Almost sure convergence of a stochastic approximation process in a convex set

Abdelkrim Bennar, Jean-Marie Monnez

ALMOST SURE CONVERGENCE OF A STOCHASTIC APPROXIMATION PROCESS IN A CONVEX SET

Abdelkrim Bennar[1], Jean-Marie Monnez[2]

[1]Université Hassan 2, Faculté des Sciences Ben M'sik Sidi Othmane, Casablanca, MAROC
e-mail : bennar1@yahoo.fr

[2]Institut ElieCartan UMR 7502, Nancy-Université, CNRS, INRIA, B.P. 239 - 54506 Vandoeuvre-lès-Nancy Cedex, FRANCE
e-mail : monnez@iecn.u-nancy.fr (Corresponding author)

Abstract : We consider a stochastic approximation process in a convex set K of $\mathbb{R}^k$: $X_{n+1} = \Pi\,(X_n - A_nY_n)$, with $E\,[A_nY_n \mid T_n] = a_nM_n(X_n)$, where Π is the projection operator on K, A_n a random matrix, a_n a positive number, M_n a function from K into $\mathbb{R}^k$ and T_n the sub-σ-algebra generated by the events before time n. We prove two theorems of almost sure convergence in the case where the equation $M_n(x) = 0$ has a set of solutions and give two applications.
AMS Subj. Classification : 62L20
Key Words : stochastic approximation, linear regression

1. Introduction

We define a stochastic approximation process (X_n) in a non-empty closed convex subset K of $\mathbb{R}^k$, named parameter space ; we consider :

. for $n \geq 1$, an observable random variable Y_n in $\mathbb{R}^p$, named observation space ; remark that the observation space may be different from the parameter space ;

. for $n \geq 1$, a (k, p) random matrix A_n ;

. the projection operator Π on K ;

. the process (X_n) in K defined recursively by

$$X_{n+1} = \Pi\,(X_n - A_nY_n)$$

All random variables are defined on a probability space $(\Omega, \mathcal{A}, P)$. Denote T_n the sub-σ-algebra of $\mathcal{A}$ generated by the events before time n ; $X_1, ...X_n$, $A_1, ..., A_n, Y_1, ..., Y_{n-1}$ are measurable with respect to T_n.

Suppose that, for $n \geq 1$, there exists a measurable function M_n from K into $\mathbb{R}^k$ and a positive number a_n such that

$$E\left[A_n Y_n \mid T_n\right] = A_n E\left[Y_n \mid T_n\right] = a_n M_n(X_n) \ a.s.$$

Let B_n be a set of solutions of the equation $M_n(x) = 0$. Define a distance $d(x, B)$ from x in $\mathbb{R}^k$ to a subset B.

We give in Section 2 two almost sure convergence theorems of $d(X_n, B_n)$ to 0. An application of each theorem is given in Section 3, concerning the estimation of a quantile interval of an unknown probability distribution and the estimation of a linear regression parameter under convex constraints.

In the following, $\langle ., . \rangle$ and $\|.\|$ are respectively the usual inner product and norm in $\mathbb{R}^k$; A' denotes the transposed matrix of A, $\lambda_{\min}(B)$ the smallest eigenvalue of B ; the abbreviation $a.s.$ means almost surely.

2. Lemmas

Let (X_n) be a stochastic process in a subset K of $\mathbb{R}^k$. Let (F_n) and (φ_n) be two sequences of measurable functions from K into $\mathbb{R}^+$ and (a_n) a sequence in $\mathbb{R}^+$. Suppose :

(H1a) There exists a random variable T in $\mathbb{R}^+$ such that $F_n(X_n) \longrightarrow T$ $a.s.$

(H1b) $\sum_1^\infty a_n \varphi_n(X_n) < \infty \ a.s.$

(H2a) Whatever $0 < \epsilon < 1$, $\sum_1^\infty a_n \inf_{\left\{x \in K, \epsilon < F_n(x) < \frac{1}{\epsilon}\right\}} \varphi_n(x) = +\infty$.

Lemma 1 *Assume H1a, b and H2a hold ; then* $F_n(X_n) \longrightarrow 0 \ a.s.$

Proof. $\omega \in \Omega$ is fixed throughout the proof, belonging to the intersection of the defined $a.s.$ convergence sets. Suppose $T(\omega) \neq 0$ and suppress ω writing.

