Math. Program., Ser. B
DOI 10.1007/s10107-007-0111-y

FULL LENGTH PAPER

Exponential weight algorithm in continuous time

Sylvain Sorin

Received: 7 May 2005 / Accepted: 25 July 2006

Abstract The exponential weight algorithm has been introduced in the framework of discrete time on-line problems. Given an observed process $\{X_m\}_{m=1,2,\ldots}$ the input at stage $m+1$ is an exponential function of the sum $S_m = \sum_{\ell=1}^{m} X_\ell$. We define the analog algorithm for a continuous time process X_t and prove similar properties in terms of external or internal consistency. We then deduce results for discrete time from their counterpart in continuous time. Finally we compare this approach to another continuous time approximation of a discrete time exponential algorithm based on the average sum S_m/m.

Keywords Exponential weight algorithm · Continuous time · Consistency

Mathematics Subject Classification (2000) 68W · 91A · 90C

1 Presentation

We consider the exponential weight algorithm and its consistency properties in the basic prediction problem for individual sequences. The purpose of the paper is to define a continuous time version of this algorithm and to prove that it satisfies analogous consistency properties. In fact this proofs are rather short and one of the main advantages of this approach is to deduce independently discrete time results from their continuous

S. Sorin
Equipe Combinatoire et Optimisation, UFR 929, Université P. et M. Curie-Paris 6,
175 Rue du Chevaleret, 75013 Paris, France

S. Sorin (✉)
Laboratoire d'Econométrie, Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France
e-mail: sorin@math.jussieu.fr

time counterpart. This interface between discrete time and continuous time processes has been very effective for a long period, see e.g. [6]. However the approach is quite different from the usual (asymptotic) approximation through differential equations (or inclusions), see e.g. [3,4] for related examples. This is due to the choice of the state variable of the process and is discussed in the last Sect. 6.

The model is as follows: $\{X_n\}$ denotes a sequence of vectors in $[0, 1]^K$. At each stage n, a predictor having observed the past realizations $X_1, \ldots, X_{n-1}$, chooses a component k_n in K. The corresponding outcome is $x_n = X_n^{k_n}$. Denote the past history $(X_1, k_1, \ldots, X_{n-1}, k_{n-1})$ by h_{n-1}, and the induced σ-algebra by $\mathcal{H}_{n-1}$. An algorithm or a strategy σ in the prediction problem is specified by the choice of $p_n(h_{n-1}) \in \Delta(K)$ (the simplex of $\mathbb{R}^K$), which is the law of k_n given the past history h_{n-1}. Note that the law of X_n may also depend on h_{n-1}.
Given a sequence $\{u_m\}$, let $\bar{u}_n = \frac{1}{n}\sum_{m=1}^n u_m$ denote the average of the n first terms. The average external regret evaluation at stage n is the vector $r_n = \{r_n^k\}_{k\in K}$ defined by:

$$r_n^k = \bar{X}_n^k - \bar{x}_n.$$

It compares the actual (average) payoff to each payoff corresponding to a constant component choice, see [8,9,11].

Definition A strategy σ satisfies external consistency if, for every process $\{X_m\}$:

$$\max_{k\in K} r_n^{k+} \longrightarrow 0 \text{ a.s., as} \quad n \longrightarrow \infty.$$

Remarks (1) In the framework of a repeated finite I-person game defined by action sets $J^i, i \in I$, let $M : J = \prod_i J^i \to \mathbb{R}$ be the payoff function of player 1, the predictor. Here, $K = J^1$ and X_n is the vector $M(., j_n^{-1}) \in \mathbb{R}^K$ corresponding to the payoff induced by the profile of actions j_n^i by each player $i, i \neq 1$ at stage n. Usually M is known, and j_n^{-1} announced to player 1 who then knows the vector X_n. In the current situation one does not assume M known by player 1 (not even $J^i, i \neq 1$), but only his own action set K, and the fact that all payoffs belong to $[0, 1]$: he is just told the vector X_n.

(2) To obtain robust results no assumption is made on the sequence $\{X_n\}$: the predictor is not "Bayesian".

The content of the paper is the following. Section 2 recalls the main results concerning the discrete time exponential weight algorithm. Section 3 introduces the continuous counterpart and its properties. Section 4 deduces, from the continuous time results, the discrete time analogs: this gives an alternative simple proof of properties of the exponentional weight algorithm. In Sect. 5 two extensions are described: restriction on the information and internal consistency. Finally Sect. 6 discusses the relation between discrete and continuous time, for algorithms based on the average sum S_n/n.

