ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

ALMOST-RATIONAL LEARNING OF NASH EQUILIBRIUM WITHOUT ABSOLUTE CONTINUITY

Thomas W.L. Norman

Number 602
April 2012

Manor Road Building, Oxford OX1 3UQ

Almost-Rational Learning of Nash Equilibrium without Absolute Continuity*

Thomas W. L. Norman

Magdalen College, Oxford

April 6, 2012

Abstract

If players learn to play an infinitely repeated game using Bayesian learning, it is known that their strategies eventually approximate Nash equilibria of the repeated game under an absolute-continuity assumption on their prior beliefs. We suppose here that Bayesian learners do not start with such a "grain of truth," but with arbitrarily low probability they revise beliefs that are performing badly. We show that this process converges in probability to a Nash equilibrium of the repeated game. *Journal of Economic Literature* Classification: C73; D83.

Key Words: Repeated games; Nash equilibrium; rational learning; Bayesian learning; absolute continuity.

*I thank Ehud Kalai and Alvaro Sandroni for useful discussions. Email thomas.norman@magd.ox.ac.uk.

1 Introduction

Kalai and Lehrer (1993) establish that rational players who know only their own payoffs in an infinitely repeated game, and who Bayesian update subjective beliefs about opponents' strategies, will eventually play a Nash equilibrium, provided that this equilibrium play is absolutely continuous with respect to their prior beliefs. Absent this "grain of truth" condition, whereby players must put positive probability on the eventual play from the outset, Bayesian learners are liable to spend eternity in disequilibrium, with their beliefs unfulfilled. Might a rational player who finds that his predictions are persistently wrong eventually come to question his predictions, and consider the possibility that his prior belief was erroneous?

The Bayesian answer is that such a "paradigm shift" (Weinstein 2011) is irrational, and should instead be anticipated in the player's prior. However, the idea that the player could, in the absence of knowledge of his opponents' payoffs, adopt a sufficiently agnostic prior to guarantee equilibrium convergence has been shown to be too demanding (Nachbar 1997, 2001, 2005, Foster and Young 2001). Moreover, Weinstein (2011) shows that, if a decision-maker's initial subjective probability of reaching a non-Bayesian paradigm shift is small, and if the dependence of his bets on "post-shift" events is bounded, the degree to which he is subject to an objective Dutch book is small.

This paper introduces the possibility of such paradigm shifts into the model of Bayesian learning in repeated games: with arbitrarily low probability, players revise their beliefs in a non-Bayesian fashion and adopt a random alternative. We show that this process converges in probability to a Nash equilibrium of the repeated game, irrespective of the players' prior beliefs. The proof of this result draws on the work of Sandroni (1998) and Foster and Young (2003). The latter exploit a similar random structure on belief revision to show that most of the time is spent in subgame-perfect equilibrium when players with limited memory learn by classical hypothesis testing. Here, we return much closer to Bayesian learning, and consider fully general strategies and beliefs rather than just those with finite memory. Sandroni's "almost absolute continuity," meanwhile, provides us with a crucial weakening of Kalai and Lehrer's (1993) absolute continuity.

2 Bayesian Learning with Paradigm Shifts

Now let us sketch the formal details of the learning model. There is a finite, n-person stage game G with action spaces X_i, $i = 1, \ldots, n$, and utility functions $u_i : X \to \mathbb{R}$,

$X = \prod_{i=1}^{n} X_i$. This stage game is infinitely repeated in discrete time $t \in \mathbb{N}$, with public observation of play, to give the repeated game G^∞.

A *history of play* is denoted ω, belonging to the set of all possible histories Ω; $\omega^t = (\omega_1^t, \ldots, \omega_n^t) \in X$ then denotes the actions taken in period t; $\overline{\omega}^t = (\omega^1, \omega^2, \ldots, \omega^t)$ the *initial history* of actions taken in periods 1 through t inclusive, belonging to the set of all possible such initial histories $\overline{\Omega}^t$, with $\overline{\Omega} = \bigcup_{t\in\mathbb{N}} \overline{\Omega}^t$; and $\Omega(\overline{\omega}^t) = \{\alpha \in \Omega \mid \overline{\alpha}^t = \overline{\omega}^t\}$ the set of all *continuations* of the initial history $\overline{\omega}^t$. Let $\overline{\Omega}(\overline{\omega}^t) = \{\overline{\alpha}^\tau \mid \tau \geq t,\ \overline{\alpha}^t = \overline{\omega}^t\}$ be the set of possible *continued initial histories* following $\overline{\omega}^t$. For every initial history $\overline{\omega}^t \in \overline{\Omega}^t$, a *cylinder with base on* $\overline{\omega}^t$ is the set $C(\overline{\omega}^t) = \{\omega \in \Omega \mid \omega = (\overline{\omega}^t, \ldots)\}$ of all histories such that the t initial elements coincide with $\overline{\omega}^t$. Let $\mathcal{I}_t$ be the σ-algebra on Ω whose elements are all finite unions of cylinders with base on $\overline{\Omega}^t$. We then have a filtration

