Stochastic stability on general state spaces[1]

Jonathan Newton [2]

School of Economics, University of Sydney

Abstract

This paper studies stochastic stability methods applied to processes with a general state space. This includes settings in which agents repeatedly interact and choose from an uncountable set of strategies. Dynamics exist for which the stochastically stable states differ to those of any reasonable finite discretization. When there are a finite number of rest points of the unperturbed dynamic, sufficient conditions for results from the finite state space literature are derived and studied. Illustrative examples are given.

Keywords: Learning, stochastic stability, general state space
JEL: C71, C72, C73

1. Introduction

The occurence of social learning and the convergence of agents' behavior via processes of adaptive behavior is well-documented within economics (e.g. Chong et al., 2006; Selten and Apesteguia, 2005). The possibility of multiple resting points for such processes naturally leads one to question which of these steady states is more plausible from an economic perspective. Strongly influenced by evolutionary game theory (Smith and Price, 1973), a literature has grown that analyses the robustness of steady states of social learning dynamics to random errors made by players in their choice of action (Kandori et al., 1993; Young, 1993a). These ideas have been applied to a variety of economic situations, including bargaining (Young, 1993b), Nash demand games (Agastya, 1997), exchange economies (Serrano and Volij, 2008), local interaction on networks and the persistence of altruistic behavior (Eshel et al., 1998).

A common approach when assessing the robustness of steady states of social learning dynamics has been that pioneered by Kandori et al. (1993) and Young (1993a). Agents are assumed to make errors independently and when they do make an error are assumed to play a strategy chosen at random from a distribution with full support on a finite set of strategies. This imposes a mathematical structure on the process that leads to clear and appealing characterization results.

Unfortunately, such results cannot be straightforwardly applied when agents' have non-finite sets of strategies.[3] Even assuming the convergence of the underlying social learning dynamic, the addition of random errors can lead to behavior which hinders efforts to obtain a clear cut characterization of the long run pattern of play. This paper takes up the task of analysing the problems and intricacies which arise and, when there are a finite number of rest points of the underlying dynamic, determines a set of sufficient conditions which enable existing results to be applied to models with continuous state spaces. These conditions include a continuity requirement on error distributions, continuity of players' responses as a function of the current state, an asymptotic stability condition and a condition which ensures a specific

[1]This paper has been and is being extensively revised, but it hasn't been cleaned up yet, so apologies for all errors. Please do not cite without permission of the author.

[2]Come rain or come shine, I can be reached at jonathan.newton@sydney.edu.au

[3]An early paper in the literature (Foster and Young, 1990) has an infinite state space and uses Gaussian error distributions. However, it differs markedly from the majority of the literature, in which the error distributions are irrelevant to the stability results as long as they have full support.

type of discontinuity does not occur at rest points of the underlying dynamic. Examples are given showing how no subset of the conditions is sufficient on its own.

Fortunately, all of these conditions are satisfied for many common models found in economics. Typical error distributions of the kind described above coupled with the continuous best responses found in many models of industrial organization will often satisfy all of the conditions. This study applies the theory to linear quadratic games and to population models in the style of Kandori et al. (1993).

A related paper is that of Feinberg (2006) which also looks at discrete time, continuous state space processes. However, the paper in question imposes the strong condition that the perturbed process be governed by transition probabilities that are continuous functions of the current state of the process. The bulk of the analysis in the current paper concerns situations where this is not the case. Schenk-Hoppe (2000) adapts the results of Freidlin and Wentzell (1998) and Ellison (2000) for use in finding stochastically stable states in a continuous strategy oligopoly model equipped with an imitation dynamic.

The paper is organized as follows. Section 2 introduces the ideas of the paper via two motivating examples. Section 3 describes the processes of interest, gives convergence results, looks at transition probabilities between steady states, and defines a useful regularity property, showing how this property allows the problems associated with non-finite state spaces to be circumvented. Section 4 gives sufficient conditions for this property to hold and discusses each of the conditions, giving examples of the problems which arise if any condition fails to hold. Section 5 gives examples. Section 6 solves an example from section 2 for which our regularity condition fails to hold. Section 7 concludes.

2. Motivating examples

This paper focuses on situations where agents follow some rule when deciding how to behave. The rule can be deterministic or random, cautious or hasty, imitative or best responding. Usually the rule is adaptive in the sense that an agent's behavior is intended to improve his lot. What really matters is that the rule has the Markov property: the past per se does not affect the future, although features of the present shaped by the past, including memories, are allowed to do so. We analyse situations where behavior over time will converge towards one of a number of steady states. As long as there is some probability of convergence to more than one steady state, this is predictively awkward. The possibility of random errors in play justifies the introduction of perturbed versions of the process which help in obtaining long run predictions. There is a well-developed literature which deals with these problems for finite state spaces[4], so the first question that must be addressed is whether there is benefit to be had from dealing directly with processes on general state spaces directly, rather than with finite discrete approximations.

2.1. Discretization can fail to represent the original process accurately

There is not always a suitable finite discretization of a process available. To illustrate, we present the following example. Consider a Markov process with state space $X = [-1, 1] \subset \mathbb{R}$. Let the process be governed by the Markov kernel $P(.,.)$. For notational ease, for $y \in X$, we identify $P(., y) := P(., \{y\})$. Let $P(.,.)$ be as follows:

$$\begin{aligned} P\left(x, \frac{x}{2}\right) &= 1 - |x| \\ P(x, 1) &= \max\{0, x\} \\ P(x, -1) &= \max\{0, -x\}. \end{aligned}$$

This process has a set of stable states $\Lambda = \{-1, 0, 1\}$: from $x^* \in \Lambda$, $P(x^*, x^*) = 1$. We examine a perturbed variant of the process in which each period, with probability $1 - \varepsilon$ the unperturbed process is followed, and with probability ε the new state is drawn from the uniform distribution $U[[-1, 1]]$. This perturbed process

[4]See Kandori et al. (1993); Young (1993a); Bergin and Lipman (1996); van Damme and Weibull (2002).

has an invariant distribution π_ε which converges to a distribution with all weight on $\{-1, 1\}$ as $\varepsilon \to 0$: the set of *stochastically stable* states is $\{-1, 1\}$.

Any discretized state space and process should satisfy some properties in order for it to be a reasonable representation of the original process. We suggest the following as reasonable restrictions on the discretized state space $X_d \subseteq X$ and the discretized unperturbed process $P_d(.,.)$:

$$(a) \quad \forall x \in X_d, A \subseteq X : \quad P^t(x, A) > 0 \implies P_d^t(x, \{z \in X_d : z \in \arg\min_{y \in X_d} d(y, A)\}) > 0.$$

$$(b) \quad \forall x : \quad P^t(x, \{\tilde{x} \in X : z \in \arg\min_{y \in X_d} d(y, \tilde{x})\}) = 0 \, \forall t \implies P_d^t(x, z) = 0 \, \forall t.$$

$$(c) \quad \Lambda \subseteq X_d.$$

In words, these conditions are that: (a) From a state $x \in X_d$, if a set $A \subseteq X$ is reached with positive probability under the original process, then the closest states to A in X_d (under the original metric) are reached with positive probability under the discretized process $P_d(.,.)$; (b) If, from a state $x \in X_d$, under the original process the set of states in X which are closer to $z \in X_d$ than to any other point in X_d is never reached with positive probability, then z is never reached with positive probability under the discretized process; (c) Stable states of the original process are states of the discretized process and therefore stable states of the discretized process by (a) and (b).

We take as a discretization of the perturbation (the uniform distribution on X) any distribution on X_d that places positive probability on at least one state in $(-1, 0)$ and at least one state in $(0, 1)$. The uniform distribution on X_d would satisfy this requirement. Now, as $\varepsilon \to 0$, the limit of π_ε places positive probability on all states in $\{-1, 0, 1\}$: discretizing the process has given us one additional stochastically stable state.

Finding the stochastically stable states of the original process in this section turns out to be simple. The reason for this is that far enough along any convergent path to a stable state, the probability under the perturbed process of moving to the basin of attraction of another given stable state is of constant order of ε. For example, any convergent path to -1 under the unperturbed process $P(.,.)$ eventually reaches -1, from where in any given period, say period t, the probability under the perturbed process of the next (from period $t+1$ onwards) basin of attraction of a stable state entered being that of state 0 is of order $\varepsilon^\infty = 0$. The probability of it being that of state 1 is of order ε. There do not exist convergent paths with different escape probabilities. We refer to this as Property C. When Property C holds, we show that variants of several results used heavily in the finite state space stochastic stability literature can be used. An important part of the current paper gives sufficient conditions under which Property C holds.

