NOTES

A NOTE ON THE MAXIMUM SAMPLE EXCURSIONS OF STOCHASTIC APPROXIMATION PROCESSES[1]

By Harold J. Kushner

Brown University

1. Introduction and summary. In this note we give a result on the maximum sample excursions of Kiefer-Wolfowitz stochastic approximation processes. The method is applicable to other stochastic approximation procedures, and under other conditions than those assumed here.

Let $y(x)$ be a scalar valued random variable with distribution function $H(y \mid x)$, where x is a scalar valued parameter. Define $M(x) = \int yH(dy \mid x)$. Let $M(x)$ be continuous and have a unique local maximum at $x = \theta$ and let a_n, c_n be sequences of positive real numbers satisfying

$$(1) \qquad \sum a_n = \infty, \qquad \sum a_n^2 c_n^{-2} < \infty, \qquad \sum a_n c_n < \infty, \qquad c_n \to 0.$$

The sequence of random variables x_n defined by

$$(2) \qquad x_{n+1} = x_n + a_n[y(x_n + c_n) - y(x_n - c_n)]/c_n, \qquad x_0 \text{ given},$$

is known as a Kiefer-Wolfowitz process and, under mild conditions on $M(x)$ and $H(y \mid x)$, x_n is known to converge to θ w.p.1. (See, e.g., Schmetterer [6] for a review of such results.)

A result of this note is an estimate of the following form, for any $m < \infty$ and even integer r,

$$(3) \qquad P[\max_{m \geqq n \geqq N} |x_n - \theta| > \epsilon] < [E(x_N - \theta)^r + \delta_{Nr}]/\epsilon^r$$

where δ_{Nr} depends on the sequences a_n and c_n and can be made *arbitrarily small for each fixed* N *and* r, while $x_n \to \theta$ w.p.1. is still insured.

As a special case of (3), let $|x_N - \theta|$ be unknown but assumed nonrandom. Let $\epsilon = \beta + (1 + \alpha)|x_N - \theta|$, $\beta > 0$ and $\alpha > 0$. Then

$$(4) \qquad P[\max_{m \geqq n \geqq N} |x_n - \theta| > (1 + \alpha)|x_N - \theta| + \beta] < [(x_N - \theta)^r + \delta_{Nr}]/[\beta + (1 + \alpha)|x_N - \theta|]^r$$

which can be made arbitrarily small by fixing r sufficiently large, and then arranging a_n and c_n so that δ_{Nr} is sufficiently small.

Aside from the intrinsic interest of (3) and (4), these results seem to have some

Received 8 March 1965; revised 27 August 1965.

[1] This research was supported in part by the National Aeronautics and Space Administration under Grant No. NGR-40-002-015 and in part by the United States Air Force through the Air Force Office of Scientific Research under Grant No. AF-AFOSR-693-64.

practical usefulness in assisting in the choice of the a_n and c_n when there is more than one local maximum of $M(x)$, or if the process (2) is used to optimize the parameters of a physical system whose performance $(M(x))$ should not be reduced below some minimum level—the value of x corresponding to this level may not be known. In both of these cases we may wish to limit the excursions to some given multiple or function of $|x_0 - \theta|$, with a high probability, while still being certain that $x_n \to \theta$ w.p.1.

2. A lemma. We require the following

LEMMA. *Let w_j, $j \geqq 1$, be a sequence of non-negative random variables with $Ew_j < \infty$, and let B_n be the minimum σ-field over which $w_1, \cdots, w_n$ are measurable. Let R_n be a sequence of non-negative random variables, measureable over B_n, such that*, w.p.1.,

$$E^{B_n}w_{n+1} - w_n \leqq R_n, \tag{5}$$

where E^{B_n} is the expectation conditioned upon B_n. Let

$$\textstyle\sum_1^\infty ER_n < \infty. \tag{6}$$

Define

$$V_n = w_n + E^{B_n}\textstyle\sum_n^\infty R_j. \tag{7}$$

Then $\{(V_n, B_n\}, n \geqq 1\}$ is a non-negative super-martingale and, for any $m < \infty$,

$$P[\max_{m \geqq n \geqq N} V_n > \epsilon] < EV_N/\epsilon. \tag{8}$$

REMARK. (3) is obtained from the lemma by letting, for r even, $(x_n - \theta)^r = w_n$, and, under the conditions below, exhibiting a sequence R_n which satisfies (5) and (6). Also $\delta_{Nr} = E\sum_N^\infty R_n$ will have the property below (3).

