The Stochastic Response Dynamic: A New Approach to Learning and Computing Equilibrium in Continuous Games

Amit Gandhi
University of Chicago
Graduate School of Business

Abstract

This paper describes a novel approach to both learning and computing Nash equilibrium in continuous games that is especially useful for analyzing structural game theoretic models of imperfect competition and oligopoly. We depart from the methods and assumptions of the traditional "Calculus" approach to computing equilibrium. Instead, we use a new and natural interpretation of games as conditionally specified probability models to construct a simple stochastic learning process for finding a Nash equilibrium. We call this stochastic process the *stochastic response dynamic* because of its resemblance to the classical Cournot best response dynamic. The stochastic response dynamic has economic meaning as a formal learning model, and thus provides a rationale for equilibrium selection, which is a major problem for the empirical application of structural game theoretic models. In addition, by way of its learning path, it provides an easy guide to check for the presence of multiple equilibria in a game. Finally, unlike the Calculus approach, the stochastic response dynamic does not depend upon the differentiability properties of the utility functions and makes no resort to the first order conditions of the underlying game.

1 Introduction

The traditional approach to computing a pure strategy Nash equilibrium in continuous games [1] is to use certain standard tools from Calculus to set up and solve a potentially nonlinear system of equations. More precisely, the "Calculus approach" finds a Nash equilibrium by using a Newton-like method to search for a state of the game that simultaneously solves each player's first order condition.

The Calculus approach suffers from some important shortcomings that hinder its overall usefulness for economic analysis. First, numerical methods of solving systems of equations do not correspond to any kind of economically meaningful learning process, i.e., the underlying search for equilibrium is not decentralized. At every step of the search, Newton's method updates each player's action in a way that depends on the utilities of all the players. On the other hand, a decentralized approach to finding equilibrium requires that each player's choice of action depend only upon his own utility function. Decentralizing the search for equilibrium is the key feature of a learning model in Economics[12].

For normative applications of game theory to the design of multi-agent systems [20], a decentralized approach finding Nash equilibrium is essential. For positive applications, where structural game theoretic models are estimated against data, and then used to analyze counterfactual states of the world, such as mergers or the introduction of new products, a learning model that effectively computes equilibrium offers a natural solution to the problem of multiple equilibria and equilibrium selection. A major motivation for developing models of learning in games is to offer a framework for saying whether some equilibria are more likely than others, depending on whether they can be learned in some natural way [8].

The usefuleness of the Calculus approach for applied work is also limited by the differentiability assumptions it imposes on the game. If utility functions of the underlying game are not differentiable, and/or the first order conditions are not well behaved, then the Calculus approach can readily fail to find a Nash equilibrium.

This paper describes a novel approach to learning and computing Nash equilibrium that abandons the methods and assumptions of the traditional Calculus approach. Instead, we use a new and natural interpretation of games as conditionally specified probability models to construct a simple stochastic process for finding a Nash equilibrium. We call this stochastic process the *stochastic response dynamic*. The stochastic response dynanamic, unlike the Calculus approach, searches for Nash equilibrium in a decentralized way[2]. Furthermore, the

[1] The term continuous games shall mean games with continous action sets and continuous utility functions. Throughout the paper, all games are implicitly assumed to be continuous games.

[2] The bulk of the literature on learning in games is focused on finite games and mixed strategy equilibrium. Our focus on learning pure strategy equilibrium in continuous games, which is the primary game theoretic framework used in empirical work, is a unique feature of the present work.

stochastic response dynamic does not require the utility functions of the game to be differentiable and makes no resort to the first order conditons whatsoever.

The stochastic response dynamic is reminiscent of the well known *best response dynamic*. The best response dynamic, which dates back to the inception of game theory [3], is a simple (perhaps the simplest) model of learning in continuous games. Starting from any initial state of the game, it specifies that the players will take turns choosing their best response to the state of the game in the previous period. Of course, a player's best response to a state of the game depends only upon his own utility function, and hence the process is decentralized. Further, if this process of iterative best responses settles down to a single state, then we know it has settled down to a Nash equilibrium.

Unfortunately the best response dynamic does not suffice as a general tool for finding equilibrium because there is little guarantee that the process will ever converge. Only under very special monotinicity assumptions on the game, such as supermodularity, can such a guarantee be made[3].

The stochastic response dynamic is essentially a stochastic version of the best response dynamic. Instead of players taking turns exactly best responding to the state of the game in the previous period, the stochastic response dynamic specifies that players behave by using stochastic responses. A stochastic response consists of the following. A player first randomly samples a candidate response from his action set. With this candidate response in hand, our player can either continue to use his action from the previous period (and thereby reject his candidate response), or he can switch his action to the candidate response (and thereby accept the candidate response). Rather than necessarily choosing the option that yields higher utility, our player *probabilistically* chooses between his two possible responses. This choice probability is governed by the classical binary logit model.

The variance parameter in the logit choice probability can be interpreted as the amount of noise entering the stochastic response dynamic. If the level of noise is infinitely large, then our player flips a fair coin in order choose between responding using his previous period action or responding using his candidate response. As the level of noise decreases, the choice probabilities increasingly favor the response that yields the higher utility. In the limit, as the level of noise approaches zero, our player necessarily chooses the response that yields the higher utility.

Thus the stochastic response dynamic maintains the decentralized nature of the best response dynamic, but replaces best responses with our notion of a stochastic response. The key benefit achieved by replacing best responses with stochastic responses is that it overcomes the major shortcoming of the best response dynamic, namely its lack of convergence. For a fixed noise level, the stochastic response dynamic transforms the previously deterministic best response dynamic into a stochastic process. The form of the stochastic process is a Markov chain, which unlike the deterministic best response dynamic, *can* be

[3]This lack of convergence problem holds true for fictitious play more generally, the best response dynamic being an extreme form of fictitious play.

assured to uniquely converge. The convergence however is no longer to a single state of the game, but rather to a probability distribution over the possible states of the game. In the language of Markov chains, the stochastic response dynamic is uniquely ergodic and converges strongly (in total variation norm) to a unique equilibrium distribution.

Our main results concern the relationship between the equilibrium distribution of the stochastic response dynamic, and the Nash equilibria of the underlying game. Essentially, we argue that for the class of games over which we wish to to compute, the Nash equilibria of the game are the modes, or "centers of gravity" of the equilibrium distribution. Furthermore we argue that as we decrease the variance parameter of the logit choice probability, i.e., as we decrease the amount of noise entering the stochastic response dynamic, the equilibrium distribution becomes increasingly concentrated over its modes, and thus in the limit, converges (in law) to a degenerate distribution over the Nash equilibria states of the game. In the case of a unique Nash equilibrium, this limiting distribution is simply a point mass over the Nash equilibrium.

Thus simulating the stochastic response dynamic at a fixed noise level asymptotically produces a random draw from the corresponding equilibrium distribution. As the noise gets smaller, the probability that this equilibrium draw is further than an epsilon away from a Nash equilibrium goes to zero. Computing Nash equilibrium amounts to simulating the stochastic response dynamic for a fixed number of iterations, then decreasing the amount of noise according to some rate, and repeating the simulation. Simulating the stochastic response dynamic at a given noise level will result in a sample being drawn from its equilibrium distribution. As we reduce the noise, this sample will be increasingly concentrated over Nash equilbrium states of the game. Once the simulation produces a sample draw near one of the equilibria at a sufficiently small noise level, the simulation becomes "stuck" near this equilibria (i.e. all the candidate responses will be rejected), and thus we have approximately computed an equilibria.

This process of adding noise or "heat" to a system, and reducing this heat so as to bring the system to its equilibrium resting state, has a long established history in thermodynamics. More generally, simulating Markov chains so as to compute scientific quantities of interest has gained much momentum in the Statistics and Physics literature over the past 50 years [10]. Usually known as "Markov Chain Monte Carlo", the approach has been used to solve computational problems that cause difficulty for gradient based numerical methods (i.e. the Calculus based approaches). The stochastic response dynamic follows in the spirit of the Markov Chain Monte Carlo literature. In fact, the stochastic response dynamic is nothing more than an adaptation of the classic Metropolis algorithm taken from statistical physics literature, except modified so as to be applicable to strategic games.

The connection to the MCMC literature provides insights for improving the computational performance of the stochastic response dynamic. Markov Chain Monte Carlo is generally quick (efficient) to provide an approximate solution, but can be slower at providing a highly accurate one. A common approach in the literature to improve upon this is to take a hybrid approach, using MCMC to generate a "close" answer to a problem, and then letting a traditional Calculus driven approach hone the accuracy of this approximate answer [1]. Thus In terms of computing Nash equilibrium, the stochastic response dynamic can quickly generate an approximate equilibrium, which can be input as the starting value for a nonlinear equations solver that computes the solution to the system of first order equations.

The plan of the paper is as follows. In section 2 we review strategic games and Nash equilibrium, and define the class of games $\mathcal{G}$ over which we wish to compute, namely continuous/quasiconcave games. Section 3 reviews the fairly simple minded, but nevertheless decentralized approach to finding equilibrium, the best response dynamic. In section 4, we see how the Calculus approach solves the lack of convergence problem of the best response dynamic. However the Calculus approach comes at the cost of not being decentralized, and requiring that the first order conditions both exist and are well behaved.

In section 5, we introduce the stochastic response dynamic, which replaces the best responses from the best response dynamic with stochastic responses. In section 6 we provide an intuitive view as to why the stochastic response dynamic converges to Nash equilibrium by introducing a restrictive assumption on the game known as compatibility. In section 7, we drop the compatibility assumption and replace it with the more natural Nash connectedness assumption, which we conjecture holds true generically for games in $\mathcal{G}$ (and can prove it holds true for 2 player games in the class). The convergence properties of the stochastic response dynamic are shown to hold equally under connectedness as compatibility. In section 8 we begin to illustrate the computational behavior of the stochastic response dynamic using some simple Cournot games. And finally, in section 9 we illustrate it using a more challenging Bertrand problem involving capacity constraints for which the standard differentiability assumptions of the Calculus approach are not met.

2 Strategic Games and Nash equilibrium

A strategic game models the problem faced by a group of players, each of whom must independently take an action before knowing the actions taken by the other players. Formally, a strategic game G consists of three components:

1. A set of players N.
2. A set of actions A_i for each player $i \in N$
3. A utility function $u_i : \times_{j \in N} A_j \mapsto \mathbb{R}$ for each player $i \in N$.

The key feature of G, what makes it a game rather than a collection of individual decisions, is that each player i has a utility function u_i over the possible combinations of actions of all the players $A = \times_{j \in N} A_j$ instead of just over their own actions A_i. We shall refer to A as the *state space* of the game.

