BULL. AUSTRAL. MATH. SOC. 62L20, 60F15
VOL. 34 (1986) 335-342

ON THE ALMOST SURE CONVERGENCE OF A GENERAL STOCHASTIC APPROXIMATION PROCEDURE

S.N. EVANS AND N.C. WEBER

A set of conditions for the almost sure convergence of a stochastic iterative procedure is given. The conditions are framed in terms of the behaviour of the random adjustment made at the n-th step rather than in terms of some underlying regression model.

1. Introduction

Iterative techniques for finding approximations to the roots, points at which extrema occur, or maxima and minima of known functions have a very long and extensive history. Stochastic approximation methods were proposed as a means of solving such problems when we have only scant information about the general nature of the function of interest and our "observations" of the function are clouded by experimental noise or error. Such procedures appear to have been first discussed in a paper by Hotelling [7]. However the seminal papers in the area of stochastic approximation are Robbins and Monro [10], which discusses a stochastic approximation technique for estimating the roots of a regression function, and Kiefer and Wolfowitz [8], which is concerned with procedures for finding minima and

Received 20 December, 1985.

maxima of regression functions.

The area of stochastic approximation developed rapidly following these two papers and the reader is referred to Wasan [14] for a coverage of the limit results and modifications to the basic schemes developed in the eighteen years following the Robbins and Monro paper. Wasan also contains an extensive bibliography for this period. In recent years the papers by Goodsell and Hanson [5], Ljung [9], Ruppert [11] and Solo [12] have relaxed the conditions required for the convergence of stochastic approximation procedures and widened the areas in which these iterative estimation techniques can be applied.

In this note a general iterative procedure is given which contains the A_o class of stochastic approximation methods proposed by Burkholder [2]. The A_o class includes the Robbins-Monro procedures and Kiefer-Wolfowitz procedures as well as extreme point estimation techniques like those given by Friedman [4].

We will consider the strong convergence of this general procedure. In contrast to the usual approach, our conditions for convergence are not framed in terms of a fixed underlying regression function but are written so that the emphasis is on the iterative procedure itself. Thus the result here applies to a wide range of situations and, for example, covers stochastic approximation techniques applied when the observed process is based on a sequence of regression functions which are themselves converging to some limiting function.

2. Notation and Result

Let $\{V_n\}$ be a sequence of random variables defined on some probability space (Ω, F, P) and define an iterative scheme $\{X_n\}$ by first taking an initial random variable X_o defined on (Ω, F, P) and setting

$$X_{n+1} = X_n - V_n \qquad n = 0, 1, \ldots .$$

We will be interested in the conditions on the variables $\{V_n\}$ that ensure the almost sure convergence of $\{X_n\}$.

https://doi.org/10.1017/S0004972700010236 Published online by Cambridge University Press

Define the σ-fields $F_n = \sigma\{X_0, X_1, \ldots, X_{n+1}\}$, $n = 0. 1, \ldots$.

Let

$$M_n = E(V_n | F_{n-1}) \qquad n = 1, 2, \ldots$$

and

$$Z_n = V_n - M_n .$$

THEOREM. *Under conditions (i) - (iv) below* X_n *converges to 0 with probability one.*

(i) For each $\epsilon > 0$,

$$P(|X_n| > \epsilon, X_n M_n < 0 \text{ infinitely often}) = 0 .$$

(ii) $$|M_n|(1 + |X_n|)^{-1} \to 0 \text{ almost surely}$$

(iii) There exists some p, $1 < p \le 2$ *such that*

$$\Sigma_{n=1}^{\infty} E|Z_n|^p < \infty .$$

(iv) $P(0 < \lim\inf|X_n| \le \lim\sup|X_n| < \infty, \Sigma_{n=1}^{\infty}|M_n| < \infty) = 0$.

The above conditions are framed in terms of X_n and the conditional expectation of the adjustment at time n, M_n . Each condition addresses a separate feature of the process. Condition (i) states that if X_n is away from 0 then X_n and the expectation of the 'correction term', given the past history, have the same sign ensuring the correction term pushes the $\{X_n\}$ process towards 0 .

Condition (ii) ensures the regressed corrections are not too large and so prevents wild oscillations in the $\{X_n\}$ process. Dvoretzky [3] gives an example based on the Robbins-Monro procedures, where M_n is a polynomial in X_n , to show that a condition of this type is necessary.

Condition (iii) controls the oscillations caused by the noise term, the V_n terms cannot vary too wildly, while condition (iv) ensures the $\{X_n\}$ process cannot get caught away from 0 . Condition (iv) states that if the process is away from 0 then the regressed corrections are large enough to force a return to zero, provided the adjustments are in the right

https://doi.org/10.1017/S0004972700010236 Published online by Cambridge University Press

direction.

