The Annals of Statistics
1973, Vol. 1, No. 6, 1175-1184

ON OPTIMAL ESTIMATION METHODS USING STOCHASTIC APPROXIMATION PROCEDURES[1,2]

BY DAN ANBAR[3]

University of California, Berkeley

The problem of estimating the zero of a regression function by means of Robbins Monro type of stochastic approximation procedures is discussed. Optimality of the procedures is defined in terms of asymptotic variance. The discussion is restricted to the case of identically distributed errors. In that case we suggest transforming the observed random variables in order to minimize the asymptotic variance of the estimators. The optimal transformation turns out to depend on the underlying distribution of the errors and on the slope of the regression function at the zero.

0. Summary. The problem of estimating the zero of a regression function by means of Robbins–Monro type of stochastic approximation procedures, is discussed.

Optimality of the procedures is defined in terms of asymptotic variance.

The discussion is restricted to the case of identically distributed errors. In that case we suggest to transform the observed random variables in order to minimize the asymptotic variance of the estimators. The optimal transformation turns out to depend on the underlying distribution of the errors and on the slope of the regression function at the zero.

1. Introduction. Investigations concerning stochastic approximation procedures were initiated by the work of Robbins and Monro [9]. Since their pioneer work, a great deal of effort has been made by many authors to generalize and improve the results of Robbins and Monro and to study the properties of the procedures.

In this work we will be concerned with the Robbins–Monro (RM) procedure.

Let $M(x)$ be a real-valued measurable function. Let θ be the unique solution

---

Received October 1971; revised June 1972.

[1] This research is based on the author's Ph. D. dissertation submitted at the University of California at Berkeley, June 1971.

This research was sponsored by the Office of Naval Research, under Contract, NONR14-69-A-0200-1038.

[2] The results in this paper were obtained independently by Sami N. Abdelhamid [1], who applied similar methods to a more general situation which includes also the Kiefer-Wolfowitz case.

[3] This work was partially done while the author was visiting at Princeton University. The author is currently at Tel Aviv University, Israel.

Key words and phrases. Robbins-Monro, slope of a regression function, stochastic approximation, optimal estimation, nonlinear regression, asymptotic optimality, sequential design, asymptotic optimal sequential design.

of the equation

$$M(x) = \theta .$$

Let $\{F_x\}$ be a family of distribution functions, and let $Y_x \sim F_x$ be a random variable with

$$EY_x = M(x)$$

for all x. Let X_1 be an arbitrary real number (in fact X_1 can be taken to be any random variable. See Sacks [10]). For $n \geqq 1$ define

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

where $\{a_n, n \geqq 1\}$ is a sequence of nonnegative real numbers satisfying

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

and Y_n is a random variable whose conditional distribution given $(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$ is F_{x_n}.

It is known that under certain conditions the process defined by (1) converges to θ almost surely and in the square mean (Blum [4]). Sacks [10] had shown that under further conditions, and for the choice

$$a_n = An^{-1} , \qquad n = 1, 2, \cdots, \tag{3}$$

$n^{\frac{1}{2}}(X_n - \theta)$ converges in law to a normal random variable with mean zero and variance

$$A^2\sigma^2(2A\alpha - 1)^{-1} \tag{4}$$

where α is the derivative of M at θ and $\sigma^2 = \lim_{x\to\theta} \mathrm{Var}\,(Y_x)$, which is assumed to be finite.

Clearly the user of the RM procedures would like to minimize the asymptotic variance of the procedure he uses.

In this work we deal with the problem of minimizing the asymptotic variance of the Robbins–Monro procedures by means of transforming the observed random variables. Our approach is as follows:

Consider a class $\mathscr{G}$ of Borel measurable transformations from the real line to itself such that the regression functions $M_g(x) = Eg(Y_x)$ have a unique zero at the same point θ as $M(x)$ for all g in $\mathscr{G}$. We ask the question: For a given family of underlying distribution functions F_x, what is the function g in $\mathscr{G}$, and the corresponding $A(g, F)$, which minimizes the asymptotic variance of the process

$$X_{n+1} = Xn - A_n^{-1}g(Y_n) , \qquad n = 1, 2, \cdots \tag{5}$$

with $X_1 = x_1$ an arbitrary real number?

