Learning within a Markovian Environment[*]

Javier Rivas[†]
European University Institute

May 30, 2007

Abstract

We investigate the behavior of two learning rules, Stochastic Best Response (SBR) and Replicator Dynamics, in a model with aggregate and time-correlated shocks to payoffs. The main difference between the two behavior of the two rules is that under SBR corners are not absorbing. We study a setting where there are two actions and many states of nature and the transition between states follows a Markov chain. We find that the SBR converges to a behavior similar to probability matching. On the other hand, the Replicator Dynamics selects the optimal action only if the average payoff of both actions is different enough.

JEL Classification Number: C73.

Keywords: Adaptive Learning, Stochastic Best Response, Replicator Dynamics, Markov Chains, Probability Matching.

1 Introduction

Despite the many applications Markov processes have for modeling real life environments, few papers are devoted to study the evolutionary properties of models with such processes. To our knowledge, only Ben-Porath et al. (1993) and Rustichini (1999) deal with this issue.

Ben-Porath et al. (1993) present a model that is framed within an environment that changes. They study two types of changing environment. One in which the change is deterministic and another in which the changes in environment follow a Markov chain. In their model, players actions are subject to random mutations. They characterize the mutation rate that maximizes population growth in the long run. On the other hand, Rustichini (1999) presents a paper that focuses on the optimality of two different population dynamics within a Markovian environment.

[*] Part of this work was written during my visit to the University of Wisconsin-Madison. I'm grateful to the faculty at UW and the participants in the Theory Lunch. I would like to thank Karl Schlag and Larry Samuelson for useful discussions and comments. I would also like to thank the participants at the Micro Workshop at EUI and Mark Le Quement for useful comments. All remaining errors are mine.

[†] javier.rivas@eui.eu. Department of Economics, European University Institute. Via della Piazzuola 43, 50133 Florence (Italy). http://www.eui.eu/Personal/Researchers/JavierRivas/index.htm.

In his model, the environment changes according to a Markov chain and for any state in the chain there is a unique action that maximizes payoff. Rustichini studies the optimality properties of linear and exponential (logit) adjustment process under different informational settings about payoffs of actions. In section 5.1 we dedicate more time to discussing Rustichini's work.

In this paper we present a model where the environment changes according to a Markov chain and for any state in the chain there is only one action that maximizes payoff. Players do not know how does the payoff matrix look like. What is more, they don't know how many states there are or what are the transition probabilities between states. To motivate the idea behind this setting, consider the following practical example. Assume there is a firm that faces a random demand. The demand follows a Markov process with m states and the firm does not know the value of m. The probability of transiting from state i to state j is given by $\theta_{ij} \in (0, 1]$. We can think of this demand as a demand that follows fashions. In some states the product 1 is on fashion and most of the costumers buy product 1. Of much better product sells will depend of the specific state of the demand. On the other hand, in the rest of states product 2 is on fashion most of the costumers buy product 2. Assume that the profit for the firm of producing good i when state is j equals π_{ij}. Assume for simplicity that the cost of producing each good is zero. However, the firm has a fixed productive capability of measure 1. This means that the amounts of good 1 and good 2 produced must always add up to 1 or less. The firms' problem is to choose how much of either good to produce. However, the firm does not know how many states nature has. Neither she knows what are the values of π_{ij} and θ_{ij} for all $i \in \{1, 2\}$ and $j \in \{1, \ldots, m\}$. The firm just observers that his profits from selling either product change over time. In this paper we show how the two different learning rules we consider will behave in such setting.

Within the setting we just presented, we study the behavior of two learning rules: Stochastic Best Response and Replicator Dynamics. The two rules are very well known and have been widely used in the literature because of their appealing interpretation and their match with real life behavior. Intuitively, the main difference between the two rules is that the Stochastic Best Response is an individual leaning rule (or non-selection rule) while the Replicator Dynamics is a population learning rule (or selection rule). The implications of this differences for our analysis will be clear later on. With this paper we aim at exploring how the two rules we consider behave in a setting with markovian changes in the state of nature. In particular, the question we want to address is the following: Are the rules considered here able to select, in the long run, the best action?

The Stochastic Best Reply is an individual learning rule because players do not learn form their neighbors but from their own experiences. It is a reinforcement rule in that more successful actions today are likely to be adapted for tomorrow. Under the Stochastic Best Reply, at any period players observe the performance of both actions. A player increases the probability of playing a given action for the next period if and only if that action yielded higher payoff than the others in the current period. The increase in the probability of playing the action will depend in the difference between the payoffs achieved by all actions and in the current probability of playing the action. In particular, the Stochastic Replicator Dynamics incorporates opportunity cost considerations. If action 1 yields more payoff today and the player is already playing it with high probability, the increase in the probability of playing action 1 is small because the players feels she is already doing quite well. That is, the opportunity cost of not increasing the probability of playing action 1 is low. However, if action 1 yields more payoff today than the other actions but the player is playing it with very low probability, then the increase in the probability of playing action 1 is big because the player feels she is doing quite badly. That is, the opportunity cost of not increasing the probability of playing action 1 is high. The opportunity cost consideration has its more extreme manifestation in the case where the player use the deterministic Best Response. Under the deterministic Best Response the player plays tomorrow with probability 1 the action that yielded higher payoff today. The deterministic Best Response is just the degenerated case of the Stochastic Best Responses.

Under the Replicator Dynamics, players learn by observing actions and payoffs of other players. Every period, each player observes the average payoff of the population. If the payoff of a player is higher than the average payoff of the population, then she will increase the probability of playing the action she played. The magnitude of this increase will depend on the difference between her payoff and the average payoff of the population.

Learning rules can be divided into two categories. The individual (or non-selection) learning rules and the population (or selection) learning rules. Concerning the individual learning rules, we focus our attention to the Stochastic Best Response. The Stochastic Best Response gathers many of the aspects that real life agents' decision exhibit. Mainly, it captures the idea of reinforcement learning. That is, actions that were more successful today are more likely to adapted for tomorrow. For two excellent detailed expositions on reinforcement learning and its relationship with real life behavior the reader is referred to Roth and Erev (1995) and Erev and Roth (1998) and Camerer and Ho (1999).

Concerning the population learning rules, we decided to use the Replicator Dynamics. The Replicator Dynamics is the most commonly used learning process for population learning in economics and in biology. The literature on selection learning rules is extensive. Other representative rules may include various types of imitation, etc. However, it has been shown that the behavior of many of these rules converges to that of the replicator dynamics. For example, Schlag (1998) shows that the Proportional Imitation Rule converges to replicator dynamics. Moreover the Replicator Dynamics can be derived from behavioral axioms as shown by Easley and Rustichini (1999). Hence, by focusing on the Replicator Dynamics we are indirectly considering other population learning rules.

In our results, we show the following two facts. First, when the agents play according to the Stochastic Best Response, the long run behavior of the population resembles to what is know as probability matching. Probability matching prescribes that if an action is better than the other x percent of the time, then the population will play it with a frequency of x. We show that if an action is on average better than the other, then the population will be playing it more often, but not always. The frequency by which the each action is played will depend on two things. First, the difference in average payoffs between the two actions. Second, on the probabilities that the limiting distribution of the Markov chain for states puts on each state. The probability matching behavior is clearly suboptimal. While some experimental papers report that this behavior is observed in real life (see, for example, Rubinstein (2002), Siegel and Goldstein (1959)). There seem not to be consensus whether probability matching is in fact present in the behavior of real life agents (see, for instance, Vulkan (2000) and Shanks et al. (2002)).

Our results for the Stochastic Best Response are closely related by the findings by Kosfeld et. al. (2002). They study a setting where a finite set of players repeatedly play a normal-form game. Players adapt their strategies by increasing the probability of playing a certain action only is this action is a best reply to the actions played by the other agents. Hence, the rule they use a particular case of the Stochastic Best Response in which the magnitude of the payoffs is irrelevant for the updating of strategies. Kosfeld et. al. (2002) find that the system converges to a best-reply matching equilibrium. In a best-reply matching equilibrium each player plays an action with a probability that is equal to the probability that this action is a best response to the actions of the other players. Our setting is different from theirs in that players do not play against other players but against nature. The probability matching behavior we find in games against nature is the equivalent to the best-reply matching equilibrium Kosfeld et. al. (2002) find. In Section 5.3 we discuss this issues in more depth.

In our second result, when players play accordingly to Replicator Dynamics, we show that the population may end up playing a suboptimal action forever. On the other hand, if the difference between the average payoff of the two actions is high enough, then the population will for sure end up playing the action that has higher average payoff. That is, if the differences in payoffs are not too high, the population may end up playing either action in the long run. Hence, the system may exhibit lock out on a suboptimal action.

A rule that has recently attracted some attention is the Börgers et. al. (2004) best monotone learning rule (BMS rule henceforth). Under the BMS rule, if the payoff achieved by the action played is higher than an endogenous aspiration level, then the probability of playing the action increases. How this aspiration level is formed depends on the initial probability of playing each action. A rule is defined to be monotone if the expected probability of playing the action that is best given today's state increases. A rule is said the be the best monotone rule if the expected increase in playing the best action form period to another is highest among all monotone rules. Börgers et. al. (2004) derive a rule that is the best monotone individual learning rule when the environment changes independently of its past values and foregone payoffs are not observed. We show that the BMS rule has a very similar behavior in the long run that the Replicator Dynamics. Whether BMS rule is more likely to select the best action in the long run than the Replicator Dynamics will depend on the specific value of the parameters of the model. A deeper analysis of BMS rule is presented in Section 5.2.

The contribution to the literature of our work is twofold. First, we have mentioned already the fact that very few papers study the situation in which the future realization of the state of nature depends on its past realizations. Most of the papers on learning consider either that the environment doesn't change or that it changes independently of past realizations. This is probably due to the technical difficulties involved in dealing with correlated realizations of states. In this paper we show how this difficulties can, at least partially, be overcome. Our proofs for the result for the Stochastic Best Response show how one can deal with dependent randomness by showing that for any possible realization of states of nature, the position of the system in the future can be approximated by simply the differences in speed of convergence towards each action.

The proof of our result for the Replicator Dynamics learning extends the result on Ellison and Fudemberg (1995) to the case of dependent state realizations. We show that the behavior of a system that evolves according to a Markov Chain can be approximated by the behavior of a system in which the probability of each state occurring is independent and equal to the limiting distribution of the Markov Chain. Hence, our first contribution to the literature is that we introduce new techniques for dealing with correlated states of nature.

Our second contribution to the literature is of an informative nature. The two rules we are considering here have been widely studied before. However, to our knowledge, there is no paper that shows of these rules will perform in a setting with correlated states. While our conclusion for the Replicator Dynamics may not seem very unexpected, we find the fact that Stochastic Best Response yields probability matching behavior quite surprising.

The rest of the paper is organized as follows. Section 2 presents the model. The two learning rules considered are introduced in Section 3. Results are developed in Section 4. Section 5 presents a discussion and a deeper comparison of our work with the existing literature. Finally, Section 6 concludes.

2 The Model

Consider a continuum of identical players of measure 1. Every period $t = 1, 2, \ldots$ players in the population have to choose between action 1 or action 2[1]. The payoff of each player at time t depends on her action and on the value of a random variable s_t. The variable s_t follows a Markov process P with states m states. The probability of transiting from state i to state j is given by $\theta_{ij} \in [0, 1]$. We assume the Markov chain to be ergodic. Hence, if $\theta_{ij} = 0$ for some i, j then there exists a sequences of states $k_1, k_2, \ldots, k_n$ such that $\theta_{ik_1}, \theta_{k_1,k_2}, \ldots, \theta_{k_n,j} \neq 0$. We define $\mu \in [0, 1]^m$ as the limiting distribution of the Markov Chain P where μ_i is the weight the limit distribution puts in state i.