PROOF. $B_{n+1} \supset B_n$. Since $V_n \geqq 0$ and $R_n \geqq 0$ and $EV_m \leqq EV_n < \infty$ for $m > n \geqq 1$, and

$$E^{B_n}V_{n+1} - V_n$$
$$= E^{B_n}w_{n+1} - w_n + E^{B_n}E^{B_{n+1}}\textstyle\sum_{n+1}^\infty R_j - E^{B_n}\sum_n^\infty R_j \leqq R_n - R_n \leqq 0$$

with probability one, we have that $\{(V_n, B_n), n \geqq 1\}$ is a non-negative super-martingale. (8) is the non-negative super-martingale version of Theorem VII 3.2 of Doob [3]. Q.E.D.

3. Assumptions and terminology. Write $\theta = 0$ for simplicity. Redefine B_n to be the minimum σ-field with respect to which $x_0, \cdots, x_n$ are measurable. Write (2) as

$$\begin{aligned} x_{n+1} &= x_n + a_nM_{c_n}(x_n) + a_n\xi_n/c_n \\ M_c(x) &= [M(x+c) - M(x-c)]/c \\ \xi_n &= [y(x_n+c_n) - M(x_n+c_n)] - [y(x_n-c_n) - M(x_n-c_n)]. \end{aligned}$$

Subsequent assumptions are part of those used by Derman [2], some of whose results we will use. Let K_0, K, C_0 be positive real numbers. Assume

$$(9) \qquad E^{B_n}\xi_n = 0, \qquad E^{B_n}\xi_n^2 \leqq 2\sigma^2 < \infty$$

with probability one. For $0 < c < C_0 < \infty$, let

$$(10) \qquad -cKx^2 \leqq [M(x+c) - M(x-c)]x \leqq -cK_0x^2$$

$$(11) \qquad a_n = A/n^{1-\epsilon}, \qquad c_n = C/n^{\frac{1}{2}-\eta} < C_0 ; \quad \eta > \epsilon > 0, \eta + \epsilon < \tfrac{1}{2}.$$

For each integer $r > 0$, there is a positive real number $M_r < \infty$ such that

$$(12) \qquad E^{B_n}|y(x_n) - M(x_n)|^r \leqq M_r/2.$$

$$(13) \qquad KA \leqq 1, \qquad A > 0.$$

4. Main result.

THEOREM. *Let* $\theta = 0$. *Assume* (9)–(13). *Then, for all integral* m, N *such that* $\infty > m \geqq N \geqq 1$, *and even integer* r,

$$(14) \qquad P[\max_{m \geqq n \geqq N} |x_n| > \epsilon] < (Ex_N^r + \delta_{Nr})/\epsilon^r,$$

where δ_{Nr} *is finite and tends to zero as* $A \to 0$. *Also*, $x_n \to 0$ *with probability one.*

PROOF. By (10), $M_{c_n}(x_n) = -K_nx_n$, where $0 < K_0 \leqq K_n \leqq K < \infty$. Thus $x_{n+1} = (1 - a_nK_n)x_n + a_n\xi_n/c_n$.

$$(15) \qquad \begin{aligned} E^{B_n}x_{n+1}^r - x_n^r &= [(1-a_nK_n)^r - 1]x_n^r \\ &\quad + \textstyle\sum_1^r \binom{r}{i}(a_n/c_n)^i(1-a_nK_n)^{r-i}x_n^{r-i}E^{B_n}\xi_n^i \\ &\leqq \textstyle\sum_2^r \binom{r}{i}(a_n/c_n)^iM_i|x_n^{r-i}|. \end{aligned}$$

The last inequality follows from the use of (9) (to set $E^{B_n}\xi_n = 0$), and successive majorizations using (13) (yielding $0 \leqq a_nK_n \leqq 1$), and (12) (yielding $E^{B_n}(\xi_n^i) \leqq M_i$).