We solve this problem for the case of the translation-parameter family of underlying distribution functions. Namely, we assume that $F_x(y) = F(y - M(x))$ for a given distribution function F. We define the family $\mathscr{G}$ (Section 3) and show that if F possesses a differentiable density function f w.r. to Lebesgue measure and the function $d/dx \log f(x)$ is in $\mathscr{S}$, then it is the (unique) solution of the minimization problem. Unfortunately it turns out that the corresponding optimal A depends on the slope of the regression function at $x = \theta$. Of course, we would not like to assume that the slope is known. This led us to look for a proper estimator of the slope. Such an estimator is suggested and its properties discussed.

Various authors have dealt with problems similar to the above-mentioned one in processes related to Robbins and Monro's.

Albert and Gardner [2] investigated procedures under a somewhat more general setup than that of Robbins and Monro's, but their approach is different.

The part of Albert and Gardner's work which related most closely to our work is the one which deals with asymptotic efficiency, and its improvement via transformation of the parameter space. They apply a certain transformation to the parameter space (the interval to which θ is assumed to belong). Then they construct a process on the new parameter space and show that the induced process on the original space is asymptotically normally distributed with asymptotic variance smaller than that of a corresponding process defined on the original parameter space. The transformation and the improved process depend explicitly on the functions F_n, the expectations of the observed random variables Y_n.

It is also shown that the improved process is most efficient (in the sense that it has a minimal asymptotic variance) if and only if the errors are normally distributed.

In the Robbins–Monro setup, it is not assumed that the function M is known and therefore Albert and Gardner's results concerning the improvement of the efficiency do not apply.

Some of Albert and Gardner's results, such as Theorem 4.1 on page 39 of [2], could be employed for our purposes. However, since their setup is more general than ours, we would obtain more satisfactory results by applying our methods which are, of course, suited to our problem.

The necessity to estimate the slope α in order to improve the efficiency was first recognized by Venter [11]. Venter's results were extended and generalized by Fabian [5]. Both Venter and Fabian constructed their estimators of the slope by taking pairs of observations at each stage and then estimating the slope in a natural way. A different approach was used by the author in [3]. The author has constructed a stochastic approximation analogue to the usual linear regression estimator of the slope. This way there is no need to take the observations by pairs.

All three authors mentioned above proved that their estimators were strongly consistent and were able to construct modified RM procedures which were asymptotically most efficient.

2. Preliminaries and assumptions. Throughout this work we will be dealing with the Robbins–Monro procedure with $a_n = An^{-1}$. When we refer to the Robbins–Monro (RM) procedures, we will mean the procedures using the Y's from the original underlying distribution, and we will use the term Transformed RM (TRM) procedures for the ones defined by (5).

The following assumptions will be referred to repeatedly (compare with Sacks [10]).

ASSUMPTION (A 1). *M is a Borel-measurable function; $M(\theta) = 0$ and $(x - \theta)M(x) > 0$ for all $x \neq \theta$.*

ASSUMPTION (A 2). *For some positive constants K, K_1 and for all x*

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

ASSUMPTION (A 3). *$M(x) = \alpha(x - \theta) + \delta(x, \theta)$ for all x where $\delta(x, \theta) = o(|x - \theta|)$ as $x - \theta \to 0$, and $\alpha > 0$.*

ASSUMPTION (A 4). (a) $\operatorname{Sup}_x EZ^2(x) < \infty$. (b) $\lim_{x \to \theta} EZ^2(x) = \sigma^2$ *where* $Z(x) \equiv Y_x - M(x)$.

ASSUMPTION (A 5).

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

Under Assumptions (A 1) (A 2) and (a) of (A 4) (in fact (A 2) is replaced by a slightly weaker condition), Blum [4] has proved that X_n as defined by (1) converges to θ a.s. and in square mean. Sacks [10] proved the following:

THEOREM (Sacks). *Suppose Assumptions* (A 1) *through* (A 5) *are satisfied. If A is such that $2AK > 1$, then $n^{\frac{1}{2}}(X_n - \theta)$ converges in law to a normal variable with mean zero and variance*

$$A^2\sigma^2(2A\alpha - 1)^{-1} .$$

3. Optimization problem. In this section we assume that $\{F_x\}$ is a translation parameter family of distribution functions, which are absolutely continuous with respect (w.r.) to Lebesgue measure, i.e.,

$$F_x(y) = F(y - M(x))$$

and F possesses a density function f w.r. to Lebesgue measure.

