Statistica Sinica 4(1994), 361-372

ALMOST SURE CONVERGENCE OF STOCHASTIC APPROXIMATION PROCEDURES

Gang Li

Ohmeda

Abstract: In this paper we investigate the convergence of the Robbins-Monro procedure $X_{n+1} = X_n - a_n(Y_n - \alpha)$. The following along with some related results are obtained.

Let $\xi_j = Y_j - M(X_j)$ be the error in the jth observation. A necessary and sufficient condition for the almost sure convergence of $\{X_n\}$ is

$$\lim_{n\to\infty} \sup_{k\geq n} \frac{\left|\sum_{j=n}^{k} a_j\xi_j\right|}{1+\sum_{j=n}^{k} a_j} = 0 \quad \text{a.s.}$$

If $\{\xi_j\}$ is an i.i.d. sequence, $p \geq 1$, $E\xi_j = 0$, and $a_j = j^{-\frac{1}{p}}$ for $j \geq 1$, then the above is true if and only if $E|\xi_1|^p < \infty$.

Key words and phrases: Robbins-Monro procedure, almost sure convergence, martingale differences.

1. Introduction

Let M be an unknown real valued function defined on the real line and suppose that at each real number x unbiased observations of $M(x)$ can be taken. Suppose θ is the unique root of $M(x) = \alpha$ for some known α and one wishes to find θ. Robbins and Monro (1951) suggested a recursive procedure for approximating θ. In the Robbins-Monro (RM) procedure one starts at some X_1 and recursively defines X_n, the estimate of θ at the nth step, by

$$X_{n+1} = X_n - a_n(Y_n - \alpha),$$

where Y_n is observed at X_n, i.e., it is an unbiased observation of $M(X_n)$. The convergence of $\{X_n\}$ to θ in various modes has been studied extensively by many authors. Many results are presented in Nevel'son and Has'minskiĭ (1976).

Most of the results on the RM-procedure suggest that the convergence of X_n to θ is associated with the errors of observation $\xi_n = Y_n - M(X_n)$. Blum

(1954) and Krasulina (1969) gave several results on almost sure convergence of the RM procedure, requiring almost sure convergence of $\sum_{n=1}^{\infty} a_n\xi_n$. Goodsell and Hanson (1976) obtained a result stating that X_n can converge almost surely even though the series above does not do so. Dvoretzsky (1956) and Krasulina (1969) showed that, for $p \in (1,2]$, $E|X_n - \theta|^p \to 0$ if $\sum_{n=1}^{\infty} a_n^p|\xi_n|^p < \infty$.

In this paper the relation between the errors of observation and convergence of X_n to θ is obtained. In Section 2, a necessary and sufficient condition is given which, for a given choice of $\{a_n\}$, states the relation between the errors of observation and the almost sure convergence of X_n to θ. Mean-square convergence is also considered there. In Section 3, almost sure convergence of martingale difference sequences is considered. Section 4 contains proofs of the results in Section 2. Section 5 contains proofs of the results in Section 3.

2. Convergence of Robbins-Monro Procedure

Let M be a real valued function defined on the real line. Suppose M is Borel-measurable and satisfies the following conditions:

(M1) $$|M(x) - \alpha| < c + d|x| \quad \text{for some } c > 0 \text{ and } d > 0,$$

(M2) $$\psi(\delta) = \inf_{|x-\theta|\geq\delta} |M(x) - \alpha| > 0 \quad \text{for every } \delta > 0, \text{ and}$$

(M3) $$M(x) < \alpha \quad \text{for} \quad x < \theta \quad \text{and} \quad M(x) > \alpha \quad \text{for} \quad x > \theta.$$

Let $\{a_n\}$ be a sequence of positive numbers such that

(A) $$a_n \to 0 \quad \text{as} \quad n \to \infty \quad \text{and} \quad \sum_{n=1}^{\infty} a_n = \infty.$$

Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that X_1 is arbitrary,

$$X_{n+1} = X_n - a_n(Y_n - \alpha) \quad \text{for} \quad n \geq 1, \quad \text{and} \tag{2.1a}$$

$$E\{Y_n|X_1,\ldots,X_n;Y_1,\ldots,Y_{n-1}\} = E\{Y_n|X_n\} = M(X_n) \quad \text{a.s. for } n \geq 1. \tag{2.1b}$$

Let $\xi_j = Y_j - M(X_j)$.