2.2. Multiple convergent paths

The next example can be considered as a model in which there are two possible focal points for a social norm. There are n agents who contribute some real amount. If at least some threshold number of agents contribute at least some focal amount (which we take to be 1) then agents converge towards that level of contribution. Otherwise they converge towards a zero contribution. Consider a state space $X = [0, x_{max}]^n \subset \mathbb{R}_+^n$, $x_{max} > 1$, $n \in \mathbb{N}$. For some $k < n$, $k \in \mathbb{N}$ define for all $x \in X$:

$$I(x) = \{i \in \{1, \ldots, n\} : x_i \geq 1\}.$$

Define $P(.,.)$ as follows:

$$\text{If} \quad |I(x)| < k \quad \text{then} \quad P\left(x, \frac{x}{2}\right) = 1$$

$$\text{If} \quad |I(x)| \geq k \quad \text{then} \quad P\left(x, \frac{x+1}{2}\right) = 1.$$

The process has stable states $\Lambda = \{0^n, 1^n\}$. We examine a perturbed variant of the process in which each period, with probability $1 - \sum_{r=1}^n \varepsilon^i$ the unperturbed process is followed, and with probability ε^r the new state is drawn from the distribution:

$$G_r(x,.) \sim U[\{\bar{x} \in X : (|\{i : x_i = \bar{x}_i\}| = n - r)\}].$$

That is, with probability ε^r, exactly r agents randomly choose their contribution from a uniform distribution on $[0, x_{max}]$. There exist convergent paths to 1^n with $|I(x)| \in \{k, k+1, \dots n\}$. From a convergent path to 1^n for which $|I(x)| = k$, a move to the basin of attraction of 0^n in a single period is an event with probability of order ε. From a convergent path to 1^n for which $|I(x)| = n$, a move to the basin of attraction of 0^n in a single period is an event with probability of order ε^{n-k+1}. Property C does not hold. The only stochastically stable state of this process turns out to be 0^n. Showing that this is the case is complicated by Property C not holding, so recourse to more general methods is necessary (see section 6). For completeness, we note that if $k < n - k + 1$, then any reasonable finite discretization of the process leads to 1^n being selected as the unique stochastically stable state.

3. A general model of perturbed adaptive behavior

A very general model is presented: the unperturbed dynamic can be any Markov process on any metric space, the only assumption being the nonemptiness of, and convergence of the process to, a set of stable states. The perturbed model then allows a broad class of perturbations which includes independent random errors such as are found in the traditional stochastic stability literature, but also allows correlated errors and any type of state dependent behavior.[5]

3.1. Quantitative characterization

The first step is to model an unperturbed dynamic which gives the behavior of agents in the absence of random errors. Let Φ be a Markov process on a metric space X with kernel $P(x, A)$, $x \in X$, $A \in \mathcal{B}(X)$, where $\mathcal{B}(X)$ is the Borel σ-algebra.

Definition. The set of *stable* states is defined:

$$\Lambda := \{x \in X : P(x, x) = 1\}.$$

Assume $\Lambda \neq \emptyset$.

Definition. The *basin of attraction* W_i of $x_i^* \in \Lambda$ is:

$$W_i := \left\{x \in X : \text{ for every open } V \supset \{x_i^*\} : P^t(x, V) \to 1 \text{ as } t \to \infty\right\}.$$

Define $W := \bigcup_i W_i$. We now introduce an assumption which guarantees that wherever you start in the state space you end up arbitrarily close to some element of Λ — the unperturbed process Φ converges to a stable set. This assumption is necessary to the purpose of this paper, which is to give tools by which to select from several stable states. If convergence were not assumed, then equilibrium selection would become a secondary issue.

Convergence assumption

$$\forall x_0 \in X : \ P^t(x_0, W) \to 1 \text{ as } t \to \infty$$

Definition. The *basin of possible attraction* $\mathcal{W}_i$ of Λ_i is:

$$\mathcal{W}_i := \left\{ x \in X : \sum_{t=1}^{\infty} P^t(x, W_i) > 0 \right\}$$

[5] For example, perturbations could include coalitional behavior such as that found in Newton (2012).

Φ can be taken to represent some unperturbed dynamic which describes the evolution of the strategies of players in a game, with each entry in a vector $x \in X$ describing the strategy chosen by a player in some previous time period. In this context the kernel $P(x, A)$ can be taken to be some sort of (not necessarily continuous) best response or imitation dynamic. Let $\{\Phi_\varepsilon\}_\varepsilon$ be a family of Markov processes on the state space X with kernels $P_\varepsilon(x, A)$. Define:

$$P_\varepsilon(x, A) = \left(1 - \sum_{i=1}^{M} \varepsilon^i\right) P(x, A) + \sum_{i=1}^{M-1} \varepsilon^i G_i(x, A) + \varepsilon^M G_M(A)$$

where $M \in \mathbb{N}$, $G_i(x, .)$, $G_M(.)$ are probability measures on $\mathcal{B}(X)$. For $A \in \mathcal{B}(X)$, $G_i(., A)$ are non-negative $\mathcal{B}(X)$-measurable functions on X. As a sum of $\mathcal{B}(X)$-measurable functions, $P_\varepsilon(., A)$ is $\mathcal{B}(X)$-measurable. Note that $P_\varepsilon(x, X) = 1$ is satisfied. $\{\Phi_\varepsilon\}_\varepsilon$ is a subset of the class of Puiseux Markov processes. For more on Puiseux Markov processes see Appendix A.[6]

In the most common models of stochastic stability, which we refer to as *independent error* models (Young, 1993a; Kandori et al., 1993), the state is composed of strategy profiles, each player has an independent probability of making an error with probability ε and players who make errors play strategies chosen from a given distribution with full support. Such a process satisfies the definition above.[7] We now show that for any given value of ε, Φ_ε has a unique invariant measure. This measure is predictive in the sense that it gives the frequencies with which given sets of states will be observed in the long run.

Definition. A measure φ on $\mathcal{B}(X)$ is an *irreducibility measure* and Φ_ε is *φ-irreducible* if for all $x \in X$, whenever $\varphi(A) > 0$, there exists some $t > 0$, possibly depending on both A and x, such that $P^t_\varepsilon(x, A) > 0$.

Definition. A set $A \in \mathcal{B}(X)$ is *petite* if there exists a nontrivial measure ν and a probability distribution a on $\mathbb{Z}_+$ such that $\forall x \in A$:

$$\sum_t a(t) P^t_\varepsilon(x, .) \geq \nu(.)$$

Proposition 1. *Φ_ε has a unique invariant probability measure π_ε.*

Proof. $G_M(.)$ is an irreducibility measure on $\mathcal{B}(X)$ as for any $A \in \mathcal{B}(X)$ with $G_M(A) > 0$ we have $P_\varepsilon(x, A) \geq \varepsilon^M G_M(A) > 0$ for all $x \in X$. Letting $a(1) = 1$, $\nu(.) = \varepsilon^M G_M(.)$ we see that the set X is petite as for all $x \in X$: $P_\varepsilon(x, .) \geq \varepsilon^M G_M(.) = \nu(.)$. As X is the entire state space, $\tau_X := \min\{t \geq 1 : \Phi^t_\varepsilon \in X\} \equiv 1$. Combined with irreducibility and petiteness of X this implies the existence of a unique invariant probability measure π_ε for Φ_ε.[8][9] □

Proposition 2. *For all $x \in X$:*

$$\sup_{A \in \mathcal{B}(X)} |P^t_\varepsilon(x, A) - \pi_\varepsilon(A)| \to 0 \qquad \textit{as } t \to \infty.$$

Proof. As $G_M(.)$ is state-independent, the process is aperiodic. This and the uniqueness and finiteness of π_ε imply the result.[10] □

[6] Setting $\varepsilon = e^{\frac{-1}{\eta}}$, the processes in this paper satisfy: $\lim_{\eta \to 0} \eta P^\eta(x, U) = -\inf_{y \in U} \rho(x, y)$ for a function $\rho(., .) : X \to X$, $\rho(., .)$ is not necessarily continuous and so does not necessarily satisfy the conditions of Kifer (1990). For example, independent error models do not satisfy continuity of $\rho(., .)$. Despite this, much of the analysis that follows is similar to that of Kifer (1990).

[7] Independent error models do not have $P_\varepsilon(x, A)$ continuous in x for any given $A \in \mathcal{B}(X)$ and so do not satisfy the assumptions of Feinberg (2006).

[8] Meyn and Tweedie (1994), Theorem 3.2.

[9] This implies Lemma 2 of Schenk-Hoppe (2000). Similarly, Proposition 3 and its corollary imply Lemma 3 of Schenk-Hoppe (2000).

[10] Meyn and Tweedie (2009), Theorem 13.0.1.

