ON A STOCHASTIC APPROXIMATION METHOD

By K. L. Chung[1]

Cornell University and Syracuse University

1. Summary. Asymptotic properties are established for the Robbins-Monro [1] procedure of stochastically solving the equation $M(x) = \alpha$. Two disjoint cases are treated in detail. The first may be called the "bounded" case, in which the assumptions we make are similar to those in the second case of Robbins and Monro. The second may be called the "quasi-linear" case which restricts $M(x)$ to lie between two straight lines with finite and nonvanishing slopes but postulates only the boundedness of the moments of $Y(x) - M(x)$ (see Sec. 2 for notations). In both cases it is shown how to choose the sequence $\{a_n\}$ in order to establish the correct order of magnitude of the moments of $x_n - \theta$. Asymptotic normality of $a_n^{1/2}(x_n - \theta)$ is proved in both cases under a further assumption. The case of a linear $M(x)$ is discussed to point up other possibilities. The statistical significance of our results is sketched.

2. Introduction. Let $M(x)$ be a fixed but unknown function and α a given (known) constant such that

$$M(x) = \alpha \tag{1}$$

has a unique (unknown) root $x = \theta$. Suppose that to each value x corresponds a random variable $Y = Y(x)$ with distribution function $\Pr[Y(x) \leqq y] = H(y \mid x)$, such that

$$M(x) = \int_{-\infty}^{\infty} y \, dH(y \mid x)$$

is the mathematical expectation of Y for the given x.

The Robbins-Monro procedure is defined as follows. Let $\{a_n\}$, $n \geqq 1$, be a fixed sequence of positive constants such that

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

We define a nonstationary Markov chain $\{x_n\}$ by taking x_1 to be an arbitrary constant and setting recursively

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

where y_n is a random variable whose distribution function, for given $x_1, \cdots, x_n, y_1, \cdots, y_{n-1}$, is $H(y \mid x_n)$. The moments of $x_n - \theta$ will be denoted as follows:

$$b_n^{(r)} = E[(x_n - \theta)^r] \qquad b_n^{(2)} = b_n$$
$$\beta_n^{(r)} = E[| x_n - \theta |^r].$$

Received 8/24/53.

[1] Research under contract with the Office of Naval Research.

Under certain assumptions regarding the nature of the functions $M(\cdot)$ and $H(\cdot \mid \cdot)$, to be specified in a moment, Robbins and Monro showed that $b_n \to 0$, namely that x_n tends to θ in mean of order 2, for every sequence $\{a_n\}$ satisfying (1), thus providing a stochastic solution of the equation (1).

Robbins and Monro made two overall[2] assumptions: namely that

(A) $M(x) \lesseqgtr \alpha$ according as $x \lesseqgtr \theta$ (our Assumption (0) in Sec. 4); and

(B) the random variables $Y(x)$ are uniformly (in x) bounded with probability one (our Assumption (III) in Sec. 4).

Furthermore they needed either one of the following two sets of conditions:

(i) $\inf_{x \neq \theta} \mid M(x) - \alpha \mid \geqq \delta > 0$;

(iia) $M(x)$ is nondecreasing and $M'(\theta) > 0$.

Wolfowitz [2] weakened the overall assumptions by keeping (A) but assuming only that $M(x)$ is bounded and that $Y(x)$ has a bounded (in x) variance. Furthermore he needed either the condition (i) above or

(iib) $M(x)$ is strictly increasing in a neighborhood of $x = \theta$ and is bounded away from α outside every such neighborhood.

Under these assumptions he proved that x_n tends to θ in probability. Later several authors proved that x_n tends to θ with probability one, under conditions yet unknown to us. We are not concerned with this question here. Very recently L. Schmetterer [7] gave some upper bounds for b_n, under assumptions which are essentially those of our second case; see footnote 5.

In this paper we shall study the finer properties of the process $\{x_n\}$, especially with regard to the moments of $x_n - \theta$ and the limiting (nondegenerate) distribution of $x_n - \theta$, suitably normalized. We shall not deal with the case (i), but shall treat two different cases.

Our first case (Sec. 4) requires a set of conditions which is similar[3] to that in the second case of Robbins and Monro. In addition to their overall assumption we assume that

(ii) $M'(\theta) > 0$ and $M(x)$ is bounded away from α outside every neighborhood of $x = \theta$; (assumptions (I) and (II) in Sec. 4).

With these assumptions and a suitable choice of $\{a_n\}$ we can obtain upper bounds for the absolute moments $\beta_n^{(r)} = E[\mid x_n - \theta \mid^r]$ (Theorem 1). An important consequence, much used thereafter, is given in Theorem 2. In order to obtain lower bounds for $\beta_n^{(r)}$ (Theorem 3) we need the new assumption ((IV) in Sec. 4) that the variance of $Y(x)$ is bounded below uniformly in x. It is interesting to note that our choice of a_n is $a_n = 1/n^{1-\epsilon}$, with an ϵ which has to be greater than some positive constant depending on $M(\cdot)$ and $H(\cdot \mid \cdot)$, but fortunately always compatible with $\epsilon < \frac{1}{2}$. The upper and lower bounds for $\beta_n^{(r)}$ are at first not of the same order of magnitude, but they lead to Theorem 4 which in turn sharpens the bounds to their correct order (Theorem 5).

The question arises what happens if an intransigent statistician refuses to use our prescribed (range of) ϵ and insists on using the simpler $a_n = c/n$, as suggested

---

[2] This adjective need not be taken literally.

[3] In fact, Theorems 1 and 2 below are proved under weaker conditions than theirs.

by Robbins and Monro. Our answer is that our method still leads to some estimate, in fact that $\beta_n^{(r)}$ has an order between $1/n$ and $1/(\log n)^{c'}$ if c is chosen sufficiently large (where c' depends on c and tends to infinity with it), but we do not know even if it has a definite asymptotic order.

Returning to the moments, it seems vain hope to obtain a precise asymptotic formula for them without further hypotheses. We shall content ourselves with the "obvious" by strengthening our last assumption to requiring that $Y(x)$ has a constant variance independent[4] of x. With this added force (Assumption (V) in Sec. 4) our method works smoothly and we finish with an asymptotically normal distribution for $a_n^{1/2}(x_n - \theta)$ (Theorem 6).

We now turn to our second case (Sec. 5) which is disjoint from the first one. Here assumption (B) is replaced by the weaker one that

(C) $Y(x) - M(x)$ has bounded (in x) moments up to order p, where p is an even integer.

Assumption (A) is kept but (ii) is replaced by

(iii) $M'(\theta) > 0$; $M(x)$ is bounded in any finite interval; and

$$0 < \varliminf_{|x| \to \infty} \frac{M(x)}{x} \leqq \varlimsup_{|x| \to \infty} \frac{M(x)}{x} < \infty.$$

Admittedly this last condition is pretty strong and our only excuse is that of inability, and an invitation to weaken it by a better method. Our choice[5] of $\{a_n\}$ is now $a_n = c/n$ with $c > 1/2K$, where $K = \inf_{x \neq \theta}[M(x) - \alpha)/(x - \theta] > 0$. With the assumptions (C) and (iii) we can prove that $\beta_n^{(r)}$ is at most of the order $n^{-r/2}$ for $0 \leqq r \leqq p$ (Theorem 7). If $p \geqq 6$ in (C) we can prove that $\beta_n^{(r)}$ is exactly of the order $n^{-r/2}$ for $2 \leqq r \leqq p$ (Theorem 8). If we can take $p = \infty$ in (C) and further if $Y(x)$ has a constant variance independent of x, then just as in the first case[4] we can prove asymptotic normality of $a_n^{1/2}(x_n - \theta)$ (Theorem 9).

The method we use in discussing both cases is elementary and depends on some simple analytical lemmas which we collect in Section 3.

In Section 6 we discuss by a different method the case where $M(\cdot)$ is a linear function. Under the assumption that $Y(x) - M(x)$ has a fixed distribution function independent of x, the problem is reduced to a classical problem in probability theory. Various easy conclusions are then drawn which show that $x_n - \theta$ may have an asymptotic distribution which is stable but not normal, or it may have no asymptotic distribution whatsoever. The main interest of these observations is to serve as a foil to our previous results.

In Section 7 we discuss briefly the statistical implications of our results.