2 The exponential weight algorithm: discrete time

Notation Given a vector $x \neq 0$ in $\mathbb{R}_+^K$, $\ell[x]$ denotes its normalization in the simplex $\Delta(K)$:

$$\ell[x]^k = \frac{x^k}{\sum_{j=1}^K x^j}.$$

Definition The discrete exponential weight (EW) algorithm (see e.g. [1,10,17]) with positive parameter A, $EW(A)$, is defined by $p_{n+1}(h_n) = p_{n+1} = \ell[\{p_n^k e^{AX_n^k}\}_k]$ or equivalently

$$p_{n+1}^k = \frac{exp(A\sum_{m=1}^n X_m^k)}{\sum_j exp(A\sum_{m=1}^n X_m^j)}.$$

Recall that p_{n+1} describes the law of the random choice k_{n+1}.
An alternative definition is $EW^*(\alpha)$,where α is a positive parameter, and $p_{n+1} = \ell[\{(1+\alpha)^{S_n^k}\}_k]$ with $S_n = \sum_{m=1}^n X_m$.

For sake of completeness and to compare with the continuous time argument, we reproduce the basic property, following [1].

Proposition 2.1 $\sigma(n) = EW^*(1/\sqrt{n})$ *satisfies conditional expected external consistency in the following sense*: *there exists a constant* M *such that*, *for any component* k *and any process* $\{X_m\}$:

$$\bar{X}_n^k - \frac{1}{n}\sum_{m=1}^n E_{\sigma(n)}(x_m|\mathcal{H}_{m-1}) \leq M/\sqrt{n}. \tag{1}$$

Proof Let $W_n = \sum_k (1+\alpha)^{S_n^k}$, hence (recall that $0 \leq X_m^k \leq 1$)

$$\begin{aligned}
\frac{W_{n+1}}{W_n} &= \sum_k \frac{(1+\alpha)^{S_n^k}(1+\alpha)^{X_{n+1}^k}}{W_n} \\
&= \sum_k p_{n+1}^k (1+\alpha)^{X_{n+1}^k} \\
&\leq \sum_k p_{n+1}^k (1+\alpha X_{n+1}^k) \\
&= 1 + \alpha\langle p_{n+1}, X_{n+1}\rangle.
\end{aligned}$$

It follows that

$$\log\left(\frac{W_n}{W_0}\right) \leq \alpha \sum_{m=1}^n \langle p_m, X_m\rangle$$

and

$$\sum_{m=1}^{n} \langle p_m, X_m \rangle \geq \frac{1}{\alpha}(S_n^k \log(1+\alpha) - \log K)$$

since $W_n \geq (1+\alpha)^{S_n^k}$, for all k in K.
Thus for α small enough one has

$$\frac{1}{n}\sum_{m=1}^{n} \langle p_m, X_m \rangle \geq \bar{X}_n^k(1-\alpha/2) - \log K/\alpha n.$$

The choice of $\alpha = 1/\sqrt{n}$ leads to

$$\max_k \bar{X}_n^k - \frac{1}{n}\sum_{m=1}^{n} \langle p_m, X_m \rangle \leq M/\sqrt{n}$$

for some constant M.
Note that $\langle p_m, X_m \rangle = E_{\sigma(n)}(x_m|\mathcal{H}_{m-1})$ so that the above inequality gives the required result. □

Obviously this implies:

Corollary 2.1 *$\sigma(n)$ satisfies expected consistency:*

$$E_{\sigma(n)}\left(r_n^{k+}\right) \to 0 \quad as \quad n \to \infty.$$

One way to obtain almost sure convergence to 0 is to use the next basic martingale property (see e.g. [13]).

