*The Annals of Probability*
1976, Vol. 4, No. 6, 890–901

# ALMOST SURE CONVERGENCE FOR THE ROBBINS–MONRO PROCESS

BY C. A. GOODSELL AND D. L. HANSON

*The University of Wisconsin, Oshkosh and*
*The State University of New York at Binghamton*

In this paper we investigate the almost sure convergence of the Robbins–Monro process $x_{n+1} = x_n - a_n(y_n - \alpha)$ under assumptions about the conditional distribution of $y_n$ given $x_n$ which involve the existence of first moments or something closely related. The process $x_n$ can converge almost surely even when the series $\sum_{n=1}^{\infty} a_n[y_n - E\{y_n \mid x_n\}]$ does not do so.

**1. Introduction.** Suppose one wishes to find the unique number $\theta$ at which a fixed function $m(\cdot)$ takes on the value $\alpha$, that the function $m(\cdot)$ is unknown, but that at each real number, $x$, a random variable whose expectation is $m(x)$ can be observed. Robbins and Monro [10] proposed starting at some $x_1$, observing $y_1$ at $x_1$, and recursively generating $x_{n+1}$ from $x_n$ and $y_n$ by the relationship

$$x_{n+1} = x_n - a_n(y_n - \alpha) \tag{1}$$

where $y_n$ is observed at $x_n$. This problem and variations of it have been studied extensively. The convergence almost everywhere, in mean-square, and in probability of $x_n$ to $\theta$ have been studied as have the asymptotic distribution of $x_n$ and optimality properties of (1).

In this paper we restrict our attention to the almost sure convergence of $x_n$ to $\theta$. We use (1) but allow the function $m(\cdot)$ to vary with time (with $n$). The convergence almost surely of $x_n$ to $\theta$ and the strong law of large numbers are closely related, however, past results on the almost sure convergence of $x_n$ to $\theta$ have assumed the finiteness of $E\{|y_n|^p \mid x_n\}$ for some $p > 1$, usually $p = 2$ (see Krasulina [8] for results with $1 < p \leqq 2$), while in the i.i.d. case the strong law of large numbers requires only the finiteness of first moments. In this paper we deal, as nearly as possible, with the case $p = 1$.

Section 2 contains a precise statement of our model—which is essentially that of Burkholder [3] in his Theorem 1—and statements of our results. In Section 3 these results are compared with other results in the literature and our assumptions are discussed. Our proofs are in Section 4. Section 5 contains some remarks about sharpness.

**2. Our model and results.** Let $x_1, y_1, y_2, \cdots$ be random variables on a probability space $(\Omega, \Sigma, P)$. The symbol $\omega$ will denote a point in $\Omega$. Let $a_n$ be a sequence of real numbers, and for $n \geqq 1$ define

$$x_{n+1} = x_n - a_n y_n . \tag{2}$$

Received October, 1974; revised April 9, 1976.
*AMS* 1970 *subject classifications.* Primary 62L20; Secondary 60G99.
*Key words and phrases.* Stochastic approximation, Robbins–Monro process.

(We assume—strictly for notational convenience—that the $\alpha$ and $\theta$ referred to in Section 1 are both zero. Thus we will be concerned with the almost sure convergence of $x_n$ to zero, not to $\theta$.) We assume that for all $\omega$ and $n$ we have $P\{y_n \leqq t \mid x_1, \cdots, x_n; y_1, \cdots, y_{n-1}\}(\omega) = P\{y_n \leqq t \mid x_n\}(\omega)$ and that the only dependence of $P\{y_n \leqq t \mid x_n\}(\omega)$ on $\omega$ is through $x_n(\omega)$; we use the notation $P\{y_n \leqq t \mid x_n = x\}$ and also $E\{\cdot \mid x_n = x\}$. We further assume that $E\{|y_n| \mid x_n = x\} < \infty$ for all real $x$. We define $m_n(x) = E\{y_n \mid x_n = x\}$ and $\xi_n = y_n - m_n(x_n)$.

In that which follows we use three notational conventions. If $A$ is a finite set, we use $\# A$ to denote the number of elements in $A$. We use $C$ to denote all positive constants whose exact numerical values do not matter; using this notation $1 + C \leqq C$ makes sense. We will use $a/bc$ and $a/b(c + d)$ instead of $a/(bc)$ and $a/(b(c + d))$ with some regularity to cut down on the number of parentheses needed.

We assume

(3a) $$a_n > 0 \quad \text{for all} \quad n,$$

(3b) there exists a $C > 0$ such that for every $a > 0$ we have $\#\{n \mid a_n \geqq a\} \leqq C/a$,

(4a) for every $a > 0$ there exists an $N$ such that $x m_n(x) \geqq 0$ for all $n \geqq N$ and $|x| \geqq a$,

(4b) for every $0 < a < b < \infty$ we have
$$\sum_{n=1}^{\infty} a_n \inf_{a \leqq x \leqq b} m_n(x) = -\sum_{n=1}^{\infty} a_n \sup_{-b \leqq x \leqq -a} m_n(x) = \infty, \qquad \text{and}$$

(5) there exists a $C > 0$ such that for all $x$ and all $n$ we have $|m_n(x)| \leqq C(1 + |x|)$.