$$\mathcal{I}_0 \subset \cdots \subset \mathcal{I}_t \subset \cdots \subset \mathcal{I},$$

where $\mathcal{I}_0$ is the trivial σ-algebra and $\mathcal{I}$ is the σ-algebra generated by the algebra of initial histories $\mathcal{I}^0 \equiv \bigcup_{t\geq 0} \mathcal{I}_t$.

A *behavior strategy* $a_i : \overline{\Omega} \to \Delta_i$ for player i is a mapping from the set of all possible initial histories into the simplex of his action mixtures Δ_i. Let $a_i(\overline{\omega}^t)(x)$ denote the probability that a_i prescribes for the action $x \in X_i$, after the initial history $\overline{\omega}^t \in \overline{\Omega}^t$, and let $a = (a_1, \ldots, a_n)$ be a typical strategy profile. Given a strategy profile a and an initial history $\overline{\omega}^t \in \overline{\Omega}^t$, the induced strategy profile $a_{\overline{\omega}^t}$ is defined by $a_{\overline{\omega}^t} = a(\overline{\alpha}^\tau | \overline{\omega}^t)$ for any $\overline{\alpha}^\tau \in \overline{\Omega}(\overline{\omega}^t)$.

Each player i has a *belief* $b^i = (b_1^i, \ldots, b_n^i)$ regarding the strategy profile, with $b_i^i = a_i$; let $b_{-i}^i = (b_1^i, \ldots, b_{i-1}^i, b_{i+1}^i, \ldots, b_n^i)$ denote i's beliefs about his opponents' strategies. Let μ_{b^i} be the conditional probability measure over histories induced by the belief b^i. It will be useful for our purposes to think of a player's belief as belonging to the space $\mathcal{B} \triangleq \prod_{k=1}^{\infty} \Delta^{|\mathcal{I}_k|}$, where $\Delta^{|\mathcal{I}_k|}$ is an $|\mathcal{I}_k|$-dimensional simplex; b^i induces a probability on each element of $\mathcal{I}_k$, $k = 1, \ldots, \infty$. Note that, since $\overline{\Omega}^t$ has finitely many elements, there are finitely many cylinders with base on $\overline{\Omega}^t$, and hence each $\mathcal{I}_t$ has finitely many elements.

Let $\rho_i \in (0, 1)$ be individual i's discount factor, and given an initial history $\overline{\omega}^{t-1}$ and a continuation $\omega \in \Omega(\overline{\omega}^{t-1})$, the future stream of utilities discounted to period t is given by

$$U_i^t(\omega) = (1 - \rho_i) \sum_{t'=t}^{\infty} \rho_i^{t'-t} u_i(\omega^{t'}).$$

Individual i's expected utility at time t over all continuations is

$$\mathrm{E}(U_i^t(\omega) \mid b^i, \overline{\omega}^{t-1}) = \frac{\int_{\Omega(\overline{\omega}^{t-1})} U_i^t(\omega) d\mu_{b^i}}{\int_{\Omega(\overline{\omega}^{t-1})} d\mu_{b^i}}.$$

Letting $U_i^t(a_i, b^i_{-i}) = \mathrm{E}(U_i^t(\omega) \mid b^i, \overline{\omega}^{t-1})$, a_i is then a *best response* if $U_i^t(a_i, b^i_{-i}) \geq U_i^t(a'_i, b^i_{-i})$, $\forall a'_i$. Given a small $\eta > 0$, a_i is an *η-best response* if

$$U_i^t(a_i, b^i_{-i}) \geq U_i^t(a'_i, b^i_{-i}) - \eta, \quad \forall a'_i.$$

We assume that, when he chooses his strategy, player i plays an η-best response, determined by a *perturbed best response function* $A_i^\eta(b^i_{-i})$ resulting from random payoff perturbations (McFadden 1981, Anderson, de Palma, and Thisse 1992), and hence we will have use for the subspace $\mathcal{B}_{-i} \subset \mathcal{B}$ of beliefs such that $b^i_i = A_i^\eta(b^i_{-i})$.

