

An introduction to stochastic approximation

Richard Combes

October 11, 2013

1 The basic stochastic approximation scheme

1.1 A first example

We propose to start the exposition of the topic by an example. The arguments are given in a crude manner. Formal proofs will be given in section 2. This example is taken from the very article [6] which introduced stochastic approximation. Consider $x \in \mathbb{R}$ the parameter of a system and $g(x) \in \mathbb{R}$ an output value from this system when parameter x is used. We assume g to be a smooth, increasing function. An agent wants to determine sequentially $x^* \in \mathbb{R}$ the value such that the system output equals a target value g^* . If for all x , the value of $g(x)$ can be observed directly from the system, then determining g^* could be solved by a simple search technique such as binary search or golden ratio search. Here we assume that only a *noisy* version of g can be observed. Namely, at time $n \in \mathbb{N}$, the decision maker sets the parameter equal to x_n , and observes $Y_n = g(x_n) + M_n$ with M_n a random variable denoting noise, with $\mathbb{E}[M_n] = 0$.

In order to determine $g(x)$, a crude approach would be to sample parameter x repeatedly and average the result, so that the effect of noise would cancel out, and apply a deterministic line search (such as binary search).

[6] proposed a much more elegant approach. If $x_n > x^*$, we have that $g(x_n) > g^*$, so that diminishing x_n by a small amount proportional to $g^* - g(x_n)$ would guarantee $x_{n+1} \in [x^*, x_n]$. Therefore, define ϵ_n a sequence of small positive numbers, and consider the following update scheme:

$$x_{n+1} = x_n + \epsilon_n(g^* - Y_n) = x_n + \epsilon_n(g^* - g(x_n)) + \epsilon_n M_n.$$

The first intuition is that if the noise sequence is well behaved (say $\{M_n\}$ is i.i.d Gaussian with mean 0 and variance 1) and $\epsilon_n = 1/n$, then the law of large numbers would guarantee that the noise “averages out” so that for large n noise can be ignored altogether. Namely, define $S_n = \sum_{k \geq n} M_k/k$, then $\text{var}(S_n)$ is upper bounded by $\sum_{k \geq n} 1/k^2 \rightarrow_{n \rightarrow +\infty} 0$, so that S_n should be negligible. (Obviously this reasoning is heuristic and to make it precise we have to use a law of large numbers-like result ...)

Now assume no noise ($M_n \equiv 0$ for all n), and $\epsilon_n = 1/n$, g smooth with a strictly positive first derivative upper bounded by g' . Removing the noise term M_n :

$$g(x_{n+1}) = g(x_n) + \frac{g'(x_n)}{n}(g^* - g(x_n)).$$

By the fundamental theorem of calculus: $(1/n)|g^* - g(x_n)| \leq (g'/n)|x^* - x_n|$. So for $n \geq g'$, we have either $x_n \leq x_{n+1} \leq x^*$ or $x_n \geq x_{n+1} \geq x^*$. In both cases, $n \mapsto |g(x_n) - g^*|$ is decreasing for large n .

It is also noted that:

$$\frac{x_{n+1} - x_n}{n} = (g^* - g(x_n)),$$

so that x_n appears as a discretization (with discretization steps $\{1/n\}$ of the following ordinary differential equation (o.d.e.):

$$\dot{x} = g^* - g(x).$$

This analogy will be made precise in the next subsection.

1.2 The associated o.d.e

We now introduce the so-called o.d.e. approach popularized by [5], which allow to analyze stochastic recursive algorithms such as the one considered by Robbins in his original paper. It is noted that [6] did not rely on the o.d.e. method and used direct probabilistic arguments.

The crude reasoning above suggests that the asymptotic behavior of the random sequence $\{x_n\}$ can be obtained by determining the asymptotic behavior of a corresponding (deterministic) o.d.e. In this lecture we will consider a sequence $x_n \in \mathbb{R}^d$, $d \geq 1$, and a general update scheme of the following form:

$$x_{n+1} = x_n + \epsilon_n(h(x_n) + M_n),$$

with $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$. We define the associated o.d.e.:

$$\dot{x} = h(x).$$

We will prove that (with suitable assumptions on h , the noise and step sizes) if the o.d.e. admits a continuously differentiable Liapunov function V , then we have that $V(x_n) \rightarrow_{n \rightarrow +\infty} 0$ almost surely. We recall that V is a Liapunov function if it is positive, radially unbounded, and strictly diminishing along the solutions of the o.d.e.