The goal of studying a strategic game is to predict what action each player in the game will take. That is, to predict the outcome $a^* \in A$ of G. A solution concept for the study of strategic games is a rule that associates a game G with a prediction $a^* \in A$ of its outcome. Nash equilibrium is the central solution concept in Game Theory. Its centrality arises from the fact that it is the only possible solution concept consistent with the assumptions that the players in the game themselves know the solution (they are intelligent), and make decisions rationally [18]. Thus a (pure strategy) Nash equilibrium [4] is an outcome $a^* \in A$ such that for every player i,

$$a_i^* \in \arg\max_{a_i \in A_i} u_i(a_{-i}^*, a_i).$$

A game G may not have a Nash equilibrium or may have more than one Nash equilibrium. In these cases, predicting the outcome requires more than just applying the concept of Nash equilibrium, such as adjusting the model to account for private information or eliminating any "unreasonable" equilibria. [19]

Yet even if a unique Nash equilibrium $a^* \in A$ exists, there is still a difficulty with predicting the outcome of a game G to be a^*. It is possible that a player i is indifferent between his Nash equilibrium action a_i^* and another action $a_i' \in A_i$, $a_i' \neq a_i^*$, such that

$$a_i' \in \arg\max_{a_i \in A_i} u_i(a_{-i}^*, a_i).$$

In this event, it is difficult to state what reason player i has to choose a_i^* instead of a_i' besides preserving the equilibrium outcome a^*. If however no such a_i' exists, which is to say the set

$$\arg\max_{a_i \in A_i} u_i(a_{-i}^*, a_i)$$

is the singelton set $\{a_i^*\}$, then the outcome $a^* \in A$ is an even more compelling prediction. Such an outcome is said to be a *strict* Nash equilibrium, and is a generic feature of pure strategy Nash equilibrium [11]. For a strict Nash equilibrium, the definition can be written simply as

$$a_i^* = \arg\max_{a_i \in A_i} u_i(a_{-i}^*, a_i).$$

We consider the family of games $\mathcal{G}$ where every $G \in \mathcal{G}$ is continuous and strictly quasiconcave. By continuous, we mean the action sets A_i of G are closed bounded rectangles in a Euclidean space, and the utility functions u_i are continuous. By strictly quasiconcave, we mean that the utility functions u_i are strictly quasiconcave in own action. This standard set up ensures that a pure strategy Nash equilibrium of G exists, and all pure strategy equilibria are strict equilibria. Our computational results suggest that strict quasiconcavity can be relaxed to ordinary quasiconcavity so long as the equilibria remain strict, but we maintain strict quasiconcavity assumption throughout the current discussion. We are interested in computing the equilibrium of games in $\mathcal{G}$, and henceforth the term "game" will describe a $G \in \mathcal{G}$.

[4] The term Nash equilibrium in this paper shall always mean a pure strategy Nash equilibrium

3 The Best Response Dynamic

For any player i, define player i's *best response function* $br_i : A \mapsto A_i$ as a map from any state of the game $a \in A$ to player i's best response [5],

$$br_i(a) = \arg\max_{a_i \in A_i} u_i(a_i, a_{-i}).$$

We can now define a function $br : A \mapsto A$ using the vector of best response functions for each player. That is, for $a \in A$,

$$br(a) = (br_i(a), \ldots, br_n(a)).$$

It is clear that a Nash equilibrium of the game $a^* \in A$ is nothing more than a fixed point of the function br, i.e., $br(a^*) = a^*$. Expressing Nash equilibrium as the fixed point of br allows for the use standard theorems on the existence of fixed points to prove the existence of Nash equilibrium [4].

Expressing Nash equilibrium in this way also suggests a natural computational strategy: simply start at any state of the game $a^0 \in A$, and iterate the mapping br until a fixed point of it has been reached. A similar such process was first put forth by Cournot [3] as a dynamic way to motivate the Nash equilibrium concept itself. Cournot envisioned player taking turns best responding to one another. If such a process of asynchronous best responses settles down to a single point, then that point would be an equilibrium point. While the actual algorithmic steps of the best response dynamic can conceivably take many forms, to fix ideas, we specify a particular form.

Algorithm 1 (The Best Response Dynamic)**.**

1. *Pick an initial state* $a^0 \in A$ *and set the counter* $t = 1$.
2. *Pick a player* $i \in N$ *at random.*
3. *Find player* i*'s best response* $a_i^t = br_i(a^{t-1})$ *to the old state* $a^{t-1} \in A$.
4. *Set the new state* $a^t = (a_1^{t-1}, \ldots, a_{i-1}^{t-1}, a_i^t, a_{i+1}^{t-1}, \ldots, a_n^{t-1})$.
5. *Update the counter* $t = t + 1$.
6. *Repeat 2–5.*

Leaving aside the potential difficulty with actually computing a player's best response in step 3 of the algorithm, the best response dynamic is conceptually simple, intuitive, and most importantly, *decentralized*. We follow [12], and describe a process for finding Nash equilibrium as decentralized if "the adjustment of a player's strategy does not depend on the payoff functions (or utility functions) of the other players (it may depend on the other players' strategies, as well as on the payoff function of the player himself)." Since computing a player's

[5]Note that by the strict quasiconcavity assumption, the best response map is indeed a function.

best response $br_i(a)$ to a state of the game $a \in A$ requires only knowledge of his utility function u_i, the best response dynamic is clearly decentralized. Decentralized dynamics are the essential feature of learning models in Economics [12].

However the best response dynamic suffers from the fact that only under special monotinicity conditions on the utility functions u_i of the game, and the starting state $a^0 \in A$ of the process, will it actually converge [8]. It is this general lack of convergence of the best response dynamic, as well as a list of other such proposed simple learning dynamics that employ the best response functions[12], that has made computing a Nash equilibrium a non-trivial problem.

4 The Calculus Approach

As an alternative to decentralized learning dynamics, a "centralized" solution to the problem computing Nash equilibrium is to use certain tools from Calculus[6] to find a fixed point of br. More precisely, the Calculus approach uses a Newton-like method to search for a state of the game that satisfies each player's first order condition. In order for such an approach to work, some simplifying assumptions have to be made of the underlying game G beyond the fact it belongs to $\mathcal{G}$. These assumptions have become quite standard, and can be seen here echoed by Tirole in his well known book [22] on industrial organization:

> Often, u_i has nice differentiability properties. One can then obtain a pure-strategy equilibrium, if any, by differentiating each player's payoff function with respect to his own action ... The first-order conditions give a system of N equations with N unknowns, which, if solutions exist and the second order condition for each player holds, yield the pure-strategy Nash equilibrium.

The Calculus approach finds a fixed point of br (and hence a Nash equilibrium) by seeking a state a^* that solves the system of first order conditions

$$\begin{aligned} \frac{d}{da_1}u_1(a^*) &= 0 \\ \frac{d}{da_2}u_2(a^*) &= 0 \\ &\vdots \\ \frac{d}{da_n}u_n(a^*) &= 0. \end{aligned} \tag{1}$$

Let us succicntly express this sysem of equations as $f(a^*) = 0$. The most prevalent numerical method of solving the system (1) is by way of the Newton-Raphson method or some quasi-Newton variant of it [13]. This leads to the following search strategy for finding Nash equilibrium.

[6]We take the term Calculus to refer to broad spectrum of tools from Mathematical Analysis.

Algorithm 2 (The Calculus Approach)**.**

1. *Pick an initial state* $a^0 \in A$ *and set the counter* $t = 1$*.*
2. *Adjust from the old state* $a^{t-1} \in A$ *to a new state* $a^t \in A$ *according to the rule*
$$a^t = a^{t-1} + \mathbf{H}(a^{t-1})f(a^{t-1}). \tag{2}$$
3. *Update the counter* $t = t + 1$*.*
4. *Repeat 2–3.*

In (2), the standard Newton-Rhapson method uses

$$\mathbf{H}(a^{t-1}) = -[f'(a^{t-1})]^{-1},$$

where $f'(a^{t-1})$ is the Jacobian matrix of second derivatives at a^{t-1}. Quasi-Newton methods usually use an easier to compute matrix of values $\mathbf{H}(a^{t-1})$ at a^{t-1}. However regardless of weather a Newton or Quasi-Newton method is employed, it is clear that the dynamics of the Calculus approach are not decentralized – the adjustment of player $i's$ action at iteration t, $a_i^t - a_i^{t-1}$, depends on the value of every player j's own derivative of utility $f_j(a^{t-1})$ at a^{t-1}.

Aside from the problem of being centralized, and thus not interpretable as a learning process, the Calculus approach also does not generally find equilibrium for all the games in our class $\mathcal{G}$. In order for this Calculus approach to work, that is, in order to solve the fixed point problem $br(a^*) = a^*$ by means of solving the system $f(a) = 0$ using Algorithm 4, some further restrictions on $\mathcal{G}$ are needed. These further restrictions can generally be lumped under one of the two following categories:

Neccesity Each utility function u_i is differentiable with respect to own action a_i. if the utility functions u_i are not differentiable, then player i's best response is not necessarily a solution to his first order conditions.

Practicality The function f must be "well behaved" enough so that Newton's method asymptotically converges to a^*. One way to ensure convergence is to assume f is continuously differentiable and the Jacobian matrix f' is invertible for all points in an open subset of $\mathbb{R}^n$ that contains the solution $a^* \in \mathbb{R}^n$ to $f(a) = 0$. Further, starting values for the algorithm must be chosen that are contained in a sufficiently small ball centered at a^*. The radius of the ball that must contain the starting values of course depend on the structure of f [13].

Thus the Calculus approach offers a centralized solution to the problem of computing Nash equilibrium for the restricted class of games $\mathcal{G}' \subset \mathcal{G}$ whose utility functions and first order conditions well behaved. However, if any of the assumptions of the Calculus approach are not met, then finding a fixed point of br, and hence Nash equilibrium, becomes problematic once again.

5 The Stochastic Response Dynamic

5.1 Introduction

We now introduce a new approach to computing Nash equlibrium, which we shall call the *stochastic response dynamic*. The stochastic response dynamic provides a decentralized approach to computing equilibrium that avoids the first order conditions of the game entirely, which are critical to the Calculus approach. Thus under the stochastic response dynamic, every time a player updates his action, he does so using only knowledge of his own payoff function, and can thus act independently of the other players. Moreover, only knowledge of the payoff function itself and not its derivatives are needed.

The stochastic response dynamic follows the exact same algorithmic steps as the best response dynamic, except that best responses are replaced by a certain form of stochastic responses. Thus starting from any initial state, players continue to take turns adjusting their actions, except instead of best responding to the previous state of the game, a player now stochastically responds. Depending on the level of variance of these stochastic responses, the stochastic response dynamic can be made more or less "noisy".

A key feature of the stochastic response dynamic is that it patches up the lack of convergence problem of the best response dynamic. By replacing best responses with stochastic responses, it transforms the previously deterministic best response dynamic into a stochastic process. This stochastic process *can* be assured to converge. However the concept of "convergence" is no longer the same as was the case with the best response dynamic. Rather than deterministically converging to a single state of the game, the stochastic response dynamic converges in law to an equilibrium *probability distribution* over the state space of the game.

Our main results aims to establish a computationally relevant relationship between the equilibrium distribution of stochastic response dynamic and the Nash equilibrium of the underlying game. For games in $\mathcal{G}$ that additionally satisfy a natural game theoretic requirement that we call *Nash connectedness*, which we conjecture to be true for all games in $\mathcal{G}$, but at present can only prove for two player scalar action games, we establish the following main result. As the level of "noise" in the process gets small, the equilibrium distribution of the stochastic response dynamic converges (in distribution) to a denegerat distribution over the set of Nash equilibria of the game.

Thus simulating the stochastic response dynamic at a fixed noise level asymptotically produces a random draw from the corresponding equilibrium distribution. As we let the noise level approach zero, the resulting sequence of equilibrium draws converges in distribution to a draw from the set of Nash equilibria of the game.

Computing Nash equilibrium thus amounts to simulating the stochastic response dynamic for a number of iterations, then decreasing the amount of noise according to some rate, and repeating the simulation. For a fixed noise level, simulating the stochastic response dynamic will result in a sample being drawn from the equilibrium distribution. As we decrease the noise level, this sample will be increasingly concentrated over the Nash equilibria states of the game. In the limit, as the noise approaches zero, the sample approaches being entirely concentrated over the Nash equilibria states of the games, and the simulation gets stuck near an equilibrium point.