3.2. Stationary distributions with small error probabilities

In the preceding subsection it was shown that Φ_ε has a unique invariant measure π_ε which depends on ε. In order to predict long term behavior under small error probabilities it helps to analyse the limit as $\varepsilon \to 0$. In the following proposition it is shown that as ε gets small, π_ε places arbitrarily small probability mass on sets of states which are not close to the steady states of the unperturbed process. Because all of our measures have $\mathcal{B}(X)$ as their domain and all our functions are $\mathcal{B}(X)$-measurable, we can use the property that an integral over a sum of measures is equal to the sum of integrals over those measures.[11]

Proposition 3. *For any $\eta \in (0,1)$, $A \in \mathcal{B}(X)$ such that $\Lambda \cap cl(A) = \emptyset$ there exists $\hat{\varepsilon}$ such that $\pi_\varepsilon(A) < \eta$ for all $\varepsilon < \hat{\varepsilon}$.*

Proof. Let $\lambda(.)$ be an arbitrary probability measure on $\mathcal{B}(X)$. By the convergence assumption there exists $K \in \mathbb{N}$ such that for all $k > K$:

$$\int_X P^k(x,A)\lambda(dx) < \frac{\eta}{2}$$

so for $k > K$:

$$\int_X P_\varepsilon^k(x,A)\lambda(dx) = \int_X \int_X P_\varepsilon(x,dy)P_\varepsilon^{k-1}(y,A)\lambda(dx)$$

$$= \int_X \int_X \left(1 - \sum_{i=1}^M \varepsilon^i\right) P(x,dy)P_\varepsilon^{k-1}(y,A)\lambda(dx) + \dots$$

$$= \int_X \int_X \dots \int_X \left(1 - \sum_{i=1}^M \varepsilon^i\right)^k P(x,dy_1)P(y_1,dy_2)\dots P(y_{k-2},dy_{k-1})P(y_{k-1},A)\lambda(dx) + \dots$$

$$= \int_X \left(1 - \sum_{i=1}^M \varepsilon^i\right)^k P^k(x,A)\lambda(dx) + \dots$$

$$< \int_X P^k(x,A)\lambda(dx) + \int_X \left(1 - \left(1 - \sum_{i=1}^M \varepsilon^i\right)^k\right)\lambda(dx)$$

$$= \int_X P^k(x,A)\lambda(dx) + 1 - \left(1 - \sum_{i=1}^M \varepsilon^i\right)^k < \frac{\eta}{2} + 1 - \left(1 - \sum_{i=1}^M \varepsilon^i\right)^k$$

which is less than η for small enough ε as $\lim_{\varepsilon\to 0} \varepsilon^i = 0$ for all i. □

Corollary 1. *A limiting distribution (in the sense of weak convergence of measures) of π_ε exists: $\pi_\varepsilon \Rightarrow \pi$. π is an invariant distribution of the unperturbed process Φ.*

Proof. The proof of existence of a limit is relegated to Appendix A. By Proposition 3 the only states in X which can have positive probability in such a limit are in Λ. □

So if the invariant distributions of the perturbed process converge then they converge to a particular invariant distribution of the unperturbed process. The addition of the random errors to the model can be seen as a way of selecting between the alternative invariant distributions of the unperturbed process. We make the following assumption.

[11] See Fremlin (2001), 234H(c).

Assumption

$|\Lambda| < \infty$

The case of $|\Lambda| = \infty$ is analysed in other work.[12]. Note that $|\Lambda| < \infty$ implies that $\Lambda = \{x_1^*, \ldots x_J^*\}$ for some $J \in \mathbb{N}$. The invariant distribution in many examples turns out to place all of the probability mass on a single steady state, predicting that in the long run the process should be observed to be at or near that state almost all of the time.

Definition. Steady states $x^* \in X$ with $\pi(x^*) > 0$ are called *stochastically stable.*

The rest of the paper devotes itself to the question of how to find stochastically stable states and the analysis of intricacies that can arise due to having an non-finite state space.

3.3. Transition probabilities between steady states

In order to find stochastically stable states it will be necessary to determine the magnitudes of the transition probabilities between the basins of attraction of different steady states. These magnitudes are given as powers of ε and it will be necessary to use Bachmann-Landau asymptotic notation. In words, the notation expresses the ideas of f being bounded by g below, above, and both above and below respectively:

$$f(\varepsilon) \in \Omega(g(\varepsilon)) \Leftrightarrow \exists k > 0, \bar{\varepsilon} \text{ s.t. } \forall \varepsilon < \bar{\varepsilon} : kg(\varepsilon) \leq |f(\varepsilon)|$$

$$f(\varepsilon) \in O(g(\varepsilon)) \Leftrightarrow \exists k > 0, \bar{\varepsilon} \text{ s.t. } \forall \varepsilon < \bar{\varepsilon} : |f(\varepsilon)| \leq kg(\varepsilon)$$

$$f(\varepsilon) \in \Theta(g(\varepsilon)) \Leftrightarrow \exists k_1, k_2 > 0, \bar{\varepsilon} \text{ s.t. } \forall \varepsilon < \bar{\varepsilon} : k_1 g(\varepsilon) \leq |f(\varepsilon)| \leq k_2 g(\varepsilon)$$

Define:

$$\tau_A := \min\{t \geq 1 : \Phi_\varepsilon^t \in A\}, \qquad T_\varepsilon(x, W_j) := Pr_x(\tau_{W_j} < \tau_{W_i} \ \forall i)$$

where Pr_x denotes the probability of events conditional on the chain beginning with $\Phi_\varepsilon^0 = x$. τ_A is the time of first return of the process to a set A. $T_\varepsilon(x, W_j)$ should be interpreted as the probability that, starting from state x, the process Φ_ε enters the basin of attraction (under Φ) of x_j^* before it enters the basin of attraction of any other stable state.

Define $W_j^\delta = W_j \cap B_\delta(x_j^*)$, $\mathcal{W}_j^\delta = \mathcal{W}_j \cap B_\delta(x_j^*)$:

$$V^\delta(x_k^*, x_j^*) := \min\left\{ i : \lim_{\varepsilon \to 0} \frac{T_\varepsilon(x, W_j)}{\varepsilon^i} > 0 \ \forall x \in W_k^\delta \right\} \wedge \infty$$

$$V^{\delta-}(x_k^*, x_j^*) := \min\left\{ i : \exists x \in W_k^\delta : \lim_{\varepsilon \to 0} \frac{T_\varepsilon(x, W_j)}{\varepsilon^i} > 0 \right\} \wedge \infty$$

which can be interpreted as resistances measuring the difficulty of moving from the basin of attraction of x_k^* to the basin of attraction of x_j^*. $V^\delta(x_k^*, x_j^*)$ and $V^{\delta-}(x_k^*, x_j^*)$ measure respectively the most unlikely and easiest way in which a move from W_j^δ to W_k^δ could occur without passing through W_i for any $i \neq j, k$.

Note that $\forall\, x \in W_k^\delta$:

$$T_\varepsilon(x, W_j) \in \Omega(\varepsilon^{V^\delta(x_k^*, x_j^*)}), \qquad T_\varepsilon(x, W_j) \in O(\varepsilon^{V^{\delta-}(x_k^*, x_j^*)})$$

[12] Schenk-Hoppe (2000) is a good introduction to the ideas involved.

Define limiting resistances as $\delta \to 0$:

$$V(x_k^*, x_j^*) := \lim_{\delta \to 0} V^\delta(x_k^*, x_j^*), \qquad V^-(x_k^*, x_j^*) := \lim_{\delta \to 0} V^{\delta-}(x_k^*, x_j^*)$$

These limits exist as V^δ and $V^{\delta-}$ are monotonic (decreasing and increasing respectively) as $\delta \to 0$ and $V^\delta, V^{\delta-} \in \{1, \ldots, M, \infty\}$ so there must exist $\hat{\delta} > 0$ so that for all $\delta < \hat{\delta}$:

$$V^\delta(x_k^*, x_j^*) = V(x_k^*, x_j^*), \qquad V^{\delta-}(x_k^*, x_j^*) = V^-(x_k^*, x_j^*)$$

Definition (Property C) *Property C is said to hold when* $V(x_k^*, x_j^*) = V^-(x_k^*, x_j^*)$ *for all* k, j.

When Property C holds we obtain an important regularity in the magnitude of the transition probabilities between the basins of attraction of steady states when the process is close to the steady state itself. Assuming Property C we have that $\forall x \in W_k^\delta$, $\delta < \hat{\delta}$:

$$T_\varepsilon(x, W_j) \in \Theta(\varepsilon^{V(x_k^*, x_j^*)})$$

The reason this regularity is important is that it allows a single magnitude of transition probability to characterize transitions from the neighborhood of one stable state to another. As there are a finite number of stable states there are a finite number of such transitions and the problem of determining the stochastically stable states is reduced to a discrete problem: the sets W_k^δ for which $\pi_\varepsilon(W_k^\delta) \nrightarrow 0$ as $\varepsilon \to 0$ are determined solely by the values of $V(.,.)$. Results from the finite state space stochastic stability literature can then be used.