1.3 Instances of stochastic approximation algorithms

Algorithms based on stochastic approximation schemes have become ubiquitous in various fields, including signal processing, optimization, machine learning and economics/game theory. There are several reasons for this:

- Low memory requirements: the basic stochastic approximation is a Markovian update: the value of x_{n+1} is a function of x_n and the observation at time n . So its implementation requires a small amount of memory.
- Influence of noise: stochastic approximation algorithms are able to work with noise, so that they are good candidates as “on-line” optimization algorithms which work with the noisy output of a running system. Furthermore the convergence of a stochastic approximation scheme is determined by inspecting a deterministic o.d.e. which is simpler to analyze and does not depend on the statistics of the noise.

- Iterative updates: Once again since they are Markovian updates, stochastic approximation schemes are good models for collective learning phenomena where a set of agents interact repeatedly and update their behavior depending on their most recent observation. This is the reason why results on learning schemes in game theory rely heavily on stochastic approximation arguments.

We give a few examples of stochastic algorithms found in the literature.

1.4 Stochastic gradient algorithms

Stochastic gradient algorithms allow to find a local maximum of a cost function whose value is only known through noisy measurements, and are commonplace in machine learning (on-line regression, training of neural networks, on-line optimization of Markov decision processes etc). We consider a function $f : \mathbb{R} \rightarrow \mathbb{R}$ which is strongly convex, twice differentiable with a unique minimum x^* . f cannot be observed directly, nor can its gradient ∇f . At time n we can observe $f(x_n) + M_n$. Therefore it makes sense to approximate ∇f by finite differences, with a suitable discretization step. Consider the scheme (due to Kiefer and Wolfowitz [4]):

$$x_{n+1} = x_n - \epsilon_n \frac{f(x_n + \delta_n) - f(x_n - \delta_n)}{2\delta_n},$$

The associated o.d.e is $\dot{x} = -\nabla f(x)$ which admits the Liapunov function $V(x) = f(x) - f(x^*)$. With the proper step sizes (say $\epsilon_n = n^{-1}$, $\delta_n = n^{-1/3}$) it can be proven that the method converges to the minimum $x_n \rightarrow_{n \rightarrow \infty} f(x^*)$ almost surely.

1.5 Distributed updates

In many applications, the components of x_n are not updated simultaneously. This is for instance the case in distributed optimization when each component of x_n is controlled by a different agent. This is also the case for on-line learning algorithms for Markov Decision Processes such as Q-learning.

For instance assume that at time n , a component $k(n)$ uniformly distributed $\{1, \dots, d\}$ is chosen, and only the $k(n)$ -th component of x_n is updated:

$$x_{n+1,k} = \begin{cases} x_{n,k} + \epsilon_n (h_k(x_n) + M_{n,k}) & , k = k(n) \\ x_{n,k} & , k \neq k(n) \end{cases}.$$

Then it can be proven that the behavior of $\{x_n\}$ can be described by the o.d.e $\dot{x} = h(x)$. Namely, the asymptotic behavior of $\{x_n\}$ is the same as in the case where all its components are updated simultaneously. This is described for instance in [1][Chap 7].

1.6 Fictitious play

Fictitious play is a learning dynamic for games introduced by [2], and studied extensively by game theorists afterwards (see for instance [3]). Consider 2 agents playing a matrix game. Namely, at time $n \in \mathbb{N}$, agent $k \in \{1, 2\}$ chooses action $a_n^k \in \{1, \dots, \mathcal{A}\}$, and receives a

reward A_{a_1, a_2}^k , where A^1, A^2 are two \mathcal{A} by \mathcal{A} matrices with real entries. Define the empirical distribution of actions of player k at time n by :

$$p^k(a, n) = \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{a_t^k = a\}.$$

A natural learning scheme for agent k is to assume that at time $n+1$, agent k' will choose an action whose probability distribution is equal to $p^{k'}(\cdot, n)$, and play the best action. Namely agent k assumes that $\mathbb{P}[a_{n+1}^{k'} = a] = p^{k'}(a, n)$, and chooses the action maximizing his expected payoff given that assumption.

