Journal of Statistical Planning and Inference 2 (1978) 153–163.

A STOCHASTIC NEWTON–RAPHSON METHOD

Dan ANBAR*
Case Western Reserve University and *Tele Aviv University*

Received 26 July 1977
Recommended by Shelley Zacks

Abstract: A stochastic approximation procedure of the Robbins–Monro type is considered. The original idea behind the Newton–Raphson method is used as follows. Given n approximations $X_1,\ldots,X_n$ with observations $Y_1,\ldots,Y_n$, a least squares line is fitted to the points $(X_m, Y_m),\ldots,(X_n, Y_n)$ where $m<n$ may depend on n. The $(n+1)$st approximation is taken to be the intersection of the least squares line with $y=0$. A variation of the resulting process is studied. It is shown that this process yields a strongly consistent sequence of estimates which is asymptotically normal with minimal asymptotic variance.

AMS 1970 subject classification: 62L20, 62L05.

Key words and phrases: Stochastic Approximation, Robbins–Monro Procedure, Efficient Estimation.

1. Introduction

The Newton–Raphson (NR) method for approximating the zero of a function is probably the best known approximation method. The idea is a very simple one. Let $M(x)$ be a differentiable function and θ be the unique solution of the equation $M(x)=0$. Pick a point x_1 arbitrarily, draw the tangent line to M at $x=x_1$ and let x_2 be the point of intersection of this line with the x-axis. Continuing in this manner one obtains a sequence of points $x_1, x_2, \ldots$ where $x_{n+1}=x_n-M(x_n)/M'(x_n)$ $n=1,2,\ldots$. When the function M satisfies some regularity conditions, the sequence $\{x_n\}$ converges to the root θ.

The first attempt to consider a stochastic version of the NR method was made by Robbins and Monro (1951). They considered the case where M can only be observed statistically. Namely, at each point x one can observe a random variable $Y(x)$ such that $\mathrm{E}\,Y(x)=M(x)$. They suggested the following procedure. Let x_1 be a real number. Let $X_1=x_1$ and for $n=1,2,\ldots$, $X_{n+1}=X_n-a_nY_n$ where Y_n is a random variable with conditional distribution given $(X_1=x_1, X_2=x_2,\ldots, X_n=x_n)$ equals to the distribution of $Y(x_n)$ and $\{a_n\}$, $n=1,2,\ldots$ is a sequence of real numbers. This process is known as the Robbins–Monro (RM) process. The analogy between the RM process and the NR method is clear. The function M evaluated at x_n is replaced in the RM case by its estimate Y_n and the sequence of derivatives by a fixed sequence of real numbers. It was shown that under certain

*Supported by the Office of Naval Research under Contract N00014-75-C-0529, Project NR 042-276 at Case Western Reserve University.

regularity conditions and a proper choice of the sequence $\{a_n\}$, the RM process converges to θ a.s. and in the mean square. (e.g. Robbins and Monro (1951), Blum (1954), Dvoretzky (1956), Robbins and Siegmund (1971)). Chung (1954) and later Sacks (1958) have found conditions under which the RM is asymptotically normal. Sacks (1958) investigated the case where $a_n = An^{-1}$, $n = 1, 2, \dots$, where A is some positive constant. He has shown that under some general conditions the sequence $n^{1/2}(X_n - \theta)$ converges in law to a normal random variable with mean zero and variance $A^2\sigma^2/(2A\alpha - 1)$ where $\sigma^2 = \lim_{x\to\theta} \mathrm{Var}\, Y(x)$ and $\alpha > 0$ is the derivative of M at $x = \theta$. It is immediate to check that the asymptotic variance is minimized when $A = \alpha^{-1}$. However, without the knowledge of M and θ, α cannot be computed. Thus the problem of constructing efficient process (i.e. with minimal asymptotic variance) for estimating θ becomes natural. This problem was considered first by Albert and Gardner (1967) and by Sakrison (1965).[1] Both Albert and Gardner and Sakrison have replaced the constant A by a stochastic sequence estimating α^{-1}. Since in both cases the estimating sequence depends on M, their methods are useful only when M is a known function. The case where M is unknown was considered by Venter (1967). Venter's method requires that at the nth stage ($n = 1, 2, \dots$) the statistician makes two observations Y_n' and Y_n'' at $X_n - c_n$ and $X_n + c_n$ where X_n is the nth approximation and $\{c_n\}$ is a given null sequence of positive numbers. This way of collecting the data made it possible for Venter to construct an efficient adaptive process. Although Venter's method is asymptotically efficient, the present author feels that taking the observations two at a time may make it unattractive in situations where the total number of experiments allowed is small. This raises a natural question. After the nth stage of the experiment the statistician has observed the first n approximations $X_1, X_2, \dots, X_n$ and the observations on M at these points $Y_1, Y_2, \dots, Y_n$. Is it not possible to use these data for determining the next approximation X_{n+1} in an efficient way? The purpose of this paper is to provide a positive answer. Motivated by the original NR method one might fit a least squares line to the points (X_1, Y_1), $(X_2, Y_2), \dots, (X_n, Y_n)$ and take the point of intersection of this line with the x-axis as the next approximating point. This yields the following procedure. Let X_1 be a random variable to be chosen by the statistician. For $n = 1, 2, \dots$ let