Whenever the function $F$ appears it will be used in an expression bounding the tails of the conditional distribution of $\xi_n$ or of $y_n$ given $x_n$. We will always assume that

(6a) $$F \text{ is nonincreasing}, \quad F(0) = 1, \quad \text{and} \quad \lim_{x \to \infty} F(x) = 0.$$

If we have an $F$ which satisfies (6a) and the remaining properties which we will want it to satisfy, then it can easily be modified so as to be continuous as well. We will assume that

(6b) $$P\{|\xi_n| \geqq s \mid x_n = x\} \leqq F\left(\frac{s}{1 + |x|}\right) \quad \text{for all} \quad x \quad \text{and all} \quad s \geqq 0,$$

or equivalently that $P\{|\xi_n| \geqq t(1 + |x_n|) \mid x_n\} \leqq F(t)$ for all $t \geqq 0$.

THEOREM 1. *Assume the model presented in this section including* (3), (4), (5), *and* (6). *Suppose one of the following two sets of conditions holds*:

(7) $$\int_1^{\infty} t \log t |dF(t)| < \infty; \qquad \textit{or}$$

(8) $$\int_1^{\infty} t |dF(t)| < \infty \qquad \textit{and}$$

(9) *for every* $C > 0$ *there exists a* $d > 0$, *a positive integer* $N$, *and an* $\varepsilon$ *in* $(0, 1)$ *such that for all* $n \geqq N$ *and all* $|x| \geqq C$ *we have*

$$\sup_{t \geqq d} |E\{\xi_n I_{\{|\xi_n| \leqq t/a_n\}} \mid x_n = x\}| \leqq (1 - \varepsilon)|m_n(x)| . \tag{9a}$$

*Then* $x_n \to 0$ *almost surely.*

REMARK. (9a) holds if either (i) the conditional distribution of $\xi_n$ given $x_n = x$ is symmetric, or if (ii) $\int_{t > d/a_n} t|dF((t)/1 + |x|)| \leqq (1 - \varepsilon)|m_n(x)|$.

COROLLARY. *Assume the model presented in this section including* (3), (5), (6a), *and* (8). *In addition assume*

$$\textstyle\sum_{n=1}^\infty a_n = \infty , \tag{3c}$$

(4c) *for every* $a > 0$ *there exists a positive integer* $N$ *such that*

$$\inf_{x \geqq a; n \geqq N} m_n(x) > 0 \quad \textit{and} \quad \sup_{x \leqq -a; n \geqq N} m_n(x) < 0 ,$$

*and either*

(4d) *for every* $C > 0$ *there exists a positive integer* $N$ *and an* $\varepsilon > 0$ *such that if* $|x| \geqq C$ *and* $n \geqq N$ *then* $\varepsilon|x| \leqq |m_n(x)|$,

*or*

(6c) $P\{|\xi_n| \geqq s \mid x_n = x\} \leqq F(s)$ *for all* $s \geqq 0$, *all* $n$, *and all* $x$.

*Then* $x_n \to 0$ *almost surely.*

PROOF. It can be shown in a straightforward manner that the conditions of Theorem 1 hold.

In addition to these results obtaining—from basic considerations—the almost sure convergence of $x_n$ to 0, we will also prove the following theorem obtaining almost sure convergence from a weaker form of convergence.

THEOREM 2. *Assume the model presented in this section including* (3), (6a), *and* (8). *Assume that*

(10) *there exists a* $b > 0$ *such that*

(10a) $m_n(x)$ *exists and is finite for all* $|x| \leqq b$,

(10b) $\sup_{n; |x| \leqq b} |m_n(x)| < \infty$,

(10c) *for all* $a$ *in* $(0, b)$ *we have*

$$\textstyle\sum_{n=1}^\infty a_n \inf_{a \leqq x \leqq b} m_n(x) = -\sum_{n=1}^\infty a_n \sup_{-b \leqq x \leqq -a} m_n(x) = \infty ,$$

(10d) *for every* $a$ *in* $(0, b)$ *there exists an* $N$ *such that* $x m_n(x) \geqq 0$ *for all* $n \geqq N$ *and* $a \leqq |x| \leqq b$, *and*

(10e) $P\{|\xi_n| \geqq s \mid x_n = x\} \leqq F(s)$ *for all* $s \geqq 0$ *and all* $|x| \leqq b$.

*If* $x_n \to 0$ *in probability, then* $x_n \to 0$ *almost surely.*

The main point of interest about Theorem 2 is that once one has $x_n \to 0$, then it takes additional assumptions *near zero only* to get convergence almost surely. Note also that the conditions (10a)—(10e) of Theorem 2 are guaranteed by either of the two sets of conditions of Theorem 1.

It should also be pointed out that if $\theta$ is in $[a, b]$, and if we define recursively

$$x_{n+1} = aI_{\{x_n - a_n y_n < a\}} + (x_n - a_n y_n)I_{\{a \leqq x_n - a_n y_n \leqq b\}} + bI_{\{b < x_n - a_n y_n\}},$$

then theorems can be proved which provide the almost sure convergence of $x_n$ to zero and which are similar to, but simpler to prove than, the ones stated here. For example, condition (5) would reduce to $\sup_{a \leqq x \leqq b} |m_n(x)| \leqq C$ and (6b) would reduce to $P\{|\xi_n| \geqq s \,|\, x_n\} \leqq F(s)$ for all $x$ in $[a, b]$. Theorems of this sort exist in the literature. Robbins and Siegmund [11, page 244] have introduced a different form of "truncation" for controlling the oscillation of the sequence $x_n$.

