ACCELERATED STOCHASTIC APPROXIMATION[1]

By Harry Kesten

Cornell University

1. Summary. Using a stochastic approximation procedure $\{X_n\}$, $n = 1, 2, \cdots$, for a value θ, it seems likely that frequent fluctuations in the sign of $(X_n - \theta) - (X_{n-1} - \theta) = X_n - X_{n-1}$ indicate that $| X_n - \theta |$ is small, whereas few fluctuations in the sign of $X_n - X_{n-1}$ indicate that X_n is still far away from θ. In view of this, certain approximation procedures are considered, for which the magnitude of the nth step (i.e., $X_{n+1} - X_n$) depends on the number of changes in sign in $(X_i - X_{i-1})$ for $i = 2, \cdots, n$. In theorems 2 and 3,

$$X_{n+1} - X_n$$

is of the form $b_n Z_n$, where Z_n is a random variable whose conditional expectation, given $X_1, \cdots, X_n$, has the opposite sign of $X_n - \theta$ and b_n is a positive real number. b_n depends in our processes on the changes in sign of

$$X_i - X_{i-1} (i \leqq n)$$

in such a way that more changes in sign give a smaller b_n. Thus the smaller the number of changes in sign before the nth step, the larger we make the correction on X_n at the nth step. These procedures may accelerate the convergence of X_n to θ, when compared to the usual procedures ([3] and [5]). The result that the considered procedures converge with probability one may be useful for finding optimal procedures. Application to the Robbins-Monro procedure (Theorem 2) seems more interesting than application to the Kiefer-Wolfowitz procedure (Theorem 3).

2. Statement of the theorem. The formulation of the theorem is similar to that of the theorem given by Dvoretzky [2]. Let θ be a real number and

$$T_n (n = 1, 2, \cdots)$$

be measurable transformations. Let X_1 and $Y_n (n = 1, \cdots)$ be random variables[2] and $\{a_n\}$ a sequence of positive numbers and define

$$X_{n+1}(\omega) = T_n(X_1(\omega), \cdots, X_n(\omega)) + b_n(\omega) Y_n(\omega). \tag{1}$$

The sequence $\{b_n(\omega)\}$ is selected in the following way from the sequence $\{a_n\}$

Received March 28, 1957.

[1] Note added in proof: The author learned recently that investigation of the above procedure had been suggested by Professor H. Robbins long ago.

[2] X_n, Y_n, and Z_n denote random variables, whereas x_n is used to denote values taken by the random variables.

$$(2)\qquad \begin{aligned} b_1 &= a_1, \\ b_2 &= a_2, \\ b_n &= a_{t(n)}, \end{aligned}$$

where

$$(3)\qquad t(n) = 2 + \sum_{i=3}^{n} \iota[(X_i - X_{i-1})(X_{i-1} - X_{i-2})]$$

and

$$\iota(x) = \begin{cases} 1 & \text{if} \quad x \leqq 0 \\ 0 & \text{if} \quad x > 0. \end{cases}$$

Thus, every time $(X_i - X_{i-1})$ differs in sign from $(X_{i-1} - X_{i-2})$ we take another a_n.

Let $\alpha_n(x_1, \cdots, x_n)$, $\beta_n(x_1, \cdots, x_n)$, $\gamma_n(x_1, \cdots, x_n)$ be nonnegative functions and put

$$(4)\qquad \epsilon_N = \sup_{\{x_k\}} \sum_{n=N}^{\infty} \beta_n(x_1, \cdots, x_n),$$

$$(5)^3\qquad \rho(\delta) = \inf_{n=1,2,\cdots} \inf_{\substack{|x_n - \theta| \geqq \delta \\ x_1, \cdots, x_{n-1} \text{ arbitrary}}} \frac{\gamma_n(x_1, \cdots, x_n)}{b_n}.$$

THEOREM 1. *If*

$$(6)\qquad |T_n(x_1, \cdots, x_n) - \theta| \leqq \begin{cases} (1 + \beta_n(x_1, \cdots, x_n))|x_n - \theta| \\ - \gamma_n(x_1, \cdots, x_n) \text{ when } (T_n - \theta)(x_n - \theta) > 0 \\ \alpha_n(x_1, \cdots, x_n) \text{ when } (T_n - \theta)(x_n - \theta) \leqq 0, \end{cases}$$

$$(7)\qquad \lim_{t(n) \to \infty} \alpha_n(x_1, \cdots, x_n) = 0$$ *uniformly, for all sequences* $x_1, x_2, \cdots$ *with* $t(n) \to \infty$,

$$(8)^3\qquad \lim_{n\to\infty} \frac{(x_n - \theta)\beta_n(x_1, \cdots, x_n)}{b_n} = 0$$ *uniformly, for all sequences* $x_1, x_2, \cdots$,

and

$$(9)\qquad \lim_{N\to\infty} \epsilon_N = 0,$$

$$(10)\qquad \rho(\delta) > 0 \quad \textit{for every positive } \delta,$$

$$(11)\qquad \sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \qquad \textit{and} \qquad a_{n+1} \leqq a_n,$$

[3] In (5), (8), and in (13), b_n depends on $x_1, \cdots, x_n$ as given in (2).

(12)[4] $E(Y_n \mid X_1, \cdots, X_n) = 0,$

$$E(Y_n^2 \mid X_1, \cdots, X_n) \leqq \sigma^2 \text{ with probability } 1,$$

(13)[3]
$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{\substack{0<|x_n-\theta|\leqq\tau \\ x_1,\cdots,x_{n-1} \text{ arbitrary}}} \cdot P\{T_n(X_1, \cdots, X_n) + b_n Y_n \geqq X_n \mid X_1 = x_1, \cdots, X_n = x_n\} > 0,$$

$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{\substack{0<|x_n-\theta|\leqq\tau \\ x_1,\cdots,x_{n-1} \text{ arbitrary}}} \cdot P\{T_n(X_1, \cdots, X_n) + b_n Y_n < X_n \mid X_1 = x_1, \cdots, X_n = x_n\} > 0,$$

then X_n converges to θ with probability 1.

