ASYMPTOTIC DISTRIBUTION OF STOCHASTIC APPROXIMATION PROCEDURES[1]

By Jerome Sacks[2]

California Institute of Technology

1. Introduction. Beginning with the paper of Robbins and Monro [11] much work has been done in stochastic approximation. The Robbins-Monro procedure (see [11] or Section 3 below) for finding the root of a regression equation and the Kiefer-Wolfowitz procedure (see [9] or Section 4 below) for finding the maximum of a regression function have been the chief objects of investigation. The investigations that have been carried out on these procedures have been along two lines: the first being concerned with conditions under which the procedures (i.e., the sequence $\{X_n\}$ of approximating random variables) converge, in some sense, and the second being concerned with the speed of convergence and the asymptotic distribution of the procedures. For details concerning these investigations we refer the reader to the literature; some account of them may be found in Sections 3, 4, and 5 when they relate to the context. In particular the results relating to conditions for convergence are all subsumed in the work of Dvoretzky [7], Wolfowitz [12], and Block [1].

Chung [5] was the first to give any results about the asymptotic distribution of these procedures in his treatment of the Robbins-Monro procedure, and his methods (see the next paragraph) have been the basis for all work done heretofore in this direction. Hodges and Lehmann [8] improved some of Chung's results. Derman [6] used Chung's methods to obtain some results for the Kiefer-Wolfowitz procedure and Burkholder [4] extended Chung's methods to obtain further results on the asymptotic distribution of the Kiefer-Wolfowitz procedure.

Chung's method for obtaining his results on the asymptotic normality of the appropriately normalized sequence $\{X_n\}$ is to compute sufficiently fine estimates for the moments of $X_n - \theta$ (θ is the root of the regression equation) and then to apply the method of moments. As we noted above all previous work on the asymptotic distribution of the two procedures in question has been based on Chung's methods. The main feature of the present work is that we do away with the method of moments by, instead, utilizing a central limit theorem for dependent random variables and obtain more general and more complete results about the asymptotic normality of $\{X_n\}$ for both procedures by using a different method of proof—the method of proof we use may be seen by referring to that portion of the proof of Theorem 1 which lies between (3.8) and (3.9c). In addition, in Examples 1 and 2 in Section 4 we show that some of the results obtained here are best possible in a certain sense.

Received June 3, 1957; revised December 2, 1957.

[1] This research was done in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Cornell University and was sponsored by the Office of Naval Research and the Air Force.

[2] Now at the Department of Mathematical Statistics, Columbia University.

One of the complications arising from use of the method of moments is that the computations needed there are not feasible unless $\{a_n\}$ and $\{c_n\}$ (see (3.0) and (4.0)) are of the special type $a_n = an^{-t}$, $c_n = cn^{-r}$. While we take $a_n = an^{-1}$ in Sections 3, 4, and 5 (see Section 6 for further remarks on this choice of a_n) one of the reasons why we can obtain better results for the Kiefer-Wolfowitz procedure than heretofore obtained is that the method of proof we use permits a wider choice for $\{c_n\}$. Other desirable features of the method of proof presented here are that restrictions are needed only on the second moments of $Z(x)$ (the method of moments requires restrictions on all moments) and that the method can be used without difficulty on some multi-dimensional analogues (see Section 5) of the procedures.

In Section 3 we treat the Robbins-Monro procedure and in Section 4 we discuss the Kiefer-Wolfowitz procedure. Section 5 is devoted to some multi-dimensional analogues of the procedures. Section 6 discusses some further consequences and extensions of the results of earler sections. In Section 2 we collect some lemmas and computations which are used repeatedly in later sections.

The author would like to take this opportunity to acknowledge his debt to Professors J. Kiefer and J. Wolfowitz for their direction and assistance during the course of this research.

2. Preliminaries. In this section we will collect and prove several simple results which are used repeatedly in later sections. In addition, we will state and prove the central limit theorem which we use in succeeding sections. In what follows D_1, D_2, etc., will denote constants appropriately chosen to suit the context in which they appear.

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

$$\sum a_n = \infty, \qquad \sum a_n^2 < \infty \tag{2.0}$$

Except for Lemma 1 it will always be assumed in this section that $a_n = an^{-1}$ for some $a > 0$. Let

$$\begin{aligned} \beta_{mn} &= \prod_{j=m+1}^{n} (1 - a_j) && 0 \leqq m < n \\ &= 1 && m = n \end{aligned} \tag{2.1}$$

It is well known that

$$(1 - \epsilon_m) \exp\left\{-\sum_{j=m+1}^{n} a_j\right\} \leqq \beta_{mn} \leqq (1 + \epsilon_m) \exp\left\{-\sum_{j=m+1}^{n} a_j\right\} \tag{2.2}$$

for all $n \geqq m$, where $\epsilon_m \to 0$ as $m \to \infty$. In particular, if, for $a > 0$, $a_n = an^{-1}$ we have

$$(1 - \epsilon_m')m^a n^{-a} \leqq \beta_{mn} \leqq (1 + \epsilon_m')m^a n^{-a} \tag{2.3}$$

where $\epsilon_m' \to 0$ as $m \to \infty$.

LEMMA 1. *Let $\{W_m\}$ be a sequence of real numbers converging to W where W may* *be taken to be* ∞. *Then, for any positive integer* m_0,

$$\lim_{n\to\infty} \sum_{m=m_0}^{n} a_m \beta_{mn} W_m = W$$

PROOF. For any fixed m it follows from (2.2) that $\lim_{n\to\infty} \beta_{mn} = 0$. Since $a_m\beta_{mn} = \beta_{mn} - \beta_{m-1,n}$ we have, for any fixed m_1,

$$\lim_{n\to\infty} \sum_{m=m_1}^{n} a_m \beta_{mn} = \lim_{n\to\infty} (1 - \beta_{m_1-1,n}) = 1.$$

The conclusion of Lemma 1 now follows quite easily.

Let $\{c_m\}$ be a sequence of positive real numbers. For each n let

$$h_n = \left(\sum_{m=1}^{n} a^2 c_m^{-2} m^{-2} \beta_{mn}^2\right)^{-1/2}.$$

LEMMA 2. *Let* $a > 1/2$ *and suppose that* $c_m \leqq c < \infty$ *for all* m. *Then, if* m_0 *is some fixed positive integer*

$$\lim_{n\to\infty} h_n \beta_{m_0 n} = 0.$$

PROOF. If $m_0 > a - 1$ let $m_1 = m_0$; otherwise, let m_1 be the smallest integer greater than $a - 1$. To prove the lemma it is obviously sufficient to prove that $h_n\beta_{m_1 n} \to 0$ as $n \to \infty$. Using (2.2) we obtain

$$\begin{aligned} h_n^2 \beta_{m_1 n}^2 &\leqq D_1 h_n^2 n^{-2a} \leqq D_2 n^{-2a} \left(\sum_{m=m_1}^{n} am^{-1}\beta_{mn} c_m^{-2} m^{a-1} n^{-a}\right)^{-1} \\ &= D_2 \left(\sum_{m=m_1}^{n} am^{-1}\beta_{mn} c_m^{-2} m^{a-1} n^{a}\right)^{-1} \end{aligned} \tag{2.4}$$

If $a > 1/2$ then $n^a m^{a-1} c_m^{-2} > c^{-2} m^{2a-1}$ which goes to ∞ as $m \to \infty$ and hence, by Lemma 1, the last term in (2.4) goes to 0 as $n \to \infty$.

LEMMA 3. *Let* $a > 1/2$ *and suppose that* $c_m \leqq c < \infty$ *for all* m. *Let* $\{W_m\}$ *be a sequence of real numbers converging to* W *where* W *may be* ∞. *Then, if* m_0 *is a fixed positive integer*

$$\lim_{n\to\infty} h_n^2 \sum_{m=m_0}^{n} a^2 c_m^{-2} m^{-2} \beta_{mn}^2 W_m = W.$$

PROOF. The proof is easily accomplished upon noting that, for any fixed m_1

$$\lim_{n\to\infty} h_n^2 \sum_{m=m_1}^{n} a^2 m^{-2} c_m^{-2} \beta_{mn}^2 = 1.$$

LEMMA 4. *Let* $\{d_m\}$ *be a sequence of positive numbers such that*

$$d_m d_{m+1}^{-1} = 1 + \epsilon_m m^{-1} \qquad \text{where } \epsilon_m \to 0 \text{ as } m \to \infty. \tag{2.5}$$

Let $q > -1$. *Then, for any positive integer* m_0,

$$\sum_{m=m_0}^{n} d_m m^q \sim (1 + q)^{-1} d_n n^{q+1}$$

Examples of sequences satisfying (2.5) are easy to obtain. For example, $d_m = (\log m)^u$ satisfies (2.5) for all real u. Note also that if $\{d_m\}$ satisfies (2.5) then $\{d_m^\rho\}$ also satisfies (2.5) for any real number ρ.

PROOF. Since $\sum_{m=1}^n m^q \sim (1+q)^{-1}n^{q+1}$ what we have to show will be accomplished if we show that

$$\frac{\sum_{m=1}^n d_m m^q}{d_n \sum_{m=1}^n m^q} - 1 = \frac{\sum_{m=1}^{n-1} (d_m - d_{m+1}) \sum_{j=1}^m j^q + d_n \sum_{j=1}^n j^q}{d_n \sum_{m=1}^n m^q} - 1$$

$$= \frac{\sum_{m=1}^{n-1} d_m d_n^{-1}(d_m d_{m+1}^{-1} - 1) \sum_{j=1}^m j^q}{\sum_1^n m^q} = A_n \quad \text{(say)}$$

goes to 0 as $n \to \infty$. Using (2.5) we see that for n sufficiently large

$$d_n = d_1 \prod_{j=1}^{n-1} d_{j+1} d_j^{-1} \geqq D_3 e^{\sum_{j=1}^{n-1} \epsilon_j j^{-1}} \geqq D_4 e^{-\epsilon \sum_{j=m_1}^{n-1} j^{-1}} \geqq D_5 m_1^\epsilon n^{-\epsilon}$$

where m_1 is chosen so that $|\epsilon_j| \leqq \epsilon < q + 1$ for all $j \geqq m_1$. Thus $d_n n^{q+1} \to \infty$ as $n \to \infty$ and, therefore, in order to show that $A_n \to 0$, we can start the outer sum in the numerator of A_n at $m = m_1$.

By use of (2.5) we have, for $n > m \geqq m_1$,

$$d_m d_n^{-1} = \prod_{j=m}^{n-1} d_j d_{j+1}^{-1} \leqq \prod_{j=m}^{n-1} (1 + \epsilon j^{-1}) \leqq D_6 n^\epsilon m^{-\epsilon}$$

and

$$|d_m d_{m+1}^{-1} - 1| \leqq \epsilon m^{-1}$$

That A_n must go to 0 now follows because for all n

$$\frac{\sum_{m=m_1}^n D_6 n^\epsilon m^{-\epsilon} \epsilon m^{-1} \sum_{j=1}^m j^q}{\sum_{m=1}^n m_q} \leqq \epsilon D_7 n^{\epsilon-q-1} \sum_{m_1}^n m^{-\epsilon-1} m^{q+1} \leqq \epsilon D_8$$

and because ϵ is arbitrary.

LEMMA 5. *Let $c_m = d_m m^{-r}$ where $r \geqq 0$ and where $\{d_m\}$ satisfies* (2.5). *Let a be a real number greater than 1/2. Then, for any positive integer m_0, and any positive number ρ,*

$$\sum_{m=m_0}^n a^2 c_m^{-\rho} m^{-2} \beta_{mn}^2 \sim a^2(2a + r\rho - 1)^{-1} c_n^{-\rho} n^{-1} \tag{2.6}$$

as $n \to \infty$. In particular, if $m_0 = 1$ and $\rho = 2$ (2.6) becomes

$$h_n^2 \sim a^{-2}(2a + 2r - 1)nc_n^2 \tag{2.7}$$

PROOF. Let $\epsilon > 0$ and let m_1 be large enough so that in (2.3) $\epsilon_m' < \epsilon$ for $m > m_1$. Then, using (2.3) and Lemma 4—the conditions of Lemma 4 are satisfied if one takes into account the conditions stated here and the remarks following the statement of Lemma 4—we obtain

$$
\begin{aligned}
a^2 \sum_{m=m_0}^{n} c_m^{-\rho} m^{-2} \beta_{mn}^2 &\leqq (1+\epsilon)a^2 \sum_{m_1}^{n} c_m^{-\rho} m^{2a-2} n^{-2a} + O(n^{-2a}) \\
&= (1+\epsilon)a^2 \sum_{m_1}^{n} d_m^{-\rho} m^{2a+\rho r-2} n^{-2a} + O(n^{-2a}) \qquad (2.8)\\
&\leqq (1+\alpha_n)a^2(2a+\rho r-1)^{-1} d_n^{-\rho} n^{2a+\rho r-1} n^{-2a} + O(n^{-2a}) \\
&= (1+\alpha_n)a^2(2a+\rho r-1)^{-1} c_n^{-\rho} n^{-1} + O(n^{-2a})
\end{aligned}
$$

where $\alpha_n \to \epsilon$ as $n \to \infty$. Similar calculation produces

$$
a^2 \sum_{m_0}^{n} c_m^{-\rho} m^{-2} \beta_{mn}^2 \geqq (1+\alpha_n)a^2(2a+\rho r-1)^{-1} c_n^{-\rho} n^{-1}
$$

Since $n^{1-2a} c_n^{\rho} \to 0$ as $n \to \infty$ and since ϵ is arbitrary we have achieved the desired result.

