Mach Learn (2007) 67:77–115
DOI 10.1007/s10994-006-0219-y

Online calibrated forecasts: Memory efficiency versus universality for learning in games

Shie Mannor · Jeff S. Shamma · Gürdal Arslan

Received: 29 September 2005 / Revised: 24 July 2006 / Accepted: 13 August 2006 / Published online: 27 September 2006

Abstract We provide a simple learning process that enables an agent to forecast a sequence of outcomes. Our forecasting scheme, termed tracking forecast, is based on tracking the past observations while emphasizing recent outcomes. As opposed to other forecasting schemes, we sacrifice universality in favor of a significantly reduced memory requirements. We show that if the sequence of outcomes has certain properties—it has some internal (hidden) state that does not change too rapidly—then the tracking forecast is weakly calibrated so that the forecast appears to be correct most of the time. For binary outcomes, this result holds without any internal state assumptions. We consider learning in a repeated strategic game where each player attempts to compute some forecast of the opponent actions and play a best response to it. We show that if one of the players uses a tracking forecast, while the other player uses a standard learning algorithm (such as exponential regret matching or smooth fictitious play), then the player using the tracking forecast obtains the best response to the *actual play* of the other players. We further show that if both players use tracking forecast, then under certain conditions on the game matrix, convergence to a Nash

Editors: Amy Greenwald and Michael Littman

S. Mannor (✉)
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Quèbec, Canada H3A-2A7
e-mail: shie.mannor@mcgill.ca

J. S. Shamma
Department of Mechanical and Aerospace Engineering, University of California - Los Angeles, 37-146 Engineering IV, UCLA, Los Angeles, CA 90095-1597
e-mail: shamma@ucla.edu

G. Arslan
Department of Electrical Engineering, University of Hawaii at Manoa, 440 Holmes Hall, 2540 Dole Street, Honolulu, HI 96822
e-mail: gurdal@hawaii.edu

equilibrium is possible with positive probability for a larger class of games than the class of games for which smooth fictitious play converges to a Nash equilibrium.

Keywords Learning in games · Forecasting · Calibration · Fictitious play · Prediction of universal sequences · Stochastic approximation · The ODE method

1 Introduction

In supervised learning, we typically make predictions about future outcomes of a certain sequence based on past observations from a sequence of outcomes. These predictions are "guesses" of the next value that will be observed. By contrast, a *forecast* is a *probability measure* on the next observation. As an example, consider prediction of increase or decrease in some financial market index such as a bit indicating if NASDAQ's QQQQ will be up or down. A prediction will be a single bit ("up" or "down"), while a forecast will be a probabilistic estimate of the form "QQQQ will be up with probability 70%". Similarly to regret minimization (Foster & Vohra, 1999), online learning (Borodin & El-Yaniv, 1998), and prediction with expert advice (Auer et al., 2002; Vovk, 1998), the objective of a forecasting algorithm is to provide a "consistent" estimate in hindsight. Roughly explained, this translates to requiring that the empirical frequencies of QQQQ going up, when the forecaster predicted QQQQ would go up with probability p, is approximately p. A forecasting scheme which is consistent in hindsight is called "calibrated" (Foster & Vohra, 1997). Having a calibrated forecast is beneficial in several ways. First, it allows the agent to choose the best response to the predicted outcome. Second, the agent may consider other risk measures which might be more valuable than greedily choosing the best action leading to highest reward. Third, calibrated forecasting rules enable multiple agents to converge to a reasonable joint play in a game situation, as explained below.

A natural approach to learning or adaptation in repeated matrix games is to have each player compute some sort of forecast of opponent actions and play a best response to this forecast. Accordingly, the limiting behavior of player actions strongly depends on the specific method for forecasting. For example, in fictitious play, as well as smooth fictitious play, forecasts are simply the empirical frequencies of opponents' actions. In some special classes of games, player strategies converge to a Nash equilibrium, but in general, the limiting behavior need not exhibit convergence (e.g., (Fudenberg & Levine, 1998)). Placing more stringent requirements on the forecasts can result in stronger convergence properties for general games. In particular, if players use calibrated forecasts, then player strategies asymptotically converge to the set of correlated equilibria (Foster & Vohra, 1997). When players use calibrated forecasts of *joint* actions, then player strategies converge to the convex hull of Nash equilibria (Kakade & Foster, 2004). See (Sandroni, Smorodinsky, & Vohra, 2003) and references therein for further discussion on calibrated forecasting as well as its generalizations.

A major drawback of calibrated forecasts is the computational burden. In particular, existing methods of computing calibrated forecasts require a discretized grid of points extracted from a probability simplex of appropriate dimension for approximate calibration. This leads to memory requirements of $O(1/\varepsilon^n)$, where ε is the required approximation level of the forecasting scheme and n is the size of the outcome alphabet. If one is interested in obtaining exact calibration rather than approximate,

the grid must be gradually (and slowly) refined. Moreover, with the exception of the elegant forecasting rule for binary sequences of Fudenberg and Levine (1999), the computational burden for every step of existing forecasting algorithms is significant. As a result, current calibrated forecasting algorithms cannot be considered for operation on-line. Another drawback of calibrated forecasts is the lack of convergence rates. We do not address convergence rates in this work.

Motivated by the importance of calibration in prediction problems and in learning in games, we explore the possibility of calibration without discretization. We introduce a "tracking forecast" that sacrifices universality in favor of a significantly reduced computational burden. Specifically, the tracking forecast has the same computational requirement as computing empirical frequencies. We show that the tracking forecast is calibrated for special classes of sequences, and we discuss some consequences of tracking forecasts in repeated matrix games.

1.1 Outline of the paper

The setup of the prediction problem and some relevant literature review is provided in Section 2. We recall the formal definition of calibrated forecasts and a relaxation of it called weakly calibrated forecasts (following (Kakade & Foster, 2004)). This sense of calibration is more natural in the context of our analysis as it involves smooth test functions instead of indicators.

This paper has two main contributions. We first discuss calibrated forecasting in isolation, while assuming that an agent tries to forecast a given sequence. We then use the developed results to consider learning in games where the multiple players use tracking forecasts (or other algorithms) to predict each other's moves.

In Section 3 we present and analyze the tracking forecast algorithm. This algorithm has the same computational burden as computing empirical frequencies, and essentially forecasts the outcomes to have the same distribution as a weighted mean of recent observations. We show that if the outcome sequence is generated by a (hidden) state process that does not change too rapidly, then the tracking forecast is weakly calibrated. The case of binary outcomes receives a special treatment as it turns out that for this case, no assumptions on the sequence are needed. Finally we outline a simple method to "produce" a calibrated scheme (as opposed to weakly calibrated) from a weakly calibrated tracking scheme. This can be done by adding a small random perturbation to the forecast.

In Section 4 we consider learning in repeated games. We show that if one of the players uses a best response strategy against the forecasted action of the other, and if the other agent uses some "slow algorithm" like regret matching, gradient play, or smooth fictitious play, the player using tracking forecast will play a best response against the actual moves of the second player. We further consider the case of self-play, where both players use a best response to a combination of tracking forecasts and empirical frequencies. We show that in this case, convergence to a Nash equilibrium is enabled in a larger class of games than standard smooth fictitious play. Numerical simulations that illustrate convergence and divergence issues are presented in Section 5.

The analysis of the results presented in this paper relies on the stochastic approximation algorithm. The main idea is to study the convergence of a discrete time stochastic iteration using the stability of an associated ODE (ordinary differential equation). This

approach is not new to machine learning—it was used to prove the convergence of reinforcement learning algorithms (Tsitsiklis, 1994) as well as to analyze certain learning in games framework (Benaim, Hofbauer, & Sorin, 2003; Fudenberg & Levine, 1998). However, since it is not a mainstream technique in machine learning, we provide some discussion and required results in Appendix A.

Notation:

We will use the following notations throughout the paper.

– $a(k) = o(b(k))$ denotes that

$$\lim_{k\to\infty} a(k)/b(k) = 0,$$

for real sequences $a(k), b(k) > 0$, $k = 0, 1, 2, \ldots$.

– $|x|$ denotes the 2-norm in $\mathbb{R}^n$:

$$|x| = \sqrt{\left(\sum_i x_i^2\right)}.$$

– For a square matrix, $\|M\|$ denotes the matrix norm:

$$\|M\| = \max_{x\in\mathbb{R}^n} \frac{|Mx|}{|x|}.$$

– Boldface $\mathbf{1}$ denotes the vector

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n.$$

– Δ denotes the $n-1$ dimensional simplex in $\mathbb{R}^n$:

$$\Delta = \{s \in \mathbb{R}^n : s \geq 0, \text{ componentwise, and } \mathbf{1}^{\mathrm{T}} s = 1\}.$$

– vert$[\Delta]$ denotes the set of vertices of the simplex:

$$\text{vert}[\Delta] = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \cdots\ , \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

– $\Pi_\Delta : \mathbb{R}^n \to \Delta$ denotes the Euclidean projection to the simplex Δ, i.e.,

$$\Pi_\Delta[x] = \arg\min_{s\in\Delta} |x - s|.$$

– $\texttt{rand}[s] \in \text{vert}[\Delta]$ denotes a random realization according to the probability distribution $s \in \Delta$. Let v^i denote the ith vertex. Then the probability that $\texttt{rand}[s] = v^i$ is given by the ith component s_i.
– $\mathcal{H} : \Delta \to \mathbb{R}$ denotes the entropy function

$$\mathcal{H}(s) = -s^{\mathrm{T}} \log(s).$$

– $\sigma : \mathbb{R}^n \to \Delta$ denotes the "logit" or "soft-max" function

$$(\sigma(x))_i = \frac{e^{x_i}}{e^{x_1} + \cdots + e^{x_n}}.$$

For notational simplicity, we will not write explicitly the underlying dimension in $\mathbb{R}^n$ of $\mathbf{1}$, Δ, or $\text{vert}[\Delta]$.

2 Calibrated forecasts

In this section we review the concept of calibration. We start from the basic setup in Section 2.1. We then recall the classical definition of (strong) calibration in Section 2.2. Finally we present a relaxed version of calibration, termed weak calibration, in Section 2.3.

2.1 Online prediction set-up

At every stage, $k = 0, 1, 2, \ldots$, there is an outcome, $x(k)$, that belongs to a finite set, $\mathcal{X}$, with n elements. We will associate $\mathcal{X}$ with the set of vertices of the simplex, so that $x(k) \in \text{vert}[\Delta]$.

A forecaster observes outcomes sequentially, and at stage k, makes a forecast, $f(k) \in \Delta$, of the current outcome based on previously observed outcomes, $\{x(0), x(1), \ldots, x(k-1)\}$. Note that the forecast, $f(k)$, may belong to the entire *simplex*, Δ, i.e., not just a vertex. Accordingly, we interpret the ith element of $f(k)$ as the forecasted *probability* that the ith element of $\mathcal{X}$ will occur at stage k. In general, we allow for the possibility of *randomized* forecasts, where $f(k)$ is a non-deterministic function of the observed outcomes.

2.2 Calibration

We now define criteria under which a forecasting scheme is considered to be "calibrated". The discussion follows that of Kakade and Foster (2004).

For any $p \in \Delta$ and $\delta > 0$, define the indicator function

$$I_{p,\delta} : \Delta \to \{0, 1\}$$

as

$$I_{p,\delta}(f) = \begin{cases} 1, & |f - p| \leq \delta; \\ 0, & \text{otherwise.} \end{cases}$$

This function indicates whether a forecast, f, is within a specified tolerance of a specified point in the simplex.

Now, define the calibration error of a sequence $\mathbf{x}$ with respect to an indicator function, $I_{p,\delta}$, as

$$e_{p,\delta}(K, \mathbf{x}) = \frac{1}{K+1} \sum_{k=0}^{K} I_{p,\delta}(f(k))(x(k) - f(k)). \tag{1}$$

The calibration error with respect to $I_{p,\delta}$ compares the predicted frequency with the actual realized frequency when the prediction is δ close to p.