Whereas a best response is automatically optimal in all periods, an η-best response in period t may be far from optimal in the distant future.[1] There is thus an important distinction to be drawn between *uniform η-optimization*—requiring an η-best response in all possible continued initial histories—and the *ex ante η-optimization* we employ here, which requires an η-best response only when the strategy is adopted (at the beginning of the game or following a paradigm shift). As Nachbar (1997) notes, ex ante η-optimization may not be sufficient to guarantee Kalai–Lehrer (1993) convergence to approximate Nash equilibrium play, but it is sufficient for our purposes, and indeed Theorem 1 would not hold under uniform ε-optimization. We view the ex ante concept as quite consistent with the paradigm-shift model, and note moreover that η becomes arbitrarily small in our convergence result.

We employ Sandroni's (1998) notion of weak closeness of strategy profiles: a probability measure μ is *weakly ε-close* to $\tilde{\mu}$ at time t if

$$d(\mu, \tilde{\mu}) = \sum_{k=t}^{\infty} 2^{-(k-t+1)} \left(\sup_{A \in \mathcal{I}_k} |\mu(A) - \tilde{\mu}(A)| \right) \leq \varepsilon.$$

Given two strategy profiles a and $\tilde{a}$, *a plays weakly ε-like $\tilde{a}$* if μ_a is weakly ε-close to $\mu_{\tilde{a}}$. Intuitively, if two strategy profiles play weakly ε-like one another, then they induce two probability measures on histories that assign similar probabilities for all measurable events except perhaps those that may only be observed in the distant future. This is to be contrasted with the stronger notion of distance captured by the sup norm, $\|a - \tilde{a}\| \triangleq$

[1] See Lehrer and Sorin (1998) for a discussion of the resulting issues.

$\sup_{A\in\mathcal{I}} |\mu_a(A) - \mu_{\tilde{a}}(A)|$, closeness in which requires that similar probabilities be assigned for all measurable events. We assume each $\mathrm{E}(U_i^t(\omega) \mid b^i, \overline{\omega}^{t-1})$ to be continuous in b^i (with respect to d); two beliefs that are weakly close generate similar expected utilities, which is quite reasonable under discounting. We also assume each perturbed best response function $A_i^\eta(b^i_{-i})$ to be continuous in b^i_{-i}, essentially requiring smoothness in the underlying payoff perturbations.

Given two strategy profiles a and $\tilde{a}$, a is *absolutely continuous with respect to* $\tilde{a}$ (denoted $\mu_a \ll \mu_{\tilde{a}}$) if $\mu_{\tilde{a}}(A) > 0$ implies $\mu_a(A) > 0$ for every $A \in \mathcal{I}$. Given perturbed best responses a and beliefs $b^1, \ldots, b^n$, the strategy profile $k^i = (k^i_1, \ldots, k^i_{i-1}, a_i, k^i_{i+1}, \ldots, k^i_n)$ is an *ε-perturbation in player i's beliefs* at time t if

- $(a_i)_{\overline{\omega}^t}$ is an ε-best response to $(k^i_{-i})_{\overline{\omega}^t}$, $i = 1, \ldots, n$, and
- $k^i_{\overline{\omega}}$ plays weakly ε-like $b^i_{\overline{\omega}}$ for every $\overline{\omega} \in \overline{\Omega}(\overline{\omega}^t)$.

Note that this differs from Sandroni's (1998) Definition 4.8, requiring an ε-best response in period t only, rather than in every subgame; this is a consequence of ex ante ε-optimization. The strategy profile a is *ε-absolutely continuous with respect to b^i* if there exists an ε-perturbation of player i's beliefs, k^i, such that a is absolutely continuous with respect to k^i. a is *almost absolutely continuous with respect to b^i* if, for every $\varepsilon > 0$, a is ε-absolutely continuous with respect to b^i. Sandroni showed that almost absolute continuity of optimal strategies with respect to beliefs is necessary and sufficient for them eventually to play weakly like a Nash equilibrium.[2]

Each player's belief is Bayesian updated each period in light of observed play. However, given such an updating process, the resulting belief may come to be seen as an unlikely basis for observed play, and questioned; for example, the observed initial history of play may have zero probability under b^i. In response to this, we assume that, with some probability $\pi_i^{t,\tau} \in [0, \varepsilon]$, player i's belief changes in a non-Bayesian fashion within the next τ periods; in particular, he undergoes a *paradigm shift* whereby a new belief is chosen at random according to a probability measure $f_i^t(\cdot \mid \overline{\omega}^{t-1})$ defined on $\mathcal{B}_{-i}$, and a new perturbed best response is given by A_i^η. Moreover, f_i is assumed to be *flexible* in