Proposition 2.2 *Let U_m be a sequence of uniformly bounded (even in L^2) random variables on a probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $\mathcal{F}_m$. Assume $E(U_m|\mathcal{F}_{m-1}) = 0$, then*

$$\frac{1}{n}\sum_{m=1}^{n} U_m \to 0 \quad as \quad n \to \infty.$$

Let now σ be defined as follows: given a sequence n_m going to ∞, use $\sigma(n_1)$ for $K_1 \geq n_2/\sqrt{n}_1$ blocks of size n_1 (where the entry on block m for running $\sigma(n_1)$ is $X_{mn_1+\ell}$, $\ell = 0, ..., n_1 - 1$) then inductively $\sigma(n_m)$ for $K_m \geq n_{m+1}/\sqrt{n}_m$ blocks of size n_m. Propositions 2.1 and 2.2 thus imply the following result:

Theorem 2.1 *σ satisfies external consistency.*

Remarks (1) The optimal choice of the parameter, A or α, in Proposition 2.1, is a function of the length n of the process (see the discussion in Sect. 6).
(2) To implement the algorithm the actual past play of the predictor (namely the sequence $\{k_m\}$) is not used. (This property also holds true for other algorithms like Fictitious Play or Smooth Fictitious Play [11]).

(3) Invariance by translation: one can use the vector of total payoffs $\{S_n^k\}_k$ or the vector of total regrets $\{\sum_{m=1}^n r_m^k = S_n^k - \sum_{m=1}^n x_m\}_k$ to define p_n, since $p_{n+1} = \ell[\{p_n^k e^{A(X_n^k + Y_n)}\}_k]$ for any $Y_n \in \mathbb{R}$. This property is specific to the exponential weight and is not shared by regret procedures based on Blackwell's approachability [5,14]).

(4) It is enough to satisfy a property like: for all n, there exists $\sigma(n)$ such that

$$\sum_{m=1}^{n} \left(X_m^k - \langle p_m, X_m\rangle\right) \leq o(n)$$

(uniformly in $\{X_m\}$) to obtain Theorem 2.1 by defining σ through concatenation as above and then using Proposition 2.2.

3 The multiplicative weight algorithm: continuous time

Given a measurable process X_t, $t \geq 0$, with values in $[0, 1]^K$, let $S_t = \int_0^t X_s ds = t\bar{X}_t$.

Definition A measurable process $p_t \in \Delta(K)$ is a continuous time exponential weigth algorithm $(CTEW)$ if it satisfies

$$p_t = \ell[\{\exp\left(S_t^k\right)\}_k].$$

Let $W_t = \sum_k \exp\left(S_t^k\right)$ so that $p_t^k W_t = \exp\left(S_t^k\right)$.

Theorem 3.1 *Conditional expected external consistency holds for CTEW in the sense that, for any $T > 0$ and any k:*

$$\frac{1}{T}\left(\int_0^T X_s^k ds - \int_0^T \langle p_s, X_s\rangle ds\right) \leq \frac{\log K}{T}.$$

Proof One has

$$\dot{W}_t = \sum_k \exp\left(S_t^k\right) X_t^k = \sum_k W_t\ p_t^k\ X_t^k = \langle p_t, X_t\rangle W_t.$$

Hence

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

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

$$\int_0^t \langle p_s, X_s\rangle ds \geq \int_0^t X_s^k ds - \log W_0.$$

□

Remark The interpretation is that the conditional law of the choice k_t at time t, given the past, is p_t. Let $x_t = X_t^{k_t}$ be the outcome at time t, then the average regret vector at time t is

$$r_t^k = \bar{X}_t^k - \bar{x}_t$$

and the expectation of $\bar{x}_t$ is given by

$$E(\bar{x}_t) = E\left(\frac{1}{t}\int_0^t \langle p_s, X_s\rangle ds\right).$$

The previous proof holds as well while replacing the integral S_t of the process by the integral of the (conditional expected) regret R_t defined by

$$R_t^k = \int_0^t \left(X_s^k - \langle p_s, X_s\rangle\right) ds.$$

Note that p_t satisfies also

$$p_t = \ell\left[\left\{\exp\left(R_t^k\right)\right\}_k\right].$$

Let $V_t = \sum_k \exp\left(R_t^k\right)$. Then:

$$\dot{V}_t = \sum_k \exp\left(R_t^k\right)\left(X_t^k - \langle p_t, X_t\rangle\right) = \sum_k V_t\ p_t^k\ \left(X_t^k - \langle p_t, X_t\rangle\right) = 0.$$

Hence V_t is constant and $V_t \geq \exp R_t^k$, for every k, implies

$$\int_0^t \langle p_s, X_s\rangle ds \geq \int_0^t X_s^k ds - \log V_0.$$

The same computation as above extends to the following framework:

Proposition 3.1 *Let P be a $\mathcal{C}^1$ function from $\mathbb{R}^K$ to $\mathbb{R}$ with $\nabla P \geq 0$ and $\neq 0$, such that $x^k \to +\infty$, for some component k, implies $P(x) \to +\infty$. If p_t satisfies*

$$p_t^k = \frac{\nabla^k P\left(\left\{\int_0^t \left(X_s^j - \langle p_s, X_s\rangle\right) ds\right\}_j\right)}{\sum_i \nabla^i P\left(\left\{\int_0^t \left(X_s^j - \langle p_s, X_s\rangle\right) ds\right\}_j\right)}$$

then conditional expected external consistency holds.