If a player chooses action i and the state equals j then she gets a payoff π_{ij}. We assume that $\pi_{ij} \in [0, 1]$ for all $i, j \in \{1, 2\}$. Further we assume that there is no weakly dominant action. That is, there exists no $i \in \{1, 2\}$ such that $\pi_{ij} \geq \pi_{-ij}$ for all $j \in \{1, \ldots, m\}$. Without loss of generality we assume that for some $h < m$, $\pi_{1j} \geq \pi_{2j}$ for $j \leq h$ and $\pi_{2j} > \pi_{1j}$ for $j > h$. That

[1] Increasing the number of actions will make computations and exposition less transparent and won't add any extra strategic value to the case with 2 actions.

is, in the first h states action 1 yields at least the same payoff as action 2. In the rest of state, action 2 yields more payoff than action 1.

One could think of adding idiosyncratic perturbations to payoffs by adding ε_{ht} to each π_{ij}. Where ε_{ht} are normally distributed zero mean random variables that are independent across players h and time t. However, since the Stochastic Best Response can treat payoffs in a non linear way (exponential, square, etc.) it is not true that the process will converge to the same value as for the case without noise. The reason is the same as why, for instance, $E(x^2) \neq E((x+\varepsilon))$ with $E(\varepsilon) = 0$. However, it can be easily verified that adding noise makes no change in the results for Replicator Dynamics and the Stochastic Best Reply when payoffs enter linearly in the learning rule.

Players have very limited information about the game they are playing. They don't know the payoff matrix nor how nature evolves or how many states there are. The only thing players know is that they have two actions at their disposal. Furthermore, players have limited memory. In particular, the only information they keep from period to period is the current strategy they are playing. Next we explain this in more detail.

The timing within each time period works as follows. First, players choose actions according to their strategies. Then, nature decides the state. Third, payoff are realized and players observe their payoff. At this stage, only under the Stochastic Best Response, players also observe foregone payoffs[2]. Finally, players updates their strategies. A strategy is the probability of playing each action. When updating their strategies, players use the following information: their strategy at the beginning of the period, the action they played and the payoff they got and, under the Stochastic Best Response, the payoff the other action would have yielded (foregone payoffs). Hence, the only information that players carry over to the next period is the strategies they have. That is, players have limited memory about payoffs or the actions they played in the past. The only information that players have about past payoff realizations is via their current strategy. Hence, their current strategy can be seen as an aggregation or summary of past payoff experiences.

Given that the two learning rules we consider only take into account the present realization of payoffs, players having full memory won't change our results. One might argue that if players had long memory they could observe different payoff realizations and easily have a significant information about the world they are living in (transition probabilities, etc.). However, this is not true since players do not know how many states nature has. The number of states of nature can be huge and players do not know how many states there are. Thus, players having long, but limited, memory won't guarantee them to learn about the environment they are living in. Hence, the question if players are able to select the best action in the long run is still open. Moreover, experimental results show that players do not tend to have very long memory (see, for example Erev and Roth (1998)). Furthermore, as showed by Rustichini (1999), even if players had infinite memory, it is not true that they will learn the best action for sure. This issue is

[2] Whether players observe forgone payoff under the Replicator Dynamics case or not is completely irrelevant as will be clear later on

further discuss in section 5.

A learning rule is a function b such that $b : [0,1] \times \{1,2\}^2 \times [0,1]^2 \to [0,1]$. That is, a function that maps three arguments, strategy for the present period, action played and payoff gotten and action not played and foregone payoff, into the strategy from the next period. The functional form of the function b will depend on the specific learning rule we consider.

Denote by z the probability that a player from the population plays action 2. The variable z_t refers to the value of z at the end of period $t-1$ and the beginning of period t. Note that given our setting, the variable z is a continuous Makorv process on $[0,1]$ and this process is ergodic.

Since we are dealing with a continuum of population, Law of Large Numbers applies and we have that z is also the fraction of players playing action 2 deterministically. In an abuse of notation, throughout the paper we will refer to z as both the probability for a single player of playing action 2 and the fraction of the population playing action 2.

3 The Learning Rules

3.1 Stochastic Best Response

To save notation we simply write $z_{t+1,i,j}$ to denote the value of z_{t+1} given that action i yielded a higher payoff at the current period and the current state of nature is j. The Stochastic Best Response is characterized by

$$\begin{aligned} z_{t+1,1,j} &= z_t + z_t \mu f(\Pi_j) \\ z_{t+1,2,j} &= z_t + (1-z_t)\mu f(\Pi_j) \end{aligned}$$

where $\mu \in [0, \hat{\mu}]$ is a learning speed parameter. The use of the learning speed parameter will be discussed later in the paper. The value of $\hat{\mu}$ is set such that z_{t+1} remains always between 0 and 1. Π_j are the payoff gotten and the forgone payoff of both actions at state j and f is a function on the payoffs. Function f must satisfy a the following assumption.

Assumption 1.

$$\begin{aligned} f(\Pi_j) &\leq 0 \textit{ for } j \in \{1,\dots,h\} \textit{ with strict inequality if and only if } 0 < z_t \leq 1, \\ f(\Pi_j) &\geq 0 \textit{ for } j \in \{h+1,\dots,m\} \textit{ with strict inequality if and only if } 0 \leq z_t < 1. \end{aligned}$$

As an example, we present a rule in which payoffs enter exponentially in the function f.

$$z_{t+1,1,j} = z_t + z_t \mu \frac{e^{\pi_{2j}} - e^{\pi_{1j}}}{e^{\pi_{1j}} + e^{\pi_{2j}}} \tag{1}$$

$$z_{t+1,2,j} = z_t + (1-z_t)\mu \frac{e^{\pi_{2j}} - e^{\pi_{1j}}}{e^{\pi_{1j}} + e^{\pi_{2j}}} \tag{2}$$

Another example could be the following.

$$\begin{aligned} z_{t+1,1,j} &= z_t + z_t \mu(\pi_{2j} - \pi_{1j}) \\ z_{t+1,2,j} &= z_t + (1-z_t)\mu(\pi_{2j} - \pi_{1j}) \end{aligned}$$

The intuition behind the Stochastic Best Response is the following. Each period, every player observes the payoff of the action chosen and the payoff of the other action. Then she updates her strategy in the following way. She increases the probability of playing action 1 in the next period if and only if action 1 yielded higher payoff than action 2 in the current period. The increase in the probability of playing action 1 will depend on the difference in payoffs between the two actions.

A different interpretation of this same rule uses the fact that z can be considered as the fraction of population playing action 2 deterministically. Under this consideration, at every period, any player that didn't play the best action, will change her action (best response to the environment) with some probability. The probability of changing action depends on the difference in payoff between the two actions. The Stochastic Best Response is an individual learning rule because actions played by other players have no effect on the updating of the own's strategy. That is, players learn only from own experiences.

3.2 Replicator Dynamics

The functional form for the Replicator Dynamics is the same independently on which action yielded higher payoff. We write $z_{t+1,j}$ to denote the value of the variable z at time t given that the state at time t was j. The Replicator Dynamics is characterized by the following equation.

$$z_{t+1,j} \quad = \quad z_t + z_t(\pi_{2j} - ((1 - z_t)\pi_{1j} + z_t\pi_{2j}))$$

The intuition behind the Replicator Dynamics is the following. Every period, each player observes the average payoff of the population (this is the term $((1 - z_t)\pi_{1j} + z_t\pi_{2j})$ in case state equals j). If her payoff is bigger than the average payoff of the population, then she will increase the probability of playing the action she played. The magnitude of this increase will depend on the difference between her payoff and the average payoff of the population. Another interpretation of the replicator Dynamics is based on imitation. Every period, every player observes the action and payoff of another player and imitates the action of the other player if and only if the other player got a higher payoff than herself.

4 Results

4.1 Stochastic Best Response

Before going to the formal results, we present a small discussion on the behavior of the learning rule. First, note that under the two rules we consider, the learning process always puts more weight tomorrow in the strategy that was more successful today. The biggest difference in the behavior of the two rules that we consider lies in the way they behave when z is close to the corners (0 and 1). In particular, under the Stochastic Best Response the corners are not absorbing. On the other hand, under the Replicator Dynamics both corners are absorbing.

Assume for this discussion that there are only 2 states of nature. Under the Stochastic Best Response, the speed at which a player adapts an action slows down as the probability of playing *that* action increases. That is, consider that action 1 is played with a high probability and that today action 1 yielded a higher payoff than action 2. Then the increase in the probability of playing action 1 will be low. On the other hand, consider that action 1 is played with a low probability and today action 1 yielded higher payoff than action 2. In this case the probability of playing action 1 next period will have a high increase. Figure 1 shows how z changes over time under the Stochastic Best Response.

Figure 1: Stochastic Best Response

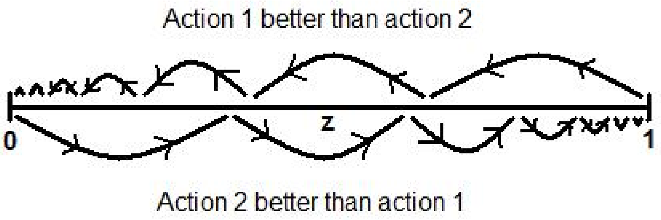


Figure 1 shows the movements of the probability of playing action 2 (z) as a response to an action being better than the other in the current period. As we mentioned before, assume that an action is played with a high probability. Then the increase in playing that action in case it yielded a higher payoff than the other action at the present period is low. This characteristic of the Stochastic Best Reply will be exploited for the proofs of our results..

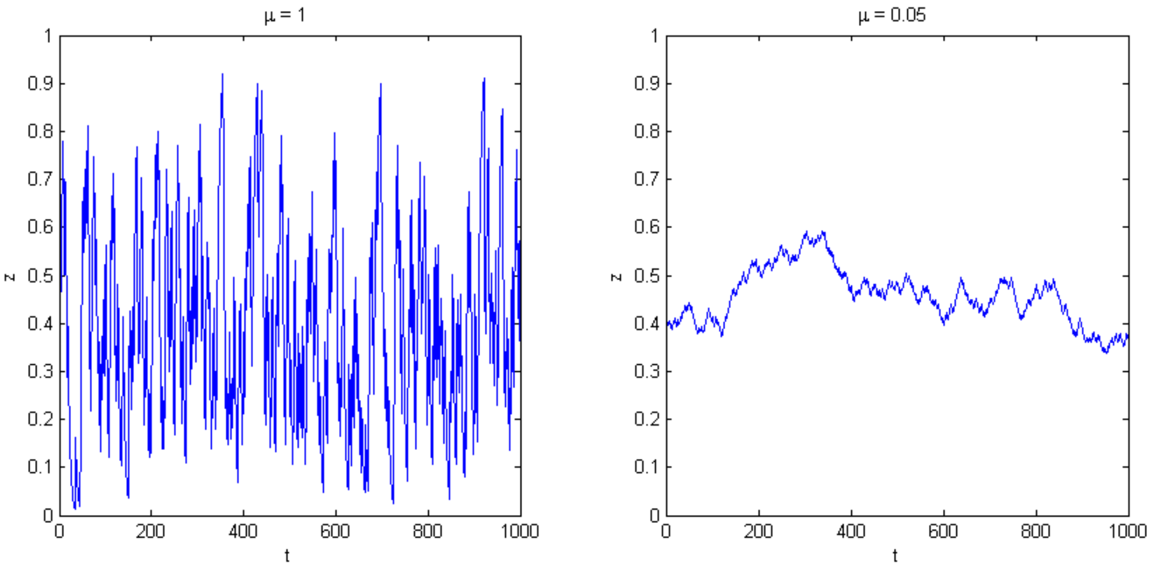
As one could possibly guess already, the Stochastic Best Response won't converge to any of the corners. The variable z will be always going back and forth. To study convergency, we consider the limit case when μ, which can be view as the size of the changes in z, goes to 0. Once such a limit is taken, the Stochastic Best Response converges to a single point. This issue can be seen much clearer by looking at Figure 2, where a simulation is conducted. The specific rule used is given by 1 and 2. The value of the parameters is set to $m = 2$, $\pi_{11} = 0.5, \pi_{12} = 0.3, \pi_{21} = 0.1, \pi_{22} = 0.6$ and $\theta_{12} = \theta_{21} = 0.3$. The initial value of z was set to $z_0 = 0.4$.

