The Annals of Mathematical Statistics
1968, Vol. 39, No. 4, 1327–1332

ON ASYMPTOTIC NORMALITY IN STOCHASTIC APPROXIMATION[1]

BY VÁCLAV FABIAN

Michigan State University and Czechoslovak Academy of Sciences

1. Summary and introduction. A new method, simpler than previous methods due to Chung (1954) and Sacks (1958), is used to prove Theorem 2.2 below, which implies in a simple way all known results on asymptotic normality in various cases of stochastic approximation. Two examples of application are concerned with Venter's (1967) extension of the RM method and Fabian's (1967) modification of the KW process. Previously, although there was no difficulty in adopting one or the other method, the proofs in various cases had to be done almost *ab initio* or skipped leaving a gap (see Venter (1967)).

The new proof is similar to that of Chung except that the basic recurrence relation is used to obtain the asymptotic characteristic function rather than limits of all moments. We remark that Lemma 2.1, a simple corollary to Chung's lemma is used only to obtain condition (2.2.4) which is weaker than (2.2.3) if $\alpha = 1$ and which corresponds to the usual Lindeberg condition. Both conditions (2.2.3) and (2.2.4) are weaker than the corresponding condition (3.4) in Sacks (1958).

In what follows $(\Omega, \mathcal{S}, P)$ will be a probability space, relations between and convergence of random variables, vectors, and matrices will be meant with probability one unless specified otherwise.

We shall write $X_n \sim \mathcal{L}$ if X_n is asymptotically $\mathcal{L}$-distributed and $X_n \sim Y_n$, for two sequences of random vectors, if for any $\mathcal{L}$, $X_n \sim \mathcal{L}$ if and only if $Y_n \sim \mathcal{L}$.

The indicator function of a set A will be denoted by χA, the expectation and conditional expectation by E and E_F, respectively. R^k is the k-dimensional Euclidean space the elements of which are considered to be column vectors, $R = R^1$, $R^{k\times k}$ is the space of all real $k \times k$ matrices. The symbols $\mathbf{R}$, $\mathbf{R}^k$, $\mathbf{R}^{k\times k}$, denote sets of all measurable transformations from $(\Omega, \mathcal{S})$ to R, R^k $R^{k\times k}$, respectively. The unit matrix in $R^{k\times k}$ is denoted by I and $\| \ \|$ is the Euclidean norm. With h_n a sequence of numbers, $o(h_n)$, $O(h_n)$, $o_u(h_n)$, $O_u(h_n)$ denote sequences g_n, G_n, q_n, Q_n, say, of elements in one of the sets $\mathbf{R}$, $\mathbf{R}^k$, $\mathbf{R}^{k\times k}$ such that $h_n^{-1}g_n \to 0$, $\|h_n^{-1}G_n\| \leqq f$ for an $f \varepsilon \mathbf{R}$ and all n, $h_n^{-1}q_n \to 0$ uniformly on a set of probability one, $\|h_n^{-1}Q_n\| \leqq K$ for a $K \varepsilon R$ and all n. In special cases $o(h_n)$ may be constant on Ω and considered as a sequence with elements in R, R^k or $R^{k\times k}$. Similarly in other cases.

For Chung's lemma, which will be frequently referred to, or used later without reference, see Fabian ((1967), Lemma 4.2); note that it holds with $\beta = 0$, too.

Received 4 October 1967.

[1] Prepared under NSF Grant GP-7362.

2. The result.

2.1. LEMMA. *Suppose* $A > 0$, $h_n, b_n \varepsilon R$,

(2.1.1) $$b_{n+1} = (1 - An^{-1})b_n + n^{-1}h_n.$$

Then $b_n \to 0$ *if and only if* $n^{-1}\sum_{j=1}^n h_j \to 0$.