3.4. Previous results restated for uncountable state spaces

Let $L = \{1, \ldots, \nu\}$. A graph on L is an *i-graph* if each $j \neq i$ has a single exiting directed edge, and the graph has no cycles. Let $\mathcal{G}(i)$ denote the set of all i-graphs. Define:

$$\mathcal{V}(i) = \min_{g \in \mathcal{G}(i)} \sum_{(k \to j) \in g} V(x_k^*, x_j^*),$$

$$L_{min} = \{i \in L : \mathcal{V}(i) = \min_{j \in L} \mathcal{V}(j)\}.$$

The key result of Young (1993a) and Kandori et al. (1993) can now be stated.

Proposition 4. *If Property C holds then:*

$$\pi(x_i^*) > 0 \Leftrightarrow i \in L_{min}$$

Proof. See appendix. □

Results of Ellison (2000) also carry through. First we define some concepts, the radius $R(.)$ and the coradius $CR(.)$, which are the natural extensions of the similarly named concepts of Ellison (2000) to larger state spaces and which for finite state spaces simplify to the definitions of radius and coradius found in his paper:

$$R(x_k^*) := \min_j V(x_k^*, x_j^*)$$

Letting $\{\bar{x}_1^*, \bar{x}_2^*, \ldots, \bar{x}_m^*\}$ be an ordered set of steady states such that $\bar{x}_1^* = x_j^*$ and $\bar{x}_m^* = x_k^*$ we can define:

$$CR(x_k^*) := \max_j \min_{\substack{\{\bar{x}_1^*, \ldots, \bar{x}_m^*\} \\ \bar{x}_1^* = x_j^* \\ \bar{x}_m^* = x_k^*}} \sum_{i=1}^{m-1} V(\bar{x}_i^*, \bar{x}_{i+1}^*)$$

where the minimum is over all possible paths of steady states starting at x_j^* and finishing at x_k^*. $R(x_k^*)$ can be interpreted as measuring how easy it is to escape from W_k^δ, $\delta < \hat{\delta}$. $CR(x_k^*)$ can be interpreted as giving a measure of how difficult it can be to get to W_k^δ from anywhere in the state space.

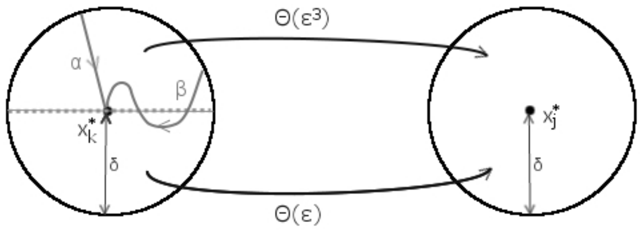

Figure 1: Property C does not hold. $V(x^*_k, x^*_j) = 3$ and $V^-(x^*_k, x^*_j) = 1$. If convergent paths were all similar to path α then the two semicircles could be treated as separate steady states, but the possibility of paths such as β complicates estimates of transition probabilities.

Proposition 5. *If Property C holds and:*

$$R(x^*_k) > CR(x^*_k)$$

then x^*_k is the unique stochastically stable state.

Proof. Choose $\delta < \hat{\delta}$, W^δ_k such that $\pi_\varepsilon(W^\delta_k) > 0$. Then (Meyn and Tweedie (2009), Theorem 10.4.9)) for all $i \neq k$ we have:

$$\pi(W^\delta_i) = \int_{W^\delta_k} \pi(dy) E_y \left[\sum_{t=0}^{\tau_{W^\delta_k}-1} \mathbb{I}(\Phi^t_\varepsilon \in W^\delta_i) \right]$$

and:

$$E_y \left[\sum_{t=0}^{\tau_{W^\delta_k}-1} \mathbb{I}(\Phi^t_\varepsilon \in W^\delta_i) \right] \in O(\varepsilon^{R(x^*_k) - CR(x^*_k)})$$

as probability of leaving W^δ_k is $\Theta(\varepsilon^{R(x^*_k)})$ and length of time spent outside W^δ_k before returning to W^δ_k is $O(\varepsilon^{-CR(x^*_k)})$. So the integrand above approaches zero as $\varepsilon \to 0$ and $\pi(W^\delta_i) \to 0$ for all $i \neq k$. So $\pi(W^\delta_k) \to 1$ and x^*_k is the stochastically stable state. □

4. Sufficient conditions for unique transition times

Given the usefulness of Property C in allowing the use of theorems from the finite state space literature, a natural question to ask is under what conditions it holds and whether or not these conditions are plausible and commonly satisfied. The next two propositions concern themselves with finding sufficient conditions for Property C. The first of these makes use of a strong continuity requirement on error distributions $G_i(.,.)$.

Proposition 6. *If there exists $\hat{\delta} > 0$ such that for all $\delta < \hat{\delta}$, $A \in \mathcal{B}(X)$:*

$$\forall x_1, x_2 \in B_\delta(x^*_k), \quad G_i(x_1, A) > 0 \Leftrightarrow G_i(x_2, A) > 0$$

then $V(x^*_k, x^*_j) = V^-(x^*_k, x^*_j)$ for all $j \neq k$.

Proof. Notice that by definition $V(x^*_k, x^*_j) \geq V^-(x^*_k, x^*_j)$. Assume $V(x^*_k, x^*_j) = a > b = V^-(x^*_k, x^*_j)$. Choose any $\delta < \hat{\delta}$. Then $V^-(x^*_k, x^*_j) = b$ implies that there exists $x_e \in W^\delta_k$ such that $\lim_{\varepsilon \to 0} \frac{T_\varepsilon(x_e, W_j)}{\varepsilon^b} > 0$. It follows that for some $\underline{b} \leq b$, $G_{\underline{b}}(x_e, C) > 0$, where $C \subset X$ is such that from any point in C the probability that the next basin of attraction of a steady state entered by the process is W_j is $\Theta(\varepsilon^{b-\underline{b}})$. Now, for all $x_f \in W^\delta_k$, as $x_e, x_f \in B_\delta(x^*_k)$ and $\delta < \hat{\delta}$, it follows that $G_{\underline{b}}(x_f, C) > 0$ which implies that $\lim_{\varepsilon \to 0} \frac{T_\varepsilon(x_f, W_j)}{\varepsilon^b} > 0$, that is $V(x^*_k, x^*_j) \leq b$. Contradiction. □

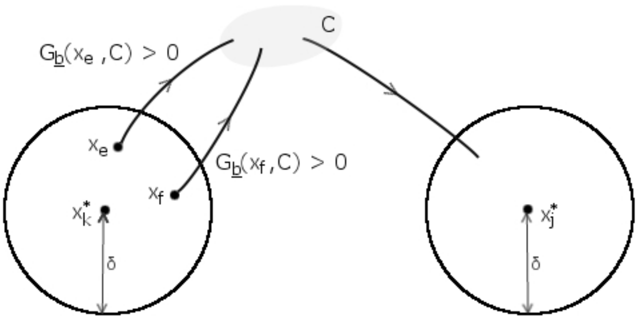

Figure 2: Proposition 6. $\delta < \hat{\delta}$ so a $\Theta(\varepsilon^{\underline{b}})$, $\underline{b} < b$ probability of transiting from x_e to set C implies a $\Omega(\varepsilon^{\underline{b}})$ probability of transiting from any $x_f \in W_k^\delta$ to set C.

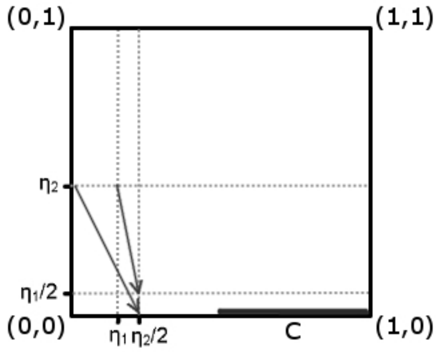

Figure 3: $(\frac{\eta_2}{2}, 0))$ differs from all points in C by one coordinate, $(\frac{\eta_2}{2}, \frac{\eta_1}{2})$ by two coordinates.

The condition in Proposition 6 is a strong one and is not satisfied by independent error models of stochastic stability. To see this take a game where each of two players chooses a number in the interval $[0, 1]$, the best response is to choose a number equal to half the number chosen by the other player and the unperturbed dynamic Φ is to naively best respond to what the other player played in the previous period (i.e. players have one period memory). The perturbed dynamic Φ_ε has players independently making errors with probability ε and when this happens a player chooses an action from a distribution over $[0, 1]$ with full support. Let $x = (x_1, x_2)$ and define:

$$C = \left\{ x : x_1 \geq \frac{1}{2}, x_2 = 0 \right\}$$

Then for $\eta_1, \eta_2 > 0$ we have $G_1((0, \eta_2), C) > 0$ but $G_1((\eta_1, \eta_2), C) = 0$ no matter how small is η_1 (Figure 3). So the condition of Proposition 6 is not satisfied.