By studying the behavior of the system when μ is made arbitrarily small we are characterizing the asymptotic behavior of z. When μ is taken to zero the adjustment in the strategies is made arbitrarily small. However, as time goes to infinity the number of times the adjustments take place is also made arbitrarily big. For papers that use this continuous time limit approximation in setting somewhat different form ours see, for example, Börgers and Sarin (1997) and Benaïm and Weibull (2003).

The following proposition characterizes the convergence of the Stochastic Best Response when μ is arbitrarily small.

Proposition 1. *Define* $\tilde{z} = \frac{\sum_{i=1}^{h} \mu_i f(\Pi_j)}{\sum_{i=1}^{m} \mu_i f(\Pi_j)}$. *For any* $\varepsilon > 0$ *there exists a* $\bar{\mu} \in (0, \hat{\mu}]$ *such that if*

Figure 2: Simulation - Stochastic Best Response



$\mu < \bar{\mu}$ *then*

$$\lim_{h \to \infty} \left| z_{t+h} - \tilde{z} \right| < \varepsilon.$$

The proof is presented in the appendix. Below we present a sketch of the proof. The point $\tilde{z}$ corresponds to the point where an expected increase in z_t due to action 2 yielding higher payoff at time t than action 1 would be equivalent to the expected decrease in z_t from action 1 yielding more payoff than action 2. That is, with $m = 2$, $\tilde{z}$ is such that $|z_{t+1,1,1} - z_t| = |z_{t+1,2,2} - z_t|$. In Figure 1, the point $\tilde{z}$ would be such that the size of the arrows (or jumps) towards the left form a given point z is the same as the size of the arrows towards the right from this same point z. Hence, $\tilde{z}$ is the point where the marginal movements towards action 1 and towards action 2 is equalized. One can easily check that if action 1 is better in the long run, which happens if $\sum_{i=1}^{h} \mu_i f(\Pi_j) > \sum_{i=h+1}^{m} \mu_i f(\Pi_j)$, then it will be played more often than action 2. However, it will never be the case that action 1 is played with frequency 1. In our simulation above we have that $\tilde{z} = 0.43$. That is, in the long run at any given period action 2 is played with probability of 0.43. The behavior under Stochastic Best Response in the limit when μ goes to 0 resembles the intuition behind the probability matching behavior explained in the introduction. This behavior is clearly suboptimal as if $\sum_{i=1}^{h} \mu_i f(\Pi_j) > \sum_{i=h+1}^{m} \mu_i f(\Pi_j)$ then the z that maximizes payoff is $z = 0$. That is, playing action 1 always is better than playing action 1 with probability $1 - \tilde{z}$ for $\tilde{z} > 0$.

We now present an sketch of the proof of our result. For studying the convergence of the variable z we first show that it suffices to study the convergence of a variable y that evolves in a world with just 2 states of nature and symmetric transition matrix. Define $h < m$, as the maximum j such that $\pi_{1j} \geq \pi_{2j}$. Define y as $y_t = z_t$ and

$$y_{t+1} = \begin{cases} y_t + 2y_t \sum_{i=1}^{h} \mu_i f(\Pi_j) \text{ with probability } 1/2 \\ y_t + 2(1-y_t) \sum_{i=h+1}^{m} \mu_i g(\Pi_j) \text{ with probability } 1/2. \end{cases}$$

The following result, whose proof is presented in the appendix, states that both z and y converge in probability to the same value.

Lemma 1. *For any $\varepsilon > 0$ there exists a $\hat{\mu} > 0$ and a $k \in \mathbb{N}$ such that for any $\mu < \hat{\mu}$ and $h > k$ we have that*

$$P\left(|z_{t+h} - y_{t+h}| > \varepsilon\right) = 0.$$

The intuition is as follows. Assume that we are at time k. Hence, for $t > k$ big enough we can rewrite the evolution of the variable z as follows.

$$E_k(z_{t+1}) = \begin{cases} z_t + z_t f(\Pi_1) \text{ with probability } \mu_1 \\ \vdots \\ z_t + z_t f(\Pi_h) \text{ with probability } \mu_h \\ z_t + (1 - z_t) f(\Pi_h) \text{ with probability } \mu_{h+1} \\ \vdots \\ z_t + (1 - z_t) f(\Pi_m) \text{ with probability } \mu_m \end{cases}$$

Note that the variable y evolves according to the expected movement in the long run of the variable z. It can be easily seen that $E_k(z_{t+1}) = E_t(y_{t+1})$. By time invariance of both z and y we can extend this to $E_k(z_{t+h}) = E_t(y_{t+h})$.

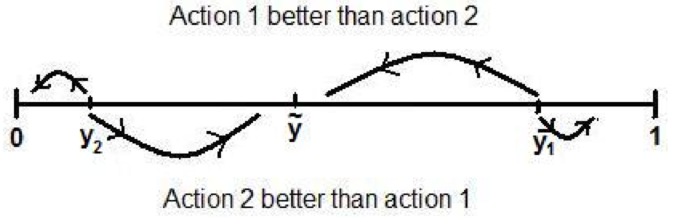
Furthermore, by making μ arbitrarily small we make the variance of both random variables z and y to shrink to zero. Thus, we must have that their limiting distribution puts weight on a single point. In other words, z and y converge in probability to a fix value $\tilde{z}$ and $\tilde{y}$ respectively. Since $E_k(z_{t+h}) = E_(y_{t+h})$ for t big enough and for all $h > 0$, we must have that $\tilde{z} = \tilde{y}$. Hence, instead of studying the convergence of the variable z we focus on the convergence of the variable y.

For simplicity of the exposition, assume for now that $\theta < 0.5$. Imagine that we start at a point y to the right of $\tilde{y}$. From that point take a sequence of the next k states of nature where states 1 occurred at least the same times as state 2. Call this sequence $\{s_{t+h}\}_{h=1}^k$. Take now the sequence that is exactly opposite to $\{s_{t+h}\}_{h=1}^k$ except for s_t and call it $\{\bar{s}_{t+h}\}_{h=1}^k$. That is, if the second state in the sequence $\{s_{t+h}\}_{h=1}^k$ is 2 then the second state in the sequence $\{\bar{s}_{t+h}\}_{h=1}^k$ would be 1 and so on. Given that the transition probability between states is symmetric, the probability that the sequence $\{\bar{s}_{t+h}\}_{h=1}^k$ occurs equals $\frac{\theta}{1-\theta}$ times the probability that the sequence $\{s_{t+h}\}_{h=1}^k$ occurs. The only different between the two probabilities is due to the value of the state of nature at time t.

For any point y to the right of $\tilde{y}$, we will have that the one-step movement towards action 1 is bigger than the movement towards action 2. In graphical terms the arrows towards the left are longer than the arrows towards the right if you are at a point to the right of $\tilde{y}$. Figure 3 presents a representation of this situation. A point like $y_1 > \tilde{y}$ is such that the one-step movements towards the left (towards action 1) are bigger than the one-step movements towards the right (action 2). Exactly the opposite happens to a point such as $y_2 < \tilde{y}$.

Consider now our sequences $\{s_{t+h}\}_{h=1}^k$ and $\{\bar{s}_{t+h}\}_{h=1}^k$. It can be shown that the value of z after the sequence $\{s_{t+h}\}_{h=1}^k$ occurs is smaller than its initial value z_t. However, it is not

Figure 3: Movement under Stochastic Best Response



necessarily true that the value of z after the sequence $\{\bar{s}_{t+h}\}_{h=1}^{k}$ occurs is bigger than its initial value z_t. It can be shown that the change in z starting off at $z_t > \tilde{z}$ and after the sequence $\{s_{t+h}\}_{h=1}^{k}$ happens, is significantly higher than the change in z starting off at z_t and after the sequence $\{\bar{s}_{t+h}\}_{h=1}^{k}$ occurs. In particular, the value of $\frac{\theta}{1-\theta}y_{t+k}$ under the sequence $\{s_{t+h}\}_{h=1}^{k}$ plus the value of y_{t+k} under the sequence $\{\bar{s}_{t+h}\}_{h=1}^{k}$ is smaller than y_t. This is true for any sequence $\{\bar{s}_{t+h}\}_{h=1}^{k}$ where states 1 occurred at least the same times as state 2.

Hence, the sum over all possible sequences $\{s_{t+h}\}_{h=1}^{k}$ of $\frac{\theta}{1-\theta}y_{t+k}(\{s_{t+h}\}_{h=1}^{k})+y_{t+k}(\{\bar{s}_{t+h}\}_{h=1}^{k})$ is smaller than y_t. That is, if we start at $y_t > \tilde{y}$, then the expected value of y after k periods, will be smaller than y_t[3]. That is, for $y_t > \tilde{y}$ and some k, $E_t(y_{t+k}) < y_t$. This same reasoning applies to show that for $y_t < \tilde{y}$ and some k, $E_t(y_{t+k}) > y_t$. With this information, we can conclude that for some k $E_t(y_{t+k}) = \tilde{y}$. The reasoning why we can extend this fact to any k bounded from below is more technical and only presented in the appendix.

Once we have that for k big enough $E_t(y_{t+k}) = \tilde{y}$. It only remains to explain the intuition why when μ goes to zero, y_{t+h} goes to $\tilde{y}$ as h increases. It can be easily seen by the equation of the Stochastic Best Response that as we decrease μ, the statistical variance of y decreases as well. Hence, since $E_t(y_{t+k}) = \tilde{y}$ and the variance of y tends to zero as μ goes to zero, we must have that y_{t+h} tends to $\tilde{y}$ as h goes to infinity and μ goes to zero.

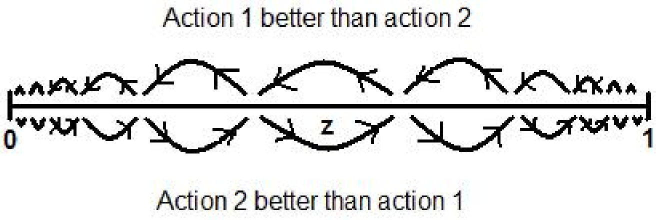
4.2 Replicator Dynamics

The Replicator Dynamics has a completely different behavior than the Stochastic Best Response. Under the replicator dynamics, the changes in the variable z become smaller as it gets closer to *either* bound. For example, assume that $m = 2$ and consider that action 1 is played with a high probability. Then the change in z will be small independently of whether action 1 yielded higher payoff than action 2 or the other way around. Figure 4 shows how the transitions work for the Replicator Dynamics.

As we see, the process will spend almost no time in intermediate values of z. This will allows to draw our conclusions from analyzing only the behavior of z in the neighborhoods of

[3] The value of k must be bounded above and below. Intuitively, we need k to be bounded below so that the value of the current state of nature is almost not influencing the value of y_{t+k}. We need k to be bounded above so that y does not hit a $\tilde{y}$. For any sequence of nature and any value of y_t there exists a $\bar{\mu}$ such that for $\mu < \bar{\mu}$ we can always find such a k. See formal proof in the Appendix for the details.