PROOF. From (2.1.1), $n(b_{n+1} - b_n) + Ab_n = h_n$ and so

$$\sum_{j=1}^n h_j = \sum_{j=1}^n j(b_{j+1} - b_j) + A\sum_{j=1}^n b_j = (A - 1)\sum_{j=1}^n b_j + nb_{n+1}.$$

This shows the necessity of $n^{-1}\sum_{j=1}^n h_j \to 0$. To show the sufficiency, let $\beta_n = n^{-1}\sum_{j=1}^n b_j$, $\epsilon_n = n^{-1}\sum_{j=1}^n h_j \to 0$. Then $b_{n+1} = (1 - A)\beta_n + \epsilon_n$ and it suffices to show that $\beta_n \to 0$. But $\beta_{n+1} = (1 - (n+1)^{-1})\beta_n + (n+1)^{-1}b_{n+1} = (1 - A(n+1)^{-1})\beta_n + (n+1)^{-1}\epsilon_n$ so that the desired conclusion follows from Chung's lemma.

2.2. THEOREM. *Suppose* k *is a positive integer,* $\mathfrak{F}_n$ *a non-decreasing sequence of* σ*-fields,* $\mathfrak{F}_n \subset \mathfrak{S}$; *suppose* $U_n, V_n, T_n \varepsilon \mathrm{R}^k$, $T \varepsilon R^k$, $\Gamma_n, \Phi_n \varepsilon \mathrm{R}^{k\times k}$, $\Sigma, \Gamma, \Phi, P \varepsilon R^{k\times k}$, Γ *is positive definite,* P *is orthogonal and* $P'\Gamma P = \Lambda$ *diagonal. Suppose* $\Gamma_n, \Phi_{n-1}, V_{n-1}$ *are* $\mathfrak{F}_n$*-measurable,* $C, \alpha, \beta \varepsilon R$ *and*

(2.2.1) $$\Gamma_n \to \Gamma, \qquad \Phi_n \to \Phi, \qquad T_n \to T \quad \text{or} \quad E\|T_n - T\| \to 0,$$

(2.2.2) $$E_{\mathfrak{F}_n}V_n = 0, \qquad C > \|E_{\mathfrak{F}_n}V_nV_n' - \Sigma\| \to 0,$$

and, with $\sigma_{j,r}^2 = E\chi\{\|V_j\|^2 \geqq rj^\alpha\}\|V_j\|^2$, *let either*

(2.2.3) $$\lim_{j\to\infty}\sigma_{j,r}^2 = 0 \qquad \text{for every } r > 0,$$

or

(2.2.4) $$\alpha = 1, \qquad \lim_{n\to\infty} n^{-1}\sum_{j=1}^n \sigma_{j,r}^2 = 0 \qquad \text{for every } r > 0.$$

Suppose that, with $\lambda = \min_i \Lambda^{(ii)}$, $\beta_+ = \beta$ *if* $\alpha = 1$, $\beta_+ = 0$ *if* $\alpha \neq 1$,

(2.2.5) $$0 < \alpha \leqq 1, \qquad 0 \leqq \beta, \qquad \beta_+ < 2\lambda$$

and

(2.2.6) $$U_{n+1} = (I - n^{-\alpha}\Gamma_n)U_n + n^{-(\alpha+\beta)/2}\Phi_nV_n + n^{-\alpha-\beta/2}T_n.$$

Then the asymptotic distribution of $n^{\beta/2}U_n$ *is normal with mean* $(\Gamma - (\beta_+/2)I)^{-1}T$ *and covariance matrix* PMP' *where*

(2.2.7) $$M^{(ij)} = (P'\Phi\Sigma\Phi'P)^{(ij)}(\Lambda^{(ii)} + \Lambda^{(jj)} - \beta_+)^{-1}.$$

PROOF. Let $\mathfrak{N}(m, S)$ be the normal distribution with mean m and covariance matrix S, let $\mathfrak{N} = \mathfrak{N}((\Gamma - (\beta_+/2)I)^{-1}T, PMP')$. As the first step we shall show that we may, without loss of generality, assume that

(2.2.8) $$\Gamma = \Lambda,\ P = I,\ \beta = \beta_+ = 0,\ \|\Gamma_n - \Lambda\| = o_u(1),\ \Phi = \Phi_n = I,$$
$$T_n = U_1 = 0.$$