[2]Recall that beliefs $(b^1, \ldots, b^n)$ and optimal strategies $a = (a_1, \ldots, a_n)$ *play eventually weakly ε-like a Nash equilibrium* if there exists a set $A \in \mathcal{I}$ such that

i. $\mu_a(A) = 1$.

ii. For every $\alpha \in A$, there exists a period $t(\alpha)$ such that for all $t \geq t(\alpha)$, $a_{\overline{\alpha}^t}$ and $b^i_{\overline{\alpha}^t}$, $i = 1, \ldots, n$, play weakly ε-like a Nash equilibrium.

If $(b^1, \ldots, b^n)$ and a play eventually weakly ε-like a Nash equilibrium for all $\varepsilon > 0$, they are said to *play eventually weakly like a Nash equilibrium.*

the sense that, for each $\nu > 0$, the f_i-measure of any ν-ball of beliefs in $\mathcal{B}_{-i}$ is bounded below by a strictly positive number $f_*(\nu) > 0$. The following result entitles us to make this assumption.

Lemma 1 *There exists a flexible probability measure f_i.*

Proof. Given $t \geq 0$, $\Delta^{|\mathcal{I}_t|}$ is compact in the topology induced by the sup norm. d metrizes the standard product topology on $\mathcal{B}_{-i}$, in which $\mathcal{B}_{-i}$ is then compact by Tychonoff's Theorem. Consider a cover $\mathcal{U}$ of $\mathcal{B}_{-i}$ composed of infinitely many open η-balls. By compactness of $\mathcal{B}_{-i}$, $\mathcal{U}$ has a finite subcover $\mathcal{V}$. Let f_i allocate measure 1 uniformly over the elements of $\mathcal{V}$. ■

Note that the use of weak closeness is important for this result; under the stronger sup norm, compactness of the belief space fails and the lemma breaks down.

Given beliefs $(b^1, \ldots, b^n)$ and perturbed best responses $a = (a_1, \ldots, a_n)$, we assume that there exists a $\overline{\tau} = \overline{\tau}(a, b)$ such that, for all i, t and τ, $\pi_i^{t,\tau} = 0$ if a is ε-absolutely continuous with respect to b^i at time t, and $\pi_i^{t,\overline{\tau}} \geq \underline{\pi} > 0$ otherwise. Informally, as ε vanishes, we can think of a player undergoing a paradigm shift with positive probability only if the observed initial history of play is impossible under his belief.

3 Equilibrium Convergence

Since $(\mathcal{B}, d)$ is compact, it is also separable, and hence we may say that random variables $\{Y_s\}$ on $(\mathcal{B}, d)$ *converge in probability* if and only if

$$\lim_{s \to \infty} \Pr(d(Y_s, Y) > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Theorem 1 *Beliefs $(b^1, \ldots, b^n)$ and perturbed best responses a converge in probability to a Nash equilibrium of G^∞.*

Proof. Given $\varepsilon > 0$, suppose that beliefs and optimal strategies have not weakly ε-converged to a Nash equilibrium at time t. Then, at worst, there is at least one player i such that a is not ε-absolutely continuous with respect to b^i. Any such player i then undergoes a paradigm shift by period $t + \overline{\tau}$ with probability at least $\underline{\pi}$, resulting in a new belief $\tilde{b}^i$, which plays weakly ν-like a fixed point of A^η (an η-Nash equilibrium) with probability at least $f_*(\nu)$. Any player not undergoing a paradigm shift during a $\overline{\tau}$-length timeframe does so with probability at least $1 - \varepsilon$. It follows that there exists a $\hat{\tau}$ such that

all beliefs play weakly ν-like an η-Nash equilibrium a^* by period $t+\hat{\tau}$ with probability at least $\lambda = (\underline{\pi} f_*(\nu))^n (1-\varepsilon)^{n(n-1)}$.