Theorem 2.1. *Suppose M satisfies* (M1)-(M3), *$\{a_n\}$ satisfies* (A), *and $\{X_n\}$ is defined as in* (2.1). *If*

$$\lim_{n\to\infty} \sup_{k\geq n} \frac{\left|\sum_{j=n}^{k} a_j\xi_j\right|}{1 + \sum_{j=n}^{k} a_j} = 0 \quad \text{a.s.} \tag{2.2}$$

then

$$X_n \to \theta \quad \text{a.s. } as \quad n \to \infty. \tag{2.3}$$

Furthermore, if M is continuous at θ then (2.3) *implies* (2.2).

It is obvious that (2.2) is true if the series $\sum_{j=1}^{\infty} a_j \xi_j$ converges almost surely, a condition required in most of the literature. In Section 3, we investigate the conditions under which (2.2) holds. For mean square convergence we have the following result which is similar to Theorem 2.1.

Theorem 2.2. *Let $\{X_n\}$ be defined as in* (2.1) *and suppose $\{a_n\}$ satisfies* (A). *If M satisfies* (M1), (M3), *and*

(M2′) $\quad |M(x) - \alpha| \geq K|x - \theta| \quad$ *for some* $K > 0$ *and all* x,

and if

$$\lim_{n\to\infty} \sup_{k \geq n} \frac{\sum_{j=n}^{k} a_j^2 E\xi_j^2}{1 + \sum_{j=n}^{k} a_j} = 0, \tag{2.4}$$

then $E(X_n - \theta)^2 \to 0$. Conversely, if M satisfies (M3) *and*

(M1′) $\quad |M(x) - \alpha| \leq K|x - \theta| \quad$ *for some* $K > 0$ *and all* x,

and if $E(X_n - \theta)^2 \to 0$, then (2.4) *holds.*

3. Almost Sure Convergence Theorems for Martingale Difference Sequences

Let $\mathcal{F}_n$ be the sigma-algebra generated by $X_1, \ldots, X_n$; $Y_1, \ldots, Y_{n-1}$.

Theorem 3.1. *Suppose $\{\xi_n, \mathcal{F}_n, n \geq 1\}$ is a martingale difference sequence, $p \geq 1$,*

$$P\{|\xi_n| > t | \mathcal{F}_{n-1}\} < \Phi(t) \quad \text{a.s. } for \quad n = 1, 2, \ldots, \tag{3.1}$$

$$\lim_{t\to\infty} \Phi(t) = 0, \quad and \quad \int_0^{\infty} t^p |d\Phi(t)| < \infty. \tag{3.2}$$

Then

$$\lim_{n\to\infty} \sup_{k \geq n} \frac{\left|\sum_{j=n}^{k} j^{-\frac{1}{p}} \xi_j\right|}{1 + \sum_{j=n}^{k} j^{-\frac{1}{p}}} = 0 \quad \text{a.s.} \tag{3.3}$$

Remark 3.1. Under the condition of Theorem 3.1, when $p \in (0, 2)$, the series $\sum_1^{\infty} n^{-1/p} \xi_n$ converges almost surely (cf. the proof of Theorem 3.2.3 of Stout (1974)), so (3.3) holds. But for $p \geq 2$, the series $\sum_1^{\infty} n^{-1/p} \xi_n$ may diverge almost surely, even when ξ_n are i.i.d., for example, $P\{\xi_1 = \pm 1\} = 1/2$.

Remark 3.2. If ξ_n are i.i.d. then (3.3) implies that $E|\xi_1|^p < \infty$. In fact (3.3) implies $n^{-1/p}\xi_n \to \infty$ a.s., and consequently $E|\xi_1|^p < \infty$.

Theorem 3.2. *Suppose* $\{\xi_n, \mathcal{F}_n, n \geq 1\}$ *is a martingale difference sequence,* $p > 2$, *and* $E|\xi_n|^p < C < \infty$ *for all* n. *Then for some* $\delta > \frac{2}{p+2}$,

$$\lim_{n\to\infty} \sup_{k\geq n} \frac{\left|\sum_{j=n}^{k} j^{-2/(p+2)}(\log j)^{-\delta}\xi_j\right|}{1+\sum_{j=n}^{k} j^{-2/(p+2)}(\log j)^{-\delta}} = 0 \quad \text{a.s.} \tag{3.4}$$

4. Proofs for Results in Section 2

For notational convenience assume that $\alpha = \theta = 0$, continue to let $\xi_n = Y_n - M(X_n)$, and denote $x \wedge y = \min(x, y)$. Our proofs are corollaries to the following lemma.