**3. A brief discussion of our assumptions and comparison with the literature.** As the name suggests, the Robbins-Monro process was introduced by Robbins and Monro [10] who proved mean-square convergence of $x_n$ to zero (actually, to $\theta$) under the assumption that the $y_n$'s are bounded, the assumption that $0 < C_1/n < a_n \leqq C_2/n$ for all $n$, and fairly strong assumptions on $m$ ($m_n \equiv m$). Blum [1] was the first to prove almost sure convergence; he assumed

(11) $$|m(x)| \leqq c + d|x| \quad \text{for all} \quad x,$$

(12) $$E\{(y_n - m(x_n))^2 \,|\, x_n\} \leqq \sigma^2 < \infty \quad \text{for all} \quad n \quad \text{and all values of} \quad x_n,$$

(13) $$xm(x) > 0 \quad \text{for} \quad x \neq 0, \qquad \text{and}$$

(14) $$\inf_{a \leqq |x| \leqq b} |m(x)| > 0 \quad \text{for all} \quad 0 < a < b < \infty.$$

Numerous authors have worked on variations of this problem. Dvoretsky [5], for example, has worked on a stochastic approximation scheme which is quite general. (A paper by Schmetterer [12] and a book by Wasan [13] contain extensive bibliographies.)

Our model is essentially that of Burkholder [3, Theorem 1]. He assumed conditions (4a) and (5) and a condition just slightly stronger than our condition (4b). He also assumed the existence of, and a uniform bound on, the conditional variances of the $y_n$'s, and that

(15) $$\sum_{n=1}^{\infty} a_n^2 < \infty.$$

If our $a_n$'s were reordered so as to be nonincreasing in $n$, then (3b) would imply that $a_n \leqq C/n$ for all $n$ so that our (3b) is more restrictive than Burkholder's (15).

Most authors who have worked on this problem have assumed that the conditional distributions of the $y_n$'s have finite variances though their methods of proof may not have required such an assumption. Krasulina [8] assumed only that for some $p > 1$

(16) $$E\{|y_n - m(x_n)|^p \,|\, x_n\} \leqq C_p < \infty \quad \text{uniformly in} \quad n \quad \text{and} \quad x_n.$$

It is our purpose in Theorem 1 to come as close to the case $p = 1$ as possible.

If we could first argue that $\sum_{n=1}^{\infty} a_n \xi_n$ converges to a finite limit almost surely, then the rest of the proof of Theorem 1 would follow easily using essentially Blum's methods [1], [2]. We were, of course, unable to do this; in fact, this method of proof would seem to be precluded entirely in the case $p = 1$ as long as we allow the conditional distributions of the $\xi_n$'s given $x_n$ to grow linearly in $x_n$ as in (6b). If the bound $F[s/(1 + |x|)]$ in (6b) were replaced by $F(s)$, then these conditional distributions would be uniformly bounded in $x_n$. Then, when (7) holds, either the lemma numbered (10) of Dubins and Freedman [4] or a minor modification of Theorem D on page 387 of Loève [9] gives the almost sure convergence of $\sum_{n=1}^{\infty} a_n \xi_n$ to a finite limit. The series $\sum_{n=1}^{\infty} a_n \xi_n$ may not converge to a finite limit almost surely if (7) does not hold. Gladyshev [7] and Robbins and Siegmund [11] allow the conditional distributions of the $\xi_n$'s given $x_n$ to grow linearly in $x_n$ but they assume finite variances and we have been unable to use either method of proof in our situation.

For intuitive purposes it is convenient to write (2) in the form $x_{n+1} = x_n - a_n m_n(x_n) - a_n \xi_n$, to think of $a_n m_n(x_n)$ as a correction term causing a return to zero, and to think of $a_n \xi_n$ as an error term. Then the particular inequality in (3a) is simply a convention, but the combination of (3a) and (4a) guarantees that "eventually" the correction term is in the right direction. Assumption (4b) guarantees that the process will not get caught away from zero lacking enough "remaining correcting force" to return to a neighborhood of zero. If the error term is zero, then it is "obvious" that some sort of bounds must be put on the $m_n$'s in order to prevent wilder and wilder oscillations of the sequence $x_n$; assumption (5) is a standard assumption for eliminating this potential problem; Dvoretsky [5, Remark 1, page 51] gives an example of this oscillation problem. (Engelhardt [6, Theorem 1] shows that a similar linear growth condition on the conditional standard deviations of the $\xi_n$'s given $x_n$ (first used by Gladyshev [7]) is reasonably "tight.")

Our remaining assumptions essentially gurantee the proper behavior of the sequence $a_n \xi_n$, either by itself or in relation to the sequence $m_n$. We originally tried to prove Theorem 1 assuming (8) as our only assumption on the function $F$ bounding the tails of our conditional distributions. Unfortunately, this is insufficient; an imbalance in the tails—positive and negative—can cause almost sure convergence to fail. Our assumption (7) bounds the imbalance sufficiently and, indirectly, guarantees the almost sure convergence of $\sum_{n=1}^{\infty} a_n \xi_n$ to a finite limit. (9a) relates the allowable imbalance to the particular sequence $a_n$ and the particular mean functions $m_n$ involved; it does not necessarily guarantee the almost sure convergence $\sum_{n=1}^{\infty} a_n \xi_n$ but instead guarantees that the $m_n$'s are large enough to compensate.