By H1a, there exist $0 < \epsilon_1 < 1$ and an integer $N(\epsilon_1)$ such that for $n > N(\epsilon_1), \epsilon_1 < F_n(X_n) < \frac{1}{\epsilon_1}$.

This implies $\varphi_n(X_n) \geq \inf_{\left\{x \in K, \epsilon_1 < F_n(x) < \frac{1}{\epsilon_1}\right\}} \varphi_n(x)$; then by H2a,

$$\sum_1^\infty a_n \varphi_n(X_n) = \infty,$$

a contradiction with H1b. Thus $T(\omega) = 0$. ■

Suppose now :

(H1c) $\|X_{n+1}-X_n\| \longrightarrow 0$ $a.s.$

(H3a) For all $0<\epsilon_1<1$, for all $\epsilon>0$, there exists $\eta>0$ such that $(\|x_1-x_2\|<\eta) \Rightarrow \left(\sup_n \sup_{\left\{\epsilon_1<F_n(x_1)<\frac{1}{\epsilon_1}\right\}} |\varphi_n(x_1)-\varphi_n(x_2)|<\epsilon\right)$

(H3b) There exist a positive integer r, a sequence of integers (n_l), for all $0<\epsilon<1$ an integer $L(\epsilon)$ such that $n_{l+1} \leq n_l + r$ and

$$b(\epsilon) = \inf_{l>L(\epsilon)} \inf_{\left\{x\in K,\, \epsilon<F_{n_l}(x)<\frac{1}{\epsilon}\right\}} \sum_{j\in I_l} \varphi_j(x) > 0$$

with $I_l = \{n_l, n_l+1, ..., n_{l+1}-1\}$

(H2b) $\sum_l \min_{j\in I_l} a_j = \infty$.

Lemma 2 *Assume H1a, b, c, H2b and H3a, b hold ; then* $F_n(X_n) \longrightarrow 0$ $a.s.$

Proof. $\omega \in \Omega$ is fixed throughout the proof, belonging to the intersection of the defined $a.s.$ convergence sets. Suppose $T(\omega) \neq 0$. Below ω is omitted.

By H1a, there exist $0<\epsilon_1<1$ and an integer $N(\epsilon_1)$ such that for $n > N(\epsilon_1), \epsilon_1 < F_n(X_n) < \frac{1}{\epsilon_1}$.

By H3b, there exists an integer $L(\epsilon_1)$ such that for $l > L(\epsilon_1)$,

$$\sum_{j\in I_l} \varphi_j(X_{n_l}) > b(\epsilon_1).$$

It follows that there exists $m_l \in I_l$ such that

$$\varphi_{m_l}(X_{n_l}) > \frac{b(\epsilon_1)}{r}.$$

Consider the decomposition

$$\begin{aligned} \varphi_{m_l}(X_{m_l}) &= \varphi_{m_l}(X_{n_l}) + \varphi_{m_l}(X_{m_l}) - \varphi_{m_l}(X_{n_l}). \\ \varphi_{m_l}(X_{m_l}) &> \frac{b(\epsilon_1)}{r} - \left|\varphi_{m_l}(X_{m_l}) - \varphi_{m_l}(X_{n_l})\right|. \end{aligned}$$

Let $\epsilon>0$; by H3a, there exists $\eta>0$ corresponding to ϵ_1 and ϵ ; by H1c, we have for l sufficiently large :

$$\|X_{m_l} - X_{n_l}\| < \eta \quad ; \quad \epsilon_1 < F_{m_l}(X_{m_l}) < \frac{1}{\epsilon_1}.$$

By H3a, this implies :

$$\left|\varphi_{m_l}(X_{m_l}) - \varphi_{m_l}(X_{n_l})\right| < \epsilon \text{ ;}$$

$$\varphi_{m_l}(X_{m_l}) > \frac{b(\epsilon_1)}{r} - \epsilon.$$

Choose $\epsilon < \frac{b(\epsilon_1)}{r}$. By H2b, $\sum_l a_{m_l}\varphi_{m_l}(X_{m_l}) = +\infty$. Then

$$\sum_n a_n\varphi_n(X_n) = +\infty,$$

a contradiction with H1b. Thus $T(\omega) = 0$. ■

3. Theorems of almost sure convergence

Consider the process (X_n) as defined in section 1 :

$$\begin{aligned} X_{n+1} &= \Pi(X_n - A_nY_n) \\ E\left[A_nY_n \mid T_n\right] &= a_nM_n(X_n)\ a.s. \end{aligned}$$

Denote $d(x,B)$ a distance from $x \in \mathbb{R}^k$ to a subset B.