We define

$$g^k(\cdot, p') = \max_{p \in \mathcal{P}} \sum_{1 \leq a \leq \mathcal{A}} \sum_{1 \leq a' \leq \mathcal{A}} p(a) A_{a, a'}^k p'(a'),$$

with \mathcal{P} the set of probability distributions on $\{1, \dots, \mathcal{A}\}$. $g^k(\cdot, p')$ is the probability distribution of the action of k maximizing the expected payoff, knowing that player k' will play an action distributed as p' .

The empirical probabilities can be written recursively as:

$$(n+1)p^k(a, n+1) = np^k(a, n) + \mathbf{1}\{a_n^k = a\},$$

so that:

$$p^k(a, n+1) = p^k(a, n) + \frac{1}{n+1} (\mathbf{1}\{a_n^k = a\} - p^k(a, n)).$$

Using the fact that $\mathbb{E}[\mathbf{1}\{a_n^k = a\}] = g^k(\cdot, p^{k'})$, we recognize that the probabilities p are updated according to a stochastic approximation scheme with $\epsilon_n = 1/(n+1)$, and the corresponding o.d.e. is

$$\dot{p} = g(p) - p.$$

It is noted that such an o.d.e may have complicated dynamics and might not admit a Liapunov function without further assumptions on the structure of the game (the matrices A^1 and A^2).

2 Convergence to the o.d.e limit

In this section we prove the basic stochastic approximation convergence result for diminishing step sizes with martingale difference noise. This setup is sufficiently simple to grasp the proof techniques without relying on sophisticated results. The only prerequisites are: the (discrete-time) martingale convergence theorem and two basic results on o.d.e., namely Gronwall's inequality and the Picard - Lindelöf theorem. We largely follow the exposition given by Borkar in [1][Chap 2].

2.1 Assumptions

We denote by \mathcal{F}_n the σ -algebra generated by $(x_0, M_0, \dots, x_n, M_n)$. Namely \mathcal{F}_n contains all the information about the history of the algorithm up to time n .

We introduce the following assumptions:

- (A1) (Lipshitz continuity of h) There exists $L \geq 0$ such that for all $x, y \in \mathbb{R}^d$ $\|h(x) - h(y)\| \leq L\|x - y\|$.
- (A2) (Diminishing step sizes) We have that $\sum_{n \geq 0} \epsilon_n = \infty$ and $\sum_{n \geq 0} \epsilon_n^2 < \infty$.
- (A3) (Martingale difference noise) There exists $K \geq 0$ such that for all n we have that $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = 0$ and $\mathbb{E}[|M_{n+1}|^2|\mathcal{F}_n] \leq K(1 + \|x_n\|)$.
- (A4) (Boundedness of the iterates) We have that $\sup_{n \geq 0} \|x_n\| < \infty$ almost surely.
- (A5) (Liapunov function) There exists a positive, radially unbounded, continuously differentiable function $V : \mathbb{R}^d \rightarrow \mathbb{R}$ such that for all $x \in \mathbb{R}^d$, $\langle \nabla V(x), h(x) \rangle \leq 0$ with strict inequality if $V(x) \neq 0$.

(A1) is necessary to ensure that the o.d.e. has a unique solution given an initial condition, and that the value of the solution after a given amount of time depends continuously on the initial condition. (A2) is necessary for almost sure convergence, and holds in particular for $\epsilon_n = 1/n$. (A3) is required to control the random fluctuations of x_n around the solution of the o.d.e. (using the martingale convergence theorem), and holds in particular if $\{M_n\}_{n \in \mathbb{N}}$ is independent with bounded variance. (A4) is essential, and can (in some cases) be difficult to prove. We will discuss how to ensure that (A4) holds in the latter sections. (A5) ensures that all solutions of the o.d.e. converge to the set of zeros of V , *and that this set is stable* (in the sense of Liapunov). Barely assuming that all solutions of the o.d.e. converge to a single point does not guarantee convergence of the corresponding stochastic approximation.

2.2 The main theorem

We are now equipped to state the main theorem.