---

[4] The following weaker assumption suffices. The variance $E[(Y(x) - M(x))^2]$ as a function of x is continuous and nonvanishing at $x = \theta$. Since by (III) it is bounded in x and by Theorem 1 below $x_n \to \theta$ in probability it follows that $E[(y_n - M(x_n))^2] \to \sigma^2$ as $n \to \infty$ so that (4.15) implies that $e_n \to \sigma^2$ and this is all we need. This weaker assumption is culled from an unpublished MS. by J. L. Hodges, Jr. and E. L. Lehmann.

[5] Schmetterer [7] gave upper bounds for b_n for $a_n = n^{-\epsilon}$, $0 < \epsilon < \frac{1}{2}$; $a_n = n^{-1}$ (in this case the order of magnitude of the upper bound depends on K); and $a_n = cn^{-1}$ with $c > (2K)^{-1}$. The last case is covered by Theorem 7 below.

The formulas in each of the 7 sections are numbered separately. A formula in the same section is referred to simply by its number; a formula in a previous section is referred to by prefixing the section number, for example, (4.6) is formula (6) in Section 4.

3. Lemmas. In this section we state and prove our principal mathematical tools. They seem to be new and can be further elaborated, but we state only what will be needed. The c's are numerical constants.

LEMMA 1. *Suppose that* $\{b_n\}$, $n \geqq 1$, *is a sequence of real numbers such that for* $n \geqq n_0$,

$$b_{n+1} \leqq \left(1 - \frac{c}{n}\right) b_n + \frac{c_1}{n^{p+1}} \tag{1}$$

where $c > p > 0$, $c_1 > 0$. *Then*

$$b_n \leqq \frac{c_1}{c - p} \frac{1}{n^p} + 0\left(\frac{1}{n^{p+1}} + \frac{1}{n^c}\right). \tag{2}$$

PROOF. There is no loss of generality if we take $n_0 = 1$. We have

$$\frac{1}{(n+1)^p} - \left(1 - \frac{c}{n}\right)\frac{1}{n^p} = \frac{c}{n^{p+1}} - \left(\frac{1}{n^p} - \frac{1}{(n+1)^p}\right) = \frac{c-p}{n^{p+1}} + 0\left(\frac{1}{n^{p+2}}\right).$$

Hence for some $c_2 > 0$,

$$\frac{c_1}{n^{p+1}} \leqq \frac{c_1}{c-p}\left[\frac{1}{(n+1)^p} - \left(1 - \frac{c}{n}\right)\frac{1}{n^p}\right] + \frac{c_2}{n^{p+2}}. \tag{3}$$

Similarly but more roughly, for some $c_3 > 0$,

$$\frac{c_2}{n^{p+2}} \leqq c_3\left[\frac{1}{(n+1)^{p+1}} - \left(1 - \frac{c}{n}\right)\frac{1}{n^{p+1}}\right].$$

Using these inequalities in (1) and rearranging terms, we obtain

$$b_{n+1} - \frac{c_1}{c-p}\frac{1}{(n+1)^p} - \frac{c_3}{(n+1)^{p+1}} \leqq \left(1 - \frac{c}{n}\right)\left[b_n - \frac{c_1}{c-p}\frac{1}{n^p} - \frac{c_3}{n^{p+1}}\right].$$

Let the quantity on the left side be denoted by b'_{n+1}. If for some $n > c$ we have $b'_n \leqq 0$, then this is true for all subsequent n, namely

$$b_n \leqq \frac{c_1}{c-p}\frac{1}{n^p} + \frac{c_3}{n^{p+1}}.$$

Otherwise for every $n > n_1 > c$ we have

$$0 < b'_n \leqq b'_{n_1} \prod_{m=n_1}^{n-1}\left(1 - \frac{c}{m}\right) = O\left(\frac{1}{n^c}\right).$$

In either case (2) is true.

LEMMA 2. *Suppose that* $\{b_n\}$, $n > 1$, *is a sequence of real numbers such that for* $n \geqq n_0$,

$$b_{n+1} \geqq \left(1 - \frac{c}{n}\right) b_n + \frac{c_1}{n^{p+1}}$$

where $c > p > 0$, $c_1 > 0$. *Then*

$$b_n \geqq \frac{c_1}{c - p}\frac{1}{n^p} + 0\left(\frac{1}{n^{p+1}} + \frac{1}{n^c}\right).$$

PROOF. The proof is entirely similar to that of Lemma 1. The point is that (3) may be changed into

$$\frac{c_1}{n^{p+1}} \geqq \frac{c_1}{c - p}\left[\frac{1}{(n+1)^p} - \left(1 - \frac{c}{n}\right)\frac{1}{n^p}\right].$$

LEMMA 3. *Suppose that* $\{b_n\}$, $n \geqq 1$, *is a sequence of real numbers such that for* $n \geqq n_0$,

$$b_{n+1} \geqq \left(1 - \frac{c}{n^s}\right) b_n + \frac{c_1}{n^t} \tag{4}$$

where $0 < s < 1$, $s < t$, $c > 0$, $c_1 > 0$. *Then*

$$\varliminf_{n\to\infty} n^{t-s} b_n \geqq \frac{c_1}{c}.$$

PROOF. We may take $n_0 = 1$. We have

$$\frac{1}{(n+1)^{t-s}} - \left(1 - \frac{c}{n^s}\right)\frac{1}{n^{t-s}} \leqq \frac{c}{n^t}.$$

Hence we have

$$\frac{c_1}{n^t} \geqq \frac{c_1}{c}\left[\frac{1}{(n+1)^{t-s}} - \left(1 - \frac{c}{n^s}\right)\frac{1}{n^{t-s}}\right].$$

Using this inequality in (4), we obtain

$$b_{n+1} - \frac{c_1}{c(n+1)^{t-s}} \geqq \left(1 - \frac{c}{n^s}\right)\left(b_n - \frac{c_1}{cn^{t-s}}\right).$$

If for some $n > c^{1/s}$ we have $b_n \geqq c_1/cn^{t-s}$, then this is true for all subsequent n. Otherwise for every $n > n_1 > c^{1/s}$ we have

$$\left|b_n - \frac{c_1}{cn^{t-s}}\right| \leqq \left|b_{n_1} - \frac{c_1}{cn_1^{t-s}}\right| \prod_{m=n_1}^{n-1}\left(1 - \frac{c}{m^s}\right) = O\left(\frac{1}{n^q}\right)$$

for every $q > 0$. The lemma follows in either case.

LEMMA 4. *Suppose that* $\{b_n\}$, $n \geqq 1$, *is a sequence of real numbers such that for* $n \geqq n_0$,

$$b_{n+1} \leqq \left(1 - \frac{c_n}{n^s}\right) b_n + \frac{c'}{n^t} \tag{5}$$

where $0 < s < 1$, $s < t$, $c_n \geqq c > 0$, $c' > 0$. *Then*

$$\varlimsup_{n\to\infty} n^{t-s} b_n \leqq \frac{c'}{c}.$$

PROOF. We have, if c'' is any number $> c'$,

$$\frac{c'}{n^t} \leqq \frac{c''}{c_n}\left[\frac{1}{(n+1)^{t-s}} - \left(1 - \frac{c_n}{n^s}\right)\frac{1}{n^{t-s}}\right] \leqq \frac{c''}{c}\left[\frac{1}{(n+1)^{t-s}} - \left(1 - \frac{c_n}{n^s}\right)\frac{1}{n^{t-s}}\right]$$

for all $n > n_0(c'')$. Using this inequality in (5), we obtain

$$b_{n+1} - \frac{c''}{c}\frac{1}{(n+1)^{t-s}} \leqq \left(1 - \frac{c_n}{n^s}\right)\left(b_n - \frac{c''}{c}\frac{1}{n^{t-s}}\right).$$

The rest follows as in the proof of Lemma 3.

4. First case. In this section we treat a case which is essentially the second, "more interesting," case of Robbins and Monro. The various assumptions needed will be listed below. Not all of them are used in every theorem we shall prove.