Definition 2.1. A forecasting scheme is ε**-calibrated** if for all outcome sequences,

$$\mathbf{x} = \{x(0), x(1), x(2), \ldots\},$$

and all indicator functions, $I_{p,\delta}$, the calibration error satisfies

$$\limsup_{K \to \infty} |e_{p,\delta}(K, \mathbf{x})| \leq \varepsilon \tag{2}$$

almost surely.

The statement "almost surely" in the definition refers to the set of realizations of randomization during forecasting. Note that a probabilistic structure has *not* been imposed on the space of outcome sequences. A sequence is called *calibrated* if it is ε-calibrated for every $\varepsilon > 0$. Prior work (Dawid, 1985; Oakes, 1985) has shown that there does not exist a deterministic forecasting scheme that satisfies the calibration criterion for *all* outcome sequences, and so randomized forecasting is necessary.

The standard intuition behind the calibration criterion is as follows. Define

$$N(K, p, \delta, \mathbf{x}) = \{0 \leq k \leq K : I_{p,\delta}(f(k)) = 1\},$$

and let $n(K, p, \delta, \mathbf{x})$ denote the number of elements of $N(K, p, \delta, \mathbf{x})$. In words, $n(K, p, \delta, \mathbf{x})$ denotes the number of times the forecast, $f(k)$, approximately equaled the specified value, p, between time 0 to time K, and the set $N(K, p, \delta, \mathbf{x})$ denotes the set of stages where this occurred. The calibration error can be rewritten as

$$e_{p,\delta}(K, \mathbf{x}) \approx \frac{n(K, p, \delta, \mathbf{x})}{K+1} \left(\left(\frac{1}{n(K, p, \delta, \mathbf{x})} \sum_{k \in N(K,p,\delta,\mathbf{x})} x(k) \right) - p \right).$$

(The $\approx$ sign is due to the fact that the forecast f may be slightly different than p on $N(K, p, \delta, \mathbf{x})$.) We see that there are two ways for the calibration error to vanish. First, the forecasted value of p may be rarely used in that

$$\limsup_{K\to\infty} \frac{n(K, p, \delta, \mathbf{x})}{K+1} = 0.$$

If this is not the case, then we require that for large K,

$$\frac{1}{n(K, p, \delta, \mathbf{x})} \sum_{k\in N(K,p,\delta,\mathbf{x})} x(k) \approx p,$$

implying that the empirical frequency of the outcomes over the stages where the forecast was (approximately) p is consistent with the forecast of p.

2.3 Weak calibration

Following (Kakade & Foster, 2004), we now state a relaxed version of calibration called "weak"[1] calibration. Let $\mathcal{W}$ denote the set of Lipschitz continuous functions $w : \Delta \to \mathbb{R}^+$.

Now define the calibration error with respect to a test function, $w \in \mathcal{W}$, as

$$e_w(K, \mathbf{x}) = \frac{1}{K+1} \sum_{k=0}^{K} w(f(k))(x(k) - f(k)). \tag{3}$$

This is the same form as the previously defined calibration error, but with the indicator function, $I_{p,\delta}(\cdot)$, now replaced by the test function, $w(\cdot)$. Note that indicator functions are excluded from being test functions because they fail the Lipschitz continuity requirement. It is convenient to think of the test functions as "bump" functions that are smoothed versions of indicator functions.

Definition 2.2. A forecasting scheme is **weakly calibrated** if for all outcome sequences, $\mathbf{x} = \{x(0), x(1), x(2), \ldots\}$, and all test functions, $w \in \mathcal{W}$, the calibration error satisfies

$$\limsup_{K\to\infty} |e_w(K, \mathbf{x})| = 0, \tag{4}$$

almost surely.

As before, randomized forecasts are allowed. However, unlike the case of strong calibration, it is possible to derive a *deterministic* forecasting scheme (Kakade & Foster, 2004) that is weakly calibrated for all outcome sequences, and so randomization

[1] The term "weak" is based on the notion of weak convergence of measures, cf. (Kakade & Foster, 2004) for a discussion.

is no longer necessary. Reference (Kakade & Foster, 2004) goes on to show how to use randomization in conjunction with a weakly calibrated forecast to achieve strong calibration. (See the forthcoming Section 3.6.) We will suppress the **x** henceforth to reduce notational clutter.

3 Calibration with bounded memory

In this section we present and analyze our forecasting algorithm. We start by discussing the complexity of existing calibration schemes in Section 3.1. We then focus the attention on a restricted class of sources for which weak calibration will be possible using a "simple" scheme in Section 3.2. We present the forecasting algorithm in Section 3.3, and proofs that it is calibrated for certain sources in Section 3.4. We provide a three letter example where tracking forecast is not weakly calibrated in Section 3.5. We finally show a simple procedure that achieves strong calibration in Section 3.6.

3.1 The complexity of calibration

Randomized forecasting schemes that achieve calibration are presented in Foster and Vohra (1998), Fudenberg and Levine (1999) and Hart and Mas-Colell (2001), and a deterministic forecasting scheme that achieves weak calibration is presented in Kakade and Foster (2004). Generalizations of calibration are discussed in Sandroni, Smorodinsky, and Vohra (2003) and references therein.

In general, all existing algorithms may be written in a state-space form

$$z(k+1) = G(z(k), x(k), k),$$
$$f(k) = H(z(k), k),$$

for suitably defined (possibly randomized) functions, $G(\cdot)$ and $H(\cdot)$, where z is some "state space" variable representing the memory needed from one period to the other. The dimension of the state-space variable, $z(k)$, constitutes a minimum memory requirement, and hence gives an indication of the complexity, to execute these algorithms.

The algorithms presented in the above works are "universal" in that the calibration criterion is satisfied for *all* outcome sequences. This universality apparently has a significant cost in terms of memory requirements. Namely, achieving calibration for any fixed set, $\mathcal{X}$, requires an ever increasing memory. Some of the above works present versions that achieve ε-calibration. In all of these works, the memory requirements of $z(k)$ typically are associated with a node of a discretization of the simplex where the amount of memory required is $O(1/\varepsilon^n)$, where n is the number of possible outcomes. Satisfying the calibration criterion (rather than ε calibration) is achieved through a slow progressive refinement of this discretization. Accordingly, the dimension of $z(k)$ increases without bound. We will not review the specifics of these algorithms here, but it is worth emphasizing that the tracking forecast presented below requires memory that is linear in n, and independent of ε.

3.2 Classes of outcome sequences

Our objective is to explore a trade-off between universality and complexity. In particular, we will consider calibration of *special classes* of outcome sequences. This will be achieved with a forecasting scheme, called *tracking forecast*, that has the same computational and memory requirements as computing running averages or empirical frequencies. Hence, these forecasts are easily computed online. The price of the complexity reduction is that the forecasting scheme is no longer universal.

Let $\mathcal{O}$ denote a class of outcome sequences, i.e., a subset of the space of $\mathcal{X}$-valued infinite sequences.

Definition 3.1. A forecasting scheme is **ε-calibrated over the set $\mathcal{O}$** if the calibration criterion (2) is satisfied for the set of outcome sequences

$$\{x(0), x(1), x(2), \ldots\} \in \mathcal{O}.$$

Similarly, a forecasting scheme is **weakly calibrated over the set $\mathcal{O}$** if the weak calibration condition (4) is satisfied for all outcome sequences belonging to $\mathcal{O}$.

The following sequence classes will be of particular interest.

Bounded rate sequences: Any sequence generated by the following recursion:

$$\begin{aligned} X(k+1) &= X(k) + a(k)(F(X(k), f(k)) + M(k)) \\ p(k) &= h(X(k), f(k)) \\ x(k) &= \texttt{rand}[p(k)], \end{aligned} \tag{5}$$

where

- $F: \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^d$ is Lipschitz.
- $a(k) = 1/(k+1)$.
- $h: \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz.
- For any fixed X, the function $f \mapsto h(X, f)$ is continuously differentiable, and

$$\sup_{X \in \mathbb{R}^d, f \in \Delta} \|\nabla_f h(X, f)\| < \gamma$$

  with $\gamma < 1$.
- $M(k) \in \mathbb{R}^d$ is a uniformly bounded random sequence with

$$\mathbb{E}\,[M(k) | X(k), e_w(k), f(k)] = 0$$

  for any calibration error function e_w.
- $\sup_k |X(k)| < \infty$ for all Δ-valued sequences, $\{f(0), f(1), f(2), \ldots\}$.

Bounded rate sequences have their own "internal" state space dynamics (the $X(k)$). These dynamics can depend on random fluctuations (via $M(k)$), on the previous state,

and on the forecasting itself (F can take $f(k)$ as an argument). The probability $p(k)$ that dictates the selection of the sequence $x(k)$ depends on that state space (and also on the forecast f, but in a weak way). The crucial requirement is that $X(k)$ cannot change too fast. As a result, $p(k)$ cannot change too rapidly as well. It is worth noting that when we assume that the sequence has bounded rate, we will never assume that the forecaster has access to the specifications of the sequence. In particular, F and h are assumed unknown.

Many algorithms for playing a repeated game generate bounded rate sequences. We mention several such algorithms in Section 4.1.2.

Relatively bounded rate sequences: Any sequence generated by the same recursion above and under the same assumptions, *except* that $a(k) = 1/(k+1)^\eta$ for some $1/2 < \eta < 1$.

In a relatively bounded sequence, the state variable can move a bit faster. Still, the rate of change to the state variable $X(k)$ and consequently to the probability governing the sequence $p(k)$ is slow.

Binary sequences: Any sequence over a binary outcome space, i.e., $n = 2$.

For binary sequences, we do not require any particular structure.

3.3 Tracking forecasts

The *tracking forecast* (Foster, 2005) is defined by

$$f(k+1) = f(k) + \left(\frac{1}{k+1}\right)^\rho (x(k) - f(k)), \tag{6}$$

where

$$0 < \rho < 1.$$

For $\rho = 1$, this becomes the same as the online computation of a running average, or *empirical frequencies*, i.e.,

$$q(k+1) = q(k) + \left(\frac{1}{k+1}\right)(x(k) - q(k)), \tag{7}$$

or, equivalently,

$$q(K) = \frac{1}{K+1}\sum_{k=0}^{K} x(k).$$

In terms of the previous discussion on calibration complexity, we see that the memory requirement of the tracking forecast is fixed to be the number of elements, n, of the outcome space. No discretization of the simplex is required. Our main result states that the tracking forecast is in fact calibrated over the above sequences.

Theorem 3.1. *The tracking forecast is weakly calibrated over the following classes of sequences:*

1. *Bounded rate sequences for any* $1/2 < \rho < 1$*;*
2. *Relatively bounded rate sequences for any* $1/2 < \rho < \eta < 1$*;*
3. *Binary sequences for any* $0 < \rho < 1$.

As a consequence of the proof of Theorem 3.1 in the case of bounded rate and relatively bounded rate sequences, we will see that the tracking forecasts actually satisfy a much more stringent condition than weak calibration. Namely, if the outcome sequence is generated according to $x(k) = \texttt{rand}[p(k)]$, then

$$\lim_{k\to\infty} (f(k) - p(k)) = 0, \qquad \text{almost surely.} \tag{8}$$

In other words, the tracking forecast converges to the actual stage-by-stage probability distribution that generates the outcome sequence. This consequence reflects an additional benefit from sacrificing universality, and it will have an important implication for learning in games (see Theorem 4.1).