As for the first three parts of (8), it suffices to observe that

$$\tilde{U}_n = (n - 1)^{\beta/2}P'U_n$$

again satisfy the assumptions of the theorem, but with $\tilde{\Gamma} = \tilde{\Lambda} = \Lambda - (\beta_+/2)I$, $\tilde{P} = I$, $\tilde{\Phi} = P'\Phi$, $\tilde{T} = P'T$. Next we can assume that (i) $\Gamma_n = \Lambda + o_u(1)$, $\Phi_n = \Phi + o_u(1)$. Indeed, suppose the theorem holds under the additional assumption (i). Then for every $\epsilon > 0$ we can, using Egorov theorem, construct a new process $\tilde{U}_n$ which differs from U_n only on a set of probability less than ϵ and for which all assumptions of the theorem together with (i) hold, with Σ, Φ, T, Λ unchanged. Then $\tilde{U}_n \sim \mathfrak{N}$ and because ϵ was arbitrary, $U_n \sim \mathfrak{N}$. Similarly we can assume $T_n \to T$ uniformly if $T_n \to T$ and, because of (1), we may assume that $E\,\|T_n - T\| \to 0$. With $\Phi_n = \Phi + o_u(1)$, and because (2) implies $E_{\mathfrak{F}_n}\|V_n\|^2 = O_u(1)$, we can change V_n to $\Phi_n V_n$ and assume $\Phi_n = \Phi = I$.

Note that if $\tilde{U}_n$ satisfy (6) with Z_n subtracted from the right hand side then $\Delta_n = U_n - \tilde{U}_n$ satisfy $\Delta_{n+1} = (I - n^{-\alpha}\Gamma_n)\Delta_n + Z_n$ and

$$\|\Delta_{n+1}\| \leqq (1 - n^{-\alpha}[\lambda + o_u(1)])\|\Delta_n\| + \|Z_n\|. \tag{2.2.9}$$

Setting $Z_n = 0$, $\tilde{U}_1 = 0$ gives ((9) and Chung's lemma) $\|\Delta_n\| \to 0$. If $Z_n = n^{-\alpha}(T_n - T)$ then $E\,\|Z_n\| = o(n^{-\alpha})$ and $E\,\|\Delta_n\| \to 0$. In both cases $\tilde{U}_n \sim U_n$ and we may assume $U_1 = 0$, $T_n = T$. Setting $Z_n = n^{-\alpha}T$ gives first $\|\Delta_n\| = O(1)$; then $\Delta_{n+1} = (I - n^{-\alpha}\Lambda)\Delta_n + n^{-\alpha}T + n^{-\alpha}(\Lambda - \Gamma_n)\Delta_n$ with the last term being $o(n^{-\alpha})$. A coordinatewise application of Chung's lemma gives $\Delta_n \to \Lambda^{-1}T$. Thus $U_n \sim \mathfrak{N}$ if and only if $\tilde{U}_n \sim \mathfrak{N}(0, M)$ and we can assume $T = 0$ and accept the whole assumption (8).

Using (2) and the measurability of Γ_n and U_n with respect to $\mathfrak{F}_n$, we obtain easily $E\,\|U_n\|^2 = O(1)$; setting then $Z_n = n^{-\alpha}(\Gamma_n - \Lambda)U_n$ we get $E\,\|Z_n\| = o(n^{-\alpha})$ and $E\,\|\Delta_n\| \to 0$, and we may assume

$$U_{n+1} = (I - n^{-\alpha}\Lambda)U_n + n^{-\alpha/2}V_n. \tag{2.2.10}$$

Next we choose a sequence $\delta_n \to 0$ such that (3) or (4) holds with δ_j substituted for r. Choosing a $\tau > 0$ and restricting t to $\{t; \|t\| \leqq \tau\}$ we obtain from a standard relation (see, e.g., Feller (1966), proof of Theorem 1, XV.6)

$$E\,|E_{\mathfrak{F}_n} e^{in^{-\alpha/2}t'V_n} - 1 + \tfrac{1}{2}n^{-\alpha}t'\Sigma t| = \|t\|\,n^{-\alpha}h_n \tag{2.2.11}$$

with

$$\begin{aligned}\|t\|\,n^{-\alpha}h_n \leqq{}& E\,|\tfrac{1}{2}n^{-\alpha}t'(\Sigma - E_{\mathfrak{F}_n}V_nV_n')t| \\ &+ En^{-\alpha}(t'V_n)^2(\|t\|\,\delta_n^{\frac{1}{2}} + \chi\{n^{-\alpha}(t'V_n)^2 \geqq \|t\|^2\,\delta_n\}) \\ &\leqq \|t\|^2\,o(n^{-\alpha}) + \|t\|^3\,o(n^{-\alpha}) + n^{-\alpha}\,\|t\|^2\,\sigma_{n\delta_n}^2\end{aligned}$$

so that $h_n \leqq o(1) + \tau\sigma_{n,\delta_n}^2$ and either $h_j \to 0$ or $\alpha = 1$ and $n^{-1}\sum_{j=1}^n h_j \to 0$.