We shall now state and prove a central limit theorem which we use later in an essential way. The multi-dimensional version we give (see Lemma 6) is a direct generalization of the one-dimensional result which may be found in Loeve [10], p. 377 C. The proof we give is likewise a direct generalization of the proof given in [10]. In Sections 3 and 4 it will suffice to consider only the one-dimensional case; we make use of the result for higher dimensions in Section 5.

With all vectors considered as elements of q-dimensional Euclidean space we adopt the following notation. If x, y are vectors $[x, y]$ will denote their inner product. The norm of a vector x we denote by $|x|$ and, of course, is equal to $[x, x]^{1/2}$. If B is a $q \times q$ matrix we define in the usual way,

$$
\| B \| = \sup_{|x|=1} [Bx, Bx]^{1/2}
$$

The obvious facts that $|Bx| \leqq \|B\| \, |x|$ and that $\|B_1 B_2\| \leqq \|B_1\| \, \|B_2\|$ will be useful below. I will denote the identity $q \times q$ matrix. B' and x' will denote the transposes of the matrix B and vector x respectively. Unless otherwise indicated a vector is to be considered a column vector.

Let $\{U_{nk} ; 1 \leqq k \leqq n, n \geqq 1\}$ be a family of vector random variables, the distribution of U_{nk} being denoted by F_{nk}. Let $V_{nk} = (U_{n1}, \cdots, U_{n,k-1})$ and suppose that $E(U_{nk} \mid V_{nk}) = 0$ with probability one. Denote the covariance matrix of U_{nk} by s_{nk} i.e., $s_{nk} = E(U_{nk}U'_{nk})$. Let $r_{nk} = E(U_{nk}U'_{nk} \mid V_{nk})$. Let $U_n = \sum U_{nk}$, $s_n = \sum s_{nk}$, and $r_n = \sum r_{nk}$ where all three summations are over $1 \leqq k \leqq n$. For $\epsilon > 0$ define $\phi_{nk}^{\epsilon} = 1$ if $|U_{nk}| > \epsilon$, $\phi_{nk}^{\epsilon} = 0$ otherwise.

LEMMA 6. *If*

$$
\lim_{n\to\infty} \sum_{k=1}^{n} E(\| r_{nk} - s_{nk} \|) = 0 \tag{2.9}
$$

$$
\sup_{n} \sum_{k=1}^{n} E(| U_{nk} |^2) < \infty, \tag{2.10}
$$

and, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{n} E(|U_{nk}|^2 \phi_{nk}^{\epsilon}) = 0, \tag{2.11}$$

and if $s_n \to s$, i.e., $\|s_n - s\| \to 0$, *then* U_n *is asymptotically normal with mean* 0 *and covariance matrix* s.

PROOF. Let F and G be q-dimensional distribution functions with characteristic functions f and g and finite covariance matrices C and D respectively, and let $H = F - G$. Let θ_1, θ_2 denote quantities whose absolute value is less than 1. Let $\Delta = \{x \mid |x| \leqq \epsilon\}$ and let Δ' be the complement of Δ. Then, for fixed t and $\epsilon < 1/|t|$,

$$\begin{aligned} |f(t) - g(t)| &\leqq \left| \int [t, x]\, dH(x) \right| + \frac{1}{2} \left| \int_\Delta [t, x]^2\, dH(x) \right| \\ &\qquad + \left| \int_\Delta \theta_1 [t, x]^3\, dH(x) \right| + \left| \int_{\Delta'} \theta_2 [t, x]^2\, dH(x) \right| \\ &\leqq \left| \int [t, x]\, dH(x) \right| + \left| \int [t, x]^2\, dH(x) \right| \\ &\qquad + \epsilon |t| \int [t, x]^2\, d(F + G) + 3 \int_{\Delta'} [t, x]^2\, d(F + G) \\ &\leqq \left| \int [t, x]\, dH(x) \right| + |t|^2 \| C - D \| \\ &\qquad + \epsilon |t|^3 \int |x|^2\, d(F + G) + 3 |t|^2 \int_{\Delta'} |x|^2\, d(F + G) \end{aligned} \tag{2.12}$$

Let G_{nk} denote the normal distribution with mean 0 and covariance matrix s_{nk}. Let $\{Y_{nk}; 1 \leqq k \leqq n, n \geqq 1\}$ be a family of independent random variables with the distribution of Y_{nk} being G_{nk}. In addition, take $\{Y_{nk}\}$ to be independent of $\{U_{nk}\}$. It is easy to see that $Y_n = Y_{n1} + \cdots + Y_{nn}$ is asymptotically normal with mean 0 and covariance matrix s.

Let f_{nk}, f_n, g_{nk}, and g_n denote the characteristic functions of U_{nk}, U_n, Y_{nk}, and Y_n respectively. Let

$$f_{nk}^*(t) = E(e^{i[t, U_{nk}]} \mid V_{nk}).$$

To prove the lemma it is clearly sufficient to prove that, for each fixed t,

$$\lim_{n\to\infty} |f_n(t) - g_n(t)| = 0.$$

Let $W_{nk} = U_{n1} + \cdots + U_{n,k-1} + Y_{n,k+1} + \cdots + Y_{nn}$ for $1 < k < n$, $W_{n1} = Y_{n2} + \cdots + Y_{nn}$, $W_{nn} = U_{n1} + \cdots + U_{n,n-1}$. Then

$$\begin{aligned} |f_n(t) - g_n(t)| &= |E(e^{i[t, U_n]} - e^{i[t, Y_n]})| \\ &= \left| E \sum_{k=1}^{n} (e^{i[t, U_{nk}]} - e^{i[t, Y_{nk}]}) e^{i[t, W_{nk}]} \right| \end{aligned} \tag{2.13}$$

$$\leqq \sum_{k=1}^{n} E \mid f_{nk}^*(t) - g_{nk}(t) \mid$$

From (2.12), (2.13), and the fact that $E(U_{nk} \mid V_{nk}) = 0$ we obtain

$$(2.14) \quad \begin{aligned} \mid f_n(t) - g_n(t) \mid \leqq \mid t \mid^2 \sum_{k=1}^{n} E \parallel r_{nk} - s_{nk} \parallel + 2\epsilon \mid t \mid^3 \sum_{k=1}^{n} E \mid U_{nk} \mid^2 \\ + 3 \mid t \mid^2 \sum_{k=1}^{n} E(\phi_{nk}^{\epsilon} \mid U_{nk} \mid^2) + 3 \mid t \mid^2 \sum_{k=1}^{n} \int_{\Delta'} \mid x^2 \, dG_{nk} \end{aligned}$$

As $n \to \infty$ the first and third terms on the right-hand side of (2.14) go to 0 because of (2.9) and (2.11), the second term is $O(\epsilon)$ because of (2.10), and the last term goes to 0 because G_{nk} is normal with covariance matrix s_{nk} and $\|s_{nk}\|$ goes to 0 as $n \to \infty$ uniformly in $k \leqq n$. Since ϵ is arbitrary this finishes the proof of the lemma.

3. The Robbins-Monro Procedure. Let M be a fixed function such that the equation $M(x) = \alpha$ has a unique solution $x = \theta$. For each x let $Y(x)$ be a random variable with $EY(x) = M(x)$. The Robbins-Monro procedure for "finding" θ is defined as follows. Let $\{a_n, n > 0\}$ be a sequence of positive numbers such that

$$(3.0) \qquad \sum a_n = \infty, \qquad \sum a_n^2 < \infty.$$

Let X_1 be some fixed number (X_1 may be taken to be an arbitrary random variable for what follows since, if $EX_1^2 < \infty$, the same proofs will hold, while, if $EX_1^2 = \infty$, the results are obtained by truncating X_1 and using the results for the case $EX_1^2 < \infty$) and define $\{X_n, n > 1\}$ by the recursion

$$(3.1) \qquad X_{n+1} = X_n - a_n(Y(X_n) - \alpha)$$

where $Y(X_n)$ is a random variable whose conditional distribution given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Y(x_n)$. Letting $Z(x) = Y(x) - M(x)$ (3.1) becomes

$$(3.2) \qquad X_{n+1} = X_n - a_n[M(X_n) - \alpha + Z(X_n)],$$

$EZ(x) = 0$ for all x, and the conditional distribution of $Z(X_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Z(x_n)$. We note for future use that, as a consequence of this,

$$(3.3) \qquad E(Z(X_n) \mid Z(X_1), \cdots, Z(X_{n-1})) = 0$$

with probability one.

For Theorem 1 we make the following assumptions about $M(x)$ and $Z(x)$. The connection between these assumptions and those made by previous authors is pointed out below.

ASSUMPTION (A1). M is a Borel-measurable function; $M(\theta) = \alpha$ and

$$(x - \theta)(M(x) - \alpha) > 0$$

for all $x \neq \theta$.

Assumption (A2). For some positive constants K and K_1, and for all x

$$K \mid x - \theta \mid \leqq |M(x) - \alpha| \leqq K_1 \mid x - \theta \mid$$

Assumption (A3). For all x

$$M(x) = \alpha + \alpha_1(x - \theta) + \delta(x, \theta)$$

where $\delta(x, \theta) = o(|x - \theta|)$ as $x - \theta \to 0$ and where $\alpha_1 > 0$.

Assumption (A4).

$$\text{(a)} \sup_x EZ^2(x) < \infty; \qquad \text{(b)} \lim_{x \to \theta} EZ^2(x) = \sigma^2$$

Assumption (A5).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x-\theta|<\epsilon} \int_{\{|Z(x)|>R\}} Z^2(x)\, dP = 0$$

When $X_n \to \theta$ with probability one (for example, under (A1), (A2′), and (a) of (A4) as shown by Blum [2], and when (a) of (A4) holds, (A5) implies

$$\lim_{R \to \infty} \sup_k \int_{\{|Z(X_k)|>R\}} Z^2(X_k)\, dP = 0 \tag{3.4}$$

which is actually what is used in the proof below. The reason we state (A5) in the way we have is that it appears as a more natural condition than (3.4). Simple conditions which imply (A5) are given by

$$\{Z(x)\} \text{ are identically distributed} \tag{3.5}$$

or

$$\sup_{|x-\theta|<\epsilon} E \mid Z(x) \mid^{2+v} < \infty \tag{3.6}$$

for some $\epsilon > 0$ and some $v > 0$.

Assumption (A2) can be weakened to

Assumption (A2′). For all x and some positive constant K_1

$$|M(x) - \alpha| \leqq K_1 \mid x - \theta|$$

and, for every t_1, t_2 such that $0 < t_1 < t_2 < \infty$,

$$\inf_{t_1 \leqq |x-\theta| \leqq t_2} \mid M(x) - \alpha \mid > 0.$$

In Theorem 1′ we obtain the result of Theorem 1 with (A2′) replacing (A2)—the truncation device used in the proof there is due to Hodges and Lehmann [8].

Under (A1), (A2), (A3), the assumption that $EZ^2(x) = \sigma^2$ for all x, and the assumption that (3.6) hold for $\epsilon = \infty$ and all v Chung obtained the result of Theorem 1 (this is what is referred to in [5] as the "second case"). Hodges and Lehmann proved the result of Theorem 1 under (A1), (A2′), (A3), (A4), and the assumption that (3.6) hold for some $\epsilon > 0$ and all v. Thus Theorem 1′ includes these earlier results by virtue of the greater generality of (A5) over related conditions made by previous authors.

As before D_1, D_2, etc., will denote positive constants appropriately chosen for the context in which they appear.

THEOREM 1. *Suppose that Assumptions* (A1) *through* (A5) *are satisfied. Let* $a_n = An^{-1}$ *for* $n > 0$ *where* A *is such that* $2KA > 1$. *Then* $n^{1/2}(X_n - \theta)$ *is asymptotically normally distributed with mean* 0 *and variance* $A^2\sigma^2(2A\alpha_1 - 1)^{-1}$.

PROOF. There is no loss in assuming that $\alpha = \theta = 0$. Abbreviating $\delta(X_n, 0)$, $M(X_n)$ and $Z(X_n)$ by δ_n, M_n, and Z_n respectively, and using (A3) we rewrite (3.2) and obtain

$$X_{n+1} = (1 - A\alpha_1 n^{-1})X_n - An^{-1}\delta_n - An^{-1}Z_n \tag{3.7}$$

Let $a = A\alpha_1$ and let β_{mn} be as in (2.1) with the a_j in (2.1) replaced by aj^{-1}. Iteration of (3.7) then yields

$$X_{n+1} = \beta_{on}X_1 - A\sum_{m=1}^{n} m^{-1}\beta_{mn}\delta_m - A\sum_{m=1}^{n} m^{-1}\beta_{mn}Z_m \tag{3.8}$$

Let $h_n = (\sum_{m=1}^{n} a^2m^{-2}\beta_{mn}^2)^{-1/2}$. Then, by Lemma 5,

$$h_n \sim (2a - 1)^{1/2}a^{-1}n^{1/2}$$

Hence, proving that $n^{1/2}X_n$ is asymptotically normal with mean 0 and variance $A^2\sigma^2(2a - 1)^{-1}$ is equivalent to proving

(3.9) h_nX_n is asymptotically normal with mean 0 and variance $A^2\sigma^2a^{-2}$.