Our condition, Nash connectedness, essentially states that if any closed subset X of the joint action space of the game is also closed under single player better responses (i.e. if any one player chooses a better response to any state of the game in X, then the resulting new state is also in X), then X must contain a Nash equilibrium a^*. The reason this is a natural game theoretic assumption is because Nash equilibrium is defined as a closed set (namely a singleton $\{a^*\}$) that is closed under single player better responses. If there is a closed set X that is closed under better responses, but does not contain Nash equilibrium, then this set X functions just like a Nash equilibrium. That is, if the state of the game is in X, then no player has the incentive to leave X. Thus X is an equilibrium "solution" of the game that is bounded away from any Nash equilibria. This of course undermines the predictive value of Nash equilibrium.

So long as such sets X cannot exist, which we conjecture to be true of games in $\mathcal{G}$ but can only prove for two player games in $\mathcal{G}$ at present, then the stochastic response dynamic computes equilibrium. We present our case that Nash connectedness appears to be satisfied for games in $\mathcal{G}$ via computational examples that demonstrate the stochastic response dynamics in a range of environments.

5.2 Stochastic Response Functions

Our new approach, the stochastic response dynamic, replaces the best responses of the best response dynamic with stochastic responses. However what is a stochastic response? Recall the behavioral assumption that underlies the idea of a best response. Player i reacts to the current state of the game $a^t \in A$ by choosing a response a_i^{t+1} from his action set A_i that maximizes his utility holding fixed the actions of the other players a_{-i}^t, i.e.,

$$a_i^{t+1} = \arg \max_{a_i \in A_i} u_i(a_{-i}^t, a_i).$$

A stochastic response on the other hand, is generated by a different underlying behavior than utility maximization. Instead of best responding, Player i reacts to the current state of the game $a^t = (a_1^t, \ldots, a_i^t, \ldots, a_n^t) \in A$ by first drawing a candidate response response a_i' uniformly at random from A_i. Player i then chooses his response a_i^{t+1} by *probabilistically* choosing between his candidate response a_i' and his current action a_i^t, *holding fixed* the current actions of the other players a_{-i}^t. If player i chooses to set his response $a_i^{t+1} = a_i'$, then we say he has chosen to accept his candidate response. If player i chooses to set his response $a_i^{t+1} = a_i^t$, then we say player i has chosen to reject his candidate response.

To complete the description of stochastic responses, we let the choice probability of accepting the candidate response be governed by the classical binary logit choice model [23]. Thus conditional upon drawing the candidate response a_i', the probability that player i accepts his candidate response a_i', when the current state of the game is a^t, is given by

$$\mathbf{P}_i(a_i', a^t) = \frac{1}{1 + \exp[-(u_i(a_{-i}^t, a_i') - u_i(a_{-i}^t, a_i^t))/T]}. \tag{3}$$

The key parameter entering the logit choice probability in (3) is the variance parameter $T > 0$. As T gets arbitrarily large, the logit choice probabilities approach one half for each of the options a_i' and a_i^t. As T decreases, the choice probabilities increasingly favor the option that yields higher utility. In the limit, as the T tends towards zero, the player's behavior approaches necessarily choosing the option that yields the higher utility. Thus we can think of the variance parameter T in the logit model as the amount of "noise" entering a stochastic response. To make the dependence of the logit choice probabilities on T explicit, we modify the above notation in (3) from $\mathbf{P}_i(a_i', a^t)$ to $\mathbf{P}_i^T(a_i', a^t)$.

We can now formalize the idea of stochastic responses, and the stochastic response dynamic, by defining *stochastic response functions.* For a measureable set S, let ΔS denote the set of probability measures over S. Player i's stochastic response function $sr_i^T : A \mapsto \Delta A_i$ maps a state of the game $a \in A$ to a probability distribution $sr_i^T(a) = sr_i^T(\,\cdot \mid a\,)$ over player i's set of own actions A_i. Thus for each measurable $B \subset A_i$,

$$sr_i^T(\,B \mid a) \in [0,1]$$

is the probability of a player i responding to the state of game $a \in A$ by choosing an action in B.

Using our description of behavior underlying stochastic responses, and letting $|A_i|$ denote the volume of player i's strategy set (which recall is a closed rectangle in $\mathbb{R}^{m_i}$), we have

$$\begin{aligned} sr_i^T(\,B \mid a\,) &= \int_B \frac{1}{|A_i|}\mathbf{P}_i^T(x,a)\,dx + \left(1 - \int_{A_i} \frac{1}{|A_i|}\mathbf{P}_i^T(x,a)\,dx\right)\mathbf{1}_{a_i}(B) && (4)\\ &= f_i^T(B,a) + g_i^T(a)\mathbf{1}_{a_i}(B). && (5)\end{aligned}$$

The term $\mathbf{1}_{a_i}(B)$ is an indicator function for the condition $a_i \in B$. The term $f_i^T(B, a_i)$ reflects the probability of drawing a candidate $a' \in B$, and accepting a' when the current state of the game is $a \in A$. The term $g_i^T(a)$ reflects the probability of player i not accepting any possible candidate response a' when the current state of the game is a, and hence continuing to respond using the current action a_i.

5.3 Stochastic Response Dynamic

Having established the meaning of stochastic response functions sr_i^T as maps from a state of the game $a \in A$ to a particular probability distribution $sr_i^T(a) = sr_i^T(\,\cdot \mid a\,)$ over own actions A_i, we can now formally define the stochastic response dynamic. The stochastic response dynamic replaces the best responses of the best response dynamic with stochastic responses.

Algorithm 3 (The Stochastic Response Dynamic)**.**

1. *Pick an initial state* $a^0 \in A$ *and set the counter* $t = 1$.
2. *Pick a player* $i \in N$ *at random.*
3. *Sample player* i*'s stochastic response* $a_i^t \sim sr_i(a^{t-1})$ *to the old state* $a^{t-1} \in A$.
4. *Set the new state* $a^t = (a_1^{t-1}, \ldots, a_{i-1}^{t-1}, a_i^t, a_{i+1}^{t-1}, \ldots, a_n^{t-1})$.
5. *Update the counter* $t = t + 1$.
6. *Repeat 2–5.*

The central feature of the stochastic response dynamic, in contrast to the best response dynamic, is that it defines a *Markov chain* over the state space A of the game. Consider the sequence of states $\{a^t\}_{t\in\mathbb{N}}$ generated by the stochastic response dynamic. For every time $t \in \mathbb{N}$, tomorrow's state a^{t+1} differs from todays state a^t in the action of only one randomly chosen player i, i.e., $a_{-i}^{t+1} = a_{-i}^t$. In addition, the randomly chosen player i's action tomorrow $a_i^{t+1} \in A_i$ is a random variable drawn from player i's stochastic response $sr_i^T(a^t)$ to the state of the game today a^t. Thus for every $t \in \mathbb{N}$, a^{t+1} is a random variable taking a value in the state space A, whose probability distribution depends on the realization of $a^t \in A$. Hence by definition, the discrete time process $\{a^t\}_{t\in\mathbb{N}}$ generated by the stochastic response dynamic is a Markov chain.

Moreover, the Markov chain $\{a^t\}_{t\in\mathbb{N}}$ is a *time homogenous* Markov chain. That is, the "law of motion" that relates the probability distribution of tomorrow's state $a^{t+1} \in A$ to the realization of today's state $a^t \in A$ does not depend upon the time t, which follows from the fact the stochastic response functions are time invariant. Thus we can express the law of motion governing the Markov chain $\{a^t\}_{t\in\mathbb{N}}$ by way of a time invariant transition kernel,

$$K^T(\, a, B\,) = P^T(\, a^{t+1} \in B \mid a^t = a\,),$$

for any $a \in A$ and measurable $B \subset A$ and time $t \in \mathbb{N}$. Note that we have used the superscript T to denote dependence of our Markov transition kernel on the "noise level" entering the stochastic response dynamic. For any state $a \in A$, $K^T(a, \cdot)$ is a probability distribution over the state space A.

We now complete our description of the stochastic response dynamic as a Markov chain by showing how the transition kernel K^T is constructed from the steps comprising the stochastic response dynamic . If the game is currently at the state $a \in A$, then the stochastic response dynamic first picks a player i at random, and then replaces the i^{th} component of the state vector $a = (a_1, \ldots, a_i, \ldots, a_n)$ with a random draw from player i's stochastic response function $sr_i(a)$. Thus conditional upon player i being the chosen player, we can define the *conditional* transition kernel $K_i^T(a, B)$ for any measurable $B \subset A$,

$$K_i^T(a, B) = sr_i^T(\{b_i \in A_i : (a_{-i}, b_i) \in B\} \mid a).$$

The unconditional transition kernel K^T is then simply a mixture over the conditional transition kernels, i.e.,

$$K^T(a, B) = \frac{1}{n} \sum_{i=1}^{n} K_i^T(a, B). \tag{6}$$

Thus for a fixed "noise level" T, which recall indexes the level of variability of the logit choice probability in the stochastic response dynamic produces a Markov chain $\{a^t\}_{t \in \mathbb{N}}$ whose motion is described by the transition kernel K^T.

5.4 Convergence to Stationarity

We are now ready to consider the asymptotic behavior of the stochastic response dynamic. The key feature of this behavior is that it avoids the lack of convergence problem of the best response dynamic. To see this, suppose we simulate the stochastic response dynamic for a fixed noise level T. Thus we start the process at an arbitrary state $a^0 \in A$, and grow the process by iteratively drawing $a^t \sim K^T(a^{t-1}, \cdot)$. A natural question to ask of such a process, and a central questions in the study of Markov chains more generally [17], is weather the behavior of the random variables a^t becomes "stable" as t grows large? In the case of the stochastic response dynamic, unlike the best response dynamic, we are able to answer this question in the affirmative.

Recall that in the case of the best response dynamic, stability of the process $\{a^t\}_{t \in \mathbb{N}}$ means that it determinstically converges to a stable point a^* in A, i.e., a point a^* such that if the initial state of the process $a^0 = a^*$, then all the following states of the process a^t would remain equal to a^*. The stable point a^* would thus correspond to a Nash equilibrium. The main problem with the best response dynamic however is that such a convergence to equilibrium cannot generally be assured for most starting values. Further conditions on the shape of the best response functions are needed to ensure convergence, which are no longer natural primitives to impose on the game.

In the case of the stochastic response dynamic however, stability of the process $\{a^t\}_{t \in \mathbb{N}}$ means that it converges to a stable *probability distribution* P^T over A, i.e., a distribution P^T such that if the initial state of the process a^0 is distributed according to P^T, then all the following states of the process a^t would remain distributed according to P^T. Recall from Markov chain theory that if the initial state a^0 of the chain $\{a^t\}_{t \in \mathbb{N}}$ is generated according to some probability measure Q^0 over A, then marginal probability distribution Q^t of a^t is given inductively by

$$Q^t(B) = \int K^T(x, B) Q^{t-1}(dx),$$

for measurable $B \subset A$.

The key feature of the stochastic response dynamic is that for any choice of Q^0, the marginal distributions Q^t converges *strongly* (i.e. in total variation norm) to a unique limiting probability measure[7] P^T. This follows from the fact, as we shall show, that the Markov chain associated with the transition kernel K^T satisfies strong *ergodic* properties. In particular, the Markov chain defined by K^T is *positive*, *irreducible*, and *Harris recurrent* [17]. From these properties, it follows that for a sufficienctly large number of iterations $j_T \in \mathbb{N}$, the state of the process a^{j_T} will be distributed as approximately (for arbitrary small approximation error in total variation norm) P^T, i.e.,

$$a^{j_T} \sim P^T. \tag{7}$$

We refer to the distribution P^T as the *equilibrium* of the stochastic response dynamic. Moreover the equilibrium distribution P^T will also be a stable or invariant probability distribution for the process $\{a^t\}_{t\in\mathbb{N}}$ in the sense defined above.