**4. Proofs.** We begin with two lemmas involving the sequence $a_n$ and the moment assumptions (7) and (8).

LEMMA 1. *Assume* (3), (6a), *and* (8). *Then for every* $d > 0$ *we have*

$$\sum_{n=1}^{\infty} F(d/a_n) < \infty \tag{17}$$

*and*

$$\sum_{n=1}^{\infty} a_n^2 \int_{[0,d/a_n]} y^2 |dF(y)| < \infty . \tag{18}$$

PROOF.

$$\begin{aligned}\sum_{n=1}^{\infty} F(d/a_n) &\leqq \sum_{M=1}^{\infty} [\sum_{\{n \mid M-1<d/a_n \leqq M\}} F(M-1)] \\ &= \sum_{M=1}^{\infty} \#\{n \mid M-1 < d/a_n \leqq M\} \sum_{k=M}^{\infty} [F(k-1) - F(k)] \\ &= \sum_{k=1}^{\infty} [F(k-1) - F(k)] \#\{n \mid d/k \leqq a_n\} \\ &\leqq C \sum_{k=1}^{\infty} k[F(k-1) - F(k)] < \infty .\end{aligned}$$

This proves (17). To prove (18) we perform similar but more complicated manipulations.

$$\begin{aligned}&\sum_{n=1}^{\infty} a_n^2 \int_{[0,d/a_n]} y^2 |dF(y)| \\ &\quad \leqq \sum_{M=1}^{\infty} [\sum_{\{n \mid M-1<d/a_n \leqq M\}} a_n^2 \int_{[0,M]} y^2 |dF(y)|] \\ &\quad \leqq C + C \sum_{M=2}^{\infty} \#\{n \mid M-1 < d/a_n \leqq M\} M^{-2} \sum_{k=1}^{M} k^2 [F(k-1) - F(k)] \\ &\quad \leqq C + C \sum_{n=1}^{\infty} k^2 [F(k-1) - F(k)] \sum_{M=k}^{\infty} M^{-2} \#\{n \mid M-1 < d/a_n \leqq M\} .\end{aligned} \tag{19}$$

Note that

$$\begin{aligned}&\sum_{M=k}^{\infty} M^{-2} \#\{n \mid M-1 < d/a_n \leqq M\} \\ &\quad \leqq \sum_{j=k}^{\infty} \left(\frac{1}{j^2} - \frac{1}{(j+1)^2}\right) \sum_{M=k}^{j} \#\{n \mid M-1 < d/a_n \leqq M\} \\ &\quad \leqq C \sum_{j=k}^{\infty} j^{-3} \#\{n \mid d/j \leqq a_n\} \leqq C \sum_{j=k}^{\infty} j^{-2} \leqq C/k .\end{aligned}$$

Thus (19) is finite.

LEMMA 2. *Assume* (3), (6a), *and* (7). *Then for every* $d > 0$

$$\sum_{n=1}^{\infty} a_n \int_{(d/a_n,\infty)} t |dF(t)| < \infty . \tag{20}$$

PROOF.

$$\begin{aligned}&\sum_{n=1}^{\infty} a_n \int_{(d/a_n,\infty)} t |dF(t)| \\ &\quad = \sum_{n=1}^{\infty} a_n [(d/a_n) F(d/a_n) + \int_{(d/a_n,\infty)} (t - d/a_n) |dF(t)|] \\ &\quad = C + \sum_{M=1}^{\infty} [\sum_{\{n \mid M-1<d/a_n \leqq M\}} a_n \int_{M-1}^{\infty} F(t)\, dt] \\ &\quad \leqq C + \sum_{M=2}^{\infty} \#\{n \mid M-1 < d/a_n \leqq M\} \frac{d}{M-1} \sum_{k=M-1}^{\infty} F(k) \\ &\quad \leqq C + C \sum_{k=1}^{\infty} F(k) \sum_{M=1}^{k} M^{-1} \#\{n \mid M < d/a_n < M+1\} \\ &\quad = C + C \sum_{j=1}^{\infty} [F(j) - F(j+1)] \sum_{M=1}^{j} \frac{j-M+1}{M} \#\{n \mid M < d/a_n \leqq M+1\} \\ &\quad \leqq C + C \sum_{j=1}^{\infty} j[F(j) - F(j+1)][\sum_{M=1}^{j} M^{-1} \#\{n \mid d/a_n \leqq M+1\} \\ &\qquad - \sum_{M=1}^{j} M^{-1} \#\{n \mid d/a_n \leqq M\}]\end{aligned}$$

$$\leqq C+C\sum_{j=1}^{\infty} j[F(j)-F(j+1)]\Big[\sum_{M=1}^{j-1}\Big(\frac{1}{M}-\frac{1}{M+1}\Big)\#\{n\,|\,d/a_n\leqq M+1\}$$
$$+\,j^{-1}\#\{n\,|\,d/a_n\leqq j+1\}\Big]$$
$$\leqq C+C\sum_{j=1}^{\infty} j[F(j)-F(j+1)][\sum_{M=1}^{j-1} M^{-2}(C/M)+C]$$
$$\leqq C+C\sum_{j=1}^{\infty} j[F(j)-F(j+1)][\log j+C]<\infty\,.$$

Proof of Theorem 1. The proof has several distinct parts to it. Lemma 3 is used later in the proof. In Lemma 4 we show that if $0<\alpha<\beta<\infty$ then $P\{\liminf x_n<\alpha;\ \limsup x_n>\beta\}=0$ and $P\{\liminf x_n<-\beta;\ \limsup x_n>-\alpha\}=0$. A standard argument using these shows that if we allow $+\infty$ and $-\infty$ as limits then $P\{\lim x_n \text{ exists}\}=P\{\lim x_n\in[-\infty,+\infty]\}=1$. In Lemma 5 we use arguments similar to those used in Lemma 4 to eliminate the probability of all finite limits except zero, i.e., to show that $P[\lim x_n\in\{-\infty,0,+\infty\}]=1$. The remainder of the proof uses a moment argument to eliminate $+\infty$ and $-\infty$.