For all n, let F_n be a function from $\mathbb{R}^k$ into $\mathbb{R}^+$ twice continuously differentiable, with gradient G_n and hessian matrix H_n ; by the Taylor formula, there exists $0 < \mu_n < 1$ such that

$$F_n(X_n - A_nY_n) = F_n(X_n) - \langle G_n(X_n), A_nY_n\rangle + \frac{1}{2}\langle A_nY_n, H_n(X_n - \mu_nA_nY_n)A_nY_n\rangle$$

Denote $V_n = \frac{1}{2}E\left[\langle A_nY_n, H_n(X_n - \mu_nA_nY_n)A_nY_n\rangle \mid T_n\right]$.

Suppose:

(H4a) For all n, F_n is twice continuously differentiable

(H4b) For all $\epsilon > 0$, there exists $\nu(\epsilon) > 0$ and for all n, there exists a subset B_n of K such that

$$\inf_n \inf_{\{d(x,B_n)>\epsilon\}} F_n(x) > \upsilon(\epsilon)$$

(H4c) There exist two sequences of positive numbers (γ_n) and (δ_n) such that $\sum_1^\infty \gamma_n < \infty$, $\sum_1^\infty \delta_n < \infty$ and for all n and x,

$$F_{n+1}(\Pi x) \le (1+\delta_n)F_n(x) + \gamma_n$$

(H5) For all n, there exist two random variables D_n and E_n in $\mathbb{R}^+$, measurable with respect to T_n, such that

$$\begin{aligned} \sum_1^\infty D_n &< \infty, \sum_1^\infty E_n < \infty, \\ V_n &\le D_nF_n(X_n) + E_n \quad a.s. \end{aligned}$$

(H6) $\sum_1^\infty \langle G_n(X_n), a_n M_n(X_n)\rangle^- < \infty \quad a.s.$

(H7) For all $0 < \epsilon < 1$, $\sum_1^\infty a_n \inf_{\left\{x\in K, \epsilon < F_n(x) < \frac{1}{\epsilon}\right\}} \langle G_n(x), M_n(x)\rangle^+ < \infty$.

Remark that in the case where B_n is reduced to a single element θ of $\mathbb{R}^k$ not depending on n, if we take $F_n(x) = d^2(x,\theta) = \|x-\theta\|^2$, then assumptions H4a, b, c hold and $G_n(x) = 2(x-\theta)$, $H_n(x) = 2I$ (I : identity matrix), $V_n = E\left[\|A_n Y_n\|^2 \mid T_n\right]$.

Theorem 3 *Assume H4a, b, c, H5, H6 and H7 hold ; then $F_n(X_n) \longrightarrow 0$ and $d(X_n, B_n) \longrightarrow 0$ a.s.*

We use in the proof the Robbins-Siegmund lemma [4] :

Lemma 4 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and (T_n) an increasing sequence of sub-σ-algebras of $\mathcal{A}$. For $n \geq 1$, let z_n, β_n, ξ_n and ζ_n be non-negative T_n-measurable random variables such that $E[z_{n+1} \mid T_n] \leq z_n(1+\beta_n) + \xi_n - \zeta_n$. Suppose $\sum_1^\infty \beta_n < \infty$, $\sum_1^\infty \xi_n < \infty$ a.s. Then $\lim_{n\longrightarrow\infty} z_n$ exists and is finite and $\sum_1^\infty \zeta_n < \infty$ a.s.*

Proof. By H4a, c and H5, we have :

$$
\begin{aligned}
F_{n+1}(X_{n+1}) &\leq (1+\delta_n)F_n(X_n - A_nY_n) + \gamma_n. \\
E[F_{n+1}(X_{n+1}) \mid T_n] &\leq (1+\delta_n)(F_n(X_n) - \langle G_n(X_n), a_nM_n(X_n)\rangle + V_n) + \gamma_n \\
&\leq (1+\delta_n)(1+D_n)F_n(X_n) + (1+\delta_n)E_n \\
&\quad +(1+\delta_n)\langle G_n(X_n), a_nM_n(X_n)\rangle^- + \gamma_n \\
&\quad -(1+\delta_n)\langle G_n(X_n), a_nM_n(X_n)\rangle^+ \quad a.s.
\end{aligned}
$$