If $\{a_n\}$ is a sequence of positive constants converging to 0 as $n \to \infty$, writing $V_n = a_n \tilde{V}_n$, for some sequence of random variables $\{\tilde{V}_n\}$, it is easy to see that the above formulation includes the A_o processes introduced by Burkholder. If $X_{n+1} = X_n - a_n \tilde{V}_n$, $n = 0, 1, \ldots$ and $\tilde{M}_n = E(\tilde{V}_n | F_{n-1})$ then condition (ii) may be written as

(ii') $\limsup |\tilde{M}_n| (1 + |X_n|)^{-1} < \infty$ almost surely.

If for each $\epsilon > 0$ there is some $\delta > 0$, such that $|X_n| > \epsilon$ implies $|\tilde{M}_n| > \delta$, then condition (iv) can be replaced by the much simpler condition

(iv') $\Sigma_{n=1} a_n = \infty$.

The proof of the theorem essentially follows that given by Blum [1] for the strong convergence of the Robbins-Monro method. The proof has to be modified substantially because we do not insist on a fixed underlying regression model to generate the V_n process. Before proving the theorem we will establish the following Lemma.

LEMMA.. *Under conditions (i) - (iii) of the theorem,* $\{X_n\}$ *converges almost surely to some random variable* X.

Proof. Write $X_{n+1} = X_n - Z_n - M_n$ so

$$X_{n+1} + \Sigma_{j=1}^{n} M_j = X_1 - \Sigma_{j=1}^{n} Z_j .$$

Clearly $\{\Sigma_{j=1}^{n} Z_j, F_n\}$ is a martingale. From Von Bahr and Esseen [13],

$$E(|\Sigma_{j=1}^{n} Z_j|^p) \le 2\ \Sigma_{j=1}^{n} E|Z_j|^p < \infty \quad ,$$

using (iii). Thus for some $p \in (1,2]$,

$$\sup E(|\Sigma_{j=1}^{n} Z_j|^p) < \infty$$

and so by the martingale convergence theorem (see, for example, Hall and Heyde [6], p. 17), $X_1 - \Sigma_{j=1}^{n} Z_j$ and hence $X_{n+1} + \Sigma_{j=1}^{n} M_j$ converges almost surely to an integrable random variable as $n \to \infty$.

https://doi.org/10.1017/S0004972700010236 Published online by Cambridge University Press

Next we show that

$$P(\lim X_n = \infty) = P(\lim X_n = -\infty) = 0 \ .$$

Note

$$\{X_n \to \infty\} \subset [\{X_n \to \infty\} \cap \ \liminf \{M_n \geq 0\}] \cup$$

$$[\liminf \{X_n > 1\} \cap \limsup \{M_n < 0\}] \ .$$

Now

$$P(\{X_n \to \infty\} \cap \liminf \{M_n \geq 0\})$$

$$\leq P(X_{n+1} + \textstyle\sum_{j=1}^{n} M_j \to \infty)$$

$$= 0 \ , \tag{1}$$

since $X_{n+1} + \sum_{j=1}^{n} M_j$ converges almost surely. Further, if $\omega \in \liminf \{X_n > 1\} \cap \limsup \{M_n < 0\}$ then we can choose $m_1 < m_2 < \ldots$ such that $M_{m_k}(\omega) < 0$ and $X_{m_k}(\omega) > 1$ for all k. But this implies

$$X_{m_k} M_{m_k} < 0$$

and so from (i)

$$P(\liminf \{X_n > 1\} \cap \limsup \{M_n < 0\}) = 0 \ . \tag{2}$$

Thus from (1) and (2) we can conclude $P(\lim X_n = \infty) = 0$.

A similar argument shows that $P(\lim X_n = -\infty) = 0$.

We complete the proof that $\{X_n\}$ converges almost surely, as in Blum [1], by showing that for any $a, b \in \mathbb{R}$, $a < b$

$$P(\liminf X_n < a < b < \limsup X_n) = 0 \ . \tag{3}$$

By symmetry it suffices to consider the two cases

I $\qquad 0 < c < \liminf X_n < a < b < \limsup X_n \ ,$

and

II $\qquad \liminf X_n \leq 0 < b < \limsup X_n \ .$

Case I. Using (i) and the result that $\{X_{n+1} + \sum_{j=1}^{n} M_j\}$ converges almost surely, we can choose $m > n$ sufficiently large so that

$$|X_m - X_n + \textstyle\sum_{j=n}^{m-1} M_j| \leq (b-a)/3 \ ; \tag{4}$$

$$c < X_n < a \ , \ X_m > b \ \text{ and } \ a \leq X_j \leq b \ , \ n < j < m \ ; \tag{5}$$

$$\text{and } M_j > 0 \ \text{ for } \ n \leq j \leq m \ . \tag{6}$$

https://doi.org/10.1017/S0004972700010236 Published online by Cambridge University Press

From (4) and (6) we have

$$X_m - X_n \le (b-a)/3 - \sum_{j=n}^{m-1} M_j \le (b-a)/3$$

but from (5), $X_m - X_n > (b-a)$ and so we have a contradiction. Thus

$$P(0 < c < \liminf X_n < a < b < \limsup X_n) = 0$$

Case II. Given b, choose δ such that $0 < \delta < b/(3 + 2b)$.