Proof Let R_t be defined, as above, by

$$R_t^k = \int_0^t \left(X_s^k - \langle p_s, X_s\rangle\right) ds.$$

One obtains, with $M_t = \sum_i \nabla^i P(R_t)$

$$\begin{aligned}\frac{d}{dt}P(R_t) &= \sum_k \nabla^k P(R_t)\left(X_t^k - \langle p_t, X_t\rangle\right)\\ &= M_t \sum_k p_t^k \left(X_t^k - \langle p_t, X_t\rangle\right)\\ &= 0\end{aligned}$$

Hence $P(R(t)) = P(0)$ so that each R_t^k is bounded from above and conditional expected external consistency follows. □

The previous case corresponds to $P(x) = \sum_k \exp x^k$; for similar “potential" approaches see [7,15].

4 Convergence

Given a discrete process $\{X_m\}$ and a corresponding EW algorithm $\{p_m\}$ the aim is to get a bound on

$$\frac{1}{n}\sum_{m=1}^{n}\left(X_m^k - \langle p_m, X_m\rangle\right)$$

from an evaluation of

$$\frac{1}{T}\int_0^T \left(Y_s^k - \langle q_s, Y_s\rangle\right) ds$$

where Y_t is a continuous process constructed from X_m and q_t is a $CTEW$ algorithm associated to Y_t.

Proposition 4.1 *Given a discrete time process* $\{X_m\} \in [0,1]^K, m = 1, \ldots, n,$ *and* $T > 0$, *there exists a measurable continuous time process* $\{Y_t\} \in [0,1]^K$, $t \in [0, T]$, *such that*

$$\frac{1}{n}\sum_{m=1}^{n} X_m^k = \frac{1}{T}\int_0^T Y_t^k dt$$

and

$$\frac{1}{n}\sum_{m=1}^{n}\langle p_m, X_m\rangle e^{-T/n} \leq \frac{1}{T}\int\limits_0^T \langle q_t, Y_t\rangle dt \leq \frac{1}{n}\sum_{m}\langle p_m, X_m\rangle e^{T/n}$$

where $\{p_m\}$ *is an* $EW(T/n)$ *associated to* $\{X_m\}$ *and* q_t *is a CTEW associated to* $\{Y_t\}$.

Proof Let $T > 0$. Divide the interval $[0, T]$ into n subintervals with equal lenght $\delta = T/n$ and define, from the discrete time sequence $\{X_m\}$, $m = 1, \ldots, n$, the continuous time process $\{Y_t\}$ on $[0, T]$ by the step function $Y_t = X_m$ for $t \in [(m-1)T/n, mT/n)$. Obviously

$$\frac{1}{n}\sum_{\ell=1}^{m} X_\ell^k = \frac{1}{T}\int\limits_0^{mT/n} Y_t^k dt.$$

Let $\{\hat{p}_t\}$ be the continuous time process defined from a discrete one $\{p_m\}$ as above : $\hat{p}_t = p_m$ for $t \in [(m-1)T/n, mT/n)$. Clearly also

$$\frac{1}{n}\sum_{m=1}^{n}\langle p_m, X_m\rangle = \frac{1}{T}\int\limits_0^T \langle \hat{p}_t, Y_t\rangle dt.$$

It remains to handle the difference between $\{\hat{p}_t\}$ and $\{q_t\}$ which is a $CTEW$ associated to $\{Y_t\}$. For this choose $\{p_m\}$ as the $EW(T/n)$ associated to $\{X_m\}$. Then, for $t \in [(m-1)T/n, mT/n)$ one has on the one hand

$$\hat{p}_t^k = p_m^k = \frac{\hat{W}_t^k}{\hat{W}_t}$$

with $\hat{W}_t^k = \exp\left[(T/n)\sum_{u=1}^{m-1} X_u^k\right] = \exp\left(\int_0^{(m-1)T/n} Y_s^k ds\right)$ and $\hat{W}_s = \sum_k \hat{W}_s^k$ and on the other

