$$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty. \tag{4.3}$$

Define a recursive approximation scheme as follows. Let x_1 be arbitrary and define

$$x_{n+1} = x_n + a_n z_n \tag{4.4}$$

where $z_n = +1$ if $y_n \leqq \alpha$ and $z_n = -1$ if $y_n > \alpha$, and y_n is a random variable distributed according to $H(y \mid x_n)$. Then, by applying Theorem 1 with $\alpha = 0$ and $y_n = -z_n$, we obtain

THEOREM 3. *If conditions* (4.1), (4.2), *and* (4.3) *hold, then* $P\{\lim x_n = \theta\} = 1$.

I should like to thank Mr. Lucien LeCam for many helpful discussions concerning this problem. I should also like to thank the referee for pointing out that the condition of uniform boundedness of $M(x)$ in Section 2 could be replaced by the present condition **(2.1).**

REFERENCES

[1] H. ROBBINS AND S. MONRO, "A stochastic approximation method," *Ann. Math. Stat.* Vol. 22 (1951), pp. 400–407.
[2] J. WOLFOWITZ, "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 457–461.
[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, "On almost sure convergence," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 279–303.

A NOTE ON THE ROBBINS-MONRO STOCHASTIC APPROXIMATION METHOD[1]

BY GOPINATH KALLIANPUR

Institute for Advanced Study and University of California, Berkeley[2]

Introduction. The almost certain convergence of the RM process and related stochastic approximation procedures is proved by Blum [1] in a paper appearing elsewhere in this issue. In the present note we consider the method originally proposed by Robbins and Monro [2] with a further restriction on the constants a_n. Our aim is to obtain, by elementary methods, an estimate of the order of magnitude of $b_n = E\,(x_n - \theta)^2$. This estimate is sharp enough to enable us to prove strong convergence for certain types of sequences a_n. The method adopted in [1], while being more general, does not yield information about the behavior

Received 1/19/53, revised 12/7/53.

[1] This work was begun with the partial support of the Office of Naval Research when the author was at the Statistical Laboratory, University of California, Berkeley.

[2] At present at the Indian Statistical Institute.

of b_n for large n. Using the notation and assumptions of [2] we state our results in the following

THEOREM. *Let* $G_1 n^{-\delta} \leqq a_n \leqq G_2 n^{-\delta}$ *for all* n, *where* G_1 *and* G_2 *are positive constants and* $2/3 < \delta \leqq 1$. *If either*

(i) $2/3 < \delta < 1$, *and*

$$K > \frac{1}{2}(G_2/G_1)(C + |\alpha|), \tag{1}$$

or (ii) $\delta = 1$, *and*

$$K > 2(G_2/G_1)(C + |\alpha|), \tag{2}$$

then

$$P(\lim_{n\to\infty} x_n = \theta) = 1. \tag{3}$$

The appropriate estimates for the order of magnitude of b_n are given by (9), (15), (16) and (18) below.

Proof. We shall briefly indicate the proof of (i). Let r be a positive constant less than $[2KG_1 - (C + |\alpha|)G_2]/(1 - \delta)$, and let $A = (C + |\alpha|)G_2/(1 - \delta)$. Then using (21) of [2] and the inequality

$$1 + 2^{-\delta} + \cdots + j^{-\delta} \leqq j^{1-\delta}/(1 - \delta) \qquad \text{(for all } j \text{ large enough)},$$

we have

$$A_j \leqq (A + r)\, j^{1-\delta} \qquad \text{(for } j \text{ sufficiently large)}. \tag{4}$$

From (4) and the easily verifiable relations [2]

$$b_{j+1} = b_j - 2a_j d_j + a_j^2 e_j, \qquad d_j \geqq b_j K/A_j, \quad \text{and} \quad e_j \leqq (C + |\alpha|)^2 < \infty$$

we obtain

$$b_{j+1} \leqq b_j q_j + M_1 j^{-2\delta} \qquad (j \geqq m), \tag{5}$$

where $q_j = 1 - B\cdot j^{-1}$, and $B = 2KG_1/(A + r) > 1 - \delta$ by the choice of r. Here and in the sequel the letter M with or without a suffix denotes a constant independent of n. Putting $j = m, m + 1, \cdots, n$ successively in (5) and setting $Q_n = \prod_{j=m}^{n} q_j$ and $R_n = 1 + m^{-2\delta}Q_m^{-1} + \cdots + n^{-2\delta}Q_n^{-1}$, we have

$$b_{n+1} \leqq M_2 \cdot Q_n \cdot R_n \cdot \tag{6}$$

Also, since $\sum_{j=1}^{n} j^{-1} \sim \log n$,

$$Q_n = O(n^{-B}). \tag{7}$$

For the estimation of R_n we consider three different possibilities:

a) $B > 2\delta - 1$. Easy computation shows that

$$R_n = O(n^{1+B-2\delta}) \tag{8}$$

which together with (6) and (7) leads to the result

(9) $$b_n = O(n^{1-2\delta}).$$

Now choose β to satisfy

(10) $$(2\delta - 1)^{-1} < \beta < (1 - \delta)^{-1},$$

with $2\delta - 1 > 1 - \delta$ since $\delta > 2/3$. For $k = 1, 2, \cdots$ define the subsequence $n_k = [k^\beta]$. Then Tchebycheff's inequality yields the simple estimate

(11) $$P(|x_{n_k} - \theta| > \epsilon) = O[k^{-\beta(2\delta-1)}].$$

Since $\beta(2\delta - 1) > 1$ from (10), using (11) and applying the Borel-Cantelli lemma, we have

(12) $$\lim_{k\to\infty} x_{n_k} = \theta$$

with probability one. For $n_k \leqq n < n_{k+1}$,

(13) $$|x_n - \theta| \leqq |x_n - x_{n_k}| + |x_{n_k} - \theta| \leqq M_3 \sum_{j=n_k}^{n_{k+1}} j^{-\delta} + |x_{n_k} - \theta|.$$

Since $n_{k+1} - n_k = O(k^{\beta-1})$

(14) $$\sum_{j=n_k}^{n_{k+1}} j^{-\delta} = O(k^{\beta-1-\beta\delta}) = o(1)$$

as n and hence k tends to infinity. The last remark follows from the right side of (10). Relations (12), (13) and (14) establish (3).

b) $1 - \delta < B < 2\delta - 1$. In this case, since $2\delta - B > 1$, $R_n = O(1)$ from which we obtain

(15) $$b_n = O(n^{-B}).$$

c) $B = 2\delta - 1$. This gives $R_n = O(\log n)$ and

(16) $$b_n = O(n^{-B} \cdot \log n).$$

Combining b) and c) we may write

(17) $$b_n = O(n^{-\mu})$$

where it is understood that $\mu = B$ in case *b*) and $1 - \delta < \mu < B$ in case c). The rest of the proof is the same as in a).

In the proof of (ii) similar computations give the following estimate for b_n:

(18) $$b_n = O(\log n)^{-\mu}, \qquad \text{where } \mu = KG_1/(C + |\alpha|)G_2 - 1.$$

The proof of (3) is accomplished by taking $n_k = [\exp k^\beta]$ $(\mu^{-1} < \beta < 1)$.

REFERENCES

[1] Julius R. Blum, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 390–394.

[2] Herbert Robbins and Sutton Monro, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.