Lemma 4.1. *Suppose* M *satisfies* (M1)-(M3), $\{a_n\}$ *satisfies* (A), *and* $\{x_n\}$ *and* $\{b_n\}$ *are sequences of real numbers.*
(i) *If*

$$x_{n+1} \leq x_n - a_n M(x_n) + a_n b_n \quad \textit{for all} \quad n \geq 1 \quad \textit{and} \tag{4.1}$$

$$\limsup_{n\to\infty} \sup_{k\geq n} \frac{\sum_{j=n}^{k} a_j b_j}{1+\sum_{j=n}^{k} a_j} \leq 0, \tag{4.2}$$

then

$$\limsup_{n\to\infty} x_n \leq 0. \tag{4.3}$$

(ii) *If* M *is continuous at* 0,

$$x_{n+1} \geq x_n - a_n M(x_n) + a_n b_n \quad \textit{for all} \quad n \geq 1, \quad \textit{and} \tag{4.4}$$

$$\lim_{n\to\infty} x_n = 0, \tag{4.5}$$

then (4.2) *holds.*

Proof of Lemma 4.1. *Proof of* (i)*:* Suppose (4.1) and (4.2) hold. Let $\delta = \liminf_{n\to\infty} x_n$. First we show that $\delta \leq 0$. If $\delta > 0$ then there exists a positive integer N such that $x_n > \frac{\delta}{2}$ for all $n > N$. Note that $x_{n+1} \leq -\sum_{j=1}^{n} a_j(M(x_j) - b_j) + x_1$ and recall the assumption that $\sum_1^\infty a_n = \infty$. Then from (M2) and (M3) we have

$$\begin{aligned} \liminf_{n\to\infty} \frac{\sum_{j=1}^{n} a_j b_j}{\sum_{j=1}^{n} a_j} &\geq \liminf_{n\to\infty} \frac{x_{n+1} + \sum_{j=1}^{n} a_j M(x_j) - x_1}{\sum_{j=1}^{n} a_j} \\ &\geq \psi\left(\frac{\delta}{2}\right) > 0. \end{aligned} \tag{4.6}$$

However (4.2) and the assumption that $\sum a_n = \infty$ give

$$\limsup_{n\to\infty} \frac{\sum_{j=1}^{n} a_j b_j}{\sum_{j=1}^{n} a_j} \leq \limsup_{m\to\infty} \sup_{n\geq m} \frac{\sum_{j=m}^{n} a_j b_j}{\sum_{j=1}^{n} a_j} \leq 0 \tag{4.7}$$

which contradicts (4.6). Thus $\delta \leq 0$.

Suppose $\limsup_{n\to\infty} x_n > b > a > 0$. From (4.2) there exists an N such that

$$\sup_{k\geq n} \frac{\sum_{j=n}^{k} a_j b_j}{1 + \sum_{j=n}^{k} a_j} < \frac{\psi(a)}{3} \wedge \frac{b-a}{3} \quad \text{for all} \quad n > N. \tag{4.8}$$

Choose m and n such that $N < n < m$,

$$\begin{cases} x_n < a \quad \text{and} \quad x_m > b, \\ a \leq x_j \leq b \quad \text{for } n < j < m, \text{ and} \\ a_n < \dfrac{1}{3d} \wedge \dfrac{b-a}{3(c+\psi(a))}. \end{cases} \tag{4.9}$$

It follows from (4.1), (4.8), (4.9), and the assumption following (i) that

$$\begin{aligned} b - a \ < \ x_m - x_n &\leq - \sum_{j=n}^{m-1} a_j M(x_j) + \sum_{j=n}^{m-1} a_j b_j \\ &\leq \ - \sum_{j=n}^{m-1} a_j \psi(a) + \sum_{j=n}^{m-1} a_j b_j + a_n \psi(a) - a_n M(x_n) \\ &\leq \ \frac{\psi(a)}{3} \wedge \frac{b-a}{3} + a_n \psi(a) - a_n M(x_n). \end{aligned} \tag{4.10}$$

If $x_n \geq 0$, then $M(x_n) \geq 0$, so (4.10) and (4.9) give the contradiction

$$b - a < \frac{b-a}{3} + a_n \psi(a) < \frac{2}{3}(b-a).$$

If $x_n < 0$, then (M1) gives $|M(x_n)| \leq c + d|x_n| \leq c + d(x_m - x_n)$; thus (4.10) and (4.9) give

$$x_m - x_n < \frac{b-a}{3} + a_n(d(x_m - x_n) + c + \psi(a)) < \frac{2}{3}(b-a) + \frac{1}{3}(x_m - x_n)$$

which contradicts $b - a < x_m - x_n$ from (4.9).