Proof of Convergence. Without loss of generality we take $\theta = 0$. Also we assume in the following $E \mid X_1 \mid < \infty$. This can be done, because replacing X_1 by

$$X_1^1 = \begin{cases} X_1 & \text{if } \mid X_1 \mid < A \\ A & \text{if } \mid X_1 \mid \geqq A \end{cases}$$

changes the process only with a probability equal to

$$P\{\mid X_1 \mid > A\}.$$

By taking A large enough, this probability becomes arbitrary small. We frequently do not write all the arguments of the functions, e.g., we write β_n for $\beta_n(x_1, \cdots, x_n)$. We shall first prove several lemmas. From

$$E(Y_n \mid X_1, \cdots, X_n) = 0$$

and $E(Y_n^2 \mid X_1, \cdots, X_n) \leqq \sigma^2$ follows immediately.

Lemma 1. *There exists a function $p(\delta)$ with $0 < p(\delta) < 1$ for $\delta > 0$, and such that*

$$P\left\{Y_n \geqq \frac{\delta}{2} > 0 \mid X_1, \cdots, X_n\right\} \leqq 1 - p(\delta) < 1,$$

$$P\left\{Y_n \leqq -\frac{\delta}{2} < 0 \mid X_1, \cdots, X_n\right\} \leqq 1 - p(\delta) < 1.$$

Lemma 2.

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \geqq \frac{-\rho^1(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; \quad X_n \geqq \delta \text{ and } t(n) \geqq k\right\} \leqq 1 - p(\rho^1(\delta)),$$

[4] $P\{\cdot\mid\cdot\}$ and $E(\cdot\mid\cdot)$ denote conditional probabilities and conditional expectations respectively.

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \leqq \frac{\rho^1(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; \quad X_n \leqq -\delta \quad \text{and} \quad t(n) \geqq k\right\} \leqq 1 - p(\rho^1(\delta)),$$

where

$$\rho^1(\delta) = \begin{cases} \rho(\delta) & \text{when} \quad \rho(\delta)a_k \leqq \delta \\ \dfrac{\delta}{a_k} & \text{when} \quad \rho(\delta)a_k > \delta. \end{cases}$$

Proof. Since $t(n) \geqq k$, we have $b_n \leqq a_k$ and for $X_n \geqq \delta$

$$X_{n+1} \leqq \max[0, (1 + \beta_n)X_n - \gamma_n] + b_n Y_n \leqq X_n + b_n\left[\frac{\beta_n X_n}{b_n} - \rho^1(\delta)\right] + b_n Y_n = X_n + b_n\left[\frac{\beta_n}{b_n} X_n - \rho^1(\delta) + Y_n\right].$$

So by (8) and Lemma 1, we have

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \geqq -\frac{\rho^1(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; \quad X_n \geqq \delta, \quad t(n) \geqq k\right\}$$
$$\leqq \liminf_{n\to\infty} P\left\{Y_n \geqq \frac{\rho^1(\delta)}{2} - \epsilon \,\middle|\, X_1, \cdots, X_n; \quad X_n \geqq \delta, \quad t(n) \geqq k\right\}$$

for every $\epsilon > 0$.

Application of Lemma 1 gives the first inequality. Similarly we prove the second part of the lemma.

Lemma 3. *For every k and N*

$$P\{t(n) = k \quad \textit{for} \quad n \geqq N \quad \textit{and} \quad X_n \nrightarrow 0\} = 0$$

(i.e., *when $X_{n+1} - X_n$ does only change sign a finite number of times, then X_n converges to θ*).

Proof. When $t(n)$ is constant for $n \geqq N$, then X_n is monotonic for $n \geqq N$. Therefore $\{X_n\}$ converges (possibly to $+\infty$ or $-\infty$). Let the limit be positive, say X. But by Lemma 2 for every $\delta > 0$ and $\epsilon < [\rho^1(\delta)a_k]/2$,

$$\lim_{N\to\infty} P\{X_{n+1} - X_n > -\epsilon \quad \text{and} \quad X_n \geqq \delta \quad \text{and} \quad t(n) \geqq k \quad \text{for all } n \geqq N \cdot | \delta \leqq X_N \quad \text{and} \quad t(N) \geqq k\} = 0,$$

so the probability that $X > \delta$ is zero. Similarly the probability $X < -\delta$ is zero. Since δ is an arbitrary positive number, this proves the lemma.

This lemma allows us to limit ourselves in the sequel to those sequences with $t(n) \to \infty$ and therefore $b_n \to 0$.