Let $\mathscr{F}$ be a family of distribution functions which are absolutely continuous w.r. to Lebesgue measure, and such that

$$0 > \int_{-\infty}^{\infty} \left[\frac{d}{dy} \log f(y)\right]^2 f(y)\, dy < \infty \tag{6}$$

for all F in $\mathscr{F}$.

Let $\mathscr{G}$ be a class of Borel-measurable functions g satisfying

(G_1): $M_g(x) = \int_{-\infty}^{\infty} g(t)f(t - M(x))\, dt$ satisfies conditions (A 1), (A 2), and (A 3) with $M(x)$ replaced by $M_g(x)$, for all $F \in \mathscr{F}$.

(G_2): $Z_g(x) = g(Y_x) - M_g(x)$ satisfies conditions (A4), (A5), with $Z(x)$ replaced by $Z_g(x)$ for all $F \in \mathscr{F}$.

(G_3): (i) $d/dx \int_{-\infty}^{\infty} g(y)f(y+x)\,dy|_{x=0} = \int_{-\infty}^{\infty} g(y)f'(y)\,dy$,
(ii) $\int_{-\infty}^{\infty} g(y)f'(y)\,dy < 0$,
(iii) $\int_{-\infty}^{\infty} g^2(y+x)f(y)\,dy$ is continuous at $x = 0$, for all $F \in \mathscr{F}$.

REMARK 1. Sufficient conditions for (i) to hold are the following:

(i′) (a) The left-hand side of (i) of G_3 exists.
(b) $\int_{-\infty}^{\infty} |g(y)f'(y+x)|\,dy$ is bounded in some neighborhood of $x = 0$, and
(c) $\int_{-\infty}^{\infty} g(y)f'(y+x)\,dy$ is continuous at $x = 0$.

To see this let us compute the left-hand side of (i):

$$\begin{aligned}\frac{d}{dx}\int_{-\infty}^{\infty} g(y)f(y+x)\,dy\,|_{x=0} &= \lim_{\delta\to 0}\frac{1}{\delta}\int_{-\infty}^{\infty} g(y)[f(y+\delta)-f(y)]\,dy\\ &= \lim_{\delta\to 0}\frac{1}{\delta}\int_{-\infty}^{\infty}\{g(y)\int_y^{y+\delta} f'(t)\,dt\}\,dy\\ &= \lim_{\delta\to 0}\frac{1}{\delta}\int_{-\infty}^{\infty}\{g(y)\int_0^{\delta} f'(y+t)\,dt\}\,dy\,.\end{aligned}$$

By (a) and (b) of (i′), Fubini's theorem applies and we have

$$\begin{aligned}&= \lim_{\delta\to 0}\frac{1}{\delta}\int_0^{\delta}\int_{-\infty}^{\infty}\{g(y)f'(y+t)\,dy\}\,dt\\ &= \int_{-\infty}^{\infty} g(y)f'(y)\,dy \quad \text{by (c) of (i′)}\,.\end{aligned}$$

REMARK 2. If $\mathscr{F}$ consists of all symmetric distribution functions such that (c) holds and $f'(y) < 0$ for $y > 0$, then (ii) of (G_3) holds for all odd functions g for which $yg(y) \geqq 0$ for all y. We mention this particular example because of the bearing it has on the minimax result of Section 4. There we consider the family of ε-contaminated normal distribution functions with a symmetric contaminating distribution H. Conditions (G_1), (G_2), and (G_3) guarantee that the process (5) converges a.s. to θ and that $n^{\frac{1}{2}}(X_n - \theta)$ converges in law to a normal random variable with mean zero and variance

$$\sigma^2_{A,g} = \frac{A^2 \lim_{x\to\theta} E[g(Y_x) - M_g(x)]^2}{2A[(d/dx)M_g(x)]_{x=0} - 1}\,. \tag{7}$$