Using (3.8) it is clear that we can show (3.9) by proving

$$h_n\beta_{on} \to 0 \tag{3.9a}$$

$$h_n \sum_{m=1}^{n} am^{-1}\beta_{mn}\delta_m \to 0 \text{ in probability} \tag{3.9b}$$

(3.9c) $h_n \sum_{m=1}^{n} am^{-1}\beta_{mn}Z_m$ is asymptotically normal with mean 0 and variance σ^2.

(3.9a) follows immediately from Lemma 2 with $c_m = 1$ for all m. To prove (3.9c) we will invoke Lemma 6 with $q = 1$ and $U_{nk} = h_nak^{-1}\beta_{kn}Z_k$. To see that we can do so observe first that by (3.3)

$$E(U_{nk} \mid U_{n1}, \cdots, U_{n,k-1}) = E(U_{nk} \mid Z_1, \cdots, Z_{k-1}) = 0.$$

Let $\phi_{nk} = 1$ if $|U_{nk}| \geqq \epsilon$ and $\phi_{nk} = 0$ otherwise, and observe that in order to verify (2.11) we have to check that $\sum_{k=1}^{n} E(\phi_{nk}U_{nk}^2) \to 0$ or, what is the same,

$$h_n^2 \sum_{k=1}^{n} a^2k^{-2}\beta_{kn}^2 E(\phi_{nk}Z_k^2) \to 0 \tag{3.10}$$

Noticing, by Lemma 5 and (2.3) that $\phi_{nk} = 1$ implies, for some $\epsilon' > 0$, that $|Z_k| \geqq \epsilon' n^{a-1/2}k^{1-a} \geqq \epsilon' k^{1/2}$, we apply (3.4) which is obtained from (A5) and obtain

$$\lim_{k\to\infty} E(\phi_k' Z_k^2) = 0 \tag{3.11}$$

where $\phi_k' = 1$ if $|Z_k| \geqq \epsilon' k^{1/2}$ and $\phi_k' = 0$ otherwise. Since $\phi_k' \geqq \phi_{nk}$, applying Lemma 3 with $c_m = 1$ for all m and using (3.11) shows that (3.10) is valid. Verifying (2.9) is equivalent to showing

$$\lim_{n\to\infty} h_n^2 \sum_{k=1}^{n} a^2 k^{-2} \beta_{kn}^2 E \mid E'[Z^2(X_k)] - EZ^2(X_k) \mid = 0 \tag{3.12}$$

where E' denotes conditional expected value with the conditioning being by V_{nk}. Use again of Lemma 3 shows that it is sufficient to prove

$$\lim_{k\to\infty} E \mid E'[Z^2(X_k)] - EZ^2(X_k) \mid = 0 \tag{3.13}$$

But (3.13) follows easily by observing that the expression between the absolute value signs is uniformly bounded ((a) of (A4)) so that Lebesgue's theorem is applicable, and by observing that (b) of (A4) together with the convergence of X_k to θ w.p.1 imply

$$\lim_{k\to\infty} E'[Z^2(X_k)] = \lim_{k\to\infty} EZ^2(X_k) = \sigma^2. \tag{3.14}$$

(3.14) and Lemma 3 also serve to show that (2.10) is satisfied with

$$\lim_{n\to\infty} s_n = \sigma^2 \tag{3.15}$$

This completes the verification that Lemma 6 is applicable and therefore establishes (3.9c).

To prove (3.9b) we require the estimate that $EX_n^2 = O(n^{-1})$. This estimate is obtained by Chung [5] but we obtain it here for completeness. The methods are essentially the same.

Squaring both sides of (3.2), taking expected values, and using (A4) we get

$$EX_{n+1}^2 = E(X_n - An^{-1}M_n)^2 + O(n^{-2}) \tag{3.16}$$

Then, by (A1) and (A2), for ϵ sufficiently small so that $2KA - \epsilon > 1$, and for n sufficiently large, say $n > N_1$,

$$\begin{aligned} EX_{n+1}^2 &\leqq (1 - 2KAn^{-1} + A^2K_1^2n^{-2})EX_n^2 + O(n^{-2}) \\ &\leqq (1 - (2KA - \epsilon)n^{-1})EX_n^2 + D_1n^{-2}. \end{aligned} \tag{3.17}$$

Let $p = 2KA - \epsilon$ and let β_{mn}' be defined by (2.1) with $a_j = pj^{-1}$. Choose N_1 large enough so that $p < N_1$ (this is to guarantee that $\beta_{mn}' > 0$ for $m \geqq N_1$ so that (3.18) can hold). Iteration of (3.17) yields

$$\begin{aligned} EX_{n+1}^2 &\leqq D_1 \sum_{m=N_1+1}^{n} m^{-2}\beta_{mn}' + \beta_{N_1n}' EX_{N_1+1}^2 \leqq D_2 n^{-1} + D_3 n^{-p} \\ &= O(n^{-1}) \end{aligned} \tag{3.18}$$

which is the estimate we require.

Let $t > 0$. Since $\delta(x) = o(|x|)$, for $t > 0$ we can find $\epsilon > 0$ with the property that

$$|\delta(x)| \leqq t^2|x| \qquad \text{for } |x| \leqq \epsilon. \tag{3.19}$$

As was pointed out above $X_n \to 0$ w.p.1; hence, we can choose N_2 so that

$$P\{|X_j| \leqq \epsilon, j \geqq N_2\} > 1 - t. \tag{3.20}$$

Let N_3 be larger than N_1 and N_2 and such that $a < N_3 + 1$. Then, denoting $h_n \sum_{m=N_3}^{n} am^{-1}\beta_{mn}\delta_m$ by V_n and $h_n \sum_{m=N_3}^{n} am^{-1}\beta_{mn} |X_m|$ by V_n^*, and using (3.20), (3.19), a Chebyshev-type inequality, (3.18), and Lyapounov's inequality, and (2.3) we have for $n > N_3$,

$$\begin{aligned} P\{|V_n| > t\} &\leqq t + P\{|V_n| > t; |X_j| \leqq \epsilon, j \geqq N_3\} \\ &\leqq t + P\{t^2 V_n^* > t\} \leqq t + tEV_n^* \\ &\leqq t + D_4 t h_n \sum_{N_3}^{n} m^{-1}\beta_{mn} m^{-1/2} \leqq D_5 t. \end{aligned} \tag{3.21}$$

(3.21) together with the fact that $h_n\beta_{mn} \to 0$ for any fixed m (Lemma 2) establishes (3.9b) and finishes the proof of the theorem.

THEOREM 1′. *Suppose that Assumptions* (A1), (A2′), (A3), (A4), *and* (A5) *are satisfied. Let* $a_n = An^{-1}$ *where* A *is such that* $A\alpha_1 > 1/2$. *Then* $n^{1/2}(X_n - \theta)$ *is asymptotically normally distributed with mean* 0 *and variance* $A^2\sigma^2(2A\alpha_1 - 1)^{-1}$.

PROOF. We assume with no loss of generality that $\alpha = \theta = 0$. Let $t > 0$ be such that $A(\alpha_1 - t) > 1/2$. Let $K = \alpha_1 - t$. Then we can find an $\epsilon > 0$ such that for $|x| \leqq \epsilon$

$$K|x| \leqq |M(x)| \leqq K_1|x|. \tag{3.22}$$

Define $M'(x) = M(x)$ if $|x| \leqq \epsilon$, $M'(x) = Kx$ if $|x| > \epsilon$.

Since under (A1), (A2′), and (A4), $X_n \to 0$ w.p.1 we can find N so that for $u > 0$,

$$P\{|X_j| \leqq \epsilon, j \geqq N\} > 1 - u. \tag{3.23}$$

Let $X_1' = X_{N+1}$ and define $\{X_n', n \geqq 1\}$ by the recursion

$$X_{n+1}' = X_n' - a_{n+N}M'(X_n') - a_{n+N}Z(X_n') \tag{3.24}$$

It is clear that the assumptions of Theorem 1′ together with (3.22) show that Theorem 1 is applicable to X_n', M', $\{a_{n+N}\}$. Hence, for all y,

$$\lim_{n\to\infty} P\{(N + n)^{1/2}X_{n+1}' < y\} = F(y) \tag{3.25}$$

where F is the normal distribution function with mean 0 and variance $A^2\sigma^2(2A\alpha_1 - 1)^{-1}$. Using (3.23) and (3.25) we obtain

$$\begin{aligned} \overline{\lim_{n\to\infty}} P\{n^{1/2}X_n < y\} &= \overline{\lim_{n\to\infty}} P\{(n + N)^{1/2}X_{n+N} < y\} \\ &\leqq \lim_{n\to\infty} P\{(n + N)^{1/2}(X_{n+N} - X_n') \end{aligned}$$

$$(3.26) \quad \begin{aligned} &+ (n + N)^{1/2}X'_n < y; |X_j| \leqq \epsilon, j \geqq N\} + u \\ &= \lim_{n\to\infty} P\{(n + N)^{1/2}X'_n < y\} + u \\ &= F(y) + u \end{aligned}$$

Similarly, we obtain

$$(3.27) \qquad \underline{\lim}_{n\to\infty} P\{n^{1/2}X_n < y\} \geqq F(y) - u$$

Since u and y are arbitrary putting (3.26) and (3.27) together finishes the proof of the theorem.

4. The Kiefer-Wolfowitz Procedure. Let M be a fixed function with a unique maximum at $x = \theta$ (by making the obvious alterations in what follows we can replace "maximum" by "minimum"). For each x let $Y(x)$ be a random variable with $EY(x) = M(x)$. The Kiefer-Wolfowitz procedure for locating the maximum is defined as follows. Let $\{a_n\}$, $\{c_n\}$ be two sequences of positive numbers such that

$$(4.0) \qquad \sum a_n = \infty, \qquad c_n \to 0, \qquad \sum a_n^2 c_n^{-2} < \infty$$

Let X_1 be a fixed number (by the same reasoning as in Sections 3 X_1 can be taken to be an arbitrary random variable for what follows) and define $\{X_n, n \geqq 2\}$ by the recursion

$$(4.1) \qquad X_{n+1} = X_n - a_n c_n^{-1}[Y(X_n - c_n) - Y(X_n + c_n)]$$

where $Y(X_n \pm c_n)$ is a random variable whose conditional distribution given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Y(x_n \pm c_n)$. It is usually assumed that $Y(X_n - c_n)$ and $Y(X_n + c_n)$ are conditionally independent i.e., for all Borel sets A and B $P\{Y(X_n + c_n) \varepsilon A, Y(X_n - c_n) \varepsilon B \mid X_n\} = P\{Y(X_n + c_n) \varepsilon A \mid X_n\} P\{Y(X_n - c_n) \varepsilon B \mid X_n\}$. Though this is commonly the case in practice we do not make this assumption since it is unnecessary to do so. Whatever assumptions we do need to make about the joint distribution of $Y(X_n - c_n)$ and $Y(X_n + c_n)$ are contained in (B5). Letting $Z(x) = Y(x) - M(x)$ and writing M_n for $M(X_n - c_n) - M(X_n + c_n)$ and Z_n for $Z(X_n - c_n) - Z(X_n + c_n)$, (4.1) becomes

$$(4.2) \qquad X_{n+1} = X_n - a_n c_n^{-1}(M_n + Z_n),$$

$EZ(x) = 0$ for all x, and the conditional distribution of Z_n given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Z(x_n - c_n) - Z(x_n + c_n)$. We note that, as a consequence of this,

$$(4.3) \qquad E(Z_n \mid Z_1, \cdots, Z_{n-1}) = E(Z_n \mid X_1, \cdots, X_n) = 0$$

with probability one.

We now make the assumptions we require for Theorem 2. Other assumptions relevant to later theorems are listed further on. The connection between these assumptions and those made by previous authors is pointed out below. As before D_1, D_2, etc. will denote appropriately chosen positive constants.

ASSUMPTION (B1). $M(x)$ is a Borel-measurable function, has a unique maximum at $x = \theta$, and, for $0 < t_0 < t_1 < t_2 < \infty$,

$$\inf_{\substack{t_1 \leqq |x-\theta| \leqq t_2 \\ 0 < \epsilon \leqq t_0}} \frac{(x - \theta)(M(x - \epsilon) - M(x + \epsilon))}{\epsilon} > 0 \tag{4.4}$$

In addition, for all x and suitable D_1 and D_2,

$$|M(x + 1) - M(x)| < D_1 + D_2|x| \tag{4.5}$$

ASSUMPTION (B2). For all x

$$M(x) = \alpha_0 - \alpha(x - \theta)^2 + \delta(x, \theta)$$

where α_0 is some real number, $\alpha > 0$, and $\delta(x, \theta) = o(|x - \theta|^2)$ as $x - \theta \to 0$.