Define R_n as the (non-negative) majorant (the last line of (15)). Define $D_{nr} = E^{B_n} \sum_n^\infty R_j$. Suppose that $ED_{1r} < \infty$. Then, by the lemma, $\{(x_n^r + D_{nr}, B_n), n \geqq 1\}$ is a non-negative super-martingale and, since $D_{nr} \geqq 0$,

$$P[\max_{m \geqq n \geqq N} |x_n| > \epsilon] = P[\max_{m \geqq n \geqq N} x_n^r > \epsilon^r]$$
$$\leqq P[\max_{m \geqq n \geqq N} (x_n^r + D_{nr}) \geqq \epsilon^r].$$

The super-martingale inequality (8) now yields (14), where $\delta_{Nr} = ED_{Nr}$. Under the supposition, it is clear that $\delta_{nr} \to 0$, as $A \to 0$.

Define $b_n^{(r)} = E|x_n|^r$. Under (9) to (13), Derman ([2], Lemma 1) has proved that, for r even (Actually Derman [2] used $A = C = 1$, and the factor $(A/C^2)^{r/2}$ does not appear in [2]. (16) follows immediately by noting that, with arbitrary positive real A and C, $A^2\sigma^2/C^2$ must replace σ^2 and AK_0 must replace K_0 in [2].)

$$(16) \quad \limsup_n n^{r(\eta-\epsilon/2)}b_n^{(r)} \leqq (r-1)(r-3)\cdots 3\cdot 1\cdot(\sigma^2/K_0)^{r/2}(A/C^2)^{r/2}.$$

Since $b_n^{(r)} \leqq [b_n^{(2r)}]^{\frac{1}{2}}$, for each integer $r > 0$, there is a positive real number $Q_r < \infty$ such that

$$\limsup_n n^{r(\eta-\epsilon/2)} b_n^{(r)} \leqq Q_r(A\sigma^2/C^2K_0)^{r/2}. \tag{17}$$

By (17), (15) and (11), we have $ED_{Nr} < \infty$. Also, both EX_N^r and δ_{Nr} tend to zero, as $N \to \infty$. Thus, the right side of (14) tends to zero with N and, hence, $x_n \to 0$ w.p.1. Q.E.D.

5. Remarks. Relations such as (3) and (4) are derivable for other stochastic approximation procedures, and for the Kiefer-Wolfowitz procedure under other conditions. The technique is the same: for an even r, compute $E^{B_n}x_{n+1}^r - x_n^r = R_n$. If $\sum_n^\infty E|R_j| < \infty$, the pair $\{x_n^r + E^{B_n}\sum_n^\infty |R_j|, B_n\}$ is a non-negative super-martingale. If $\sum_n^\infty E|R_j| \to 0$ as A, or some other parameter, goes to zero, analogs of (3) and (4) are available.

If x is a vector, then we look for super-martingales of the form

$$V_n = M^r(x_n) + E^{B_n}\sum_n^\infty |E^{B_j}M^r(x_{j+1}) - M^r(x_j)|.$$

If x_t is a continuous parameter stochastic approximation, which is also a Markov process, then similar relations are possible, provided that an infinitesimal operator of the x_t process can be suitably defined. With the use of Dynkin's formula (Dynkin [4], Theorem 2) an appropriate super-martingale may be defined. (See Kushner [5] for a closely related continuous parameter problem and method).

REFERENCES

[1] Chung, K. L. (1954). On a stochastic approximation method. *Ann. Math. Statist.* **25** 463–483.

[2] Derman, C. (1956). An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure. *Ann. Math. Statist.* **27** 532–536.

[3] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[4] Dynkin, E. B. (1956). Infinitesimal operators of Markov processes. *Theor. Prob. Appl.* **1** 35–54.

[5] Kushner, H. J. (1966). On the theory of stochastic stability. Technical Report 65-1, Center for Dynamical Systems, Brown Univ. To appear in *Advances in Control Theory*, **4**, Academic Press, New York.

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