By H4c, H5 and H6, the assumptions of the preceding lemma hold ; then there exists a random variable T in $\mathbb{R}^+$ such that $F_n(X_n) \longrightarrow T$ a.s. and $\sum_1^\infty \langle G_n(X_n), a_nM_n(X_n)\rangle^+ < \infty$ a.s.

Let $\varphi_n(x) = \langle G_n(x), M_n(x)\rangle^+$. The assumptions H1a, b and H2a of lemma 1 hold. Then $F_n(X_n) \longrightarrow 0$ a.s.

By H4b, it follows that $d(X_n, B_n) \longrightarrow 0$ a.s. ■

Prove now a second theorem.

Suppose :

(H4d) For all $0 < \epsilon < 1$, $\sup_n \sup_{\left\{\epsilon < F_n(x) < \frac{1}{\epsilon}\right\}} \|G_n(x)\| < \infty$

(H4e) For all $\epsilon > 0$, there exists $\eta > 0$ such that
$(\|x_1 - x_2\| < \eta) \Rightarrow (\sup_n \|G_n(x_1) - G_n(x_2)\| < \epsilon)$

(H8a) For all $0 < \epsilon < 1$, $\sup_n \sup_{\left\{\epsilon < F_n(x) < \frac{1}{\epsilon}\right\}} \|M_n(x)\| < \infty$

(H8b) For all $\epsilon > 0$, there exists $\eta > 0$ such that $(\|x_1 - x_2\| < \eta) \Rightarrow (\sup_n \|M_n(x_1) - M_n(x_2)\| < \epsilon)$

(H8c) There exist a positive integer r, a sequence of integers (n_l), for all $0 < \epsilon < 1$ an integer $L(\epsilon)$ such that $n_{l+1} \leq n_l + r$ and

$$b(\epsilon) = \inf_{l > L(\epsilon)} \inf_{\left\{x \in K,\, \epsilon < F_{n_l}(x) < \frac{1}{\epsilon}\right\}} \sum_{j \in I_l} \langle G_j(x), M_j(x) \rangle^+ > 0$$

with $I_l = \{n_l, n_l + 1, ..., n_{l+1} - 1\}$

(H2b) $\sum_l \min_{j \in I_l} a_j = \infty$.

Remark that in the case where $B_n = \{\theta\}$ and $F_n(x) = \|x - \theta\|^2$, assumptions H4d, e hold.

Theorem 5 *Assume H2b, H4a, b, c, d, e, H5, H6, H8a, b, c hold ; then in the set* $\{A_n Y_n \longrightarrow 0\}$, $F_n(X_n) \longrightarrow 0$ *and* $d(X_n, B_n) \longrightarrow 0$ *a.s.*

Proof. Following the proof of theorem 3, we have by H4a, c, H5, H6 : $F_n(X_n) \longrightarrow T$ and $\sum_1^\infty \langle G_n(X_n), a_n M_n(X_n) \rangle^+ < \infty$ *a.s.*

Apply lemma 2 with $\varphi_n(x) = \langle G_n(x), M_n(x) \rangle^+$.

H1a, b and H3b hold. H1c holds in the set $\{A_n Y_n \longrightarrow 0\}$ as

$$\|X_{n+1} - X_n\| = \|\Pi(X_n - A_n Y_n) - \Pi X_n\| \leq \|X_n - A_n Y_n - X_n\| = \|A_n Y_n\|$$

As $|a^+ - b^+| \leq |a - b|$, we have :

$$\begin{aligned} |\varphi_n(x_1) - \varphi_n(x_2)| &\leq |\langle G_n(x_1), M_n(x_1) \rangle - \langle G_n(x_2), M_n(x_2) \rangle| \\ &\leq |\langle G_n(x_1), M_n(x_1) - M_n(x_2) \rangle| \\ &\quad + |\langle G_n(x_1) - G_n(x_2), M_n(x_2) - M_n(x_1) \rangle| \\ &\quad + |\langle G_n(x_1) - G_n(x_2), M_n(x_1) \rangle| . \end{aligned}$$

By H4d, e and H8a, b, assumption H3a holds.