Lemma 3. *Assume the model presented in Section* 2 *including* (3), (5), (6) *and* (8). *Then if* $0\leqq\alpha<\beta<\infty$

$$P\{x_n\leqq\alpha \ \textit{ and } \ x_{n+1}\geqq\beta \text{ i.o.}\}=0 \tag{21}$$

*and*

$$P\{x_n\geqq-\alpha \ \textit{ and } \ x_{n+1}\leqq-\beta \text{ i.o.}\}=0\,.$$

Proof. The proofs of the two conclusions are essentially the same. We will prove only the first of the two. For that it suffices to show that $\sum_{n=1}^{\infty}P\{x_n\leqq\alpha,\ x_{n+1}\geqq\beta\}<\infty$. Note that

$$\begin{aligned}\{x_{n+1}\geqq\beta,\ x_n\leqq\alpha\}&=\{(x_n-\beta)/a_n-m_n(x_n)\geqq\xi_n,\ x_n\leqq\alpha\}\\&\subset\{(x_n-\beta)/a_n+C(1+|x_n|)\geqq\xi_n,\ x_n\leqq\alpha\}\\&=\Big\{(1+|x_n|)\Big[\frac{-|\beta-x_n|}{a_n(1+|x_n|)}+C\Big]\geqq\xi_n,\ x_n\leqq\alpha\Big\}\\&\subset\{(1+|x_n|)(-\lambda/a_n+C)\geqq\xi_n,\ x_n\leqq\alpha\}\end{aligned}$$

where $\lambda=\min\{1,(\beta-\alpha)/(1+\alpha)\}$. Since (3b) implies $a_n\to0$, for $n$ sufficiently large $C<\lambda/2a_n$, and thus $P\{x_{n+1}\geqq\beta\,|\,x_n\leqq\alpha\}\leqq F(\lambda/2a_n)$. Consequently

$$\sum_{n=1}^{\infty}P\{x_n\leqq\alpha,\ x_{n+1}\geqq\beta\}\leqq C+\sum_{n=1}^{\infty}F(\lambda/2a_n)$$

which is finite from Lemma 1.

Lemma 4. *Assume the conditions of Theorem* 1. *If* $0<\alpha<\beta<\infty$ *then*

$$\begin{aligned}&P\{\liminf x_n<\alpha \ \textit{ and } \ \limsup x_n>\beta\}\\&\qquad=P\{\liminf x_n<-\beta \ \textit{ and } \ \limsup x_n>-\alpha\}=0\,.\end{aligned}\tag{22}$$

Proof. The proofs that the two probabilities in (22) are zero are essentially the same. We will prove only that the first of the two is zero. For $n=1,2,\cdots$ define $\eta_n=a_n\xi_nI_{\{\alpha/2\leqq x_n\leqq\beta\}}$. Fix $d>0$; if (9) is applicable let $d$ be associated with $C=\alpha/2$ in (9). Define $\eta_n{}^d=\eta_nI_{\{|\eta_n|\leqq d\}}$. Whether $x_n\in[\alpha/2,\beta]$ or not, we have $P\{\eta_n \neq \eta_n{}^d \mid x_n\} \leqq F(\lambda/a_n)$ where $\lambda = d/(1+\beta)$. Lemma 1 gives $\sum_{n=1}^{\infty} P\{\eta_n \neq \eta_n{}^d\} < \infty$ so that $P\{\eta_n \neq \eta_n{}^d \text{ i.o.}\} = 0$.

Let $z_n = \eta_n{}^d - E\{\eta_n{}^d \mid x_n\}$. Note that $E\{z_n \mid x_n\} = 0$ and that the $z_n$'s are bounded by $2d$. A straightforward calculation and an application of Lemma 1 show that $\sum_{n=1}^{\infty} E\{z_n{}^2 \mid x_n\} < \infty$. Then either the lemma numbered (10) of Dubins and Freedman [4] or a minor modification of Theorem $D$ on page 387 of Loève [9] will give the almost sure convergence of $\sum_{n=1}^{\infty} z_n$ to a finite limit.

Define the sets $A$, $B$, $C$, and $D$ as follows:

$$A = \{\omega \mid x_n(\omega) \leqq \alpha/2 \text{ and } x_{n+1}(\omega) \geqq \alpha \text{ only finitely often}\},$$
$$B = \{\omega \mid \textstyle\sum_{k=1}^{\infty} z_k(\omega) \text{ converges to a finite limit}\},$$
$$C = \{\omega \mid \eta_n(\omega) \neq \eta_n{}^d(\omega) \text{ only finitely often}\}, \text{ and}$$
$$D = \{\omega \mid \liminf x_n(\omega) < \alpha,\ \limsup x_n(\omega) > \beta\}.$$

Note that $P(A) = P(B) = P(C) = 1$ so that $P(A \cap B \cap C \cap D) = P(D)$. We will show that $A \cap B \cap C \cap D$ is empty and that will complete the proof of this lemma.

Now suppose that $\omega \in A \cap B \cap C \cap D$. Choose $N$ large enought that for $m \geqq n \geqq N$ we have

(23) $$x_{n+1}(\omega) \geqq \alpha \quad \text{implies} \quad x_n(\omega) \geqq \alpha/2\,,$$

(24) $$\left|\textstyle\sum_{k=n}^{m} z_k(\omega)\right| < (\beta - \alpha)/2\,,$$

(25) $$\eta_n(\omega) = \eta_n{}^d(\omega)\,,$$

(26) $$m_n(x) \geqq 0 \qquad \text{for all} \quad x \geqq \alpha/2\,,$$

(27a) if (7) holds (using Lemma 2), then

$$\textstyle\sum_{k=m}^{n} (1+\beta) a_k \int_{(d/a_k(1+\beta),\infty)} s|dF(s)| < (\beta-\alpha)/2\,,$$

and

(27b) if (9) holds, then for all $|x| \geqq \alpha/2$, our particular $d$, and some fixed $\varepsilon$ in (0, 1) we have (9a).