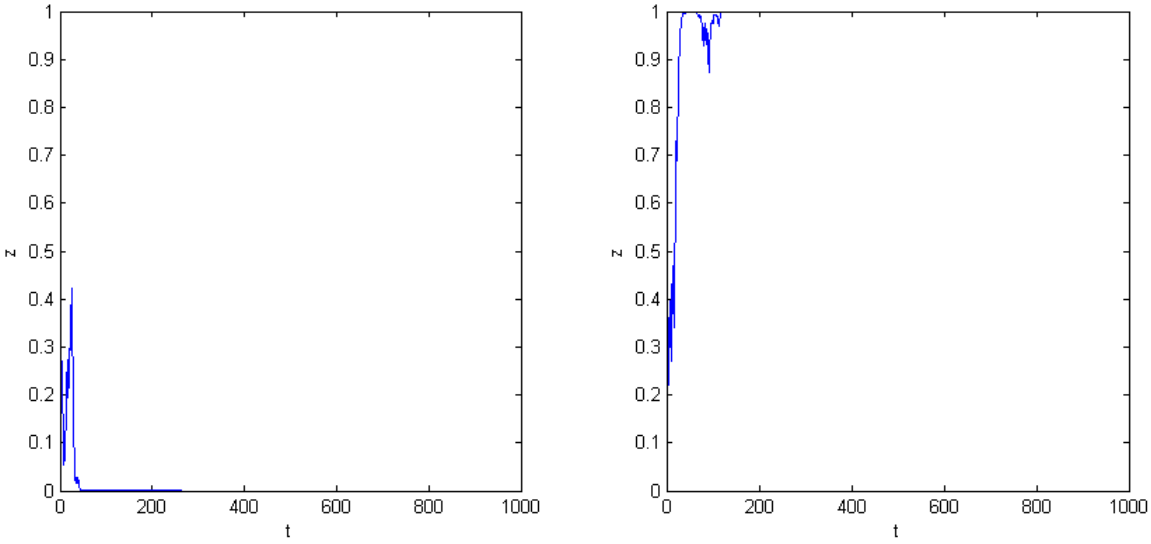
Figure 4: Replicator Dynamics



its bounds. In this respect, our analysis will rely partially on the approach by Ellison and Fudemberg (1995).

Figure 5 shows the same simulation as Figure 2 but with the Replicator Dynamics Learning instead of the Stochastic Best Response. That is, the parameters used are the same but the learning rule is different. Figure 5 shows the result of the same simulation performed with two different random seeds.

Figure 5: Simulation - Replicator Dynamics



As we see, the Replicator Dynamics quickly converges to a situation in which all the population plays the same action a fraction 1 of the time. An interesting thing to note is that it is not necessarily the case that the action selected by the Replicator Dynamics coincides with the best action. The simulation in the right hand size shows a situation in which the Replicator Dynamics converges to a situation in which everybody is playing the suboptimal action. As we show below, this fact will be the result of the two actions performing not too differently in terms of payoffs in the long run.

Define $\gamma_j = 1 + \pi_{2j} - \pi_{1j}$ and $\hat{\gamma}_j = 1 - \pi_{2j} + \pi_{1j}$ for all $j \in \{1, \ldots, m\}$. The following proposition characterizes the convergence of the variable z when it evolves according to the

Replicator Dynamics equation.

Proposition 2. *Consider the two inequalities:*

$$\sum_{j=1}^{m} \mu_j \log \gamma_j > 0 \tag{3}$$

$$\sum_{j=1}^{m} \mu_j \log \hat{\gamma}_j > 0. \tag{4}$$

1. *If both (3) and (4) hold, then z does not converge to any value.*

2. *If (3) holds but (4) does not then z converges to 1.*

3. *If (4) holds but (3) does not then z converges to 0.*

4. *If neither (3) nor (4) hold then z converges to either 0 and 1, both with positive probability.*

An important fact that the proposition above is revealing is that the process may fail to converge to the best action. Action 1 is weakly better than action 2 in the long run if and only $\sum_{j=1}^{m} \mu_j \pi_{1j} \geq \sum_{j=1}^{m} \mu_j \pi_{2j}$. This condition can be rewritten as $\sum_{j=1}^{m} (\mu_j + \mu_j(\pi_{2j} - \pi_{1j})) \leq 1$. Which in turn can be rewritten as $\sum_{j=1}^{m} \mu_j \gamma_j \leq 1$. Similarly, action 1 is weakly better than action 2 in the long run if and only $\sum_{j=1}^{m} \mu_j \hat{\gamma}_j \geq 1$. However, even if $\sum_{j=1}^{m} \mu_j \hat{\gamma}_j \geq 1$ holds it may happen that 4 does not hold. For making this point more clear consider the case in which $m = 2$ and $\mu_1 = \mu_2 = 0.5$. That is, there are only two states of nature and both states are equally likely in the long run. The following proposition characterizes the convergence of z in this case when action 1 is better in the long run than action 2.

Proposition 3. *Assume $m = 2$, $\mu_1 = \mu_2 = 0.5$ and $\pi_{11} + \pi_{12} > \pi_{21} + \pi_{22}$.*

- *If*

$$\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} - (\pi_{11} - \pi_{21})(\pi_{22} - \pi_{12}) > 0$$

then the process converges to $z = 0$.

- *If*

$$\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} - (\pi_{11} - \pi_{21})(\pi_{22} - \pi_{12}) \leq 0$$

then the process converges to either $z = 0$ or $z = 1$, both with positive probability.

Proof. We can rewrite Proposition 2 for the case with $m = 2$ and $\mu_1 = \mu_2 = 0.5$ as follows. Consider the two inequalities:

$$\frac{1-p}{p} > -\frac{\log \gamma_1}{\log \gamma_2} \tag{5}$$

$$\frac{1-p}{p} > -\frac{\log \hat{\gamma}_2}{\log \hat{\gamma}_1}. \tag{6}$$

1. If both (5) and (6) hold, then z does not converge to any value.

2. If (5) holds but (6) does not then z converges to 1.

3. If (6) holds but (5) does not then z converges to 0.

4. If neither (5) nor (6) hold then z converges to either 0 and 1, both with positive probability.

Moreover, the condition in (5) can be rewritten as

$$-\frac{\log(1-b)}{\log(1+a)} < 1$$

with $1 > b > a > 0$. The equation above can be rewritten as

$$\begin{aligned} \log(1-b) + \log(1+a) > 0 \\ 1-b > \frac{1}{1+a}. \end{aligned}$$

Hence, (5) implies $-b + a - ab > 0$ which is impossible if $1 > b > a > 0$. Therefore, if $\pi_{11} + \pi_{12} > \pi_{21} + \pi_{22}$ condition (5) can not hold. Thus, we must have that either z converges to 0 or z can converge to both 0 and 1. □

For the process to select the best action, the two actions need to perform significantly different. That is, having $\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} > 0$ (action 1 better than action 2) is not enough for the Replicator Dynamics to select the best action. We must have $\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} - (\pi_{11} - \pi_{21})(\pi_{22} - \pi_{12}) > 0$.

Here we present the intuition for the proof when there are only 2 states of nature. The proof of Proposition 1 relies partially on the analysis by Ellison and Fudemberg (1995). Their trick is the following. Assume, as in their paper, that there are only 2 states and the realization of states is independent. Let p be the probability by which of state 2 occurs. Since the process spends almost no time at its intermediate value, only the boundaries, it suffices to examine the convergence of the variable z when it is close to its boundary values (0 and 1). Imagine that z is arbitrarily close to 0. Then, we can rewrite the equations for the Replicator Dynamics as follows,

$$\begin{aligned} z_{t+1,1} &= \gamma_1 z_t + o(z_t) \\ z_{t+1,2} &= \gamma_2 z_t + o(z_t) \end{aligned}$$

where $\gamma_1 = 1 + \pi_{21} - \pi_{11}$, $\gamma_2 = 1 + \pi_{22} - \pi_{12}$ and $o(z_t)$ it's a term that goes to zero at a higher rate than z_t. Since action 1 is better in the long run than action 2, that is $\pi_{11} + \pi_{12} > \pi_{21} + \pi_{22}$, we have that $\gamma_2 > 1 > \gamma_1 > 0$.

The variable z converges to 0 if and only if the variable $x = \log(z)$ converges to $-\infty$. The process for x when z is close to 0 can be approximated by

$$\begin{aligned} x_{t+1,1} &= x_t + \log(\gamma_1) \\ x_{t+1,2} &= x_t + \log(\gamma_2) \end{aligned}$$

where $x_{t+1} = x_{t+1,1}$ with probability $1-p$ and $x_{t+1} = x_{t+1,2}$ with probability p. Therefore, $E_t(x_{t+1}) = (1-p)\log(\gamma_1) + p\log(\gamma_2) + x_t$. Hence, if $(1-p)\log(\gamma_1) + p\log(\gamma_2) > 0$ then $E_t(x_{t+1}) > x$ which implies that x is a super-martingale. Thus, by the Martingale Convergence Theorem, if $(1-p)\log(\gamma_1) + p\log(\gamma_2) > 0$, then x cannot converge to $-\infty$.

The result by Ellison and Fudemberg (1995) is presented here for the readers convenience.[4]

Lemma 2 (Ellison and Fudemberg (1995)). *Let x_t be a Markov Process on (0,1) with*

$$x_{t+1} = \left\{ \begin{array}{c} H_1(x_t) \textit{ with probability } 1-p \\ H_2(x_t) \textit{ with probability } p \end{array} \right\}.$$

Suppose that $H_i(x_t) = \gamma_i x_i + o(x_t)$, with $\gamma_2 < 1 < \gamma_1$.

(a) If

$$\frac{1-p}{p} > -\frac{\log(\gamma_1)}{\log(\gamma_2)}.$$

then x_t cannot converge to 0 with positive probability.

(b) If

$$\frac{1-p}{p} < -\frac{\log(\gamma_1)}{\log(\gamma_2)}.$$

then there are strictly positive δ and ϵ such that $Prob[x_t \to 0|x_0 \leq \delta] \geq \epsilon$.

(c) If

$$\frac{1-p}{p} > -\frac{\log(\gamma_1)}{\log(\gamma_2)}.$$

there is a $x^* > 0$ such that for all $x_0 > 0$, $Prob[x_t < x^* \forall t|x_0] = 0$.

The difference between Ellison and Fudemberg setting and ours is that in our model the nature evolves according to a Markov chain not as independent realizations of a bernoulli random variable.

So how do we get from our setting to profit from their result? Under our setting, the state of nature tomorrow depends on the state of nature today. However, the state of nature many periods ahead is independent of the state of nature today. This means that by the law of large numbers, we can use Ellison and Fudemberg's setting taking their probability of each state being realized p as the limiting distribution of the Markov chain that governs states under our setting.

5 Discussion

5.1 The Infinite Memory Case - Rustichini (1999)

As mentioned in the introduction, Rusthichini (1999) presents a model where the environment changes according to a Markov change and for any state in the chain there is a unique action that

[4] What we here write as p is written as $1-p$ in the original paper. Moreover, what in the original papers is written as γ_1 (γ_2) we write it here as γ_2 (γ_1).

maximizes payoff. Rustichini considers the case of many actions and many states as opposite to just two actions. The two main difference between his analysis and ours are the rules he considers and the fact that he assumes that players have infinite memory. Rustichini explores two situations, one in which players observe only own payoffs (partial information) and another one in which players also observe foregone payoffs (full information). He shows that exponential procedures (Logit learning) selects in the long run the best action under the full information case but not under the partial information case. On the other hand, linear procedures select the best action in the partial information case but not under the full information set. The exponential procedures are defined as follows. Under full information denote by Π_{it} the payoff that action i yielded at time t. Under partial information, write Π_{it} to denote the payoff that action i yielded at time t if it was played by the agent, write $\Pi_{it} = 0$ otherwise. The exponential procedure is defined by

$$z_{t+1} = \frac{e^{\sum_{h=0}^{t} \Pi_{2h}}}{e^{\sum_{h=0}^{t} \Pi_{1h}} + e^{\sum_{h=0}^{t} \Pi_{2h}}}.$$

On the other hand, the linear procedure is defined by

$$z_{t+1} = \frac{\sum_{h=0}^{t} \Pi_{2h}}{\sum_{h=0}^{t} \Pi_{1h} + \sum_{h=0}^{t} \Pi_{2h}}.$$

5.2 Börgers et. al. (2004) Best Monotone Learning Rule

In this subsection we use our analysis above to study the convergence of the Best Monotone rule in Börgers, T., Morales, A. and Sarin, R. (2004) (BMS rule henceforth). The BMS rule is a fairly recent rule that is attractive because of its optimal properties. In particular, it is the best monotone rule (to be explained later) among all the individual learning rules in setting were foregone payoffs are not observed.