ASSUMPTION (B3). For some $c_0 > 0$ there exist positive constants K_1 and K_2 such that, for all x and all c for which $0 < c \leqq c_0$,

$$K_1(x - \theta)^2 \leqq (x - \theta)[M(x - c) - M(x + c)]c^{-1} \leqq K_2(x - \theta)^2$$

ASSUMPTION (B4). For every $\epsilon > 0$ there exists $c_\epsilon > 0$ such that, for all c satisfying $0 < c \leqq c_\epsilon$ and all x satisfying $|x - \theta| < c$,

$$|\delta(x - c, \theta) - \delta(x + c, \theta)|c^{-1} \leqq \epsilon|x - \theta|$$

ASSUMPTION (B5).

$$\sup_x EZ^2(x) = s < \infty \tag{4.6}$$

$$\lim_{\substack{x \to \theta \\ a \to 0}} E[Z(x - a) - Z(x + a)]^2 = \sigma^2. \tag{4.7}$$

In case $Z(X_m - c_m)$ and $Z(X_m + c_m)$ are uncorrelated we can replace (4.7) by

$$\lim_{x \to \theta} EZ^2(x) = \sigma^2/2 \tag{4.8}$$

ASSUMPTION (B6).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x| < \epsilon} \int_{\{|Z(x)| > R\}} Z^2(x)\, dP = 0$$

When $X_n \to \theta$ w.p.1 (for example, under (B1) and (4.6) as shown by Burkholder [4] and Dvoretzky [7]—Blum [2] proved convergence w.p.1 earlier but under stronger restrictions) (B6) implies

$$\lim_{R \to \infty} \sup_k \int_{\{|Z_k| > R\}} Z_k^2\, dP = 0 \tag{4.9}$$

The remarks made about (3.4) and (A5) pertain here to (4.9) and (B6) and, as with (A5), (B6) is satisfied if either (3.5) or (3.6) is fulfilled.

In Theorem 2′ we obtain the same result as in Theorem 2 with (B3) replaced by the weaker restriction

ASSUMPTION (B3′). For some $c_0 > 0$ there exist positive constants K_1 and K_2 such that, for all x in some neighborhood of θ and all c for which $0 < c \leqq c_0$,

$$K_1(x - \theta)^2 \leqq (x - \theta)[M(x - c) - M(x + c)]c^{-1} \leqq K_2(x - \theta)^2$$

(B3) ((B3′)) is used only for Theorem 2 (2′); it is replaced by a different con dition for later theorems. (B4) which is also used only for Theorems 2 and 2′ is fulfilled whenever M satisfies (B2), (B3′), and has a continuous second derivative in some neighborhood of θ with $M''(\theta) = -2\alpha$ (i.e., $\delta''(\theta) = 0$). When (B2), (B3′), and (B4) hold simultaneously it is redundant to require the lower inequality in (B3′).

It is easy to see that (B3) (also (B3′)) implies that M is symmetric in some neighborhood of θ; in fact, $M(\theta - c) = M(\theta + c)$ for all $c < c_0$. If (B3) is satisfied and the interval of symmetry is known, i.e., c_0 is known, Burkholder was able to show that modifying the Kiefer-Wolfowitz procedure by taking $c_n = c_0$ for all n will yield, under certain additional restrictions, the fact that $n^{1/2}X_n$ is asymptotically normal with mean 0 and a certain variance. It is easy to check that this result can be obtained, under Assumptions (B1) through (B6) and the assumption that $M(x - c_0) - M(x + c_0)$ is differentiable at $x = \theta$, by using Theorem 1, replacing the $M(x)$ in Theorem 1 by $[M(x - c_0) - M(x + c_0)] / c_0$. Since this modification depends on knowing c_0 it will usually be undesirable. Theorem 2 (also 2′) gives a result using the Kiefer-Wolfowitz procedure which has the advantage of not depending on c_0. This gain, however, is offset, if c_0 is known, by the fact that, in general, for the Kiefer-Wolfowitz procedure X_n can never be $O_p(n^{-1/2})$ (see Example 1). However, as noted in the remarks following the proof of Theorem 2, $\{a_n\}$ and $\{c_n\}$ can be chosen so that X_n is arbitrarily close to being $O_p(n^{-1/2})$ without ever attaining it.

Under a stronger set of Assumptions than (B1) through (B6) Derman [6] proves a weaker result than the one we prove in Theorem 2. Using Chung's methods he shows that for any $t < 1/2$ there exist sequences $\{a_n\}$ and $\{c_n\}$ such that $n^t(X_n - \theta)$ is asymptotically normal with mean 0 and a certain variance.

THEOREM 2. *Suppose that Assumptions* (B1) *through* (B6) *are satisfied. Let* $AK_1 > 1/2$ *and take*

$$a_n = An^{-1} \tag{4.10}$$

Let $\{c_n\}$ *be a sequence of positive numbers satisfying* (4.0) *with* $a_n = An^{-1}$, *and the assumptions of Lemma* 5 *with* $r = 0$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normally distributed with mean* 0 *and variance* $\sigma^2A^2(8\alpha A - 1)^{-1}$.

PROOF. With no loss of generality we assume $\alpha_0 = \theta = 0$. Abbreviating $\delta(X_n - c_n) - \delta(X_n + c_n)$ by δ_n and using (B2), we rewrite (4.2) and obtain

$$X_{n+1} = (1 - 4\alpha An^{-1})X_n - An^{-1}c_n^{-1}\delta_n - An^{-1}c_n^{-1}Z_n \tag{4.11}$$

Let $a = 4\alpha A$. Using the notation of (2.1) with $a_j = aj^{-1}$, iteration of (4.11) yields

$$X_{n+1} = \beta_{on} X_1 - A \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m - A \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} Z_m \tag{4.12}$$

Let $h_n = (\sum_{m=1}^{n} a^2 m^{-2} c_m^{-2} \beta_{mn}^2)^{-1/2}$. By Lemma 5 with $r = 0$ and $\rho = 2$ we have

$$h_n^2 \sim a^{-2}(2a - 1)nc_n^2 . \tag{4.13}$$

Hence, what we wish to prove is that

(4.14) $a(2a - 1)^{-1/2} h_n X_n$ is asymptotically normal with mean 0 and variance $A^2\sigma^2(8\alpha A - 1)^{-1}$.

After multiplying both sides of (4.13) by $(2a - 1)^{-1/2} a h_n$ it becomes clear that (4.14) will be proved if we can prove

$$h_n \beta_{on} \to 0, \tag{4.15a}$$

$$h_n \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m \to 0 \quad \text{in probability,} \tag{4.15b}$$

and

(4.15c) $h_n \sum_{m=1}^{n} a m^{-1} c_m^{-1} \beta_{mn} Z_m$ is asymptotically normal with mean 0 and variance σ^2.

Lemma 2 shows that (4.15a) holds. We establish (4.15c) by using the same argument used to prove (3.9c) in Theorem 1. The details being the same we omit the argument except to note that, by Lemma 3, and (B5),

$$\sigma_n^2 = h_n^2 \sum_{m=1}^{n} a^2 m^{-2} c_m^{-2} \beta_{mn}^2 EZ_m^2 \to \sigma^2$$

Note that up to this point the only assumptions used have been (B1), (B2), (B5), and (B6). This observation will enable us to begin the proofs of later theorems at the point where we have to verify (4.15b).

To establish (4.15b) we require an estimate of EX_n^2 which we now obtain.

Squaring both sides of (4.2), taking expected values, and making use of (B3) and (B5) we obtain, for $n > N_0$ where N_0 is large enough so that $c_{N_0} < c_0$,

$$EX_{n+1}^2 \leqq (1 - 2AK_1 n^{-1} + A^2 K_2^2 n^{-2}) EX_n^2 + sA^2 c_n^{-2} n^{-2} \tag{4.16}$$

Let $u > 0$ be such that $2AK_1 - u > 1$ and denote $2AK_1 - u$ by p. Then, for sufficiently large n, say $n \geqq N_1$, (4.16) implies

$$EX_{n+1}^2 \leqq (1 - pn^{-1}) EX_n^2 + D_1 c_n^{-2} n^{-2} \tag{4.17}$$

Put $a_j = pj^{-1}$ in (2.1) and denote the β_{mn} thus obtained by β'_{mn}. Then, iterating (4.17) and using (2.3) and Lemma 4 with $c_n^{-2} = d_n$ and $q = p - 2$, we obtain, for $n > N_1$,

$$\begin{aligned} EX_{n+1}^2 &\leqq \beta'_{N_1 n} EX_{N_1+1}^2 + D_1 \sum_{m=N_1+1}^{n} m^{-2} c_m^{-2} \beta'_{mn} \\ &\leqq D_2 n^{-p} + D_3 \sum_{m=N_1+1}^{n} m^{p-2} c_m^{-2} n^{-p} = O(c_n^{-2} n^{-1}) \end{aligned} \tag{4.18}$$

which is the desired estimate.

For each integer m define ϕ_m to be 1 if $| X_m | \leqq c_m$ and $\phi_m = 0$ for $| X_m | > c_m$. Then, to prove (4.15b) it is sufficient to prove

$$h_n \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m \phi_m \to 0 \quad \text{in probability} \tag{4.19a}$$

$$h_n \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m (1 - \phi_m) \to 0 \quad \text{in probability} \tag{4.19b}$$

For $\epsilon > 0$ it is a consequence of (B4), that, for m sufficiently large, say $m > N_2$, $| \phi_m \delta_m c_m^{-1} | \leqq \epsilon | X_m |$. Use of Lemma 2 and a Chebyshev-type inequality now show that (4.19a) is implied by

$$h_n \sum_{m=N_2}^{n} m^{-1} \beta_{mn} E^{1/2}(X_m^2) = O(1) \tag{4.20a}$$

Since $\delta_m c_m^{-1} = O(| X_m |)$ (a consequence of (B3)) and

$$E(| X_m | (1 - \phi_m)) \leqq P^{1/2}\{ | X_m | > c_m \} E^{1/2}(X_m^2)$$

it follows, in similar fashion, that (4.19b) is implied by

$$h_n \sum_{m=1}^{n} m^{-1} \beta_{mn} E^{1/2}(X_m^2) P^{1/2}\{ | X_m | > c_m \} = o(1) \tag{4.20b}$$

Since our assumptions on $\{c_n\}$ imply that $c_n n^{1/4} \to \infty$ (see the proof of Lemma 4 where it is shown that $d_n n^{q+1} \to \infty$ for $q + 1 > 0$) we have $P\{ | X_m | > c_m \} \leqq c_m^{-2} EX_m^2 = 0(c_m^{-4} m^{-1}) \to 0$ as $m \to \infty$. Hence (4.20b) will follow from (4.20a) by an argument like that in Lemma 3.

To show (4.20a) observe that by (2.3), (4.18), and Lemma 4 with $d_m = c_m^{-1}$ and $q = a - 3/2$ ($a - 3/2 > -1$ because our assumptions imply that $4\alpha \geqq K_1$ and hence $a = 4\alpha A \geqq AK_1 > 1/2$) we have, for $n > N_3 = \max (N_1, N_2)$,

$$h_n \sum_{N_3}^{n} m^{-1} \beta_{mn} E^{1/2}(X_m^2) \leqq D_4 h_n n^{-a} \sum_{N_3}^{n} m^{a-3/2} c_m^{-1} \leqq D_5 h_n n^{-1/2} c_n^{-1}. \tag{4.21}$$

Use of (4.13) and Lemma 2 yields (4.20a) thus completing the proof of Theorem 2.

THEOREM 2′. *Suppose that all the conditions of Theorem 2 are satisfied except that* (B3) *is replaced by* (B3′). *Then* $n^{1/2} c_n (X_n - \theta)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2 A^2 (8\alpha A - 1)^{-1}$.

We omit the proof since it follows from Theorem 2 by use of the same truncation argument used in obtaining Theorem 1′ from Theorem 1.

It is easily checked that, for any sequence $\{f_n\}$ of positive numbers approaching 0, there exists a sequence $\{c_n\}$ satisfying the conditions of Theorem 2 and such that $c_n \geqq f_n$. Thus Theorem 2 says that under its conditions we can always find sequences $\{a_n\}$ and $\{c_n\}$ satisfying (4.0) and such that X_n is arbitrarily close to being $O_p(n^{-1/2})$ without ever attaining it. The question then arises as to whether it is possible to choose $\{a_n\}$ and $\{c_n\}$ satisfying (4.0) and such that $X_n = O_p(n^{-1/2})$. The answer to this is, in general, negative. To see this we give the following example.

EXAMPLE 1. Let $M(x) = -x^2/4$. For each x let $Z(x)$ be normally distributed with mean 0 and variance 1/2, and let $Z(X_m - c_m)$ and $Z(X_m + c_m)$ be independent. Then $\{Z_m\}$ is a sequence of independent normal random variables with mean 0 and variance 1. Note also that $\{Z_m\}$ and X_1 are independent. Let $\{a_n\}$ and $\{c_n\}$ be sequences of positive numbers satisfying (4.0). We now show that it is impossible that, for any infinite sequence $\{n_k\}$ of distinct integers,

$$\lim_{y\to\infty} \lim_{k\to\infty} P\{n_k^{1/2} X_{n_k+1} < y\} = 1. \tag{4.22}$$

For, if it were possible, writing $X_{n+1} = \beta_{on} X_1 - \sum_{m=1}^{n} a_m c_m^{-1} \beta_{mn} Z_m$, we would have

$$\lim_{y\to\infty} \lim_{k\to\infty} P\left\{ n_k^{1/2} \sum_{m=1}^{n_k} a_m c_m^{-1} \beta_{mn_k} Z_m < y \right\} = 1$$

which, by the normality of Z_m, implies that

$$n_k \sum_{m=1}^{n_k} a_m^2 c_m^{-2} \beta_{mn_k}^2 = O(1).$$

But this is impossible by Lemma 1 and the fact that

$$\left(\sum_{m=1}^{n_k} a_m c_m^{-1} \beta_{mn_k} \right)^2 \leqq n_k \sum_{m=1}^{n_k} a_m^2 c_m^{-2} \beta_{mn_k}^2$$

For Theorem 3 we drop (B3) and (B4) and substitute in their place

ASSUMPTION (B7). There exist positive numbers ϵ, c_0, and K_1 with $\epsilon > c_0$ such that, for all $c \leqq c_0$ and all x satisfying $c < |x - \theta| < \epsilon$

$$(x - \theta)[M(x - c) - M(x + c)]c^{-1} > K_1(x - \theta)^2 \tag{4.23}$$

(B2) and (B7) are both implied by the condition (which we refer to hereafter as the derivative condition) that M has a continuous second derivative in a neighborhood of θ with $M''(\theta) = -2\alpha$. Under (B1), the derivative condition, (B5), and the assumption that (3.6) hold for *all* $v > 0$ and some $\epsilon > 0$, Burkholder produces, for every $t < 1/4$, sequences $\{a_n\}$ and $\{c_n\}$ for which $n^t(X_n - \theta)$ is asymptotically normal with mean 0 and a certain variance. Theorem 3 shows that under weaker restrictions the same is true for $t = 1/4$.