Again using (i), (ii) and the result that

$\{X_{n+1} + \sum_{j=1}^{n} M_j\}$ converges almost surely, choose

$m > n$ sufficiently large such that

$$|X_m - X_n + \sum_{j=n}^{m-1} M_j| \le (b/3) ; \tag{7}$$

$$X_n < (b/3) , X_m > b \text{ and } (b/3) \le X_j \le b , n < j < m ; \tag{8}$$

$$M_j > 0 , n < j < m ; \tag{9}$$

and

$$|M_n|(1 + |X_n|)^{-1} < \delta . \tag{10}$$

From (7), (9) and (10),

$$X_m - X_n \le (b/3) - M_n < (b/3) + \delta(1 + |X_n|) .$$

If $|X_n| < (b/3)$ then

$$X_m - X_n \le (b/3) + \delta(1 + b/3) < (2b/3)$$

contradicting the construction in (8).

If $X_n < -(b/3)$ then

$$\begin{aligned} X_m - X_n &\le (b/3) + \delta(1 + |X_n|) \\ &\le (b/3) + \delta(1 + X_m - X_n) \end{aligned}$$

so

$$X_m - X_n < (2b/3)$$

again contradicting (8). Thus

$$P(\liminf X_n \le 0 < b < \limsup X_n) = 0,$$

completing the proof of the lemma.

https://doi.org/10.1017/S0004972700010236 Published online by Cambridge University Press

Proof of the Theorem. From the lemma we have that X_n converges to some random variable X almost surely. To conclude $P(X = 0) = 1$ it suffices to show that for any $a, b \in \mathbb{R}$ with $0 < a < b$ or $a < b < 0$,

$$P(\liminf \{X_n \in [a,b]\}) = 0$$

Suppose $0 < a < b$. Fix n and consider the set $\bigcap_{m\geq n}\{X_m \in [a,b]\}$. If $\omega \in \bigcap_{m\geq n}\{X_m \in [a,b]\}$ then $|X_m(\omega)| > a$, $m\geq n$ and so from (i) and (iv) $\sum_{j=1}^{\infty} M_j(\omega) = \infty$, since $M_j(\omega) \geq 0$ for all but at most a finite number of $j \geq n$. But this implies

$$X_{n+1}(\omega) + \sum_{j=1}^{n} M_j(\omega) \to \infty \quad \text{as} \quad n \to \infty \quad ,$$

so $P(\bigcap_{m\geq n}\{X_m \in [a,b]\}) = 0$ as $X_{n+1} + \sum_{j=1}^{n} M_j$ converges with probability one. Thus for $0 < a < b$

$$P(\liminf \{X_n \in [a,b]\}) = 0 \ .$$

A similar argument handles the case $a < b < 0$ and completes the proof.

References

[1] J.R. Blum, "Approximation methods that converge with probability one", *Ann. Math. Statist.* 25 (1954), 382-386.

[2] D.L. Burkholder, "On a class of stochastic approximation procedures", *Ann. Math. Statist.* 27 (1956), 1044-1059.

[3] A. Dvoretzky, "On stochastic approximation", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I (1956), 39-55.

[4] S. Friedman, "On stochastic approximations", *Ann. Math. Statist.*, 34 (1963), 343-346.

[5] C.A. Goodsell and D.L. Hanson, "Almost sure convergence for the Robbins-Monro process", *Ann. Probab.* 4 (1976), 890-901.

[6] P. Hall and C.C. Heyde, *Martingale Limit Theory and its Application* (Academic Press, New York, 1980).

[7] H. Hotelling, "Experimental determination of the maximum of a function", *Ann. Math. Statist.* 12 (1941), 20-46.

[8] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function", *Ann. Math. Statist.* 23 (1952), 462-466.

[9] L. Ljung, "Strong convergence of a stochastic approximation algorithm", *Ann. Statist.* 6 (1978), 680-696.

[10] H. Robbins and S. Monro, "A stochastic approximation method", *Ann. Math. Statist.* 22 (1951), 400-407.

[11] D. Ruppert, "Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise", *Ann. Probab.* 10 (1982), 178-187.

[12] V. Solo, "Stochastic approximation with dependent noise", *Stoch. Processes and their Applications*, 13 (1982), 157-170.

[13] B. VonBahr and C.G. Esseen, "Inequalities for the r^{th} absolute moment of a sum of random variables, $1 \leq r \leq 2$", *Ann. Math. Statist.* 36 (1965), 299-303.

[14] M.T. Wasan, *Stochastic approximation*, (Cambridge University Press, 1969).

S.N. Evans,
Statistical Laboratory,
University of Cambridge,
16 Mill Lane,
Cambridge,
England.

N.C. Weber,
Department of Mathematical Statistics,
University of Sydney,
New South Wales 2006
Australia.