In the following, $K_0, K_1, \cdots$ are positive constants which depend only on the nature of the functions $M(\cdot)$ and $H(\cdot \mid \cdot)$, and $K_7', K_9', \cdots$ are positive constants which depend moreover on the choice of $\{a_n\}$. If these constants happen to depend also on some new parameter δ (say), this dependence will be indicated by the usual parentheses. They are numbered in order of appearance. We use also the customary O and o notations, as we have already done in Section 3, it being understood that the constants involved may depend on $M(\cdot)$, $H(\cdot \mid \cdot)$, and $\{a_n\}$. The initial value x_1 of the process is supposed to be a fixed constant or at least a random variable which is bounded by a fixed constant with probability one.

ASSUMPTION (0). $M(\cdot)$ is a Borel measurable function; $M(\theta) = \alpha$, and $(x - \theta)(M(x) - \alpha) > 0$ for all $x \neq \theta$. This assumption will be used throughout the paper and will not always be explicitly mentioned.

ASSUMPTION (I). We have $M'(\theta) > 0$, namely, as $x - \theta \to 0$,

$$M(x) = \alpha + \alpha_1(x - \theta) + o(|x - \theta|) \qquad 0 < \alpha_1 < \infty.$$

ASSUMPTION (II). For every $\delta > 0$ we have

$$\inf_{|x-\theta|>\delta} |M(x) - \alpha| = K_0(\delta) > 0.$$

ASSUMPTION (III). For all x we have

$$\Pr(|Y(x) - \alpha| \leqq K_1) = 1.$$

ASSUMPTION (IV). For all x, we have

$$E[(Y(x) - M(x)]^2 \geqq K_2 > 0.$$

In this section we set for $n \geqq 1$

(1) $$a_n = 1/n^{1-\epsilon}, \qquad 0 < \epsilon < \tfrac{1}{2},$$

where ϵ is to be chosen later.

We record some simple consequences of the assumptions and the choice of $\{a_n\}$.

First, it follows from (III) that

$$\text{(IIIa)} \qquad | M(x) - \alpha | \leqq K_1$$

and that all moments of $Y(x)$ are bounded by constants K which depend on the order of the moment but not on x.

The Robbins-Monro procedure yields

$$x_n = x_1 + \sum_{k=1}^{n-1} a_k(\alpha - y_k).$$

From (III) we have with probability one

$$\text{(2)} \qquad | x_n - \theta | \leqq | x_1 - \theta | + K_1 \sum_{k=1}^{n-1} \frac{1}{k^{1-\epsilon}} \leqq \frac{K_3}{\epsilon} n^{\epsilon}.$$

From (I) it follows that to every γ, $0 < \gamma < 1$, there corresponds a $\delta = \delta(\gamma)$ such that

$$| M(x_n) - \alpha | \geqq \gamma\alpha_1 | x_n - \theta | \qquad \text{if } | x_n - \theta | \leqq \delta.$$

From (II) and (2), it follows that the conditional probability is one that

$$| M(x_n) - \alpha | \geqq [K_0(\delta)\epsilon/K_3 n^{\epsilon}] | x_n - \theta | \quad \text{if } | x_n - \theta | > \delta.$$

Namely, the inequality for $| M(x_n) - \alpha |$ holds almost everywhere on the set $| x_n - \theta | > \delta$. Together we have with probability one

$$\text{(3)} \qquad | M(x_n) - \alpha | \geqq K_4 \epsilon n^{-\epsilon} | x_n - \theta |.$$

This constant K_4 is of extreme importance in the analysis below. Of course, as given in (3), it does not depend on n. However, for $n \to \infty$ we can determine its asymptotic value as follows. For a given γ, $0 < \gamma < 1$, let $\delta_0(\gamma)$ be the supremum of all δ such that

$$| M(x) - \alpha | \geqq \gamma\alpha_1 | x - \theta | \quad \text{for } | x - \theta | \leqq \delta.$$

Then for $n \geqq n_0(\gamma, \delta', x_1 - \theta)$ the K_4 in (3) may be taken to be $K_0(\delta_0(\gamma))/K_1 - \delta'$ where $\delta' > 0$ is arbitrarily small. If $\lim_{\gamma \to 0} \delta_0(\gamma) = \delta_0$, the K_4 in (3) may be taken asymptotically as $n \to \infty$ to be $K_0(\delta_0)/K_1$.

Finally, we note that from (I) and (IIIa) it follows that

$$\text{(4)} \qquad | M(x) - \alpha | \leqq K_5 | x - \theta |.$$

THEOREM 1. *Suppose that the assumptions (I), (II) and (III) are satisfied. If* $a_n = n^{-(1-\epsilon)}$, with $1/2(1 + K_4) < \epsilon < \frac{1}{2}$, *then for each real* $r > 0$, we have

$$\text{(5)} \qquad \beta_n^{(r)} \leqq K_6'(r) n^{-(r/2)(1-2\epsilon)}$$

PROOF. We write as in [1]

$$d_n = E[(x_n - \theta)(M(x_n) - \alpha)], \qquad e_n = E[(y_n - \alpha)^2].$$

The proof is divided into three stages: (i) $r = 2$, (ii) all even integer r, (iii) all real r. It is obvious that (iii) follows from (ii) by Lyapunov's (or Hölder's) inequality.

(i) The following equation is given in [1]:

$$b_{n+1} = b_n - 2a_nd_n + e_na_n^2. \tag{6}$$

By (3), we have, using (0), $d_n \geqq K_4\epsilon n^{-\epsilon}b_n$. By (III), we have $e_n \leqq K_7$. Using these in (6), we obtain

$$b_{n+1} \leqq \left(1 - \frac{2K_4\epsilon}{n^\epsilon}\frac{1}{n^{1-\epsilon}}\right)b_n + \frac{K_7}{n^{2-2\epsilon}} = \left(1 - \frac{2K_4\epsilon}{n}\right)b_n + \frac{K_7}{n^{1+\lambda}} \tag{7}$$

where we have put, once for all, $0 < \lambda = 1 - 2\epsilon < 1$. If $\epsilon > 1/2(1 + K_4)$, then $2K_4\epsilon > \lambda$. Applying Lemma 1 to (7) with $c = 2K_4\epsilon$, $c_1 = K_7$, $p = \lambda$, we obtain

$$b_n \leqq \frac{K_7}{2(1 + K_4)\epsilon - 1}\frac{1}{n^\lambda} + O\left(\frac{1}{n^{2K_4\epsilon}} + \frac{1}{n^{\lambda+1}}\right).$$

Thus (5) is proved for $r = 2$.

(ii) We use induction on even r. Suppose then that r is even and that

$$b_n^{(t)} \leqq K_6'(t)n^{-(t/2)(1-2\epsilon)} \qquad 2 \leqq t \leqq r - 2. \tag{5 bis}$$

Recalling that $x_{n+1} - \theta = x_n - \theta + a_n(\alpha - y_n)$, we have, as a generalization of (6), for all integer $r \geqq 1$,

$$\begin{aligned} b_{n+1}^{(r)} &= E\left[\int_{-\infty}^{\infty}\{x_n - \theta - a_n(y - \alpha)\}^r\,dH(y \mid x_n)\right] \\ &= b_n^{(r)} - ra_nE[(x_n - \theta)^{r-1}(M(x_n) - \alpha)] + \sum_{t=2}^{r}(-1)^t\binom{r}{t}J_t \end{aligned} \tag{8}$$

where

$$J_t = J_t(r) = a_n^tE[(x_n - \theta)^{r-t}(y_n - \alpha)^t]. \tag{9}$$

By (III) we have $|J_t| \leqq K_8n^{-t(1-\epsilon)}b_n^{(r-t)}$. Hence by the induction hypothesis (5 bis) we have $|J_t| \leqq K_9'n^{-(t+r\lambda)/2}$. Therefore we have

$$\left|\sum_{t=2}^{r}(-1)^t\binom{r}{t}J_t\right| \leqq \frac{K_{10}'}{n^{1+r\lambda/2}}.$$

By (3), we have since r is even (using Assumption (0))

$$E[(x_n - \theta)^{r-1}(M(x_n) - \alpha)] \geqq K_4\epsilon n^{-\epsilon}b_n^{(r)}.$$

Using these inequalities in (8), we obtain

$$b_{n+1}^{(r)} \leqq (1 - rK_4\epsilon/n)b_n^{(r)} + K_{10}'n^{-(1+r\lambda/2)}. \tag{10}$$

Our choice of ϵ makes $rK_4\epsilon > r\lambda/2$. Applying Lemma 1 to (10) with $c = rK_4\epsilon$, $c_1 = K_{10}'$, $p = r\lambda/2$, we obtain

$$b_n^{(r)} \leqq \frac{2K_{10}'}{r[2(1 + K_4)\epsilon - 1]}\frac{1}{n^{r\lambda/2}} + O\left(\frac{1}{n^{rK_4\epsilon}} + \frac{1}{n^{1+r\lambda/2}}\right).$$

This completes the induction, thus proving (5) for all even integers r.