3.4 Proofs for Theorem 3.1

The proof of each of the statements in Theorem 3.1 first casts the system as a stochastic approximation algorithm. The general iteration of the stochastic approximation algorithm is:

$$Y(k+1) = Y(k) + a(k)\,(F(Y(k)) + M(k))\,,$$

where $a(k)$ is a decreasing learning rate, F is some Lipschitz function and $M(k)$ is typically a random noise. Under certain conditions, $Y(k)$ converges to y^*, where y^* is an equilibrium point[2] of the ODE $\dot{y} = F(y)$. Appendix A contains a more precise statement of the results needed here. After casting the relevant system as a stochastic approximation algorithm, each proof contains specialized analysis that is usually concerned with proving that the ODE is stable.

3.4.1 Bounded rate sequences

We will write the combined equations for a bounded rate sequence (5), tracking forecast (6), and calibration error (3), in such a way to apply the ODE method of stochastic approximation presented in Appendix A. In particular, the form of these equations will satisfy the "two time scale stochastic approximation" setup of Proposition A.2. In two time scale stochastic approximation, there are two iterations on different time scales. The analysis of the fast iteration assumes that the variables which are modified by the slow iteration are fixed, while the slow iteration assumes that the variables modified by the fast iteration reach their equilibrium points.

[2] An ODE $\dot{y} = F(y)$ has an equilibrium y^* if $F(y^*) = 0$, so that the constant function $y(t) \equiv y^*$ is a solution of the ODE.

Step 1: Casting as a stochastic approximation problem. By writing the calibration error (3) in a recursive form, the overall discrete time iterations are

$$\begin{aligned} X(k+1) &= X(k) + \left(\frac{1}{k+1}\right)(F(X(k), f(k)) + M(k)) \\ e_w(k+1) &= e_w(k) + \left(\frac{1}{k+1}\right)(w(f(k))(x(k) - f(k)) - e_w(k)) \\ f(k+1) &= f(k) + \left(\frac{1}{k+1}\right)^{\rho}(x(k) - f(k)) \\ x(k) &= \texttt{rand}[h(X(k), f(k)]. \end{aligned} \tag{9}$$

Since

$$\mathbb{E}[x(k)] = h(X(k), f(k)),$$

we can rewrite these equations as

$$\begin{aligned} X(k+1) &= X(k) + \left(\frac{1}{k+1}\right)(F(X(k), f(k)) + M(k)) \\ e_w(k+1) &= e_w(k) + \left(\frac{1}{k+1}\right)(w(f(k))(h(X(k), f(k)) - f(k)) - e_w(k) + \tilde{M}(k)) \\ f(k+1) &= f(k) + \left(\frac{1}{k+1}\right)^{\rho}(h(X(k), f(k)) - f(k) + N(k)). \end{aligned}$$

where

$$\tilde{M}(k) = w(f(k))(x(k) - h(X(k), f(k)))$$

and

$$N(k) = x(k) - h(X(k), f(k)).$$

These satisfy

$$\mathbb{E}\left[\tilde{M}(k) \mid X(k), e_w(k), f(k)\right] = 0$$

and

$$\mathbb{E}[N(k) \mid X(k), e_w(k), f(k)] = 0.$$

Also, by assumption,

$$\mathbb{E}[M(k) \mid X(k), e_w(k), f(k)] = 0,$$

and so the resulting equations fall under the framework of Proposition A.2.

Step 2: Analysis of the fast time scale. After showing that the two time scale framework holds for this problem, we set out to analyze the possible solutions of the ODE. We start with the "fast" iteration that considers the tracking forecast.

$$\dot{f}(t) = h(\bar{x}, f(t)) - f(t), \tag{10}$$

where $\bar{x}$ is fixed. We will show that for any $\bar{x}$, the differential Eq. (10) has a unique globally asymptotically stable equilibrium[3] $f^* = \phi(\bar{x})$ for some Lipschitz continuous function $\phi(\cdot)$.

By assumption, for any $\bar{x}$,

$$\|\nabla_f h(\bar{x}, f)\| < \gamma < 1,$$

which implies that $f \mapsto h(\bar{x}, f)$ is a contraction (e.g., (Khalil, 2001, Lemma 3.1)). That is, for any f_1 and f_2,

$$|h(\bar{x}, y_1) - h(\bar{x}, y_2)| \leq \gamma |y_1 - y_2|,$$

with $\gamma < 1$. According to the contraction mapping theorem (e.g., (Bertsekas & Tsitsiklis, 1989, Section 3.1)), for any $\bar{x}$, the equation

$$f = h(\bar{x}, f)$$

has a unique solution, f^*, which we can write as

$$f^* = \phi(\bar{x}).$$

The Lipschitz assumption on h assures that ϕ is also Lipschitz continuous. To see this, consider

$$f_1 = h(x_1, f_1) = \phi(x_1)$$
$$f_2 = h(x_2, f_2) = \phi(x_2).$$

Let L_h be the Lipschitz constant of $x \mapsto h(x, f)$. We can write

$$h(x_2, f_2) = h(x_2, f_1) + h(x_2, f_2) - h(x_2, f_1),$$

and so

$$\begin{aligned} |f_2 - f_1| &= |h(x_2, f_1) - h(x_1, f_1) + h(x_2, f_2) - h(x_2, f_1)| \\ &\leq |h(x_2, f_1) - h(x_1, f_1)| + |h(x_2, f_2) - h(x_2, f_1)| \\ &\leq L_h |x_2 - x_1| + \gamma |f_2 - f_1|. \end{aligned}$$

[3] An equilibrium y^* of the ODE $\dot{y} = F(y)$ is called a globally asymptotically stable equilibrium, if for any initial condition $y(0)$ the solution of the ODE, $y(t)$, satisfies that $|y(t) - y^*| \to 0$ as $t \to \infty$ and if for every $\epsilon > 0$ there exists $\delta > 0$ such that if $y(t)$ is a solution to the ODE satisfying $|y(0) - y^*| < \delta$ then $|y(t) - y^*| < \epsilon$ for all $t \geq 0$.

Since $\gamma < 1$, this implies that

$$|f_2 - f_1| = |\phi(x_2) - \phi(x_1)| \leq \frac{L_h}{1-\gamma} |x_2 - x_1|,$$

which shows that ϕ is Lipschitz continuous.

To show that $\phi(\bar{x})$ is a globally asymptotically stable equilibrium of (10), consider the Lyapunov function candidate

$$V(f) = \frac{1}{2}(f - \phi(\bar{x}))^{\mathrm{T}}(f - \phi(\bar{x})).$$

Along solutions of (10),

$$\begin{aligned}
\frac{d}{dt}V(f(t)) &= (f(t) - \phi(\bar{x}))^{\mathrm{T}} \dot{f}(t) \\
&= (f(t) - \phi(\bar{x}))^{\mathrm{T}}(h(\bar{x}, f(t)) - f(t)) \\
&= (f(t) - \phi(\bar{x}))^{\mathrm{T}}(h(\bar{x}, f(t)) \underbrace{-h(\bar{x}, \phi(\bar{x})) + \phi(\bar{x})}_{=0} - f(t)) \\
&\leq \gamma |f(t) - \phi(\bar{x})|^2 - |f(t) - \phi(\bar{x})|^2
\end{aligned}$$

This implies that

$$\frac{d}{dt}V(f(t)) \leq -2(1-\gamma)V(f(t)).$$

Standard methods from nonlinear systems analysis (e.g., (Khalil, 2001)) establish that $V(f(t))$ decreases exponentially and that $\phi(\bar{x})$ is a globally asymptotically stable equilibrium.

Step 3: Analysis of the slow time scale. Following Proposition A.2, we now consider the differential equation

$$\begin{aligned}
\dot{x}(t) &= F(x(t), \phi(x(t))) \\
\dot{e}_w(t) &= -e_w(t) + w(\phi(x(t))(h(x(t), \phi(x(t))) - \phi(x(t))).
\end{aligned}$$

Since

$$\phi(x(t)) = h(x, \phi(x(t))),$$

we have that

$$\dot{e}_w(t) = -e_w(t),$$

and so the set

$$\{(x, e_w) \mid e_w = 0\}$$

is a global asymptotically stable attractor.[4]

With all of the conditions of Proposition A.2 are satisfied, we can conclude that $\lim_{k\to\infty} e_w(k) = 0$ almost surely. It is worth emphasizing that the proof holds for every test function w.

Note that Proposition A.2 further implies that

$$\lim_{k\to\infty} (f(k) - \phi(X(k))) = 0,$$

almost surely. Since

$$\phi(X(k)) = h(X(k), \phi(X(k))),$$

then we can also conclude the convergence stated in (8).

3.4.2 Relatively bounded rate sequences

The proof for relatively bounded rate sequences progresses similarly to bounded rate sequences. We first write the discrete time iterations similarly to the iterations (9).

$$\begin{aligned}
X(k+1) &= X(k) + \left(\frac{1}{k+1}\right)^{\eta} (F(X(k), f(k)) + M(k)) \\
\tilde{e}_w(k+1) &= \tilde{e}_w(k) + \left(\frac{1}{k+1}\right)^{\eta} (w(f(k))(x(k) - f(k)) - \tilde{e}_w(k)) \\
f(k+1) &= f(k) + \left(\frac{1}{k+1}\right)^{\rho} (x(k) - f(k)) \\
x(k) &= \texttt{rand}[h(X(k), f(k)],
\end{aligned}$$

with $1/2 < \rho < \eta < 1$. The same arguments used previously for bounded rate sequences when applied to the above iterations establish that $\lim_{k\to\infty} \tilde{e}_w(k) = 0$ almost surely. The definition of calibration requires, however, that $\lim_{k\to\infty} e_w(k) = 0$ almost surely, where

$$e_w(k+1) = e_w(k) + \left(\frac{1}{k+1}\right)(w(f(k))(x(k) - f(k)) - e_w(k)).$$

[4] An ODE $\dot{y} = F(y)$ has a stable attractor $\mathcal{Z}$, if for any $\varepsilon > 0$, there exists a $\delta > 0$, such that $\inf_{z\in\mathcal{Z}} |y(0) - z| < \delta$ implies that $\inf_{z\in\mathcal{Z}} |y(t) - z| < \varepsilon$, for all $t \geq 0$. A stable attractor, $\mathcal{Z}$, is globally asymptotically stable if furthermore, for any initial conditions, the solution of the ODE satisfies that $\inf_{z\in\mathcal{Z}} |y(t) - z| \to 0$.

We now proceed to prove that $\lim_{k\to\infty} \tilde{e}_w(k) = 0$ implies that $\lim_{k\to\infty} e_w(k) = 0$. The following technical lemma is the well known Kronecker lemma (e.g., (Durrett, 1991)).

Lemma 3.1 (Kronecker). *Let $s(k)$ be a sequence of real numbers and $S(k) = s(1) + \cdots + s(k)$. Let $\alpha(k)$ be a sequence of positive real numbers with $\lim_{k\to\infty} \alpha(k) = 0$. If $\sum_{k=1}^{K} \alpha(k)s(k)$ converges to a finite limit as $K \to \infty$, then $\alpha(k)S(k) \to 0$ as $k \to \infty$.*

Since

$$\tilde{e}_w(k+1) - \tilde{e}_w(k) = \left(\frac{1}{k+1}\right)^{\eta} (w(f(k))(x(k) - f(k)) - \tilde{e}_w(k)),$$

we have that

$$\sum_{k=1}^{K} \left(\frac{1}{k+1}\right)^{\eta} (w(f(k))(x(k) - f(k)) - \tilde{e}_w(k)) = \tilde{e}_w(K+1) - \tilde{e}_w(1)$$

converges to a finite limit. By Kronecker's lemma,

$$\left(\frac{1}{k+1}\right)^{\eta} \sum_{k=0}^{K} (w(f(k))(x(k) - f(k)) - \tilde{e}_w(k))$$

converges to zero. Consequentially, the running average

$$\left(\frac{1}{k+1}\right) \sum_{k=0}^{K} (w(f(k))(x(k) - f(k)) - \tilde{e}_w(k))$$

also converges to zero. Since $\tilde{e}_w(k) \to 0$ almost surely, we have that the average

$$\lim_{K\to\infty} \left(\frac{1}{k+1}\right) \sum_{k=0}^{K} (w(f(k))(x(k) - f(k))) = 0$$

almost surely. But this average is the calibration error, $e_w(K)$, which completes the proof.