Lemma 4. *Let δ be a fixed positive number. Then there exist positive numbers*

n_0 and t_0 such that, whenever $n \geqq n_0$, $t(n) \geqq t_0$ and $| X_n | \geqq \delta$, one has

$$E\{ | X_{n+1} | \, | X_1, \cdots, X_n ; \quad | X_n | \geqq \delta\} \leqq | X_n | - \frac{b_n}{4} \rho(\delta).$$

PROOF. Choose t_0 such that $\alpha_n(X_1, \cdots, X_n) \leqq \delta/2$ for $t(n) \geqq t_0$ and

$$a_{t_0} \leqq \min \left(\frac{2\delta}{4\sigma + \rho(\delta)}, \frac{\delta\rho(\delta)}{16\sigma^2}, \frac{\delta}{\rho(\delta)} \right). \tag{14}$$

Then $b_n \leqq a_{t_0}$ for $t(n) \geqq t_0$. We distinguish two cases

(a)
$$| T_n(X_1, \cdots, X_n) | \leqq \frac{\delta}{2}.$$

$$E\left\{ | X_{n+1} | \, | X_1, \cdots, X_n ; | X_n | \geqq \delta, | T_n | \leqq \frac{\delta}{2} \right\}$$
$$\leqq \frac{\delta}{2} + b_n E | Y_n | \leqq \frac{\delta}{2} + b_n \sigma \leqq \delta - \frac{b_n \rho(\delta)}{4} \leqq | X_n | - \frac{b_n \rho(\delta)}{4},$$

(b)
$$| T_n(X_1, \cdots, X_n) | > \frac{\delta}{2}.$$

As $\alpha_n(X_1, \cdots, X_n) \leqq \delta/2$ for $t(n) \geqq t_0$, we must have $T_n \cdot X_n > 0$ (cf. (6)). Let $X_n \geqq \delta$. Denote the distribution function of $Y_n(X_1, \cdots, X_n)$ by $H_n(y \mid X)$. As $X_{n+1} = T_n + b_n Y_n$, we have by (12) and (14)

$$E\left\{ | X_{n+1} | \,\middle|\, X_1, \cdots, X_n ; X_n \geqq \delta, T_n \geqq \frac{\delta}{2} \right\}$$
$$= \int_{-T_n/b_n}^{\infty} (T_n + b_n y) \, dH_n(y \mid X) - \int_{-\infty}^{-T_n/b_n} (T_n + b_n y) \, dH_n(y \mid X)$$
$$\leqq T_n + b_n \left[\int_{-T_n/b_n}^{\infty} y \, dH_n(y \mid X) - \int_{-\infty}^{-T_n/b_n} y \, dH_n(y \mid X) \right]$$
$$= T_n - 2b_n \int_{-\infty}^{-T_n/b_n} y \, dH_n(y \mid X)$$
$$\leqq T_n + 2b_n \left[\int_{-\infty}^{-T_n/b_n} y^2 \, dH_n(y \mid X) \int_{-\infty}^{-T_n/b_n} dH_n(y \mid X) \right]^{1/2}$$
$$\leqq T_n + 2b_n \sigma \frac{b_n \sigma}{T_n} \leqq T_n + \frac{4b_n^2 \sigma^2}{\delta} \leqq T_n + b_n \frac{\rho(\delta)}{4}.$$

But by (8),

$$| T_n | \leqq | X_n | + b_n \left\{ \frac{\beta_n | X_n |}{b_n} - \rho(\delta) \right\} \leqq | X_n | - b_n \frac{\rho(\delta)}{2}$$

for sufficiently large n, say $n \geqq n_0$. For $X_n \leqq -\delta$, the proof is similar. Thus, in all cases

$$E\left\{| X_{n+1} | \middle| X_1, \cdots, X_n ; | X_n | \geqq \delta\right\} \leqq | X_n | - \frac{b_n \rho(\delta)}{4}.$$

LEMMA 5. *For every* $0 < \delta < \delta' < \delta''$

$$P\{\delta < \liminf | X_n | < \delta' \quad \textit{and} \quad \delta'' < \limsup | X_n | \quad \textit{and} \quad t(n) \to \infty\} = 0.$$

PROOF. Choose t_0 and n_0, corresponding to δ as introduced in the preceding lemma. Assume now

$$P\{\delta < \liminf | X_n | < \delta' \quad \text{and} \quad \delta'' < \limsup | X_n | \quad \text{and} \quad t(n) \to \infty\} > 0.$$

Then there exist an $n_1 \geqq n_0$ and $t_1 \geqq t_0$ such that

$$(15) \quad P\{\delta < \liminf | X_n | < \delta' \quad \text{and} \quad \delta'' < \limsup | X_n | \quad \text{and} \quad | X_n | > \delta \quad \text{for all} \quad n \geqq n_1 \quad \text{and} \quad b_{n_1} = a_{t_1}\} > 0.$$

Now introduce a new process.

$$Z_i = | X_i | \quad \text{if} \quad i = 1, \cdots, n_1$$

and

$$Z_{n_1+i} = \begin{cases} | X_{n_1+i} | & \text{if} \quad \delta < Z_{n_1+j} \quad \text{for} \quad j = 0, 1, \cdots, i-1 \quad \text{and} \quad b_{n_1} = a_{t_1}, \\ 0 & \text{otherwise.} \end{cases}$$

Unless $b_{n_1} \neq a_{t_1}$ or $| X_j | \leqq \delta$ for some $j \geqq n_1$, we have always $Z_i = | X_i |$, and thus, by (15), also

$$(16) \qquad P\{\delta \leqq \liminf Z_n < \delta' \quad \text{and} \quad \delta'' < \limsup Z_n\} > 0.$$

But by Lemma 4

$$0 \leqq E(Z_{n+1} | X_1, \cdots, X_n) \leqq EZ_n \qquad \text{for} \quad n \geqq n_1.$$

So application of the semimartingale convergence theorem (Loève [4], p. 393) shows that (16) cannot be true. This proves the lemma.

LEMMA 6. $P\{\liminf | X_n | = \infty$ *and* $t(n) \to \infty\} = 0.$

PROOF. If the proposition were not true, we could find, analogous to the last proof, a process Z_i with

$$(17) \quad 0 \leqq E(Z_{n+1} | X_1, \cdots, X_n) \leqq EZ_n \quad \text{for sufficiently large } n, \quad \text{say} \quad n \geqq n_2,$$

and

$$(18) \qquad P\{\liminf Z_n = \infty\} > 0.$$

But as we took $E | X_1 | < \infty$, one would have

$$(19) \qquad E | Z_{n_2} | < \infty.$$

However, (17) and (19) together are in contradiction with (18). This proves the lemma.