$$X_{n+1} = \bar{X}_n - \bar{Y}_n / b_{1n}, \tag{1}$$

where

$$b_{1n} = \sum_{i=1}^{n} (X_i - \bar{X}_n) Y_i \Big/ \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \qquad \text{for } n \geq 2$$

and b_{11} is an arbitrary positive number. Since M is not assumed to be linear it seems preferable to consider the process given by

$$X_{n+1} = X_n - Y_n / n b_{m(n), n-1}, \quad n > m(n) \tag{2}$$

[1] The author is indebted to V. Fabian, for pointing out this reference to him.

with X arbitrary and

$$b_{m,n}=\sum_{i=m}^{n}(X_i-\bar{X}_{m,n})Y_i\Big/\sum_{i=m}^{n}(X_i-\bar{X}_{m,n})^2 \tag{3}$$

where

$$\bar{X}_{m,n}=(n-m)^{-1}\sum_{i=m+1}^{n}X_i.$$

As indicated, m may be taken to depend upon n. This is not needed in any of the proofs. However, one might want to base the estimator (3) of α only upon the later part of the sequence $\{X_n\}$ and ignore the beginning. This can obviously be done if m is chosen to go to infinity with n. As will be seen later all the results hold in this was provided $m(n)=o((\log n)^{1/2+\varepsilon})$ for every $\varepsilon>0$.

In this paper a truncated version of (2) is considered. It is shown that this process yields a strongly consistent sequence of estimators for θ which is asymptotically efficient.

2. The results

For the purpose of easy reference let us list all the assumptions which will be needed in the sequel.

(M1) $M(x)$ is a measurable function satisfying $(x-\theta)\,M(x)>0$ for all $x\neq\theta$.

(M2) For every $0<\varepsilon<1$ $\inf_{\varepsilon<|x-\theta|<\varepsilon^{-1}}|M(x)|>0$.

(M3) $M(x)=\alpha(x-\theta)+\delta(x-\theta)$ for all x where $0<\alpha<\infty$ and $\delta(x)=o(x)$.

(M4) There exist positive numbers $0<K<K_1<\infty$ such that for all $x\neq\theta$.

(i) $|M(x)|<K_1|x-\theta|$,

(ii) $|M(x)|>K|x-\theta|$.

Let $Y(x)$ be a random variable such that $\mathrm{E}\,Y(x)=M(x)$. It is assumed that for every $-\infty<x<\infty$ the random variable $Y(x)$ is observable. Let $Z(x)=Y(x)-M(x)$.

(Z1) (i) $\sup_x \mathrm{E}|Z(x)|^{2+\delta}<c_1<\infty$ for some $\delta>0$

(ii) $\inf_x \mathrm{E}\,Z^2(x)>c_0>0$.

(Z2) $\lim_{x\to\theta}\mathrm{E}\,Z^2(x)=\sigma^2$.

(Z3) $\lim_{R\to\infty}\lim_{\varepsilon\to0+}\sup_{|x-\theta|<\varepsilon}\int_{\{|Z(x)|>R\}}Z^2(x)\,\mathrm{d}F=0$.

(Z4) There exist $\lambda>0$ and $\gamma>0$ such that for all x

$\mathrm{P}\{|Z(x)|>\lambda\}>\gamma$.

D

Remark. Condition (Z1)(ii) is obviously implied by (Z4). Both conditions are included, however, since some of the results which are of independent interest (such as Lemma 4) make use only of the weaker version (Z1)(ii).