THEOREM 2. *Under the same hypotheses as in Theorem* 1, *we have, for every* $\delta > 0, r \geqq 0$ and $q > 0$,

$$\int_{|x_n-\theta|>\delta} | x_n - \theta |^r \, dPr = O(n^q).$$

PROOF. By (2) and Chebychev's inequality, we have for every positive integer s,

$$\int_{|x_n-\theta|>\delta} | x_n - \theta |^r \, dPr \leqq K'_{11} n^{r\epsilon} \Pr (| x_n - \theta | > \delta) \leqq K'_{11} n^{r\epsilon} \beta_n^{(2s)}/\delta^{2s} = O(n^{r\epsilon - s\lambda}).$$

It remains only to choose s so that $s\lambda - r\epsilon > q$.

THEOREM 3. *Suppose Assumptions (I), (III) and (IV) are satisfied. If* $a_n = n^{-(1-\epsilon)}$ *with* $0 < \epsilon < \frac{1}{2}$, *then for each integer* $r \geqq 2$, *we have*

$$\underline{\lim}_{n\to\infty} n^{(1-\epsilon)r/2} \beta_n^{(r)} \geqq K_{12}. \tag{11}$$

PROOF. By Lyapunov's inequality we need only prove (11) for $r = 2$. By (4), we have $d_n \leqq K_5 b_n$. By (IV), we have $e_n \geqq K_2$. Using these in (6), we obtain

$$b_{n+1} \geqq (1 - 2K_5 n^{-(1-\epsilon)}) b_n + K_2 n^{-(2-2\epsilon)}.$$

Applying Lemma 3 with $s = 1 - \epsilon$, $t = 2 - 2\epsilon$, $c = 2K_5$, $c_1 = K_2$, we obtain

$$\underline{\lim}_{n\to\infty} n^{1-\epsilon} b_n \geqq K_2/2K_5.$$

REMARK. If in Theorem 3 we choose $a_n = c/n$ with a sufficiently large c, we obtain $b_n \geqq K_{13}(c)/n$. We do not need this result in this section; but see Section 5.

THEOREM 4. *Suppose that the Assumptions (I) to (IV) are satisfied. If* $a_n = n^{-(1-\epsilon)}$ *with* $1/2(1 + K_4) < \epsilon < \frac{1}{2}$, *we have*

$$\lim_{n\to\infty} (d_n/b_n) = \alpha_1 = M'(\theta).$$

PROOF. Given any small $\eta > 0$, there exists a $\delta = \delta(\eta)$ such that

$$| M(x) - \alpha - \alpha_1(x - \theta) | \leqq \eta \, | x - \theta | \qquad \text{for } | x - \theta | \leqq \delta.$$

Hence

$$\begin{aligned}
&\int_{|x_n-\theta|<\delta} (x_n - \theta)(M(x_n) - \alpha) \, dPr \\
&\qquad = \alpha_1 \int_{|x_n-\theta|>\delta} (x_n - \theta)^2 \, dPr + \eta' \int_{|x_n-\theta|<\delta} (x_n - \theta)^2 \, dPr \\
&\qquad = \alpha_1 b_n - \alpha_1 \int_{|x_n-\theta|>\delta} (x_n - \theta)^2 \, dPr + \eta'' b_n
\end{aligned}$$

where $|\eta''| \leqq |\eta'| \leqq \eta$. On the other hand, by (IIIa) and Theorem 2

$$\int_{|x_n-\theta|>\delta} (x_n - \theta)(M(x_n) - \alpha)\, dPr \leqq K_1 \int_{|x_n-\theta|>\delta} |x_n - \theta|\, dPr = O(n^q)$$

for every $q > 0$. It follows from Theorem 3 that

$$\int_{|x_n-\theta|>\delta} (x_n - \theta)(M(x_n) - \alpha)\, dPr = o(b_n).$$

Combining these results, we obtain $d_n = \alpha_1 b_n + \eta'' b_n + o(b_n)$. This proves Theorem 4 since η is arbitrarily small.

THEOREM 5. *Under the same assumptions as in Theorem 4, we have*

(12) $$K_2/2\alpha_1 \leqq \varliminf_{n\to\infty} n^{1-\epsilon} b_n \leqq \varlimsup_{n\to\infty} n^{1-\epsilon} b_n \leqq K_7/2\alpha_1 .$$

Moreover for every integer $r \geqq 2$,

(13) $$0 < K_{14}(r) \leqq \varliminf_{n\to\infty} n^{(1-\epsilon)r/2} \beta_n^{(r)} \leqq \varlimsup_{n\to\infty} n^{(1-\epsilon)r/2} \beta_n^{(r)} \leqq K_{15}(r) < \infty.$$

PROOF. By Theorem 4, for any small $\delta > 0$, if $n > n_0(\delta)$ we have

$$(\alpha_1 - \delta) b_n \leqq d_n \leqq (\alpha_1 + \delta) b_n .$$

Recalling that we have $K_2 \leqq e_n \leqq K_7$, we have for $n > n_0(\delta)$ on the one hand,

$$b_{n+1} \leqq (1 - 2(\alpha_1 - \delta) n^{-(1-\epsilon)}) b_n + K_7 n^{-(2-2\epsilon)};$$

and on the other hand

$$b_{n+1} \geqq (1 - 2(\alpha_1 + \delta) n^{-(1-\epsilon)}) b_n + K_2 n^{-(2-2\epsilon)}.$$

Lemmas 3 and 4 yield

$$K_2/2(\alpha_1 + \delta) \leqq \varliminf_{n\to\infty} n^{1-\epsilon} b_n \leqq \varlimsup_{n\to\infty} n^{1-\epsilon} b_n \leqq K_7/2(\alpha_1 - \delta).$$

This proves (12) since δ is arbitrary. The left half of (13) follows now from Lyapunov's inequality. The proof of the right half of (13) is entirely similar to part (ii) of the proof of Theorem 1, modified according to the proof of Theorem 3. We find that, if r is even and if $\lim_{n\to\infty} n^{(t/2)(1-\epsilon)} b_n^{(t)} \leqq K_{16}$ for $2 \leqq t \leqq r - 2$, then for all $n > n_1(\delta, \epsilon)$ we have

$$b_{n+1}^{(r)} \leqq (1 - r(\alpha_1 - \delta) n^{-(1-\epsilon)}) b_n^{(r)} + K_{17} n^{-(1-\epsilon)(1+r/2)}$$

where K_{17} does not depend on ϵ. This explains the reason why the constants $K_{14}(r)$ and $K_{15}(r)$ do not depend on ϵ. Naturally, they are equivalent to

$$K'_{18}(r) n^{-(1-\epsilon)r/2} \leqq \beta_n^{(r)} \leqq K'_{19}(r) n^{-(1-\epsilon)r/2}$$

since in this form ϵ enters through the absorbed error terms.

It is possible to give explicit bounds for the constants $K_{14}(r)$ and $K_{15}(r)$ by proceeding inductively. However, it seems more interesting to study a case in which there is an asymptotic formula for b_n, in other words, where the limits in (13) may be replaced by a unique limit. For this purpose we shall make a further assumption which strengthens (IV).

ASSUMPTION (V). For all x

$$E[(Y(x) - M(x))^2] = \sigma^2 > 0$$

where σ^2 is a constant which does not depend on x. This assumption states that all the distributions $H(y \mid x)$ have the same (positive) variance. Even the much stronger assumption, that there is a fixed distribution $H_0(y)$ such that $H(y \mid x) = H_0(y - M(x))$, seems reasonable in many applications.