The BMS rule is derived from the Cross (1973) learning rule. Under Cross learning rule, players increase the probability of playing the action they just played. The increase is proportional to the payoff archived by the action they chose. Note that the probability of playing the action the player chose increases for the next period even if the payoff achieved was low. Another important point to note here is that under BMS or Cross learning rule players do not observe foregone payoffs[5]. As opposite to the Cross learning rule, BMS rule incorporates an endogenous aspiration level. If the payoff achieved by the action played is higher than the aspiration level, then the probability of playing the action increases. How this aspiration level is formed depends on the initial probability of playing each action. Börgers, Morales and Sarin (2004) show that this rule is the best monotone rule.

A rule is defined to be monotone if the expected probability of playing the action that is best given today's state increases. Since Börgers et. al. (2004) study a setting in which the realization of states is independent of its past value, the action that is best today is the action

[5] One could assume that players observe foregone payoffs but don't use this extra information in the way they update their strategies

that that is best forever. In our setting the action that is best today may not be the best action tomorrow due to the markovian evolution of the states of nature. This particular difference will have big consequences in the optimality properties of the BMS rule. A rule is said the be the best monotone rule if the expected increase in playing the best action form period to another is highest among all monotone rules.

In the Börgers et. al. (2004) setting there is a single decision maker instead of a continuum of population. Let $\sigma^1_{t+1,j}$ be the probability by which the decision maker plays action 2 at time $t+1$ given that action 1 was played at time t and state at t equals j. Note that since the decision maker does not observe forgone payoffs she conditions the way she updates her strategies in the action she played but not in which action yielded higher payoff. Let z_0 be given exogenously and define $A = -\frac{\min\{1-z_0,z_0\}}{\max\{1-z_0,z_0\}}$, $B = \frac{1}{\max\{1-z_0,z_0\}}$. The BMS rule when there are two actions available is defined as follows.

$$\sigma^1_{t+1,j} = \sigma_t - \sigma_t\left(A + B\pi_{1j}\right)$$
$$\sigma^2_{t+1,j} = \sigma_t + \left(1-\sigma_t\right)\left(A + B\pi_{1j}\right)$$

Notice the similarity between BMS rule and the Stochastic Best Response rule. The main difference lies in the fact that under the Stochastic Best Response, the probability of playing action i increases if and only if that action was better than the other. On the other hand, under BMS rule, the expected probability of playing action i increases if the action yielded higher payoff than the endogenous aspiration level determined by A and B.

Since we are dealing with a continuum of population, Law of Large applies and we can define our variable z as follows. Let $z^1_{t+1,j}$ be the probability by which the population (alternatively, the fraction of people in the population) that plays action 2 at time $t+1$ given that action 1 was played at time t and state at t equals j. Then we can write the evolution of the variable z as follows.

$$z_{t+1,j} = z_t\left(1 + (1-z_t)B(\pi_{2j} - \pi_{1j})\right)$$

For studying the convergence of the variable z under the BMS rule we proceed in a similar fashion as in the Replicator Dynamics case. We focus on the convergence of z when it is close to the corners. First, if z is arbitrarily close to the borders we can rewrite the value of $z_{t+1,j}$ as

$$z_{t+1,j} = z_t\beta_j$$

where $\beta_j = (1 + B(\pi_{2j} - \pi_{1j}))$. Once we have this, using the same steps as in the proof of Proposition 1 we can get the following result.

Proposition 4. *Define $\beta_j = (1 + B(\pi_{2j} - \pi_{1j}))$ and $\hat{\beta}_j = (1 - B(\pi_{2j} - \pi_{1j}))$. Consider the two inequalities:*

$$\sum_{j=1}^{m} \mu_j \log \beta_j > 0 \tag{7}$$

$$\sum_{j=1}^{m} \mu_j \log \hat{\beta}_j > 0. \tag{8}$$

1. *If both (7) and (8) hold, then z does not converge to any value.*
2. *If (7) holds but (8) does not then z converges to 1.*
3. *If (8) holds but (7) does not then z converges to 0.*
4. *If neither (7) nor (8) hold then z converges to either 0 and 1, both with positive probability.*

The only difference in the conditions for convergence between BMS rule and the Replicator Dynamics lie in the parameter B. Next we compare which of the two rules is more likely to select the best action in the long run.

We have that $B = \frac{1}{\max\{1-z_0, z_0\}}$, thus, $1 \leq B \leq 2$. Hence, $\gamma_j \geq \beta_j$ for $j \in \{1, \ldots, h\}$ and $\gamma_j \leq \beta_j$ for $j \in \{h+1, \ldots, m\}$. That is, under BMS rule the process moves faster towards the action that is better given today's state as compared with the Replicator Dynamics. However, the action that is better given today's state needs not to be the action that is better in the long run. Hence, moving faster towards the action that was better given today's state maybe a bad thing if that action is not the one that is actually the best in the long run. Moreover, what matters for the convergence of z in the long run are not the values of γ_j and β_j but the values of $\log \gamma_j$ and $\log \beta_j$. We find that whether BMS rule is more likely to select the best action in the long run for sure than the Best Response ill depend on the specific payoffs and transition matrix we are dealing with.

5.3 Relating our results for the Stochastic Best Response with Kosfeld et. al. (2002)

Kosfeld et. al. (2002) present a setting were a finite set of players play a normal-form game. Each period players update their strategies myopically in the following way. They increase the probability of playing an action if and only if that action is a best response to the action played by the other players. In case there are many actions that are a best response, the increase in probability is shared among these actions. Formally, let $\sigma_i^t(s_j)$ be the probability by which player i plays action j at time t. Define s_{-i} as the actions played by all the players but i. Finally, let $B_i(s_{-i})$ be the set of actions that are a best response to s_{-i} for player i. The change in the strategies of every player i is governed by

$$\sigma_i^{t+1}(s_i) = \begin{cases} (1-\theta)\sigma_i^t(s_i) + \theta/|B_i(s_{-i})| & \text{if } s_i \in B_i(s_{-i}), \\ (1-\theta)\sigma_i^t(s_i) & \text{otherwise.} \end{cases}$$

where $\theta \in (0,1)$ is exogenously given.

Comparing this rule with the Stochastic Best Response there are two points worth noting. First, in our model players play against nature and not against themselves. Hence, in Kosfeld et. al. (2002) setting, players best responds the actions of other players while in our setting players best respond the actions of nature. Second, and most important, in our setting how fast an action is adapted depends on whether it is a best response to the environment and in the payoffs it yields. On the other hand, in Kosfeld et. al. (2002) how fast an action is adapted depends only on whether it is a best response to the actions of the other players.

Kosfeld et. al. (2002) show that the continuous time limit of their process converges to a so called best-reply matching equilibrium. The best reply matching equilibrium is a situation in which for every player the probability of playing a given action is equal to the probability of that action being a best response given the strategies of the other players. It is clear from this definition that the set of pure strategy best reply matching equilibria coincides with the set of pure strategies Nash equilibria.

Their result and our result for the Stochastic Best Response have the same intuition behind and in some setting are equivalent. Consider the Stochastic Best Response in which the magnitude of the payoffs doesn't matter. That is, consider the following equation for the evolution of the variable z.

$$z_{t+1,1,j} = z_t + z_t\mu \tag{9}$$

$$z_{t+1,2,j} = z_t + (1 - z_t)\mu \tag{10}$$

In Proposition 1 we show that the process above converges in probability to $\tilde{z} = \frac{\sum_{i=1}^{h}\mu_i f(\Pi_j)}{\sum_{i=1}^{m}\mu_i f(\Pi_j)} = \frac{\sum_{i=1}^{h}\mu_i}{\sum_{i=1}^{m}\mu_i}$. That is, z, which is the probability of playing action 2, converges to the limiting probability that action 2 is a best response to the environment. Hence, the population strategies are matching the nature's strategies, exactly what the best-reply matching equilibrium would prescribe.

In our results for the Stochastic Best Response we consider a much bigger set of rules than Kosfeld et. al. (2002) do[6]. However, for the specific rule in which the magnitude of payoffs do not matter, as in equations (9) and (10), our result is a particular case of theirs.

6 Conclusions

In this paper we investigate the behavior of two well known and widely used learning rules in a model with aggregate and correlated shocks to payoffs. In particular, the payoff of each possible action depends on the state of nature. The transition between states follows a Markov Chain and, hence, there is correlation between today's state and tomorrow's state of nature. The learning rules we studied, Stochastic Best Response and Replicator Dynamics, represent in our opinion the paradigm of individual and population learning rules. Our contribution to the literature relies on the fact that we studied the behavior of these two rules in a setting where the realization of the state of nature is correlated with the past.

The literature has focused its attention to the study of such rules only in setting where the realization of states (or the shocks to payoffs) is independent. The most likely reason being the technical complexities involved in dealing with correlated realization of states.

[6] In particular, they only consider one rule

There are several questions that we have been unable to answer in this paper. For the Stochastic Best Response, we had to assume that the transition matrix between states is symmetric. The characterization of the behavior of the Replicator Dynamics we presented is only partial. That is, in the cases where z can can converge to either end point, we are unable to specify with which probability z converges to each of the two end points.

We expect that in the future more papers dealing with correlated environments will appear. The present piece of work has tried to shed some light on the techniques one could use for dealing with such environments.

References

1. Ben-Porath, E., Dekel, E. & Rustichini, A. (1993): "On the Relationship between Mutation Rates and Growth Rates in Changing Environment". *Games and Economic Behavior* 5 (4), 576-603.

2. Benaïm & Weibull (2003): "Deterministic Approximation of Stochastic Evolution in Games". *Econometrica* 71 (3), 873-903.

3. Börgers, T., Morales, A. & Sarin, R. (2004): "Expedient and Monotone Learning Rules". *Econometrica* 71 (2), 383-405.

4. Börgers, & Sarin, R. (1997): "Learning Through Reinforcement and Replicator Dynamics". *Journal of Economic Theory* 77, 1-14.

5. Camerer, C. & Ho, T. H. (1999): "Experienced-Weighted Attraction Learning in Normal Form Games". *Econometrica* 67 (4), 827-874.

6. Cross, J. (1973): "A Stochastic Learning Model of Economic Behavior". *The Quarterly Journal of Economics* 87, 239-266.

7. Easley, D. & and Rustichini, A. (1999): "Choice without Beliefs". *Econometrica* 67 (5), 1157-1184.

8. Ellison, G. & Fudenberg, D. (1995): "Word-of-Mouth Communication and Social Learning". *The Quarterly Journal of Economics* 110 (1), 93-125.

9. Erev and Roth (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria". *The American Economic Review* 88 (4), 848-881.

10. Fudenberg, D. & Harris, C. (1992): "Evolutionary Dynamics with Aggregate Shocks". *Journal of Economic Theory* 57, 420-441.

11. Hopkins, E. (2002): "Two Competing Models of How People Learn in Games". *Econometrica* 70 (6), 2141-2166.

12. Kosfeld, M., Droste, E. & Voorneveld, M. (2002): “A Myopic Adjustment Leading to Best Reply Matching”. *Games and Economic Behavior* 40, 270-298.

13. Rubinstein, A. (2002): “Irrational Diversification in Multiple Decision Problems”. *European Economic Review* 46, 1369-1378.

14. Rustichini, A. (1999): “Optimal Properties of Stimulus Response Learning Models”. *Games and Economic Behavior* 29, 244-273.

15. Roth and Erev (1995): “Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term”. *Games and Economic Behavior* 8, 164-212.

16. Schlag, K. (1998). “Why Imitate, and if so, How? A Boundedly Rational Approach to Multi-Armed Bandits”. *Journal of Economic Theory* 78 (1), 130-156.

17. Shanks, D., Tunney, R. & McCarthy, J. (2002): “A Re-examination of Probability Matching and Rational Choice”. *Journal of Behavioral Decision Making* 15, 233-250.

18. Siegel, S. & Goldstein, D. A. (1959): “Decision Making Behavior in a Two-Choice Uncertain Outcome Situation”. *Journal of Experimental Psychology* 57 (1), 37-42.

19. Vulkan, N. (2000): “An Economist’s Perspective on Probability Matching”. *Journal of Economic Surveys* 14 (1), 101118.