Choose particular values of $m$ and $n$ with $N \leqq n \leqq m$ so that $x_n(\omega) < \alpha$, $\alpha \leqq x_k(\omega) \leqq \beta$ for $n < k \leqq m$, and $x_{m+1}(\omega) > \beta$. From (23) we see that $x_n(\omega) \geqq \alpha/2$. Then

(28) $$\begin{aligned}\beta - \alpha &< x_{m+1}(\omega) - x_n(\omega)\\ &= -\textstyle\sum_{k=n}^{m} [a_k \xi_k(\omega) - \eta_k(\omega)] - \sum_{k=n}^{m} [\eta_k(\omega) - \eta_k{}^d(\omega)]\\ &\quad - \textstyle\sum_{k=n}^{m} z_k(\omega) - \sum_{k=n}^{m} [a_k m_k(x_k(\omega)) + E\{\eta_k{}^d \mid x_k\}(\omega)]\\ &< 0 + 0 + (\beta-\alpha)/2\end{aligned}$$

(28a) $$\quad - \textstyle\sum_{k=n}^{m} [a_k m_k(x_k(\omega)) + E\{\eta_k{}^d \mid x_k\}(\omega)]\,.$$

If (7) holds then, since $E\{\xi_k \mid x_k\} = 0$, we have $|E\{\eta_k{}^d \mid x_k\}| \leqq a_k \int_{(d/a_k,\infty)} t|dF((t)/(1+|x_k|))|$ which is bounded, if $|x_k| \leqq \beta$, by $(1+\beta) a_k \int_{(d/a_k(1+\beta),\infty)} s|dF(s)|$. Using (27a) we get the following contradiction from (28):

$$\beta - \alpha < (\beta-\alpha)/2 - \textstyle\sum_{k=n}^{m} E\{\eta_k{}^d \mid x_k\}(\omega) \leqq \beta - \alpha\,.$$

If (9) holds then, using (9a),

$$\begin{aligned}(28\text{a}) &\leqq -\textstyle\sum_{k=n}^{m} a_k m_k(x_k(\omega)) + \sum_{k=n}^{m} a_k|E\{\xi_k I_{\{|\xi_k|\leqq d/a_k\}}\,|\,x_k\}| \\ &\leqq -\textstyle\sum_{k=n}^{m} \varepsilon a_k m_k(x_k(\omega)) < 0\end{aligned}$$

so that (28) gives the contradiction $\beta - \alpha < (\beta - \alpha)/2$.

LEMMA 5. *Assume the conditions for Theorem* 1. *Then*

$$P[\lim x_n \in \{0, +\infty, -\infty\}] = 1 . \tag{29}$$

PROOF. It follows from Lemma 4 and a standard argument that $P\{\lim x_n \in [-\infty, +\infty]\} = P\{\liminf x_n = \limsup x_n\} = 1$. If the conclusion of this lemma were false there would exist $0 < \alpha < \beta < \infty$ and a $\delta > 0$ such that $\alpha + \delta < \beta - \delta$ and such that either $P\{\lim x_n \in [\alpha + \delta, \beta - \delta]\} > 0$ or $P\{\lim x_n \in [-\beta + \delta, -\alpha - \delta]\} > 0$. We will prove that the former is impossible.

Define $\eta_n$, $\eta_n{}^d$, $z_n$, and the sets $B$ and $C$ as in the proof of Lemma 4. Let $A = \{\lim x_n \in [\alpha + \delta, \beta - \delta]\}$. As in the proof of Lemma 4, $P(B) = P(C) = 1$ so that $P(A \cap B \cap C) = P(A)$. We will show that $A \cap B \cap C$ is empty so that $P(A) = 0$.

Fix $\omega \in A \cap B \cap C$. Choose $N$ so that for all $m \geqq n \geqq N$ we have (24), (25), (26),

$$\alpha < x_n(\omega) < \beta , \tag{30}$$

and, if appropriate, (27a) or (27b). Then, using the expansion in (28),

$$-(\beta - \alpha) < x_{m+1}(\omega) - x_n(\omega) < 0 + 0 + (\beta - \alpha)/2 + (\text{expression } (28\text{a})) . \tag{31}$$

In the following we treat $E\{\eta_k{}^d\,|\,x_k\}$ as in the proof of Lemma 4; when (7) holds, (31) gives (using (27a))

$$\textstyle\sum_{k=n}^{m} a_k m_k(x_k(\omega)) < 2(\beta - \alpha)$$

which gives a contradiction for large enoug $m$; when (9) holds, (31) gives

$$\textstyle\sum_{k=n}^{m} a_k m_k(x_k(\omega)) \leqq 3(\beta - \alpha)/2 + \sum_{k=n}^{m} a_k(1 - \varepsilon)m_k(x_k(\omega))$$

which again gives a contradiction for large enough $m$.

REMAINDER OF THE PROOF OF THEOREM 1. From Lemma 5 we see that it suffices to prove that $P\{\lim x_n = +\infty\} = 0$ and $P\{\lim x_n = -\infty\} = 0$. We will prove only the former.