Assumption (M4) implies that the derivative of M at $x=\theta$ satisfies $K<\alpha<K_1$. It is therefore assumed that two numbers $K\leqq\alpha_1<\alpha_2\leqq K_1$ are known and $\alpha_1<\alpha<\alpha_2$.

Let X_1 be a random variable. Define successively $X_2, X_3, \ldots$ by

$$X_{n+1}=X_n-A_{mn}n^{-1}Y_n, \quad n=1,2,\ldots, \tag{4}$$

$$m=m(n)=o((\log n)^{1/2+\varepsilon}) \quad \text{for every } \varepsilon>0,$$

where

$$A_{mn}^{-1}=\begin{cases}\alpha_1 & \text{if } b_{m,n-1}<\alpha_1,\\ b_{m,n-1} & \text{if } \alpha_1<b_{m,n-1}<\alpha_2,\\ \alpha_2 & \text{if } \alpha_2\leqq b_{m,n-1}.\end{cases} \tag{5}$$

$b_{m,n}$ is given by (3), and Y_n is a random variable with conditional distribution given $(X_1, Y_1, Y_2, \ldots, Y_{n-1})$ the same as the distribution of YX_n).

Theorem 1. *If conditions* (M1), (M2), (M4) (i) *and* (Z1) (i) *are satisfied then* X_n *given by* (4) *and* (5) *converges almost surely to* θ.

The proof is omitted since it follows from Robbins and Siegmund's Application 2 (1971, Sec. 4).

Remark. It can be easily verified that (5) can be replaced by truncating $b_{m,n}$ in (α_n, β_n) where $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of positive numbers satisfying $\Sigma\alpha_n^{-2}n^{-2}<\infty$ and $\Sigma\beta_n^{-1}n^{-1}=\infty$. e.g. $\alpha_n=c_1n^{-\beta}$ $(0<\beta<\frac{1}{2})$, $\beta_n=c_2(\log n)^{\delta}$ $(0<\delta\leqq 1)$. The proof of consistency remains unchanged. However, for our purposes (5) is suitable.

The next aim is to prove that $b_{m,n}$ is a strongly consistent estimator of α. For this purpose a few auxiliary lemmas are needed. Without loss of generality it is assumed that $\theta=0$. Also for the sake of simplicity in the notation A_{mn} will be written as A_n with the first index suppressed.

Lemma 1. *Let* $\{X_n\}$ *be given by* (4) *and* (5) *with* $\alpha_2<2K$. *If Conditions* (M1), (M4) *and* (Z1) *are satisfied then there exist constants* $0<c_3<c_4<\infty$ *such that for all* n *sufficiently large* $c_3n^{-1}<\mathrm{E}X_n^2<c_4n^{-1}$.

The proof of the right-hand side inequality follows the standard argument. The reader is referred for details to Sacks (1958) p. 382. The proof of the left-hand side inequality follows the same lines and makes use of conditions (M1), (M4) (ii) and (Z1) (ii). The restriction $\alpha_2<2K$ is not needed for the later part of the proof. The details are omitted.

Lemma 2. *Let* $\{X_n\}$ *be given by* (4) *and* (5). *Assume that the conditions of Lemma*

1 *and Conditions* (M3) *and* (Z4) *hold. If* $m=m(n)$ *is such that* $m(n)=o((\log n)^{1/2+\varepsilon})$ *as* $n\to\infty$ *for every* $\varepsilon>0$, *then* $\sum_{k=1}^{n} X_k^2\to\infty$ *a.s. at least as fast as* $\log n$.

Proof. By (M3) it follows that $M(x)$ can be written as $M(x)=(\alpha+\psi(x))x$ where $\psi(x)\to 0$ as $x\to 0$. Thus

$$X_{n+1}=(1-d_n n^{-1})X_n-A_n n^{-1}Z_n \tag{6}$$

where

$$d_n=A_n(\alpha+\psi(X_n)).$$

By Theorem 1,

$$a_1<\alpha a_2^{-1}\leqq\liminf_n d_n\leqq\limsup_n d_n\leqq\alpha\alpha_1^{-1}<a_2 \quad \text{a.s.}$$

and

$$\tfrac{1}{2}<a_1<1<a_2.$$

Thus for $n\geqq n_0=n_0(\omega)$ and almost all ω,

$$\begin{aligned} X_{n+1}^2&=(1-d_n n^{-1})^2X_n^2+A_n^2n^{-2}Z_n^2-2A_n n^{-1}(1-d_n n^{-1})X_nZ_n\\ &\geqq(1-a_2n^{-1})X_n^2+A_n^2n^{-2}Z_n^2-2A_n n^{-1}(1-d_n n^{-1})X_nZ_n. \end{aligned}\tag{7}$$