Theorem 1. *Assume that (A1) - (A5) hold, then we have that:*

$$V(x_n) \rightarrow_{n \rightarrow \infty} 0, \text{ a.s.}$$

The proof of theorem 1 is based on an intermediate result stating that the sequence $\{x_n\}$ (suitably interpolated) remains arbitrarily close to the solution of the o.d.e. We define $\Phi_t(x)$ the value at t of the unique solution to the o.d.e. starting at x at time 0. Φ is uniquely defined because of (A1) and the Picard-Lindelöf theorem. We define $t(n) = \sum_{k=0}^{n-1} \epsilon_k$, and $\bar{x}(t)$ the interpolated version of $\{x_n\}_{n \in \mathbb{N}}$. Namely for all n , $\bar{x}(t(n)) = x_n$, and \bar{x} is linear by parts. We define $x^n(t) = \Phi_{t-t(n)}(x_n)$ the o.d.e. trajectory started at x_n at time $t(n)$.

Lemma 1. *For all $T > 0$, we have that:*

$$\sup_{t \in [t(n), t(n)+T]} \|\bar{x}(t) - x^n(t)\| \rightarrow_{n \rightarrow \infty} 0 \text{ a.s.}$$

Proof of lemma 1: Since the result holds almost surely we consider a fixed sample path throughout the proof. Define $m = \inf\{k : t(k) > t(n) + T\}$ so that we can prove the result for $T = t(m) - t(n)$ and consider the time interval $[t(n), t(m)]$. Consider $n \leq k \leq m$, we are going

to start by bounding the difference between \bar{x} and x^n at time instants $t \in \{t(n), \dots, t(m)\}$, that is $\sup_{n \leq k \leq m} |x_k - x^n(t(k))|$.

We start by re-writing the definition of x_k and $x^k(t(k))$:

$$x_k = x_n + \sum_{u=n}^{k-1} \epsilon_u h(x_u) + \sum_{u=1}^{k-1} \epsilon_k M_k$$

and by the fundamental theorem of calculus:

$$\begin{aligned} x^n(t(k)) &= x_n + \int_{t(n)}^{t(k)} h(x^n(v)) dv \\ &= x_n + \sum_{u=n}^{k-1} \int_{t(u)}^{t(u+1)} h(x^n(v)) dv \\ &= x_n + \sum_{u=n}^{k-1} \epsilon_u h(x^n(t(u))) + \int_{t(u)}^{t(u+1)} [h(x^n(v)) - h(x^n(t(u)))] dv \end{aligned}$$

we recall that $\int_{t(u)}^{t(u+1)} dv = \epsilon_u$.

Our goal is to upper bound the following difference, decomposed into 3 terms:

$$C_k = \|x^n(t(k)) - x_k\| \leq A_k + \sum_{u=n}^{k-1} B_u + \sum_{u=n}^{k-1} L \epsilon_u C_u, \quad (1)$$

with:

$$\begin{aligned} A_k &= \left\| \sum_{u=n}^{k-1} \epsilon_u M_u \right\|, \\ B_k &= \int_{t(u)}^{t(u+1)} \|h(x^n(v)) - h(x^n(t(u)))\| dv. \end{aligned}$$

The stochastic term

We first upper bound A_k , the stochastic term in (1). Define $S_n = \sum_{u=0}^n \epsilon_u M_u$. It is noted that $A_k = S_k - S_n$. S_n is a martingale since:

$$\mathbb{E}[S_{n+1} - S_n | \mathcal{F}_n] = \mathbb{E}[\epsilon_{n+1} M_{n+1} | \mathcal{F}_n] = 0.$$

From (A3), $\mathbb{E}[\|M_{n+1}\|^2 | \mathcal{F}_n] \leq K(1 + \sup_k \|x_k\|) < \infty$. Therefore the sequence $\{S_n\}$ is a square integrable martingale:

$$\sum_{n \geq 0} \mathbb{E}[\|S_{n+1} - S_n\|^2 | \mathcal{F}_n] \leq K(1 + \sup_n \|x_n\|) \sum_{n \geq 0} \epsilon_n^2 < \infty.$$

Using the martingale convergence theorem (lemma 2), we have that S_n converges almost surely to a finite value S_∞ . This implies that:

$$A_k \leq \|S_k - S_n\| \leq \|S_k - S_\infty\| + \|S_n - S_\infty\| \leq 2 \sup_{n' \geq n} \|S_{n'} - S_\infty\| \xrightarrow{n \rightarrow \infty} 0, \text{ a.s.}$$

Therefore, until the end of the proof we choose n large enough so that $A_k \leq \delta/2$ for all $k \geq n$ with $\delta > 0$ arbitrarily small.