Now let us denote $B_n = I - n^{-\alpha}\Lambda$, $\varphi_n(t) = Ee^{itU_n}$, $\psi_1(t) = 1$, $\psi_{n+1}(t) = \psi_n(B_nt)(1 - \tfrac{1}{2}n^{-\alpha}t'\Sigma t)$. Then a rearrangement of terms, (11) and $\mathfrak{F}_n$-measurability of U_n gives

$$\begin{aligned}|\varphi_{n+1}(t) - \psi_{n+1}(t)| ={}& |E\{[e^{it'B_nU_n} - \psi_n(B_nt)](1 - \tfrac{1}{2}n^{-\alpha}t'\Sigma t) \\ &+ e^{it'B_nU_n}[e^{itn^{-\alpha/2}V_n} - 1 + \tfrac{1}{2}n^{-\alpha}t'\Sigma t]\}| \\ \leqq{}& |1 - \tfrac{1}{2}n^{-\alpha}t'\Sigma t|\,|\varphi_n(B_nt) - \psi_n(B_nt)| + \|t\|\,n^{-\alpha}h_n.\end{aligned}$$

If $|\varphi_n(t) - \psi_n(t)| \leqq \Delta_n \|t\|$ then an analogous relation holds for $n + 1$, with $\Delta_{n+1} = \|B_n\| \Delta_n + n^{-\alpha}h_n = (1 - n^{-\alpha}\lambda)\Delta_n + n^{-\alpha}h_n$ for sufficiently large n. Lemma 2.1 or Chung's lemma imply $\Delta_n \to 0$, and this implies $\varphi_n(t) - \psi_n(t) \to 0$ for every t.

Hence the asymptotic distribution of U_n is completely determined by α, Λ, Σ. The proof is completed by observing that $U_n \sim \mathfrak{N}$ in the particular case when $V_1, V_2, \cdots$ are independent and $\mathfrak{N}(0, \Sigma)$. In this case U_n is normal, $EU_n = 0$ and $EU_{n+1}U_{n+1}' = (I - n^{-\alpha}\Lambda)EU_nU_n'(I - n^{-\alpha}\Lambda) + n^{-\alpha}\Sigma \to M$, the convergence following from a coordinatewise application of Chung's lemma.

3. Examples of application. First we shall state (with slight changes) the assumptions for the modification of the Robbins-Monro method, proposed by Venter (1967). In this procedure the goal is to approximate a zero point θ of a function h, using estimates of $h'(\theta)$ to achieve the best asymptotic speed. For the original proof, motivation and comments see Venter (1967). Note that (4) is much less restrictive than Venter's analogous condition which requires a knowledge of an interval containing d and not containing 0, and that the slight change in the definition of A_n simplifies considerations by making Y_n conditionally $(\mathfrak{F}_n)$ independent of d_n.

3.1. ASSUMPTION. h is a function defined on R, there is a $\theta \varepsilon R$ such that for any closed finite intervals $I \subset (-\infty, 0)$, $J \subset (0, +\infty)$ we have $\sup h(I) < 0$ and $\inf h(J) > 0$. The function h has a bounded second derivative in a neighborhood of θ, $d = h'(\theta) > 0$ and there are $A, B \varepsilon R$ such that

$$|h(x)| \leqq A|x - \theta| + B. \tag{3.1.1}$$

3.2. THEOREM. *Let Assumption* 3.1 *hold, let* X_n, Y_n, Z_n *be random variables'* $\mathfrak{F}_n = \sigma(X_1, Y_1, \cdots, Y_{n-1}, Z_1, \cdots, Z_{n-1})$ (i.e. *the smallest σ-field with respect to which the indicated variables are measurable*), $M_n = E_{\mathfrak{F}_n}Y_n$, $N_n = E_{\mathfrak{F}_n}Z_n$, $V_n = Y_n - M_n$, *let*