3.4.3 Binary sequences

Unlike the previous proofs, we will *not* assume a model of how the outcome sequence is being generated. Instead, we will take advantage of the redundancy in the binary case and the fact the dynamics of the forecast and the calibration error are essentially scalar. The proof is based on stochastic approximation convergence results presented in Appendix A.

The discrete time iterations for the calibration error of a specific test function, $w \in \mathcal{W}$, and the tracking forecast are

$$e_w(k+1) = e_w(k) + \left(\frac{1}{k+1}\right)(w(f(k))(x(k) - f(k)) - e_w(k))$$
$$f(k+1) = f(k) + \left(\frac{1}{k+1}\right)^{\rho}(x(k) - f(k)).$$

By definition, the components of both $x(k)$ and $f(k)$ sum to unity, i.e., $\mathbf{1}^{\mathrm{T}}x(k) = \mathbf{1}^{\mathrm{T}}f(k) = 1$. There are several consequences of this constraint in the binary case:

– We need only show that the first component of $e_w(k)$ converges to zero to establish weak calibration. This is because sum of the components of $e_w(k)$ will equal zero by Definition 3.
– We can write $f(k)$ in terms of its first component only, i.e.,

$$f(k) = \begin{pmatrix} f_1(k) \\ f_2(k) \end{pmatrix} = \begin{pmatrix} f_1(k) \\ 1 - f_1(k) \end{pmatrix}.$$

– The test function $w(f(k))$ is really a function of a scalar quantity, i.e.,

$$w(f(k)) = w\left(\begin{pmatrix} f_1(k) \\ 1 - f_1(k) \end{pmatrix}\right).$$

Temporary notation: In the rest of the proof, we will consider only the first component of $e_w(k)$, the first component of $f(k)$, and the scalar domain view of $w(f(k))$. In order to avoid cumbersome notation, we will not make this explicit.

We can write the equation for the tracking forecast as

$$(x(k) - f(k)) = \frac{f(k+1) - f(k)}{\left(\frac{1}{k+1}\right)^{\rho}}.$$

Substituting this into the equation for the calibration error results in

$$e_w(k+1) = e_w(k) + \underbrace{\left(\frac{1}{k+1}\right)}_{a(k)}\left(-e_w(k) + \underbrace{w(f(k))\left(\frac{f(k+1) - f(k)}{\left(\frac{1}{k+1}\right)^{\rho}}\right)}_{M(k)}\right).$$

If we can show that the product of $a(k)$ and $M(k)$, defined above, satisfies the Kushner-Clark condition (25), then the resulting differential equation in stochastic approximation will be

$$\dot{e}_w(t) = -e_w(t).$$

Therefore, the discrete iterations for $e_w(k)$ converge to zero.

Towards this end, let us inspect

$$a(k)M(k) = \left(\frac{1}{k+1}\right) w(f(k)) \left(\frac{(f(k+1) - f(k))}{\left(\frac{1}{k+1}\right)^{\rho}}\right)$$
$$= \left(\frac{1}{k+1}\right)^{1-\rho} w(f(k))(f(k+1) - f(k)). \tag{11}$$

Given any test function, $w \in \mathcal{W}$, define $v : [0, 1] \to \mathbb{R}^+$ by

$$v(x) = \int_0^x w(s)ds.$$

Since w is Lipschitz continuous, and since $v(x_2) - v(x_1) = \int_{x_1}^{x_2} w(s)ds$, we have

$$w(x_1)(x_2 - x_1) - L_w(x_2 - x_1)^2 \leq v(x_2) - v(x_1) \leq w(x_1)(x_2 - x_1) + L_w(x_2 - x_1)^2,$$

which in turn implies that

$$v(x_2) - v(x_1) - L_w(x_2 - x_1)^2 \leq w(x_1)(x_2 - x_1)$$
$$\leq v(x_2) - v(x_1) + L_w(x_2 - x_1)^2. \tag{12}$$

Substituting v in place of w in (11) and using (12) we obtain

$$a(k)M(k) \leq \left(\frac{1}{k+1}\right)^{1-\rho} (v(f(k+1)) - v(f(k))) \tag{13a}$$
$$+ \left(\frac{1}{k+1}\right)^{1-\rho} L_w(f(k+1) - f(k))^2. \tag{13b}$$

The second term (13b) is absolutely summable since

$$(f(k+1) - f(k))^2 \leq \left(\frac{1}{k+1}\right)^{2\rho}$$

by definition. Therefore, this term satisfies the Kushner-Clark condition (25).

An application of Lemma A.1 establishes that the first term (13a) also satisfies the Kushner-Clark condition. A similar analysis holds for the reverse inequality (replacing the $+$ with $-$ in (13b)).

Finally, Proposition A.1 implies that the calibration error, $e_w(k)$, converges to zero. Note that we used deterministic arguments to establish the Kushner-Clark condition, and so we have a deterministic guarantee of convergence (as opposed to "almost surely").

3.5 Non binary sequences

In this section we show that the fact that tracking forecast is calibrated for binary alphabet cannot be extended to richer alphabets. Specifically, we show using a counterexample that tracking forecast is not calibrated for a trinary alphabet prediction problem. We use a somewhat informal language in this example since an exact argument would be tedious. Our example is similar to the Shapley polygon (Shapley, 1964).

Consider a three letter sequence where a tracking forecast is used with some $0 < \rho \leq 1$. We will construct a policy for Nature that guarantees that the tracking forecast is not weakly calibrated. For a visual interpretation of the example, see Fig. 1. Suppose that in this example the following values are given to points in the simplex: $A = (4/7, 1/7, 2/7)$, $B = (2/7, 4/7, 1/7)$, and $C = (1/7, 2/7, 4/7)$.

Nature's strategy is to start by following A for a long time. By following we mean that either Nature randomizes between the three letters as prescribed by A or that Nature uses some deterministic time sharing policy so that the empirical frequency is approximately A. In either case, up to small fluctuations (which we will ignore), we assume that the tracking forecast is A after sufficiently long time (we will denote this time by k_0). We will write this as $f(k) \approx A$ and take it to mean that up to some small $\epsilon > 0$ we have that $|f(k) - A| < \epsilon$. Now, Nature starts playing e_2 repeatedly. We claim that after sufficiently long time, k_B, the tracking forecast is $f(k_B) \approx B$. Indeed, B is on the segment between A and e_2 so that after observing the second letter once we have that $f(k_0 + 1) \approx A + (1/(k_0 + 1))^{\rho}(e_2 - A) = (1 - (1/(k_0 + 1))^{\rho})A + (1/(k_0 + 1))^{\rho} e_2$ which is on the segment between A and e_2. As Nature repeatedly plays action 2, $f(k)$ will traverse the segment between A and e_2. We can therefore find k_B such that $(Ak_0 + e_2(k_B - k_0))/k_B \approx B$. Note that the exact value of k_B depends on ρ. When the tracking forecast reaches B, that is $f(k_B) \approx B$, Nature switches to playing e_3. After a long enough time, k_C, we have that $f(k_c) \approx C$ (again, this is because C is on the segment between B and e_3). Now, Nature switches to playing e_1 until time k_A where $f(k_A) \approx A$. It follows that from time k to k_B the observation has empirical frequency of e_2 while the tracking forecast prediction ranges from A to B. Similarly, from time k_B until k_C the actual empirical frequency is e_3 while the forecast is between B to C. Finally, from time k_C to k_A the actual empirical

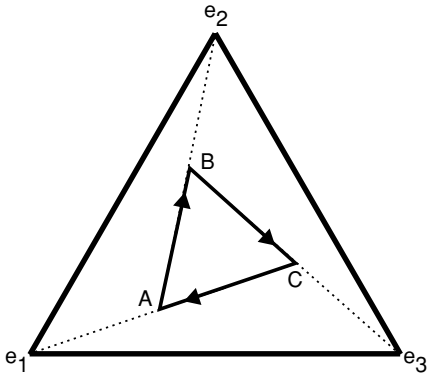

Fig. 1 A visual illustration of the Shapley polytope used in the 3 letters counterexample

frequency is e_1 while the tracking forecast is between C and A. Nature can now repeat the strategy again by using e_1 followed by e_2 followed by e_3 in a similar way to before. This process can proceed ad-infinitum.

In terms of weak calibration as defined in Eq. (3), we can take three smooth testing rules: a smoothed one for each of the segments AB, BC, CA (a smoothed inflated indicator, to account for errors due to finite samples). It is easy to verify that the tracking forecast is not weakly calibrated.

3.6 Randomized tracking forecasts and strong calibration

In this section we outline a method to obtain a strongly calibrated forecast based on tracking forecast. Kakade and Foster (2004) show how to use randomization in conjunction with a weakly calibrated forecast to achieve calibration. Since their calibration approach is based on a fine discretization of the simplex, their randomization is based on rounding to vertices of the discretization of the simplex. In contrast, our approach is based on tracking forecast and would therefore be calibrated only for a restricted classes of sequences. In order to convert weak to strong calibration, we add some small additive noise and show that while this adversely affects the accuracy of the weak calibration, it allows us to obtain strong calibration.

The *randomized* tracking forecast, $\tilde{f}(k)$, is defined as

$$\tilde{f}(k) = \Pi_{\Delta}[f(k) + h(k)], \tag{14}$$

where

- $f(k)$ is the usual tracking forecast (6).
- $h(k)$ is an independent and identically distributed random vector with uniformly distributed elements over an interval $[-\bar{h}, \bar{h}]$, for some $\bar{h} > 0$.

The projection operator, Π_{Δ} assures that the randomized tracking forecast lies in the simplex.

Now recall the calibration error with respect to an indicator function, $I_{p,\delta}$, defined in (1), applied to the randomized tracking forecast. The calibration error can be written in the recursive form,

$$e_{p,\delta}(k+1) = e_{p,\delta}(k) + \left(\frac{1}{k+1}\right)(I_{p,\delta}(\tilde{f}(k))(x(k) - \tilde{f}(k)) - e_{p,\delta}(k)).$$

Define

$$w(f) = \mathbb{E}\left[I_{p,\delta}(\Pi_{\Delta}[f + h])\right],$$

where the expectation is taken over h. We can now rewrite

$$e_{p,\delta}(k+1)$$
$$= e_{p,\delta}(k) + \left(\frac{1}{k+1}\right)(w(f(k))(x(k) - f(k)) - e_{p,\delta}(k) + M_1(k) - M_2(k)), \tag{15}$$

where

$$M_1(k) = (I_{p,\delta}(\Pi_\Delta[f(k) + h(k)]) - w(f(k)))(x(k) - f(k)),$$
$$M_2(k) = I_{p,\delta}(\tilde{f}(k))(\Pi_\Delta[f(k) + h(k)] - f(k)).$$

By construction,

$$\mathbb{E}\left[M_1(k) \mid f(k), x(k), e_{p,\delta}(k)\right] = 0.$$

Furthermore,

$$|M_2(k)| \leq C\bar{h},$$

for some constant, C, that does not depend on the outcome sequence or indicator function.

It is clear from (15) that if the tracking forecast, f, is weakly calibrated, then the randomized tracking forecast, $\tilde{f}$, will be ε-calibrated for $\varepsilon = C\bar{h}$ — *provided that* the function w is Lipschitz continuous.