In our shift parameter case it follows from (iii) of (G_3) that

$$\begin{aligned}\lim_{x\to\theta} E[g(Y_x) - M_g(x)]^2 &= \lim_{x\to\theta}\{\int_{-\infty}^{\infty} g^2(y+M(x))f(y)\,dy - M_g^{\,2}(x)\}\\ &= \int_{-\infty}^{\infty} g^2(y)f(y)\,dy\,.\end{aligned}$$

$$\frac{d}{dx}M_g(x) = \frac{d}{dx}\int_{-\infty}^{\infty} g(y)f(y - M(x))\,dy\,.$$

By (i) of (G_3) we may differentiate under the integral sign and may interchange integration and limit operations. Thus,

$$\frac{d}{dx} M_g(x)|_{x=0} = -\alpha \int_{-\infty}^{\infty} g(y)f'(y)\, dy ,$$

which is positive by (ii) of (G_3). Therefore we have:

$$\sigma^2_{A,g} = \frac{A^2 \int_{-\infty}^{\infty} g^2(y)f(y)\, dy}{-2A\alpha \int_{-\infty}^{\infty} g(y)f'(y)\, dy - 1} . \tag{8}$$

Considering $\sigma^2_{A,g}$ as a function of A, we obtain by minimizing w.r. to A

$$\sigma^2_{A,g} \geqq \frac{\int_{-\infty}^{\infty} g^2(y)f(y)\, dy}{\alpha^2\{\int_{-\infty}^{\infty} g(y)f'(y)\, dy\}^2} . \tag{9}$$

Now, by the Cauchy—Schwarz inequality

$$[\int_{-\infty}^{\infty} g(y)f'(y)\, dy]^2 \leqq [\int_{-\infty}^{\infty} g^2(y)f(y)\, dy] \cdot \left[\int_{-\infty}^{\infty} \left(\frac{f'(y)}{f(y)}\right)^2 f(y)\, dy\right] .$$

Thus,

$$\sigma^2_{A,g} \geqq \frac{1}{\alpha^2 \int_{-\infty}^{\infty} [f'(y)/f(y)]^2 f(y)\, dy} = \frac{1}{\alpha^2 I_0^2(f)} , \tag{10}$$

where $I_0^2(f)$ is the Fisher information number. From (10) we obtain immediately the following:

THEOREM 1. *If the underlying distribution function F is absolutely continuous* w.r. *to Lebesgue measure such that* (6) *holds, and if*

$$g_0(y) = c \frac{d}{dy} [\log f(y)] \tag{11}$$

is in $\mathscr{G}$ for some $c \neq 0$, the g_0 minimizes $\alpha^2_{A,g}$. Furthermore, g_0 is unique up to a multiplicative constant.

PROOF. To show that g_0 minimizes $\sigma^2_{A,g}$ it is sufficient to show that

$$\sigma^2_{A,g_0} = \frac{1}{\alpha^2 I_0^2(f)} , \qquad \text{for some } A.$$

But

$$\int_{-\infty}^{\infty} g_0^2(y)f(y)\, dy = c \int_{-\infty}^{\infty} g_0(y)f'(y)\, dy .$$

Therefore,

$$\frac{\int_{-\infty}^{\infty} g_0^2(y)f(y)\, dy}{\{\int_{-\infty}^{\infty} g_0(y)f'(y)\, dy\}^2} = \frac{c}{\int_{-\infty}^{\infty} g_0(y)f'(y)\, dy} = \frac{1}{I_0^2(f)} .$$

Therefore

$$\sigma_F{}^2 = \sigma^2_{A_0,g_0} = \frac{1}{\alpha^2 I_0^2(f)} \tag{12}$$

where

$$A_0 = -\frac{1}{\alpha \int_{-\infty}^{\infty} g_0(y)f'(y)\, dy} > 0 . \tag{13}$$

Uniqueness follows from the fact that equality in the Cauchy–Schwarz inequality is obtained if and only if there exist constants $c \neq 0$ and d such that

$$g(y) = c\frac{f'(y)}{f(y)} + d\,.$$

But the requirement $g \in \mathscr{G}$ implies $\int_{-\infty}^{\infty} g(y)f(y)\,dy = 0$. Since

$$\int_{-\infty}^{\infty} \frac{f'(y)}{f(y)} f(y)\,dy = \int_{-\infty}^{\infty} f'(y)\,dy = 0\,,$$

it follows that $d = 0$. This completes the proof.