Let $d > 0$ and, if (9) holds, let $d$ be such that for large enough $n$ we have for $|x| \geqq 2d$

$$\sup_{t\geqq d} |E\{\xi_n I_{\{|\xi_n|\leqq t/a_n\}}\,|\,x_n = x\}| \leqq (1 - \varepsilon)|m_n(x)| . \tag{32}$$

Let $b > 0$. Define $D_b = \{|x_1| \leqq b\}$, $B_k = \{x_k \geqq 2d\}$, $C_k = \{|a_k \xi_k| \leqq |x_k|/2\}$, and $A_{N,n} = D_b \cap \bigcap_{k=N}^{n} B_k C_k$.

A straightforward induction argument shows that $E|x_n|I_{D_b} < \infty$ for all $n$. Choose $N$ large enough that, using (4a) and (5), if $x \geqq d$ and $n \geqq N$ then

$m_n(x) \geqq 0$ and $x - a_n m_n(x) > x/2$; in addition, if (9) holds, choose $N$ large enough that (32) holds if $n \geqq N$ and $|x| \geqq 2d$. Note that if $\omega \in B_n C_n$ and $n \geqq N$ then $x_{n+1} > x_n/2 - a_n \xi_n \geqq 0$. If $n > N$ then

$$\begin{aligned} Ex_{n+1} &\leqq E(x_n - a_n m_n(x_n)) I_{A_{N,n-1}} I_{B_n} \\ &\quad - a_n E\{I_{A_{N,n-1}} I_{B_n} E\{\xi_n I_{C_n} \mid x_1, \cdots, x_n\}\} \\ &\leqq Ex_n I_{A_{N,n-1}} - a_n E\{I_{A_{N,n-1}} I_{B_n}[m_n(x_n) - |E\{\xi_n I_{C_n} \mid x_n\}|]\} . \end{aligned} \tag{33}$$

If (9)—and thus (32)—holds, then for $\omega$ in $B_n$ we have $|E\{\xi_n I_{C_n} \mid x_n\}| \leqq \sup_{t \geqq d} |E\{\xi_n I_{\{|\xi_n| \leqq t/a_n\}} \mid x_n\}|$ so that $Ex_{n+1} I_{A_{N,n}} \leqq Ex_n I_{A_{N,n-1}} - \varepsilon a_n E\{I_{A_{N,n-1}} I_{B_n} m_n(x_n)\} \leqq Ex_n I_{A_{N,n-1}}$. It follows that $\sup_{n>N} Ex_n I_{A_{N,n-1}} < \infty$. If (7) holds then using (33) we get

$$\begin{aligned} Ex_{n+1} I_{A_{N,n}} &\leqq Ex_n I_{A_{N,n-1}} + a_n E\left\{I_{A_{N,n-1}} I_{B_n} \int_{t > |x_n|/2a_n} t \left| dF\left(\frac{t}{1 + |x_n|}\right)\right|\right\} \\ &\leqq (Ex_n I_{A_{N,n-1}})(1 + \beta_n) + \beta_n \end{aligned}$$

where $\beta_n = a_n \int_{t > d/a_n(1+2d)} t\, |dF(t)|$. We know that $\sum_{n=1}^{\infty} \beta_n < \infty$ from Lemma 2. An induction argument shows that if $n > N$ then

$$\begin{aligned} Ex_{n+1} I_{A_{N,n}} &\leqq (1 + Ex_{N+1} I_{A_{N,N}}) \prod_{k=N+1}^{n} (1 + \beta_k) - 1 \\ &\leqq (1 + Ex_{N+1} I_{A_{N,N}}) \prod_{k=N+1}^{\infty} (1 + \beta_k) - 1 < \infty . \end{aligned}$$

Again we have $\sup_{n>N} Ex_n I_{A_{N,n-1}} < \infty$. Let $B^*_N = \bigcap_{k=N}^{\infty} B_k$, $C_N^* = \bigcap_{k=N}^{\infty} C_k$, and $A_N = D_b B_N^* C_N^* = \bigcap_{n=N}^{\infty} A_{N,n}$. If either (7) or (9) holds we have $\sup_{n>N} Ex_n I_{A_N} < \infty$ so that $P[A_N \cap \{x_n \to +\infty\}] = 0$. Since $\bigcap_N B_N^* \supset \{x_n \to +\infty\}$ and $P(C_N^*) \to 1$ as $N \to \infty$, it follows that $P(D_b \cap \{x_n \to +\infty\}) = 0$. Since this holds for all $b > 0$ we have $P\{x_n \to +\infty\} = 0$.

PROOF INDICATION FOR THEOREM 2. One proves that

$$\text{if} \quad 0 < \alpha < \beta < b \quad \text{then} \quad P\{\liminf |x_n| < \alpha,\ \limsup |x_n| > \beta\} = 0 . \tag{34}$$

The convergence in probability of $x_n$ to zero implies $P\{\liminf |x_n| = 0\} = 1$. This combined with (34) gives $P\{x_n \to 0\} = 1$.

One first proves an analog of Lemma 3, that for $0 \leqq \alpha < \beta < b$

$$P\{-b \leqq x_n \leqq \alpha,\ x_{n+1} \geqq \beta \ \text{ i.o.}\} = P\{-\alpha \leqq x_n \leqq b,\ x_{n+1} \leqq -\beta \ \text{ i.o.}\} = 0 ,$$

using essentially the proof of Lemma 3. One then proves that

$$P\{\liminf |x_n| \leqq \alpha, \limsup x_n \geqq \beta\} = P\{\liminf |x_n| \leqq \alpha, \liminf x_n \leqq -\beta\} = 0 .$$

The proof is very much like the proof given for Lemma 4. The sets $A$ and $D$ must be redefined as follows:

$$A = \{\omega \mid -b \leqq x_n(\omega) \leqq \alpha/2,\ x_{n+1}(\omega) \geqq \alpha \ \text{ only finitely often}\}$$

and

$$D = \{\omega \mid \liminf |x_n(\omega)| < \alpha, \limsup x_n(\omega) > \beta\} .$$

**5. Some remarks about sharpness.** Condition (6b) allows the conditional distribution of the errors to "spread out" linearly in $|x_n|$, i.e., in distance from zero (or from the desired root). This would allow, for example, for a proportional measurement error if $m(x)$ were linear. It is an intuitively appealing hypothesis, and in addition it is the analog of hypotheses made about the errors when the existence of second moments has been assumed. Unfortunately, the techniques of proof which we used to deal with (6 b) obscure some more basic considerations.