THEOREM 3. *Suppose that Assumptions* (B1), (B2), (B5), (B6) *and* (B7) *are satisfied with* $K_1 \leqq 4\alpha$ *in* (B7). *Let* $c_n = n^{-1/4}$ *and* $a_n = An^{-1}$ *where A is such that*

$AK_1 > 1/4$. *Then* $n^{1/4}(X_n - \theta)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2 A^2 (8\alpha A - 1)^{-1}$.

PROOF. Let $\alpha_0 = \theta = 0$. If we prove Theorem 3 when (B7) is strengthened so that (4.23) holds for all x satisfying $|x| > c$ and, in addition, $|M(x - c) - M(x + c)| \leqq K_2 |x|$ for all $|x| > \epsilon$ and all $c \leqq c_0$ then, by using the truncation device used in the proof of Theorem 1′, we will be able to establish Theorem 3 with (B7) as it stands. By the remarks made in the proof of Theorem 2 we will be finished with this proof if we can verify (4.15b).

As previously we require an estimate of EX_n^2 obtained as follows. Let $\phi_n = 1$ if $|X_n| \leqq c_n$ and $\phi_n = 0$ if $|X_n| > c_n$. Let $t > 0$. Then, it is a consequence of (B2) that, for n sufficiently large, say $n > N_1$

$$|\delta_n| \phi_n \leqq tc_n^2 \tag{4.24}$$

Squaring (4.2), taking expected values, and using (B5), (B2), and the strengthened form of (B7) yields

$$\begin{aligned} EX_{n+1}^2 &\leqq E\phi_n(X_n - An^{-1}c_n^{-1}M_n)^2 + E(1 - \phi_n)(X_n - An^{-1}c_n^{-1}M_n)^2 \\ &\quad + D_8 n^{-2}c_n^{-2} \\ &\leqq E\phi_n(1 - 4\alpha An^{-1})^2 X_n^2 + D_9 E\phi_n |\delta_n X_n| n^{-1}c_n^{-1} \\ &\quad + D_{10}E\phi_n \delta_n^2 n^{-2}c_n^{-2} \\ &\quad + E(1 - \phi_n)(1 - 2K_1An^{-1} + A^2K_2^2n^{-2}c_n^{-2})X_n^2 + D_8n^{-2}c_n^{-2} \end{aligned} \tag{4.25}$$

Let u and w be positive numbers such that $2K_1A - w > 1/2$ and $2K_1A - w < 8\alpha A - u$, and let $2K_1A - w$ be denoted by p. Choose $N_2 > N_1$ so that, if $n > N_2$, $A^2K_2^2n^{-2}c_n^{-2} < wn^{-1}$ and $16\alpha^2A^2n^{-2} < un^{-1}$. Then, for all $n > N_2$, we have from (4.24) and (4.25)

$$\begin{aligned} EX_{n+1}^2 &\leqq (1 - pn^{-1})EX_n^2 + D_{11}n^{-2}c_n^{-2} + tD_9n^{-1}c_n^2 \\ &= (1 - pn^{-1})EX_n^2 + D_{12}n^{-3/2} \end{aligned} \tag{4.26}$$

Iteration of (4.26) now shows that, for $n > N_2$,

$$EX_{n+1}^2 = O(n^{-1/2}) \tag{4.27}$$

which is the desired estimate.

(B2) and the fact that $X_n \to 0$ w.p.1 imply that

$$\lim_{n\to\infty} \delta_n(X_n^2 + c_n^2)^{-1} = 0$$

w.p.1. Hence an argument like that in Theorem 1((3.20) et seq) shows that in order to verify (4.15b) it is sufficient to prove that, for any integer $N > N_2$

$$h_n \sum_{m=N}^{n} m^{-1}c_m^{-1}\beta_{mn}(EX_m^2 + c_m^2) = O(1). \tag{4.28}$$

Putting $c_m = m^{-1/4}$ in (4.28) and using (4.27) this follows quite easily, thus finishing the proof of Theorem 3.

If we make some further assumptions about M we will be able to improve on the $n^{1/4}$ obtained in Theorem 3. To this end note that from (B2) we have

$$|\delta(x)| \leqq \epsilon_x(x-\theta)^2 \text{ where } \epsilon_x \to 0 \text{ as } x \to \theta.$$

Assumption (B8) which we now specify is an assumption about ϵ_x.

Assumption (B8). There exist positive numbers c_0, ρ, and R such that, for all $c \leqq c_0$,

$$\sup_{|x-\theta|\leqq c} \epsilon_x \leqq Rc^\rho$$

If $\delta(x, \theta) = O(|x-\theta|^3)$ for x near θ it is easy to see that (B8) is satisfied for appropriate R and c_0 and $\rho = 1$; thus, the case of most interest is when $\rho = 1$. (B8) is very closely related to Burkholder's condition of "local-evenness"—for $\rho \geqq 0$, M is called ρ-locally-even if

$$\limsup_{\epsilon\to 0} f(\epsilon)\epsilon^{-1-\rho} < \infty$$

where $f(\epsilon) = \sup\{x \mid M(x-\epsilon) - M(x+\epsilon) \leqq 0\}$. It is easy to verify that when M is continuous in a neighborhood of θ and (B8) is satisfied then M is ρ-locally-even. In fact, when M satisfies the derivative condition, requiring M to be ρ-locally-even is equivalent to requiring $\delta(x, \theta) - \delta(-x, \theta) = O(|x-\theta|^{2+\rho})$ as $x - \theta \to 0$. The disadvantage in assuming the slightly more restrictive (B8) rather than local-evenness is allayed by the fact that (B8) appears as a more natural condition.

Under (B1), the derivative condition, (B5), the assumption that (3.6) hold for all $v > 0$ and some $\epsilon > 0$, and the assumption that M is ρ-locally-even, Burkholder proves that, for any $t < (1+\rho)/(4+2\rho)$, there exist sequences $\{a_n\}$ and $\{c_n\}$ such that $n^t(X_n - \theta)$ is asymptotically normal. Theorem 4 replaces the condition of local-evenness by (B8), weakens the other assumptions made by Burkholder, and gives a stronger conclusion, e.g., with $\rho = 1$ in (B8), $a_n = An^{-1}$, and $c_n = (n^{1/6}\log n)^{-1}$, Theorem 4 shows that $n^{1/3}(\log n)^{-1}(X_n - \theta)$ is asymptotically normal.

Theorem 4. *Suppose that Assumptions* (B1), (B2), (B5), (B6), (B7), *and* (B8) *are satisfied with* $K_1 \leqq 4\alpha$ *in* (B7). *Let* $a_n = An^{-1}$ *where* A *is such that* $AK_1 > 1$ *and let* $c_n = d_n n^{-r}$ *with* $d_n \to 0$ *and satisfying* (2.5) *of Lemma* 4 *and with* $r = (4+2\rho)^{-1}$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2A^2(8\alpha A - 1)^{-1}$.

Proof. Let $\alpha_0 = \theta = 0$. By the reasoning in the first paragraph of the proof of Theorem 3 we have only to verify (4.15b). To do so we require an estimate of EX_n^2 which is obtained in much the same way as (4.26) is obtained. In fact, using (B8) to replace (4.24) by

$$|\phi_n\delta_n| = O(c_n^{2+\rho}), \tag{4.29}$$

a repetition of the argument given in Theorem 3 shows that

$$EX_{n+1}^2 \leqq (1 - pn^{-1})EX_n^2 + D_{13}n^{-1}c_n^{2+\rho} + D_{14}n^{-2}c_n^{-2} \tag{4.30}$$

Iterating (4.30) and applying Lemma 4 and (2.3)—note that $p = 2K_1A - w > 1$ and that $p > r(2 + \rho)$—yields, for $n > N$ where N is chosen sufficiently large (how large can be determined by inspecting the proof of Theorem 3),

$$\begin{aligned} EX_{n+1}^2 &\leqq O(n^{-p}) + D_{13} \sum_{m=N}^{n} m^{-1} c_m^{2+\rho} \beta'_{mn} + D_{14} \sum_{m=N}^{n} m^{-2} c_m^{-2} \beta'_{mn} \\ &= 0(n^{-p}) + O(c_n^{2+\rho}) + O(n^{-1} c_n^{-2}) \\ &= O(c_n^{2+\rho}) + O(n^{-1} c_n^{-2}) \end{aligned} \tag{4.31}$$

which is the desired estimate.

To prove (4.15b), it is sufficient to prove that for N_1 sufficiently large

$$h_n \sum_{m=N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} \phi_m \delta_m \rightarrow 0 \quad \text{in probability} \tag{4.32}$$

$$h_n \sum_{m=N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} (1 - \phi_m) \delta_m \rightarrow 0 \quad \text{in probability.} \tag{4.33}$$

By (4.29), (2.7) of Lemma 5, and Lemma 4—note that $4\alpha A > r(1 + \rho)$—we obtain, for $n > N_1$,

$$\left| h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} \phi_m \delta_m \right| = O\left(h_n \sum_{N_1}^{n} m^{-1} c_m^{1+\rho} \beta_{mn} \right) = O(n^{1/2} c_n^{2+\rho}) \tag{4.34}$$

which, by the choice of $\{c_n\}$, proves (4.32).

To show (4.33) we proceed as follows. Let $\mu_m = 1$ if $| X_m | \leqq c_0$ (c_0 here is the same as in (B8)) and let $\mu_m = 0$ otherwise. Using (B8) and, from (B2), the fact that $| \delta_m | = O(| X_m |)$ if $| X_m | > c_0$, we then have, for N_1 sufficiently large so that in particular $c_m < c_0$ for $m \geqq N_1$

$$\begin{aligned} & h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} (1 - \phi_m) \, | \delta_m | \\ &= h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} (1 - \phi_m) \mu_m \, | \delta_m | \\ &\qquad + h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} (1 - \phi_m)(1 - \mu_m) \, | \delta_m | \\ &= O\left(h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} X_m^2 \right) + O\left(h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} (1 - \mu_m) \, | X_m | \right) \end{aligned} \tag{4.35}$$

Now

$$h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} E\{ (1 - \mu_m) \, | X_m | \} \leqq h_n c_0^{-1} \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} EX_m^2$$

and since, by (4.31) and Lemma 4 (note that $4\alpha A > 1$),

$$h_n \sum_{N_1}^{n} m^{-1} c_m^{-1} \beta_{mn} EX_m^2 = O(n^{1/2} c_n^{2+\rho}) + O(n^{-1/2} c_n^{-2}),$$

our choice of c_n shows that the right hand side of (4.35) goes to 0 in probability. This establishes (4.33) and finishes the proof of the theorem.

Focusing our attention for the present on the case $\rho = 1$ (this is by no means necessary since all ensuing remarks can be suited to the cases where $\rho \neq 1$), we can ask whether or not it is possible to find sequences $\{a_n\}$ and $\{c_n\}$ satisfying (4.0) and a sequence $\{g_n\}$ such that, under Assumptions (B1), (B2), and (B5) to (B8), $g_n X_{n+1}$ is asymptotically normal and $g_n^{-1} = O(n^{-1/3})$. Example 2, which we now give, shows that the answer to this question is no.

EXAMPLE 2. Let $\{a_n\}$ and $\{c_n\}$ be sequences satisfying (4.0). For $0 < C < 1/6$ let $M(x)$ be defined as follows.

$$\begin{aligned} M(x) &= -x^2/4 + x^3 \quad \text{if} \quad |x| \leqq C \\ &= -x^2/4 + C^3 \quad \text{if} \quad x > C \\ &= -x^2/4 - C^3 \quad \text{if} \quad x < -C \end{aligned} \tag{4.36}$$

For each x let $Z(x)$ be normally distributed with mean 0 and variance 1/2 and let $Z(X_m - c_m)$ and $Z(X_m + c_m)$ be independently distributed. Thus $\{Z_m\}$ is a sequence of independent normal random variables with mean 0 and variance 1 and Z_m and $X_{m'}$ are independent if $m \geqq m'$. It is clear that (B1), (B2), (B5), (B6), (B7), and (B8) with $\rho = 1$ are all satisfied. Suppose that $\{g_n\}$ is a sequence of real numbers such that $g_n X_{n+1}$ converges in distribution to the normal distribution with mean 0 and variance v with $v \geqq 0$. Since $|g_n| X_{n+1}$ is then also asymptotically normal with mean 0 and variance v we can assume to begin with that, for all n, $g_n \geqq 0$. We will show that $\limsup_{n\to\infty} n^{-1/3} g_n = 0$.