Proposition 7 gives conditions for Property C that can be satisfied by independent error models of stochastic stability. This is important as such models are commonly found in the literature. Firstly, a weaker continuity requirement is placed on the error distributions than is the case in Proposition 6. This requirement is satisfied by independent error models. The weakening of the continuity requirement requires the introduction of further conditions. Secondly, the weak Feller property is imposed on the transition kernel of the unperturbed dynamic Φ. In a game theoretic context this corresponds to imposing continuity of the response correspondence of the underlying game. Thirdly, an asymptotic stability condition is imposed in the neighbourhoods of the steady states. Fourthly, a condition is given which restricts the behaviour of the process at a steady state according to the behaviour of the process at nearby states. Following the proof of the proposition a series of examples illustrates the role each of these conditions plays in the proof.

First define an attainability property for each $x^* \in X^*$ which holds when there is a positive probability of ending up spending time in the basin of attraction of x^* when the process is started from some arbitrary point in the state space.

Definition (Attainable steady states). For a given steady state x_i^*, let

$$B_i = \{x \in X : \exists t \, s.t. \, P_\varepsilon^t(x, W_i) > 0\}.$$

Define the set of attainable steady states $\mathcal{A} = \{x_i^* \in \Lambda : G_M(B_i) > 0\}$.

Condition (iii) of Proposition 7 is implied by (but does not imply) asymptotic stability under Φ of any $x^* \in \mathcal{A}$. Note $x_i^* \in \Lambda \setminus \mathcal{A}$ implies $\pi_\varepsilon(\mathcal{W}_i) = 0$.

Proposition 7. *If the following conditions hold then $V(x_k^*, x_j^*) = V^-(x_k^*, x_j^*)$ for all $j \neq k$.*

(i) For all $A \in \mathcal{B}(X)$ open, $x_1 \in X$, there exists $\delta_{Ax_1} > 0$ such that if $x_2 \in X$ satisfies $|x_1 - x_2| < \delta_{Ax_1}$ then:

$$G_i(x_1, A) > 0 \Rightarrow G_i(x_2, A) > 0$$

(ii) $P(x, .)$ is weak Feller, i.e. for all $x \in X$:

$$P(y, .) \Rightarrow P(x, .) \text{ as } y \to x$$

(iii) For all $x_i^* \in \mathcal{A}$, there exists $\tilde{\delta} > 0$ such that $\mathcal{W}_i^{\tilde{\delta}}$ is open.

(iv) For all j, k:

$$T_\varepsilon(x_k^*, W_j) \in \Theta(\varepsilon^{V^-(x_k^*, x_j^*)})$$

First a useful claim is proved.

Claim. *Assume the assumptions of Proposition 7 hold. Let $Q \in \mathcal{B}(X)$ be an open set, $b \in \{0, \dots, M\}$. Define:*

$$D_b = \{x \in X : \exists t > 0 \text{ s.t. } P_\varepsilon^t(x, Q) \in \Omega(\varepsilon^b)\}.$$

Then D_b is open.

Proof. The proof proceeds by induction. Assume that the claim holds for $b < k$: for a given Q, $D_0, D_1, \dots, D_{k-1}$ are open. Define:

$$R = \{x \in X : G_j(x, D_{k-j}) > 0 \text{ for some } j \in \{1, \dots, k\}\} \cup D_{k-1}$$

then R is open by (i). Notice that:

$$P_\varepsilon^t(x, Q) \in \Omega(\varepsilon^k) \qquad \forall\, x \in R$$

and as a consequence D_k can be written as:

$$D_k = \left\{x \in X : P^t(x, R) > 0 \text{ for some } t\right\}.$$

which is open by (ii). Showing that D_0 is open will complete the proof

$$D_0 = \{x \in X : P_\varepsilon^t(x, Q) \in \Omega(1) \text{ for some } t\}.$$

$$= \{x \in X : P^t(x, Q) > 0 \text{ for some } t\}.$$

which is open by (ii).

□

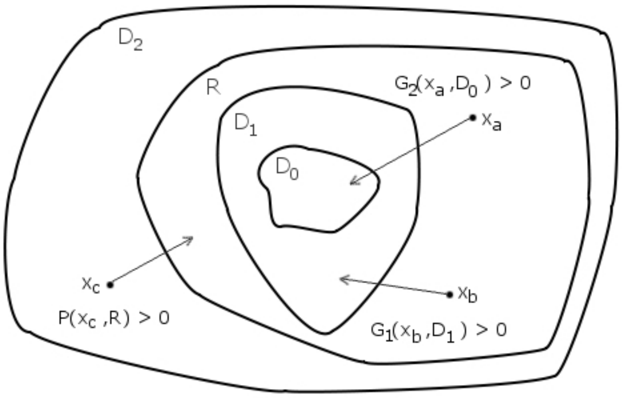

Figure 4: Given D_0, D_1, the sets R and then D_2 are constructed.

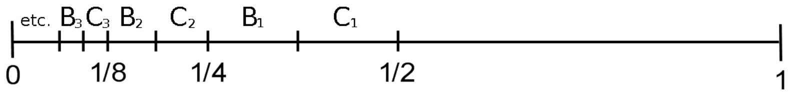

Figure 5: Example: (ii) does not hold

This paves the way for the proof of Proposition 7.

Proof of proposition 7. Assume $V^-(x_k^*, x_j^*) = b < \infty$. This implies $x_j^* \in \mathcal{A}$. Define $D_0, D_1, \ldots, D_M$ as above with $Q = \mathcal{W}_j^{\tilde{\delta}}$ for some $\tilde{\delta}$ which we can choose by (iii). As $V^-(x_k^*, x_j^*) = b$, we know from (iv) that $T_\varepsilon(x_k^*, W_j) \in \Theta(\varepsilon^b)$. This implies that $x_k^* \in D_b$. We know from the claim that D_b is an open set so $\hat{\delta}$ can be chosen such that:

$$x \in B_{\hat{\delta}}(x_k^*) \implies x \in D_b \implies P_\varepsilon^t(x, \mathcal{W}_j) \in \Omega(\varepsilon^b) \text{ for some } t.$$

This implies that $V(x_k^*, x_j^*) \leq b = V^-(x_k^*, x_j^*)$. $\square$

The rest of this section consists of stylized examples designed to demonstrate how Property C can fail if any of the four conditions of Proposition 7 do not hold.

4.1. Example: (ii) does not hold

Define $X = [0,1]^3$. Let the unperturbed process Φ represent a game played repeatedly by 3 players with strategy spaces $[0,1]$. Each period each player plays a best response to the actions of the other two players in the previous period. Define $C_i = \left(\frac{3}{4}\frac{1}{2^i}, \frac{1}{2^i}\right]$. Define $C = \bigcup_{i=1}^{\infty} C_i$. Let $B_i = \left(\frac{1}{2}\frac{1}{2^i}, \frac{3}{4}\frac{1}{2^i}\right]$, $B = \bigcup_{i=1}^{\infty} B_i$. Let best response functions be symmetric and anonymous:

$$BR(a,b) = \begin{cases} C_{i+1}, & \text{if } a \in C_i, b \notin C_j \quad \text{for some } i, j \\ C_{i+1}, & \text{if } a \in C_i, b \in C_j \quad \text{for some } i, j, i \geq j \\ 1, & \text{if } a \in (\frac{1}{2}, 1], b \notin C \\ B_{i+1}, & \text{if } a \in B_i, b \in B_j \quad \text{for some } i, j, i \geq j \\ 0, & \text{if } a = b = 0. \end{cases}$$

Note that $x_0^* = (0,0,0)$ and $x_1^* = (1,1,1)$ are the only stationary states of the unperturbed dynamic. Let the perturbed process Φ_ε be such that each player independently with probability ε plays an action chosen randomly from $[0,1]$ instead of playing a best response. $V(x_0^*, x_1^*) = 3$. $V^-(x_0^*, x_1^*) = 1$. There is in effect one convergent path to x_0^* via states in B from which it is easy to escape and another convergent path via states in C from which it is difficult to escape. As $\varepsilon \to 0$, $\pi_\varepsilon \Rightarrow \pi$ where $\pi((0,0,0)) = 1$.

Intuitively, (ii) not holding carries the implication that even when two states are extremely close to one another there is no guarantee that similar random shocks will lead to similar responses by the players. It is always possible to choose two states x_1, x_2 arbitrarily close to one another such that from x_1 a single player choosing a given random action would lead to completely different short run behaviour of the process to that which would occur if from x_2 exactly the same player chose exactly the same random action.

4.2. Example: (i) does not hold

For some steady states $x_k^* = 0, x_j^* = 1$, let $W_k \supset B_\delta(x_k^*)$ for some δ, so (iii) is satisfied. The error process guarantees that (iv) is also satisfied:

$$G_1(x,x) = 1, \text{ if the first non-zero digit in the decimal expansion of } x \text{ is } 1.$$

$$G_1(x, W_j) = 1, \text{ otherwise}$$

$$G_2(W_j) = 1$$

Then $V^-(x_k^*, x_j^*) = 1$ and $V(x_k^*, x_j^*) = 2$ no matter whether or not (ii) is satisfied.