The discretization term, maximal slope of x^n

In order to upper bound B_u , we prove that for $t \in [t(u), t(u+1)]$, $x^n(t)$ can be approximated by $x^n(t(u))$ (up to a term proportional to ϵ_u). To do so we have to bound the maximal slope of $t \mapsto x^n(t)$ on $[t(n), t(m)]$. We know that $x^n(t(n)) = x_n \leq \sup_{n \in \mathbb{N}} \|x_n\|$ which is finite by (A4). Using the fact that h is Lipschitz and applying Gronwall's inequality (lemma 2) there exists a constant $K_T > 0$ such that:

$$\|h(x^n(t))\| \leq \|h(0)\| + L\|x^n(t)\| \leq K_T, \quad t \in [t(n), t(m)].$$

We have used the fact that h is Lipschitz so it grows at most linearly: for all x , $\|h(x) - h(0)\| \leq L\|x\|$, so that $\|h(x)\| \leq \|h(0)\| + L\|x\|$. Therefore by the fundamental theorem of calculus, for $t \in [t(u), t(u+1)]$:

$$\|x^n(t) - x^n(t(u))\| \leq \int_{t(u)}^{t(u+1)} \|h(x^n(v))\| dv \leq \epsilon_u K_T.$$

In turn, using the Lipschitz continuity of h we have that:

$$B_u \leq \int_{t(u)}^{t(u+1)} L\|x^n(v) - x^n(t(u))\| dv \leq \epsilon_u^2 L K_T.$$

By (A2), $\sum_{u \geq n} \epsilon_u^2 \rightarrow_{n \rightarrow +\infty} 0$, and so $\sum_{u=n}^{k-1} B_u \leq \sum_{u \geq n} B_u \rightarrow_{n \rightarrow +\infty} 0$. Until the end of the proof, we consider n large enough so that $\sum_{u \geq n} B_u \leq \delta/2$.

The recursive term

Going back to (1), by the reasoning above, we have proven that:

$$C_k \leq \delta + L \sum_{u=0}^{k-1} \epsilon_u C_u.$$

Using the fact that $\sum_{u=n}^{k-1} \epsilon_u \leq t(m) - t(n) = T$, and applying the discrete time version of Gronwall's inequality (lemma 3):

$$\sup_{n \leq k \leq m} C_k \leq \delta e^{LT}.$$

By letting δ arbitrary small we have proven that:

$$\sup_{n \leq k \leq m} \|x_k - x^n(t(k))\| \rightarrow_{n \rightarrow \infty} 0$$

Error due to linear interpolation

In order to finish the proof, we need to provide an upper bound for $\|\bar{x}(t) - x^n(t)\|$ when $t \notin \{t(n), \dots, t(m)\}$. Consider $n \leq k \leq m$, and $t \in [t(k), t(k+1)]$. Since \bar{x} is linear by parts (by definition), there exists $\lambda \in [0, 1]$ such that:

$$\bar{x}(t) = \lambda x_k + (1 - \lambda)x_{k+1}.$$

Applying the fundamental theorem of calculus twice, $x^n(t)$ can be written:

$$\begin{aligned} x^n(t) &= x^n(t(k)) + \int_{t(k)}^t h(x^n(v))dv \\ &= x^n(t(k+1)) - \int_t^{t(k+1)} h(x^n(v))dv \end{aligned}$$

Therefore the error due to linear interpolation can be upper bounded as follows:

$$\begin{aligned} \|\bar{x}(t) - x^n(t)\| &\leq \lambda \|x_k - x^n(t(k))\| + (1 - \lambda) \|x_{k+1} - x^n(t(k+1))\| \\ &\quad + \lambda \int_{t(k)}^t \|h(x^n(v))\|dv + (1 - \lambda) \int_t^{t(k+1)} \|h(x^n(v))\|dv, \end{aligned}$$

and we obtain the announced result:

$$\sup_{t \in [t(n), t(m)]} \|\bar{x}(t) - x^n(t)\| \leq \sup_{n \leq k \leq m} \|x_k - x^n(t(k))\| + \epsilon_n C_T \xrightarrow{n \rightarrow +\infty} 0$$

which concludes the proof. □

We can proceed to prove the main theorem.

