TIME AVERAGE REPLICATOR AND BEST REPLY DYNAMICS

JOSEF HOFBAUER, SYLVAIN SORIN AND YANNICK VIOSSAT

Abstract. Using an explicit representation in terms of the logit map we show, in a unilateral framework, that the time average of the replicator dynamics is a perturbed solution, hence an asymptotic pseudo-trajectory of the best reply dynamics.

1. Presentation

The two prime examples of deterministic evolutionary game dynamics are the *replicator dynamics* (RD) and the *best response dynamics* (BRD). In the framework of a symmetric 2 person game with $K \times K$ payoff matrix A played within a single population, the *replicator equation* is given by

$$\dot{x}_t^k = x_t^k \left(e^k A x_t - x_t A x_t \right), \quad k \in K \qquad (RD) \tag{1}$$

with x_t^k denoting the frequency of strategy k at time t. It was introduced in [22] as the basic selection dynamics for the evolutionary games of Maynard Smith [19], see [15] for a summary. The interpretation is that in an infinite population of replicating players, the per capita growth rate of the frequencies of pure strategies is linearly related to their payoffs.

The *best reply dynamics*

$$\dot{z}_t \in BR(z_t) - z_t, \qquad t \geq 0 \qquad (BRD) \tag{2}$$

was introduced in [8] and studied further in [12], [15], [4]. Here $BR(z)$ denotes the set of all pure and mixed best replies to the strategy profile z. The interpretation is that in an infinite population of players, in each small time interval, a small fraction of players revises their strategies and changes to a best reply against the present population distribution. It is the prototype of a population model of rational (but myopic) behaviour.

(BRD) is closely related to the fictitious play process introduced by Brown [3]. In the framework of a bimatrix game and continuous time this is given, for all $t \geq 0$, by $X_t = \frac{1}{t}\int_0^t x_s ds$, $Y_t = \frac{1}{t}\int_0^t y_s ds$ such that $x_t \in BR_1(Y_t)$, $y_t \in BR_2(X_t)$. This implies that $Z_t = (X_t, Y_t)$ satisfies

Date: June 2007.

the *continuous fictitious play* equation

$$\dot{Z}_t \in \frac{1}{t}\left(BR(Z_t) - Z_t\right), \quad t > 0 \qquad (CFP) \tag{3}$$

which is equivalent to (BRD) via a change in time $Z_{e^s} = z_s$.

Despite the different interpretation and the different dynamic character there are amazing similarities in the long run behaviour of these two dynamics, that have been summarized in the following heuristic principle, see [7] and [12].

For many games, the long run behaviour ($t \to \infty$) of the time averages $X_t = \frac{1}{t}\int_0^t x_s ds$ of the trajectories x_t of the replicator equation is the same as for the BR trajectories.

In this paper we will provide a rigorous statement that largely explains this heuristics. We show that for any interior solution of (RD), for every $t \geq 0$, x_t is an approximate best reply against X_t and the approximation gets better as $t \to \infty$. This implies that X_t is an asymptotic pseudo trajectory of (BRD) and hence the limit set of X_t has the same properties as a limit set of a true orbit of (BRD), i.e. it is invariant and internally chain transitive under (BRD). The main tool to prove this is via the logit map which is a canonical smoothing of the best response correspondence. We show that x_t equals the logit approximation at X_t with error rate $\frac{1}{t}$.

2. Unilateral processes

The model will be in the framework of an N-person game but we consider the dynamics for one player, without hypotheses on the behavior of the others. Hence, from the point of view of this player, he is facing a (measurable) vector outcome process $\mathcal{U} = \{U_t, t \geq 0\}$, with values in the cube $C = [-c, c]^K$ where K is his action's set and c is some positive constant. U_t^k is the payoff at time t if k is the action at that time. The cumulative vector outcome up to stage t is thus $S_t = \int_0^t U_s ds$ and its time average is denoted $\bar{U}_t = \frac{1}{t} S_t$.
br denotes the (payoff based) best reply correspondence from C to the simplex Δ on K, defined by

$$\mathbf{br}(U) = \{x \in \Delta; \langle x, U\rangle = \max_{y \in \Delta}\langle y, U\rangle\}$$

The $\mathcal{U}$-fictitious play process (FPP) is defined on Δ by

$$\dot{X}_t \in \frac{1}{t}[\mathbf{br}(\bar{U}_t) - X_t] \tag{4}$$

The $\mathcal{U}$-replicator process (RP) is specified by the following equation on Δ:

$$\dot{x}_t^k = x_t^k[U_t^k - \langle x_t, U_t\rangle], \qquad k \in K. \tag{5}$$

Explicitely, in the framework of a N-player game with payoff for player 1 defined by a function G from $\prod_{i\in N} K^i$ to $I\!R$, with $X^i = \Delta(K^i)$, one has $U^k_t = G(k, x^{-1}_t)$.
If all the players follow a (payoff based) fictitious play dynamics, each time average strategy satisfies (4). For $N = 2$ this is (CFP).
If all the players follow the replicator dynamics then (5) is the replicator dynamics equation.