$$q_t^k = \frac{W_t^k}{W_t}$$

with $W_t^k = \exp\left(\int_0^t Y_s^k ds\right)$ and $W_t = \sum_k W_t^k$.
Thus, since $0 \leq Y_s^k \leq 1$, one deduces

$$\hat{W}_s^k \leq W_s^k \leq \hat{W}_s^k e^\delta$$

hence also

$$\hat{W}_s \leq W_s \leq \hat{W}_s e^\delta$$

so that

$$\hat{p}_s^k e^{-\delta} \leq q_s^k \leq \hat{p}_s^k e^{\delta}$$

and

$$\left(\frac{1}{T}\int_0^T \langle \hat{p}_s, Y_s\rangle ds\right) e^{-\delta} \leq \frac{1}{T}\int_0^T \langle p_s, Y_s\rangle ds \leq \left(\frac{1}{T}\int_0^T \langle \hat{q}_s, Y_s\rangle ds\right) e^{\delta}$$

as well. □

We thus obtain an alternative proof of Proposition 2.1 that we recall:

Lemma 4.1 *There exists a EW algorithm satisfing*

$$\frac{1}{n}\sum_{m=1}^{n}\left(X_m^k - \langle p_m, X_m\rangle\right) \leq Mn^{-1/2}.$$

Proof Given n, choose $T = \sqrt{n}$ so that:

- the bound in the continuous version is of the order $1/T = 1/\sqrt{n}$

$$\frac{1}{T}\int_0^T \left(Y_t^k - \langle q_t, Y_t\rangle\right) dt \leq \frac{\log K}{\sqrt{n}}$$

- and the error term with the discrete approximation of the order of $e^{\delta} - 1 \sim \delta = T/n = 1/\sqrt{n}$

$$\frac{1}{n}\sum_{m=1}^{n}\langle p_m, X_m\rangle \geq \frac{1}{T}\left(\int_0^T \langle q_t, Y_t\rangle dt\right) - L/\sqrt{n}$$

so that the result follows from Theorem 3.1 and Proposition 4.1.

□

Remark The choice of $T = \sqrt{n}$ amounts to taking $EW(1/\sqrt{n})$, hence as in Sect. 3 the procedure is not uniform.

5 Extensions

The same analysis applies to similar setups. We consider briefly two of them: the case of partial information where only the outcome x_n is known and the internal consistency criteria.

5.1 Partial information

Consider the framework of Sect. 1 but where the vector X_n is not revealed ex-post, only the actual chosen component x_n is announced. The aim is to define an algorithm having similar properties as in Sect. 2 but depending only on the available information. We follow [1,2].
In discrete time, define inductively the vector $\hat{X}_n$ by

$$\hat{X}_n^k = \begin{cases} \frac{X_n^k}{p_n^k} & \text{if } k = k_n \\ 0 & \text{otherwise.} \end{cases}$$

Let $\gamma \in (0, 1)$ be a parameter to be tuned later on and define first $\hat{S}_n = \sum_{m=1}^{n} \gamma \hat{X}_m / K$, then

$$\hat{p}_{n+1}^k = \frac{\exp\left(\hat{S}_n^k\right)}{\sum_j \exp\left(\hat{S}_n^j\right)}.$$

Finally the strategy at stage $n+1$ is

$$p_n^k = (1-\gamma)\hat{p}_n^k + \frac{\gamma}{K}.$$

Note that

$$E\left(\hat{X}_n^k | X_1, \ldots, X_n\right) = X_n^k$$

and

$$x_n = \langle p_n, \hat{X}_n \rangle$$

hence it is enough, using Proposition 2.2 to bound

$$\frac{1}{n}\sum_{m=1}^{n}\left(\hat{X}_m^k - \langle p_m, \hat{X}_m \rangle\right).$$

The analysis in continuous time is as follows.
Given $\{X_s\}$, let $\{\hat{X}_s\}$ and $\{p_s\}$ satisfy

$$\int_0^t x_s ds = \int_0^t p_s^{k_s} \hat{X}_s^{k_s} ds = \int_0^t \langle p_s, \hat{X}_s \rangle ds$$
$$p_s^k = (1-\gamma)\hat{p}_s^k + \gamma/K$$

and $\hat{p}_s$ be adapted to $\hat{X}_s$ as in the usual $CTEW$. In particular one obtains

$$\int_0^t x_s ds \geq (1-\gamma) \int_0^t \langle \hat{p}_s, \hat{X}_s \rangle ds$$