Let $\phi_{1m}, \phi_{2m}, \cdots, \phi_{5m}$ be random variables taking on the values 0 and 1 only, with the value 1 being taken on as follows:

$$\begin{aligned} \phi_{1m} &= 1 \quad \text{if} \quad |X_m - c_m| \leqq C, |X_m + c_m| \leqq C \\ \phi_{2m} &= 1 \quad \text{if} \quad X_m - c_m > C \\ \phi_{3m} &= 1 \quad \text{if} \quad X_m - c_m \leqq C < X_m + c_m \\ \phi_{4m} &= 1 \quad \text{if} \quad X_m + c_m < -C \\ \phi_{5m} &= 1 \quad \text{if} \quad X_m - c_m < -C \leqq X_m + c_m \end{aligned}$$

Let N_0 be such that, for all $m > N_0$, $c_m < C/2$ and, in addition, suppose that N_0 is large enough so that $a_m < 1$ for all $m > N_0$—the latter requirement is to guarantee that $\beta_{mn} > 0$ for all $n \geqq m > N_0$. Since for all $m > N_0$, $\sum_{i=1}^5 \phi_{im} \equiv 1$, it follows from (4.36) that $m > N_0$ implies

$$\begin{aligned} M_m = M(X_m - c_m) - M(X_m + c_m) &= \sum_{i=1}^{5} M_m \phi_{im} \\ &= c_m X_m - 2c_m^3 \phi_{1m} - 6c_m X_m^2 \phi_{1m} + ((X_m - c_m)^3 - C^3)\phi_{3m} \\ &\quad - (C^3 + (X_m + c_m)^3)\phi_{5m}. \end{aligned}$$

Observe that none of the last three terms is positive. Abbreviating $-a_m/c_m$ times their sum by G_m we obtain from (4.2)

$$X_{n+1} = (1 - a_m)X_m + 2c_m^2 a_m \phi_{1m} + G_m - a_m c_m^{-1} Z_m \tag{4.37}$$

Iterating (4.37) we obtain, for $n > N \geqq N_0$,

$$\begin{aligned} X_{n+1} &= \beta_{Nn} X_{N+1} + 2 \sum_{N+1}^{n} a_m c_m^2 \beta_{mn} \phi_{1m} + \sum_{N+1}^{n} \beta_{mn} G_m - \sum_{N+1}^{n} a_m c_m^{-1} \beta_{mn} Z_m \\ &= \beta_{Nn} X_{N+1} + G_{1nN} + G_{2nN} + G_{3nN} \end{aligned} \tag{4.38}$$

where G_{1nN}, G_{2nN}, and G_{3nN} are abbreviations for the terms in the corresponding positions in the previous line and we note that G_{2nN} is never negative and that G_{3nN} is normally distributed with mean 0 and variance $\sum_{N+1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2$.

We will show that $\lim\sup_{n\to\infty} n^{-1/3} g_n = 0$ by contradicting the assumption that there exists a positive constant D_{15} and a subsequence $\{n_k\}$ such that $n_k^{-1/3} g_{n_k} \geqq D_{15}$ for all k. We may assume that $\{n_k\}$ consists of all the positive integers since the argument below remains valid if we begin by restricting ourselves to the subsequence $\{n_k\}$ for which $n_k^{-1/3} g_{n_k} \geqq D_{15}$. Let

$$H_{nN} = \left(\sum_{N+1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2\right)^{-1/2} \text{ and } G'_{nN} = 2 \sum_{N+1}^{n} a_m c_m^2 \beta_{mn} .$$

We will arrive at the contradiction by showing first that the asymptotic normality of $g_n X_{n+1}$ implies that $g_n H_{nN_0}^{-1} = O(1)$ as $n \to \infty$ and $\lim\sup_{n\to\infty} g_n G'_{nN} = o(1)$ as $N \to \infty$, and then showing (see (4.45) et seq) the impossibility of having simultaneously $H_{nN_0}^{-1} = O(n^{-1/3})$ and $\lim\sup_{n\to\infty} n^{1/3} G'_{nN} = o(1)$ as $N \to \infty$.

Let E_m be the set $\{ \mid X_j \mid \leqq C/2, j > m\}$. Since $X_m \to 0$ w.p.1, $1 - P\{E_m\} = \epsilon_m \to 0$ as $m \to \infty$. Since $c_m < C/2$ for all $m > N_0$ we have, for all such m,

$$E_m \subset \{\phi_{1j} = 1, j > m\}. \tag{4.39}$$

For $v \geqq 0$ let F_v denote the normal distribution with mean 0 and variance v. We consider two cases according as $v = 0$ or $v > 0$.

Case 1: $v = 0$. To begin with we obtain from (4.38), the fact that $G_{1nN} + G_{2nN}$ is never negative, and the independence of X_{N+1} and G_{2nN} that, for all $n > N_0$,

$$\begin{aligned} P\{X_{n+1} > 0\} &= P\{\beta_{N_0 n} X_{N_0+1} + G_{1nN_0} + G_{2nN_0} + G_{3nN_0} > 0\} \\ &\geqq P\{\beta_{N_0 n} X_{N_0+1} + G_{3nN_0} > 0\} \\ &\geqq P\{X_{N_0+1} > 0, G_{3nN_0} > 0\} \\ &= \tfrac{1}{2} P\{X_{N_0+1} > 0\} \end{aligned} \tag{4.40}$$

We will show that for some N $\lim_{n\to\infty} g_n G'_{nN} = 0$. Since G'_{nN} is decreasing in N this will imply $\lim_{N\to\infty} \lim\sup_{n\to\infty} g_n G'_{nN} = 0$.

Suppose that for all N $\lim\sup_{n\to\infty} g_n G'_{nN} > 0$. Then, for each N, there exists a positive constant D_{16} and a sequence $\{n_k\}$ such that, for all k, $g_{n_k} G'_{n_k N} > D_{16}$. Let $G^*_{nN} = g_n G_{3nN}$. Then, since $g_n X_{n+1} \to 0$ in probability we have, by (4.38), (4.39) and (4.40),

$$0 = 1 - F_0(D_{16}) \geqq \lim_{k\to\infty} P\{g_{n_k}\beta_{Nn_k}X_{N+1} + G^*_{n_kN} + g_{n_k}G_{1n_kN} + g_{n_k}G_{2n_kN} > D_{16}; E_N\}$$

$$\geqq \lim_{k\to\infty} P\{g_{n_k}\beta_{Nn_k}X_{N+1} + G^*_{n_kN} > 0; E_N\} \tag{4.41}$$

$$\geqq \lim_{k\to\infty} P\{X_{N+1} > 0, G^*_{n_kN} > 0\} - \epsilon_N$$

$$\geqq \tfrac{1}{4}P\{X_{N_0+1} > 0\} - \epsilon_N$$

Since N can be chosen large enough so that the right-hand side of (4.41) is strictly positive we have a contradiction, thus proving that $g_nG'_{nN} = o(1)$ for some N.

To show that $g_nH^{-1}_{nN_0} = O(1)$ assume, to the contrary, that $g_{n_k}H^{-1}_{n_kN_0} \to \infty$ for some sequence $\{n_k\}$. Since $H_{nN_0}\, G_{3nN_0}$ is normally distributed with mean 0 and variance 1 we would have, for any y,

$$\lim_{k\to\infty} P\{G^*_{n_kN_0} > y\} = \tfrac{1}{2}$$

Hence

$$0 = 1 - F_0(y) \geqq \lim_{k\to\infty} P\{g_{n_k}\beta_{N_0n_k}X_{N_0+1} + G^*_{n_kN_0} > y\}$$

$$\geqq \lim_{k\to\infty} P\{G^*_{n_kN_0} > y\}P\{X_{N_0+1} > 0\}$$

$$\geqq \tfrac{1}{2}P\{X_{N_0+1} > 0\}$$

which is a contradiction, thus proving that $g_nH^{-1}_{nN_0} = O(1)$.

Case 2: $v > 0$. The argument used in Case 1 to show that $g_nH^{-1}_{nN_0} = O(1)$ can be used here with (4.42) becoming

$$1 - F_v(y) \geqq \tfrac{1}{2}P\{X_{N_0+1} > 0\}$$

Let $y \to \infty$ and we obtain a contradiction to the assumption that $g_nH^{-1}_{nN_0}$ is not $O(1)$.

To show that $\limsup_{n\to\infty} g_n\, G'_{nN} = o(1)$ suppose, to the contrary, there exists a sequence $\{N_j\}$, a positive number D_{17}, and a sequence $\{n_{kj}\}$ such that, for all j, k, $g_{n_{kj}}\, G'_{n_{kj}N_j} > D_{17}$. Let T be a random variable independent of

$$\{G_{3nN}\,,\, n > N, N > 0\}$$

and having F_v as its distribution function. Then

$$\sup_y |\, P\{g_nX_{n+1} < y\} - P\{T < y\}\,| < t_n \tag{4.43}$$

where $t_n \to 0$ as $n \to \infty$. Then, letting $u_{kj} = g_{n_{kj}} \beta_{N_j n_{kj}} g_{N_j}^{-1}$, we have

$$
\begin{aligned}
1 - F_v(D_{17}) &\geqq \limsup_{k\to\infty} P\{u_{kj} g_{N_j} X_{N_j+1} + G^*_{n_{kj}N_j} > 0; E_{N_j}\} \\
&\geqq \limsup_{k\to\infty} P\{u_{kj} g_{N_j} X_{N_j+1} + G^*_{n_{kj}N_j} > 0\} - \epsilon_{N_j} \\
&\geqq \limsup_{k\to\infty} P\{u_{kj} T + G^*_{n_{kj}N_j} > 0\} - \epsilon_{N_j} - t_{N_j} \\
&= \tfrac{1}{2} - \epsilon_{N_j} - t_{N_j}
\end{aligned}
\tag{4.44}
$$

For j large enough it is clear that $\epsilon_{N_j} + t_{N_j} < F_v(D_{17}) - \frac{1}{2}$ which gives the desired contradiction.

To conclude the argument we have to show that it is impossible to have $\lim_{N\to\infty} \limsup_{n\to\infty} n^{1/3} G'_{nN} = 0$ and $n^{1/3} H^{-1}_{nN_0} = O(1)$. If it were possible we would have, for $N \geqq N_0$,

$$
\begin{aligned}
\left(\sum_{m=N+1}^{n} a_m^{3/2} \beta_{mn}^{3/2}\right)^2 &\leqq \left(\sum_{m=N+1}^{n} a_m c_m^2 \beta_{mn}\right)\left(\sum_{m=N+1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2\right) \\
&\leqq \left(\sum_{N+1}^{n} a_m c_m^2 \beta_{mn}\right)\left(\sum_{N_0+1}^{n} a_m^2 c_m^{-2} \beta_{mn}\right) \\
&\leqq \epsilon_{nN} n^{-1}
\end{aligned}
\tag{4.45}
$$

where $\lim_{N\to\infty} \limsup_{n\to\infty} \epsilon_{nN} = 0$. Hence, using Holder's inequality,

$$
\left(\sum_{N+1}^{n} a_m \beta_{mn}\right)^3 \leqq \left(\sum_{N+1}^{n} a_m^{3/2} \beta_{mn}^{3/2}\right)^2 (n - N) \leqq \epsilon_{nN}(n - N)n^{-1}
\tag{4.46}
$$

Applying Lemma 1 with $W_m = W = 1$ we conclude that, for each $N > N_0$,

$$
1 \leqq \limsup_{n\to\infty} \epsilon_{nN}
$$

which is impossible.

The reason that we cannot have $n^{1/3} X_n$ asymptotically normal in Example 2 is clearly the upsetting character of G_{1nN}. The following example shows how we can obtain $n^{1/3}$, and even better, by considering the asymptotic behavior of $\{X_{n+1} - G'_{n0}\}$ instead of $\{X_n\}$. What this indicates is that the "bias" term, G'_{n0}, around which X_n becomes rapidly concentrated, is the dominant error. Of course, the improvement in the order of convergence is of little practical use since it is $X_n - \theta$ which matters.

EXAMPLE 3. Let M be as in Example 2 and let $Z(x)$ satisfy (B5) and (B6). Note that M satisfies (B8) with $\rho = 1$. Let $a_n = An^{-1}$ for $A > 1$ and let $c_n = n^{-1/8} d_n$ where d_n satisfies the conditions of Lemma 4 and $d_n \to 0$. We will show that $n^{1/2} c_n (X_{n+1} - G'_{n0})$ is asymptotically normal with mean 0 and variance $\sigma^2 A^2 (2A - 3/4)^{-1}$. By (4.36) and the kind of argument used several times before we will succeed in doing so if we can show

$$h_n \sum_{N+1}^{n} a_m c_m^2 \beta_{mn}(1 - \phi_m) = o_p(1) \tag{4.41}$$

$$h_n \sum_{N+1}^{n} a_m \beta_{mn} X_m^2 = o_p(1) \tag{4.42}$$

where ϕ_m is 1 or 0 according as $|X_m| \leqq C - c_m$ or not, and where $\beta_{mn} = \prod_{j=m+1}^{n} (1 - a_j)$. By use of Chebyshev's inequality (4.41) and (4.42) will be proved if we show

$$h_n \sum_{N+1}^{n} a_m \beta_{mn} EX_m^2 = o(1). \tag{4.43}$$

Using (4.31) and Lemma 4 (note that $A > 1$) we obtain

$$\begin{aligned} h_n \sum_{N+1}^{n} m^{-1} \beta_{mn} EX_m^2 &= O\left(h_n \sum_{1}^{n} m^{-1} c_m^3 \beta_{mn}\right) + O\left(h_n \sum_{1}^{n} m^{-2} c_m^{-2} \beta_{mn}\right) \\ &= O(n^{1/2} c_n^4) + O(n^{-1/2} c_n^{-1}) = o(1) \end{aligned} \tag{4.44}$$

which proves (4.43). Note that we can do better than $n^{1/3}$ and, in fact, we can get arbitrarily close to $n^{3/8}$.