Proof of (ii): Suppose M is continuous at θ and that (4.4) and (4.5) hold. Suppose

$$\limsup_{n\to\infty}\sup_{l\geq n}\frac{\sum_{j=n}^{l} a_j b_j}{1+\sum_{j=n}^{l} a_j} > 0. \tag{4.11}$$

Then there exists an $\varepsilon > 0$ and subsequences $\{n_k\}$ and $\{m_k\}$ such that for all k we have $n_k \leq m_k$ and

$$\frac{\sum_{j=n_k}^{m_k} a_j b_j}{1+\sum_{j=n_k}^{m_k} a_j} \geq \varepsilon. \tag{4.12}$$

Choose δ so that $\sup_{|x|\leq\delta}|M(x)| < \varepsilon$ and $\delta < \frac{\varepsilon}{4}$; then choose $N > 0$ so that for $n > N$ we have $|x_n| < \delta$. Then for $n_k > N$ we have

$$|x_{m_k+1} - x_{n_k}| < 2\delta < \frac{\varepsilon}{2}. \tag{4.13}$$

However it follows from (4.4) and (4.12) that

$$\begin{aligned} x_{m_k+1} - x_{n_k} &\geq -\sum_{j=n_k}^{m_k} a_j M(x_j) + \sum_{j=n_k}^{m_k} a_j b_j \\ &\geq -\sum_{j=n_k}^{m_k} a_j |M(x_j)| + \sum_{j=n_k}^{m_k} a_j b_j \\ &\geq -\sum_{j=n_k}^{m_k} a_j \varepsilon + \left(1+\sum_{j=n_k}^{m_k} a_j\right)\varepsilon \\ &\geq \varepsilon \end{aligned}$$

which contradicts (4.13). Thus (4.11) leads to a contradiction and (4.2) holds.

Proof of Theorem 2.1. Let

$$\begin{aligned} \Omega_1 &= \left\{\omega : \limsup_{n\to\infty}\sup_{k\geq n}\frac{\sum_{j=n}^{k} a_j \xi_j(\omega)}{1+\sum_{j=n}^{k} a_j} \leq 0\right\}, \\ \Omega_2 &= \left\{\omega : \limsup_{n\to\infty}\sup_{k\geq n}\frac{-\sum_{j=n}^{k} a_j \xi_j(\omega)}{1+\sum_{j=n}^{k} a_j} \leq 0\right\}; \end{aligned}$$

$$A = \{\omega : \limsup X_n(\omega) \leq 0\}, \text{ and } B = \{\omega : \limsup(-X_n(\omega)) \leq 0\}.$$

Suppose (2.2) holds, then $P(\Omega_1 \cap \Omega_2) = 1$. By Lemma 4.1, $\Omega_1 \subset B$ and $\Omega_2 \subset A$. Therefore $P(A \cap B) = 1$ and (2.3) holds.

Conversely, suppose (2.3) holds so that $P(A \cap B) = 1$, and suppose M is continuous at 0. By Lemma 4.1, $A \cap B \subset \Omega_1$ and $A \cap B \subset \Omega_2$. Thus $P(\Omega_1 \cap \Omega_2) = 1$ and (2.2) holds.

Proof of Theorem 2.2. From (M1), (M2′), (M3) and (2.1b) we have , for n sufficiently large,

$$\begin{aligned} EX_{n+1}^2 &= E(X_n - a_n M(X_n))^2 + a_n^2 E\xi_n^2 \\ &\le (1 - Ka_n)EX_n^2 + a_n^2 c^2 + a_n^2 E\xi_n^2. \end{aligned}$$

By Lemma 4.1 (with $M =$ the identity there), (2.4) implies $EX_n^2 \to 0$.

Similarly, from (M1′) and (M3) $EX_{n+1}^2 \ge (1 - 2Ka_n)EX_n^2 + a_n^2 E\xi_n^2$. By Lemma 4.1 $EX_n^2 \to 0$ implies (2.4).