Thus for any game in our class, the stochastic response dynamic solves the instability problem of the best response dynamic by way of simulating a Markov chain with transition kernel K^T that has an equilibrium distribution P^T. However the fact that it solves the lack of stability problem of the best response dynamic does not yet mean it necessarily solves our problem, which is to find Nash equilibrium.

The equilibrium distribution P^T of the stochastic response dynamic does not yet bear any obvious relation to the Nash equilibria of the underlying game. However our main result shows that for our class $\mathcal{G}$ of games, which we conjecture satisfies the natural game theoretic condition that we call *Nash connectedness*, P^T is in fact related to the Nash equilibrium of the underlying game in a computationally quite relevant way, which we explain next.

5.5 Convergence to Nash

Thus far we have considered the stochastic response dynamic for only a *fixed* noise level T. We have argued that for each fixed noise level $T > 0$, the stochastic response dynamic with kernel K^T has a unique equilibrium distribution P^T. For T infinitely large, which corresponds to the stochastic response dynamic becoming infinitely noisy, the equilibrium distribution P^T becomes maximally variant, and approaches a uniform distribution over the state space A.

However as the "noise" in the stochastic response dynamic gets smaller and smaller, the sequence of equilibrium distributions P^T converges in distribution to a limit distribution P. Our main result is that for games in $\mathcal{G}$ that further satisfy a condition we call *Nash Connectedness*, which we conjecture holds true for all games in $\mathcal{G}$, and prove holds true for two player games in $\mathcal{G}$, the support of P consits entirely of Nash equilibria. Thus for small T, the equilibrium

[7] We maintain the T index as reminder that this limiting measure will depend on the noise level T of the stochastic response dyanamic.

distribution P^T becomes increasingly concentrated over the the Nash equilibria states of the game. Thus if we produce a draw from the distribution P^T for small T, the probability that this draw if further than any epsilon from a Nash equilibrium goes to zero as T goes to zero.

In light of the main result, computing Nash equilibrium via the stochastic response dynamic becomes clear. Start at some large $T > 0$, and simulate the stochastic response dynamic. The subsequent values a_T^t for $t = 1, 2, \ldots, etc$ of the process can then be approximately regarded as a sample from its equilibrium distribution P^T. For large T, this sample will resemble a uniform distribution over A. After some pre-specified number of iterations (call it the threshold number of iterations), decrease the noise level T by some factor, and repeat.

As T gets smaller, the equilibrium distribution P^T, and hence the simulated values a_T^t becomes more concentrated over the Nash equilibrium states of the game. Eventually, the process will get "stuck" in a state $a^{**} \in A$ for which no candidate responses are accepted after the threshold number of attempts. This state a^{**} is by definition an approximate Nash equilibrium since no player is willing to move under the stochastic response dynamic.

This process of learning Nash equilibrium, which combines the basic stochastic response dynamic with a "noise reduction" schedule, bears a resemblance to "annealing" methods in physics. Such annealing methods first heat up a physical system so as to make its particles move around rapidly, which is the analogue of the equilibrium distribution of the stochastic response dynamic. The system is then cooled, i.e., made less noisy, so as to bring the system to its equilibrium motionless or crystal state – which is analogous to the stochastic response dynamic coming to rest on a Nash equilibrium.

6 Where Does it Come From and Why Does it Work

We have just made some bold claims about the behavior of the stochastic response dynamic for our class of games $\mathcal{G}$. In this section we seek to present the basic idea that motivates thee stochastic response dynamic as a new way to learn and compute equilibrium. This motivation provides proof of its functionality in a somewhat more specialized setting than the entire class $\mathcal{G}$. We then lay the foundations for the generalizability of the proof to the entire class of games $\mathcal{G}$. Finally, we conclude with computational experiments.

The fundamental new idea is to recognize that strategic games bear a very natural relationship to conditionally specified probability models. A conditionally specified probability model is a system of conditional probability density functions $p_i(x_i \mid x_{-i})$ for $i = 1, \ldots n$ that describes the joint distribution of a random vector $(X_1, \ldots, X_n)$. We now show how it is possible to transform any given strategic game (even though we are only interestd in games in $\mathcal{G}$) into such a system of conditional probability density functions (conditional pdf's for short) without "changing" the game (i.e., preserving the Nash equilibria).

Using this transformation, the motivation for the stochastic response dynamic becomes clear.

6.1 Games as Conditional Probabilities

The first step is to recognize that for the purposes of Nash equilibrium, the variation of each utility function u_i over the entire space A is not significant, but only the variation of u_i over own actions A_i for each fixed profile a_{-i} of the other players' actions. It then becomes natural to speak of each player having a conditional utility function $u_i(\, a_i \mid a_{-i}) = u_i(a_{-i}, a_i)$ over the axis A_i for each fixed profile $a_{-i} \in A_{-i}$ of other player actions. Thus instead of viewing a game as n utility functions, we can equivalently think of it as a system of n conditional utility functions $u_i(\, a_i \mid a_{-i}\,)$. A Nash equilibrium $a^* \in A$ is a point that maximizes each player's conditional utility function,

$$a_i^* = \arg\max_{a_i \in A_i} u_i(\, a_i \mid a_{-i}^*\,).$$

However once we express the game as a system of conditional utility functions, it begins to "look" very similar to a situation from standard probability theory. We will make this similarity apparent by first normalizing each conditional utility function $u_i(\, a_i \mid a_{-i}\,)$ to be strictly positive and integrate to one as a function of a_i. One way of doing this is using the tranformation

$$\begin{aligned} p_i^T(\, a_i \mid a_{-i}\,) &\propto \exp\left(\frac{u_i(\, a_i \mid a_{-i}\,)}{T}\right) \\ &= c_{a_{-i}}^T \exp\left(\frac{u_i(\, a_i \mid a_{-i}\,)}{T}\right) \end{aligned} \tag{8}$$

The parameter T is a positive real number that has the same interpretation as the noise parameter in the stochastic response dynamic, but we shall make this connection later. For now, notice only that a small T makes the conditional pdf p_i^T more "peaked" around its mode (p_i^T is unimodal because of the strict quasiconcavity assumption made of games in $\mathcal{G}$), and for large T, p_i^T becomes flat (uniform) over A.

Notice that there is a constant $c_{a_{-i}}^T$ of proportionality (depending on both T and a_{-i}) needed so that the resulting normalized conditional utility function $p_i^T(\, a_i \mid a_{-i}\,)$ satisfies the property of integrating to one over A_i. At no point do we ever need to compute this constant, but it is useful to be aware that it exists given the assumptions we have made about the game (recall A is compact and thus a normalization constant exists).

The next step is to notice that normalizing the conditional utility functions does not change Nash equilibrium. That is, $a^* \in A$ also satisfies

$$a_i^* = \arg\max_{a_i \in A_i} p_i^T(\, a_i \mid a_{-i}^*\,). \tag{9}$$

Thus we have gone from n utility functions u_i to n conditional utility functions $u_i(\, a_i \mid a_{-i}\,)$ to n normalized conditional utility functions $p_i^T(\, a_i \mid a_{-i}\,)$.

The advantage to us of working with $p_i^T(\, a_i \mid a_{-i}\,)$ is that it not only provides an equivalent model of the original strategic game, but it also serves as a *conditionally specified probability model*.

That is, each $p_i^T(\, a_i \mid a_{-i}\,)$ can be interpreted as a conditional probability density function, since as a function of a_i, it is strictly positive and integrates to one. Thus one can postulate the existence of a vector of random variables $(X_1, \ldots, X_n)$ having support $A = A_1 \times \cdots \times A_n$, and whose conditional distributions are described by the conditional probability density functions $p_i^T(\, a_i \mid a_{-i}\,)$. We also observe that the conditional pdf's $p_i^T(\, a_i \mid a_{-i}\,)$ are equal to the *quantal response functions* of McKelvey and Palfrey [21] (which are distinct from our stochastic response functions), but generalized to continuous action sets.

The system of conditional pdf's $p_i^T(\, a_i \mid a_{-i}\,)$, or quantal response functions, *conditionally specifies* a probability model for a random vector $X = (X_1, \ldots, X_n)$ having support $A = A_1 \times \cdots \times A_n$. That is, the system $p_i^T(\, a_i \mid a_{-i}\,)$ of conditional pdf's specifies all the conditional distributions of the random vector $(X_1, \ldots, X_n)$.

6.2 Compatibility

When we are given the n utility functions u_i that define a game, there is no general way to combine them into a single function u over A whose characteristics bear any obvious relation to the Nash equilibria of the game. The closest thing we have to such a method of combining the utility functions is the Calculus approach, which forms the system of first order conditions and associates Nash equilibria with the roots of the system.

However the key innovation in thinking of a game in terms of the conditional pdf's $p_i^T(\, a_i \mid a_{-i}\,)$ is that it allows us to use tools from probability to combine the utilities. In particular, let us assume that the system of conditional pdf's $p_i^T(\, a_i \mid a_{-i}\,)$ that we derive from G are *compatible*. That is, assume that there exists a joint pdf $p^T(a_1, \ldots, a_n)$ that has conditional distributions equal to $p_i^T(\, a_i \mid a_{-i}\,)$. More formally, the conditional pdf's p_i^T are compatible if there exists a real valued function q^t over A such that for any i, and any $a_{-i} \in A_{-i}$,

$$p_i^T(\, a_i \mid a_{-i}\,) = K_{a_{-i}} q^T(a_1, \ldots, a_n), \tag{10}$$

for a constant $K_{a_{-i}}$. If such a function exists, then it is clear that the joint pdf p^T having $p_i^T(\, a_i \mid a_{-i}\,)$ as its conditional pdfs is given by $p^T(a_1, \ldots, a_n) = cq^T(a_1, \ldots, a_n)$ for a constant of proportionality that enables p^T to integrate to 1. Thus (10) expresses the usual relation that the joint distribution of a random vector $(X_1, \ldots, X_n)$ is proportional to its conditionals. By a basic theorem in probability theory, known as the *Hammersley-Clifford Theorem* [9], if such a joint distribution p^T exists, then under mild regularity conditions, which are met by our conditionals p_i^T because they have full support A_i, this joint is uniquely determined.

In fact, for our conditional system, if such a joint p^T exists for a single T, say $T = 1$, it follows that a joint p^T will exist for all $T > 0$. We now state this as a lemma.

Lemma 1. *If the conditional pdf's* $p_i^T(\, a_i \mid a_{-i}\,)$ *that we derive from* G *are compatible for* $T = 1$*, then they are compatible for all* T*.*

Proof. Consider any i and any $a_{-i} \in A_{-i}$. Then by (8),

$$p_i^T(\, a_i \mid a_{-i}\,) = \frac{c_{a_{-i}}^T}{c_{a_{-i}}^1} p_i^1(\, a_i \mid a_{-i}\,)^{1/T}$$

However since the conditional pdf's for $T = 1$ are compatible, then for some $K_{a_{-i}}$

$$p_i^T(\, a_i \mid a_{-i}\,) = K_{a_{-i}}^{1/T} \frac{c_{a_{-i}}^T}{c_{a_{-i}}^1} p_i^1(a_1, \ldots, a_n)^{1/T}.$$

Thus it follows that the joint pdf $p^T(a_1, \ldots, a_n)$ for T is equal to $c_T p^1(a_1, \ldots, a_n)^{1/T}$, for a constant c_T that allows the expression to integrate to 1. □

Furthermore we note that that if the conditional pdf's are compatible for $T = 1$, then the following holds.

