extended abstract

Calibrated Forecasts: Efficiency versus Universality

Gürdal Arslan
Dept of Electrical Engineering
University of Hawaii
gurdal@hawaii.edu

Shie Mannor
Dept of Electrical & Computer Engineering
McGill University
shie@ece.mcgill.ca

Jeff S. Shamma
Dept of Mechanical and Electrical Engineering
University of California Los Angeles
shamma@ucla.edu

1 Overview

One approach to learning/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 special classes of games, player strategies converge to a Nash equilibrium, but in general, the limiting behavior need not exhibit convergence (e.g., [FL98]).

Placing more stringent requirements on the forecasts can result in stronger convergence properties for general games. In particular, if players use "calibrated" forecasts [FV97], 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 [KF04]. See [SSV03] and references therein for further discussion on calibrated forecasting as well as its generalizations.

A drawback of calibrated forecasts is the computational requirement. In particular, existing methods of computing calibrated forecasts require a state space that evolves over a discretized grid of points extracted from a probability simplex of appropriate dimension. Furthermore, the asymptotic convergence results cited earlier require a progressive refinement of this grid. As a result, calibration for a multi-move opponent can be a considerable computational burden.

This talk presents some work in progress that explores 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 possible consequences of tracking forecasts in repeated matrix games.

2 Set-Up

The analysis presented here is motivated by stochastic discrete-time iterations but will be entirely for deterministic continuous-time systems. The Ordinary Differential Equation (ODE) method of stochastic approximation establishes a connection between the discrete-time and continuous-time settings. At this point, we only make formal connections between the two. Incorporating methods from [BHS03, Bor97] would provide an analytical basis.

At each stage, $n = 0, 1, 2, ...$, there is a finite-alphabet outcome, $x(n)$, represented as a simplex vertex. There is also a forecast, $f(n) \in \Delta$, where Δ is the probability simplex. The information available to make the forecast $f(n)$ includes the outcomes $\{x(0), x(1), ..., x(n-1)\}$, but does not include $x(n)$.

Following [KF04], let $w : \Delta \to R^+$ be Lipschitz continuous. Define the calibration error with respect to w as

$$e_w(N) = \frac{1}{N}\sum_{n=1}^{N} w(f(n))(x(n) - f(n)).$$

A forecast is "weakly calibrated" if for all sequences, $x(\cdot)$, and all test functions, $w(\cdot)$,

$$\lim_{N\to\infty} e_w(N) = 0.$$

Reference [KF04] presents a universal algorithm that constructs a weakly calibrated forecast for arbitrary sequences.

For a given test function, $w(t)$, we can write the calibration error recursively as the running average

$$e_w(n+1) = e_w(n) + \frac{1}{n+1}\Big(w(f(n))(x(n) - f(n)) - e_w(n)\Big).$$

The continuous-time calibration error differential equation suggested by the ODE methods of stochastic approximation is

$$\dot{e}_w(t) = -e_w(t) + w(f(t))(p(t) - f(t)).$$

Here, $p(t) \in \Delta$ may be identified with the underlying probability density function behind the outcome $x(t)$. As before, it is desired that

$$\lim_{t\to\infty} e_w(t) = 0$$

for all $p(\cdot)$. The information restriction in continuous time is that the forecast, $f(\cdot)$, at time t, cannot depend on the value of the input, $p(\cdot)$, at time t. The main concern is that any candidate continuous time forecasting algorithm must allow a discrete-time "translation" of the continuous-time forecast. This will be made clear in the coming definition of the forecast.

3 Tracking Forecast

Typically, universally calibrated forecasts depend explicitly on the calibration errors associated with a collection of test functions. This collection of test functions constitutes the discretization typically used in calibration.

We now present a forecast that does not depend on any specific e_w. Consider the following "tracking forecast",

$$\dot{f}(t) = \lambda(p(t) - f(t)),$$

where $\lambda \gg 1$. The terminology is due to the following property. If $|\dot{p}|$ is uniformly bounded, i.e.,

$$|\dot{p}(t)| \leq \dot{p}_{\max},$$

then $f(t)$ "tracks" $p(t)$ in the sense that the tracking error, $p(t) - f(t)$, asymptotically satisfies

$$|p(t) - f(t)| \leq \frac{1}{\lambda}|\dot{p}_{\max}|.$$

A discrete-time implementation is

$$f(n+1) = f(n) + \frac{\lambda}{n+1}\Big(x(n) - f(n)\Big).$$

One issue is quantifying how large $\lambda \gg 1$ should be. A conjecture is that this may be sidestepped via the implementation[1]

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

where $0 < \rho < 1$.

3.1 Bounded rate sequences

The tracking property of this forecast results in an "approximate" calibration property for bounded rate sequences, i.e., sequences that satisfy $|\dot{p}(t)| \leq \dot{p}_{\max}$. One example of a bounded rate sequence is the actions of a player using smooth fictitious play (cf., section 4.1). For a given test function, $w(\cdot)$, define $w_{\max} = \max_{f\in\Delta} w(f)$. Then for bounded rate sequences,

$$\limsup_{t\to\infty} |e_w(t)| \leq \frac{1}{\lambda}\dot{p}_{\max} w_{\max}.$$

3.2 Binary Sequences

It is possible to remove the bounded rate condition for binary sequences. In the case of binary sequences, there is single degree of freedom in that $f(t) = \big(\, \tilde{f}(t) \quad 1 - \tilde{f}(t)\,\big)$ and $p(t) = (\, \tilde{p}(t) \quad 1 - \tilde{p}(t)\,)$.