THEOREM 6. *Suppose that the Assumptions (I), (II), (III) and (V) are satisfied. If* $a_n = n^{-(1-\epsilon)}$ *with* $1/2(1 + K_4) < \epsilon < \frac{1}{2}$, *then we have for every integer* $r \geqq 1$

$$\lim_{n\to\infty} n^{(1-\epsilon)r/2} b_n^{(r)} = \begin{cases} 0 & \text{if } r = 2s - 1 \\ (\sigma^2/2\alpha_1)^s(2s-1)(2s-3)\ldots\cdot 3\cdot 1 & \text{if } r = 2s. \end{cases}$$

Consequently the random variable $n^{(1-\epsilon)/2}(x_n - \theta)$ *tends in distribution to the normal distribution with mean* 0 *and variance* $(\sigma^2)/(2\alpha_1)$.

PROOF. (i) $r = 1$. We have

$$b_{n+1}^{(1)} = E[(x_n - \theta) - a_n(y_n - \alpha)] = b_n^{(1)} - a_nE[M(x_n) - \alpha]. \tag{14}$$

By (I), there exists a $\delta = \delta(\alpha_1/2)$ such that

$$\int_{|x_n-\theta|<\delta} (M(x_n) - \alpha)\, dPr = (\alpha_1 + \eta') \int_{|x_n-\theta|<\delta} (x_n - \theta)\, dPr$$
$$= (\alpha_1 + \eta')b_n^{(1)} - (\alpha_1 + \eta') \int_{|x_n-\theta|>\delta} (x_n - \theta)\, dPr$$

where $|\eta'| \leqq \alpha_1/2$. Hence by Theorem 2, for every $q > 0$ we have

$$\int_{|x_n-\theta|<\delta} (M(x_n) - \alpha)\, dPr = (\alpha_1 + \eta')b_n^{(1)} + O(n^q).$$

On the other hand we have by Theorem 2,

$$\left| \int_{|x_n-\theta|>\delta} (M(x_n) - \alpha)\, dPr \right| \leqq K_1 \Pr(|x_n - \theta| > \delta) = O(n^q).$$

Together we have

$$E[M(x_n) - \alpha] \geqq \alpha(n)b_n^{(1)} + O(n^{-q})$$

where $\alpha(n) = \alpha_1 - (\alpha_1/2) \operatorname{sgn} b_n^{(1)} \geqq \alpha_1/2$. Using this inequality in (14), we obtain

$$b_{n+1}^{(1)} \leqq (1 - \alpha(n)n^{-(1-\epsilon)})b_n^{(1)} + K'_{20}n^{-q}.$$

Applying Lemma 4 with $s = 1 - \epsilon$, $t = q$, $c_n = \alpha(n)$, $c = \alpha_1/2$, $c' = K'_{20}$ we obtain $b_n^{(1)} = O(n^{-p})$ for every $p > 0$.

(ii) $r = 2$. We have, by (V)

$$e_n = E[(y_n - M(x_n))^2 + (M(x_n) - \alpha)^2] = \sigma^2 + E[(M(x_n) - \alpha)^2]. \tag{15}$$

By (4), $E[(M(x_n) - \alpha)^2] \leqq K_5^2 b_n$. Hence given any $\delta > 0$, there exists $n_0 = n_0(\delta)$ such that if $n > n_0$,

$$\sigma^2 - \delta \leqq e_n \leqq \sigma^2 + \delta.$$

Moreover by Theorem 4, this n_0 may be also chosen so that if $n > n_0$

$$(\alpha_1 - \delta)n_n \leqq d_n \leqq (\alpha_1 + \delta)b_n .$$

Using these in (6), we obtain

$$\left(1 - \frac{2(\alpha_1 + \delta)}{n^{1-\epsilon}}\right) b_n + \frac{\sigma^2 - \delta}{n^{2-2\epsilon}} \leqq b_{n+1} \leqq \left(1 - \frac{2(\alpha_1 - \delta)}{n^{1-\epsilon}}\right) b_n + \frac{\sigma^2 + \delta}{n^{2-2\epsilon}} .$$

Applying Lemmas 3 and 4 we obtain

$$\frac{\sigma^2 - \delta}{2(\alpha_1 + \delta)} \leqq \varliminf_{n \to \infty} n^{1-\epsilon} b_n \leqq \varlimsup_{n \to \infty} n^{1-\epsilon} b_n \leqq \frac{\sigma^2 + \delta}{2(\alpha_1 - \delta)} .$$

Since δ is arbitrary it follows that $\lim_{n\to\infty} n^{1-\epsilon} b_n = \sigma^2/2\alpha_1$.

(iii) Induction. Let r be an integer >2. It follows from (I) that, for every $\eta > 0$, there is a $\delta = \delta(\eta)$ such that

$$\begin{aligned} E[(x_n - \theta)^{r-1}(M(x_n) - \alpha)] &= (\alpha_1 + \eta') \int_{|x_n - \theta| \leqq \delta} (x_n - \theta)^r \, dPr + O(n^{-q}) \\ &= (\alpha_1 + \eta') b_n^{(r)} + O(n^{-q}) \end{aligned}$$

where $|\eta'| \leqq \eta$ and $q > 0$ is arbitrary, by Theorem 2. Moreover, by (V), (4) and Theorem 5

$$\begin{aligned} E[(x_n - \theta)^{r-2}(y_n - \alpha)^2] &= E[\{\sigma^2 + (M(x_n) - \alpha)^2\}(x_n - \theta)^{r-2}] \\ &= \sigma^2 b_n^{(r-2)} + O(\beta_n^{(r)}) = \sigma^2 b_n^{(r-2)} + O(n^{-(1-\epsilon)r/2}). \end{aligned}$$

Hence by (9),

$$J_2 = \sigma^2 b_n^{(r-2)}/n^{-2(1-\epsilon)} + O(n^{-(2+r/2)(1-\epsilon)})$$

and if $3 \leqq t \leqq r$

$$|J_t| = O(n^{-t(1-\epsilon)}\beta_n^{(r-t)}) = O(n^{-(r+t)(1-\epsilon)/2}) = O(n^{-(r+3)(1-\epsilon)/2}).$$

Substituting these estimates into (8), we obtain

$$(16) \quad b_{n+1}^{(r)} = (1 - r(\alpha_1 + \eta')n^{-(1-\epsilon)})b_n^{(r)} + \binom{r}{2} \sigma^2 b_n^{(r-2)} n^{-2(1-\epsilon)} + O(n^{-(r+3)(1-\epsilon)/2}).$$

Now assume as our induction hypothesis that $\lim_{n\to\infty} n^{(r-2)(1-\epsilon)/2} b_n^{(r-2)} = B_{r-2}$. It follows that if $n > n_0(\eta)$,

$$\begin{aligned} b_{n+1}^{(r)} = (1 - r(\alpha_1 + \eta')n^{-(1-\epsilon)})b_n^{(r)} &+ \binom{r}{2} (\sigma^2 + \eta'')B_{r-2} n^{-(1+r/2)(1-\epsilon)} \\ &+ O(n^{-(1-\epsilon)(r+3)/2}) \end{aligned}$$

where $|\eta'| \leqq \eta$ and $|\eta''| \leqq \eta$. A fast application of Lemmas 3 and 4 yields

$$B_r = \lim_{n\to\infty} n^{(1-\epsilon)r/2} b_n^{(r)} = \binom{r}{2} \sigma^2 B_{r-2}/r\alpha_1 = (r - 1)\, \sigma^2 B_{r-2}/2\alpha_1 .$$

From (i) and (ii) above we have $B_1 = 0$ and $B_2 = \sigma^2/2\alpha_1$, hence it follows now by induction that for each integer $r \geqq 1$

$$B_r = \begin{cases} 0 & \text{if } r = 2s - 1, \\ (\sigma^2/2\alpha_1)^s(2s-1)(2s-3)\cdots 3\cdot 1 & \text{if } r = 2s. \end{cases}$$

This proves the first part of Theorem 6, the second part is a well known consequence. We may remark, for the benefit of future textbook writers, that here is another instance in which the method of moments seems to apply more easily than the more modern method of characteristic functions. (see however Sec. 6). This method of moments is not mentioned in several books on probability and statistics.

