Games and Economic Behavior 109 (2018) 401–412

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Note

Sampled fictitious play is Hannan consistent

Zifan Li [1], Ambuj Tewari [*,2]

Department of Statistics, University of Michigan, 311 West Hall, 1085 South University, Ann Arbor, MI 48109-1107, USA

ARTICLE INFO

Article history:
Received 31 July 2016
Available online 5 February 2018

JEL classification:
C73

Keywords:
Adaptive heuristics
Learning
Repeated games
Hannan consistency
Fictitious play

ABSTRACT

Fictitious play is a simple and widely studied adaptive heuristic for playing repeated games. It is well known that fictitious play fails to be Hannan consistent. Several variants of fictitious play including regret matching, generalized regret matching and smooth fictitious play, are known to be Hannan consistent. In this note, we consider sampled fictitious play: at each round, the player samples past times and plays the best response to previous moves of other players at the sampled time points. We show that sampled fictitious play, using Bernoulli sampling, is Hannan consistent. Unlike several existing Hannan consistency proofs that rely on concentration of measure results, ours instead uses anti-concentration results from Littlewood–Offord theory.

1. Introduction

In the setting of repeated games played in discrete time, the (unconditional) regret of a player, at any time point, is the difference between the payoffs she would have received had she played the best, in hindsight, constant strategy throughout, and the payoffs she did in fact receive. Hannan (1957) showed the existence of procedures with a "no-regret" property: procedures for which the average regret per time goes to zero for a large number of time points. His procedure was a simple modification of fictitious play: random perturbations are added to the cumulative payoffs of every strategy so far and the player picks the strategy with the largest perturbed cumulative payoff. No regret procedures are also called "universally consistent" (Fudenberg and Levine, 1998, Section 4.7) or "Hannan consistent" (Cesa-Bianchi and Lugosi, 2006, Section 4.2).

It is well known that smoothing the cumulative payoffs before computing the best response is crucial to achieve Hannan consistency. One way to achieve smoothness is through stochastic smoothing, or adding perturbations. Without perturbations, the procedure becomes identical to fictitious play, which fails to be Hannan consistent (Cesa-Bianchi and Lugosi, 2006, Exercise 3.8). Besides Hannan's modification, other variants of fictitious play are also known to be Hannan consistent, including (unconditional) regret matching, generalized (unconditional) regret matching and smooth fictitious play (for an overview, see Hart and Mas-Colell (2013, Section 10.9)).

In this note, we consider another variant of fictitious play, namely sampled fictitious play. Here, the player samples past time points using some (randomized) sampling scheme and plays the best response to the moves of the other players restricted to the set of sampled time points. Sampled fictitious play has been considered by other authors in different

* Corresponding author.
E-mail addresses: zifanli@umich.edu (Z. Li), tewaria@umich.edu (A. Tewari).

[1] Zifan Li is a Ph.D. student in the Department of Statistics and Data Science at Yale University. This research is part of his undergraduate honors thesis undertaken with the guidance of the second author when he was at the University of Michigan.

[2] Ambuj Tewari is an associate professor in the Department of Statistics and the Department of Electrical Engineering and Computer Science (by courtesy) at the University of Michigan.

https://doi.org/10.1016/j.geb.2018.01.005
0899-8256/

contexts. Kaniovski and Young (1995) established convergence to Nash equilibrium in 2×2 games. Gilliland and Jung (2006) provided regret bounds for the game of matching pennies. Lambert et al. (2005) considered games with identical payoffs for all players and use sampled fictitious play to solve large-scale optimization problems. To the best of our knowledge, it is not known whether sampled fictitious play is Hannan consistent without making any assumptions on the form of the game and payoffs. The purpose of this note is to show that it is indeed Hannan consistent when used with a natural sampling scheme, namely Bernoulli sampling.