From Lemmas 3, 5, and 6 one may conclude that with probability 1 either that $\liminf |X_n| = 0$ or $|X_n|$ converges to a finite positive number. We now prove that the last possibility has probability zero.

LEMMA 7.

$$P\{|X_n| \text{ converges to } X \text{ and } 0 < \delta < X < \delta' < \infty \text{ and } t(n) \to \infty\} = 0.$$

PROOF. Choose n_0 and t_0 corresponding to δ, as introduced in Lemma 4. Assume now

$$(20)\quad P\{|X_n| \text{ converges to } X \text{ and } 0 < \delta < X < \delta' < \infty \text{ and } t(n) \to \infty\} > 0.$$

Again there exist an $n_1 \geqq n_0$ and a $t_1 \geqq t_0$ such that

$$(21)\quad P\{\delta < |X_n| < \delta' \text{ for all } n \geqq n_1 \text{ and } b_{n_1} = a_{t_1}\} > 0 \text{ and } a_{t_1}\rho(\delta) \leqq \delta.$$

By Lemma 2 we can choose n_1 and t_1 so that at the same time for $n \geqq n_1$

$$(22)\quad P\left\{X_{n+1} - X_n \geqq -\frac{\rho(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; X_n \geqq \delta, t(n) \geqq t_1\right\} \leqq 1 - \frac{p(\rho(\delta))}{2},$$

$$(23)\quad P\left\{X_{n+1} - X_n \leqq \frac{\rho(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; X_n \leqq -\delta, t(n) \geqq t_1\right\} \leqq 1 - \frac{p(\rho(\delta))}{2}.$$

As before we construct a new process.

$$Z_i = |X_i| \quad \text{if} \quad i = 1, \cdots, n_1$$

and

$$Z_{n_1+i} = \begin{cases} |X_{n_1+i}| \text{ if } \delta < Z_{n_1+j} < \delta' & \text{for } j = 0, \cdots, i-1 \text{ and } b_{n_1} = a_{t_1}, \\ Z_{n_1+i-1} - a_{t_1+i-1}\dfrac{\rho(\delta)}{4} & \text{otherwise.} \end{cases}$$

From (21) follows

$$(24)\quad P\{\delta < Z_n < \delta' \text{ for all } n \geqq n_1\} > 0,$$

and thus,

$$(25)\quad P\left\{\left|\sum_{k=n_1}^{n} (Z_{k+1} - Z_k)\right| < 2(\delta' - \delta) \text{ for all } n \geqq n_1\right\} > 0.$$

Denote

$$E(Z_{k+1} - Z_k \mid X_1, \cdots, X_k) \quad \text{by} \quad m_k(X_1, \cdots, X_k) \qquad (= m_k \text{ for short}).$$

By Lemma 4 and the construction of the Z-process,

$$m_k(X_1, \cdots, X_k) \leqq -\frac{c_k}{4}\rho(\delta) \text{ for } k \geqq n_1, \tag{26}$$

where

$$c_{n_1+i} = \begin{cases} b_{n_1+i} & \text{if } \delta < Z_j < \delta' \quad \text{for} \quad j = 0, \cdots, i \quad \text{and} \quad b_n = a_{t_1}, \\ a_{n_1+i} & \text{otherwise.} \end{cases}$$

Further for $k \geqq n_1$,

$$\operatorname{var}(Z_{k+1} - Z_k \mid X_1, \cdots, X_k) \leqq c_k^2\left[\frac{\rho^2(\delta)}{16} + EY_k^2\right] \leqq c_k^2 C, \tag{27}$$

where

$$C = \frac{\rho^2(\delta)}{16} + \sigma^2.$$

In addition

$$\sum_{k=n_1}^{n} c_k \geqq \sum_{k=0}^{n-n_1} a_{t_1+k}. \tag{28}$$

By (25),

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1} - Z_k - m_k)\right| \geqq \left|\sum_{k=n_1}^{n} m_k\right| - 2(\delta' - \delta) \quad \text{for all } n \geqq n_1\right\} > 0, \tag{29}$$

and thus, by (26) and (28),

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1} - Z_k - m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta) \quad \text{for all } n \geqq n_1\right\} > 0. \tag{30}$$

But for

$$\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta) > 0$$

we have by Tchebycheff's inequality and (27)

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1}-Z_k-m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k}-2(\delta'-\delta)\right\}$$

(32)
$$\leqq \frac{C\sum_{k=n_1}^{n} Ec_k^2}{\left\{\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k}-2(\delta'-\delta)\right\}^2},$$

(32)
$$\sum_{k=n_1}^{n} Ec_k^2 \leqq \sum_{k=0}^{n-n_1} a_{t_1+k}^2 Er_{t_1+k},$$

where

$$r_{t_1+k} = \text{number of times } c_{n_1+i} = a_{t_1+k}\,.$$

As soon as the Z_i process differs from the $|X_i|$ process, we don't keep the same a_{t_1+k} for more than one step. Therefore $Er_{t_1+k} \leqq 1 +$ expected number of times that $\{c_{n_1+i} = a_{t_1+k}$ and $\delta < Z_{n_1+j} < \delta'$ for $j = 1, \cdots, i$ and $b_{n_1} = a_{t_1}\}$ occurs.