Lemma 2. *Any local maxima (mode) of* $p^1(a_1, \ldots, a_n)$ *is a Nash equilibrium of the game.*

Proof. This follows simply by noting that any local maxima of the joint pdf p^1 globally maximizes each conditional pdf $p_i^1(\, a_i \mid a_{-i}\,)$ over own actions A_i (recall each conditional pdf is unimodal by the strict quasiconcavity assumption). Thus by (9), such a local maxima is a Nash equilibrium. By similar logic, the proposition works in the other direction as well - a Nash equilibrium must be a local maximum of p^1. □

Now we come to the fundamental observation. If any local maxima $p^1(a_1, \ldots, a_n)$ is a Nash equilibrium, and the joint pdf $p^T(a_1, \ldots, a_n) = c_T p^1(a_1, \ldots, a_n)^{1/T}$, then as T goes to 0, p^T becomes increasingly concentrated over Nash equilibrium states. In the limit, p^T converges in distribution to a degenerate distribution over the global maxima of p^1, and thus a subset of the Nash equilibrium states.

6.3 Markov Chain Monte Carlo

We have just shown that

$$p = \lim_{T \to 0} p^T$$

is a degenerate distribution over the Nash equilibria states of the game, where the limiting operation corresponds to convergence in distribution. However our original question concerned

$$P = \lim_{T \to 0} P^T,$$

where recall P^T is the stationary distribution of the stochastic response dynamic. Under the compatibility assumption, it is a fact that the stationary distribution P^T of the stochastic response dynamic *equals* the compatible joint

p^T. This is because the stochastic response dynamic is nothing more than a form of Markov Chain Monte Carlo when the system of conditional pdf's is compatible.

Markov Chain Monte Carlo is concerned with the construction of a Markov Chain M over a given state space S whose equilibrium distribution is equal to some given probability distribution of interest π over S. One of the first Markov Chain Monte Carlo algorithms to emerge was the Metropolis algorithm [16], or more precisely the "variable at a time" Metropolis algorithm [2]. If the state space can be expressed as a product $S = S_1 \times \cdots \times S_n$, then the Metropolis algorithm shows how to construct a Markov chain M whose equilibrium distribution is π. It constructs this chain by using the information contained conditionals distributions $\pi(\, s_i \mid s_{-i}\,)$.

In terms of our state space A, and our joint pdf $p^T(a_1, \ldots, a_n)$ over A that we derived from the conditional pdf's $p_i^T(\, a_i \mid a_{-i}\,)$ and the compatibility assumption, the Metropolis algorithm provides a recipe for constructing a Markov chain whose equilibrium distribution is p^T. This recipe is as follows:

Algorithm 4 (Metropolis Algorithm for Games)**.**

1. *Start the process at any state* $a^0 \in A$ *and set the counter* $t = 1$.

2. *Pick a player* $i \in N$ *at random.*

3. *Draw a candidate* a_i' *action from a uniform distribution over* A_i.

4. *Set the new state* $a^{t+1} = (a_1^t, \ldots, a_{i-1}^t, a_i', a_{i+1}^t, \ldots, a_n^t)$. *with probability*

$$\mathbf{P}_i(a_i', a^t) = \frac{1}{1 + \left(\frac{p_i^T(\, a_i^t | a_{-i}^t \,)}{p_i^T(\, a_i' | a_{-i}^t \,)} \right)}. \tag{11}$$

   and set $a^{t+1} = a^t$ *with probability* $1 - \mathbf{P}_i(a_i', a^t)$.

5. *Update the counter* $t = t + 1$.

6. *Repeat 2–4.*

The acceptance probability expressed in (11) is one version of the *Metropolis criterion*. The remarkable feature of the Metropolis algorithm is that the only information it uses about the joint distribution p^T is conveyed by the Metropolis criterion. This has proved powerful in a range of applications for which some joint distribution π over a state space $S = S_1 \times \cdots \times S_n$ needs to be studied, but π may only be known up to its set of conditional distributions $\pi_i(\, s_i \mid s_{-i}\,)$ for $i = 1, \ldots, n$, and each π_i may only be known up to a constant of proportionality.

If we rewrite the Metropolis criterion in terms of the underlying utility functions of our game by expanding the definition of p_i^T given in (8), it yields

$$\frac{1}{1 + \exp[-(u_i(a_{-i}^t, a_i') - u_i(a_{-i}^t, a_i^t))/T]}. \tag{12}$$

But (12) this is nothing more than the probability of accepting a candidate response in (3) under the stochastic response dynamic. Thus the Metropolis algorithm for games and the stochastic response dynamic are the same.

Hence under the compatibility assumption, the standard results on the Metropolos algorithm allow us to conclude that the equilibrium distribution of the stochastic response dynamic P^T equals the compatible joint distribution p^T. Thus it follows from lemma 2 that

$$\lim_{T \to 0} P^T,$$

is a degenerate distribution over Nash equilibria states of the game, which is precisely the main result we wished to establish.

However in order to arrive at this result, we have had to make extensive use of the fairly strong assumption of compatibility, which will not generally hold true for the games in our class $\mathcal{G}$ of interest. The key to understanding the generalizability of the stochastic response dynamic to the whole class of games $\mathcal{G}$ is recognizing that even without the compatibility assumption, and hence without the existence of the joint pdf p^T, we can still show that a unique limiting equilibrium distribution P^T of the stochastic response dynamic exists by way of its properties as a Markov chain. Furthermore we can still prove that the sequence of equilibrium probability distributions P^T will still have a limit P as $T \to 0$. The key question is weather the P continues to be a degenerate distribution over Nash equilibria states of the game.

Our main theorem, which we present next, answers this question in the affirmative by replacing the compatibility assumption with a much more natural game theoretic condition that we call *Nash connectedness*. Nash connectedness is a key assumption for the predictive credibility of the Nash equilibrium concept. We conjecture Nash connectedness holds generally for games in $\mathcal{G}$, although it can only be proven for two player games in $\mathcal{G}$ at present. After establishing that our main results follows by replacing compatibility with connectedness, we move towards illustrating the computational utility of the stochastic response dynamic in oligopoly problems.

7 The Main Result

We will now relax the compatibility assumption, and replace it with a considerably weaker condition called Nash connectedness, which we conjecture holds true for games in our class $\mathcal{G}$. Using this connectedness assumption, we establish that the limit of the equilibrium distributions P^T as $T \to 0$ converges to a degenerate distribution having support over Nash equilibria states of the game.

7.1 Nash Connectedness

In order to establish Nash connectedness, we first need to define the following definitions.

Definition 1. *For a game G, the relation $>$ over A is defined such that $a^j > a^k$ iff for some player i, $a^k_{-i} = a^j_{-i}$ and $u_i(a^k) > u_i(a^j)$.*

Thus $a^j > a^k$ iff and only if a single player changes his action between a^j and a^k, and this player improves his utility. We can thus say that a^j and a^k are related by a single player improvement. A subset $B \subset A$ of the joint action space of the game is said to be *closed under single player improvements* if for any $a \in B$, $b \in A$ and $b > a$ implies $b \in B$. That is, for any state in B, a single player improvement of that state is also contained in B. Thus we can alternatively define Nash equilibrium a^* as simply a singelton set $\{a^*\}$ that is closed under single player improvements (i.e. the set of single player improvements from Nash equilibrium is the empty set).

Definition 2. *A chain γ in G is a sequence $\{a^0, a^1, \ldots, a^n\}$ in A such that $a^k > a^{k-1}$ for $k = 1, \ldots, n$.*

Definition 3 (Nash Connectedness)**.** *A game G possessing a pure strategy Nash equilibrium is* Nash connected *if for any $\epsilon > 0$, and any initial state $a^0 \in A$, there exists a chain $\gamma = \{a_0, \ldots, a_n\}$ such that a_n is in the epsilon open ball centered at some Nash equilibrium $a^* \in A$.*

Whereas compatibility was clearly a somewhat artifical assumption to impose on our class $\mathcal{G}$, Nash connectedness is an entirely natural assumption to impose. We conjecture that for games in $\mathcal{G}$, connectedness will be satisfied. At present, the connectedness assumption has only been proved for 2 player games in $\mathcal{G}$ by way of an adaptation of theorem 2 in [6]. The author is at present developing a proof for the general case based on the Lefschetz fixed point theorem.

One motivating idea to see the naturalness of connectedness is to think what it would mean for it to fail. Failure of connectedness would actually undermine the predictive value of Nash equilibrium. In order to make this point, we establish the following result.

Theorem 1. *For a continuous game G that has a Nash equilibrium a^*, Nash connectedness fails if and only if there exists a closed subset $B \subset A$ of the joint action space, which does not contain a Nash equilibrium and is closed under single player improvements.*

Proof. Obviously if such a set existed, then Nash connectedness fails for any point $b \in B$, since any Nash equilibrium has an epsilon ball around it that does not intersect B, and thus no chain from within B can enter any of these epsilon balls.

The other direction works as follows. Suppose Nash connectedness fails at a point a^0. Then build the following sets inductively. Let $S^0 = \{a^0\}$. Let $S^n = \{a \in A : a > b \text{ for some } b \in S^{n-1}\}$. Then let S^∞ be the countable union over the S^n sets for $n = 1, 2, \ldots, etc$, and $\bar{S}$ be the closure of S^∞. It follows that $\bar{S}$ is closed (by construction). Moreover, since Nash connectedness is being assumed to fail, $\bar{S}$ does not contain a Nash equilibrium. We wish to show that $\bar{S}$ satisfies the property of B in the proposition, namely, it is closed under single player improvements.

Suppose that $\bar{S}$ was not closed under single player improvements, i.e. there is some point $x \in \bar{S}$ that has a single player improvement $z \in A$ not in $\bar{S}$. Then the failure must occur at a limit point x of S^∞ because otherwise x would be contained in an S^n for finite n, and thus z would be in S^{n+1}.

Let us say without loss of generality that state z is a single player improvement for player i from state x.Thus we have that $x = (x_i, x_{-i})$ and $z = (z_i, x_{-i})$ for some z_i not equal to x_i. However since x is a limit point of S^∞, then there is a sequence of points x^n for $n = 1, 2, \ldots, etc$ in S^∞ converging to x. However by definition of z, $u_i(z_i, x_{-i}) > u_i(x_i, x_{-i})$, and thus for x^n close to x it follows by continuity of u_i that $u_i(z_i, x^n_{-i}) > u_i(x^n_i, x^n_{-i})$. But this means that (z_i, x^n_{-i}) is in S^∞ for all large n. Thus the limit (z_i, x_{-i}) of (z_i, x^n_{-i}) is in $\bar{S}$ since $\bar{S}$ is the closure of S^∞. But this contradicts our assumption that $z = (z_i, x_{-i})$ was not in $\bar{S}$. Thus it must be in $\bar{S}$.

□

Here then is the pathology. Nash connectedness is equivalent to the existence of a closed set $B \subset A$, which is strictly bounded away from any Nash equilibria, but would be as potent a solution for the game as any Nash equilibrium a^*. That is, if someone announced to the players that the outcome of the game will live B, then this prediction would be self reinforcing because no player would want to deviate from B, i.e., B is closed under single player improvements. This undermines the predictive value of Nash equilibrium - there is no a priori reason to assume the outcome of the game will be in the set $\{a^*\}$ for some Nash equilibrium a^*, or B, which is bounded away from any Nash equlibrium. Both sets satisfy the same rationality requirements of being closed under single player improvements, i.e., no player has the incentive to deviate from the solution.