Figure 2 illustrates the effect of randomization on an indicator function. Let

$$I(x) = \begin{cases} 1 & -\delta \leq x \leq \delta; \\ 0 & \text{otherwise.} \end{cases}$$

Now define

$$w(x) = \mathbb{E}\left[I(x + h)\right],$$

where h is a uniformly distributed random variable over the interval $[-\bar{h}, \bar{h}]$. Then, assuming that $\bar{h} < \delta$,

$$w(x) = \begin{cases} 0 & x \in (-\infty, -\delta - \bar{h})); \\ (x + (\delta + \bar{h}))/(2\bar{h}) & x \in [-\delta - \bar{h}, -\delta + \bar{h}]; \\ 1 & x \in [-\delta + \bar{h}, \delta - \bar{h}]; \\ 1 - (x - (\delta - \bar{h}))/(2\bar{h}); & x \in [\delta - \bar{h}, \delta + \bar{h}]; \\ 0 & x \in (\delta + \bar{h}, \infty), \end{cases}$$

which is Lipschitz continuous.

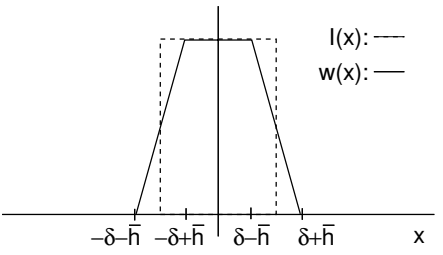

Fig. 2 Randomized indicator function

The procedure we outlined above is general, and would lead to an ε-calibrated forecast as long as the tracking forecast is weakly calibrated. We therefore have the following corollary whose proof is a combination of Theorem 3.1 and the above argument.

Corollary 3.1. *By choosing $\bar{h}$ small enough, the randomized tracking forecast* (14) *is ε-calibrated over the following classes of sequences:*

1. *Bounded rate sequences for any* $1/2 < \rho < 1$*;*
2. *Relatively bounded rate sequences for any* $1/2 < \rho < \eta < 1$*;*
3. *Binary sequences for any* $0 < \rho < 1$.

4 Opponent forecasting in repeated games

There is a substantial body of literature on the topic of learning in games and the related topic of evolutionary games. This includes several recent monographs (Fudenberg & Levine, 1998; Hofbauer & Sigmund, 1998; Samuelson, 1997; Young, 1998, 2004; Weibull, 1995). At issue in much of this work is understanding the limiting behavior of interacting players that adapt their strategies given incomplete information.

One approach to learning in games is to have each player compute some sort of forecast of opponent actions and play a best response to this forecast. Accordingly, the limiting behavior of player actions strongly depends on the specific method for forecasting. For example, in "fictitious play", forecasts are simply the empirical frequencies of opponents actions. In special classes of games, player strategies converge to a Nash equilibrium, but in general, the limiting behavior need not exhibit convergence.

Placing more stringent requirements on the forecasts can result in stronger convergence properties for general games. In particular, if players use calibrated forecasts (Foster & Vohra, 1997), then player strategies asymptotically converge to the set of correlated equilibria. When players use calibrated forecasts of *joint* actions, then player strategies converge to the convex hull of Nash equilibria (Kakade & Foster, 2004).

The computational burden of existing universal calibration algorithms effectively prohibits the implementation of such calibration based methods. In the following sections, we explore the use of tracking forecasts in learning in games.

4.1 Learning in repeated games

This section outlines our framework and notation for learning in games. Relevant references are Hart (2005), Young (2004), and Fudenberg and Levine (1998). We start with defining the game setup in our notation in Section 4.1.1 We continue by showing how several popular learning schemes can be unified within the notations in Section 4.1.2. We continue with the case where one of the players uses a tracking forecast, while the other uses one of the standard "slow" algorithms in Section 4.2. We show that the fast player obtains the best response against the slow player's actual play. We finally consider the case where the players play equally fast in Section 4.3. We show that by playing best response with calibrated forecast in a certain way, we

can obtain convergence to a Nash equilibrium under certain conditions on the reward matrix.

4.1.1 Static games

We consider a two player game in strategic form with players $\mathcal{P}_1$ and $\mathcal{P}_2$. For convenience, we assume each player has m moves. A strategy for player $\mathcal{P}_i$, is $p_i \in \Delta$. The standard interpretation is that p_i represents probabilistic (mixed) strategies. Each player selects an integer action $a_i \in \text{vert}[\Delta]$ according to the probability distribution p_i, i.e.,

$$a_i = \texttt{rand}[p_i].$$

The reward to player $\mathcal{P}_i$ when $\mathcal{P}_i$ selects action a_i and $\mathcal{P}_{-i}$ chooses a_{-i} (we adopt the notation that $-i$ denotes the "other" player) is

$$U_i(a_i, a_{-i}; p_i) = a_i^{\mathrm{T}} M_i a_{-i} + \tau \mathcal{H}(p_i),$$

which is characterized by the matrix M_i and parameter $\tau \geq 0$. The reward to player $\mathcal{P}_i$ is the element of M_i corresponding to the a_i^{th} row and a_{-i}^{th} column, plus the weighted entropy of her strategy. We add the parameterized penalty term, $\tau \mathcal{H}(p_i)$, to the reward to allow consideration of several different algorithms.

For a given strategy pair, (p_1, p_2), the expected rewards are

$$\begin{aligned} \mathcal{U}_i(p_i, p_{-i}) &= \mathbb{E}\left[a_i{}^{\mathrm{T}} M_i a_{-i}\right] + \tau \mathcal{H}(p_i) \\ &= p_i^{\mathrm{T}} M_i p_{-i} + \tau \mathcal{H}(p_i). \end{aligned}$$

Define the *best response* mappings,

$$\beta_i : \Delta \to \Delta,$$

by

$$\beta_i(p_{-i}) = \arg\max_{p_i \in \Delta} \mathcal{U}_i(p_i, p_{-i}).$$

For $\tau > 0$, the best response turns out to be the logit or soft-max function (see Notation section)

$$\beta_i(p_{-i}) = \sigma(M_i p_{-i}/\tau).$$

For $\tau = 0$, the best response amounts to selecting a maximizing simplex vertex, which need not be unique.

A Nash equilibrium is a pair $(p_1^*, p_2^*) \in \Delta \times \Delta$ such that for all $p_i \in \Delta$,

$$\mathcal{U}_i(p_i, p_{-i}^*) \leq \mathcal{U}_i(p_i^*, p_{-i}^*), \quad i \in \{1, 2\}, \tag{16}$$

i.e., each player has no incentive to deviate from an equilibrium strategy provided that the other player maintains an equilibrium strategy. In terms of the best response mappings, a Nash equilibrium is a pair (p_1^*, p_2^*) such that

$$p_i^* = \beta_i(p_{-i}^*), \quad i \in \{1, 2\}.$$

4.1.2 Repeated games

Suppose now that the game is sequentially repeated over stages $k = 0, 1, 2, \ldots$. At each stage, k, player $\mathcal{P}_i$ uses her *current* strategy, $p_i(k)$, to generate her current action, $a_i(k)$. Again, player $\mathcal{P}_i$ receives a reward, $U_i(a_i(k), a_{-i}(k); p_i(k))$, according to her utility function evaluated on the total current action profile.

Player strategies, $p_i(k)$, are updated, or adapted, at each stage according to the information available to player $\mathcal{P}_i$ over times $\{0, \ldots, k-1\}$. We assume that after each stage k, player $\mathcal{P}_i$ can observe the actions, $a_{-i}(k)$, of the other player.

The following are well known approaches to updating player strategies. The sequence of actions of each of the algorithms is a bounded rate sequence. In each of these methods, players compute the empirical frequencies (as in Eq. (7) of their actions) according to

$$q_i(k+1) = q_i(k) + \left(\frac{1}{k+1}\right)(a_i(k) - q_i(k)). \tag{17}$$

Smooth fictitious play: Players compute a smoothed ($\tau > 0$) best response to the empirical frequencies of their opponent's actions. Player $\mathcal{P}_i$ plays according to:

$$a_i(k) = \texttt{rand}[p_i(k)], \tag{18a}$$

$$p_i(k) = \beta_i(q_{-i}(k)). \tag{18b}$$

Gradient play: Players update their strategies according to the evolving gradient of the non-smoothed ($\tau = 0$) utility. Player $\mathcal{P}_i$ plays according to:

$$a_i(k) = \texttt{rand}[p_i(k)],$$

$$p_i(k) = \Pi_\Delta[q_i(k) + M_i q_{-i}(k)].$$

The terminology stems from the gradient equation (for $\tau = 0$)

$$\nabla_{p_i} \mathcal{U}_i(p_i, p_{-i}) = M_i p_{-i}.$$

Exponential regret matching: Players accumulate retrospective "regrets", $r_i(k)$, of past actions and update their strategies to reduce regret. Player $\mathcal{P}_i$ plays according to:

$$a_i(k) = \texttt{rand}[p_i(k)],$$

$$p_i(k) = \sigma(r_i(k)/\tau),$$

$$r_i(k+1) = r_i(k) + \left(\frac{1}{k+1}\right)(M_i a_{-i}(k) - a_i(k)^{\mathrm{T}} M_i a_{-i}(k) \times \mathbf{1}).$$

4.2 Slower opponents

One way to view either smooth fictitious play or gradient play is in terms of an *opponent model*. The following reflects the perspective of player $\mathcal{P}_1$ modeling player $\mathcal{P}_2$:

> Suppose our opponent, $\mathcal{P}_2$, uses a *stationary* strategy, i.e.,
>
> $$p_2(k) = s^* \in \Delta, \quad \forall k = 0, 1, 2 \ldots.$$
>
> Then, by the law of large numbers, the empirical frequencies of $\mathcal{P}_2$ will converge so that $q_2(k) \to s^*$. This, in turn, implies that our strategy, $p_1(k)$, will asymptotically approach to the best response to our opponent's strategy, i.e., $p_1(k) \to \beta_1(s^*)$.

This perspective has a strong connection to the notion of calibration over a class of outcome sequences. Namely, empirical frequencies constitute a *calibrated forecast* for the class of outcome sequences generated by stationary strategies.

We now introduce a modification of smooth fictitious play in which a player uses a tracking forecast in lieu of empirical frequencies.

Smooth fictitious play with tracking forecasts: Player $\mathcal{P}_i$ plays according to:

$$a_i(k) = \texttt{rand}[p_i(k)], \tag{19a}$$

$$p_i(k) = \beta_i(f_{-i}(k)), \tag{19b}$$

$$f_{-i}(k+1) = f_{-i}(k) + \left(\frac{1}{k+1}\right)^{\rho} (a_{-i}(k) - f_{-i}(k)). \tag{19c}$$

Theorem 4.1. *Suppose Player $\mathcal{P}_1$ plays according to tracking forecast smooth fictitious play* (19) *with $1/2 < \rho < 1$. If the outcome sequence generated by player $\mathcal{P}_2$ is a:*

- *Bounded rate sequence, or*
- *Relatively bounded rate sequence, with $\rho < \eta < 1$,*

then, almost surely,

$$\lim_{k\to\infty} (f_2(k) - p_2(k)) = 0,$$

which implies that

$$\lim_{k\to\infty} (p_1(k) - \beta_1(p_2(k))) = 0.$$

Proof: The proof follows from the same arguments used in Theorem 3.1 for bounded rate and relatively bounded rate sequences. □

In terms of the previous discussion on opponent modeling, the use of tracking forecasts can be viewed as a broader class under which a player's strategy approximates the stage-by-stage best response.

Corollary 4.1. *The conclusions of Theorem* 4.1 *hold if player* $\mathcal{P}_2$ *plays according to* (1) *Smooth fictitious play,* (2) *Gradient play,* (3) *Exponential regret matching, or* (4) *Smooth fictitious play with tracking forecast, with exponent* $\eta > \rho$.