5. Proofs for Section 3

Throughout the remainder of this paper we use C to denote positive constants whose exact numerical values are not important. $[x]$ denotes the greatest integer less than or equal to x. 1_A denotes the indicator function of the set A. We need the following lemma.

Lemma 5.1. *Suppose $\alpha \ge 2$, $\{a_n\}$ is a sequence of positive numbers, and the sequence of martingale difference $\{\xi_n, \mathcal{F}_n\}$ satisfies $P\{|\xi_n| \ge t \mid \mathcal{F}_{n-1}\} \le \Phi_n(t)$* a.s., $n = 1, 2, \ldots$. *Then there exists a constant C_α such that*

$$E\left|\sum_{i=1}^n a_i\xi_i\right|^\alpha \le C_\alpha\left(\sum_{i=1}^n a_i^\alpha \int_0^\infty t^{\alpha-1}\Phi_i(t)dt + \left(\sum_{i=1}^n a_i^2 \int_0^\infty t\Phi_i(t)dt\right)^{\alpha/2}\right).$$

Proof. It is derived by repeated use of Theorem 3.3.6 of Stout (1974) (cf. Bai and Yin (1993)).

Proof of Theorem 3.1. By Remark 3.1, we need only to prove the theorem for $p \in [1, 2)$. Let $\eta_n = \xi_n 1_{\{|\xi_n| \le n^{\frac{1}{p}}\}}$, $\eta_n^* = \eta_n - E\{\eta_n|\mathcal{F}_{n-1}\}$, and

$$A_n = \left\{\sup_{k \ge n} \frac{\left|\sum_{j=n}^k j^{-\frac{1}{p}}\eta_j\right|}{1 + \sum_{j=n}^k j^{-\frac{1}{p}}} \ne \sup_{k \ge n} \frac{\left|\sum_{j=n}^k j^{-\frac{1}{p}}\xi_j\right|}{1 + \sum_{j=n}^k j^{-\frac{1}{p}}}\right\}.$$

Then $A_n \subset \bigcup_{j=n}^\infty \{|\xi_j| > j^{\frac{1}{p}}\}$. The fact that $\int_0^\infty t^p |d\Phi(t)| < \infty$ implies that $\sum_{j=1}^\infty P\{|\xi_j| > j^{\frac{1}{p}}\} < \infty$. Consequently, by the Borel-Cantelli lemma $P(\limsup_{n\to\infty}$

$\{|\xi_n| > n^{\frac{1}{p}}\}) = 0$. Therefore $P\{\limsup_{n\to\infty} A_n\} = 0$. Note that $|E\{\eta_n|\mathcal{F}_{n-1}\}| \leq \int_{n^{1/p}}^{\infty} t|d\Phi(t)| \to 0$ a.s. so that

$$\sup_{k\geq n} \frac{\left|\sum_{j=n}^{k} j^{-\frac{1}{p}} E\{\eta_j|\mathcal{F}_{j-1}\}\right|}{1+\sum_{j=n}^{k} j^{-\frac{1}{p}}} \leq \int_{n^{1/p}}^{\infty} t|d\Phi(t)| \quad \text{a.s. as } n\to\infty.$$

Therefore, to prove Theorem 3.1, it suffices to show that

$$\sup_{k\geq n} \frac{\left|\sum_{j=n}^{k} j^{-\frac{1}{p}} \eta_j^*\right|}{1+\sum_{j=n}^{k} j^{-\frac{1}{p}}} \to 0 \quad \text{a.s.} \tag{5.1}$$

We prove (5.1) by cases where $p \in [2^l, 2^{l+1})$ for $l = 1, 2, \ldots$. Let $m = 2^{l+1}$, $T_{n,k} = \sum_{j=n}^{k} j^{-1/p}\eta_j^*$, and $b_{n,k} = 1 + \sum_{j=n}^{k} j^{-1/p}$. Let

$$\begin{aligned} P_n(\varepsilon) &= P\left\{ \sup_{k\geq n} \left|\frac{T_{n,k}}{b_{n,k}}\right| \geq \varepsilon \right\}, \\ P_n^*(\varepsilon) &= P\left\{ \sup_{k>2n} \left|\frac{T_{n,k}}{b_{n,k}}\right| \geq \varepsilon \right\}, \quad \text{and} \\ P_n^{**}(\varepsilon) &= P\left\{ \sup_{n\leq k\leq 2n} \left|\frac{T_{n,k}}{b_{n,k}}\right| \geq \varepsilon \right\}. \end{aligned}$$