5. Second case. In this section we treat the second case described in Section 2. Assumptions (0), (I) to (V) are as stated in Section 4. Others are

ASSUMPTION (VI). The function $M(x)$ is bounded in any finite interval of x, and we have

$$0 < \varliminf_{|x| \to \infty} M(x)/x \leqq \varlimsup_{|x| \to \infty} M(x)/x < \infty.$$

ASSUMPTION (VII). For a certain even integer $p \geqq 2$ we have

$$E[(Y(x) - M(x))^p] \leqq K_{21} < \infty.$$

We note that (I), (II) and (VI) imply that

$$K \mid x - \theta \mid \leqq \mid M(x) - \alpha \mid \leqq K_{22} \mid x - \theta \mid. \tag{1}$$

The constant $K > 0$ will figure prominently in what follows and so we omit its subscript. We also introduce a new constant for the upper bound on the variance of $Y(x)$:

$$E[(Y(x) - M(x))^2] \leqq K_{23} < \infty. \tag{2}$$

The existence of such a constant is of course implied by (VII); in fact $K_{23} \leqq K_{21}^{2/p}$.

In this section we set $n \geqq 1$

$$a_n = c/n \qquad 0 < c < \infty$$

where c is to be chosen later. In contrast to Section 4, the initial value x_1 may now be any random variable, bounded or not. The analysis in this section is quite similar to and somewhat simpler than that in Section 4, and we shall be more brief.

THEOREM 7. *Suppose that Assumptions (I), (II), (VI) and (VIII) are satisfied.*[6] *If $a_n = c/n$ with $c > 1/2K$, then for every positive $r \leqq p$ we have*

$$\varlimsup_{n \to \infty} n^{r/2}\beta_n^{(r)} \leqq B_r < \infty \tag{3}$$

[6] If we assume that $M(\cdot)$ is continuous everywhere then these four assumptions may be replaced by (1) and (VII).

where

$$B_r \leqq [r!/2^{r/2}(r/2)!][(K_{23}c^2/(2Kc - 1)]^{r/2}$$

for an even integer r.

REMARK. To minimize the above bound for B_r we should choose $c = 1/K$, giving $B_r \leqq [r!/2^{r/2}(r/2)!][K_{23}/K^2]^{r/2}$. However, since K is unknown it is better not to fix c.

PROOF. By (1) we have $d_n \geqq Kb_n$. On the other hand, by (1) and (2),

$$e_n = E[(y_n - M(x_n))^2] + E[(M(x_n) - \alpha)^2] \leqq K_{23} + K_{22}^2 b_n .$$

Using these in (4.6) we have

$$\begin{aligned} b_{n+1} &\leqq (1 - 2Kc/n + K_{22}^2c^2/n^2)b_n + K_{23}c^2/n^2 \\ &\leqq (1 - (2Kc - \eta)/n)b_n + K_{23}c^2/n^2 \end{aligned}$$

for $n > n_0(\eta, c)$, where $\eta > 0$ is arbitrarily small. Let $c > 1/2K$. Applying Lemma 1 we obtain

$$b_n \leqq \frac{K_{23}\, c^2}{2Kc - \eta - 1}\frac{1}{n} + O\left(\frac{1}{n^2} + \frac{1}{n^{2Kc-\eta}}\right).$$

Since c may be chosen arbitrarily close to $1/2K$, and η arbitrarily small, (3) is true for $r = 2$ with $B_2 = K_{23}c^2/(2Kc - 1)$.

Now let r be even and assume that

$$\overline{\lim_{t\to\infty}}\ n^{t/2}\beta_n^{(t)} \leqq B_t < \infty$$

for $2 \leqq t \leqq r - 2$. Consider the J_t, $2 \leqq t \leqq r - 2$, defined in (4.9). We have, by (1) and (2),

$$\begin{aligned} J_2 &= c^2n^{-2}E[|\, x_n - \theta\,|^{r-2}E\{(y_n - M(x_n))^2 + (M(x_n) - \alpha)^2 \mid x_n\}] \\ &\leqq c^2n^{-2}\{K_{23}b_n^{(r-2)} + K_{22}^2b_n^{(r)}\}. \end{aligned}$$

Similarly, if $3 \leqq t \leqq r - 2$, we have, using $(a + b)^t \leqq 2^t(a^t + b^t)$,

$$|\, J_t\,| \leqq 2^tc^tn^{-t}[K_{21}\beta_n^{(r-t)} + K_{22}^tb_n^{(r)}] = O(n^{-(r+3)/2}).$$

On the other hand, we have by (1),

$$E[(x_n - \theta)^{r-1}(M(x_n) - \alpha)] \geqq Kb_n^{(r)}.$$

Using these in (4.8) we have

$$b_{n+1}^{(r)} \leqq (1 - rKc/n + O(n^{-2}))b_n^{(r)} + \binom{r}{2}K_{23}c^2B_{r-2}n^{-(r/2+1)} + O(n^{-(r+3)/2})$$

Applying Lemma 1 we obtain

$$\overline{\lim_{n\to\infty}}\ n^{r/2}b_n^{(r)} \leqq r(r - 1)K_{23}\, c^2B_{r-2}/(2rKc - r).$$

Thus (3) is inductively true with

$$B_r \leqq [(r-1)K_{23}c^2/(2Kc-1)]B_{r-2}.$$

This yields the stated bounds for B_r and proves (3) for every even $r \leqq p$. The rest of the theorem follows from Lyapunov's inequality.

THEOREM 8. *Suppose that the Assumptions (I), (II), (IV), and (VI) are satisfied; and (VII) is satisfied with a certain $p \geqq 6$. If $a_n = c/n$ with $c > 1/2K$ then we have*

$$\varliminf_{n\to\infty} nb_n \geqq \frac{K_2 c^2}{2\alpha_1 c - 1} > 0.$$

REMARK. Note that $\alpha_1 \geqq K$ so that $2\alpha_1 c - 1 > 0$.

PROOF. As in the proof of Theorem 4, given any $\eta > 0$, there exists a $\delta = \delta(\eta)$ such that

$$\int_{|x_n-\theta|\leqq\delta} (x_n-\theta)(M(x_n)-\alpha)\,dPr = (\alpha_1+\eta'')b_n - \alpha_1\int_{|x_n-\theta|>\delta}(x_n-\theta)^2\,dPr$$

where $|\eta''| \leqq \eta$. By Theorem 7 we have

$$(4) \qquad \int_{|x_n-\theta|>\delta}(x_n-\theta)^2\,dPr \leqq \delta^{-(p-2)}\,b_n^{(p)} = O(n^{-p/2}).$$

Furthermore, by (1) and (4) we have

$$\int_{|x_n-\theta|>\delta}(x_n-\theta)(M(x_n)-\alpha)\,dPr \leqq K_{22}\int_{|x_n-\theta|>\delta}(x_n-\theta)^2\,dPr = O(n^{-p/2}).$$

Combining these results we obtain

$$(5) \qquad d_n = (\alpha_1+\eta'')b_n + O(n^{-p/2}).$$

By (IV) we have $e_n \geqq K_2$. Using these in (4.6) we have, since $p \geqq 6$,

$$b_{n+1} \geqq [1-2(\alpha_1+\eta)c/n]b_n + K_2c^2n^{-2}[1+o(1)].$$

Applying Lemma 2 we conclude that

$$\varliminf_{n\to\infty} nb_n \geqq \frac{K_2c^2}{2(\alpha_1+\eta)c-1}.$$

Since η is arbitrary Theorem 8 is proved.

COROLLARY. *We have* $\lim_{n\to\infty}(d_n/b_n) = \alpha_1$.

Proof of this follows from (5) and the theorem itself.

THEOREM 9. *Suppose that the Assumptions (I), (II), (V) and (VI) are satisfied; (VIII) is satisfied for every p with $K_{21} = K_{21}(p)$. If $a_n = c/n$ with $c > 1/2K$, then we have for every integer $r \geqq 1$*

$$(6) \qquad \lim_{n\to\infty} n^{r/2}b_n^{(r)} = \begin{cases} 0 & \text{if } r = 2s-1, \\ [\sigma^2c^2/(2\alpha_1c-1)]^s(2s-1)(2s-3)\ldots 3\cdot 1 & \text{if } r = 2s. \end{cases}$$

Consequently the random variable $n^{1/2}(x_n - \theta)$ tends in distribution to the normal distribution with mean 0 and variance $\sigma^2c^2/(2\alpha_1 c - 1)$.

PROOF. The proof of Theorem 9 is similar to that of Theorem 6, except that certain estimates are obtained in a slightly different way. We need only note the following points.