4.3 Equally fast opponents and convergence to nash equilibrium

In the previous section, player $\mathcal{P}_1$ had a sort of strategic advantage in playing a best response to a weakly calibrated forecast. We now consider *both* players using tracking forecasts.

The focus in this section is shifted away from forecasting and redirected to the issue of convergence to Nash equilibrium. Indeed, if both players are using a tracking player, then *neither* player's forecast is guaranteed to be weakly calibrated (except in 2-move games).

In particular, we will consider how one may overcome non-convergence properties that are exhibited by a broad class of strategy update mechanisms. Hart and Mas-Colell (2003) construct a game such that if players use strategies that are functions of the *current value* of the empirical frequencies, then convergence to a (mixed) Nash equilibrium cannot occur. The result strongly relies on utility functions not being shared among players. This non-convergence result is reminiscent of earlier results, such as Crawford (1985), that established non-convergence for certain special classes of strategy update mechanisms.[5]

Recent work (Arslan & Shamma, 2004; Shamma & Arslan, 2005) showed that it is possible to overcome this lack of convergence by processing the empirical frequencies in a "dynamic" manner, i.e., by allowing strategies to depend on the evolution of the empirical frequencies, and not just their current values. This is achieved by introducing auxiliary variables, upon which to base strategy adaptation. Related work is Hart and Mas-Colell (2004), which also investigates the potential benefit of introducing increased memory to enable converge to Nash equilibria.

4.3.1 Conditions for non-convergence in smooth fictitious play

The method of stochastic approximation can be used to deduce the *lack* of convergence to a Nash equilibrium (see Theorem 5.1 in Benaim & Hirsch (1999)). When both players use smooth fictitious play as in Eqs. (17)–(18), the asymptotic behavior of the discrete-time iterations may be analyzed by the differential equations

$$\dot{q}_1(t) = -q_1(t) + \beta_1(q_2(t)), \tag{20a}$$

$$\dot{q}_2(t) = -q_2(t) + \beta_2(q_1(t)). \tag{20b}$$

[5] Of course, there are special classes of games for which adaptation mechanisms such as fictitious play are known to converge. See (Hart, 2005) for further discussion.

Let (q_1^*, q_2^*) be a Nash equilibrium, which, by definition, will be an equilibrium point of (20). The local asymptotic stability of (20) may be assessed by examining the eigenvalues of the appropriate Jacobian matrix. Linearizing the right hand side of (20) at the equilibrium (q_1^*, q_2^*) results in

$$\begin{pmatrix} -I & \nabla\beta_1(q_2^*) \\ \nabla\beta_2(q_1^*) & -I \end{pmatrix}.$$

However, this Jacobian matrix does not reflect that the $q_i(t)$ are constrained to evolve on the simplex. We can write any $q_i(t) \in \Delta$ as

$$q_i(t) = q_i^* + \delta q_i(t).$$

The elements of both $q_i(t)$ and q_i^* sum to unity, therefore the elements of $\delta q_i(t)$ must sum to zero. Equivalently,

$$\delta q_i(t) = \mathcal{N}\tilde{q}_i(t),$$

for some $\tilde{q}_i(t)$, where $\mathcal{N}$ an $m \times (m-1)$ matrix (recall that m is the number of moves for each player) such that

$$\mathcal{N}^{\mathrm{T}}\mathcal{N} = I, \quad \mathcal{N}^{\mathrm{T}}\mathbf{1} = 0. \tag{21}$$

The appropriate *reduced order* Jacobian matrix is in fact

$$J = \begin{pmatrix} -I & \mathcal{N}^{\mathrm{T}}(\nabla\beta_1(q_2^*))\mathcal{N} \\ \mathcal{N}^{\mathrm{T}}(\nabla\beta_2(q_1^*))\mathcal{N} & -I \end{pmatrix}, \tag{22}$$

which reflects the dynamics being written in terms of the $\tilde{q}_i$. A more detailed discussion may be found in Shamma and Arslan (2005, Eqs. (9)–(12)).

The following is an adaptation of Theorem 5.1 in Benaim and Hirsch (1999). It is being stated for comparison to the convergence result in the next section.

Proposition 4.1. *Consider smooth fictitious play* (17)–(18) *with a Nash equilibrium* (q_1^*, q_2^*). *If any eigenvalue of the Jacobian matrix, J, in* (22) *has a positive real part, then the event*

$$\lim_{k\to\infty} q_i(k) = q_i^*, \quad i \in \{1, 2\},$$

occurs with zero probability.

4.3.2 Enabling convergence to nash equilibrium

We will show that the introduction of a tracking forecast can enable, in some games, convergence to Nash equilibrium. We first introduce a modified tracking forecast that will greatly simplify the analysis.

Modified tracking forecast: For the outcome sequence, $x(k)$, the modified tracking forecast is defined as

$$f(k+1) = f(k) + \left(\frac{\lambda}{k+1}\right)(x(k) - f(k)), \tag{23}$$

for some fixed $\lambda \gg 1$. It can happen that the modified tracking forecast, $f(k)$, will lie outside of the simplex, but this does not affect the following discussion.

The form of the modified tracking forecast serves to mimic the effect of the step size in computing the original tracking forecast as compared to the step size in computing an empirical frequency. For any $\rho < 1$,

$$\left(\frac{1}{k+1}\right)^{\rho} \Big/ \left(\frac{1}{k+1}\right) \to \infty$$

as $k \to \infty$. The modified tracking forecast reflects this ratio by using $\lambda \gg 1$. Indeed, it is possible to show that the modified tracking forecast is weakly ε-calibrated for bounded rate sequences for $\varepsilon \approx 1/\lambda$.

We now introduce another modification of smooth fictitious play that combines smooth fictitious play with tracking forecasts and standard smooth fictitious play (with empirical frequencies).

Smooth fictitious play with combined forecasts: Player $\mathcal{P}_i$ plays according to:

$$a_i(k) = \texttt{rand}[p_i(k)], \tag{24a}$$

$$p_i(k) = \beta_i((1-\gamma)q_{-i}(k) + \gamma f_{-i}(k)), \tag{24b}$$

$$f_{-i}(k+1) = f_{-i}(k) + \left(\frac{\lambda}{k+1}\right)(a_{-i}(k) - f_{-i}(k)), \tag{24c}$$

for some $0 \leq \gamma \leq 1$ and $\lambda \gg 1$.

In words, each player uses a smoothed best response to a convex combination of the empirical frequencies and tracking forecasts. For $\gamma = 0$, this is standard smooth fictitious play, and for $\gamma = 1$, this is smooth fictitious play with (modified) tracking forecasts.

Theorem 4.2. *Consider smooth fictitious play with combined forecasts* (24) *with a Nash equilibrium* (q_1^*, q_2^*). *Let* $a_i + jb_i$ *denote*[6] *the eigenvalues of the Jacobian matrix* J *in* (22) *for (standard) smooth fictitious play. Then the event*

$$\lim_{k\to\infty} q_i(k) = q_i^*, \quad i \in \{1, 2\},$$

occurs with ***strictly positive*** *probability for sufficiently large* λ *if and only if,*

1. $0 \leq \gamma \leq 1$, *if* $\max_i a_i < 0$,
2. $\max_i \frac{a_i}{a_i^2+b_i^2} < \frac{\gamma}{1-\gamma} < \frac{1}{\max_i a_i}$, *if* $\max_i a_i \geq 0$.

[6] Where $j \equiv \sqrt{-1}$.

Condition 1 in Theorem 4.2 implies that the linearization of smooth fictitious play is asymptotically stable (compare to Proposition 4.1). In this case, convergence of empirical frequencies to the Nash equilibrium in smooth fictitious play with combined forecasts is still possible for any mixture $0 \leq \gamma \leq 1$.

More important is Condition 2. This implies that, under certain conditions, smooth fictitious play with combined forecasts can converge to Nash equilibrium in situations where standard smooth fictitious play does not.

Appendix B presents the proof of Theorem 4.2. The main idea is to show that the differential equations associated with smooth fictitious play with combined forecasts in (24) closely resemble (for large λ) the differential equations for so-called "derivative action fictitious play" in Arslan and Shamma (2004), and Shamma and Arslan (2003). The hypotheses of Theorem 4.2 imply local asymptotic stability of the Nash equilibrium for the derivative action fictitious play differential equations. Because of the close resemblance, the Nash equilibrium of smooth fictitious play with combined forecasts also will be locally asymptotically stable.[7] We can then invoke Theorem 5.4 of Benaim and Hirsch (1999) to conclude convergence to Nash equilibrium with positive probability.

5 Numerical simulations

In this section we present some experiments with simple toy examples. We will consider the Shapley game that was introduced in Shapley (1964) as an example to the cycles in (non-smooth) fictitious play. The game is defined by the utility matrices

$$M_1 = M_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

The unique Nash equilibrium is

$$p_1^* = p_2^* = (1/3, \quad 1/3, \quad 1/3).$$

All of the simulations below use a smoothed best response with $\tau = 0.05$. The tracking forecasts exponent is $\rho = 0.6$.

Smooth fictitious play: Both players use smooth fictitious play (18). Figure 3 shows the empirical frequency history. The figure illustrates the well known oscillations associated with the Shapley game. The average rewards for each player is approximately $(1/2, 1/2)$. Note that the average rewards associated with the Nash equilibrium are $(1/3, 1/3)$. So although the behavior is oscillatory, the average is greater.

[7] An equilibrium point y^* of an ODE $\dot{y} = F(y)$ is locally asymptotically stable if y^* is a stable attractor and if there exists an open ball B around y^* such that if the initial conditions are in B then the solution of the ODE satisfies that $|y(t) - y^*| \to 0$.

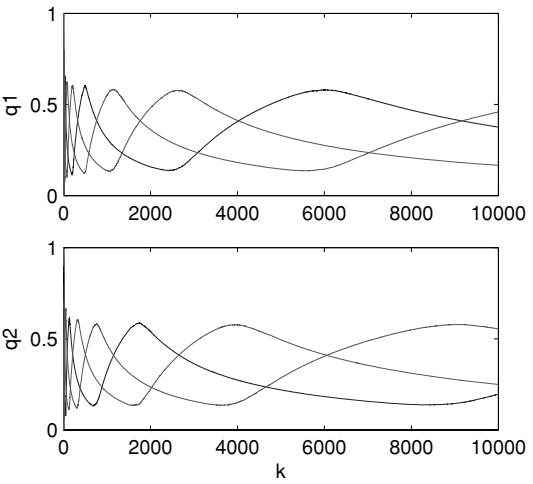

Fig. 3 Smooth fictitious play with empirical frequencies. The top figure is the empirical frequency of each of the three actions for player 1 as a function of the step. The bottom figure is the same for player 2

Tracking forecasts vs. Empirical frequencies: Player $\mathcal{P}_1$ uses smooth fictitious play with tracking forecasts (19) while Player $\mathcal{P}_2$ uses smooth fictitious play with empirical frequencies (18). In terms of Theorem 4.1, player $\mathcal{P}_2$ is "slower", and so player $\mathcal{P}_1$ asymptotically plays the best response to player $\mathcal{P}_2$'s stage-by-stage strategy. Figure 4 shows the history of empirical frequencies. It turns out that these converge to the Nash equilibrium. In fact, the governing differential equation for the empirical frequencies is

$$\dot{q_1}(t) = -q_1(t) + \beta_1(\beta_2(q_1(t))),$$
$$\dot{q_2}(t) = -q_2(t) + \beta_2(q_1(t)).$$

These equations were studied in Leslie and Collins (2003) where players adapt at different time-scales. The difference here is that in the present paper, both players