For the subsequence $n_k = [k^{\frac{p}{p-1}}]$ we show that both $\sum_k P_{n_k}^*(\varepsilon) < \infty$ and $\sum_k P_{n_k}^{**}(\varepsilon) < \infty$; from this, it will follow that $\sum_k P_{n_k}(\varepsilon) < \infty$ and thus that

$$\sup_{j\geq n_k} \left|\frac{T_{n_k,j}}{b_{n_k,j}}\right| \to 0 \text{ a.s. as } k\to\infty.$$

We will then compare $T_{n,j}/b_{n,j}$ for $n_k < n < n_{k+1}$ with terms in our subsequence, show that the difference is appropriately bounded, and hence argue that (5.1) holds.

Since $\{T_{n,k}, \mathcal{F}_k, k \geq n\}$ is a martingale and both t^{2m} and e^{tu} are convex functions, we see that $\{T_{n,k}^{2m}, \mathcal{F}_k, k \geq n\}$ and $\{\exp\{T_{n,k}u\}, \mathcal{F}_k, k \geq n\}$ are both submartingales.

Suppose $\varepsilon > 0$, $\sigma^2 = \int_0^\infty t^2|d\Phi(t)|$, and $u = \min\{\frac{\varepsilon}{2\sigma^2}, \frac{1}{2}, \frac{1}{\varepsilon}\}$. Then, almost surely,

$$E\left\{ \exp(j^{-1/p}\eta_j^* u)|\mathcal{F}_{j-1}\right\} < \exp\left(\frac{u\varepsilon}{2j^{2/p}}\right)$$

so that, using backward induction,

$$\begin{aligned} E\{\exp(T_{n,k}u)\} &= E\Big\{\exp(T_{n,k-1}u)E(\exp(k^{-1/p}\eta_k^* u) \mid \mathcal{F}_{k-1})\Big\} \\ &< \cdots < \exp\left(\frac{u\varepsilon}{2}\sum_{j=n}^{k} j^{-2/p}\right) = \exp\left(\frac{u\varepsilon}{2}b_{n,k}\right). \end{aligned} \tag{5.2}$$

Using Chow's inequality (see Stout (1974, Theorem 3.3.7)), and then (5.2), we get

$$\begin{aligned} &P\Big\{\sup_{k>2n} T_{n,k}/b_{n,k} \geq \varepsilon\Big\} = P\Big\{\sup_{k>2n} \exp\{T_{n,k}u\}/\exp\{b_{n,k}u\varepsilon\} \geq 1\Big\} \\ &\leq \sum_{k=2n+1}^{\infty} \left(e^{-b_{n,k}u\varepsilon} - e^{-b_{n,k+1}u\varepsilon}\right)Ee^{T_{n,k}u} + \limsup_{k\to\infty} Ee^{T_{n,k}u}/e^{b_{n,k}u\varepsilon} \\ &\leq \sum_{k=2n+1}^{\infty} e^{-b_{n,k}u\varepsilon}(k+1)^{-1/p}\varepsilon u \cdot \exp\left(\frac{u\varepsilon}{2}b_{n,k}\right) + 0 \\ &\leq \sum_{k=2n+1}^{\infty} \exp\left(-b_{n,k}\varepsilon\frac{u}{2}\right)k^{-\frac{1}{p}}\varepsilon u \\ &\leq C\exp\left\{-\frac{\varepsilon u}{2}\left(2^{1-\frac{1}{p}}-1\right)n^{1-\frac{1}{p}}\right\}. \end{aligned} \tag{5.3}$$

Similarly,

$$P\Big\{\sup_{k>2n}(-T_{n,k}/b_{n,k}) \geq \varepsilon\Big\} \leq C\exp\left\{-\frac{\varepsilon u}{2}\left(2^{1-\frac{1}{p}}-1\right)n^{1-\frac{1}{p}}\right\}.$$

Thus

$$P_n^*(\varepsilon) = P\Big\{\sup_{k>2n}\left|\frac{T_{n,k}}{b_{n,k}}\right| \geq \varepsilon\Big\} \leq C\exp\left\{-\frac{\varepsilon u}{2}(\sqrt{2}-1)n^{1-\frac{1}{p}}\right\}. \tag{5.4}$$