But, as in Sect. 3 one has for all k

$$\int_0^t \langle \hat{p}_s, \hat{X}_s \rangle ds \geq \int_0^t \hat{X}_s^k ds - C$$

hence finally one deduces:

Proposition 5.1

$$\int_0^t x_s ds \geq (1-\gamma) \int_0^t \hat{X}_s^k ds - C.$$

The corresponding discrete inequality is now

$$\frac{1}{n} \sum_{m=1}^{n} x_m \geq \frac{1}{n} \left[(1-\gamma) e^{-\delta K/\gamma} \sum_{m=1}^{n} \hat{X}_m^k \right] - C/T$$

with $\delta = T/n$. The choice of $T = 1/\gamma = n^{1/3}$ leads to:

Proposition 5.2 *There exists M such that*

$$\frac{1}{n} \sum_{m=1}^{n} \left(\hat{X}_m^k - \langle p_m, \hat{X}_m \rangle \right) \leq M n^{-1/3}.$$

5.2 Internal consistency

Given a history h_n, the average internal regret evaluation at stage n is defined by the $K \times K$ matrix r_n, with entries

$$r_n^{k\ell} = \frac{1}{n} \sum_{m, k_m = k} \left(X_m^\ell - X_m^k \right)$$

which corresponds to a comparison of the average payoff obtained on the dates where k was chosen, to the payoff for some other fixed component, ℓ, on these dates.

Definition A strategy σ satisfies internal consistency if, for every process $\{X_m\}$:

$$\max_{k,\ell} r_n^{k\ell+} \longrightarrow 0 \text{ a.s.}, \quad \text{as } n \longrightarrow \infty.$$

Using an analog of Proposition 2.2 it is enough to show, for example, that the quantities

$$Q_n(k,\ell) = \sum_{m=1}^{n} \nu_m^k \left(X_m^\ell - X_m^k\right)$$

are of the order $o(n)$, where ν_m^k stands for the conditional probability of playing k at stage m given the past history.

We first prove a lemma on invariant measures.

Lemma 5.1 *Given a matrix $A \in \mathbb{R}^{K\times K}$, let $\psi(A) \in \Delta(K)$ be the unique solution of*

$$\psi(A)^k \sum_\ell \exp A(k,\ell) = \sum_\ell \psi(A)^\ell \exp A(\ell,k).$$

Then ψ is Lipschitz continuous.

Proof Let $||.||$ denote the maximal norm and let $||A - B|| = \rho$. Then

$$\exp B(k,\ell)e^{-\rho} \le \exp A(k,\ell) \le \exp B(k,\ell)e^{\rho} \quad \forall k,\ell.$$

Similarly with $m(A) = \sum_{k,\ell} \exp A(k,\ell)$, one has $m(B)e^{-\rho} \le m(A) \le m(B)e^{\rho}$ for all k,ℓ hence

$$\frac{\exp B(k,\ell)}{m(B)} e^{-2\rho} \le \frac{\exp A(k,\ell)}{m(A)} \le \frac{\exp B(k,\ell)}{m(B)} e^{2\rho} \quad \forall k,\ell.$$

Since $\psi(A)$ is the unique invariant measure of the transition matrix with coefficients $M(k,\ell) = \frac{\exp A(k,\ell)}{m(A)}$ for $k \ne \ell$, one deduces, by Theorem 7.2 (Corollary) in [18]

$$\psi(B)e^{-4K\rho} \le \psi(A) \le e^{4K\rho}\psi(B).$$

Hence for ρ small enough $||\psi(A) - \psi(B)|| \le 5K\rho$. The constant being independent of A, B the result obtains: there exists L with

$$||\psi(A) - \psi(B)|| \le L||A - B||.$$

□

The continuous time approach is as follows. Given X_t, let

$$S_t(k, \ell) = \int_0^t \mu_s^k \left(X_s^\ell - X_s^k\right) ds$$

where $\mu_s = \psi(S_s)$ is the invariant measure associated to $\exp S_s$ and μ_s^k is the conditional probability of playing k at time s.