Let $\gamma_n=\prod_{j=j_0}^{n}(1-a_2j^{-1})$, $n\geqq j_0$, where j_0 is such that $1-a_2j^{-1}>0$ for all $j\geqq j_0$. Clearly $\gamma_n=h_n n^{-a_2}$ where $0<c_5\leqq h_n\leqq c_6<\infty$ uniformly in n. Let $\beta_{mn}=\gamma_m^{-1}\gamma_n$ for $n\geqq m\geqq\max(j_0,n_0)$.

Iterating (7) yields

$$X_{n+1}^2\geqq\beta_{m-1,n}^2X_m^2+\sum_{k=m}^{n}A_k^2\beta_{kn}^2k^{-2}Z_k^2-2\sum_{k=m}^{n}A_k\beta_{kn}^2k^{-1}(1-d_kk^{-1})X_kZ_k$$

for all $n\geqq m\geqq\max(j_0,n_0)$.

Thus

$$\begin{aligned}\sum_{n=m+1}^{N}X_n^2\geqq X_m^2\sum_{n=m}^{N-1}\beta_{m-1,n}^2+\sum_{n=m}^{N-1}\sum_{k=m}^{n}A_k^2\beta_{kn}^2k^{-2}Z_k^2\\ -\sum_{n=m}^{N-1}\sum_{k=m}^{n}A_k\beta_{kn}^2k^{-1}(1-d_kk^{-1})X_kZ_k.\end{aligned}\tag{8}$$

Consider first the last term of (8). By changing the order of summation one gets

$$\begin{aligned}&\sum_{n=m}^{N}\sum_{k=m}^{n}A_k\beta_{kn}^2k^{-1}(1-d_kk^{-1})X_kZ_k=\\ &=c\sum_{k=m}^{N}A_kk^{2a_2-1}(1-d_kk^{-1})h_{kN}'(k^{-2a_2+1}-N^{-2a_2+1})X_kZ_k\end{aligned}$$

$$=c\left[\sum_{k=m}^{N} A_k h'_{kN}(1-d_k k^{-1})X_k Z_k \right.$$
$$\left. -N^{-2a_2+1}\sum_{k=m}^{N} A_k k^{2a_2-1}(1-d_k k^{-1})h'_{kN} X_k Z_k\right].$$

where the h'_{kN} are uniformly bounded positive numbers.

Now

$$\mathrm{E}(A_k(1-d_k k^{-1})X_k Z_k | Z_1,\ldots,Z_{k-1})=0.$$

Furthermore

$$\mathrm{E}[(\log k)^{-(1+\varepsilon)/2}A_k(1-d_k k^{-1})X_k Z_k]^2 \leqq c'[k(\log k)^{1+\varepsilon}]^{-1}$$

by Lemma 1 and (Z1)(i) for every $\varepsilon>0$. Thus

$$\sum_{k=m}^{n} (\log k)^{-(1+\varepsilon)/2}A_k(1-d_k k^{-1})X_k Z_k$$

converges a.s. to a finite limit as $n\to\infty$ by Theorem D of Loève (1963) p. 387. Next observe that Kronecker's Lemma (Loève (1963) p. 238) can be slightly modified to read: If $\sum_n W_n = W < \infty$, $\{b_n\}$ is a monotone increasing sequence of positive numbers such that $\lim_{n\to\infty} b_n = \infty$ and h_{kn} are positive and uniformly bounded numbers, then $\lim_{n\to\infty} b_n^{-1}\sum_{k=1}^{n} h_{kn} b_k W_k = 0$. Thus,

$$\sum_{n=m}^{N}\sum_{k=m}^{n} A_k \beta_{kn}^2 k^{-1}(1-d_k k^{-1})X_k Z_k = \mathrm{o}((\log N)^{1/2+\varepsilon})$$

for every $\varepsilon>0$.