We conjecture that for games in $\mathcal{G}$, i.e. quasiconcave game for which equilibrium is known to exist, this pathology cannot arise. At present the result is only known for 2 player games in $\mathcal{G}$. The proof of this follows by way of an adaptation of Theorem 2 in [6]. The author is currently developing a proof for the general case based on an application of the Lefschetz fixed point theorem (which allows for proving fixed point results without requring the domain of the underlying function be convex).

7.2 Main Theorem

Theorem 2. *For games in $\mathcal{G}$, if Nash connectedness is satisfied, then the equilibrium distribution P^T converges in distribution to a degenerate distribution over Nash equilibria states of the game as the noise $T \to 0$.*

Proof. We present the logical flow of the proof, and for the sake of length, the mathematical details are left to the references. Without compatibility, we need to first show that an invariant probability measure P^T for the Markov kernel K^T exists. Continuity of the utility functions u_i allows us to conclude that the the kernel K^T for any $T > 0$ will will be define a weak *Feller* chain [17]. Since our state space A is compact and by property of Feller chains, we can assure a fixed point P^T for the kernel K^T exists, which will be an invariant probability measure for the process.

Secondly we need to show that for $T > 0$, P^T is the unique invariant measure, and the chain will strongly converge to P^T, i.e. in total variation norm. This follows from the irreducible and aperiodic behavior of the stochastic response dynamic. From any one point in the state space, the stochastic response dynamic can with positive probability travel to any other open set in a finite number of steps, which means it is irreducible. Moreover, there is a strictly positive probability that starting from any state, we can enter any open set of A in n iterations of the chain (where n is the number of players). Thus the chain satisfies a strong recurrency condition known as Harris recurrence. Also, since there is always some positive probability of staying at the current state after one iteration (i.e. rejecting the candidate response), the chain is also aperiodic. By standard ergodic theorems for irreducible Harris recurrent chains that admit invariant measures, found in [17], P^T is the equilibrium distribution, i.e., the unique invariant probability measure for K^T to which the marginal distributions of the process converge strongly. We also have that P^T can be described by a density function, i.e. P^T is absolutely continuous with respect to Lesbegue measure, which follows from irreducibility.

Now we wish to understand what happens to P^T as $T \to 0$. First we recognize the kernel K^T has a limit K as $T \to 0$. The underlying stochastic response functions of this limiting kernel K consist of a player randomly picking a new action and changing to this action if it weakly improves utility. This kernel K also defines a weak Feller chain based on the continuity of the u_i. Clearly a Dirac (delta) measure concentrated over a Nash equilibrium state is an invariant probability measure for K. i.e., if we start at Nash, then under K, we stay there.

Through the Nash connectedness assumption, we can show that the only invariant probability mesures of the limiting kernel K are the Nash equilibria Dirac (delta) measures and their convex mixtures. If there existed any other invariant measure, then this would imply that there exists an ergodic invariant probability measure of K that puts positive probability on an open set not containing any Nash equilibria. By Nash connectedness, this invariant probability measure must contain a Nash equilibrium a^* in its support. However this would imply that the support of one ergodic probability invariant measure strictly contained the support of another ergodic invariant probability measure (namely the support of the Dirac measure concentrated at a^*), which is impossible for Feller chains on compact metric spaces as can be shown from the main result in Chapter 4 of [24]. Thus the Dirac measures concentrated at Nash equilibria states and their convex combinations are the only invariant measures for K. Said another way, every invariant measure measure of K has a support consisting entirely of Nash equilibria.

However K will of course not be irreducible, which is why we don't run the stochastic response dynamic at K but rather use noise $T > 0$. However a standard result from [5] shows that the sequence of equilibrium distributions P^T for K^T converge weakly (in distribution) to an invariant distribution of the limiting kernel K. But all invariants of K are supported over Nash equilibria staes. This would end the proof. □

8 Computational Examples – Cournot Equilibrium

Many problems in industrial organization are formulated quite naturally as N-player games. In this section, the behavior of the stochastic response dynamic is illustrated for a few simple cases of the the Cournot oligopoly model. Follow up work to this primarily theoretical paper uses the stochastic response dynamic to perform more substantive economic analysis, i.e., merger analysis.

The Cournot oligopoly models the strategic interactions that occur among N firms who each produce the same good. If each firm i chooses an amount q_i to produce, then the total quantity $Q = q_1 + q_2 + \cdots + q_N$ produced determines the price p of the good via the inverse market demand $p(Q)$. Furthermore, each firm i incurs a cost $c_i(q_i)$ to produce its chosen amount q_i. The choice each firm i faces is strategic because the profit $\pi_i(q_i, q_{-i}) = p(Q)q_i - c_i(q_i)$ it receives from producing q_i depends on the quantities q_{-i} produced by the other firms. Thus as a game, the Cournot oligopoly consists of:

1. N firms

2. An action set $A_i = [0, \bar{q}]$ for each firm i from which it chooses an amount q_i to produce. The upper limit $\bar{q}$ on each firm's actions set A_i is understood either as a capacity limit, or as the quantity above which the price on the market for the good is zero.

3. A utility function $u_i(q_i, q_{-i})$ for each firm i that equals its profit $\pi_i(q_i, q_{-i}) = p(Q)q_i - c_i(q_i)$ from having produced q_i and its competitors having produced q_{-i}.

Assuming different forms of market demand $p(Q)$ and firm specific costs $c_i(q_i)$ leads to a variety of possible Cournot oligopoly games. The stochastic response dynamic as a computational tool is now illustrated using some specific forms that are also easily solvable by means of the more traditional Calculus approach. Our purposes in these illustrations is to just convey an image of the stochastic response dynamic in practice – thus we have kept the examples intentionally fairly simple.

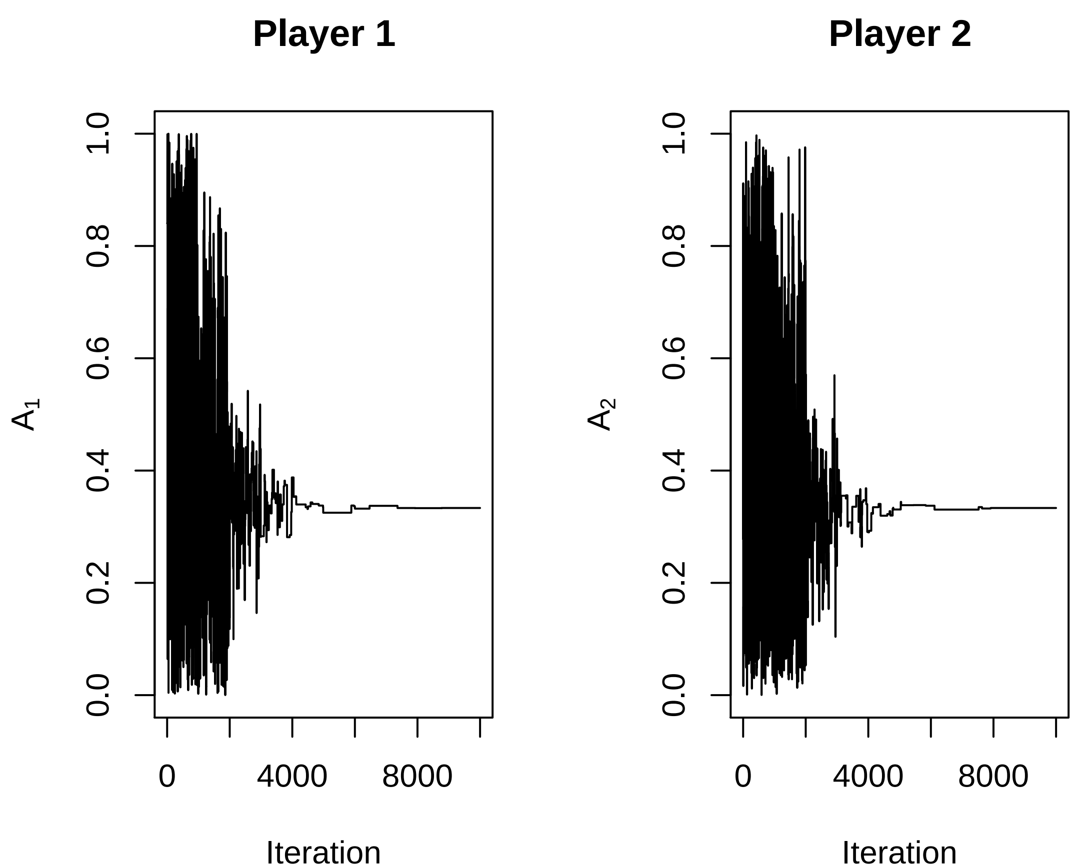
For each case of Cournot considered, the definition of the game is given, followed by the solution known by way of traditional technique, and then the stochastic response dynamic output. The algorithm is run at a starting noise level of 1, and reduced by .01 every 1000 iterations until a noise level is reached such that no candidate actions are accepted. The state of simulation in the final state is rounded to two decimal places is taken to be the computed solution of the game. We see that for each case, we successfully reproduces the known Cournot solutions.

8.1 Linear Demand No Cost Cournot

1. 2 Players
2. $A_1 = [0, 1]$ and $A_2 = [0, 1]$
3. $\pi_1(q_1, q_2) = q_1(1 - (q_1 + q_2))$ and $\pi_2(q_1, q_2) = q_2(1 - (q_1 + q_2))$

The Nash equilibrium of the game is $q_1^* = 1/3$ and $q_2^* = 1/3$.

The stochastic response dynamic resulted in the following simulation:



The simulation stopped in the state $\widehat{q_1} = 0.33$ and $\widehat{q_2} = 0.33$.

Stackelberg Cournot

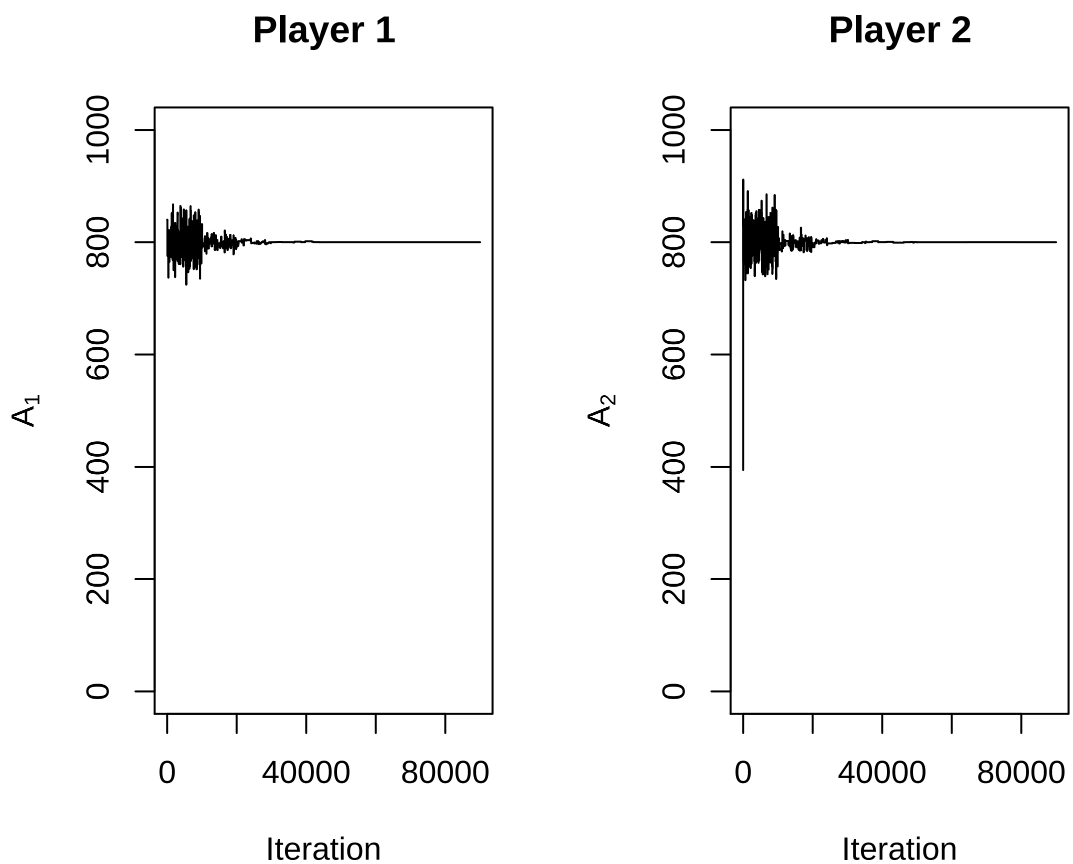
1. 2 Players

2. $A_1 = [0, 1000]$ and $A_2 = [0, 1000]$

3. $\pi_i(q_1, q_2) = q_i(100e^{-(1/10)\sqrt{q_1+q_2}})$ for $i = 1, 2$.