3. Logit rule and perturbed best reply

Define a map L from $I\!R^K$ to Δ by

$$L^k(V) = \frac{\exp V^k}{\sum_j \exp V^j}. \tag{6}$$

Given $\eta > 0$, let $[\mathbf{br}]^\eta$ be the correspondence from C to Δ with graph being the η-neighborhood for the uniform norm of the graph of $\mathbf{br}$.
The L map and the $\mathbf{br}$ correspondence are related as follows:

Proposition 3.1. *For any $U \in C$ and $\varepsilon > 0$*

$$L(U/\varepsilon) \in [\mathbf{br}]^{\eta(\varepsilon)}(U)$$

with $\eta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Proof
Given $\eta > 0$, define the correspondence D^η from C to Δ by

$$D^\eta(U) = \{x \in \Delta; (U^k + \eta < \max_{j\in K} U^j \Rightarrow x^k \leq \eta), \forall k \in K\}.$$

and note that $D^\eta \subset [\mathbf{br}]^\eta$.
Let $\varepsilon(\eta)$ satify

$$\exp(-\eta/\varepsilon(\eta)) = \eta.$$

By definition of L, one has for all (j, k)

$$L^k(U/\varepsilon) = \frac{\exp((U^k - U^j)/\varepsilon)}{1 + \sum_{\ell\neq j} \exp((U^\ell - U^j)/\varepsilon)}$$

and it follows that $\varepsilon \leq \varepsilon(\eta)$ implies

$$L(U/\varepsilon) \in D^\eta(U).$$

Define finally $\eta(\varepsilon)$ to be the inverse function of ε to get the result. ∎

Remarks
L is also given by

$$L(V) = \mathrm{argmax}_{x\in\Delta}\{\langle x, V\rangle - \sum_k x^k \log x^k\}.$$

Hence introducing the (payoff based) perturbed best reply $\mathbf{br}^{\varepsilon}$ from C to Δ defined by

$$\mathbf{br}^{\varepsilon}(U) = \operatorname{argmax}_{x\in\Delta}\{\langle x, U\rangle - \varepsilon \sum_k x^k \log x^k\}$$

one has

$$L(U/\varepsilon) = \mathbf{br}^{\varepsilon}(U)$$

and the previous property also follows from Berge's maximum theorem. The map $\mathbf{br}^{\varepsilon}$ is the logit approximation.

4. Explicit representation of the replicator process

4.1. **CEW.** The following procedure has been introduced in discrete time in the framework of on-line algorithms under the name "multiplicative weight algorithm" ([5], [18]). We use here the name (CEW) (continuous exponential weight) for the process defined, given $\mathcal{U}$, by

$$x_t = L(\int_0^t U_s ds).$$

4.2. **Properties of CEW.** The main property of (CEW) that will be used is that it provides an explicit solution of (RD).

Proposition 4.1. *(CEW) satisfies (RP).*

Proof
Straightforward computations lead to

$$\dot{x}_t^k = x_t^k U_t^k - x_t^k \sum_j \frac{U_t^j \exp \int_0^t U_v^j dv}{\sum_j \exp \int_0^t U_v^j dv}$$

which is

$$\dot{x}_t^k = x_t^k [U_t^k - \langle x_t, U_t\rangle]$$

hence gives the previous (RP) equation (5). ∎

Note that (CEW) specifies the solution starting from the barycenter of Δ.

The link with the best reply correspondence is the following.

Proposition 4.2. *CEW satisfies*

$$x_t \in [\mathbf{br}]^{\delta(t)}(\bar{U}_t)$$

with $\delta(t) \to 0$ as $t \to \infty$.