For every $r \geqq 0$, $\delta > 0$, $q > 0$ we have

$$(7) \quad \int_{|x_n-\theta|>\delta} |x_n - \theta|^r dPr \leqq \delta^{-2q} \int_{|x_n-\theta|>\delta} |x_n - \theta|^{r+2q} dPr = O(n^{-q}).$$

This follows from Theorem 7.

For every integer $r \geqq 0$ and $q > 0$, we have

$$E[(x_n - \theta)^{r-1}(M(x_n) - \alpha)]$$
$$= (\alpha_1 + \eta) \int_{|x_n-\theta|\leqq\delta} (x_n - \theta)^r dPr + O\left[\int_{|x_n-\theta|>\delta} |x_n - \theta|^r dPr\right]$$
$$= (\alpha_1 + \eta)b_n^{(r)} + O(n^{-q})$$

where $\eta = \eta(\delta)$ and $\lim_{\delta\to 0} \eta(\delta) = 0$. This follows from (1) and (7). Furthermore we have

$$E[(x_n - \theta)^{r-2}(M(x_n) - \alpha)^2] \leqq K_{22}^2\beta_n^{(r)} = O(n^{-r/2}).$$

This follows from (1) and Theorem 7.

Using the Corollary to Theorem 8 we obtain

$$b_{n+1}^{(r)} = [1 - r\alpha_1 c(1 + o(1))/n]b_n^{(r)} + \binom{r}{2}\sigma^2c^2b_n^{(r-2)}[1 + o(1)]n^{-(r/2+1)}$$

(cf. (4.16)), from which the theorem follows.

6. Linear case. In this section we consider the simplest possible $M(\cdot)$, namely, a linear function

$$M(x) = \mu x - \theta \qquad \mu \neq 0$$

where both μ and θ are unknown. Without loss of generality we may suppose $\mu > 0$ and set $\alpha = 0$. The problem is then to obtain θ/μ stochastically. This case is not covered by either [1] or [2] but is covered by our second case if $Y(x)$ has finite moments to a certain order. Here we treat it with a different method under the sole hypothesis that there exists a distribution function $F(x)$ with mean 0 such that $H(y \mid x) = F(x - M(x))$.

Let the characteristic function of $F(x)$ be $f(t)$; then that of $H(y \mid x)$ is

$$e^{it(\mu x-\theta)}f(t).$$

In other words, we have the conditional expectation

$$(1) \qquad E[e^{ity_n} \mid x_n] = e^{it(\mu x_n-\theta)}f(t).$$

Let the characteristic function of x_n be $f_n(t)$, $n \geqq 1$. Then we have, recalling that $x_{n+1} = x_n - a_n y_n$ and using (1),

$$f_{n+1}(t) = E[e^{itx_{n+1}}] = E[e^{it(x_n - a_n y_n)}]$$
$$= E[e^{itx_n} E\{e^{-ita_n y_n} \mid x_n\}] = E[e^{itx_n} e^{-ita_n(\mu x_n - \theta)} f(-a_n t)]$$
$$= E[e^{it(1-\mu a_n)x_n}] f(-a_n t) e^{ita_n\theta} = f_n((1 - \mu a_n)t) f(-a_n t) e^{ita_n \theta}.$$

It follows by recursion that

$$f_{n+1}(t) = \exp\left[\frac{it\theta}{\mu}\left\{1 - \prod_{k=1}^{n}(1 - \mu a_k)\right\}\right]$$
$$\prod_{k=1}^{n} f[-(1 - \mu a_n) \cdots (1 - \mu a_{k+1}) a_k t] f_1\left(\prod_{k=1}^{n}(1 - \mu a_k)t\right).$$

If we now choose

$$a_n = 1/n\mu, \tag{2}$$

then we have

$$\prod_{k=1}^{n}(1 - \mu a_k) = 0 \quad \text{and} \quad (1 - \mu a_n) \cdots (1 - \mu a_{k+1}) a_k = \frac{1}{\mu n}.$$

Therefore we have, for every $n \geqq 1$ and every initial $f_1(\cdot)$,

$$f_{n+1}(t) = e^{it\theta/\mu}[f(-t/\mu n)]^n. \tag{3}$$

Equation (3) determines the distribution of x_n, $n \geqq 2$, completely, at least in theory. Let $\xi_1, \xi_2, \cdots$ be independent random variables with the same distribution function $F(x)$ with mean 0. Then the characteristic function of

$$-(\xi_1 + \cdots + \xi_n)/n$$

is precisely $[f(-t/\mu n)]^n$. Thus the study of (3) is reduced to a classical problem in probability theory. We need not go into details here but shall content ourselves with mentioning the following facts.

First, since $F(x)$ has mean 0 it follows from Khintchine's weak law of large numbers (see e.g. [3], p. 253; [5], p. 138) that

$$\lim_{n\to\infty} f(-t/\mu n)^n = 0.$$

Therefore by (3)

$$\lim_{n\to\infty} f_n(t) = e^{it\theta/\mu}$$

and consequently x_n tends to θ in probability. It is curious that in this simple case the stochastic convergence of the procedure is equivalent to Khintchine's theorem.

Next it follows, by a classical result of P. Lévy ([4], p. 254; see also [5], pp.) 163, that all possible limit distributions of $x_n - \theta$ are stable laws (including the normal) of exponent α: $1 < \alpha \leqq 2$. More precisely, if there exists a sequence of positive constants $\{A_n\}$ such that $(\xi_1 + \cdots + \xi_n)/A_n$ tends in distribution[7] to the stable law $G(x)$, then $(\mu n/A_n)(\theta - x_n)$ does the same. In particular, every stable law of exponent α, $1 < \alpha \leqq 2$, is the limit distribution of $x_n - \theta$, for a suitable choice of $F(\cdot)$.

Finally, it is known (see [5], p. 186) that there exist distribution functions $F(x)$ such that $A_n(x_n - \theta)$ does not tend in distribution to a limit, whatever the sequence A_n may be.

As a last remark, it is clear from (3) that the proper choice of $\{a_n\}$ must depend to a certain extent on the unknown parameter μ. In fact, any other choice than (2), even $a_n = c/n$ with $c \neq 1/\mu$, already greatly complicates the analysis given above. It is therefore small wonder that in Sections 4 and 5 the choice of $\{a_n\}$ has to depend on the unknown.

7. Consequences. In this section we sketch briefly some statistical consequences of the results of Section 5. For brevity we state strong assumptions which may obviously be weakened. We put $\alpha = 0$ without loss of generality. Let $\mathcal{H} = \{H\}$ be a family of functions of the type denoted by $H(\cdot|\cdot)$ in Section 2. Denote by $M_H(\cdot)$, θ_H, σ_H^2, K_H etc., the $M(\cdot)$, θ, σ^2, K, etc. (if they exist) corresponding to a given H.

We assume throughout this section that, for each $H \varepsilon \mathcal{H}$, Assumptions (I), (V), and (VII) (with subscripts H) are satisfied, and that there are (known) positive numbers γ and β such that $\alpha_{1H} \geqq \gamma$ and $\sigma_H^2 \leqq \beta$ for all $H \varepsilon \mathcal{H}$. (The constancy of σ_H^2 for fixed H is not necessary; see footnote 4.) We suppose also that there is an interval I of positive length such that $\mathcal{H}$ contains the family $\mathcal{G}$ consisting in every function H for which, for some $z \varepsilon I$, $H(y - \gamma(x - z) \mid x)$ is, for all x, the normal distribution function with mean 0 and variance β. For simplicity we put $x_1 = 0$ throughout this section.

Suppose now that, for a given even integer $r \geqq 2$, the conclusion to Theorem 9 with $c = 1/\gamma$ holds *for that* r uniformly over all H in $\mathcal{H}$. It may be important for applications to note that it suffices for our argument below that for every $\epsilon > 0$ there exists an $N(\epsilon)$ such that $n^{r/2}b_n^{(r)}(1 - \epsilon)$ is $\leqq$ the right side of (5.6) for all $n > N(\epsilon)$ and all $H \varepsilon \mathcal{H}$. It is clear that usually this would have been more difficult to satisfy under the assumptions of Section 4 (especially III).