REMARK. A_0 as defined in (12) can be rewritten as

$$A_0 = -\frac{1}{\alpha c I_0^2(f)}\,. \tag{14}$$

This implies that c must be taken to be negative. The only condition which is affected by the sign of c is Assumption (A 1). It is therefore of interest to know whether (A 1) is consistent with the requirement that c should be negative, for some interesting distribution functions.

DEFINITION. A distribution function F is called (strictly) strongly unimodal if it possesses a density f w.r. to Lebesgue measure, and $-\log f(x)$ is a (strictly) convex function in some open interval (a, b) such that $-\infty \leqq a < b \leqq \infty$ and $\int_a^b f(x)\,dx = 1$.

THEOREM 2. *If F is strictly strongly unimodal and* (6) *holds, then g_0 defined by* (11) *with negative c, satisfies* (A 1).

PROOF.

$$\begin{aligned} M_{g_0}(x) &= \int_{-\infty}^{\infty} g_0(y)f(y - M(x))\,dy \\ &= c\int_{-\infty}^{\infty} \frac{f'(y)}{f(y)} f(y - M(x))\,dy\,. \end{aligned}$$

By strict strong unimodality $f'(y)/f(y)$ changes sign exactly once and

$$\{f_\theta(y)\} = \{f(y - \theta)\}$$

is a family with monotone likelihood ratio (see e.g. Lehmann [8] page 330, Example 1). Therefore as Karlin and Rubin proved in [7], $M_{g_0}(x)$ changes sign exactly once at $x = \theta$. The fact that $c < 0$ guarantees that (A 1) holds.

REMARK. (a) In the proof of Theorem 1 we have already indicated that if we use the transformation g_0, then the optimal choice of A is given by (13) and the smallest asymptotic variance by (12). In fact the argument which led us to (9) indicates that if we use any transformation g in $\mathscr{G}$, then the best choice of A for this g is

$$A_g = -[\alpha \int_{-\infty}^{\infty} g(y)f'(y)\,dy]^{-1}\,, \tag{15}$$

which leads to a process with asymptotic variance given by the right-hand side of (9).

(b) As we have seen, the optimal transformation g_0 is determined uniquely up to a constant c. However, the asymptotic variance (12), is independent of c. The independence from c becomes clear if we notice that the right-hand side of (9) is scale-invariant. We may, therefore, take $c = -1$.

EXAMPLES. (i) *Normal distribution.*

$$f(y) = (2\pi)^{-\frac{1}{2}} \exp\{-y^2/2\}, \qquad -\infty < y < \infty .$$

$$I_0^2(f) = 1$$

and therefore

$$g_0(y) = y \qquad \text{and} \qquad A_0 = \alpha^{-1} .$$

It is clear that $g_0(y) = y$ is in fact in $\mathscr{G}$ and therefore, if the underlying distribution is normal, the RM procedure is optimal.

(ii) *Double exponential.*

$$f(y) = \frac{1}{2} \exp\{-|y|\}, \qquad -\infty < y < \infty .$$

In this case

$$\frac{d}{dy} \log f(y) = -1 \qquad \text{if} \quad y > 0$$

$$= 1 \qquad \text{if} \quad y < 0$$

and

$$\left\{\frac{d}{dy} \log f(y)\right\}^2 = 1 \qquad \text{a.s.}$$

and therefore $I_0^2(f) = 1$. Hence the function under consideration is:

$$g_0(y) = -\frac{d}{dy} \log f(y) = \text{sign}(y) \qquad \text{and} \qquad A_0 = \alpha^{-1} .$$

In this case g_0 satisfies:

$$\int_{-\infty}^{\infty} |g_0(y) f'(y + a)|\, dy \equiv 1 \qquad \text{for all} \quad a$$

$$\int_{-\infty}^{\infty} g_0(y) f'(y)\, dy = -1 < 0$$

and

$$\int_{-\infty}^{\infty} g^2(y + a) f(y)\, dy \equiv 1 \qquad \text{for all} \quad a .$$

Therefore it is clear that $g_0 \in \mathscr{G}$ and is the optimal transformation.