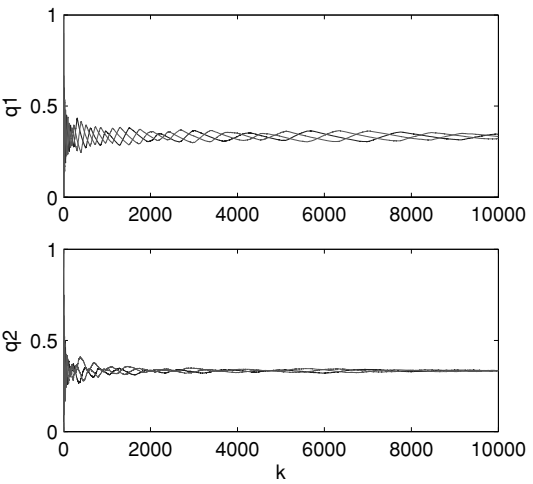

Fig. 4 Smooth fictitious play: Tracking forecast vs. Empirical frequencies

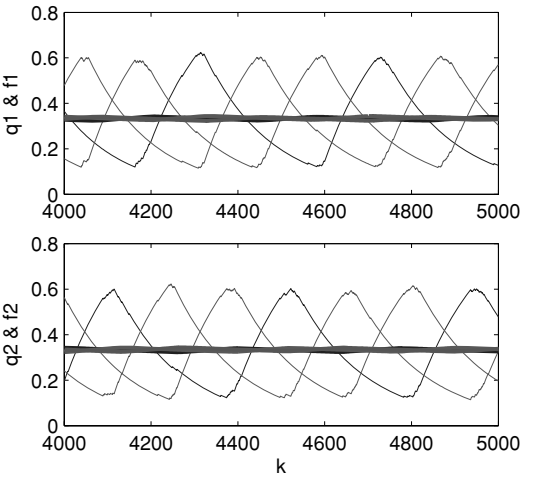

Fig. 5 Smooth fictitious play with tracking forecasts. The thick lines are the empirical frequencies while the thin lines are the tracking forecasts of each action. Top figure is for player 1 and the bottom figure is for player 2

learn at the same rate, but one player has a superior forecast. As expected, the average rewards to each player approach $(1/3, 1/3)$, which is lower than the oscillatory case of Fig. 3.

Tracking forecasts: Both players use smooth fictitious play with tracking forecasts (19). Note that the strategy updates mimic standard smooth fictitious play, but at the forecast level:

$$p_i(k) = \beta_i(f_{-i}(k)),$$
$$f_i(k+1) = f_i(k) + \left(\tfrac{1}{k+1}\right)^{\rho}(x_i(k) - f_i(k)).$$

The tracking forecasts exhibit the same oscillations observed before, while the empirical frequencies average these oscillations. Figure 5 shows a close-up of both the empirical frequencies and tracking forecasts. Once again, the average rewards are $(1/2, 1/2)$.

While the averaging effect of the empirical frequencies is expected, the convergence to the Nash equilibrium values turns out to be a consequence of a symmetry in the Shapley game and is coincidental. Figure 6 shows behavior for a *modified* Shapley game, where M_1 is changed to

$$M_1 = \begin{pmatrix} 0 & 3 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

The tracking forecasts exhibit the oscillations that would be seen in standard smooth fictitious play, but at a faster timescale. Once again, the empirical frequencies flatten out to the average behavior of the oscillatory tracking forecasts. However, the asymptotic values no longer *not* coincide with a Nash equilibrium.

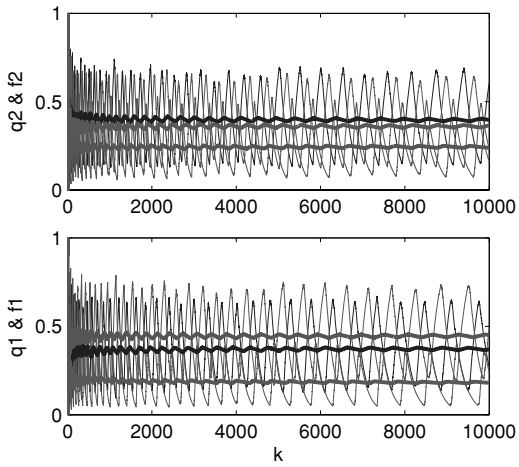

Fig. 6 Smooth fictitious play with tracking forecasts: Modified Shapley game. The thick lines are the empirical frequencies while the thin lines are the tracking forecasts of each action. Top figure is for player 1 and the bottom figure is for player 2

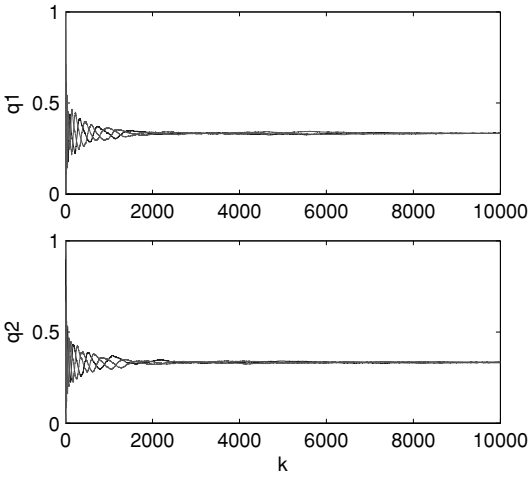

Fig. 7 Smooth fictitious play with combined forecasts (for the original Shapley game)

Combined forecast: Both players use smooth fictitious play with combined forecasts (24) with $\gamma = 0.1$. The Jacobian matrix for the Shapley game satisfies the conditions of Theorem 4.2. Figure 7 shows that the empirical frequencies converge to the Nash equilibrium. The average rewards approach $(1/3, 1/3)$. Note that while the analysis of Theorem 4.2 was for the modified tracking forecast (23), the simulations use the original tracking forecast (6).

6 Concluding remarks

The proposed tracking forecast offers a tradeoff between universality and efficiency. On the one hand, it is an online algorithm which is easy to implement. On the other hand, it is calibrated with respect to a non-trivial class of sequences. When playing a repeated game, it enables convergence to a Nash equilibrium in a wider range of games

than the range of games for which the smooth fictitious play converges. The simplicity of the framework suggests possible extensions to more complicated decision setups, like multi-stage competitive decision problems and perhaps even stochastic games (Filar & Vrieze, 1996).

The resulting algorithm is not weakly calibrated against any source as demonstrated in Section 3.5. It is an open question if there is an efficient forecasting algorithm which is weakly calibrated against all sources (complexity here is in the sense of Blum et al. (1996)). A promising approach in that respect is online convex programming (Zinkevich, 2003) where one tries to track the best solution to a combination of convex functions. The algorithm of Zinkevich (2003) is simple and efficient, however it is not clear if it is possible to construct an ϵ-calibrated scheme based on online convex programming that would not have increasing complexity with respect to ϵ.[8]

The proof technique used in this paper are based on analysis of stochastic approximation type algorithms. It should be possible to provide convergence rate results using existing results for convergence rates for standard stochastic approximation (Borkar & Meyn, 2000; Kushner & Yin, 1997) and for two time scale stochastic approximation (Konda & Tsitsiklis, 2004). Another important issue concerns characterizing the complexity/universality tradeoff. That is, determining the amount of memory requirements that are *necessarily required* by a universally calibrated scheme.

Appendix

A. The ODE method of stochastic approximation

The ODE method of stochastic approximation enables one to assess the limiting behavior of discrete time stochastic iterations via the analysis of continuous time differential equations. Reference are Benaim (1999), and Kushner and Yin (1997). For the sake of completeness, we provide some standard results from stochastic approximation below.

Proposition A.1 (Standard stochastic approximation). *Consider a random sequence, $X(k)$, generated by*

$$X(k+1) = X(k) + a(k)(F(X(k)) + M(k))$$

where

1. $F : \mathbb{R}^m \to \mathbb{R}^m$ *is a Lipschitz function.*
2. $a(k)$ *are nonnegative numbers such that*

$$\sum_k a(k) = \infty$$

[8] We thank Martin Zinkevich for helpful discussions.

and

$$\lim_{k\to\infty} a(k) = 0.$$

1. $M(k) \in \mathbb{R}^m$ *is any sequence satisfying the following. For all real* $T > 0$,

$$\lim_{n\to\infty} \sup_{\left\{\ell \geq n \,:\, \sum_{k=n}^{\ell} a(k) \leq T\right\}} \left| \sum_{k=n}^{\ell} a(k) M(k) \right| = 0. \tag{25}$$

2. $\sup_k |X(k)| < \infty$.
3. *The differential equation*

$$\dot{x}(t) = F(x(t))$$

has a unique global asymptotically stable equilibrium, x^*.

Then

$$\lim_{k\to\infty} x(k) = x^*.$$

Equation (25) is sometimes referred to as the Kushner-Clark condition (e.g, (Wang, Chong, & Kulkarni, 1996)). Typical sufficient conditions (Benaim, 1999) for (25) are that the $M(k)$ are random variables such that:

$$\mathbb{E}\left[M(k) \mid X(k)\right] = 0.$$

Furthermore, either

A1. $\sum_{k=0}^{\infty} a^2(k) < \infty$,
A2. $\sup_k \mathbb{E}[M(k)^{\mathrm{T}} M(k)] < \infty$,

or

B1. $a(k) = o(\frac{1}{\log k})$,
B2. $\sup_k |M(k)| < \infty$.

These are sufficient conditions for the Kushner-Clark to be satisfied almost surely. It is not always necessary to impose a random structure to assure these conditions. The following lemma will be useful in this regard.

Lemma A.1. *Let* $\alpha(k)$ *and* $\beta(k)$, $k = 0, 1, 2, \ldots$, *be bounded real valued sequences. Assume that*

$$\lim_{k\to\infty} \alpha(k) = 0$$

and

$$\sum_{k=0}^{\infty} |\alpha(k+1) - \alpha(k)| < \infty.$$

Then,

$$\lim_{n\to\infty} \sup_{n\le\ell} \sum_{k=n}^{\ell} \alpha(k)(\beta(k+1) - \beta(k)) = 0.$$

Proof: Rearrange the summation to show that

$$\begin{aligned}
&\sum_{k=n}^{\ell} \alpha(k)(\beta(k+1) - \beta(k)) \\
&\quad = \alpha(n)(\beta(n+1) - \beta(n)) + \alpha(n+1)(\beta(n+2) - \beta(n+1)) \\
&\qquad + \cdots + \alpha(\ell)(\beta(\ell+1) - \beta(\ell)) \\
&\quad = -\alpha(n)\beta(n) + \alpha(\ell)\beta(\ell+1) \\
&\qquad + \beta(n+1)(\alpha(n) - \alpha(n+1)) + \cdots + \beta(\ell)(\alpha(\ell-1) - \alpha(\ell)) \\
&\quad = -\alpha(n)\beta(n) + \alpha(\ell)\beta(\ell+1) + \sum_{k=n+1}^{\ell} \beta(k)(\alpha(k-1) - \alpha(k)).
\end{aligned}$$

The lemma follows from the assumptions on $\alpha(k)$ and $\beta(k)$. □

Note that $\alpha(k) = 1/(k+1)^{\rho}$ satisfies the assumptions of Lemma A.1 for any $0 < \rho$. Indeed, in that case $\alpha(k) - \alpha(k+1) \le \rho k^{-(\rho+1)}$ by the convexity of the function $x^{-\rho}$.