Appendix

Proof of Lemma 1

Define $h < m$, as the maximum j such that $\pi_{1j} \geq \pi_{2j}$. Define y as $y_t = z_t$ and

$$y_{t+1} = \begin{cases} y_t + 2y_t \sum_{i=1}^{h} \mu_i f(\Pi_j) \text{ with probability } 1/2 \\ y_t + 2(1-y_t) \sum_{i=h+1}^{m} \mu_i g(\Pi_j) \text{ with probability } 1/2. \end{cases}$$

Hence, we have that $E_k(y_{t+1}) = E_k(z_{t+1})$. Moreover, since both z and y have time invariant distributions we have that $E_k(y_{t+h}) = E_k(z_{t+h})$ for all $h \in \mathbb{N}$.

Lemma 3. *For any $\varepsilon > 0$ there exists a $\hat{\mu} > 0$ and a $k \in \mathbb{N}$ such that for any $\mu < \hat{\mu}$ and $h > k$ we have that $Var_k(y_{t+h}) < \varepsilon$ and $Var_k(z_{t+h}) < \varepsilon$.*

Proof. Given the definition of z and y, as μ is taken to zero the variance of both y and z goes to zero as well. The lemma follows. □

Hence, we have that the variable y as the same expected value in the long run than z and it’s variance collapses to zero in the long run as μ goes to zero. Thus, we can show the following result.

Lemma 1. *For any $\varepsilon > 0$ there exists a $\hat{\mu} > 0$ and a $k \in \mathbb{N}$ such that for any $\mu < \hat{\mu}$ and $h > k$ we have that*

$$P\left(|z_{t+h} - y_{t+h}| > \varepsilon\right) = 0.$$

Proof. Assume the opposite. As μ goes to zero the variance of z and y goes to zero. Hence both variables will converge in probability to a single point. If this point is different them we must have that $E_k(z_{t+m}) \neq E_k(y_{t+m})$ for some m. Which represents a contradiction. □

Proof of Proposition 1

We can rewrite the process for y as follows

$$y_{t+1,1,1} = b(y_t, \mu, 1)$$
$$y_{t+1,2,2} = b(y_t, \mu, 2)$$

Define

$$b^k(y_t, \mu, \{s_i\}_t^{t+k}) = b(b(\ldots b(b(y_t, \mu, s_t), \mu, s_{t+1}), \ldots), \mu, s_{t+k})$$

with $k \in \mathbb{N}$.

We now establish a set of facts about the function b.

Fact 1. *For all $\mu \in [0, \hat{\mu}]$ we have that*

$$\begin{aligned} b(y_t, \mu, 1) &\leq y_t \text{ with strict inequality if and only if } 0 < y_t \leq 1 \text{ and } \mu > 0, \\ b(y_t, \mu, 2) &\geq y_t \text{ with strict inequality if and only if } 0 \leq y_t < 1 \text{ and } \mu > 0, \\ \lim_{h\to\infty} b^h(y_t, \mu, \{1\}^h) &= 0 \text{ for all } y_t \in [0, 1), \\ \lim_{h\to\infty} b^h(y_t, \mu, \{2\}^h) &= 1 \text{ for all } y_t \in (0, 1]. \end{aligned}$$

Fact 2. *For any y_t, s_t and $\mu \in [0, \hat{\mu}]$, the function $b(y_t, \mu, s_t)$ is in $\mathcal{C}^2$ with respect to y_t. Furthermore,*

$$\frac{\partial\big(b(y_t, \mu, 1) - y_t\big)}{\partial y_t} \leq 0,$$
$$\frac{\partial\big(b(y_t, \mu, 2) - y_t\big)}{\partial y_t} \leq 0.$$

Fact 3.

$$b(0, \mu, 2) > 0$$
$$b(1, \mu, 1) > 0$$

Fact 4. *For any $y_t \in [0, 1]$ we have that*

$$b(y_t, 0, s_t) = y_t.$$

Fact 5. *For any y_t and s_t, the function $b(y_t, \mu, s_t)$ is twice continuously differentiable with respect to $\mu \in (0, \hat{\mu}]$. The function $b(y_t, \mu, 1)$ is weakly convex and $b(y_t, \mu, 2)$ is weakly concave with respect to $\mu \in (0, \hat{\mu}]$. Furthermore,*

$$\frac{\partial b(y_t, \mu, 1)}{\partial \mu} \leq 0 \text{ with strict inequality if and only if } 0 < y_t < 1.$$
$$\frac{\partial b(y_t, \mu, 2)}{\partial \mu} \geq 0 \text{ with strict inequality if and only if } 0 < y_t < 1.$$

Fact 6. *For $\tilde{y} = \frac{\sum_{i=1}^{h} \mu_i f(\Pi_j)}{\sum_{i=1}^{m} \mu_i f(\Pi_j)}$ we have that:*

- *For any $y_t = \tilde{y}$ then the following must hold.*

$$|b(y_t, \mu, 1) - y_t| = |b(y_t, \mu, 2) - y_t|$$

- *For any $y_t \in (\tilde{y}, 1]$ then the following must hold.*

$$|b(y_t, \mu, 1) - y_t| > |b(y_t, \mu, 2) - y_t|$$

- *For any $y_t \in [0, \tilde{y})$ then the following must hold.*

$$|b(y_t, \mu, 1) - y_t| < |b(y_t, \mu, 2) - y_t|$$

For proving Proposition 1 we first proof the following result.

Proposition 5. *For any $\varepsilon > 0$ there exists a $\bar{\mu} \in [0, \hat{\mu}]$ and $\bar{h}$ such that for $\mu < \bar{\mu}$ and $h > \bar{h}$ we have that*

$$P(|y_{t+h} - \tilde{y}| > \varepsilon) = 0.$$

First, we aim at showing that $\lim_{h\to\infty} \lim_{\mu\to 0} E_t(y_{t+h}) \to \tilde{y}$. Define the variable $x_t \in [\tilde{z}, 1]$ as follows,

$$x_t = \begin{cases} \tilde{y} & \text{if } y_t \leq \tilde{y} \\ y_t & \text{if } y_t > \tilde{y} \end{cases}$$

Similarly, define $w_t \in [0, \tilde{z}]$ as follows,

$$w_t = \begin{cases} y_t & \text{if } y_t < \tilde{y} \\ \tilde{y} & \text{if } y_t \geq \tilde{y} \end{cases}$$

Given the definition of x_t and w_t we have that for any $t, h \in \mathbb{N}$, $E_t(y_{t+h}) \leq E_t(x_{t+h})$ and $E_t(y_{t+h}) \geq E_t(w_{t+h})$. The strategy for the proof will be to show that $\lim_{h\to\infty} \lim_{\mu\to 0} E_t(x_{t+h}) \leq \tilde{y}$ and $\lim_{h\to\infty} \lim_{\mu\to 0} E_t(w_{t+h}) \geq \tilde{y}$.

First we show that $\lim_{h\to\infty} \lim_{\mu\to 0} E_t(x_{t+h}) \leq \tilde{y}$.

Definition 1. *Define*

$$\underline{x}^k = \inf\{x_t \textit{ such that } b^k(x_t,\mu,\{1\}^k) > \tilde{z}\}$$

and

$$\bar{x}^k = \sup\{z_t \textit{ such that } b^k(x_t,\mu,\{1\}^k) < 1\}.$$

Lemma 2. *For any x_t such that $x_t \geq \underline{x}^2$ and any and $\bar{x}_t \in [\underline{x}^1, 1]$ there exists an $\eta > 0$ such that $|b(x_t,\mu,1) - x_t| - |b(\bar{x}_t,\mu,2) - \bar{x}_t| > \eta$.*

Proof. Assume that $x_t < \bar{x}_t$. By Fact 2 we have that $|b(x_t,\mu,2) - x_t| \geq |b(\bar{x}_t,\mu,2) - \bar{x}_t|$. Furthermore, by Fact 6, $|b(x_t,\mu,1) - x_t| > |b(x_t,\mu,2) - x_t|$. Hence we have that $|b(x_t,\mu,1) - x_t| > |b(\bar{x}_t,\mu,2) - \bar{x}_t|$.

The proof for the cases in which $\bar{x}_t > x_t$ follow the same logic as above. Finally, we define $\eta = \arg\min_{x_t,\bar{x}_t}\{|b(x_t,\mu,1) - x_t| - |b(\bar{x}_t,\mu,2) - \bar{x}_t|\}$. Such η is strictly positive and well defined due to what we shown in the preceding paragraph and the fact that $x_t, \bar{x}_t$ are bounded. □

Define $\bar{\eta} = \min\{\eta$ such that $\{|b(x_t,\mu,1) - x_t| - |b(\bar{x}_t,\mu,2) - \bar{x}_t|\} \geq \eta$ for all $x_t, \bar{x}_t \in [\underline{x}^2, 1]\}$. From Lemma 2 we have that $\bar{\eta} > 0$.

Assume that learning is slow. In particular, μ is such that $\underline{x}^k < \frac{1}{2}$ for $k = \lceil \frac{2}{\bar{\eta}} + 1 \rceil$. In other words, learning is such that if you start somewhere below $\frac{1}{2}$, you need at least $\frac{2}{\bar{\eta}} + 1$ consecutive periods where the state is always 1 to get to 0. Note that we are still not imposing that μ should go to zero.

Finally, for any given sequence of states of nature $\{s_i\}_t^{t+k}$ define $\mathbb{1}\{s_i\}_t^{t+k} = \{\#1 \in \{s_i\}_t^{t+k}\}$.

Lemma 3. *Take $k \in \mathbb{N} \geq 2$. For any two sequences of states of nature $\{s_i\}_t^{t+k-1}$ and $\{\bar{s}_i\}_t^{t+k-1}$ such that $\mathbb{1}\{s_i\}_t^{t+k-1} > \mathbb{1}\{\bar{s}_i\}_t^{t+k-1}$ and $s_i = \{1,2\} \setminus \bar{s}_i$ for each $i \in \{t+1, \ldots, t+k-1\}$ with $s_t = \bar{s}_t$, we have that $|b^k(x_t,\mu,\{s_i\}_t^{t+k-1}) - x_t| - |b^k(x_t,\mu,\{\bar{s}_i\}_t^{t+k-1}) - x_t| > k\bar{\eta}$ for $x_t \geq \underline{x}^k$.*

Proof. Because of the fact that $\mathbb{1}\{s_i\}_t^{t+k-1} > \mathbb{1}\{\bar{s}_i\}_t^{t+k-1}$ and $s_i = \{1,2\} \setminus \bar{s}_i$ for each $i \in \{t+1, \ldots, t+k-1\}$ we know that $\mathbb{1}\{s_i\}_t^{t+k-1} \geq \frac{k}{2}$. Hence, we have the following,

$$\begin{aligned}\left|b^k(x_t,\mu,\{s_i\}_t^{t+k-1}) - x_t\right| \geq{} & \left(\mathbb{1}\{s_i\}_t^{t+k-1} - \left(k - \mathbb{1}\{s_i\}_t^{t+k-1}\right)\right) \min_{x_t \in (\underline{x}^k, 1]}\{b(x_t,\mu,1)\} \\ & + \left(k - \mathbb{1}\{s_i\}_t^{t+k-1}\right)\bar{\eta}.\end{aligned} \tag{11}$$

By the same token, we have that,

$$\begin{aligned}|b^k(x_t,\mu,\{\bar{s}_i\}_t^{t+k-1}) - x_t| \leq{} & \left(k - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1}\right) \max_{\bar{x}_t \in [0, \bar{x}^k)}\{b(\bar{x}_t,\mu,2)\} \\ & - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1}\bar{\eta}\end{aligned} \tag{12}$$

Because of the fact that $s_i = \{1,2\} \setminus \bar{s}_i$ for each $i \in \{t+1, \ldots, t+k-1\}$ and $s_t = \bar{s}_t$ it is true that

$$k - \mathbb{1}\{s_i\}_t^{t+k-1} = \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} - 1 \tag{13}$$

Hence, combining equations 11, 13 and the fact that, by definition of $\bar{\eta}$, $\min_{z_t \in (\underline{x}^k, 1]}\{b(x_t, \mu, 1)\} > \bar{\eta}$, we get the following.