Proof
Write

$$x_t = L(\int_0^t U_s ds) = L(t\,\bar{U}_t)$$
$$= L(U/\varepsilon) \in [\mathbf{br}]^{\eta(\varepsilon)}(U)$$

with $U = \bar{U}_t$ and $\varepsilon = 1/t$, by Proposition 3.1. Let $\delta(t) = \eta(1/t)$. ∎

4.3. **Time average.** We describe here the consequences for the time average process.
Define

$$X_t = \frac{1}{t}\int_0^t x_s ds$$

Proposition 4.3. *If x_t follows (CEW) then X_t satisfies*

$$\dot{X}_t \in \frac{1}{t}([\mathbf{br}]^{\delta(t)}(\bar{U}_t) - X_t)$$

with $\delta(t) \to 0$ as $t \to \infty$.

Proof
Since

$$\dot{X}_t \in \frac{1}{t}(x_t - X_t)$$

the result follows from the previous section. ∎

4.4. **Initial conditions.** The solution of (RP) starting from $x_0 \in \operatorname{int}\Delta$ is given by $x_t = L(U_0 + \int_0^t U_s ds)$ with $U_0^k = \log x_0^k$. The average process satisfies

$$\dot{X}_t \in \frac{1}{t}([\mathbf{br}]^{\delta(t)}(U_0/t + \bar{U}_t) - X_t). \tag{7}$$

which can be written as

$$\dot{X}_t \in \frac{1}{t}([\mathbf{br}]^{\alpha(t)}(\bar{U}_t) - X_t). \tag{8}$$

with $\alpha(t) \to 0$ as $t \to \infty$.

5. Consequences for games

Consider a 2 person (bimatrix) game (A, B). If the game is symmetric this gives rise to the single population replicator dynamics (RD) and best reply dynamics (BRD) as defined in section 1. Otherwise, we consider the two population replicator dynamics

$$\dot{x}_t^k = x_t^k \left(e^k A y_t - x_t A y_t\right), \quad k \in K_1 \tag{9}$$

$$\dot{y}_t^k = y_t^k \left(x_t B e^k - x_t B y_t\right), \quad k \in K_2$$

and the corresponding BR dynamics as in (3).

Let M be the state space (a simplex Δ or a product of simplices $\Delta_1 \times \Delta_2$).
We now use the previous results with the $\mathcal{U}$ process being defined by $U_t = Ay_t$ for player 1, hence $\bar{U}_t = AY_t$. Note that $\mathbf{br}(AY) = BR_1(Y)$.

Proposition 5.1. *The limit set of every replicator time average process X_t starting from an initial point $x_0 \in \operatorname{int} M$ is a closed subset of M which is invariant and internally chain transitive (ICT) under (BRD).*

Proof
Equation (8) implies that X_t satisfies a perturbed version of (CFP) hence X_{e^t} is a perturbed solution to the differential inclusion (BRD), according to Definition II in Benaim et al [1]. Now apply Theorem 3.6 of that paper. ∎

In particular this implies:

Proposition 5.2. *Let $\mathcal{A}$ be the global attractor (i.e., the maximal invariant set) of (BRD). Then the limit set of every replicator time average process X_t is a subset of $\mathcal{A}$.*

Some consequences are:

If the time averages of an interior orbit of the replicator dynamics converge then the limit is a Nash equilibrium. Indeed, by Proposition 5.1 the limit is a singleton invariant set of the (BRD), and hence a Nash equilibrium. As a consequence one obtains: If an interior orbit of the replicator dynamics converges then the limit is a Nash equilibrium. (For a direct proof see [15, Theorem 7.2.1].)

For 2 person zero-sum games, the global attractor of (BR) equals the (convex) set of NE (this is a strengthened version of Brown and Robinson's convergence result for fictitious play, due to [16]). Therefore, by Proposition 5.2 the time averages of (RD) converge to the set of NE as well. For a direct proof (in the special case when an interior equilibrium exists) see [15]. Note that orbits of (RD) in general do not converge, but oscillate around the set of NE, as in the matching pennies game.

In potential games the only ICT sets of (BRD) are (connected subsets of) components of NE, see [1, Theorem 5.5]. Hence, by Proposition 5.1 time averages of (RD) converge to such components. In fact, orbits of (RD) themselves converge.

For games with a strictly dominated strategy, the global attractor of (BRD) is contained in a face of M with no weight on this strategy. Hence time averages of (RD) converge to this face, i.e., the strictly dominated strategy is eliminated on the average. In fact, the frequency of a strictly dominated strategy under (RD) vanishes, see [15, Theorem 8.3.2].