Applying Chow's inequality to the submartingale $\{T_{n,k}^{2m}, \mathcal{F}_k, k \geq n\}$ gives

$$\begin{aligned} &\varepsilon^{2m}P\Big\{\sup_{n\leq k\leq 2n}\left(\frac{T_{n,k}}{b_{n,k}}\right)^{2m} \geq \varepsilon^{2m}\Big\} \\ &\leq \sum_{k=n}^{2n-1}\left(\frac{1}{b_{n,k}^{2m}} - \frac{1}{b_{n,k+1}^{2m}}\right)ET_{n,k}^{2m} + \frac{ET_{n,2n}^{2m}}{b_{n,2n}^{2m}}. \end{aligned} \tag{5.5}$$

The fact that $\frac{1}{(1+t)^{2m}} \geq 1 - 2mt$ when t is small enough implies that for n large enough and all $k \geq n$

$$\frac{1}{b_{n,k}^{2m}} - \frac{1}{b_{n,k+1}^{2m}} \leq \frac{2m}{b_{n,k}^{2m+1}}k^{-\frac{1}{p}}. \tag{5.6}$$

Now we apply Lemma 5.1 to estimate $ET_{n,k}^{2m}$. Define

$$\Phi_n(t) = \begin{cases} 1, & \text{if } 0 \le t < 1; \\ \Phi\left(\frac{t}{2}\right), & \text{if } 1 \le t \le n^{1/p} + \frac{1}{2}; \\ 0, & \text{if } t > n^{1/p} + \frac{1}{2}. \end{cases}$$

When n is so large that $\int_{n^{1/p}}^{\infty} t|d\Phi(t)| < 1/2$, we have $P\{|\eta_n^*| \ge t \mid \mathcal{F}_{n-1}\} \le \Phi_n(t)$. It follows from Lemma 5.1 that, for $n \le k \le 2n$,

$$\begin{aligned} ET_{n,k}^{2m} &\le C\left(\sum_{j=n}^{2n} j^{-\frac{2m}{p}} \int_0^\infty t^{2m-1}\Phi_j(t)dt + \left(\sum_{j=n}^{2n} j^{-\frac{2}{p}} \int_0^\infty t\Phi_j(t)dt\right)^m\right) \\ &\le C\left(\sum_{j=n}^{k} j^{-\frac{2m}{p}}\right)\left(1 + \int_0^{(3n)^{\frac{1}{p}}} t^{2m-1}\Phi(t)dt\right) + C\left(\sum_{j=n}^{k} j^{-\frac{2}{p}}\right)^m. \end{aligned} \tag{5.7}$$

A calculation yields that, for $\alpha \ge 2$ and $n \ge 2$ we have

$$\sum_{k=n}^{2n-1} \frac{\left(\sum_{j=n}^{k} j^{-\frac{\alpha}{p}}\right)^{\frac{2m}{\alpha}}}{\left(1 + \sum_{j=n}^{k} j^{-\frac{1}{p}}\right)^{2m+1} k^{\frac{1}{p}}} + \frac{\left(\sum_{j=n}^{2n} j^{-\frac{\alpha}{p}}\right)^{\frac{2m}{\alpha}}}{\left(1 + \sum_{j=n}^{2n} j^{-\frac{1}{p}}\right)^{2m}}$$
$$\le \begin{cases} Cn^{-\frac{2m-1}{p}}, & \text{if } \alpha = 2m; \\ Cn^{-\frac{m}{p}}, & \text{if } \alpha = 2. \end{cases} \tag{5.8}$$

Combining (5.5), (5.6), (5.7) and (5.8)

$$\begin{aligned} &\varepsilon^{2m} P_n^{**}(\varepsilon) = \varepsilon^{2m} P\left\{\max_{n \le k \le 2n} \left(\frac{T_{n,k}}{b_{n,k}}\right)^{2m} \ge \varepsilon^{2m}\right\} \\ &\le Cn^{-\frac{2m-1}{p}} + Cn^{-\frac{2m-1}{p}} \int_0^{(3n)^{1/p}} t^{2m-1}\Phi(t)dt + Cn^{-m/p}. \end{aligned} \tag{5.9}$$