An examination of the proof of Theorem 9 shows that the validity of the conclusion for a fixed r need not entail Assumption VI for *all* p. The reader may scrutinize our proofs to obtain various conditions on $\mathcal{H}$ under which the conclusion stated above holds. Note that it can hold even if the domain of values of θ_H for $H \varepsilon \mathcal{H}$ is unbounded, despite the unboundedness of $b_{1H}^{(r)}$ over $\mathcal{H}$ in that

[7] It can be shown (see [5], p. 175, Translator's note) that if there exist $\{A_n\}$ and $\{A'_n\}$ such that $(\xi_1 + \cdots + \xi_n - A'_n)/A_n$ tends in distribution to a stable law, then we may take $A'_n = 0$ for all n. This is because $\int x \, dF(x) = 0$.

case. For example, if $\mathcal{H}$ is a family like that of Section 6 (with rth moment finite) with $\gamma = \mu = 1/c$ (μ is known), then $b_{nH}^{(r)}$ is independent of H (i.e., of θ_H) for $n > 1$ (see (6.3)). Similar remarks with inequality apply if $\mathcal{H}$ consists in all H for which M_H is linear with known slope and the rth central moment of H is bounded over $\mathcal{H}$. For $r = 2$ and such an $\mathcal{H}$ (that is, $W(H, d) = (\theta_H - d)^2$ below), the result of this section was obtained in the Hodges and Lehmann manuscript cited in Section 2. Our M_H need not, of course, be linear.

Let S_n be the n-observation statistical procedure defined by using the Robbins-Monro scheme with $x_1 = 0$ and $a_n = 1/\gamma n$ for n observations $y_1, \cdots, y_n$ and then estimating θ_H by x_{n+1}. Clearly, as $n \to \infty$

$$(1) \quad \sup_{H \varepsilon \mathcal{H}} E_H \mid x_{n+1} - \theta_H \mid^r = n^{-r/2}(\beta/\gamma)^r[(r-1)(r-3)\ldots\cdot 3\cdot 1][1 + o(1)].$$

Let $H' \varepsilon \mathcal{G}$. Then the random variables $u_i = x_i - (1/\gamma)y_i$, $i = 1, 2, \cdots$, are independently and identically distributed Gaussian variables (the correlation between any two of them is easily computed to be 0) with mean θ_H, and variance β/γ. A knowledge of the values taken on by $u_1, \cdots, u_n$ is equivalent to that of those taken on by $y_1, \cdots, y_n$ (recall that $x_1 = 0$), and $z_n = n^{-1}\sum_{i=1}^{n} u_i$ is a sufficient statistic for the family $\mathcal{G}$. Since I has positive length and

$$E_{H'} \mid z_n - \theta_{H'} \mid^r = n^{-r/2}(\beta/\gamma)^r[(r-1)(r-3)\ldots\cdot 3\cdot 1] = O(n^{-r/2}) \qquad \text{all } H' \varepsilon \mathcal{G},$$

a simple modification of the argument of Wolfowitz [6] (as applied to the fixed sample-size case) shows that, if $\mathfrak{J}_n$ is the class of all procedures T requiring a total of n observations (which are taken sequentially by determining $x_1, \cdots, x_n$ *in any prescribed manner*, not necessarily that of S_n), and if δ_T is the final estimator of θ_H when the procedure T is used (thus, $\delta_{S_n} = x_{n+1}$ when the x_i are determined by the scheme S_n), then as $n \to \infty$,

$$(2) \quad \inf_{T \varepsilon \mathfrak{J}_n} \sup_{H' \varepsilon \mathcal{G}} E(\delta_T - \theta_{H'})^r = n^{-r/2}(\beta/\gamma)^r[(r-1)(r-3)\ldots\cdot 3\cdot 1][1 - o(1)].$$

We conclude from (1) and (2) that for the problem of point estimation of θ based on n observations *which may be taken in any manner*, and when the weight function (the loss when H is the "true" member of $\mathcal{H}$ and we estimate θ_H by d) is $W(H, d) = \text{const.}\cdot(\theta_H - d)^r$, *the procedure S_n is asymptotically minimax in the sense that*

$$(3) \quad \lim_{n\to\infty} \frac{\sup_{H \varepsilon \mathcal{H}} E_H W(H, \delta_{S_n})}{\inf_{T \varepsilon \mathfrak{J}_n} \sup_{H \varepsilon \mathcal{H}} E_H W(H, \delta_T)} = 1.$$

Thus, for large n the procedure S_n seems to be very satisfactory.

The above result may easily be strengthened. Assume for simplicity that the interval I characterizing $\mathcal{G}$ is the whole real line. Let $\mathfrak{J}$ now be the class of all sequential procedures T which terminate (no longer necessarily at a specified n) with probability one under all H in $\mathcal{H}$. As before, let δ_T be the estimator of θ_H when T is used, and let $E_H(N \mid T)$ be the expected number of observations before termination of the experiment when H is the "true" member of $\mathcal{H}$ and the procedure T is used. Let $c > 0$ be the cost of taking a single observation y_n, whatever be T, n, and x_n. Let

$$r(H, T, c) = cE_H(N \mid T) + E_H W(H, \delta_T) \tag{4}$$

be the risk function of T when c is the cost of experimentation. Then the results of [6] (for the sequential case) and an argument like that of the previous paragraph show that, for the setup of the previous paragraph, there is an integral valued function $\nu(\cdot)$ on the positive reals (which is easy to calculate from [6]) such that

$$\lim_{c \to 0} \frac{\inf_{T \varepsilon \mathfrak{T}} \sup_{H \varepsilon \mathcal{H}} r(H, T, c)}{\sup_{H \varepsilon \mathcal{H}} r(H, S_{\nu(c)}, c)} = 1. \tag{5}$$

Alternatively, we may state that if $\mathfrak{T}_n^*$ is the subset of $\mathfrak{T}$ for which $\sup_{H \varepsilon \mathcal{H}} E_H(N \mid T) \leqq n$, then

$$\lim_{n \to \infty} \frac{\inf_{T \varepsilon \mathfrak{T}_n^*} \sup_{H \varepsilon \mathcal{H}} E_H W(H, \delta_T)}{\sup_{H \varepsilon \mathcal{H}} E_H W(H, \delta_{S_n})} = 1. \tag{6}$$

The dual property to (6), wherein we minimax $E_H(N \mid T)$ subject to $\sup_H E_H W(H, \delta_T) \leqq w$ as $w \to 0$, can be stated similarly.

The results of the two previous paragraphs may easily be extended to weight functions other than $W(H, d) = \mid \theta_H - d \mid^r$. The results of [6] may be applied whenever $W(H, d)$ is a nondecreasing function of $\mid \theta_H - d \mid$ satisfying appropriate integrability conditions (see [6]). Thus, one need only verify that $\mathcal{H}$ satisfies a condition like that of footnote 2 above with the inequality replaced by

$$E_H W(H, x_n)/E_{H'} W(H, x_n) \leqq 1 - \epsilon$$

for some H' in $\mathcal{G}$, in order to obtain results like (3), (5) and (6) above. Particular choices of W will give results on interval estimation, etc.

Questions of optimality for the setup of Section 4 remain unanswered at present because the technique used in Section 7 is not applicable, linear $M(\cdot)$ being disallowed in Section 4 and our knowledge of the case $a_n = c/n$ being incomplete there. We need not, of course, detail the remark that the results of Section 4 (like those of Sec. 7) may still be used to obtain asymptotic confidence intervals, etc.

In conclusion, I wish to thank my colleague, Professor J. C. Kiefer, for criticism of the MS and for contributing largely to Section 7.

REFERENCES

[1] HERBERT ROBBINS AND SUTTON MONRO, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.

[2] J. WOLFOWITZ, "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 457–461.

[3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] P. LÉVY, *Calcul des Probabilités*, Gauthier-Villars, Paris, 1925.

[5] B. V. GNEDENKO AND A. N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables;* English translation by K. L. Chung, Addison-Wesley Co., Cambridge, 1954.

[6] J. WOLFOWITZ, "Minimax estimation of the mean of a normal distribution with known variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 218–230.

[7] L. SCHMETTERER, "Bemerkungen zum Verfahren der stochastischen Iteration," *Osterreich. Ing.-Arch.*, Vol. 7 (1953), pp. 111–117.