Proposition 5.3 *Under the above procedure, there exists a constant C such that*

$$S_t(k, \ell) \leq C \qquad \forall k, \ell \in K, \quad \forall t \geq 0.$$

Proof Define

$$A_t = \sum_{k,\ell} \exp S_t(k, \ell)$$

so that

$$\dot{A}_t = \sum_{k,\ell} \exp S_t(k, \ell) \mu_t^k \left(X_t^\ell - X_t^k\right) = 0$$

since the coefficient of X_t^k is precisely

$$\sum_\ell \exp S_t(k, \ell) \mu_t^k - \sum_\ell \exp S_t(\ell, k) \mu_t^\ell = 0$$

because $\mu_t = \psi(S_t)$. Hence $A_t = A_0 = K^2$ and each $S_t(k, \ell)$ is uniformy bounded from above. □

This property corresponds to conditional expected internal consistency.

The discrete procedure $EW(A)$ is defined inductively through $\nu_{m+1} = \psi(AQ_m)$.

Proposition 5.4 *For $4LT = \log n$, the discrete procedure $EW(T/n)$ satisfies*:

$$Q_n/n \leq M/\log n$$

hence internal consistency follows.

Proof Given a discrete process $\{X_m\}$, $m = 1, \ldots, n$, let $\{Y_t\}$ be the associated step process on $[0, T]$, as in Sect. 4. Hence one has, inductively

$$S_t(k, \ell) = \int_0^t \psi^k(S_s) \left(Y_s^\ell - Y_s^k\right) ds.$$

Define also $L_m = (T/n)Q_m$ which corresponds to $EW(T/n)$ for $\{X_m\}$, hence inductively

$$Q_m(k,\ell) = Q_{m-1}(k,\ell) + \psi^k((T/n)Q_{m-1})\left(X_m^\ell - X_m^k\right)$$

and let L_t be the continuous linear interpolation: for $t \in [mT/n, (m+1)T/n)$

$$L_t(k,\ell) = (T/n)\left[Q_m(k,\ell) + (t - mT/n)\psi^k((T/n)Q_m)\left(X_{m+1}^\ell - X_{m+1}^k\right)\right].$$

Thus L_t satisfies

$$L_t(k,\ell) - L_{mT/n}(k,\ell) = \int_{mT/n}^{t} \psi^k(L_{mT/n})\left(Y_s^\ell - Y_s^k\right)ds.$$

Hence one has

$$\begin{aligned} &S_t(k,\ell) - S_{mT/n}(k,\ell) - \left[L_t(k,\ell) - L_{mT/n}(k,\ell)\right] \\ &= \int_{mT/n}^{t} \left[\psi^k(S_s) - \psi^k(L_{mT/n})\right]\left(Y_s^\ell - Y_s^k\right)ds. \end{aligned}$$

Let $\rho = 4LT/n \geq 2\max\{||\psi(L_t) - \psi(L_{mT/n})||;\, mT/n \leq t \leq (m+1)T/n\}$. Then, using Lemma 5.1.

$$||S_t - L_t|| \leq \int_0^t 2L||S_s - L_s||ds + \rho T$$

from which one deduces, by Gronwall's lemma, that for $t \in [0, T]$:

$$||S_t - L_t|| \leq \rho T \exp(2LT).$$

Recall that S_t is bounded above by C hence

$$L_T(k,\ell)/T \leq C/T + (4LT/n)\exp(2LT).$$

It then follows, choosing $4LT = \log n$ that

$$Q_n/n \leq M/\log n.$$

□

6 Comments

Let us compare several discrete procedures leading to consistency and their continuous counterparts.

First consider procedures related to exponential evaluation.

The current one (EW) builds on a state parameter $z_m = S_m$ for which the updating rule is time independent: if z_m is the state at stage m and ξ_{m+1} the current observation at stage $m+1$, the new state is $z_{m+1} = h(z_m, \xi_{m+1})$, (for example $S_{m+1} = S_m + \xi_{m+1}$). This applies to a family of procedures, see e.g. [8,9], however the precision of the procedure depends on a parameter that is a function of the lenght of the process: it is not uniform. In our case the optimal value of the parameter to handle the n-stage problem is $1/\sqrt{n}$. Hence to obtain consistency the parameter has to be time dependent. Concerning the continuous time embedding, it has to be performed on a compact interval. Note that in fact in our analysis, the approximation is through a sequence of longer and longer intervals, of size $T = \sqrt{n}$, with finer and finer discretization of mesh $T/n = 1/\sqrt{n}$.

