} & \text{if } \hat{p}_i \cdot a = (1 - \hat{p}_i) \cdot b \\ \{s_2\} & \text{if } \hat{p}_i \cdot a < (1 - \hat{p}_i) \cdot b \end{cases}$$

The pseudocode for modelling the dynamics over time is as follows:

Algorithm 1 Finite population, discrete time simulation

```
Initialize population of num_agents agents
for i = 1 to n_steps do
  for all agents in population do
    Sample k1 or k2 other players from population (according to the agent's initialisation)
    Let n1 be number of pl from the sample following action 1
    p̂1 = n1 / sample_size
    p̂2 = (sample_size − n1) / sample_size
    Û1 = p̂1 ∗ a
    Û2 = p̂2 ∗ b
    Update strategy based on which estimated payoff Ûi is higher
  end for
  Calculate and print proportion of strategy 1
end for
```

We have created a web-based application for visual inspection and review of the evolution of the state. The source code is available on our Github repository.

5.2 Infinite population, continuous time space case

Let us first examine the simple sampling best response dynamic. Assume that we have infinitely many players interacting over inifinitesimal time steps. We have a vector $\mathbf{x} = (x_1, x_2, ...x_m)$ where x_j is the proportion of players following strategy s_j over time. Let $\mathbf{u} = (u_1, u_2, ...u_m)$ be the payoff vector (diagonal of the payoff matrix) where $u_h = v(s_h, s_h)$; since we consider only coordination games, $v(s_h, s_j) = 0$ when $j \neq h$.

We have a function $w_j(\mathbf{x}, \mathbf{u}, k)$ which represents the expected proportion of players following s_j at the next infinitesimally incremented time step. We can represent the rate of change as the difference of this with the previous proportion, giving us the differential equation:

$$\frac{dx_j}{dt} = w_j(\mathbf{x}, \mathbf{u}, k) - x_j$$

Suppose, X_j represents the numbers of agents following strategy s_j from amongst a sample of k players randomly sampled from the population. Then, $\mathbf{X} = (X_1, X_2, ...X_m)$ is a random vector from a multinomial distribution with k trials and with event probabilities $x_1, x_2, ...x_m$. Let $\hat{p}_h = \frac{X_h}{k}$, representing the sample estimate of the proportion of s_j in the population.

$$
\begin{aligned}
w_j(\mathbf{x}, \mathbf{u}, k) &= \mathbb{E}[\mathbb{I}(s_j \text{ is the sampling best response})] \\
&= \mathbb{P}(s_j \text{ is the sampling best response}) \\
&= \sum \mathbb{P}(s_j \text{ is the sample best response for a given } \mathbf{X}) \\
&= \sum_{X_1+X_2+\ldots+X_m=k} \frac{k!}{X_1!X_2!\ldots X_m!}(\prod_{j=1}^{m}(x_j^{X_j})) \cdot \mathbb{I}\,(s_j \text{ is the best response given } \mathbf{X}) \\
&= \sum_{X_1+X_2+\ldots+X_m=k} \frac{k!}{X_1!X_2!\ldots X_m!}(\prod_{j=1}^{m}(x_j^{X_j})) \cdot \mathbb{I}(j = \underset{h\in\{1,2,..m\}}{\arg\max}\ (\hat{p}_h u_h))
\end{aligned} \tag{1}
$$

For k_1 k_2 sampling,

$$\frac{dx_j}{dt} = W_j(\mathbf{x}, \mathbf{u}, k_1, k_2, \alpha) - x_j$$

where:

$$W_j(\mathbf{x}, \mathbf{u}, k_1, k_2, \alpha) = \alpha * w_j(\mathbf{x}, \mathbf{u}, k_1) + (1-\alpha) * w_j(\mathbf{x}, \mathbf{u}, k_2)$$

We then use this system of ordinary differential equations to model the trajectory for using computational methods, specifically the **odeint** module from the Python library **scipy**.