$$X_{n+1} = X_n - d_n n^{-1}Y_n, \tag{3.2.1}$$

$$M_n = \tfrac{1}{2}[h(X_n + c_n) + h(X_n - c_n)], \tag{3.2.2}$$

$$N_n = (2c_n)^{-1}[h(X_n + c_n) - h(X_n - c_n)]$$

with

$$c_n = cn^{-\gamma}, \qquad \tfrac{1}{4} < \gamma < \tfrac{1}{2}. \tag{3.2.3}$$

Let $A_n = (n-1)^{-1}\sum_{j=1}^{n-1} Z_j$, *and, with* $\vee$, $\wedge$ *denoting maximum and minimum, respectively,*

$$d_n = (C_1(\log(n+1))^{-1} \vee A_n^{-1}) \wedge C_2 n^{\alpha} \tag{3.2.4}$$

with $0 < C_1 < C_2$, $0 < \alpha < \frac{1}{2}$.

Let for a $\sigma > 0$ *and a* $C > 0$

$$E_{\mathfrak{F}_n}(Z_n - N_n)^2 \leqq Cc_n^{-2}, \quad E_{\mathfrak{F}_n}V_n^2 \leqq C, \tag{3.2.5}$$

and $E_{\mathfrak{F}_n}\chi\{V_n^2 \geqq rn\}V_n^2 = s_{n,r}^2(X_n)$ *with*

$$s_{n,0}^2(x_n) \to \tfrac{1}{2}\sigma^2, \quad n^{-1}\textstyle\sum_{j=1}^n s_{j,r}^2(x_n) \to 0 \tag{3.2.6}$$

for every sequence $x_n \to \theta$ *and every* $r > 0$.

Then $X_n \to \theta$, $d_n \to d^{-1}$ *and* $n^{\frac{1}{2}}(X_n - \theta)$ *is asymptotically normal* $(0, \frac{1}{2}\sigma^2 d^{-2})$.

PROOF. That $X_n \to \theta$ can be established easily, either as in Venter (1967), using a modification of Blum's (1954) argument, or Dvoretzky's (1956) proof for the RM method can be applied almost without any change when we note that assumption $X_{n+1} = T_n + Y_n$ in Dvoretzky's theorem can be weakened to $|X_{n+1}| \leqq |T_n + Y_n|$ (see relation (6) in Derman and Sacks (1959) and the proof there). The only assumption concerning d_n which is needed for this part of proof is that d_n lie in bounds given implicitly by (4).

Setting $\xi_n = Z_n - N_n$ we get from (3) and (5) that $\sum_{j=1}^{\infty}(\log j)^2 j^{-2}E\xi_j^2 < +\infty$, ξ_j are uncorrelated and (e.g. by Theorem 5.2, IV in Doob (1953)) $n^{-1}\sum_{j=1}^n \xi_j \to 0$. Thus the limit of A_n is that of $n^{-1}\sum_{j=1}^n N_j$, which by (2) is d. Hence $d_n \to d^{-1}$.

Assuming for simplicity $\theta = 0$, using (2), and expanding h up to the second derivative, we obtain that $M_n = m_n X_n + O(c_n^2)$ and $m_n \to d$. This makes it possible to rewrite (1) as

$$X_{n+1} = (1 - n^{-1}\Gamma_n)X_n + n^{-1}\Phi_n V_n + n^{-3/2}T_n \tag{3.2.7}$$

where $\Gamma_n = m_n d_n \to 1$, $\Phi_n = -d_n \to -d^{-1}$, $T_n = n^{\frac{1}{2}}O(c_n^2) \to 0$. We may then apply Theorem 2.2 with $U_n = X_n$, $k = 1$, $\alpha = 1$, $\beta = 1$, $\Gamma = 1$, $\Phi = -d^{-1}$, $P = 1$, $\Sigma = \frac{1}{2}\sigma^2$, $T = 0$. We verify easily the measurability assumption and (2.2.1). Condition (2.2.2) follows from the definition of V_n, from (5) and the first part of (6). The second part of (6) implies (2.2.4). Hence $n^{\frac{1}{2}}X_n$ is asymptotically normal with asymptotic mean 0 and variance $\frac{1}{2}d^{-2}\sigma^2/(2 - 1) = \frac{1}{2}d^{-2}\sigma^2$.