Then $F_n(X_n) \longrightarrow 0$ *a.s.* By H4b, $d(X_n, B_n) \longrightarrow 0$ *a.s.* ∎

4. Application to the estimation of a quantile interval

Let Z be a real random variable whose distribution function $F(t) = P(Z < t)$ is unknown. Suppose that there exists an interval (a, b), which is eventually reduced to a single point, such that : $F(t) = \alpha \Leftrightarrow t \in (a, b)$.

Let $m \geq 1$ be an integer and $(Z_{nj}, n \geq 1, j = 1, ..., m)$ a set of mutually independent random variables which have the same law as Z. For all x, define the random variables $I_{nj}(x)$ and $F_{nm}(x)$ such that :

$$\begin{aligned} I_{nj}(x) &= 1 \textit{ if } Z_{nj} < x,\ I_{nj}(x) = 0 \textit{ otherwise} \\ F_{nm}(x) &= \frac{1}{m} \sum_{j=1}^{m} I_{nj}(x). \end{aligned}$$

Then $E\left[F_{nm}(x)\right] = E\left[I_{nj}(x)\right] = F(x)$.

Define the stochastic approximation process (X_n) such that

$$X_{n+1} = X_n - a_n(F_{nm}(X_n) - \alpha).$$

If z_{nj} is the observed value of Z_{nj} and x_n the value of X_n, $F_{nm}(x_n)$ is the proportion of elements of $\{z_{n1}, ..., z_{nm}\}$ which are smaller than x_n.

Suppose :

(H2b') $\sum_1^\infty a_n = \infty$
(H2c) $\sum_1^\infty a_n^2 < \infty$.

Theorem 6 *Let* $d(x, (a, b)) = \inf_{\{y \in (a,b)\}} |x - y|$. *Assume H2b', c hold ; then* $d(X_n, (a, b)) \longrightarrow 0$ *a.s.*

Proof. Define the function f such that

$$\begin{aligned} f(x) &= (x-a)^2 \text{ if } x < a \\ f(x) &= 0 \text{ if } a \leq x \leq b \\ f(x) &= (x-b)^2 \text{ if } x > b. \end{aligned}$$

H4a, b, c hold for $F_n = f$ and $B_n = (a, b)$.
$|f''(x)| \leq 2$, $|F_{nm}(x) - \alpha| \leq 1$; then $V_n \leq a_n^2$; H5 holds.
$M_n(X_n) = E\left[F_{nm}(X_n) - \alpha \mid T_n\right] = F(X_n) - \alpha$;
$f'(x)(F(x) - \alpha) \geq 0$, $\inf_{\left\{\epsilon < f(x) < \frac{1}{\epsilon}\right\}} f'(x)(F(x) - \alpha) > 0$; H6 and H7 hold.
Applying theorem 3 gives $d(X_n, (a, b)) \longrightarrow 0$ *a.s.* ■

5. Application to linear regression under convex constraints

Consider a sequence (Z_n) of observable mutually independent real random variables.

Suppose that there exist an unknown vector θ in $\mathbb{R}^k$, for all n a known vector b_n in $\mathbb{R}^k$ and a real random variable R_n with $E\left[R_n\right] = 0$ such that

$$Z_n = b_n'\theta + R_n.$$

Suppose moreover that θ belongs to a non-empty closed convex set K of $\mathbb{R}^k$. For instance :

1) $\|\theta\|$ is bounded ;

2) the components of θ are non-negative.

Consider the stochastic approximation process (X_n) such that :

$$X_{n+1} = \Pi\left(X_n - a_n\frac{b_n}{\|b_n\|^2}(b_n'X_n - Z_n)\right).$$

Suppose :

(H2b)$\sum_1^\infty \min_{j\in I_l} a_j = \infty$

(H2c) $\sum_1^\infty a_n^2 < \infty$

(H2d) $\sum_1^\infty a_n^2 \frac{E\left[R_n^2\right]}{\|b_n\|^2} < \infty$

(H9) $\lambda = \inf_l \lambda_{\min}\left(\sum_{j\in I_l}\frac{b_j b_j'}{\|b_j\|^2}\right) > 0.$

Theorem 7 *Assume H2b, c, d and H9 hold ; then $X_n \longrightarrow \theta$ a.s.*

This theorem completes in the case of linear regression results of Albert and Gardner [1] (p. 103, conjectured theorem).