(iii) *Logistic distribution.*

$$f(y) = e^{-y}/(1 + e^{-y})^2 \qquad -\infty < y < \infty$$

$$-\log f(y) = y + 2 \log(1 + e^{-y}),$$

$$\frac{d}{dy}(-\log f(y)) = 1 - 2e^{-y}/(1 + e^{-y})$$

$$\frac{d^2}{dy^2}(-\log f(y)) = 2e^{-y}/(1 + e^{-y})^2 > 0$$

and therefore $-\log f(y)$ is convex and $f(y)$ is strongly unimodal. Here:

$$I_0^2(f) = \int_{-\infty}^{\infty} \left[\frac{d}{dy} \log f(y)\right]^2 f(y)\, dy$$
$$= \int_{-\infty}^{\infty} (e^{-y} - 1)^2/(e^{-y} + 1)^4 e^{-y}\, dy = \tfrac{1}{3}\,,$$

$$A_0 = 3\alpha^{-1}$$

and

$$g_0(y) = -(e^{-y} - 1)/e^{-y} + 1) = \tanh (2y)$$

which is in $\mathscr{G}$ and is optimal.

(iv) *Huber's distribution.*

(16) $$f_K(y) = Ce^{-\rho(y)}$$

where

(17) $$\rho(y) = \frac{y^2}{2} \qquad \text{if} \quad |y| \leqq K$$
$$= K|y| - \frac{K^2}{2} \qquad \text{if} \quad |y| > K\,.$$

A simple calculation shows that in this case the optimal procedure is:

(18) $$g_0(y) = y \qquad \text{if} \quad |y| < K$$
$$= K \operatorname{sign}(y) \qquad \text{if} \quad |y| > K\,.$$

4. A minimax result. In the last section we saw that for a given distribution function F, and a given transformation g, the smallest (w.r. to the choice of A) asymptotic variance was given by the right-hand side of (9). This expression is essentially the same variance expression Huber has dealt with in [6]. Therefore his minimax result applies in our case as well and we obtain:

Minimax result (Huber). If the underlying distribution F is an ε-contaminated normal, i.e., $F(t) = (1 - \varepsilon)\Phi(t) + \varepsilon H(t)$, where Φ is the normal $N(0, 1)$ distribution, and H is a symmetric distribution function, and if K is the solution of the equation:

$$(1 - \varepsilon)^{-1} = \int_{-K}^{K} \varphi(t)\, dt + 2\varphi(K)/K\,,$$

then g_0 defined by (18) determines a TRM process which minimizes the supremum of the "best" asymptotic variance (here "best" refers to the best choice of A), where the supremum is taken over the set of all ε-contaminated normal distributions.

This TRM process corresponds to the underlying distribution with density (16). The corresponding function is therefore the least favorable distribution.

Acknowledgment. I wish to express my deep gratitude to Professor Peter J. Bickel under whose supervision and guidance this research was done. I always found him available for discussions from which I greatly benefited.

REFERENCES

[1] Abdelhamid, Sami N. (1971). Transformation of observations in stochastic approximation. Ph. D. dissertation, Michigan State Univ.

[2] Albert, A. E. and Gardner, L. A. (1967). Stochastic approximation and nonlinear regression. Research monograph No. 42. The M.I.T. Press, Cambridge.

[3] Anbar, D. (1971). On optimal estimation methods using stochastic approximaion procedures. Unpublished Ph. D. dissertation, Univ. of California, Berkeley.

[4] Blum, J. R. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25** 382-386.

[5] Fabian, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Statist.* **39** 1327-1332.

[6] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

[7] Karlin, S. and Rubin, H. (1956). The theory of decision procedures of distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272-299.

[8] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley New, York.

[9] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400-407.

[10] Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373-405.

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

Department of Statistics
Tel Aviv University
Ramat Aviv
Tel Aviv, Israel