3.3. REMARK. Next we shall show the application to the multidimensional KW procedure in its modified form proposed by Fabian (1967), (1968); we shall refer for a moment to the first paper by the symbol I. Suppose the assumptions of Theorem I.5.3 hold. Then we can choose γ so as to have (3.4.1), (3.4.2) and X_n converges to θ. The rest in (3.4.3) will follow under assumptions analogous to (3.2.6). Relation (3.4.4) follows from (I.3.1.4) and from the expression for Q_n preceding Remark I.3.2, if we assume continuity of D_{s+1} at θ and if

$$A = H(\theta),\ m = 2((s+1)!)^{-1}c^s \textstyle\sum_{i=1}^{s/2} u_i v_i^{s+1} D_{s+1}(\theta). \tag{3.3.1}$$

3.4. THEOREM. *Let* k *be a positive integer,* X_n, $Y_n \in \mathrm{R}^k$, $\Sigma, A, P \in R^{k\times k}$, A *positive definite,* P *orthogonal,* $P'AP = \Lambda$ *diagonal,* $\lambda = \min \Lambda^{(ii)}$, *let* m, $\theta \in R^k$, $0 < \beta < 2\lambda a$,

$$X_{n+1} = X_n - a_n Y_n, \tag{3.4.1}$$

$\mathfrak{F}_n = \sigma(X_1, Y_1, Y_2, \cdots, Y_{n-1})$, $M_n = E_{\mathfrak{F}_n}Y_n$, $V_n = c_n(Y_n - M_n)$, $a > 0$, $c > 0$, $C > 0$

(3.4.2) $$a_n = an^{-1}, \qquad c_n = cn^{-\gamma}, \qquad \gamma = \tfrac{1}{2}(1 - \beta);$$

(3.4.3) $$X_n \to \theta, \qquad C > \|E_{\mathfrak{F}_n} V_n V_n' - \Sigma\| \to 0, \qquad n^{-1} \sum_{j=1}^{n} \sigma_{j,r}^2 \to 0$$

for every $r > 0$, *with* $\sigma_{j,r}^2$ *as in Theorem* 2.2; *and let for* X_n *in a neighborhood of* θ

(3.4.4) $$\|M_n - A(X_n - \theta) - n^{-\beta/2} m\| \leqq o(1)[n^{-\beta/2} + \|X_n - \theta\|].$$

Then the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ *is normal with mean* $-a(aA - (\beta/2)\mathrm{I})^{-1}m$ *and covariance matrix* PMP' *with*

$$M^{(ij)} = a^2 c^{-2} [P' \Sigma P]^{(ij)} / (a\Lambda^{(ii)} + a\Lambda^{(jj)} - \beta).$$

PROOF. Suppose $\theta = 0$. From (4), $M_n = A_n X_n + n^{-\beta/2} m_n$ with A_n, m_n being $\mathfrak{F}_n$-measurable and uniformly converging to A, m, respectively. Then

(3.4.5) $$X_{n+1} = (I - an^{-1} A_n) X_n - ac^{-1} n^{-(1+\beta)/2} V_n - an^{-1-\beta/2} m_n$$

and Theorem 2.2 may be applied with $\Gamma = aA$, $\Phi = -ac^{-1}I$, $T = -am$, $U_n = X_n$, giving the desired result.

REFERENCES

[1] BLUM, J. R. (1954). Multidimensional stochastic approximation methods. *Ann. Math. Statist.* **25** 737–744.

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

[3] DERMAN, C. and SACKS, J. (1959). On Dvoretzky's stochastic approximation theorem. *Ann. Math. Statist.* **30** 601–606.

[4] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[5] DVORETZKY, A. (1956). On stochastic approximation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 39–56. Univ. of California Press.

[6] FABIAN, V. (1967). Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist.* **38** 191–200.

[7] FABIAN, V. (1968). On the choice of design in stochastic approximation methods. *Ann. Math. Statist.* **39** 457–465.

[8] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*. **2**. New York.

[9] SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373–405.

[10] VENTER, J. H. (1967). An extension of the Robbins-Monro procedure. *Ann. Math. Statist.* **38** 181–190.