Blum in [3] has suggested a procedure which replaces (4.1) by

$$X_{n+1} = X_n - a_n c_n^{-1}[Y(X_n) - Y(X_n + c_n)].$$

This was suggested mainly for the multi-dimensional case which we consider in the next section but we point out here, in Example 4, that this procedure can be rather inefficient.

EXAMPLE 4. Let $M(x) = -x^2/2$ and let Z be as in Example 1. Then, using the Blum procedure,

$$X_{n+1} = \beta_{0n} X_1 - \frac{1}{2} \sum_{1}^{n} a_m c_m \beta_{mn} - \sum_{1}^{n} a_m c_m^{-1} \beta_{mn} Z_m$$

We show that if $h_n X_n = O_p(1)$ then h_n^{-1} cannot be $o(n^{-1/4})$. Theorem 2 shows, of course, that, for the Kiefer-Wolfowitz procedure, h_n^{-1} can be almost $O(n^{-1/2})$.

If $h_n^{-1} = o(n^{-1/4})$ and $X_1 < 0$ we would have

$$\sum_{1}^{n} a_m c_m \beta_{mn} = o(n^{-1/4})$$

$$\sum_{1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2 = o(n^{-1/2})$$

Hence

$$\sum_{1}^{n} a_m^{4/3} \beta_{mn}^{4/3} \leqq \left(\sum_{1}^{n} a_m c_m \beta_{mn}\right)^{2/3} \left(\sum_{1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2\right)^{1/3} = o(n^{-1/3})$$

But then

$$\sum_1^n a_m \beta_{mn} \leqq \left(\sum_1^n a_m^{4/3} \beta_{mn}^{4/3} \right)^{3/4} n^{1/4} = o(1)$$

which is impossible by Lemma 1.

Again it is the "bias" term $\sum a_m c_m \beta_{mn}$ which is the dominant error, i.e., the Blum procedure becomes rapidly concentrated about the wrong value just as in Example 3.

5. Multi-Dimensional Procedures. In this section we consider multi-dimensional analogues of the Robbins-Monro and Kiefer-Wolfowitz procedures. Since the theorems and proofs for the multi-dimensional case are quite similar to those for the one dimensional case considered in Sections 3 and 4 we will not go into great detail in this section. We first consider a q-dimensional analogue of the Robbins-Monro procedure identical with the one considered by Blum [3]. The q-dimensional analogue of the Kiefer-Wolfowitz procedure considered next differs somewhat from the procedure given by Blum—the differences are pointed out below. At the end of the section we remark on some more general q-dimensional analogues.

Let x be a q-vector and let M be a vector-valued function of x with $M(x)$ also being a q-vector. Let α be a vector and let θ be a solution of the equation $M(x) = \alpha$. Let $Y(x)$ be a vector random variable with $EY(x) = M(x)$. The Robbins-Monro procedure for "locating" θ is given as follows.

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

$$\sum a_n = \infty, \quad \sum a_n^2 < \infty \tag{5.0}$$

Let X_1 be an arbitrary vector (as in Section 3 X_1 can actually be taken to be a random variable) and define $\{X_n, n \geqq 2\}$ by the recursion

$$X_{n+1} = X_n - a_n(Y(X_n) - \alpha) \tag{5.1}$$

where $Y(X_n)$ is a random variable whose conditional distribution given $X_1 = x_1, \cdots, X_n = x_n$ is the same as $Y(x_n)$. Writing $Y(x) = M(x) + Z(x)$ we obtain from (5.1)

$$X_{n+1} = X_n - a_n(M(X_n) - \alpha) - a_n Z(X_n). \tag{5.2}$$

where, as before, the conditional distribution of $Z(X_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Z(x_n)$ and

$$E(Z(X_n) \mid X_1, \cdots, X_n) = 0 \tag{5.3}$$

w.p.1.

The assumptions we make now are easily seen to correspond to the assumptions made in Section 3—(A2*) corresponding, of course to (A2′). The notation we use is the same as that adopted in Section 2 for Lemma 6.

Assumption (A1*). M is Borel-measurable, $M(\theta) = \alpha$, and, for every $\epsilon > 0$

$$\inf_{1/\epsilon > |x-\theta| > \epsilon} [x - \theta, M(x) - \alpha] > 0$$

(A1*) is satisfied, for example, if (A3*) is satisfied with $\delta = 0$ which is, of course, much stronger than needed.

ASSUMPTION (A2*). There exists a positive constant K_1 such that, for all x,

$$| M(x) - \alpha | \leqq K_1 | x - \theta |$$

ASSUMPTION (A3*). For all x

$$M(x) = \alpha + B(x - \theta) + \delta(x, \theta)$$

where B is a positive definite $q \times q$ matrix and $| \delta(x, \theta) | = o(| x - \theta |)$ as $x - \theta \to 0$.

ASSUMPTION (A4*).

$$\sup_x E | Z(x) |^2 < \infty \tag{5.4}$$

$$\lim_{x \to \theta} EZ(x)Z'(x) = \pi \tag{5.5}$$

where π is a non-negative definite matrix and where the limit is in the sense of the norm we have defined.

ASSUMPTION (A5*).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x - \theta| < \epsilon} \int_{\{|Z(x)| > R\}} | Z(x) |^2 \, dP = 0$$

The remarks concerning Assumption (A5) in Section 3 also pertain here—we use (A5*) in conjunction with the convergence of X_n to θ w.p.1 (a consequence of (A1*), (A2*) and (5.4)) and (5.4) only to obtain

$$\lim_{R \to \infty} \sup_k \int_{\{|Z(X_k)| > R\}} | Z(X_k) |^2 \, dP = 0 \tag{5.6}$$

As before, with $Z(x)$ considered as a vector, (3.5) and (3.6) imply (A5*).

Let $b_1, \cdots, b_q$ denote the eigenvalues of B in decreasing order. Write $B = PDP^{-1}$ where P is orthogonal and D is the diagonal matrix whose diagonal elements are $b_1, \cdots, b_q$. Observe that $\inf_{|x|=1} [Bx, x] = b_q$, $\inf_{|x|=1} [Bx, Bx = b_q^2$, and $\| B \| = b_1$. Let π_{ij} be the (i, j)th element of π and let π_{ij}^* be the (i, j)th element of $\pi^* = P^{-1}\pi P$.

THEOREM 5. *Suppose that Assumptions* (A1*) *through* (A5*) *are satisfied. Let* $a_n = An^{-1}$ *where* A *is such that* $Ab_q > \frac{1}{2}$. *Then* $n^{1/2}(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1} *where* Q *is the matrix whose* (i, j)th *element is* $A^2(Ab_i + Ab_j - 1)^{-1} \pi_{ij}^*$.

PROOF. Let $\alpha = \theta = 0$. Let $u = P^{-1}x$, $M^*(u) = P^{-1}M(Pu)$, $\delta^*(u) = P^{-1}\delta(Pu)$, and $Z^*(u) = P^{-1}Z(Pu)$. Then, with x, M, δ, and Z being replaced by u, M^*, δ^*, and Z^* respectively, it is easy to see that (A1*) through (A5*) are satisfied with B replaced by D and π replaced by π^* and that (5.2) is transformed into another Robbins-Monro procedure with α replaced by $P^{-1}\alpha$. Thus, in order to prove the theorem it is sufficient to prove that, when B is diagonal, $n^{1/2}X_n$ is asymptotically normal with mean 0 and covariance matrix

$$((A^2(Ab_i + Ab_j - 1)^{-1}\pi_{ij})).$$

(A1*), (A2*), and (5.4) imply that X_n converges to 0 w.p.1 (this follows from Dvoretzky's theorem—Blum's earlier proof of convergence w.p.1 is under stronger assumptions) and, hence, using (A3*), an argument like that in Theorem 1′ shows that we can add the additional restriction that there exists a positive constant K such that $AK > 1$, $K < b_q$, and, for all x,

$$[M(x), x] \geqq K \mid x \mid^2 \tag{5.7}$$

The proof proceeds now just as in Theorem 1. Iterating (5.2) and using (A3*) we obtain

$$X_{n+1} = B_{0n} X_1 - A \sum_{m=1}^{n} m^{-1} B_{mn} \delta_m - A \sum_{m=1}^{n} m^{-1} B_{mn} Z_m \tag{5.8}$$

where

$$B_{mn} = \prod_{m+1}^{n} (I - Aj^{-1}B). \quad \text{Let} \quad h_n = \left(\sum_{1}^{n} A^2 m^{-2} \| B_{mn} \|^2 \right)^{-1/2}$$

Since $\| B_{mn} \| = (1 + \epsilon_m)(mn^{-1})^{Ab_q}$ where $\epsilon_m \to 0$ as $m \to \infty$, we have $h_n \sim (2Ab_q - 1)^{1/2} A^{-1} n^{1/2}$. Making use of (5.7) and the same argument as used to obtain (3.18) in Theorem 1 we obtain $E \mid X_m \mid^2 = O(m^{-1})$. It then follows just as in Theorem 1 that

$$h_n \left(B_{0n} X_1 - A \sum_{m=1}^{n} m^{-1} B_{mn} \delta_m \right) \to 0$$

in probability.

To conclude the proof we will apply Lemma 6 with $U_{nk} = Ah_n k^{-1} B_{kn} Z_k$. Just as in Theorem 1 we obtain quite readily that (2.19) and (2.20) are satisfied with this choice of U_{nk}. Since $\mid B_{kn} x \mid \geqq (1 + \epsilon_k)(kn^{-1})^{Ab_1} \mid x \mid$ it follows from (A5*) in the same way as in Theorem 1 that (2.21) is satisfied. We have only to compute $\lim_{n\to\infty} s_n$ to be finished. Let the (i, j)th elements of $EZ_k Z_k'$ and s_n be denoted by $\pi_{ij}^{(k)}$ and s_n^{ij} respectively. Let $\beta_{ikn} = \prod_{k+1}^{n} (1 - Ab_i j^{-1})$. Then

$$s_n^{ij} = A^2 h_n^2 \sum_{k=1}^{n} k^{-2} \beta_{ikn} \beta_{jkn} \pi_{ij}^{(k)}$$

Since $\pi_{ij}^{(k)} \to \pi_{ij}$ and $h_n^2 \sim (2Ab_q - 1) A^{-2} n$ it follows that

$$s_n^{ij} \to (2Ab_q - 1)(Ab_i + Ab_j - 1)^{-1} \pi_{ij}$$

Thus, when B is diagonal, $n^{1/2} X_n$ is asymptotically normal with mean 0 and covariance matrix $((A^2(Ab_i + Ab_j - 1)^{-1} \pi_{ij}))$, and this finishes the proof of the theorem.

We will now take up the multi-dimensional Kiefer-Wolfowitz procedure. Let x be a q-vector and let f be a real valued function of x. Let $y(x)$ be a real random variable with $Ey(x) = f(x)$. We will consider the following q-dimensional version of the Kiefer-Wolfowitz procedure for finding the point at which f has a maximum.

Let $\{a_n\}$ and $\{c_n\}$ be two sequences of positive real numbers satisfying

$$\sum a_n = \infty, \qquad \sum a_n^2 c_n^{-2} < \infty, \qquad \lim_{n\to\infty} c_n = 0 \tag{5.9}$$

For $1 \leqq i \leqq q$ let e_i be the q-vector whose ith coordinate is 1 and whose other coordinates are 0. Let $Y(x, a) = (y(x + ae_1), \cdots, y(x + ae_q))$. Let X_1 be an arbitrary q-vector and define $\{X_n, n \geqq 2\}$ by the recursion

$$X_{n+1} = X_n - a_n c_n^{-1} (Y(X_n, -c_n) - Y(X_n, c_n)) \tag{5.10}$$

where the conditional distribution of $Y(X_n, \pm c_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as $Y(x_n, \pm c_n)$. Writing $y(x) = f(x) + z(x)$, and letting

$$M(x, a) = (f(x + ae_1), \cdots, f(x + ae_q)),$$

$$Z(x, a) = (z(x + ae_1), \cdots, (z(x + ae_q)),$$

we rewrite (5.10) and obtain

$$X_{n+1} = X_n - a_n c_n^{-1}(M(X_n, -c_n) - M(X_n, c_n)) - a_n c_n^{-1}(Z(X_n, -c_n) - Z(X_n, c_n)) \tag{5.11}$$

We will denote $M(X_n, -c_n) - M(X_n, c_n)$ by M_n and $Z(X_n, -c_n) - Z(X_n, c_n)$ by Z_n. It is clear that just as in Section 4

$$E(Z_{n+1} \mid Z_1, \cdots, Z_n) = E(Z_{n+1} \mid X_1, \cdots, X_{n+1}) = 0 \tag{5.12}$$

w.p.1.

The procedure we have defined by (5.10) differs from the one considered by Blum [3] in that Blum uses $Y(X_n, 0) - Y(X_n, c_n)$ rather than $Y(X_n, -c_n) - Y(X_n, c_n)$. The advantage of the Blum procedure is that it requires at each stage $q + 1$ observations whereas the number of observations required by (5.10) at each stage is $2q$. However, as noted in Example 4, the Blum procedure is quite inefficient with respect to the rate at which it converges to θ.