4.3. Example: (iii) does not hold

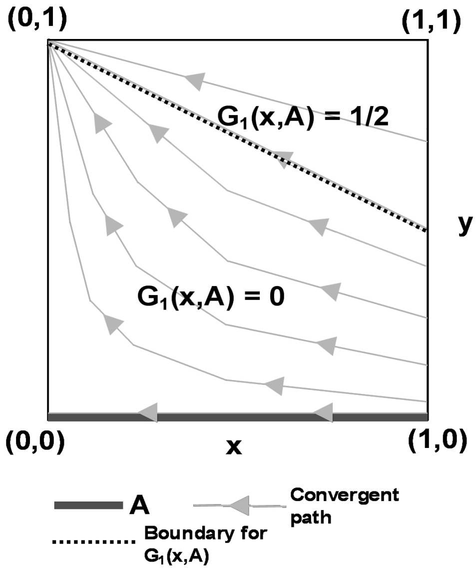

Figure 6: Example: (iii) does not hold

Define $X = [0,1]^2$. Let (x, y) describe an element of the state space. Let:

$$P\left((x,y), \left(\frac{x}{2}, y + \min\left\{\frac{1}{2}, y\right\}(1-y)\right)\right) = 1$$

then there are 2 steady states, $x_k^* = (0,1)$ and $x_j^* = (0,0)$. Let:

$$A = \{(x,y) \in X : x = 0\}$$

$$G_1(x, A) = \frac{1}{2} \text{ for } x \in U := \{(x,y) \in X : x + 2y \geq 2\},$$

and let $G_1(x,.)$ be uniform on X otherwise. Let $G_2(A) = 1$. Then (i), (ii) and (iv) hold but $V^-(x_k^*, x_j^*) = 1$ and $V(x_k^*, x_j^*) = 2$. Intuitively, although (ii) implies that from any point close to x_k^* but not in U it is possible to reach any point arbitrarily close to x_j^* with a probability of order ε, (iii) not holding means that this is not sufficient for convergence to x_j^* and the process ends up reconverging towards x_k^*.

4.4. Example: (iv) does not hold

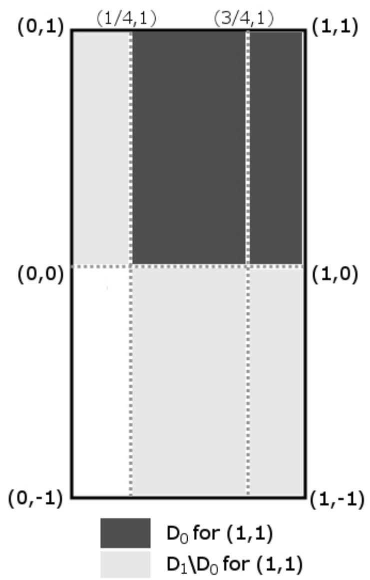

Figure 7: Example: (iv) does not hold. Although convergent paths to $(0,0)$ from above are in D_1, convergent paths from below are not so $(1,1)$ cannot always be reached with $\Omega(\epsilon)$ probabilities.

Define $X = [0,1] \times [-1,1]$. Let (x,y) describe an element of the state space. Let:

$$P\left((x,y), \left(\frac{x}{2}, \frac{y}{2}\right)\right) = \frac{\max\left\{0, \frac{3}{4} - x, \frac{1}{2} - y\right\}}{\max\left\{0, \frac{3}{4} - x, \frac{1}{2} - y\right\} + \max\left\{0, \min\left\{x - \frac{1}{4}, y\right\}\right\}}$$

$$P\left((x,y), \left(\frac{x+1}{2}, \frac{y+1}{2}\right)\right) = \frac{\max\left\{0, \min\left\{x - \frac{1}{4}, y\right\}\right\}}{\max\left\{0, \frac{3}{4} - x, \frac{1}{2} - y\right\} + \max\left\{0, \min\left\{x - \frac{1}{4}, y\right\}\right\}}$$

then there are 2 steady states, $x_k^* = (0,0)$ and $x_j^* = (1,1)$. Note that under the unperturbed dynamic the process will in each period move in a straight line towards one of the steady states. Let:

$$G_1((x,y),.) \sim U[\{(x',y') : x' = x \text{ or } y' = y\}]$$

$$G_2(.) \sim U[X]$$

Then (i), (ii) and (iii) hold but $V^-(x_k^*, x_j^*) = 1$ and $V(x_k^*, x_j^*) = 2$. Intuitively, although a convergent path to x_k^* exists in $D_1(B_{\tilde{\delta}}(x_j^*))$, the limit x_k^* is not in D_1, and so the existence of other convergent paths not in D_1 cannot be ruled out.

5. Economic examples

5.1. Linear quadratic games

We apply the theory to two player games with strategies $y_i \in \mathbb{R}_+$ and the following payoff functions:

$$u_i(y_i, y_j) = a_i y_i^2 + b_i y_i y_j + c_i y_i + d_i y_j, \qquad a_i, b_i < 0;\ c_i > 0;\ a_i, b_i, c_i, d_i \in \mathbb{R}.$$

This class of games includes public goods problems with strategic substitutes and Cournot duopolies with linear demand and quadratic costs. Oechssler and Riedel (2001) showed that under the replicator dynamic with symmetric payoff functions such games converge to the interior equilibrium in which players play $y_i^{int} = \frac{b_i c_j - 2c_i a_j}{4a_i a_j - b_i b_j}$[13]. We assume a multiplicity of equilibria: $y_i^{int} > 0$ so that $x_I = (y_1^{int}, y_2^{int})$ is a Nash equilibrium; $4a_i a_j < b_i b_j$ so that corner equilibria exist: $x_{c1} = (y_1^{cnr}, 0)$, $x_{c2} = (0, y_2^{cnr})$, $y_i^{cnr} = \frac{-c_i}{2a_i}$. Let the unperturbed dynamic be a Markov process on $X = \mathbb{R}_+^2$ in which each period each player has an equal chance to best respond. Φ has kernel:

$$P\left((y_1, y_2), \left(\max\left\{\frac{-b_1 y_2 - c_1}{2a_1}, 0\right\}, y_2\right)\right) = \frac{1}{2}; \qquad P\left((y_1, y_2), \left(y_1, \max\left\{\frac{-b_2 y_1 - c_2}{2a_2}, 0\right\}\right)\right) = \frac{1}{2}.$$

This gives:

$$W_I = \{x_I\}, \quad W_{ci} = \{(y_i, y_j) : y_i > y_i^{int}, y_j < y_j^{int}\},$$

$$\mathcal{W}_I = \{(y_i, y_j) : y_i = y_i^{int} \text{ or } y_j = y_j^{int}\}, \quad \mathcal{W}_{ci} = \{(y_i, y_j) : y_i > y_i^{int} \text{ or } y_j < y_j^{int}\}.$$

Note that $P(.,.)$ satisfies condition (ii) of proposition 7. We analyse two possible perturbed dynamics.

5.1.1. Uniform local perturbations

For some small $\delta > 0$, define:

$$G_1((y_1, y_2), .) \sim U[B_\delta((y_1, y_2)) \cap \mathbb{R}_+]$$

$$G_n((y_1, y_2), .) = G_1^n((y_1, y_2), .), \quad n = 2, \ldots, M-1$$

$$G_M((0,0)) = 1.$$

These $G_n(.,.), G_M(.)$ satisfy condition (i) of proposition 7. $G_M(.)$ is not necessary for the results of this section, although it is of interest to note that any Markovian dynamic with some small, bounded below, probability of an armageddon event will satisfy the conditions for ergodicity. $\mathcal{A} = \{x_{c1}, x_{c2}\}$ and it is clear that for some $\delta > 0$, $\mathcal{W}_{ci}^\delta$ are open so condition (iii) is also satisfied. As $\mathcal{W}_{cj}^\delta$ is open, for any escape path from a state close to x_{ci} to $\mathcal{W}_{cj}^\delta$ there is a similar path from x_{ci} itself to $\mathcal{W}_{cj}^\delta$, so condition (iv) is satisfied. Now:

$$V(x_{ci}, x_{cj}) = \min\left\{\frac{y_j^{int}}{\delta}, \frac{y_i^{cnr} - y_i^{int}}{\delta}\right\}$$

and applying proposition 4 we obtain:

Proposition 8.

$$\min\left\{y_j^{int}, y_i^{cnr} - y_i^{int}\right\} \geq \min\left\{y_i^{int}, y_j^{cnr} - y_j^{int}\right\} \Leftrightarrow \pi(x_{ci}) > 0$$

5.1.2. Proportional perturbations

For some $k \in (0,1)$, define:

$$G_1((y_1, y_2), .) \sim U[[ky_1, y_1] \times [ky_2, y_2]]$$

$$G_n((y_1, y_2), .) = G_1^n((y_1, y_2), .), \quad n = 2, \ldots, M-1$$

$$G_M((0,0)) = 1.$$

Similarly to above, conditions (i),(iii),(iv) are satisfied. Then:

$$V(x_{ci}, x_{cj}) = \frac{y_i^{cnr}}{y_i^{int}}$$

and applying proposition 4 we obtain:

[13] $y_i^{int} = \frac{-c}{2a+b}$ under symmetry.