If $\delta < X_{n_1+i} < \delta'$, then by (22),

$$P\{X_{n_1+i+1} > X_{n_1+i} \mid \delta < |X_{n_1+j}| < \delta' \quad j = 1, \cdots, i \text{ and } b_{n_1} = a_{t_1} \text{ and } X_{n_1+i} > \delta\} \leqq 1 - \frac{p(\rho(\delta))}{2},$$

and

$$P\left\{\begin{matrix} X_{n_1+i+l} > X_{n_1+i+l-1} \mid \delta < |X_{n_1+j}| < \delta' \quad j = 1, \cdots, i \\ l = 1, \cdots, s \end{matrix} \text{ and } b_{n_1} = a_{t_1} \text{ and } X_{n_1+i} > \delta\right\} \leqq \left\{1 - \frac{p(\rho(\delta))}{2}\right\}^s.$$

As we pick a new a_j as soon as $(X_i - X_{i-1})(X_{i-1} - X_{i-2}) \leqq 0$, we have: Expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \quad j = 1, \cdots, i \quad \text{and} \quad b_{n_1} = a_{t_1}\}$$

under the condition that $X_{j+1} \geqq X_j > \delta$ for the first $j \geqq n_1$ with $b_j = a_{t_1+k}$, is at most

(33)
$$2 + \left(1 - \frac{p(\rho(\delta))}{2}\right) + \left(1 - \frac{p(\rho(\delta))}{2}\right)^2 + \cdots \leqq \frac{3}{p(\rho(\delta))}.$$

The case where $X_{j+1} < X_j$ at the first time that $b_j = a_{t_1+k}$ is more difficult. Let us divide the interval (δ, δ') in

$$\left[\frac{2(\delta'-\delta)}{\rho(\delta)a_{t_1+k}}\right] + 1$$

non-overlapping intervals[5] I_t with

[5] $[a]$ is the largest integer $\leqq a$.

$$\text{length } (I_t) < \frac{\rho(\delta)a_{t_1+k}}{2}\left(t = 1, 2, \cdots, \left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1\right).$$

Expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \, j = 1, \cdots, i, \, Z_{n_1+i} \, \varepsilon \, I_t\}$$

under the condition that $X_{j+1} < X_j$, $X_j > \delta$ for first $j \geqq n_1$, with $b_j = a_{t_1+k}$, is at most

$$1 + \left(1 - \frac{p(\rho(\delta))}{2}\right) + \left(1 - \frac{p(\rho(\delta))}{2}\right)^2 \cdots = \frac{2}{p(\rho(\delta))}.$$

This can be proved analogous to (33) using (22) and the fact that

$$\text{length } (I_t) < [\rho(\delta)a_{t_1+k}]/2.$$

As there are

$$\left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1$$

intervals I_t, expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \, j = 1, \cdots, i, \quad \text{and} \quad b_{n_1} = a_{t_1}\}$$

under the condition that $X_{j+1} < X_j$, $X_j > \delta$ for the first $j \geqq n_1$ with

$$b_j = a_{t_1+k},$$

is at most

$$\frac{2\left\{\left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1\right\}}{p(\rho(\delta))}.$$

Similar estimates are valid when $X_j < -\delta$ for the first $j \geqq n_1$ with $b_j = a_{t_1+k}$.

As $a_{t_1+k} \to 0(k \to \infty)$, we can find a positive constant D such that

$$Er_{t_1+k} \leqq \frac{D}{a_{t_1+k}}.$$

By (31) and (32), it follows that

$$P\left\{\left|\sum_{n=n_1}^{n} (Z_{k+1} - Z_k - m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta)\right\}$$

(34)
$$\leqq \frac{CD \sum_{k=0}^{n-n_1} a_{t_1+k}}{\left\{\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta)\right\}^2}.$$

As $\sum_{k=0}^{\infty} a_{t_1+k} = \infty$, the right hand side of (34) tends to zero when $n \to \infty$

and therefore (29) cannot be true and

$$P\{|X_n| \text{ converges to } X \neq 0 \quad \text{and} \quad t(n) \to \infty\} = 0.$$

Combining the remark after Lemma 6, and Lemma 7 we proved

$$P\{\liminf |X_n| = 0\} = 1. \tag{35}$$

Until now we only used that a_n tends monotonically to zero and

$$\sum_{n=1}^{\infty} a_n = \infty,$$

but not yet $\sum_{n=1}^{\infty} a_n^2 < \infty$.

LEMMA 8. *Define*

$$s(n) = \begin{cases} 1 & \text{if} \quad T_n(X_1, \cdots, X_n)\cdot X_n > 0 \\ -1 & \text{if} \quad T_n(X_1, \cdots, X_n)\cdot X_n \leqq 0, \end{cases}$$

$$Y_n^1 = Y_n \prod_{j=1}^{n} s(j),$$

$$d(m, m-1) = 1$$

$$d(m, n) = \prod_{j=m}^{n} (1 + \beta_j) \qquad (n \geqq m),$$

$$S(m+1, n) = \sum_{j=m}^{n} d(j+1, n) b_j Y_j^1.$$

Then the conditions

$$\alpha_{m+j-1}(X_1, \cdots, X_{m+j-1}) \leqq \frac{\epsilon}{8} \qquad j = 1, \cdots, k,$$

$$d(m, \infty) \leqq \frac{3}{2},$$

$$|X_m| \leqq \frac{\epsilon}{4},$$

$$|X_{m+j}| > \frac{\epsilon}{4} \qquad j = 1, \cdots, k-1,$$

and

$$\sup_{n \geqq m} |S(m+1, n)| \leqq \frac{\epsilon}{16}$$

imply

$$|X_{m+j}| \leqq \frac{\epsilon}{2} \qquad j = 1, \cdots, k.$$

The proof follows immediately from Wolfowitz [6], p. 1154. We need the following

COROLLARY. *If* $t(m)$ *is so large that*

$$\alpha_{m+j-1}(X_1, \cdots, X_{m+j-1}) \leqq \frac{\epsilon}{8} \qquad j = 1, 2, \cdots$$

and if

$$d(m, \infty) \leqq \frac{3}{2},$$

then

$$P\left\{ | X_{m+j} | > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, | X_m | \leqq \frac{\epsilon}{4}\right\}$$
$$\leqq P\left\{ \sup_{n_1, n_2 \geqq m} | S(n_1 + 1, n_2) | \geqq \frac{\epsilon}{16}\right\}$$
$$\leqq \frac{32^2}{\epsilon^2} \sum_{j=m}^{\infty} \text{var } \{d(j + 1, n) b_j Y_j^1\} \leqq \left(\frac{48\sigma}{\epsilon}\right)^2 \sum_{j=m}^{\infty} Eb_j^2.$$