For our study, we assume $m = 3$ possible actions. The pseudocode for modelling the dynamics over time is as follows:

Algorithm 2 Infinite population, continuous time simulation

Input: Sampling sizes k_1, k_2, proportion α, payoffs u_1, u_2, u_3

define function $w(x_1, x_2, k)$ to calculate the transition probabilities:
Initialise an array, $Y = (Y_1, Y_2, Y_3)$.
for $i = 0$ to k **do**
  for $j = 0$ to $k - i$ **do**
    Calculate multinomial coefficient $c = \frac{k!}{j!\cdot i!\cdot(k-i-j)!}$
    Calculate sample proportions: $\hat{p}_1 = \frac{i}{k}$ $\hat{p}_2 = \frac{j}{k}$ $\hat{p}_3 = \frac{k-i-j}{k}$
    if $\hat{p}_1 \cdot u_1 > \hat{p}_2 \cdot u_2$ and $\hat{p}_1 \cdot u_1 > \hat{p}_3 \cdot u_3$ **then**
      Increment Y_1 by $c \times (x_1)^i \times (x_2)^j \times (1 - x_1 - x_2)^{k-i-j}$
    else if $\hat{p}_2 \cdot u_2 > \hat{p}_1 \cdot u_1$ and $\hat{p}_2 \cdot u_2 > \hat{p}_3 \cdot u_3$ **then**
      Increment Y_2 by $c \times (x_1)^i \times (x_2)^j \times (1 - x_1 - x_2)^{k-i-j}$
    else
      Increment Y_3 by $c \times (x_1)^i \times (x_2)^j \times (1 - x_1 - x_2)^{k-i-j}$
    end if
  end for
end for
return Y
end function

define function $W(x_1, x_2, k_1, k_2, \alpha) = \alpha \cdot w(x_1, x_2, k_1) + (1-\alpha) \cdot W(x_1, x_2, k_2)$

define system of differential equations $system(x1, x2, t)$:
$\frac{dx_1}{dt} = W(x_1, x_2)[0] - x_1$
$\frac{dx_2}{dt} = W(x_1, x_2)[1] - x_2$

Generate a grid of points within the simplex
Initialize a list to store final points of trajectories
Compute the trajectories of the dynamic starting from the given initial points for a given amount of time.
Perform clustering on final points to identify stable points
Average points within each cluster to determine stable points

---

For this model as well, we have created and hosted an interface for visual inspection and review of the evolution of the state. The source code is available on our Github repository.

6 Results

After iterating through various parameters, we find several examples of mixed equilibria (where two or more strategies coexist) in coordination games. This reproduces and verifies the previous experimental and theoretical results of Dr Arigapudi's research, and adds context and nuance to previous research which has dealt with the prevalence of pure equilibria (where only one strategy survives) in sampling best response dynamics. We present one example each of the discrete and continuous cases.
In the discrete case, we initialise the simulation at $n = 10^4, k_1 = 3, k_2 = 500, \alpha = 0.4, a = 2.4, b = 1, (x_1)_{t=0} = 0.1$ and run it for 24 time steps. The dynamic stabilises around (0.177, 0.823).