Proposition A.2 (Two time scale stochastic approximation, (Borkar, 1997)). *Consider random sequences, $X(k)$ and $Y(k)$, generated by*

$$\begin{aligned}
X(k+1) &= X(k) + a(k)(F(X(k), Y(k)) + M(k)) \\
Y(k+1) &= Y(k) + b(k)(G(X(k), Y(k)) + N(k)),
\end{aligned}$$

where

1. *$F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ are Lipschitz functions.*
2. *$M(k) \in \mathbb{R}^m$ and $N(k) \in \mathbb{R}^n$ are uniformly bounded random sequences with*

$$\begin{aligned}
\mathbb{E}[M(k) \mid X(k), Y(k)] &= 0 \\
\mathbb{E}[N(k) \mid X(k), Y(k)] &= 0.
\end{aligned}$$

3. *$a(k)$ and $b(k)$ are nonnegative numbers such that*

$$\sum_k a(k) = \infty, \quad \sum_k b(k) = \infty, \quad \sum_k a^2(k) < \infty, \quad \sum_k b^2(k) < \infty.$$

4. $a(k) = o(b(k))$.
5. $\sup_k |(X(k), Y(k))| < \infty$.
6. *For any fixed $\bar{x}$, the differential equation*

$$\dot{y}(t) = G(\bar{x}, y(t))$$

has a unique global asymptotically stable equilibrium, $\phi(\bar{x})$, with $\phi(\cdot)$ Lipschitz.
7. *The differential equation*

$$\dot{x}(t) = F(x(t), \phi(x(t)))$$

has a global asymptotically stable attractor, $\mathcal{Z}$.

Then

$$\lim_{k\to\infty} (X(k), Y(k)) \in \{(x^*, \phi(x^*)) \mid x^* \in \mathcal{Z}\}$$

almost surely.

B. Proof of Theorem 4.2

We require a more specialized result from the ODE method of stochastic approximation. Propositions A.1–A.2 provide conditions for almost sure convergence based on the global asymptotic stability of the resulting differential equations. If an equilibrium point is only locally asymptotically stable, then one can conclude that convergence to equilibrium occurs with *strictly positive* probability (see Theorem 5.4 of Benaim & Hirsch (1999)).

The relevant differential equations for smooth fictitious play with combined forecasts (24) are

$$\begin{aligned}\dot{q}_i(t) &= -q_i(t) + \beta_i(q_{-i} + \gamma(f_{-i}(t) - q_{-i}(t)))\\ \dot{f}_i(t) &= \lambda(-f_i(t) + \beta_i(q_{-i}(t) + \gamma(f_{-i}(t) - q_{-i}(t))).\end{aligned}$$

The change of variables $z_i = f_i - q_i$ results in

$$\dot{q}_i(t) = -q_i(t) + \beta_i(q_{-i}(t) + \gamma z_{-i}(t)) \tag{26a}$$

$$\frac{1}{\lambda}\dot{z}_i(t) = -z_i(t) + \beta_i(q_{-i}(t) + \gamma z_{-i}(t)) - q_i(t) - \frac{1}{\lambda}\dot{q}_i(t). \tag{26b}$$

We can compare these differential equations to those of "derivative action fictitious play". From the analysis of "derivative action fictitious play" in Shamma and Arslan (2005), the hypotheses of Theorem 4.2 assure that the dynamics

$$\dot{q}_i(t) = -q_i(t) + \beta_i(q_{-i}(t) + \gamma\lambda(q_{-i}(t) - r_{-i}(t)))$$
$$\dot{r}_i(t) = \lambda(q_i(t) - r_i(t))$$

are locally asymptotically stable at the Nash equilibrium (q_1^*, q_2^*). For these dynamics, the change of variables

$$z_i = \lambda(q_i - r_i)$$

results in

$$\dot{q}_i(t) = -q_i(t) + \beta_i(q_{-i}(t) + \gamma z_{-i}(t)) \tag{27a}$$
$$\frac{1}{\lambda}\dot{z}_i(t) = -z_i(t) + \beta_i(q_{-i}(t) + \gamma z_{-i}(t)) - q_i(t). \tag{27b}$$

We will exploit the similarity between (26) and (27) to show that local asymptotic stability of (27) implies local asymptotic stability of (26).

The relevant (reduced order) Jacobian matrix for (27) can be written as

$$\begin{pmatrix} X_1 & \gamma X_2 \\ \lambda X_1 & \lambda\gamma X_2 - \lambda I \end{pmatrix}, \tag{28}$$

where

$$X_1 = \begin{pmatrix} -I & \mathcal{N}^{\mathrm{T}}\nabla\beta_1(q_2^*)\mathcal{N} \\ \mathcal{N}^{\mathrm{T}}\nabla\beta_2(q_1^*)\mathcal{N} & -I \end{pmatrix},$$

$$X_2 = \begin{pmatrix} 0 & \mathcal{N}^{\mathrm{T}}\nabla\beta_1(q_2^*)\mathcal{N} \\ \mathcal{N}^{\mathrm{T}}\nabla\beta_2(q_1^*)\mathcal{N} & 0 \end{pmatrix},$$

and $\mathcal{N}$ is defined in (21). From Theorem 3.5 in Shamma and Arslan (2005), this matrix is stable (i.e., has negative real parts) for all sufficiently large λ as long as γ satisfies the hypotheses of Theorem 4.2.

Now let us inspect the Jacobian matrix of (26), which can be written as

$$\begin{pmatrix} X_1 & \gamma X_2 \\ (\lambda - 1)X_1 & (\lambda - 1)\gamma X_2 - \lambda I \end{pmatrix}. \tag{29}$$

An eigenvector/eigenvalue pair $(\binom{v_1}{v_2}, \mu)$ satisfies

$$X_1 v_1 + \gamma X_2 v_2 = \mu v_1,$$
$$(\lambda - 1)\mu v_1 = (\lambda + \mu)v_2.$$

Consequentially, $\binom{v_1'}{v_2'}, \mu)$ is an eigenvector/eigenvalue pair for the *perturbed* matrix

$$\begin{pmatrix} X_1 & \gamma' X_2 \\ \lambda X_1 & \lambda\gamma' X_2 - \lambda I \end{pmatrix}$$

with

$$v_1' = \frac{\lambda - 1}{\lambda} v_1, \quad v_2' = v_2,$$

and

$$\gamma' = \frac{\lambda - 1}{\lambda}\gamma.$$

For λ sufficiently large, γ' will lie in the open interval specified by Theorem 4.2. A comparison to (28) implies that the eigenvalue, μ, must have a negative real part for λ sufficiently large.

Acknowledgment Partially supported by NSERC, FQRNT, The Canada Research Chairs Program, ARO grant #W911NF-04-1-0316, and AFOSR grant #FA9550-05-1-0239.

References

Arslan, G., & Shamma, J. S. (2004). Distributed convergence to Nash equilibria with local utility measurements. In *43rd IEEE conference on decision and control* (pp. 1538–1543).

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2002). The non-stochastic multi-armed bandit problem. *SIAM Journal of Computation*, *32*, 48–77.

Benaim, M. (1999). Dynamics of stochastic approximation algorithms. In J. Azema, et al. (Eds), *Seminaire de probabilites XXXIII*, vol. 1709 (pp. 1–68). Springer-Verlag Lecture Notes in Mathematics.

Benaim, M., & Hirsch, M. W. (1999). Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, *29*, 36–72.

Benaim, M., Hofbauer, J., & Sorin, S. (2003). Stochastic approximations and differential inclusions. online: http://www.unine.ch/math/personnel/equipes/benaim/benaim_pers/bhs.pdf.

Bertsekas, D. P., & Tsitsiklis, J. N. (1989). *Parallel and distributed computation*. Prentice Hall.

Blum, L., Cucker, F., Shub, M., & Smale, S. (1996). Complexity and real computation: a manifesto. *International Journal of Bifurcation and Chaos*, *6*, 3–26.

Borkar, V. S. (1997). Stochastic approximation with two time scales. *Systems & Control Letters*, *29*(5), 291–294.

Borkar, V. S., & Meyn, S. P. (2000). The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM J. Control Optim.*, *38*(2), 447–469.

Borodin, A., & El-Yaniv, R. (1998). *Online computation and competitive analysis*. Cambridge University Press.

Crawford, V. P. (1985). Learning behavior and mixed strategy Nash equilibria. *Journal of Economic Behavior and Organization*, *6*, 69–78.

Dawid, A. P. (1985). The impossibility of inductive inference. *Journal of the American Statistical Association*, *80*, 340–341.

Durrett, R. (1991). *Probability : theory and examples*. Wadsworth.

Filar, J., & Vrieze, K. (1996). *Competitive markov decision processes*. Springer Verlag.

Foster, D. P. (2005). Personal communication in the context of binary sequences.

Foster, D. P., & Vohra, R. (1997). Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, *21*, 40–55.
Foster, D. P., & Vohra, R. (1998). Asymptotic calibration. *Biometrika*, *85*(2), 379–390.
Foster, D. P., & Vohra, R. (1999). Regret in the on-line decision problem. *Games and Economic Behavior*, *29*(1–2), 7–35.
Fudenberg, D., & Levine, D. K. (1998). *The theory of learning in games*. Cambridge, MA: MIT Press.
Fudenberg, D., & Levine, D. K. (1999). An easier way to calibrate. *Games and Economic Behavior*, *29*, 131–137.
Hart, S. (2005). Adaptive Heuristics. *Econometrica*, *73*(5), 1401–1430.
Hart, S., & Mas-Colell, A. (2001). A general class of adaptative strategies. *Journal of Economic Theory*, *98*, 26–54.
Hart, S., & Mas-Colell, A. (2003). Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, *93*(5), 1830–1836.
Hart, S., & Mas-Colell, A. (2004). Stochastic uncoupled dynamics and Nash equilibrium. Preprint, http://www.ma.huji.ac.il/˜hart/abs/uncoupl-st.html.
Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.
Kakade, S. M., & Foster, D. P. (2004). Deterministic calibration and Nash equilibrium. J. Shawe-Taylor, & Y. Singer, (Eds), *Proceedings of the 17th annual conference on learning theory* (pp. 33–48).
Khalil, H. K. (2001). *Nonlinear systems*. 3rd edn. Prentice Hall.
Konda, V. R., & Tsitsiklis, J. N. (2004). Rate of convergence of two-time-scale stochastic approximation. *Annals of Applied Probability*, *14*(2), 796–819.
Kushner, H. J., & Yin, G. G. (1997). *Stochastic approximation algorithms and applications*. Springer-Verlag.
Leslie, D. S., & Collins, E. J. (2003). Convergent multiple-timescales reinforcement learning algorithms in normal form games. *The Annals of Applied Probability*, *4*(4), 1231–1251.
Oakes, D. (1985). Self-calibrating priors do not exist. *Journal of the American Statistical Association*, *80*, 339–342.
Samuelson, L. (1997). *Evolutionary games and equilibrium selection.* Cambridge, MA: MIT Press.
Sandroni, A., Smorodinsky, R., & Vohra, R. (2003). Calibration with many checking rules. *Mathematics of Operations Research*, *28*(1), 141–153.
Shamma, J. S., & Arslan, G. (2003). A feedback stabilization approach to fictitious play. In *Proceedings of the 42nd IEEE conference on decision and control* (pp. 4140–4145).
Shamma, J. S., & Arslan, G. (2005). Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, *50*(3), 312–327.
Shapley, L. S. (1964). Some topics in two-person games. In M. Dresher, L.S. Shapley, and A.W. Tucker (Eds), *Advances in game theory* (pp. 1–29). Princeton, NJ: University Press.
Tsitsiklis, J. N. (1994). Asynchronous stochastic approximation and Q-learning. *Machine Learning*, *16*, 185–202.
Vovk, V. (1998). A game of prediction with experts advice. *Journal of Computer and Systems Sciences*, *56*(2), 153–173.
Wang, I.-J., Chong, E. K. P., & Kulkarni, S. R. (1996). Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms. *Advances in Applied Probability*, *28*(3), 784–801.
Weibull, J. W. (1995). *Evolutionary game theory.* Cambridge, MA: MIT Press.
Young, H. P. 1998. *Individual strategy and social structure.* Princeton, NJ: Princeton University Press.
Young, H. P. (2004). *Strategic learning and its limits.* Oxford University Press.
Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proc. 20th international conference on machine learning* (pp. 928–936). AAAI Press.