Proof. Let $Y_n = b_n'X_n - Z_n = b_n'(X_n - \theta) - R_n$ and $A_n = a_n\frac{b_n}{\|b_n\|^2}$.

As $E\left[R_n \mid T_n\right] = E\left[R_n\right] = 0$, $M_n(X_n) = \frac{b_n b_n'}{\|b_n\|^2}(X_n - \theta)$ *a.s.*
Remark that, for fixed n, equation $M_n(x) = 0$ has an infinity of solutions.
Denote I an identity matrix. Define $F_n(x) = \|x - \theta\|^2$; then :

$$G_n(x) = 2(x-\theta), H_n(x) = 2I, V_n = E\left[a_n^2\left\|Y_n\right\|^2 \mid T_n\right].$$

Assumptions H4a, b, c, d, e, H6, H8a, b hold.

$$V_n = E\left[a_n^2\left\|Y_n\right\|^2 \mid T_n\right] = a_n^2\left\|X_n - \theta\right\|^2 + a_n^2\frac{E\left[R_n^2\right]}{\|b_n\|^2}.$$

By H2d, assumption H5 holds.
By H9, assumption H8c holds as

$$\begin{aligned}\sum_{j\in I_l}\langle G_j(x), M_j(x)\rangle^+ &= 2\sum_{j\in I_l}\left\langle x-\theta, \frac{b_j b_j'}{\|b_j\|^2}(x-\theta)\right\rangle \\ &\geq 2\lambda\left\|x-\theta\right\|^2.\end{aligned}$$

Furthermore, as $E\left[R_n \mid T_n\right] = 0$:

$$\begin{aligned}
E\left[\|X_{n+1}-\theta\|^2 \mid T_n\right] &= \|X_n-\theta\|^2 + a_n^2 E\left[\|Y_n\|^2 \mid T_n\right] \\
&\quad -2a_n \left\langle X_n-\theta, \frac{b_n b_n'}{\|b_n\|^2}(X_n-\theta)\right\rangle \\
&\leq \left(1+a_n^2\right)\|X_n-\theta\|^2 + a_n^2 \frac{E\left[R_n^2\right]}{\|b_n\|^2}. \\
E\left[\|X_{n+1}-\theta\|^2\right] &\leq (1+a_n^2) E\left[\|X_n-\theta\|^2\right] + a_n^2 \frac{E\left[R_n^2\right]}{\|b_n\|^2}.
\end{aligned}$$

By H2c, d, there exists $t \geq 0$ such that $E\left[\|X_n-\theta\|^2\right] \longrightarrow t$. Then :

$$\begin{aligned}
\sum_1^\infty E\left[a_n^2 \|Y_n\|^2\right] &= \sum_1^\infty \left(a_n^2 E\left[\|X_n-\theta\|^2\right] + a_n^2 \frac{E\left[R_n^2\right]}{\|b_n\|^2}\right) < \infty \ ; \\
\sum_1^\infty a_n^2 \|Y_n\|^2 &< \infty \ a.s. \ ; \ a_n Y_n \longrightarrow 0 \ a.s.
\end{aligned}$$

Applying theorem 5 gives $X_n \longrightarrow \theta \ a.s.$ ■

References

[1] A.E. Albert and L.A. Gardner, *Stochastic Approximation and Nonlinear Regression,* Research Monograph 42, The M.I.T. Press, Cambridge, Massachusetts (1967).

[2] A. Bennar, *Approximation stochastique, convergence dans le cas de plusieurs solutions et étude de modèles de corrélation*, thèse de Doctorat de troisième cycle, Université de Nancy 1 (1985).

[3] J.M. Monnez, Almost sure convergence of stochastic gradient processes with matrix step sizes, *Statistics and Probability Letters,* **76** (2006), 531-536.

[4] H. Robbins and D. Siegmund, A convergence theorem for nonnegative almost supermartingales and some applications, *Optimizing methods in Statistics, J.S. Rustagi ed.*, Academic Press, New-York (1971), 233-257.