$$\begin{aligned} \left| b^k(x_t, \mu, \{s_i\}_t^{t+k-1}) - x_t \right| > \left(k - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} \right) \min_{x_t \in (\underline{x}^k, 1]} \{b(x_t, \mu, 1)\} \\ + \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} \bar{\eta} \end{aligned} \tag{14}$$

Combining 12 and 14 we get to the following result.

$$\begin{aligned} |b^k(x_t, \mu, \{s_i\}_t^{t+k-1}) - x_t| - |b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) - x_t| > \\ \left(k - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} - \mathbb{1}\{\bar{s}_i\}_t^{t+k-1} \right) \left(\min_{x_t \in (\underline{x}^k, 1]} \{b(x_t, \mu, 1)\} - \max_{\bar{x}_t \in [0, \bar{x}^k)} \{b(\bar{x}_t, \mu, 2)\} \right) + \\ 2\mathbb{1}\{\bar{s}_i\}_t^{t+k-1} \bar{\eta} \end{aligned} \tag{15}$$

Now, by the definition of $\bar{\eta}$ we have that

$$\min_{x_t \in (\underline{x}^k, 1]} \{b(x_t, \mu, 1)\} - \max_{\bar{x}_t \in [0, \bar{x}^k)} \{b(\bar{x}_t, \mu, 2)\} \geq \bar{\eta} \tag{16}$$

Finally, combining 15 and 16 we get that

$$|b^k(x_t, \mu, \{s_i\}_t^{t+k}) - x_t| - |b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k}) - x_t| > k\bar{\eta}$$

□

Lemma 4. *Take $k \in \mathbb{N} \geq 2$. For any $\theta \in (0,1)$ there exists a $\frac{2-x_t}{\bar{\eta}} + 1 < k \leq \frac{2}{\bar{\eta}} + 1$ such that for any $x_t \geq \underline{x}^k$ and any two sequences of states of nature $\{s_i\}_{t+1}^{t+k-1}$ and $\{\bar{s}_i\}_{t+1}^{t+k-1}$ such that $\mathbb{1}\{s_i\}_{t+1}^{t+k-1} > \mathbb{1}\{\bar{s}_i\}_{t+1}^{t+k-1}$ and $s_i = \{1,2\} \smallsetminus \bar{s}_i$ for each $i \in \{t+1, \ldots, t+k-1\}$ we have that $b^k(x_t, \mu, \{s_t, \{s_i\}_{t+1}^{t+k-1}\}) + b^k(x_t, \mu, \{s_t, \{\bar{s}_i\}_{t+1}^{t+k-1}\}) < x_t - \bar{\eta}$.*

Proof. The fact that $\mathbb{1}\{s_i\}_{t+1}^{t+k-1} > \mathbb{1}\{\bar{s}_i\}_{t+1}^{t+k-1}$ and $s_i = \{1,2\} \smallsetminus \bar{s}_i$ for each $i \in \{t+1, \ldots, t+k-1\}$ with $s_t = \bar{s}_t$ implies that by Lemma 3,

$$|b^k(x_t, \mu, \{s_i\}_t^{t+k-1}) - x_t| - |b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) - x_t| > k\bar{\eta} \tag{17}$$

Finally, note that $b^k(x_t, \mu, \{s_i\}_t^{t+k-1}) - x_t \leq 0$. We now distinguish two cases.

Case 1. $b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) - x_t \geq 0$.
In this case by equation 17 we have that

$$b^k(x_t, \mu, \{s_i\}_t^{t+k-1}) + b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) \leq 2x_t - k\bar{\eta} < 2 - k\bar{\eta}$$

Take $k = \lceil \frac{2-x_t}{\bar{\eta}} + 1 \rceil$. With such k the result of the lemma in this case holds true.

Case 2. $b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) - x_t < 0$.
In this case we have that, proceeding as in the previous case,

$$b^k(x_t, \mu, \{s_i\}_t^{t+k-1}) + b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) < 2b^k(x_t, \mu, \{\bar{s}_i\}_t^{t+k-1}) - k\bar{\eta} < 2 - k\bar{\eta}.$$

Take $k = \lceil \frac{2-x_t}{\bar{\eta}} + 1 \rceil$. With such k the result of the lemma in this case holds true, which completes the proof.

□

Define the probability that the sequence of events $\{s_i\}_{t+1}^{t+k-1}$ takes place given the current state of nature s_t and a given value for θ as $P(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t)$. We can write the expected value of x_{t+k} as

$$E_t(x_{t+k}) = \sum_{\{s_i\}_{t+1}^{t+k-1}} P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right) b^k\left(x_t, \mu, \{s_t, \{s_i\}_{t+1}^{t+k-1}\}\right)$$

That is, the sum over all possible sequences of states of nature of the product between the probability that a particular sequence occurs times the value of x_{t+k} if that particular sequence happens.

Define $S_{t+1}^{t+k-1} = \left\{\{s_i\}_{t+1}^{t+k-1} \in \{1,2\}^{k-1} \text{ such that } s_{t+1} = s_t\right\}$.

Lemma 5. *Let s_t be given and let $\{s_i\}_{t+1}^{t+k-1}$ be a sequence of states of nature such that $s_{t+1} = s_t$. Define $\{\bar{s}_i\}_{t+1}^{t+k-1}$ with $\hat{s}_i = \{1,2\} \setminus s_i$. Then we have:*

1. $\frac{\theta}{1-\theta} P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right) = P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$
2. $\sum_{\{s_i\}_{t+1}^{t+k-1} \in S_{t+1}^{t+k-1}} P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right) = 1 - \theta$
3. $\sum_{\{s_i\}_{t+1}^{t+k-1} \notin S_{t+1}^{t+k-1}} P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right) = \theta$

Proof. We begin by proving the first equality. Note that, given the transition probabilities, for any sequence $\{s_i\}_{t+1}^{t+k-1}$ and any values s_t and θ, the term $P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ is a polynomial of the form $\theta^s(1-\theta)^{k-s}$ with $s \leq k$.

For proceeding with the proof, we are going to construct the terms of $P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ and $P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$. Because $s_{t+1} = s_t$, the first term in $P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ is $1 - \theta$. On the other hand, $s_{t+1} = s_t$ implies $s_t \neq \hat{s}_{t+1}$ which means that the first term in $P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ is θ.

If $s_{t+2} = s_t$, then $s_{t+2} = s_{t+1}$ and the second term in $P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ is again $1 - \theta$. Reasoning as before, $s_{t+2} = s_t$ implies $s_{t+2} = s_{t+1}$ which in turn implies $\hat{s}_{t+1} = \hat{s}_{t+2}$. Hence, the second term in $P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ is also $1 - \theta$.

On the other hand, if $s_{t+2} \neq s_t$ then $s_{t+2} \neq s_{t+1}$ and the second term in $P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ equals θ. Moreover, if $s_{t+2} \neq s_t$ then $s_{t+2} \neq s_{t+1}$ which implies $\hat{s}_{t+2} \neq \hat{s}_{t+1}$ and the second term in $P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ also equals θ. If we continue in this fashion we get that all the terms from the second to the kth in $P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ and $P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ are the same. The only term that differs is the first one. In the function $P\left(\{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ the first term equals $(1-\theta)$ and in the function $P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ the first term equals θ. Hence, we can write that $\frac{\theta}{1-\theta} P\left(s_t, \{s_i\}_{t+1}^{t+k-1}|\theta, s_t\right) = P\left(s_t, \{\hat{s}_i\}_{t+1}^{t+k-1}|\theta, s_t\right)$ which completes the proof of the first equality.

The second and third equality in the lemma follow easily. For the second one, note that,

$$\sum_{\{s_i\}_{t+1}^{t+k-1}\in S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)+\sum_{\{s_i\}_{t+1}^{t+k-1}\notin S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)=1.$$

Hence, since $\frac{\theta}{1-\theta}P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)=P\left(s_t,\{\hat{s}_i\}_{t+1}^{t+k-1}|\theta,s_t\right)$, we have that

$$\sum_{\{s_i\}_{t+1}^{t+k-1}\in S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)+\frac{\theta}{1-\theta}\sum_{\{s_i\}_{t+1}^{t+k-1}\in S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)=1.$$

The third equality in the lemma follows by proceeding as above. □

Using Lemma 5 we can the rewrite the expression for the expected value of x_{t+k} at time t as follows.

$$\begin{aligned}E_t(x_{t+k})=\sum_{\{s_i\}_{t+1}^{t+k}\in S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)\Big[b^k\left(x_t,\mu,\{s_t,\{s_i\}_{t+1}^{t+k-1}\}\right)\\ +\frac{\theta}{1-\theta}b^k\left(x_t,\mu,\{s_t,\{\hat{s}_i\}_{t+1}^{t+k-1}\}\right)\Big]\end{aligned}\tag{18}$$

Similarly,

$$\begin{aligned}E_t(x_{t+k})=\sum_{\{s_i\}_{t+1}^{t+k}\notin S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)\Big[\frac{1-\theta}{\theta}b^k\left(x_t,\mu,\{s_t,\{s_i\}_{t+1}^{t+k-1}\}\right)\\ +b^k\left(x_t,\mu,\{s_t,\{\hat{s}_i\}_{t+1}^{t+k-1}\}\right)\Big]\end{aligned}\tag{19}$$

Lemma 6. *For any* $\theta\in(0,1)$ *there exists a* $\frac{2-x_t}{\bar{\eta}}+1<k\leq\frac{2}{\bar{\eta}}+1$ *such that if* $x_t\geq\underline{x}^k$ *then* $E_t(x_{t+k})<\max\{(1-\theta),\theta\}(x_t-\bar{\eta})$.

Proof. From Lemma 5 we have that

$$\max\Big\{\sum_{\{s_i\}_{t+1}^{t+k-1}\in S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right),\sum_{\{s_i\}_{t+1}^{t+k}\notin S_{t+1}^{t+k-1}} P\left(s_t,\{s_i\}_{t+1}^{t+k-1}|\theta,s_t\right)\Big\}=\max\{(1-\theta),\theta\}.\tag{20}$$

Moreover, in the sequence $\{s_i\}_{t+1}^{t+k-1}\in S_{t+1}^{t+k-1}$ state 1 occurred more times than state 2 if and only if in the sequence $\{\hat{s}_i\}_{t+1}^{t+k-1}$ state 2 occurred more times than state 1. Hence, by Lemma 3, if $\theta\leq 0.5$ we have that

$$b^k\left(x_t,\mu,\{s_t,\{s_i\}_{t+1}^{t+k-1}\}\right)+\frac{\theta}{1-\theta}b^k\left(x_t,\mu,\{s_t,\{\hat{s}_i\}_{t+1}^{t+k-1}\}\right)<(x_t-\bar{\eta}).\tag{21}$$

On the other hand, if $\theta\geq 0.5$ we have that

$$\frac{1-\theta}{\theta}b^k\left(x_t,\mu,\{s_t,\{s_i\}_{t+1}^{t+k-1}\}\right)+b^k\left(x_t,\mu,\{s_t,\{\hat{s}_i\}_{t+1}^{t+k-1}\}\right)<(x_t-\bar{\eta}).\tag{22}$$

Therefore, if $\theta \leq (\geq) 0.5$ combining equations 18 (19), 20 and 21 (22) with Lemma 4 we get the following.

$$E_t(x_{t+k}) < \max\{(1-\theta), \theta\}(x_t - \bar{\eta})$$

□

Lemma 7. *Assume* $x_t \geq \underline{z}^k$ *for* $\frac{2-x_t}{\bar{\eta}}+1 < k \leq \frac{2}{\bar{\eta}}+1$*. For any* $m \in \mathbb{N}$ *such that* $(1-\theta)(x_t - m\bar{\eta}) \geq \underline{x}^k$ *we have that* $E_t(x_{t+mk}) < \max\{(1-\theta), \theta\}(x_t - m\bar{\eta})$*.*

Proof. The proof will be carried out assuming $1-\theta > \theta$. For the case were $1-\theta < \theta$ one just has to follow the same steps.