PROOF OF THEOREM 1. In view of Lemma 3, we only have to prove

$$P\{\lim \sup | X_n | > 0 \quad \text{and} \quad t(n) \to \infty \} = 0.$$

By condition (13)

$$2\zeta = \min\Big(\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{0<|x_n|\leqq\tau} P\{X_{n+1} - X_n \geqq 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\},$$
$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{0<|x_n|\leqq\tau} P\{X_{n+1} - X_n < 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\}\Big) > 0$$

Take $\xi > 0$ and n_2 such that

$$P\{X_{n+1} - X_n \geqq 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\} > \zeta > 0,$$
$$P\{X_{n+1} - X_n < 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\} > \zeta > 0 \tag{36}$$

for $0 < | x_n | \leqq \xi$ and $n \geqq n_2$.

Choose an $\epsilon \leqq \xi$ and t_2 such that

$$\alpha_n(X_1, \cdots, X_n) \leqq \frac{\epsilon}{8} \quad \text{when} \quad t(n) \geqq t_2,$$

and let

$$d(n_2, \infty) \leqq \frac{3}{2}.$$

Let now for some $m \geqq n_2$

$$| X_m | \leqq \frac{\epsilon}{4} \quad \text{and} \quad t(m) \geqq t_2.$$

Construct the following process

$$Z_k^{(m)} = X_k \quad \text{if} \quad k = 1, \cdots, m$$

and

$$Z_{m+i}^{(m)} = T_{m+i-1}(Z_1, \cdots, Z_{m+i-1}) + c_{m+i-1}Y_{m+i-1}(X_1, \cdots, X_{m+i-1})(i = 1, 2, \cdots),$$

where the c's are determined in the following way:

$$c_m = b_m = a_{t(m)}$$

$$(37)\ c_{m+i} = \begin{cases} c_{m+i-1} & \text{if } \ |Z_{m+j}| \leqq \frac{\epsilon}{2} \quad j = 0, 1, \cdots, i \\ & \text{and } \ (Z_{m+i} - Z_{m+i-1})(Z_{m+i-1} - Z_{m+i-2}) > 0 \\ a_l & \text{if } \ |Z_{m+j}| \leqq \frac{\epsilon}{2} \quad j = 0, 1, \cdots, i \\ & \text{and } \ (Z_{m+i} - Z_{m+i-1})(Z_{m+i-1} - Z_{m+i-2}) \leqq 0 \\ & \text{and } \ c_{m+i-1} = a_{l-1} \\ a_{t(m)+i} & \text{otherwise.} \end{cases}$$

Then Er_l = expected number of times $c_{m+j} = al$ is zero when $l < t(m)$. For $l \geqq t(m)$ it is at most

$$(38) \qquad 1 + (1 - \zeta) + (1 - \zeta)^2 \cdots = \frac{1}{\zeta}.$$

In fact from (36) and (37),

$$P\{c_{m+j} = c_{m+j-1}\} \leqq 1 - \zeta.$$

Using (38) and applying the corollary of Lemma 8 to the $Z(m)$ process, and thus replacing the b's by the c's, one finds for $m \geqq n_2$,

$$P\left\{|X_{m+j}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |X_m| \leqq \frac{\epsilon}{4}, t(m) \geqq t_2\right\}$$

$$\leqq P\left\{|Z_{m+j}^{(m)}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |Z_m^{(m)}| \leqq \frac{\epsilon}{4}, t(m) \geqq t_2\right\}$$

$$\leqq \left(\frac{48\sigma}{\epsilon}\right)^2 \sum_{n=t_2}^{\infty} \frac{a_n^2}{\zeta}.$$

Now choose $t_3 \geqq t_2$ such that

$$\left(\frac{48\sigma}{\epsilon}\right)^2 \frac{1}{\zeta} \sum_{n=t_3}^{\infty} a_n^2 \leqq \frac{\epsilon}{2},$$

and $n_3 \geqq n_2$ such that

$$P\left\{\left[\,|X_n| > \frac{\epsilon}{4} \quad \text{for all} \quad n \geqq n_3 \quad \text{or} \quad t(n_3) < t_3\right] \quad \text{and} \quad t(n) \to \infty\right\} \leqq \frac{\epsilon}{2}$$

(such an n_3 exists by (35)). Then

$$P\{\limsup |X_n| > \epsilon \quad \text{and} \quad t(n) \to \infty\}$$
$$\leqq P\left\{\left[\,|X_n| > \frac{\epsilon}{4} \quad \text{for all} \quad n \geqq n_3 \quad \text{or} \quad t(n_3) < t_3\right] \quad \text{and} \quad t(n) \to \infty\right\}$$
$$+ \sum_{m=n_3}^{\infty} P\left\{X_m \quad \text{is the first after} \quad X_{n_3-1} \quad \text{with} \quad |X_m| \leqq \frac{\epsilon}{4}\right.$$
$$\left.\text{and} \quad t(n_3) \geqq t_3 \quad \text{and} \quad \max_{k \geqq m} |Z_k(m)| > \frac{\epsilon}{2}\right\}$$
$$\leqq \frac{\epsilon}{2} + \frac{\epsilon}{2} \sum_{m=n_3}^{\infty} P\left\{X_m \quad \text{is the first after} \quad X_{n_3-1} \quad \text{with} \quad |X_m| \leqq \frac{\epsilon}{4}\right\} \leqq \epsilon.$$

As the only restriction on ϵ is $\epsilon \leqq \xi$, this proves the theorem.