We now list the assumptions we require. The correspondence between these assumptions and those of Section 4 is easy to see.

ASSUMPTION (B1*). f is Borel-measurable, has a unique maximum at $x = \theta$, $|f(x + 1) - f(x)| \leqq D_1 + D_2 |x|$ for some positive constants D_1 and D_2, and, for $0 < \epsilon_0 < \epsilon_1 < \epsilon_2 < \infty$,

$$\inf_{\substack{\epsilon_1 \leqq |x-\theta| \leqq \epsilon_2 \\ 0<\epsilon\leqq\epsilon_0}} \epsilon^{-1}[M(x, -\epsilon) - M(x, \epsilon), x - \theta] > 0$$

(B1*) is satisfied, for example, if (B2*) is satisfied with $\delta = 0$; of course, this is much stronger than is needed.

ASSUMPTION (B2*). For all x

$$f(x) = \alpha_0 - [B(x - \theta), x - \theta] + \delta(x, \theta)$$

where α_0 is real, B is a positive definite $q \times q$ matrix, and $\delta(x, \theta) = o(|x - \theta|^2)$ as $x - \theta \to 0$.

ASSUMPTION (B3*). There exist positive numbers K_1, K_2, and c_0 such that, for all x in some neighborhood of θ and all c with $0 < c \leqq c_0$,

$$K_1 \mid x - \theta \mid^2 \leqq [x - \theta, (M(x, -c) - M(x, c))/c] \leqq K_2 \mid x - \theta \mid^2$$

and, for all x,

$$\left| \frac{M(x, -c) - M(x, c)}{c} \right| \leqq K_3 \mid x \mid.$$

ASSUMPTION (B4*). If $c_0 > 0$ then, for all x and c such that $\mid x \mid < c < c_0$

$$(\delta(x, -c, \theta) - \delta(x, +c, \theta))/c = o(\mid x - \theta \mid)$$

ASSUMPTION (B5*).

$$\sup_x E \mid Z(x, 0) \mid^2 < \infty \tag{5.13}$$

$$\lim_{\substack{x \to \theta \\ c \to 0}} E(Z(x, -c) - Z(x, c))(Z(x, -c) - Z(x, c))' = \pi \tag{5.14}$$

where π is a non-negative definite matrix.

ASSUMPTION (B6*).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x| < \epsilon} \int_{\{|Z(x,0)|>R\}} \mid Z(x, 0) \mid^2 dP = 0$$

As before we use (B6*) to obtain

$$\lim_{R \to \infty} \sup_k \int_{\{|Z_k|>R\}} \mid Z_k \mid^2 dP = 0 \tag{5.15}$$

and, as before, (B6*) is implied by (3.5) or (3.6) with $Z(x)$ considered as a vector of course.

ASSUMPTION (B7*). There exist positive numbers ϵ, c_0, and K_1 such that, for all $c \leqq c_0$ and all x satisfying $c < \mid x - \theta \mid < \epsilon$,

$$[x - \theta, (M(x, -c) - M(x, c))/c] > K_1 \mid x - \theta \mid^2$$

Let $\delta(x, \theta) = \epsilon_x \mid x - \theta \mid^2$

ASSUMPTION (B8*). There exist positive numbers c_0, ρ, and R such that for all $c \leqq c_0$

$$\sup_{|x-\theta| \leqq c} \epsilon_x \leqq Rc^\rho$$

As in the paragraph preceding Theorem 5 let $B = PDP^{-1}$ and let $((\pi_{ij}^*)) = \pi^* = P^{-1}\pi P$.

THEOREM 6. *Suppose Assumptions* (B1*) *through* (B6*) *are satisfied. Let* $AK_1 > 1/2$ *and choose* $a_n = An^{-1}$. *Let* $\{c_n\}$ *be a sequence of positive numbers satisfying* (5.9) *with* $a_n = An^{-1}$ *and the assumptions of Lemma* 5 *with* $r = 0$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1} *where* $Q = ((A^2(4Ab_i + 4Ab_j - 1)^{-1} \pi_{ij}^*))$.

PROOF. Let $\alpha_0 = \theta = 0$. (B1*) and (5.13) imply that X_n converges to 0 w.p.1 (this is a consequence of Dvoretzky's theorem [7]). Hence, an argument like that in Theorem 1′ shows that (B3*) can be strengthened so that it holds for all x. Rewriting (5.11) by using (B2*) and letting $a = 4A$ we obtain

$$X_{n+1} = (I - an^{-1}B)X_n - An^{-1}c_n^{-1}\delta_n - An^{-1}c_n^{-1}Z_n \tag{5.16}$$

Let $B_{mn} = \prod_{m+1}^{n} (I - aj^{-1}B)$. Then, iteration of (5.16) yields

$$X_{n+1} = B_{on}\, X_1 - A \sum_1^n m^{-1}c_m^{-1}B_{mn}\, \delta_m - A \sum_1^n m^{-1}c_m^{-1}B_{mn}\, Z_m \tag{5.17}$$

It is easy to verify that

$$\| B_{mn} \| = \| P^{-1}B_{mn}\, P \| = \left\| \prod_{m+1}^{n} (I - aj^{-1}D) \right\| \sim m^{ab_q} n^{-ab_q} \tag{5.18}$$

Also, letting $\pi_m = EZ_m Z_m'$, $D_{mn} = \prod_{m+1}^{n} (I - aj^{-1} D)$, and

$$Q_m = \left(\left(A^2nc_n^2 \sum_{m=1}^{n} m^{-2}c_m^{-2}m^{a(b_i+b_j)}n^{-a(b_i+b_j)}\pi_{ij}^*\right)\right),$$

and using (5.14), an argument like that in Lemma 3, and Lemma 4, observe that

$$\begin{aligned} &\lim_{n\to\infty} \left\| A^2nc_n^2 \sum_{m=1}^{n} m^{-2}c_m^{-2}B_{mn}\, \pi_m\, B_{mn} - PQP^{-1} \right\| \\ &= \lim_{n\to\infty} \left\| A^2nc_n^2 \sum_{m=1}^{n} m^{-2}c_m^{-2}D_{mn}\, P^{-1}\pi_m\, PD_{mn} - Q \right\| \\ &= \lim_{n\to\infty} \left\| A^2nc_n^2 \sum_{m=1}^{n} m^{-2}c_m^{-2}D_{mn}\, \pi^* D_{mn} - Q \right\| \\ &= \lim_{n\to\infty} \| Q_n - Q \| = 0 \end{aligned} \tag{5.19}$$

To prove Theorem 6 we now proceed as in the proof of Theorem 2 and show

$$n^{1/2}c_n B_{on} \to 0 \tag{5.20a}$$

$$n^{1/2}c_n \sum_{m=1}^{n} m^{-1}c_m^{-1}B_{mn}\, \delta_m \to 0 \quad \text{in probability} \tag{5.20b}$$

(5.20c) $An^{1/2}c_n \sum_{m=1}^{n} m^{-1}c_m^{-1}B_{mn}\, Z_m$ is asymptotically normal with mean 0 and covariance matrix PQP^{-1}.

(5.20a) and (5.20b) are proved in the same way that (4.15a) and (4.15b) in Theorem 2 are proved; the only way that B_{mn} need enter in this parallel proof is through its norm which we have calculated in (5.18). To show (5.20c) let $U_{nk} = An^{1/2}c_n k^{-1}c_k^{-1}\, B_{kn}Z_k$ and observe that by (B5*), (B6*), and (5.18) all the conditions of Lemma 6 are satisfied with $s = PQP^{-1}$. This completes the proof of Theorem 6.

THEOREM 7. *Suppose* (B1*), (B2*), (B5*), (B6*), *and* (B7*) *are satisfied with* $K_1 \leqq 4b_q$ *in* (B7*). *Let* $c_n = n^{-1/4}$ *and* $a_n = An^{-1}$ *where* A *is such that* $AK_1 > 1/4$. *Then* $n^{1/4}(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1}.

We omit the proofs of Theorem 7 and Theorem 8 below since they proceed from the proofs of Theorems 3 and 4 in the same fashion that the proof of Theorem 6 did from that of Theorem 2.

THEOREM 8. *Suppose that* (B1*), (B2*), *and* (B5*) *through* (B8*) *are satisfied with* $K_1 \leqq 4b_q$ *in* (B7*). *Let* $a_n = An^{-1}$ *where* $AK_1 > 1$ *and let* $\{c_n\}$ *satisfy the conditions of Lemma* 5 *with* $d_n \to 0$ *and with* $r = (4 + 2\rho)^{-1}$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1}.

The procedures given by (5.1) and (5.10) can be generalized if we replace $\{a_n\}$ by a sequence $\{T_n\}$ of matrices. When $\{T_n\}$ is a sequence of positive definite matrices such that, for all n, B and T_n are diagonalized by the same orthogonal matrix P, and when the smallest and largest eigenvalues of T_n, denoted by t_n^* and t_n^{**} respectively, satisfy (5.0) and (5.9) with a_n replaced by t_n^* and t_n^{**}, methods like those used in the earlier part of this section and in earlier sections can be used to study the asymptotic behavior of these procedures. Indeed, if $T_n = n^{-1} T$ where T is a positive definite matrix which is diagonalized by P, results like those proved in the earlier part of this section can be obtained by using the same methods as used in obtaining these results. When, for $A > 0$, $T = AI$, we are in the situation covered by those theorems. In studying (5.10) (the Kiefer-Wolfowitz procedure) we can, in addition, replace $\{c_n\}$ by a sequence $\{C_n\}$ of matrices; the remarks about $\{T_n\}$ are also relevant to $\{C_n\}$.

Since Examples 1 and 2 of Section 4 can be extended to their q-dimensional analogues we cannot hope to improve materially the results of Theorems 5, 6, 7, and 8 by using sequences $\{T_n\}$ which satisfy the second sentence of the preceding paragraph and which are more general than $\{a_nI\}$. However, if we knew B and π, then, by suitable choice of such $\{T_n\}$ we can, in general, obtain a limiting covariance matrix of smaller size than is obtainable by using merely $\{a_nI\}$. As an indication of this suppose that we are concerned with a two-dimensional Robbins-Monro procedure satisfying Assumptions (A1*) through (A5*) with

$$\pi = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}.$$

Letting

$$T_n = n^{-1}T = n^{-1}\begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix}$$

where both t_1b_1 and t_2b_2 are larger than 1/2, we compute the limiting covariance matrix to be

$$(5.21) \quad \sum_{m=1}^{n} m^{-2}T^2 \prod_{j=m+1}^{n} (I - j^{-1}TB)^2\pi = \begin{pmatrix} \dfrac{t_1^2}{(2t_1 b_1 - 1)} & 0 \\ 0 & \dfrac{2t_2^2}{(2t_2 b_2 - 1)} \end{pmatrix}$$

Choosing $t_1 = 1/b_1$ and $t_2 = 1/b_2$ will minimize the entries in the matrix in (5.21). Thus, if $b_1 \neq b_2$ we can do better by using $\{n^{-1}T\}$ than by using $\{An^{-1}\}$ since using $\{An^{-1}\}$ would correspond to the case where $t_1 = t_2$.

6. Concluding Remarks. In Sections 3, 4, and 5 we have restricted ourselves to sequences $\{a_n\}$ of the type $a_n = An^{-1}$. It is clear that arguments like the ones presented above can be given for cases where a_n is chosen to be something other than An^{-1} e.g., $a_n = An^{\epsilon-1}$. Due to Examples 1 and 2 however, the results of the previous sections are not likely to be improved very much by using these different sequences. Indeed, for the Robbins-Monro procedure it was shown in [5], Section 7 that under some restrictions, the Robbins-Monro procedure with $a_n = An^{-1}$ for a certain choice of A is optimal in the sense that it is asymptotically minimax for many weight functions. We may remark that this optimum property can be extended with no difficulty to the multi-dimensional Robbins-Monro procedure.

In [4] Burkholder considers somewhat more general processes than considered here in the sense that he permits $M(X_n)$ and $Z(X_n)$ to depend on n as well as X_n. With some modifications of the assumptions we have made this situation can be treated using the methods of Sections 3 and 4. Procedures given by Burkholder for locating points of inflection of a regression function and for finding the maximum of a density function can also be treated using our methods.

It is sometimes of interest to study the asymptotic behavior of $M(X_n) - \alpha$ for the Robbins-Monro procedure and of $M(X_n) - \alpha_0$ for the Kiefer-Wolfowitz procedure. It is easy to see that results about the asymptotic distribution of these quantities can be obtained from the results about the asymptotic distribution of $X_n - \theta$.

REFERENCES

[1] H. D. Block, *"On stochastic approximation,"* ONR Report.

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

[3] J. R. Blum, "Multidimensional stochastic approximation procedures," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 737–744.

[4] D. L. Burkholder, "On a class of stochastic approximation procedures," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1044–1059.

[5] K. L. Chung, "On a stochastic approximation method," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 463–483.

[6] C. Derman, "An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 529–532.

[7] A. Dvoretzky, "On stochastic approximation," *Proc. of the Third Berkeley Symposium*, Vol. 1, pp. 39–56.

[8] J. L. Hodges, Jr. and E. L. Lehmann, "Two Approximations to the Robbins-Monro Process," *Proc. of the Third Berkeley Symposium*, vol. 1, pp. 96–104.

[9] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.

[10] M. Loeve, *"Probability theory,"* New York (1955).

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

[12] J. Wolfowitz, "On stochastic approximation methods," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1151–1155.