Consider the strategy profile $k^i = (k_1^i, \ldots, k_{i-1}^i, \tilde{a}_i, k_{i+1}^i, \ldots, k_n^i)$, where $\tilde{a}_i = A_i^\eta(\tilde{b}^i)$, $k_j^i = (1-\eta)\tilde{b}_j^i + \eta \tilde{a}_j$ and $\eta < \varepsilon/2$. We know that $\tilde{a}_i$ is an η-best response to $\tilde{b}^i$. We also know that a_i^* is an η-best response to a_{-i}^*; hence, given any $\hat{\eta} \in (\eta, \varepsilon]$ there exists ν sufficiently small that $\tilde{a}_i$ is an $\hat{\eta}$-best response to $\tilde{a}_{-i}$, by continuity of $A_i^\eta(\cdot)$ and $\mathrm{E}(U_i^t(\omega) \mid b^i, \overline{\omega}^{t-1})$ in b^i. Thus, we can choose ν sufficiently small that $\tilde{a}_i$ is an $\varepsilon/2$-best response to k_{-i}^i and $\tilde{a}$ is an ε-Nash equilibrium. Moreover, $(k^i)_{\overline{\omega}}$ plays (strongly and hence) weakly η-like $(\tilde{b}^i)_{\overline{\omega}}$ for every $\overline{\omega} \in \overline{\Omega}$. Hence, k^i is an $\varepsilon/2$-perturbation in player i's beliefs, with respect to which $\tilde{a}$ is absolutely continuous. Therefore, with probability at least λ, $\tilde{a}$ is $\varepsilon/2$-absolutely continuous with respect to all beliefs by period $t+\hat{\tau}$. In this event, the Blackwell and Dubins (1962) theorem implies that there exists a full $\mu_{\tilde{a}}$-measure set $A(\varepsilon) \in \mathcal{I}$ such that for every $\omega \in A(\varepsilon)$, there exists a period $\bar{t}$ such that

$$(k^i(\eta))_{\overline{\omega}^\tau} \text{ plays } \varepsilon/2\text{-like } \tilde{a}_{\overline{\omega}^\tau}, \qquad \text{for all } \tau \geq \bar{t}.$$

Thus, for every $\omega \in A(\varepsilon)$, there exists a period $\bar{t}$ such that

$$(\tilde{b}^i)_{\overline{\omega}^\tau} \text{ plays weakly } \varepsilon\text{-like } \tilde{a}_{\overline{\omega}^\tau}, \qquad \text{for all } \tau \geq \bar{t}.$$

Now consider the probability of the event $\mathcal{E}_s$ that optimal strategies are not ε-absolutely continuous with respect to beliefs in an arbitrary period s. In this event, there are at least k distinct times $t_1 < \cdots < t_k \leq s$ such that the following hold:

- $t_{j+1} - t_j \geq \hat{\tau}$ for $1 \leq j < k$,
- optimal strategies are not ε-absolutely continuous with respect to beliefs at time t_j for $1 \leq j < k$,
- optimal strategies do not become ε-absolutely continuous with respect to beliefs from t_1 to t_k.

From above, the probability of this event is at most $(1-\lambda)^{k-1}$. Moreover, $k \to \infty$ as $s \to \infty$. Hence, $\lim_{s\to\infty} \Pr(\mathcal{E}_s) = 0$, and the limiting probability that beliefs and η-optimal strategies do not play weakly ε-like an ε-Nash equilibrium in period s is also zero. ■

References

Anderson, S., A. de Palma, and J. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, Massachusetts.

Blackwell, D., and L. Dubins (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 38, 882–886.

Foster, D. P., and H. P. Young (2001): "On the Impossibility of Predicting the Behavior of Rational Agents," *Proceedings of the National Academy of Sciences of the USA*, 98(22), 12848–12853.

——— (2003): "Learning, Hypothesis Testing, and Nash Equilibrium," *Games and Economic Behavior*, 45, 73–96.

Kalai, E., and E. Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1045.

Lehrer, E., and S. Sorin (1998): "ε-Consistent Equilibrium in Repeated Games," *International Journal of Game Theory*, 27, 231–244.

McFadden, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski, and D. McFadden, pp. 287–324. MIT Press, Cambridge, MA.

Nachbar, J. H. (1997): "Prediction, Optimization, and Learning in Repeated Games," *Econometrica*, 65, 275–309.

——— (2001): "Bayesian Learning in Repeated Games of Incomplete Information," *Social Choice and Welfare*, 18(2), 303–326.

——— (2005): "Beliefs in Repeated Games," *Econometrica*, 73, 459–480.

Sandroni, A. (1998): "Necessary and Sufficient Conditions for Convergence to Nash Equilibrium: The Almost Absolute Continuity Hypothesis," *Games and Economic Behavior*, 22, 121–147.

Weinstein, J. (2011): "Provisional Probabilities and Paradigm Shifts," working paper.