3. Applications.

Accelerated Robbins-Monro procedure.

THEOREM 2. *Let X_1 and $Y(x)$ be random variables and $\{a_n\}$ a sequence of positive numbers and define*

$$X_{n+1}(\omega) = X_n(\omega) - b_n(M(X_n) - \alpha) + b_n Y(X_n).$$

The sequence $\{b_n\}$ is selected in the following way from the sequence $\{a_n\}$:

$$b_1 = a_1\,,$$
$$b_2 = a_2\,,$$
$$b_n = a_{t(n)}\,,$$

(cf. (2) and (3)).

If $M(x)$ is a measurable function satisfying

(39) $$(x - \theta)(M(x) - \alpha) > 0 \qquad \text{for } x \neq \theta,$$

(40) $$\inf_{\delta \leqq |x-\theta| < \infty} |M(x) - \alpha| > 0 \qquad \text{for every} \quad \delta > 0,$$

(41) $|M(x) - \alpha| \leqq c + d\,|x - \theta|$ for some positive constants c and d,

and if

(42) $$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \text{and} \quad a_{n+1} \leqq a_n\,,$$

(43) $$E(Y(X_n)|\, X_1\,, \cdots, X_n) = 0, \qquad E(Y^2(X_n)|\, X_1\,, \cdots, X_n) \leqq \sigma^2$$

with probability 1,

$$(44)\qquad \begin{aligned} &\lim_{\tau\to 0} \inf_{0<|x-\theta|\leqq\tau} P\{Y(x) - M(x) + \alpha \geqq 0\} > 0 \\ &\lim_{\tau\to 0} \inf_{0<|x-\theta|\leqq\tau} P\{Y(x) - M(x) + \alpha < 0\} > 0, \end{aligned}$$

then

$$P\{X_n \text{ converges to } \theta\} = 1.$$

PROOF. Take

$$\alpha_n = \begin{cases} b_n(c + d \mid x_n - \theta \mid) & \text{for } b_n d > 1 \\ b_n c & \text{for } b_n d \leqq 1, \end{cases}$$

$$\beta_n \equiv 0,$$

$$\gamma_n = b_n \mid M(x_n) - \alpha \mid$$

in Theorem 1.

The process as described in Theorem 2 gives a stochastic approximation method for the point θ which uses the number of changes in sign in

$$(X_i - X_{i-1})(X_{i-1} - X_{i-2}) \qquad i = 3, \cdots, n$$

to determine $(X_{n+1} - X_n)$. We only reduce b_n and thus the magnitude of

$$X_{n+1} - X_n$$

when the last two corrections $X_n - X_{n-1}$ and $X_{n-1} - X_{n-2}$ had different signs. As indicated in the summary this process may pull X_n to θ faster (for large $| X_n - \theta |$) than the Robbins-Monro procedure. In Theorem 2 the conditions are slightly stronger than for the Robbins-Monro process as given by Blum [1]. Blum does not need

$$a_{n+1} \leqq a_n$$

or (44) and has

$$(40a)\qquad \inf_{\delta\leqq|x-\theta|\leqq\delta'} | M(x) - \alpha | > 0 \quad \text{for every} \quad 0 < \delta \leqq \delta' < \infty$$

instead of (40).

One can easily give an example to show that we cannot replace (40) by (40a) and the following example shows that (44) cannot be dispensed with.

EXAMPLE. Take

$$\theta = 0, \qquad \alpha = 0, \qquad a_n = \frac{1}{n}.$$

Let $\{x_{2n+1,0}\}(n = 0, 1, \cdots)$ be a sequence of real numbers such that

$$x_{2n+1,0} \neq x_{2m+1,0} \quad \text{for} \quad n \neq m \quad \text{and} \quad 1 \leqq x_{2n+1,0} \leqq 2.$$

Let $\{x_{2n,0}\}(n = 1, 2, \cdots)$ be a sequence of real numbers such that

$$x_{2n,0} \neq x_{2m,0} \quad \text{for} \quad n \neq m \quad \text{and} \quad -2 \leqq x_{2n,0} \leqq -1.$$

We now construct recursively sequences $\{x_{n,k}\}(k = 0, 1, \cdots)$. Put

$$Z(x) = M(x) - Y(x)$$

and

$$Z_{n,k} = Z(x_{n,k}),$$

so

$$X_{n+1} = X_n - b_n Z(X_n)$$

We start with $\{x_{1,k}\}$ by taking

$$Z_{1,0} = \begin{cases} z'_{1,0} = x_{1,0} - x_{2,0} & \text{with probability } \frac{1}{2} \\ z''_{1,0} & \text{with probability } \frac{1}{2}, \end{cases}$$

where $\frac{1}{2}x_{1,0} < z''_{1,0} < x_{1,0}$. Further take $x_{1,1} = x_{1,0} - z''_{1,0}$, and in general

$$Z_{1,k} = \begin{cases} z'_{1,k} = x_{1,k} - x_{2,0} & \text{with probability } \frac{1}{2} \\ z''_{1,k} & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$x_{1,k+1} = x_{1,k} - z''_{1,k},$$