Another exponential evalution, but where the state variable is the average sum $\bar{S}_m = S_m/m$ is used in smooth fictitious play, see [11,12]. Explicitly p_m is given by $\ell\big[\{\exp((1/\epsilon)\bar{S}_m^k)\}_k\big]$ and the corresponding procedure satisfies (approximate) consistency. Note that the udpdating rule requires the knowledge of the current stage; namely $z_{m+1} = \frac{1}{m+1}(mz_m + \xi_{m+1}) = h_m(z_m, \xi_{m+1})$. However this equation is a special case of discrete dynamics of the form

$$z_{m+1} - z_m = a_{m+1} F(z_n, \xi_{m+1})$$

with $\sum a_n = +\infty$, $\sum a_n^2 < +\infty$ and F with bounded range. Hence, see [3,4] the asymptotics of this dynamics can be studied using the asymptotics of the continuous time process

$$\dot{z}(t) = (\in)G(z(t))$$

where G satisfies $E(F(z_m, \xi_{m+1})|\mathcal{H}_m) = (\in)G(z_m)$. Thus there is no need to adapt the coefficient to the lenght of the process.

The same study through continuous time processes applies to regret dynamics [15,16] based on approachability [5] which satisfy consistency as well, but do depend on the past behavior of the predictor, see [4].

It is interesting to notice that Blackwell's original procedure, based on L^2 norm, when applied to the current framework (where one approaches an orthant) satisfies positive homogeneity of degree zero. Hence $p(h_m)$ can be defined as a function of the average regret r_m or of the total regret mr_m.

For extension to potential based algorithms, where the same machinery will work, see [7,15].

Acknowledgments Part of this work was done while visting IMPA (Rio, Brazil) and CMM (Santiago, Chile). The support of these institutions is gratefully acknowledged.

References

1. Auer, P., Cesa-Bianchi, N., Freund, Y., Shapire, R.E.: Gambling in a rigged casino: the adversarial multi-armed bandit problem. Proceedings of the 36th annual IEEE symposium on foundations of computer science, pp. 322–331 (1995)
2. Auer, P., Cesa-Bianchi, N., Freund, Y., Shapire, R.E.: The non-stochastic multiarmed bandit problem. SIAM J. Comput. **32**, 48–77 (2002)
3. Benaim, M., Hofbauer, J., Sorin, S.: Stochastic approximations and differential inclusions. SIAM J. Opt. Control **44**, 328–348 (2005)
4. Benaim, M., Hofbauer, J., Sorin, S.: Stochastic approximations and differential inclusions. Part II: applications. Math. Oper. Res. **31**, 673–695 (2006)
5. Blackwell, D.: An analog of the minmax theorem for vector payoffs. Pac. J. Math. **6**, 1–8 (1956)
6. Brezis, H., Lions, P.-L.: Produits infinis de résolvantes. Israel J. Math. **29**, 329–345 (1978)
7. Cesa-Bianchi, N., Lugosi, G.: Potential-based algorithms in on-line prediction and game theory. Mach. Learn. **51**, 239–261 (2003)
8. Foster, D., Vohra, R.: Asymptotic calibration. Biometrika **85**, 379–390 (1998)
9. Foster, D., Vohra, R.: Regret in the on-line decision problem. Games Econ. Behav. **29**, 7–35 (1999)
10. Freund, Y., Shapire, R.E.: Adaptive game playing using multiplicative weights. Games Econ. Behav. **29**, 79–103 (1999)
11. Fudenberg, D., Levine, D.K.: Consistency and cautious fictitious play. J. Econ. Dyn. Control **19**, 1065–1089 (1995)
12. Fudenberg, D., Levine, D.K.: Conditional universal consistency. Games Econ. Behav. **29**, 104–130 (1999)
13. Hall, P., Heyde, C.: Martingale limit theory and its applications. Academic, London (1980)
14. Hart, S., Mas-Colell, A.: A simple adaptive procedure leading to correlated equilibria. Econometrica **68**, 1127–1150 (2000)
15. Hart, S., Mas-Colell, A.: A general class of adaptive strategies. J. Econ. Theory **98**, 26–54 (2001)
16. Hart, S., Mas-Colell, A.: Regret-based continuous time dynamics. Games Econ. Behav. **45**, 375–394 (2003)
17. Littlestone, N., Warmuth, M.K.: The weighted majority algorithm. Inform. Comput. **108**, 212–261 (1994)
18. Seneta, E.: Non-negative matrices and Markov chains. Springer, Heidelberg (1981)