Proposition 9.

$$\frac{y_i^{cnr}}{y_i^{int}} \geq \frac{y_j^{cnr}}{y_j^{int}} \Leftrightarrow \pi(x_{ci}) > 0$$

5.2. Sampling a population

Take a two player symmetric bimatrix game in which a player has a set N of possible actions, $|N| = n$. Let there be a continuum of agents on the unit interval. The state space is defined as the proportions in which each action is played at a point in time: X is the unit $(n-1)$-simplex. In period t, independently of his previous actions, with probability $1-\alpha$ any given player plays the same action as at time $t-1$. With probability α he randomly and uniformly samples a finite number k of the actions of players in period $t-1$ before playing a best response to the mixed strategy which has each action being played with a probability equal to its proportion in his sample. If multiple best responses exist we assume that they are chosen with equal probability Denote the distribution of such best responses to an action profile x by $BR(x)$. Then:

$$P(x, \tilde{x}) = 1, \qquad \tilde{x} := (1-\alpha)x + \alpha BR(x).$$

Note that as the probability of drawing any given sample is continuous in x, $P(x, .)$ is itself continuous and condition (ii) of proposition 7 is satisfied. We restrict attention to games for which this process satisfies the convergence assumption. It is trivial to construct games which do not satisfy the convergence assumption under this process, for example the 2 by 2 matrix of zeroes.

Any steady state x^* is close to a Nash equilibrium of Γ in the following sense:

Proposition 10. *For any $\xi > 0$, there exists $\bar{k} \in \mathbb{N}_+$ such that if $k > \bar{k}$, for any $x^* \in \Lambda$ there exists a symmetric Nash equilibrium x^{NE} of Γ such that $|x^* - x^{NE}| < \xi$.*

Proof. See appendix. □

The perturbations are defined as follows. For some small $\delta > 0$, define:

$$G_1(x, .) \sim U[B_\delta(x) \cap X]$$

$$G_n(x, .) = G_1^n(x, .), \quad n = 2, \ldots, M-1$$

$$G_M(.) \sim U[X]$$

These $G_n(., .), G_M(.)$ satisfy condition (i) of proposition 7. $G_M(.)$ is not necessary for the results of this section.

Proposition 11. *If x^{NE} is a strict symmetric pure Nash equilibrium then:*

$$x^{NE} \in \mathcal{A} \quad \textit{and} \quad x^{NE} \textit{ is asymptotically stable.}$$

Proof. See appendix. □

	L	R
L	a, a	b, c
R	c, b	d, d

Figure 8: A two player strategic game. (L, L) and (R, R) assumed to be strict Nash equilibria.

In a game such as that of figure 8 in which there are two strict symmetric Nash equilibria, by proposition 11 both of these Nash equilibria correspond to steady states in $\mathcal{A}$ such that the basin of attraction includes some open ball centred on the Nash equilibrium. Close to the mixed Nash equilibrium, for large enough k, there is one steady state which is not in $\mathcal{A}$. Hence condition (iii) of proposition 7 is satisfied. Let $(p, 1-p)$ be the equilibrium strategy of the mixed Nash equilibrium. Then as long as neither p nor $1-p$ is an integer multiple of δ, condition (iv) is also satisfied.[14] Identical arguments to those in Young (1998) then give:

Proposition 12. *In a 2 by 2 game with two symmetric strict pure Nash equilibria, one of which is risk dominant, the risk dominant equilibrium is uniquely stochastically stable.*

6. What if Property C does not hold?

Analysis can be complicated in cases in which Property C does not hold. In some cases, however, a simple argument can still be used to find the stochastically stable state. This is true for the dynamic of section 2.2. The state space in that example can be partitioned into a finite collection of disjoint sets such that under $P_\varepsilon(.,.)$ any of the sets can be reached from any $x \in X$ with a probability bounded below by a strictly positive number. Such a partition is:

$$X = X_0 \cup X_k \cup X_{k+1} \ldots \cup X_n$$

where for $i = k, \ldots, n$:

$$X_i = \{x \in X : |I(x)| = i\}$$

and

$$X_0 = \{x \in X : |I(x)| < k\}$$

Bounds can be found for the transition probabilities between these sets. Freidlin and Wentzell (1998) style tree arguments can then be used. The 'flow' of probability mass from X_i to X_{i-1} for $i = k+1, \ldots, n$, and from X_k to X_0 can be shown to be of order $\Theta(\varepsilon)$, so a tree of order $\Theta(\varepsilon^{n-k+1})$ rooted at X_0 can be constructed. Any tree not rooted at X_0 must include an edge leaving X_0 of order $O(\varepsilon^k)$. Taking limits of the invariant measures then gives the conclusion that all probability mass in the limit is concentrated in X_0 and therefore, by proposition 3, on 0^n.

Proposition 13. *In the example of section 2.2 the unique stochastically stable state is* 0^n.

Proof. See appendix. □

7. Concluding remarks

This paper has demonstrated how commonly used stochastic stability methods can be applied to settings with non-finite state spaces, corresponding to situations in which economic agents choose from non-finite strategy sets. It includes sufficient conditions for the straightforward application of existing results in the literature to such settings. Moreover, the analysis of the complications that can arise with general state spaces aids understanding of the problems of which one should be aware when applying ideas of robustness to random perturbations to processes which do not satisfy all of our conditions.

Another approach when seeking to find stochastically stable states for games with non-finite strategy sets is to discretize the strategy space and the transition kernel. This is not always simple and can lead to problems such as the absence of simple closed form best response functions and nonexistence of equilibrium. Difficulties can be met when passing to the limit of any discretization as it becomes fine. Moreover, examples in this paper show that for some examples, any discretization satisfying plausible criteria leads to the selection of different equilibria to those selected when the analysis is carried out directly on the original state space and process. Sometimes it may be better to analyze stochastic stability whilst remaining in a non-finite world. This paper gives tools with which to aid that endeavour.

[14] If $p = 2\delta$, from $(0,1)$ it takes an order ε^3 event to move to the basin of attraction of $(1,0)$, but from $(\xi, 1-\xi), \xi \in (0,\delta)$, an order ε^2 event is all that is required.

Appendix A. Puiseux Markov processes

The field of Puiseux functions is all functions $\hat{f} : (0, 1) \to \mathbb{R}$ that can be represented as:

$$\hat{f}_\varepsilon = \sum_{i=L}^{\infty} a_i \varepsilon^{i/M}.$$

in some neighbourhood of zero for some integer L and some positive integer M. The *valuation* of a Puiseux function f with this representation is:

$$w(\hat{f}) = \frac{\min\{i : a_i \neq 0\}}{M}.$$

Proof of existence of limiting distribution. If all $P_\varepsilon(x, B)$, $x \in X$, $B \in \mathcal{B}(X)$, are Puiseux, then:

$${}_AP^t_\varepsilon(x, B) := P_x(\Phi^t_\varepsilon \in B, \tau_A \geq t);\ x \in X;\ A, B \in \mathcal{B}(X)$$

are Puiseux functions as Puiseux functions form a field and $P_x(\Phi^t_\varepsilon \in B, \tau_A \geq n)$ can be expanded into a summing and multiplication of Puiseux functions. So:

$$U_A(x, B) := \sum_{t=1}^{\infty} {}_AP^t_\varepsilon(x, B); \quad x \in X;\ A, B \in \mathcal{B}(X)$$

is a Puiseux function if it exists as the limit of Puiseux functions is Puiseux because Puiseux functions are a field. It does exist if we let A have positive probability under $G_M(.)$, as:

$${}_AP^t_\varepsilon(x, B) \leq (1 - G_M(A))^t < 1$$

so $U_A(x, B)$ is bounded above by the sum of a geometric series which converges. As $A \in \mathcal{B}^+(X)$ we have:

$$\pi(B) = \int_A \pi(dy) U_A(x, B)$$

which is a weighted average of Puiseux functions and therefore a Puiseux function (CHECK THIS!!!) for any ε. As $\pi(B_1)$, $\pi(B_2)$ are Puiseux functions, say $w(\pi(B_1)) \geq w(\pi(B_2)$ we have:

$$\lim_{\varepsilon \to 0} \left\{ \frac{\pi(B_1)}{\pi(B_2)} \right\} \to a \geq 0.$$

□

Appendix B. Proofs

Let $\{X_i | i \in L\}$, $|L| = \nu$, be a partition of X into finitely many disjoint sets such that:

$$\exists c_{ij} > 0 \text{ s.t. } \inf_{x \in X_i} P_\varepsilon(x, X_j) \geq c_{ij}.$$