2. Preliminaries

Consider a game in strategic form where M is the number of players, S_i is the set of strategies for player i, and $u_i : \prod_{j=1}^{M} S_i \to \mathbb{R}$ is the payoff function for player i. For simplicity assume that the payoff functions of all players are $[-1, 1]$ bounded. We also assume the number of pure strategies is the same for each player and that $S_i = \{1, \dots, N\}$. Let $S = \prod_{i=1}^{M} S_i$ be the set of M-tuples of player strategies. For $s = (s_i)_{i=1}^{M} \in S$, we denote the strategies of players other than i by $s_{-i} = (s_j)_{1 \le j \le M, j \ne i}$.

The game is played repeatedly over (discrete) time $t = 1, 2, \dots$. A learning procedure for player i is a procedure that maps the history $h_{t-1} = (s_\tau)_{\tau=1}^{t-1}$ of plays just prior to time t, to a strategy $s_{t,i} \in S_i$. The learning procedure is allowed to be randomized, i.e., player i has access to a stream of random variables $\epsilon_1, \epsilon_2, \dots$ and she is allowed to use $\epsilon_1, \dots, \epsilon_{t-1}$, in addition to h_{t-1}, to choose $s_{t,i}$. Player i's regret at time t is defined as

$$\mathcal{R}_{t,i} = \max_{k \in S_i} \sum_{\tau=1}^{t} u_i(k, s_{\tau,-i}) - \sum_{\tau=1}^{t} u_i(s_\tau).$$

This compares the player's cumulative payoff with the payoff she could have received had she selected the best constant (over time) strategy k with knowledge of the other players' moves.

A learning procedure for player i is said to be *Hannan consistent* if and only if

$$\limsup_{t \to \infty} \frac{\mathcal{R}_{t,i}}{t} \le 0 \qquad \text{almost surely.}$$

Hannan consistency is also known as the "no-regret" property and as "universal consistency". The term "universal" refers to the fact that the regret per time goes to zero irrespective of what the other players do.

Fictitious play is a (deterministic) learning procedure where player i plays the best response to the plays of the other players so far. That is,

$$s_{t,i} \in \arg\max_{k \in \{1,\dots,N\}} \sum_{\tau=1}^{t-1} u_i(k, s_{\tau,-i}). \tag{1}$$

As mentioned earlier, fictitious play is not Hannan consistent. However, consider the following modification of fictitious play, called *sampled fictitious play*. At time t, player randomly selects a subset $\mathbb{S}_t \subseteq \{1, \dots, t-1\}$ of previous time points and plays the best response to the other players' moves only over $\mathbb{S}_t$. That is,

$$s_{t,i} \in \arg\max_{k \in \{1,\dots,N\}} \sum_{\tau \in \mathbb{S}_t} u_i(k, s_{\tau,-i}). \tag{2}$$

If multiple strategies achieve the maximum, then the tie is broken uniformly at random, and independently with respect to all previous randomness. Also, if $\mathbb{S}_t$ turns out to be empty (an event that happens with probability exactly $2^{-(t-1)}$ under the Bernoulli sampling described below), we adopt the convention that the argmax above includes all N strategies.

In this note, we consider *Bernoulli sampling*, i.e., any particular round $\tau \in \{1, \dots, t-1\}$ is included in $\mathbb{S}_t$ independently with probability $1/2$. More specifically, if $\epsilon_1^{(t)}, \dots, \epsilon_{t-1}^{(t)}$ are i.i.d. symmetric Bernoulli (or Rademacher) random variables taking values in $\{-1, +1\}$, then

$$\mathbb{S}_t = \{\tau \in \{1, \dots, t-1\} : \epsilon_\tau^{(t)} = +1\} \tag{3}$$

and therefore,

$$\sum_{\tau \in \mathbb{S}_t} u_i(k, s_{\tau,-i}) = \sum_{\tau=1}^{t-1} \frac{(1 + \epsilon_\tau^{(t)})}{2} u_i(k, s_{\tau,-i}).$$

Note that the procedure defined by the combination of (2) and (3) is completely parameter free, i.e., there is no tuning parameter that has to be carefully tuned in order to obtain desired convergence properties.