Now consider the calibration error defined by

$$\dot{e}_w(t) = -e_w(t) + w(\tilde{f}(t))(\tilde{p}(t) - \tilde{f}(t)).$$

Let us assume for convenience that the test function, $w(\cdot)$, is differentiable and satisfies $w(\tilde{f}) = \nabla v(\tilde{f})$ for some differentiable function, $v(\cdot)$, bounded by $v_{\max}$. One can show that asymptotically,

$$\limsup |e_w(t)| \leq \frac{c}{\lambda} v_{\max},$$

for some constant, c.

4 Application to Learning in Games

Now consider a continuous-time repeated two-player matrix game. The dynamics (e.g., [FL98]) are as follows:

$$\dot{q}_i = -q_i + p_i,$$

where the $q_i \in \Delta$ are identified with the empirical frequencies, and the $p_i \in \Delta$ are identified with the players' current strategies. These equations are coupled since each player's strategy will be a function of the opponent's actions.

Let the utility function of the i^{th} player be characterized by the matrix M_i. The *smoothed* best response function, $\beta_i : \Delta \to \Delta$, is defined as

$$\beta_i(q) = \arg\max_{s\in\Delta} s^\top M_i q - \tau s^\top \log(s),$$

where $\tau > 0$ is a small coefficient and $\log(\cdot)$ is component-wise. Let the subscript "$-i$" denote the opponent of player i. If each player uses a smoothed best response to a forecast of the other player, then the above dynamics become

$$\dot{q}_i = -q_i + \beta_i(f_{-i}).$$

[1] Suggested by Dean Foster (personal communication) as an approach to calibrate binary sequences.

4.1 Tracking Forecast: Single Player

Suppose that player 2's strategy is as in smooth fictitious play, i.e., a smoothed best response to empirical frequencies, whereas player 1's strategy is a best response to a tracking forecast. The resulting dynamics are

$$\dot{q}_1 = -q_1 + \beta_1(f_2), \quad \dot{f}_2 = \lambda(p_2 - f_2)$$
$$\dot{q}_2 = -q_2 + \beta_2(q_1),$$

where $p_2 = \beta_2(q_1)$.

Note that for a fixed smoothing parameter, τ, the strategy of player 2 constitutes a bounded rate sequence, i.e., $\dot{p}_2(t)$ is bounded. Accordingly, for sufficiently large λ, $f_2 \approx p_2$, and so player 1 approximately plays the best response to player 2's strategy. In this case,

$$\dot{q}_1 \approx -q_1 + \beta_1(\beta_2(q_1)).$$

These are the same dynamics that emerge in [LC03], where different players adapt at different time-scales. In the present case, the two-time-scale behavior stems from one player using a more sophisticated forecast than the other.

4.2 Tracking Forecast: Both Players

In case both players use smoothed best responses to tracking forecasts, the resulting dynamics are

$$\dot{q}_i = -q_i + \beta_i(f_{-i}), \quad \dot{f}_i = \lambda(\beta_i(f_{-i}) - f_i).$$

Note that the $\dot{f}_i$ dynamics are the usual smooth fictitious play. For large λ, these equations exhibit two-time-scale behavior. In particular, if the $\dot{f}_i$ dynamics are asymptotically oscillatory, the methods of [TMN03] show that the empirical frequencies approach the running average of the best responses. More precisely, suppose that

$$f_i^* = \lim_{T\to\infty} \frac{1}{T}\int_0^T \beta_i(f_{-i}(\tau))d\tau,$$

then for large λ, $q_i(t) \approx f_i^*$ asymptotically.

4.3 Weighted Tracking Forecasts

Another possibility is where both players use smoothed best responses to a combination of empirical frequencies and tracking forecasts, e.g.,

$$\dot{q}_i = -q_i + \beta_i(\gamma q_{-i} + (1-\gamma)f_{-i})$$
$$\dot{f}_i = \lambda(\beta_i(f_{-i}) - f_i).$$

For $\gamma = 1$, this is standard smooth fictitious play, and for $\gamma = 0$, this is both players using a tracking forecast best response. It is possible that $\gamma \in (0,1)$ may result in completely different asymptotic behaviors. This set-up is reminiscent of "derivative action" fictitious play, introduced in [SA03, SA05], where it was shown that for some games for which smooth fictitious play does not converge, there exists a suitable modification such that playing the best response to an "anticipated" empirical frequency can converge (locally) to a Nash equilibrium. It is likely that a similar stability analysis is applicable to the above dynamics.

4.4 Illustrative Simulations

The talk will present several simulations that illustrate the construction of tracking forecasts and their effect in learning in games.

References

[BHS03] M. Benaim, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. 2003. online: http://www.unine.ch/math/personnel/equipes/benaim/benaim_pers/bhs.pdf.

[Bor97] V.S. Borkar. Stochastic approximation with two time scales. *Systems & Control Letters*, **29**(5):291–294, 1997.

[FL98] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.

[FV97] D.P. Foster and R.V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, **21**:40–55, 1997.

[KF04] S.M. Kakade and D.P. Foster. Deterministic calibration and Nash equilibrium. 2004. online: http://www.cis.upenn.edu/~skakade/papers/gt/calibration.pdf.

[LC03] D.S. Leslie and E.J. Collins. Convergent multiple-timescales reinforcement learning algorithms in normal form games. *The Annals of Applied Probability*, **4**(4):1231–1251, 2003.

[SA03] J.S. Shamma and G. Arslan. A feedback stabilization approach to fictitious play. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 4140–4145, 2003.

[SA05] J.S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, **50**(3):312–327, 2005.

[SSV03] A. Sandroni, R. Smorodinsky, and R. Vohra. Calibration with many checking rules. *Mathematics of Operations Research*, **28**(1):141–153, 2003.

[TMN03] A.R. Teel, L. Moreau, and D. Nesic. A unified framework for input-to-state stability in systems with two time scales. *IEEE Transactions on Automatic Control*, **48**(9):1526–1544, 2003.