For given invariant measure $\pi(.)$, let:

$$p_{ij} := \frac{1}{\pi(X_i)} \int_{X_i} P_\varepsilon(x, X_j) \pi(dx)$$

For $g \in \mathcal{G}(i)$, define:

$$vol(g) := \prod_{(j \to k) \in g} p_{jk}; \qquad Q_i := \sum_{g \in \mathcal{G}(i)} vol(g)$$

then (Freidlin and Wentzell, 1998):

$$\pi_\varepsilon(X_i) = \frac{Q_i}{\sum_{j=1}^{\nu} Q_j}.$$

Proof of proposition 4. Property C ensures we can choose δ small enough that:

$$\forall x \in W_j^\delta : \quad p_{jk} := T_\varepsilon(x, W_k) \in \Theta(\varepsilon^{V(x_k^*, x_j^*)})$$

Choosing the partition $\{W_1^\delta, \ldots, W_{\nu-1}^\delta, X \setminus \cup_i W_i^\delta\}$ we get that:

$$i \in L_{min} \Leftrightarrow \mathcal{V}(i) \leq \mathcal{V}(j)\ \forall j \neq i$$

$$\Leftrightarrow \exists g \in \mathcal{G}(i) : \quad vol(g) \in \Theta(\varepsilon^{\mathcal{V}(i)}) \quad \text{and} \quad \forall j, \forall g \in \mathcal{G}(j) : \quad vol(g) \in O(\varepsilon^{\mathcal{V}(i)})$$

$$\Leftrightarrow \exists \mathcal{V}(i) \in \mathbb{N}_0 : \ Q_i \in \Theta(\varepsilon^{\mathcal{V}(i)}),\ \sum_j Q_j \in \Theta(\varepsilon^{\mathcal{V}(i)})$$

$$\Leftrightarrow \pi(W_i^\delta) > 0 \Leftrightarrow \pi(x_i^*) > 0.$$

□

Proof of proposition 10. Assume the statement is incorrect for some $\xi > 0$. There must be an infinite sequence of steady states $\{x^*(k)\}_k$ corresponding to increasing values of k such that none of these steady states are within ξ of any Nash equilibrium. As the sequence is bounded it contains a convergent subsequence. Denote its limit by $\bar{x}^*$. As $\bar{x}^*$ is not a Nash equilibrium at least one of the actions played is not a best response to the true distribution of play. Assume that action i is one of these actions. $\bar{x}_i^* > 0$. Then there exists $\underline{k}$, $\eta > 0$, such that for all $k > \underline{k}$, $x_i^*(k) > \eta$. Note that as $x^*(k)$ is a steady state, $BR_i(x^*(k)) = x_i^*(k)$.

There exists γ such that for all x such that $|x - \bar{x}^*| < \gamma$, best responses to the true distribution of x are a subset of the best responses to the true distribution of $\bar{x}^*$. Define $f_k(\sigma)$ as the distribution of samples σ at $x^*(k)$. As $k \to \infty$: $f_k(\sigma)$ approaches in distribution the probability measure with point mass on $\bar{x}^*$. This implies:

$$\forall \eta\, \exists \bar{k} : \forall k > \bar{k} : \sum_{\sigma \notin B_\gamma(\bar{x}^*)} f_k(\sigma) < \eta,$$

which implies:

$$BR_i(x^*(k)) < \eta.$$

and we have a contradiction. □

Proof of proposition 11. Assume that $x_i^{NE} = 1$. Let x^t be such that $x_i^t = 1 - \xi, \xi \in (0,1)$. There exists $s \in \mathbb{R}$ such that if i is not a best response to the strategy σ then:

$$\sum_{j \neq i} \sigma_j \geq s.$$

The proportion of players changing strategy who will sample such a σ is:

$$\tilde{\xi} = \sum_{j=\lceil ks \rceil}^{k} \binom{k}{j} \xi^j (1-\xi)^{k-j} \in \Theta(\xi^{\lceil ks \rceil}).$$

Assuming k is large enough that $\lceil ks \rceil \geq 2$, there exists $\bar{\xi}$ such that:

$$\forall \xi < \bar{\xi} : \quad \tilde{\xi} < \xi.$$

So for any $x^t \in B_{\bar{\xi}}(x^{NE})$ we have that $x_i^t = 1 - \xi$ for some $\xi < \bar{\xi}$ and:

$$x_i^{t+1} \geq (1-\alpha)x_i^t + \alpha(1-\tilde{\xi}) = (1-\alpha)(1-\xi) + \alpha(1-\tilde{\xi}) > 1 - \xi = x_i^t.$$

So we have convergence to x^{NE} from an open ball centered on x^{NE}, so $x^{NE} \in \Lambda$. This open ball is reached with positive probability from anywhere in the state space so $x^{NE} \in \mathcal{A}$. □

Proof of proposition 13. For $i = k+1, \ldots, n$, $x \in X_i$:

$$P_\varepsilon(x, X_{i-1}) \in \Theta(\varepsilon).$$

therefore $p_{i,i-1}, \in \Theta(\varepsilon)$. Also, for $x \in X_k$:

$$P_\varepsilon(x, X_0) \in \Theta(\varepsilon).$$

and $p_{k,0}, \in \Theta(\varepsilon)$. So there exists a 0-graph $\tilde{g}$ with $vol(\tilde{g}) \in \Theta(\varepsilon^{n-k+1})$. Therefore $Q_0 \in \Theta(\varepsilon^{n-k+1})$. For $x \in X_0$, $i \neq 0$:

$$P_\varepsilon(x, X_i) \in O(\varepsilon^k)$$

so any i-graph, $i \neq 0$, has $vol(\tilde{g}) \in O(\varepsilon^{(n-k)+k}) = O(\varepsilon^n)$. Therefore Therefore $Q_i \in O(\varepsilon^n)$. Using the formula for $\pi_\varepsilon(X_i)$ we see that $\pi_\varepsilon(X_i) \to 0$ as $\varepsilon \to 0$ for all $i \neq 0$. So $\pi_\varepsilon(X_0) \to 1$ as $\varepsilon \to 0$ and by proposition 3, π_ε approaches the distribution with point mass on 0^n. □

Agastya, M., 1997. Adaptive play in multiplayer bargaining situations. Review of Economic Studies 64, 411–26.

Bergin, J., Lipman, B.L., 1996. Evolution with state-dependent mutations. Econometrica 64, 943–56.

Chong, J.K., Camerer, C.F., Ho, T.H., 2006. A learning-based model of repeated games with incomplete information. Games and Economic Behavior 55, 340–371.

van Damme, E., Weibull, J.W., 2002. Evolution in games with endogenous mistake probabilities. Journal of Economic Theory 106, 296–315.

Ellison, G., 2000. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. Review of Economic Studies 67, 17–45.

Eshel, I., Samuelson, L., Shaked, A., 1998. Altruists, egoists, and hooligans in a local interaction model. American Economic Review 88, 157–79.

Feinberg, Y., 2006. Evolutionary dynamics and long-run selection. The B.E. Journal of Theoretical Economics 6.

Foster, D., Young, H.P., 1990. Stochastic evolutionary game dynamics. Theoretical Population Biology 38, 219–232.

Freidlin, M.I., Wentzell, A.D., 1998. Random perturbations of dynamical systems, isbn 9780387983622 430 pp. Springer .

Fremlin, D.H., 2001. Measure Theory: Broad Foundations. Torres Fremlin. ISBN 978-0953812929.

Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. Econometrica 61, 29–56.

Kifer, Y., 1990. A discrete-time version of the wentzell-friedlin theory. The Annals of Probability 18, pp. 1676–1692.

Meyn, S.P., Tweedie, R.L., 1994. State-dependent criteria for convergence of markov chains. The Annals of Applied Probability 4, pp. 149–168.

Meyn, S.P., Tweedie, R.L., 2009. Markov chains and stochastic stability. Springer-Verlag, London ; New York :.

Newton, J., 2012. Coalitional stochastic stability. Games and Economic Behavior 75, 842–54.

Oechssler, J., Riedel, F., 2001. Evolutionary dynamics on infinite strategy spaces. Economic Theory 17, 141–162.

Schenk-Hoppe, K.R., 2000. The evolution of walrasian behavior in oligopolies. Journal of Mathematical Economics 33, 35–55.

Selten, R., Apesteguia, J., 2005. Experimentally observed imitation and cooperation in price competition on the circle. Games and Economic Behavior 51, 171 – 192.

Serrano, R., Volij, O., 2008. Mistakes in cooperation: the stochastic stability of edgeworth's recontracting. Economic Journal 118, 1719–1741.

Smith, J.M., Price, G.R., 1973. The logic of animal conflict. Nature , 15–18.

Young, H.P., 1993a. The evolution of conventions. Econometrica 61, 57–84.

Young, H.P., 1993b. An evolutionary model of bargaining. Journal of Economic Theory 59, 145 – 168.

Young, H.P., 1998. Individual strategy and social structure. Princeton University Press.