Let $n_k = [k^{\frac{p}{p-1}}]$. We now show that

$$\sum_{k=1}^{\infty} P_{n_k}^{**}(\varepsilon) < \infty. \tag{5.10}$$

Let $A_0 = \int_0^{3^{\frac{1}{p}} k_0^{\frac{1}{p-1}}} t^{2m-1}\Phi(t)dt$ and $A_j = \int_{3^{\frac{1}{p}} j^{\frac{1}{p-1}}}^{3^{\frac{1}{p}}(j+1)^{\frac{1}{p-1}}} t^{p-1}\Phi(t)dt$. Then

$$\int_0^{(3n_k)^{\frac{1}{p}}} t^{2m-1}\Phi(t)dt \le \int_0^{3^{\frac{1}{p}} k^{\frac{1}{p-1}}} t^{2m-1}\Phi(t)dt \le A_0 + \sum_{j=k_0}^{k-1} 3^{\frac{2m-p}{p}} (j+1)^{\frac{2m-p}{p-1}} A_j.$$

Thus

$$\sum_{k=k_0+1}^{\infty} C[k^{\frac{p}{p-1}}]^{-\frac{2m-1}{p}} \int_0^{(3n_k)^{\frac{1}{p}}} t^{2m-1}\Phi(t)dt$$
$$\leq C\int_0^{\infty} t^{p-1}\Phi(t)dt = \frac{C}{p}\int_0^{\infty} t^p|d\Phi(t)| < \infty. \tag{5.11}$$

Hence (5.10) follows from (5.9) and (5.11).

(5.4) implies $\sum_{k=1}^{\infty} P^*_{n_k}(\varepsilon) < \infty$. Together with (5.10) this gives $\sum_{k=1}^{\infty} P_{n_k}(\varepsilon) < \infty$. By the Borel-Cantelli Lemma,

$$\sup_{i\geq n_k} \frac{|T_{n_k,i}|}{b_{n_k,i}} \to 0 \quad \text{a.s.} \tag{5.12}$$

Note that $b_{n,j} \geq 1$ so that for $n_k < n < n_{k+1}$

$$\left|\frac{T_{n,j}}{b_{n,j}}\right| = \left|\frac{T_{n_k,j}}{b_{n_k,j}}\left(1+\frac{b_{n_k,(n-1)}}{b_{n,j}}\right) - \frac{T_{n_k,(n-1)}}{b_{n_k,(n-1)}}\cdot\frac{b_{n_k,(n-1)}}{b_{n,j}}\right|$$
$$\leq \sup_{i\geq n_k}\left|\frac{T_{n_k,i}}{b_{n_k,i}}\right|(1+2b_{n_k,n_{k+1}}), \tag{5.13}$$

and that

$$b_{n_k,n_{k+1}} \leq C. \tag{5.14}$$

Then (5.1) follows from (5.12), (5.13) and (5.14), and this completes the proof.

Proof of Theorem 3.2. It is similar to the proof of Theorem 3.1 (cf. the proof of Theorem 3.3.2 of Li (1993)).

Acknowledgments

This paper is based on part of my Ph.D. dissertation at the State University of New York at Binghamton, written under the supervision of Professor David L. Hanson. I am very grateful for his generous guidance and continuous encouragement. I also thank Professor Zhidong Bai, Professor G. C. Tiao, and the referee for their invaluable suggestions.

References

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of large dimensional covariance matrix. *Ann. Probab.* **21**, 1275–1294.

Blum, J. R. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25**, 382–386.

Dvoretzsky, A. (1956). On stochastic approximation. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* **1**, 39–55.

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

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

Krasulina, T. P. (1969). On stochastic approximation processes with infinite variance. *Theory Probab. Appl.* **14**, 522–526.

Li, G. (1993). Almost sure convergence of stochastic approximation procedures. Ph.D. dissertation at the State University of New York at Binghamton.

Nevel'son, M. B. (1986). On properties of moments of stochastic approximation procedures. *Theory Probab. Appl.* **30**, 407–413.

Nevel'son, M. B. and Has'minskii, R. Z. (1976). *Stochastic Approximation and Recursive Estimation.* Translations of Mathematical Monographs; V. 47. American Math. Society, Providence.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22**, 400–407.

Stout, W. F. (1974). *Almost Sure Convergence.* Academic Press.

Ying, Z. (1990). Nonlinear stochastic approximation procedures for L_p loss functions. *Ann. Statist.* **18**, 1817–1828.

Ohmeda, P.O. Box 804, 110 Allen Road, Liberty Corner, NJ 07938-0804, U.S.A.

(Received February 1992; accepted July 1993)