where

$$\tfrac{1}{2}x_{1,k} < z''_{1,k} < x_{1,k}.$$

For $n > 1$ we take

$$Z_{n,0} = \begin{cases} z'_{n,0} = (n-1)(x_{n,0} - x_{n+1,0}) & \text{with probability } \dfrac{1}{2n^2} \\ z''_{n,0} & \text{with probability } 1 - \dfrac{1}{2n^2}, \end{cases}$$

$$x_{n,1} = x_{n,0} - \frac{1}{n-1} z''_{n,0},$$

where $z''_{n,0}$ is such that

$$z''_{n,0} \cdot x_{n,0} > 0 \quad \text{and} \quad \tfrac{1}{2} \mid x_{n,0} \mid < \mid z''_{n,0} \mid < \mid x_{n,0} \mid$$

and $x_{n,1}$ is not equal to any $x_{m,i}$ with $m < n$. Further for $k > 0$,

$$Z_{n,k} = \begin{cases} z'_{n,k} = n(x_{n,k} - x_{n+1,0}) & \text{with probability } \dfrac{1}{2n^2} \\ z''_{n,k} & \text{with probability } 1 - \dfrac{1}{2n^2}, \end{cases}$$

$$x_{n,k+1} = x_{n,k} - \frac{1}{n} z''_{n,k},$$

where $z''_{n,k}$ is such that

$$z''_{n,k} \cdot x_{n,k} > 0, \qquad \tfrac{1}{2} | x_{n,k} | < | z''_{n,k} | < | x_{n,k} |$$

and $x_{n,k+1}$ is not equal to any $x_{m,i}$ with $m < n$. We take $M(x_{n,k}) = EZ_{n,k}$ and

$$Y(x_{n,k}) = Z_{n,k} - M(x_{n,k}).$$

For $x \neq x_{n,k}$ for all n, k, we take $M(x)$ and $Y(x)$ in any way such that the conditions of Theorem 2, except (44), are satisfied.

Take now $X_1 = x_{1,0}$ with probability 1. By the choice of $z'_{n,k}$, we get the value $x_{n+1,0}$ as soon as $Z_{n,k}$ takes the value $z'_{n,k}$. But for every n, with probability 1, Z will take once the value $z'_{n,k}$. Therefore with probability 1, all the values $x_{n,0}$ occur in the sequence $\{X_n\}$ and thus,

$$P\{X_n \text{ converges to } 0\} = 0.$$

Accelerated Kiefer-Wolfowitz procedure.

THEOREM 3. *Let X_1 and $Y(x)$ be random variables and let $\{a_n\}$ be a sequence of positive numbers and u some positive constant and define*

$$X_{n+1}(\omega) = X_n(\omega) - b_n[M(X_n - u) - M(X_n + u)] + b_n[Y(X_n - u) - Y(X_n + u)].$$

The sequence $\{b_n\}$ is selected in the following way from the sequence $\{a_n\}$:

$$b_1 = a_1,$$

$$b_2 = a_2,$$

$$b_n = a_{t(n)},$$

(cf. (2) and (3)). *If $M(x)$ is a measurable function, satisfying*

$$\inf_{x-\theta \geqq \delta} \{M(x - u) - M(x + u)\} > 0 \tag{45}$$

$$\inf_{x-\theta \leqq -\delta} \{M(x - u) - M(x + u)\} < 0$$

for every $\delta > 0$,

$$| M(x - u) - M(x + u)| \leqq c + d \, | x - \theta | \tag{46}$$

for some positive constants c and d, and if

$$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \text{and} \quad a_{n+1} \leqq a_n, \tag{47}$$

$$\begin{aligned} &E(Y(X_n - u) - Y(X_n + u) \mid X_1, \cdots, X_n) = 0, \\ &E((Y(X_n - u) - Y(X_n + u))^2 \mid X_1, \cdots, X_n) \leqq \sigma^2 \end{aligned} \tag{48}$$

with probability 1,

$$\begin{aligned} &\lim_{\tau \to 0} \inf_{0 < |x-\theta| \leqq \tau} P\{Y(x-u) - Y(x+u) - M(x-u) + M(x+u) \\ &\qquad\qquad\qquad \geqq 0\} > 0 \\ &\lim_{\tau \to 0} \inf_{0 < |x-\theta| \leqq \tau} P\{Y(x-u) - Y(x+u) - M(x-u) + M(x+u) \\ &\qquad\qquad\qquad < 0\} > 0, \end{aligned} \tag{49}$$

then

$$P\{X_n \text{ converges to } \theta\} = 1.$$

Proof. Take

$$\alpha_n = \begin{cases} b_n(c + d \mid x_n - \theta \mid) & \text{for } b_n d > 1 \\ b_n c & \text{for } b_n d \leqq 1, \end{cases}$$

$$\beta_n \equiv 0,$$

$$\gamma_n = b_n \mid M(x_n - u) - M(x_n + u) \mid$$

in Theorem 1.

Remark. Theorem 3 is also implied by Theorem 2. The procedure in Theorem 3 requires u to be independent of n, and therefore differs from the usual Kiefer-Wolfowitz procedure ([3]). Also condition (45) does not imply that $M(x)$ has a maximum, or if it has one, that θ is the location of the maximum. However, for every y with $| y - \theta | > u$, there exists an x with $| x - \theta | \leqq u$, such that $M(x) > M(y)$.

Acknowledgement. The author wishes to thank Professor J. Wolfowitz for suggesting the possibility of generalizing Theorem 2 and for his many other helpful and critical remarks.

REFERENCES

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

[2] A. Dvoretzky, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, 1956, pp. 39–55.
[3] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.
[4] M. Loève, *Probability Theory*, D. van Nostrand, 1955.
[5] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.
[6] J. Wolfowitz, "On stochastic approximation methods," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1151–1156.