How do our general results compare with the examples in Gaunersdorfer and Hofbauer [7]? In the rock–scissors–paper game with payoff matrix $A = \begin{pmatrix} 0 & -b_2 & a_3 \\ a_1 & 0 & -b_3 \\ -b_1 & a_2 & 0 \end{pmatrix}$, $a_i, b_i > 0$, there are two cases. If $a_1a_2a_3 \geq b_1b_2b_3$ then the NE $\hat{x}$ is the global attractor of (BRD). Hence, Proposition 5.2 implies that the time averages of (RD) converge to $\hat{x}$ as well. Note that in case of equality, $a_1a_2a_3 = b_1b_2b_3$ the orbits of (RD) oscillate around $\hat{x}$ and hence do not converge, only their time averages do. If $a_1a_2a_3 < b_1b_2b_3$ then there are two ICT sets under (BRD), $\hat{x}$

and the Shapley triangle. Then Proposition 5.1 implies that time averages of (RD) converge to one of these, whereas the limit set of all non constant orbits is the boundary of M. However, our results do not show that for most orbits, the time average converges to the Shapley triangle. This still requires a more direct argument, as in [7].

If $\hat{x} \in \operatorname{int} M$ is the global attractor of (BRD), then time averages of (RD) converge to $\hat{x}$. In the literature on (RD) the following sufficient condition for the convergence of its time averages is known: If the (RD) is *permanent*, i.e., all interior orbits have their ω–limit set contained in a compact set in $\operatorname{int} M$, then the time averages of (RD) converge to the unique interior equilibrium $\hat{x}$. (See [15, Theorem 13.5.1].) It is tempting to conjecture that, for generic payoff matrices A, permanence of (RD) is equivalent to the property of (BRD) that its global attractor equals the unique interior equilibrium.

6. External consistency

6.1. **Definition.** A procedure satisfies external consistency if for each process $\mathcal{U} \in I\!R^K$, it produces a process $x_t \in \Delta$, such that for all k

$$\int_0^t [U_s^k - \langle x_s, U_s \rangle] ds \leq C_t = o(t)$$

This property says that the (expected) average payoff induced by x_t along the play is asymptotically not less than the payoff obtained by any fixed choice $k \in K$, see [6].

6.2. **CEW.** We recall this result from [21], where the aim was to compare discrete and continuous time procedures.

Proposition 6.1. *(CEW) satifies external consistency.*

Proof

Define $W_t = \sum \exp S_t^k$. Then

$$\dot{W}_t = \sum_k \exp(S_t^k) U_t^k = \sum_k W_t \; x_t^k \; U_t^k = \langle x_t, U_t \rangle W_t.$$

Hence

$$W_t = W_0 \exp\left(\int_0^t \langle x_s, U_s \rangle ds\right).$$

Thus, $W_t \geq \exp(S_t^k)$ for every k, implies:

$$\int_0^t \langle x_s, U_s \rangle ds \geq \int_0^t U_s^k ds - \log W_0.$$

∎

6.3. **RP.** In fact a direct and more simple proof is available, see [13]:

Proposition 6.2. (RP) *satifies external consistency.*

Proof

By integrating equation (5), one obtains, on the support of x_0:

$$\int_0^t [U_s^k - \langle x_s, U_s\rangle]ds = \int_0^t \frac{\dot{x}_s^k}{x_s^k} ds = \log(\frac{x_t^k}{x_0^k}) \leq -\log x_0^k.$$

∎

Remark

The previous proof shows in fact more: for any accumulation point $\bar{x}$ of x_t, one component $\bar{x}^k$ will be positive hence the corresponding asymptotic average difference in payoffs will be 0.
Back to a game framework this implies that if player 1 follows (RP) the set of accumulation points of the empirical correlated distribution process will belong to her reduced Hannan set, see [6], [9], [10]:

$$\bar{H}^1 = \{\theta \in \Delta(S); G^1(k, \theta^{-1}) \leq G^1(\theta), \forall k \in S^1, \text{with equality for one component}\}.$$

The example due to Viossat [23] of a game where the limit set for the replicator dynamics is disjoint from the unique correlated equilibrium shows that (RP) does not satisfy internal consistency.