We proceed by computing first the value of $E_t(x_{t+2k})$. In this case we have the following.

$$E_t(x_{t+2k}) = \sum_{\{s_i\}_{t+1}^{t+2k}} P\left(s_t, \{s_i\}_{t+1}^{t+2k-1} | \theta, s_t\right) b^{2k}\left(x_t, \mu, \{s_t, \{s_i\}_{t+1}^{t+2k-1}\}\right)$$

For any sequence $\{s_i\}_{t+1}^{t+2k-1}$ define $\{s_i^1\}_{t+1}^{t+k-1}$ as the sequence with the first $k-1$ states of nature in the sequence $\{s_i\}_{t+1}^{t+2k-1}$. Similarly, define $\{s_i^2\}_{t+k}^{t+2k-1}$ as the sequence with the states of nature from the kth to the $(2k-1)$th in the sequence $\{s_i\}_{t+1}^{t+2k-1}$. Then we have the following.

$$\begin{aligned} E_t(x_{t+2k}) \quad < \quad & \sum_{\{s_i^2\}_{t+k}^{t+2k-1}} \max_{\hat{s}_{t+k}^2 \in \{1,2\}} P\left(\hat{s}_{t+k}^2, \{s_i^2\}_{t+k+1}^{t+2k-1} | \theta, \hat{s}_{t+k}^2\right) \\ & \left[\sum_{\{s_i^1\}_{t+1}^{t+k-1}} P\left(s_t, \{s_i^1\}_{t+1}^{t+k-1} | \theta, s_t\right) b^k\left(b^k(x_t, \mu, \{s_t, \{s_i^1\}_{t+1}^{t+k-1}\}), \mu, \{2, \hat{s}_{t+k}, \{s_i^2\}_{t+k+1}^{t+2k-1}\}\right)\right] \end{aligned}$$

Note that $\sum_{\{s_i^2\}_{t+k}^{t+2k-1}} \max_{\hat{s}_{t+k}^2 \in \{1,2\}} P\left(\hat{s}_{t+k}^2, \{s_i^2\}_{t+2}^{t+2k-1} | \theta, \hat{s}_{t+k}^2\right) = 1$. This is due to the fact it is the sum of the probabilities in all possible scenarios, whether we start at $\hat{s}_{t+k}^2 = 1$ or at $\hat{s}_{t+k}^2 = 2$. Define $X_t(k) = \{x_t$ such that $x_t \in [b^k(x_t, \mu, \{1\}^k), b^k(x_t, \mu, \{2\}^k)]\}$. The equation above implies the following.

$$\begin{aligned} E_t(x_{t+2k}) \quad < \quad & \sum_{\{s_i^2\}_{t+k}^{t+2k-1}} \max_{\hat{s}_{t+k}^2 \in \{1,2\}} P\left(\hat{s}_{t+k}^2, \{s_i^2\}_{t+k+1}^{t+2k-1} | \theta, \hat{s}_{t+k}^2\right) \\ & \left[E_t(x_{t+k}) + \max_{x_t \in X_t(k)}\left(b^k(x_t, \mu, \{\hat{s}_{t+k}^2, \{s_i^2\}_{t+k+1}^{t+2k-1}\}) - x_t\right)\right] \end{aligned}$$

Define $\hat{x}_t(k) = \arg\max_{x_t \in X_t(k)}\left(b^k(x_t, \mu, \{2, \{s_i^2\}_{t+1}^{t+2k}\}) - x_t\right)$. Applying lemma 6 to the equation above we get that.

$$E_t(x_{t+2k}) < (1-\theta)(x_t - \bar{\eta}) + [(1-\theta)(\hat{x}_t(k) - \bar{\eta}) - \hat{x}_t(k)]$$

Which since $(1-\theta)(\hat{x}_t(k) - \bar{\eta}) - \hat{x}_t(k) < -(1-\theta)\bar{\eta}$ implies the following.

$$E_t(x_{t+2k}) < (1-\theta)(x_t - 2\bar{\eta})$$

Showing the result of the lemma for $E_t(x_{t+3k})$, $E_t(x_{t+4k})$, etc. can be done proceeding in a similar fashion as above. The only difference is in the step where any sequence of events is divided into two. In general, $\{s_i^1\}_{t+1}^{t+(m-1)k-1}$ has to include the first $(m-1)k-1$ terms of $\{s_i\}_{t+1}^{t+mk}$ and $\{s_i^2\}_{t+(m-1)k}^{t+mk-1}$ has to include the latest k terms of $\{s_i\}_{t+1}^{t+mk}$. □

Lemma 8. *For all $x_t \in [\tilde{z}, 1]$ we have that*

$$\lim_{h\to\infty} E_t(x_{t+h}) \leq \underline{x}^k.$$

Proof. Given any $h > \frac{1}{\bar{\eta}}\left(1 - \frac{\underline{x}^k}{\max\{1-\theta,\theta\}}\right)k$ and $g < h$, we have that

$$\begin{aligned}
E_t(x_{t+h}) &\leq E_t(x_{t+h}|s_t = 2) \\
&\leq E_{t+1}(x_{t+h}|s_t = 2, s_{t+1} = 2) \\
&\leq \dots \\
&\leq E_{t+h-g}(x_{t+h}|\{s_i\}_t^{t+h-g} = \{2\}^{h-g+1}) \\
&\leq E_{t+h-g}(x_{t+h}|z_{t+h-g} = 1, s_{t+h-g} = 2) \\
&= E_t(x_{t+g}|z_t = 1, s_t = 2).
\end{aligned}$$

The last equality is due to the time invariant distribution of x_t. Fix $g = mk$ with $m = \left\lceil \frac{1}{\bar{\eta}}\left(1 - \frac{\underline{x}^k}{\max\{1-\theta,\theta\}}\right)\right\rceil$ and $k = \frac{2}{\bar{\eta}} + 1$. By the Lemma 7 we have that $E_t(x_{t+g}|x_t = 1, s_t = 2) < \max\{(1-\theta),\theta\}(x_t - m\bar{\eta})$. Which given the value of m implies that $E_t(x_{t+g}|z_t = 1, s_t = 2) < \underline{x}^k$. Hence $E_t(x_{t+h}) < \underline{x}^k$ when $h > mk$. If we take limits when h goes to infinity on both sides we the get that

$$\lim_{h\to\infty} E_t(x_{t+h}) < \underline{x}^k.$$

□

Lemma 9. *For any $k \in \mathbb{N}$ and any $\varepsilon > 0$ there exists a $\bar{\mu} > 0$ such that if $\mu < \bar{\mu}$ then $\underline{x}^k < \tilde{y} + \varepsilon$.*

Proof. By definition, $\underline{x}^k$ is the infimum x such that after k periods the process is till on the right of $\tilde{y}$ (or just reached $\tilde{y}$). By Facts 4 and 5, this value goes to $\tilde{z}$ as μ goes to 0. □

Lemma 10. *For any $\varepsilon > 0$ there exists $\bar{\mu} > 0$ such that if $\mu < \bar{\mu}$ then*

$$\lim_{h\to\infty} E_t(x_{t+h}) \leq \tilde{y} + \varepsilon.$$

Proof. Follows from lemmas 8 and 9 above. □

The lemma above can be rewritten as $\lim_{h\to\infty}\lim_{\mu\to 0} E_t(x_{t+h}) \leq \tilde{y}$. Which concludes the first part of the proof that $\lim_{h\to\infty}\lim_{\mu\to 0} E_t(y_{t+h}) \to \tilde{y}$. Next, we have to show that $\lim_{h\to\infty}\lim_{\mu\to 0} E_t(w_{t+h}) \geq \tilde{y}$. This is straightforward once one notes the fact that our analysis above can be applied to state that $\lim_{h\to\infty}\lim_{\mu\to 0} E_t(-w_{t+h}) \leq -\tilde{y}$. This considerations yield to the following lemma.

Lemma 11. *For any $\varepsilon > 0$ there exists a $\bar{\mu} \in [0, \hat{\mu}]$ such that if $\mu < \bar{\mu}$ then,*

$$\lim_{h \to \infty} \left| E_t(y_{t+h}) - \tilde{y} \right| < \varepsilon.$$

Finally, from Lemma 11 one can easily get to the desired result.

Proposition 3. *For any $\varepsilon > 0$ there exists a $\bar{\mu} \in (0, \hat{\mu}]$ and $\bar{h}$ such that for $\mu < \bar{\mu}$ and $h > \bar{h}$ we have that*

$$P(|y_{t+h} - \tilde{y}| > \varepsilon) = 0.$$

Proof. We know by Lemma 11 that for any $\varepsilon > 0$, as μ collapses to zero then $\lim_{h \to \infty} \left| E_t(z_{t+h}) - \tilde{z} \right| < \varepsilon$. By Facts 4 and 5, μ going to zero implies that the statistical variance of z goes to zero as well. Hence, given that if $\mu \to 0$ then $\lim_{h \to \infty} E_t(z_{t+h}) \to \tilde{z}$ and $Var(z) \to 0$, we must have that once the process as been undergoing for a sufficiently long amount of time, $z = \tilde{z}$ almost surely. □

Now one just has to combine Lemma 1 with Proposition 5 to get the result in Proposition 1.

Proof of Proposition 2

Whenever the process is arbitrarily close to $z = 0$, we can use the same reasoning as Ellison and Fudemberg (1995). That is, for z_t close to zero we have that,

$$z_{t+1,j} = z_t(1 + \pi_{2j} - \pi_{1j}) + o(z_t).$$

Define $\gamma_j = 1 + \pi_{2j} - \pi_{1j}$ for al $j \in \{1, \ldots, m\}$. Hence we have that $\gamma_i < 1 < \gamma_j$ for $i \leq h$ and $j > h$.

Lemma 12. *For any z, $\varepsilon > 0$ there exists a $z_t < z$ and a $\bar{t} > 0$ even such that*

$$P\left(\left| z_{t+1} - z_t \Pi_{j=1}^{m} \gamma_j^{\mu_j} \right| > \varepsilon \right) = 0$$

for $t > \bar{t}$.

Proof. Follows from the Law of Large numbers for Markov Chains: The long run distribution of the states of nature puts weight μ_j to state j. Hence, as the number of realizations of states increases, the probability of observing any sequence in which the frequency of state j is not μ_j shrinks to zero. Therefore, if the process evolves during a sufficient amount of time and z_t is sufficiently close to $z = 0$, the result follows. □

Lemma 13. *The process can not converge to 0 if*

$$\sum_{j=1}^{m} \mu_j \log \gamma_j > 0.$$

There is a positive probability that the process converges to 0 if

$$\sum_{j=1}^{m} \mu_j \log \gamma_j < 0.$$

Proof. Reasoning as in the proof of Lemma 1 in Ellison and Fudemberg (1995), the variable z can converge to zero if and only if the variable $y = \log z$ can converge to $-\infty$. Using again the proof from Lemma 1 in Ellison and Fudemberg (1995) and Lemma 12 in this appendix, the variable y can converge to $-\infty$ only if $\sum_{j=1}^{m} \mu_j \log \gamma_j < 0$. The result follows. □

For studying the situation in which the process is arbitrarily close to 1, we proceed as follows. First, we define $x_t = 1 - z_t$. Then we apply the analysis above to the variable x_t. We have the following,

$$x_{t+1} \simeq x_t \Pi_{j=1}^{m} \hat{\gamma}_j^{\mu_j}$$

However, this time we have that $\hat{\gamma}_j = 1 - \pi_{2j} + \pi_{1j}$. Hence,$\gamma_i > 1 > \gamma_j$ for $i \leq h$ and $j > h$. An analogous to the lemma above for z close to 0 is the following when z is close to 1.

Lemma 14. *The process z can not converge to 1 if*

$$\sum_{j=1}^{m} \mu_j \log \hat{\gamma}_j > 0.$$

There is a positive probability that the process z converges to 1 if

$$\sum_{j=1}^{m} \mu_j \log \hat{\gamma}_j < 0.$$

Summing up our results from the previous lemmas the results in Proposition 2 follows.