The Nash equilibrium of the game is $q_1^* = 800$ and $q_2^* = 800$ [15].

The stochastic response dynamic resulted in the following simulation.



The simulation stopped in the state $\widehat{q_1} = 800.03$ and $\widehat{q_2} = 799.98$.

Hard Cournot

This is an example of a more challenging Cournot computation. It is outside the realm of the previous examples in that it cannot be done by hand, and has been used for demonstrating numerical techniques for finding equilibrium under the traditional approach.

In this case, there are 5 firms. Each has a cost function

$$C_i(q_i) = c_i q_i + \beta_i/(\beta_i + 1)5^{-1/\beta_i} q_i^{(\beta_i+1)/\beta_i},$$

for firm $i = 1, 2, 3, 4,$ and 5. Furthermore the industry demand curve for their common product is

$$p = (5000/Q)^{1/1.1}.$$

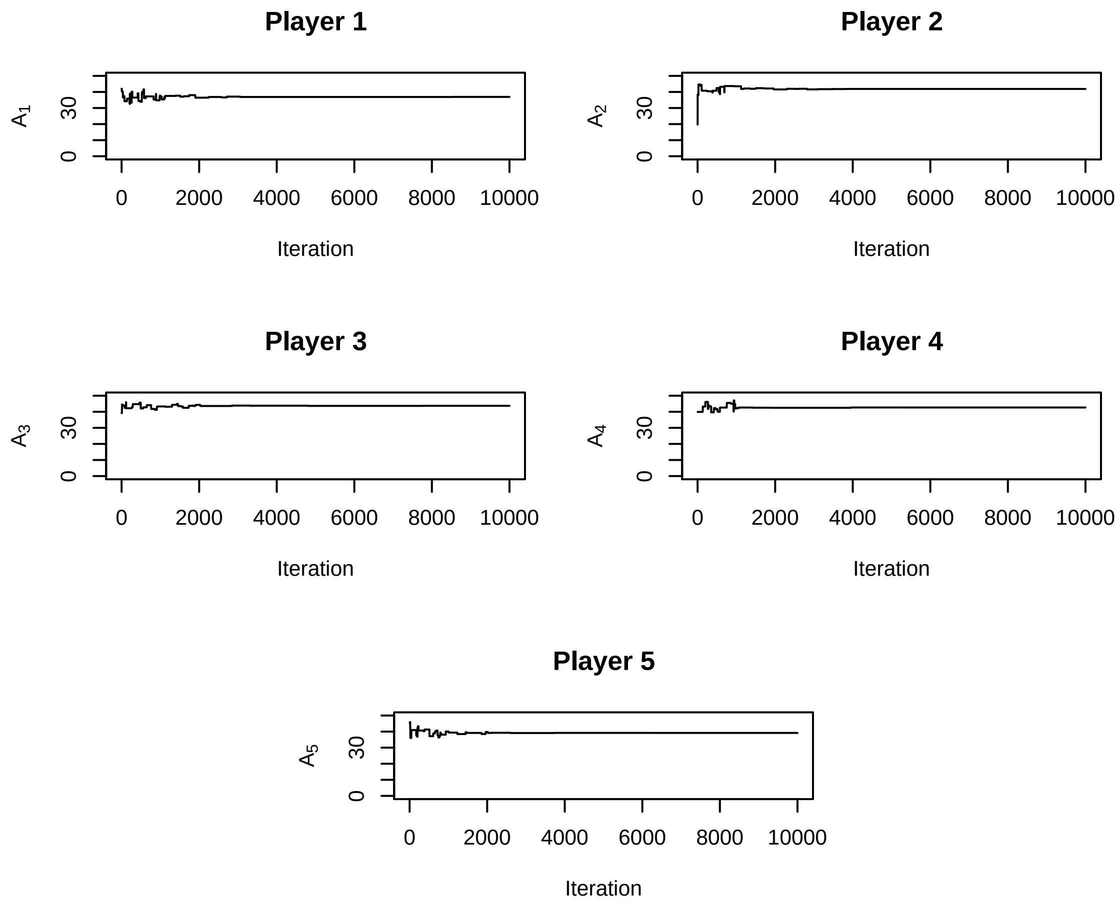
The paramters of the 5 firm specific cost functions are given in the following table:

firm i	c_i	β_i
1	10	1.2
2	8	1.1
3	6	1.0
4	4	0,9
5	2	0.8

The Nash equilibrium of this game has been numerically found [14] to be approximately

$$(q_1^* = 36.93, q_2^* = 41.82, q_3^* = 43.70, q_4^* = 42.66, q_5^* = 39.18).$$

The stochastic response dynamic resulted in the following output:



The simulation stopped in the state

$$(\widehat{q_1} = 36.92, \widehat{q_2} = 41.87, \widehat{q_3} = 43.71, \widehat{q_4} = 42.66, \widehat{q_5} = 39.18).$$

9 Capacity Constraints

The preceding examples illustrated the use of the stochastic response dynamic as a computational tool for a variety of Cournot oligopoly models. However all of these games were just as easily solvable by traditional "Caclulus" means. We now illustrate a problem that becomes "harder" for the Calclus approach but does not disturb the stochastic response dynamic.

A game in which the Calculus approach cannot proeed in the usual way occurs when the firms supplying a market have capacity constraints on the amount they can produce. Capacity constraints are an empirically relevant fact of industrial life, and thus are important to include in an oligopoly model. However capacity constraints cause a lack of differentiability in the utility functions to arise, thus violating the *necessity* assumption on the payoffs imposed by the Calculus approach (the necessity assumption was discussed in section 4).

Froeb et al. [7] have recently explored the problem of computing equilibrium in the presence of capacity constraints. They use a differentiated products Bertrand oligopoly. That is, an oligopoly model where firms are price setters rather than quantity setters, and the firms not homogenous.

To see the problem introduced by capacity constraints, let q_i be the amount supplied by each firm i, which is a function of both own price p_i and the prices of the other firms p_{-i}. If firm i faces no capacity constraints, then its profit function is

$$u_i(\, p_i, p_{-i}\,) = p_i q_i(p_i, p_{-i}) - c_i(q_i(p_i, p_{-i})).$$

The best response $br_i(p)$ of firm i to a profile of prices $p = (p_i, p_{-i})$ is chosen so as to satisfy its first order condition, i.e., firm i solves the equation

$$\frac{d}{dp_i} u_i(p_i, p_{-i}) = 0.$$

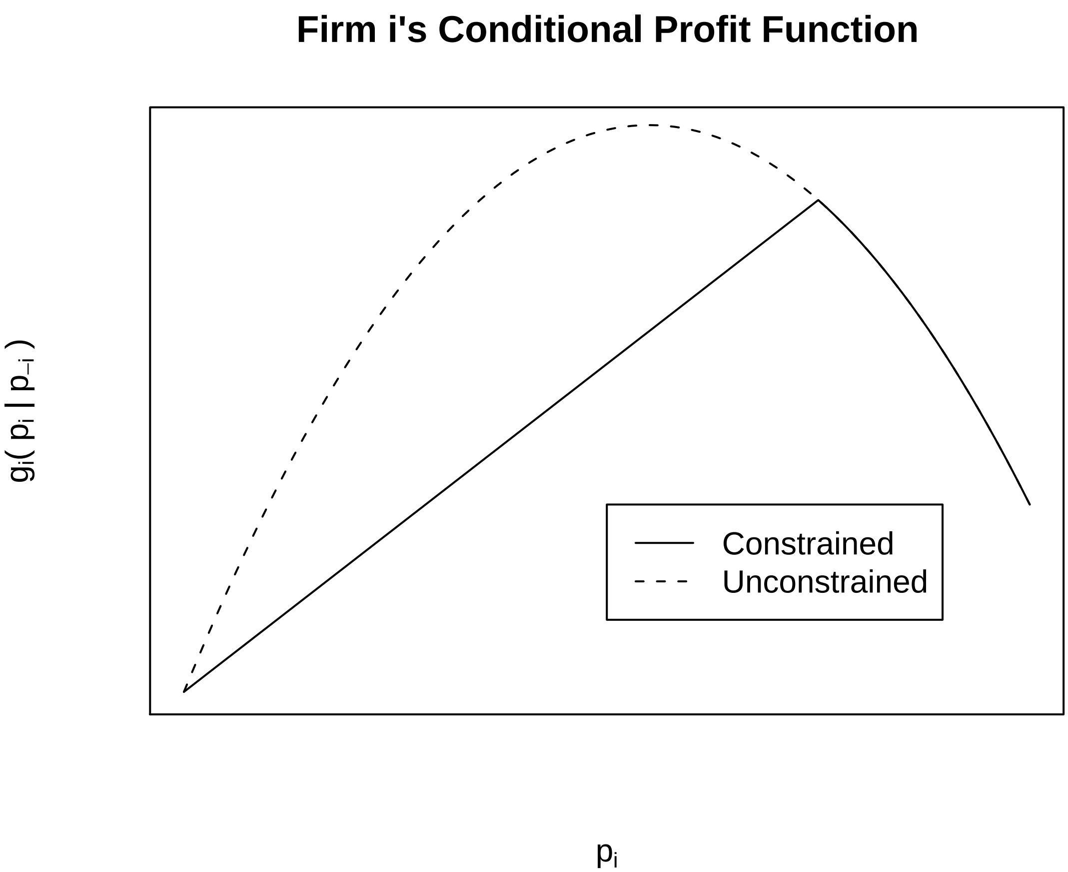
However if firm i faces a capacity constraint k_i on the amount it can produce, then the profit function becomes

$$u_i(\, p_i, p_{-i}\,) = p_i \min(q_i(p_i, p_{-i}), k_i) - c_i(\min(q_i(p_i, p_{-i}), k_i)).$$