7. Comments

We can now compare several processes in the spirit of (payoff based) fictitious play.
The original fictitious play process (I) is defined by

$$x_t \in \mathbf{br}(\bar{U}_t)$$

The corresponding time average satisfies (CFP).
With a smooth best reply process (see [17]) one has (II)

$$x_t = \mathbf{br}^{\varepsilon}(\bar{U}_t)$$

and the corresponding time average satisfies a smooth fictitious play process.
Finally the replicator process (III) satisfies

$$x_t = \mathbf{br}^{1/t}(\bar{U}_t)$$

and the time average follows a time dependent perturbation of the fictitious play process.
While in (I), the process x_t follows exactly the best reply correspondence, the induced average X_t does not have good unilateral properties.
One the other hand for (II), X_t satisfies a weak form of external consistency, with an error term $\alpha(\varepsilon)$ vanishing with ε ([6], [2]).
In contrast, (III) satisfies exact external consistency due to a both smooth and time dependent approximation of **br**.

References

[1] Benaïm M., J. Hofbauer and S. Sorin (2005) Stochastic approximations and differential inclusions, *SIAM J. Opt. and Control*, **44**, 328-348.

[2] Benaïm M., J. Hofbauer and S. Sorin (2006) Stochastic approximations and differential inclusions. Part II: applications, *Mathematics of Operations Research*, **31**, 673-695.

[3] Brown G. W.(1951) Iterative solution of games by fictitious play, in Koopmans T.C. (ed.) *Activity Analysis of Production and Allocation* , Wiley, 374-376.

[4] Cressman, R. (2003) *Evolutionary Dynamics and Extensive Form Games*, M.I.T. Press.

[5] Freund Y. and R.E. Schapire (1999) Adaptive game playing using multiplicative weights, *Games and Economic Behavior*, **29**, 79-103.

[6] Fudenberg D. and D. K. Levine (1995) Consistency and cautious fictitious play, *Journal of Economic Dynamics and Control*, **19**, 1065-1089.

[7] Gaunersdorfer A. and J. Hofbauer (1995) Fictitious play, Shapley polygons and the replicator equation, *Games and Economic Behavior*, **11**, 279-303.

[8] Gilboa I. and A. Matsui (1991) Social stability and equilibrium, *Econometrica*, **59**, 859-867.

[9] Hannan J. (1957) Approximation to Bayes risk in repeated plays, *Contributions to the Theory of Games, III*, Drescher M., A.W. Tucker and P. Wolfe eds., Princeton University Press, 97-139.

[10] Hart S. (2005) Adaptive heuristics, *Econometrica*, **73**, 1401-1430.

[11] Hart S. and A. Mas-Colell (2000) A simple adaptive procedure leading to correlated equilibria, *Econometrica*, **68**, 1127-1150.

[12] Hofbauer J. (1995) Stability for the best response dynamics, mimeo.

[13] Hofbauer J. (2004) Time averages of the replicator dynamics and correlated equilibria, preprint.

[14] Hofbauer J. and W. H. Sandholm (2002) On the global convergence of stochastic fictitious play, *Econometrica*, **70**, 2265-2294.

[15] Hofbauer J. and K. Sigmund (1998) *Evolutionary Games and Population Dynamics*, Cambridge U.P.

[16] Hofbauer J. and S. Sorin (2006) Best response dynamics for continuous zero-sum games, *Discrete and Continuous Dynamical Systems-series B*, **6**, 215-224.

[17] Hopkins E. (2002) Two competing models of how people learn in games, *Econometrica*, **70**, 2141-2166.

[18] Littlestone N. and M.K. Warmuth (1994) The weighted majority algorithm, *Information and Computation*, **108**, 212-261.

[19] Maynard Smith J. (1982) *Evolution and the Theory of Games*, Cambridge U.P.

[20] Robinson J. (1951) An iterative method of solving a game, *Annals of Mathematics*, **54**, 296-301.

[21] Sorin S. (2007) Exponential weight algorithm in continuous time, *Mathematical Programming*, Ser. B , DOI: 10.1007/s10107-007-0111-y .

[22] Taylor P. and L. Jonker (1978) Evolutionary stable strategies and game dynamics, *Mathematical Biosciences*, **40**, 145-156.

[23] Viossat Y. (2007) The replicator dynamics does not lead to correlated equilibria, *Games and Economic Behavior*, **59**, 397-407

Department of Mathematics, University Vienna, Nordbergstrasse 15, A-1090 Vienna, Austria

Equipe Combinatoire et Optimisation, UFR 929, Université P. et M. Curie - Paris 6, 175 Rue du Chevaleret, 75013 Paris and Laboratoire d'Econométrie, Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France

CEREMADE, Université Paris-Dauphine, Place du Marchal de Lattre de Tassigny, 75775 Paris, France

E-mail address: Josef.Hofbauer@univie.ac.at, sorin@math.jussieu.fr, viossat@ceremade.dauphine.fr