Condition (7) in Theorem 1 was not intended to imply that $\sum_{n=1}^{\infty} a_n \xi_n$ converges to a finite limit, however, the convergence of this series is a by-product of Theorem 1. Once we know that $x_n \to 0$ almost surely we know, in particular, that for each $\omega$ the sequence $x_n(\omega)$ is bounded. Then we can't quite replace (6b) by (6c) but we can come close. In particular, (7) is now enough to guarantee that $\sum_{n=1}^{\infty} a_n \xi_n$ converges to a finite limit.

Suppose $\{\xi_n\}$ is an i.i.d. sequence with continuous distribution function $G$. We write $x_{m+1} - x_n = -\sum_{k=n}^{m} a_k m_k(x_k) - \sum_{k=n}^{m} a_k \xi_k$ and recall that, since the correcting terms $a_k m_k(x_k)$ behave properly, $x_n$ will converge to zero (or to $\theta$) if $\sum_{n=1}^{\infty} a_n \xi_n$ converges almost surely (using Blum's methods [1], [2], for example). We write $\eta_k = a_k \xi_k I_{\{|a_k \xi_k| \leq d\}}$ for a fixed $d > 0$. Then

$$x_{m+1} - x_n = -\sum_{k=n}^{m} [a_k \xi_k - \eta_k] - \sum_{k=n}^{m} [\eta_k - E\eta_k] - \sum_{k=n}^{m} a_k m_k(x_k) - \sum_{k=n}^{m} E\eta_k . \tag{35}$$

Using Lemma 1, the assumed finiteness of $E|\xi_1|$ and our assumptions on $\{a_k\}$ guarantee that $P\{a_k \xi_k \neq \eta_k \text{ i.o.}\} = 0$ and also guarantee, via the Three Series Theorem (Theorem A on page 237 of Loève [9]), that $\sum_{k=1}^{\infty} (\eta_k - E\eta_k)$ converges almost surely. The convergence of $\{x_n\}$ thus depends on the precise relationship between the two sequences $\{a_k m_k(x_k)\}$ and $\{E\eta_k\}$. Suppose $P\{\xi_1 \leqq -d/(\max a_k)\} = 0$. Then

$$\sum_{k=1}^{\infty} E\eta_k = \sum_{k=1}^{\infty} a_k E\{\xi_k I_{\{|a_k \xi_k| \leq d\}}\} = -\sum_{k=1}^{\infty} a_k \int_{d/a_k}^{\infty} t\, dG(t) . \tag{36}$$

If, for example, for $t > 0$ we have $G'(t) = C(t \log t)^{-2}$, then $E|\xi_k| = \int |t|\, dG(t) < \infty$ but $\int_{(s,\infty)} t\, dG(t) = C(\log s)^{-1}$ so that (36) is infinite if $a_k = 1/k$. Clearly one can produce a sequence $m_k(\cdot)$ so that $-\sum_{k=n}^{m} a_k m_k(x_k) - \sum_{k=n}^{m} E\eta_k \to \infty$ as $m \to \infty$.

Now in the case given above $\sum_{k=1}^{n} a_k \xi_k \to -\infty$ almost surely since $\sum_{k=1}^{n} E\eta_k \to -\infty$. In spite of this, $\{x_n\}$ will converge almost surely if the sequence $\{a_k m_k(\cdot)\}$ behaves properly.

## REFERENCES

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

[2] Blum, J. R. (1958). A note on stochastic approximation. *Proc. Amer. Math. Soc.* **9** 404–407.

[3] Burkholder, D. L. (1956). On a class of stochastic approximation processes. *Ann. Math. Statist.* **27** 1044–1059.

[4] Dubins, Lester E. and Freedman, David A. (1965). A sharper form of the Borel-Cantelli lemma and the strong law. *Ann. Math. Statist.* **36** 800–807.

[5] DVORETSKY, A. (1956). On stochastic approximation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 39–55, Univ. of California Press.

[6] ENGELHARDT, MAX (1973). On upper bounds for variances in stochastic approximation. *SIAM J. Appl. Math.* **24** 145–151.

[7] GLADYSHEV, E. G. (1965). On stochastic approximation. *Theor. Probability Appl.* **10** 275–278.

[8] KRASULINA, T. P. (1969). On stochastic approximation processes with infinite variance. *Theor. Probability Appl.* **14** 522–526.

[9] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[10] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **25** 400–407.

[11] ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for nonnegative almost supermartingales and some applications. *Optimization Methods in Statistics*, 233–257. Academic Press, New York.

[12] SCHMETTERER, L. (1961). Stochastic approximation. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 587–609, Univ. of California Press.

[13] WASAN, M. T. (1969). *Stochastic Approximation*. University Press, Cambridge.

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF WISCONSIN
OSHKOSH, WISCONSIN 54901

DEPARTMENT OF MATHEMATICAL SCIENCES
S.U.N.Y.—BINGHAMTON
BINGHAMTON, NEW YORK 13901