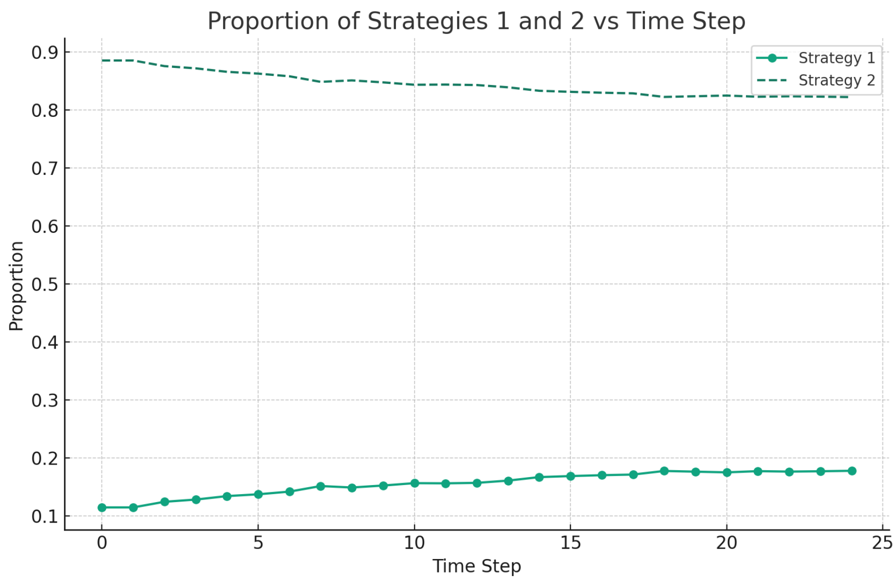

Figure 1: The trajectory of the dynamic

In the continuous case, we take $k_1 = 3, k_2 = 20, \alpha = 0.4, u_1 = 2.5, u_2 = 1, u_3 = 0.4$ and run it for 6 time units. We can observe that the system of differential equations does not converge towards the origin (0,0,1), (0,1,0) or (1,0,0) (unless it is initialised there). Instead, the equilibria are (0.79, 0, 0.21) and (0, 0.79, 0.21). This demonstrates the existence of a mixed equilibria within a continuous time, infinite population model of k_1k_2 sampling best response dynamics.

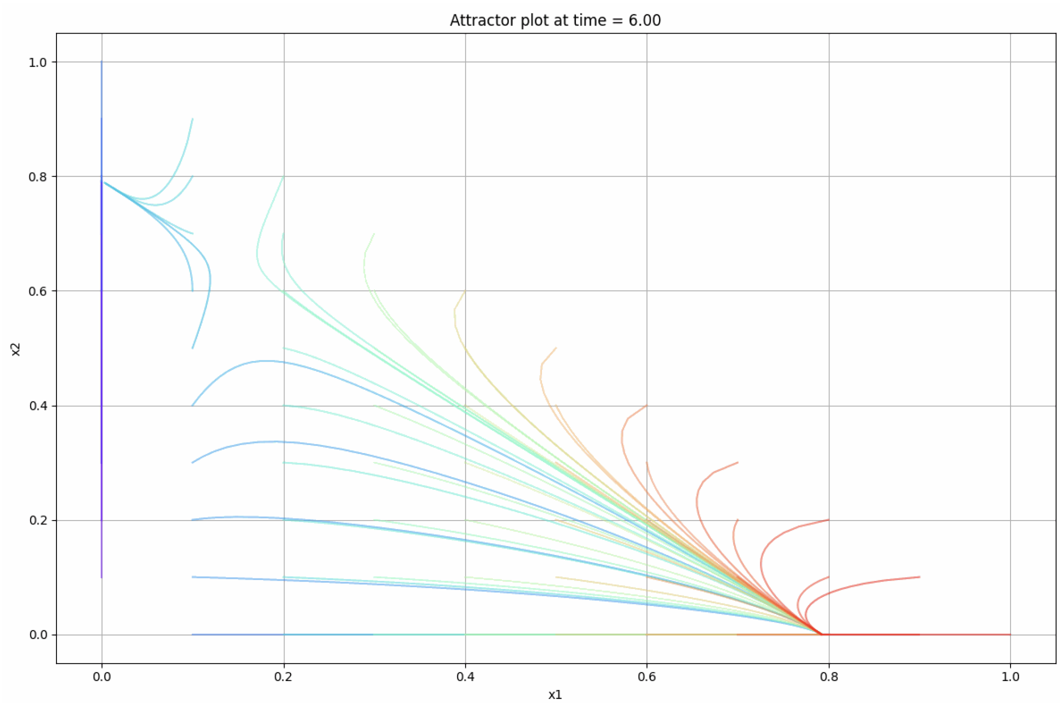

Figure 2: The trajectories of the dynamic from set of initial points.

7 References

1. Sandholm, William H. Almost global convergence to p-dominant equilibrium. Game Theory 30, 107–116 (2001). https://doi.org/10.1007/s001820100067

2. Ritzberger, Klaus, and Jörgen W. Weibull. "Evolutionary Selection in Normal-Form Games." Econometrica, vol. 63, no. 6, 1995, pp. 1371–99. JSTOR, https://doi.org/10.2307/2171774. Accessed 2 Nov. 2023.

3. Oyama, Daisuke, William H. Sandholm, and Olivier Tercieux. "Sampling Best Response Dynamics and Deterministic Equilibrium Selection." Theoretical Economics, vol. 10, no. 1, 2015, pp. 243–281. https://doi.org/10.1555-7561/20150243. Accessed 3 Nov. 2023.