Proof. Once again we work with a fixed sample path. We consider $\nu > 0$, and define the level set $H^\nu = \{x : V(x) \geq \nu\}$. Choose $\epsilon > 0$ such that if $V(x) \leq \nu$ and $\|x - y\| \leq \epsilon$, then $V(y) \leq 2\nu$. Such an ϵ exists because (by radial unboundedness) the set $\{x : V(x) \leq \nu\}$ is compact, and because of the uniform continuity of V on compact sets. Since V is continuously differentiable, $x \mapsto \langle \nabla V, h(x) \rangle$ is strictly negative on H^ν , and H^ν is closed, we define $\Delta = \sup_{x \in H^\nu} \langle \nabla V, h(x) \rangle < 0$. Denote by $V_\infty = \sup_{\|x\| \leq \sup_n \|x_n\|} V(x)$ which is finite since $\sup_n \|x_n\|$ is finite and V is continuous. Define $T = (V_\infty - \nu)/\Delta$. Then for all x such that $\|x\| \leq \sup_n \|x_n\|$ and all $t > T$, we must have $V(\Phi_t(x)) \leq \nu$. Finally, choose n large enough so that $\sup_{t \in [t(n), t(n)+T]} \|\bar{x}(t) - x^n(t)\| \leq \epsilon$ and m such that $t(m) = t(n) + T$. Then we have that $V(x^n(t+T)) \leq \nu$ and that $\|x^n(t+T) - x_m\| \leq \epsilon$, which proves that $V(x_m) \leq 2\nu$.

The reasoning above holds for all sample paths, for all $\nu > 0$, for all m arbitrarily large, so $V(x_n) \xrightarrow{n \rightarrow \infty} 0$ a.s. which is the announced result. □

3 Appendix

3.1 Ordinary differential equations

We state here two basic results on o.d.e.s used in the proof of the main theorem.

Lemma 2 (Gronwall's inequality). *Consider $T \geq 0$, $L \geq 0$ and a function $t \mapsto x(t)$ such that $\dot{x}(t) \leq L\|x(t)\|$, $t \in [0, T]$. Then we have that $\sup_{t \in [0, T]} \|x(t)\| \leq \|x(0)\|e^{LT}$.*

Lemma 3 (Gronwall's inequality, discrete case). Consider $K \geq 0$ and positive sequences $\{x_n\}$, $\{\epsilon_n\}$ such that for all $0 \leq n \leq N$:

$$x_{n+1} \leq K + \sum_{u=0}^n \epsilon_u x_u.$$

Then we have the upper bound: $\sup_{0 \leq n \leq N} x_n \leq K e^{\sum_{n=0}^N \epsilon_n}$.

3.2 Martingales

We state the martingale convergence which is required to control the random fluctuations of the stochastic approximation in the proof of the main theorem.

Consider a sequence of σ -fields $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$, and $\{M_n\}_{n \in \mathbb{N}}$ a sequence of random variables in \mathbb{R}^d . We say that $\{M_n\}_{n \in \mathbb{N}}$ is a \mathcal{F} -martingale if M_n is \mathcal{F}_n -measurable and $\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n$. The following theorem (due to Doob) states that if the sum of squared increments of a martingale is finite (in expectation), then this martingale has a finite limit a.s.

Theorem 2 (Martingale convergence theorem). Consider $\{M_n\}_{n \in \mathbb{N}}$ a martingale in \mathbb{R}^d with:

$$\sum_{n \geq 0} \mathbb{E}[|M_{n+1} - M_n|^2 | \mathcal{F}_n] < \infty,$$

then there exists a random variable $M_\infty \in \mathbb{R}^d$ such that $\|M_\infty\| < \infty$ a.s. and $M_n \rightarrow_{n \rightarrow \infty} M_\infty$ a.s.

References

- [1] Vivek S. Borkar. *Stochastic approximation: a dynamical systems viewpoint*. Cambridge University Press, 2008.
- [2] G.W Brown. Iterative solutions of games by fictitious play. *Activity Analysis of Production and Allocation*, 1951.
- [3] Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. MIT Press Books. The MIT Press, July 1998.
- [4] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):pp. 462–466, 1952.
- [5] L. Ljung. Analysis of recursive stochastic algorithms. *Automatic Control, IEEE Transactions on*, 22(4):551–575, 1977.
- [6] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, September 1951.