Depending on the profile of prices p, the constrained best response $br_i(p)$ of firm i is the solution to one of two potential equations. If the unconstrained best response p_i' results in a demand q_i that does not exceed capacity k_i, the constrained best response is also p_i', and the capacity constraint has no effect. However if the unconstrained best response p_i' results in a demand q_i that exceeds capacity k_i, then the constrained best response p_i'' is set so as to solve the equation

$$q_i(p_i, p_{-i}) - k_i = 0.$$

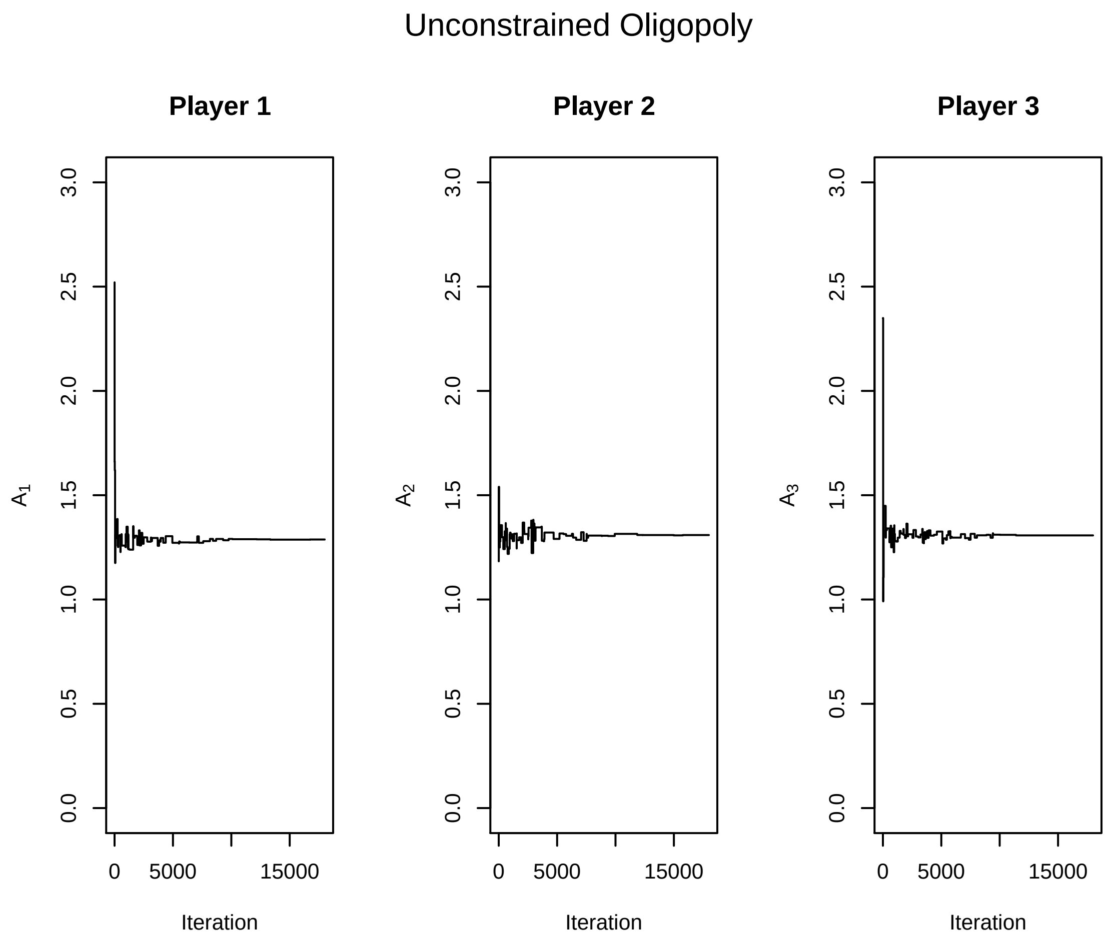
This situation is shown graphically below.

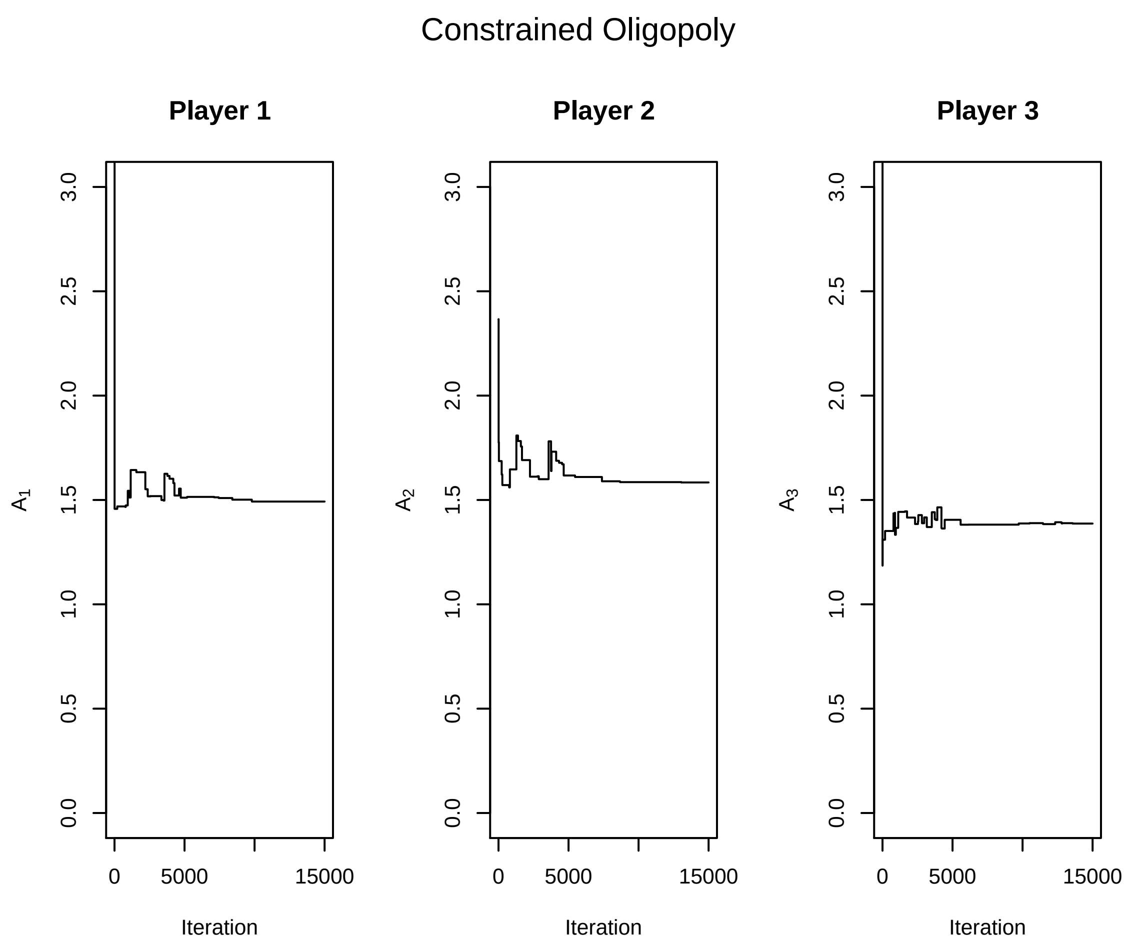


Thus each player may not be solving its first order condition in equilibrium, which forces one to modify the standard Calculus approach to computation.

However it is easy to imagine that the stochastic response dynamic would have an even easier time with the constrained problem than with the unconstrained problem. This follows from looking at the graph of the profit function for the capacity constrained producer. The capacity constraint introduces a kink into the profit function of each firm with respect to own action. Thus in equilibrium, any firm that is capacity constrained also has a profit function that is more "peaked" at the maximum than in the unconstrained case. In terms of the equilibrium distribution of the stochastic response dynamic P^T, which becomes increasingly centered over Nash equilibrium as $T \rightarrow 0$, the "mode" of P^T will become sharper via the capacity constraint. Thus we expect the stochastic response dynamic to get stuck in equilibrium more readily under constraints than without constraints.

To illustrate, the stochastic response dynamic was run on the Bertrand oligopoly model employed in Froeb et al. [7]. The stochastic response dynamic resulted in the following output, first for the unconstrained case, and then for the constrained case. As can be seen, the output for the constrained case shows the simulation getting "stuck" in a mode faster than for the unconstrained case.





10 Conclusion

In this paper we have shown a new approach to learning and computing Nash equilibrium in continuous games that abandons the methods and assumptions of the traditional Calculus driven approach. The stochastic response dynamic is unlike the Calculus based approach to finding Nash equilibrium in two important respects. First the stochastic response dynamic is decentralized, and thus corresponds to a formal model of learning in economic environemnts. Such a learning model provides a solution to equilibria selection in the presence of multiple equilibria in structural game theoretic models. Second, the stochastic response dynamic converges (probabilistically) to Nash over the general class of games $\mathcal{G}$ of interest without making any differentiability assumptions, assuming only that $\mathcal{G}$ satisfies the natural game theoretic condition we term Nash connectedness. We conjecture Nash connectedness holds true for $\mathcal{G}$, but only are able to prove for 2 player games in $\mathcal{G}$ at present. An important basis for understanding why the stochastic response dynamic would work in the manner described was the link we established between strategic games and conditionally specified probability models. Essentially this link to probability models motivates the stochastic response dynamic. After establishing the main convergence properties of the stochastic response dynamic, we proceeded to illustrate its computational behavior on a variety of oligopoly games, including those in which standard differentiability assumptions are not met and thus the Calculus approach encounters difficulty. The computational behavior conforms to out expectations. Future is work is underway towards establishing Nash connectedness for the entire class of games $\mathcal{G}$ by way of the Lefschetz fixed point theorem.

References

[1] S.P. Brooks and B.J.T. Morgan. Optimisation using simulated annealing. *The Statistician*, 44:241–257, 1995.

[2] Stephen P. Brooks. Markov chain Monte Carlo method and its application. *The Statistician*, 47:69–100, 1998.

[3] Augustine Cournot. *"Researches into the Mathematical Principles of the Theory of Wealth"*. London, Hafner, "1838".

[4] G Debreu. *Existence of Competitive Equilibrium*, volume 2, chapter 15. North-Holland, 1982.

[5] S. R. Foguel. *The Ergodic Theory of Markov Processes*. Van Nostrand Reinhold, New York, 1969.

[6] James W. Friedman and Claudio Mezzetti. Learning in games by random sampling. *Journal of Economic Theory*, 98(1):55–84, 2001.

[7] L. Froeb, S. Tschantz, and P. Crooke. Bertrand competition with capacity constraints: Mergers among parking lots. *Journal of Econometrics*, Forthcoming.

[8] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1996.

[9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[10] W. R. Gilks, S. Richardson, and D. J. (Ed) Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1998.

[11] J.C. Harsanyi. "games with randomly distributed payoffs: A new rationale for mixed-strategy equilibrium points. *Internaional Journal of Game Theory*, 2:1–23, 1973.

[12] Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to nash equilibrium. *American Economic Review*, 93(5):1830–1836, December 2003.

[13] K Judd. *Numerical Methods in Economics*. MIT Press, 1998.

[14] C. Kolstad and L. Mathiesen. Computing cournot-nash equilibrium. *Operations Research*, 1991.

[15] S. Martin. *Advanced Industrial Economics*. Blackwell, 1993.

[16] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[17] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

[18] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Havard University Press, Cambridge, Massachusetts, 1991.

[19] M.J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[20] David C. Parkes and Lyle Ungar. Learning and adaption in multiagent systems. In *Proc. AAAI'97 Multiagent Learning Workshop*, Providence, USA, 1997.

[21] M. Richard and T. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10:6–38, 1995.

[22] J. Tirole. *The Theory of Industrial Organization*. MIT Press, 1988.

[23] Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2003.

[24] Radu Zahraopol. *Invariant Probabilities of Markov-Feller Operators and their Supports*. Birkhauser Verlag, 2005.