Now

$$\sum_{n=m}^{N} \beta_{m-1,n}^2 \approx cm(1-(m/N)^{2a_2-1}) = \mathrm{o}((\log N)^{1/2+\varepsilon})$$

as $N\to\infty$ for every $\varepsilon>0$.

It remains to consider the middle term of (8). Again by changing the order of summation one gets

$$\sum_{n=m}^{N}\sum_{k=m}^{n} A_k^2\beta_{kn}^2 k^{-2}Z_k^2 \geqq c\sum_{k=m}^{N} k^{2a_2-2}(k^{-2a_2+1}-N^{-2a_2+1})Z_k^2$$
$$=c\sum_{k=m}^{N} k^{-1}(1-(k/N)^{2a_2-1})Z_k^2 > c\Delta\sum_{k=m}^{[DN]} k^{-1}Z_k^2$$

where $0<\Delta<1$ and $D=(1-\Delta)^{1/(2a_2-1)}$. Let $\lambda>0$ and $\gamma>0$ be as in Assumption (Z4). Define

$$Z_k^* = \lambda \quad \text{if} \quad |Z_k|>\lambda,$$
$$\quad =0 \quad \text{if} \quad |Z_k|\leqq\lambda.$$

Clearly $\sum k^{-2} \operatorname{var} Z_k^{*2}<\infty$ hence

$$\sum k^{-1}[Z_k^{*2}-\mathrm{E}(Z_k^{*2}|Z_1,\ldots,Z_{k-1})]<\infty \quad \text{a.s.}$$

But by (Z4)

$$\mathrm{E}(Z_k^{*2}|Z_1,\ldots,Z_{k-1})>\lambda^2\gamma.$$

Hence

$$\sum_{k=m}^{[DN]} k^{-1}\mathrm{E}(Z_k^{*2}|Z_1,\ldots,Z_{k-1})\geqq c_1 \log(N/m)$$

which implies that

$$\sum_{k=m}^{[DN]} k^{-1} Z_k^{*2}\geqq c_2 \log(N/m)$$

where c_1 and c_2 are positive constants. Since $Z_k^2\geqq Z_k^{*2}$ the proof is complete.

The following result is due to Stout (1970).

Lemma 3. *Let $\{\xi_n, \mathscr{F}_n, n\geqq 1\}$ be a martingale difference sequence. Let*

$$V_n^2=\sum_{k=1}^{n} \mathrm{E}(\xi_k^2|\mathscr{F}_{k-1}),$$

$$U_n=(2\log\log V_n^2)^{1/2},$$

$$S_n=\sum_{k=1}^{n} \xi_k, \quad n=1,2,\ldots .$$

Let κ_n be $\mathscr{F}_{n-1}$ measurable functions such that $\kappa_n\to 0$ a.s. as $n\to\infty$. If $V_n^2\to\infty$ a.s. and $|\xi_n|\leqq\kappa_n V_n/U_n$ a.s., then $\limsup_{n\to\infty} S_n/U_nV_n=1$ *a.s.*

Lemma 4. *Let X_n be given by* (4) *and* (5). *If conditions* (M1) *through* (M4) *and* (Z1) *hold, then*

$$|X_n|=\mathrm{o}(\eta_n n^{-1/2}(\log\log n)^{1/2}) \qquad \text{a.s.}$$

where $\{\eta_n\}$ is any sequence of real numbers such that $\eta_n\to\infty$ as $n\to\infty$.

Proof. Let j_1 be an integer such that $1-a_1 j_1^{-1}>0$. Iterating (6) and recalling that $a_1<d_n$ a.s. for all n, one obtains

$$|X_{n+1}|\leqq\beta'_{m-1,n}|X_n|+\left|\sum_{k=m}^{n}A_k k^{-1}\beta''_{kn}Z_k\right| \quad \text{a.s.} \tag{9}$$

for every $m\geqq j_1$, where

$$\begin{aligned}\beta'_{mn}&=\prod_{j=m+1}^{n}(1-a_1 j^{-1}) \quad \text{if} \quad n>m,\\ &=1 \qquad\qquad\qquad\quad\ \text{if} \quad m=n\end{aligned}$$

and

$$\begin{aligned}\beta''_{mn}&=\prod_{j=m+1}^{n}(1-d_j j^{-1}) \quad \text{if} \quad n>m,\\ &=1 \qquad\qquad\qquad\quad\ \text{if} \quad n=m.\end{aligned}$$

Similar to the definition of γ_n, let $\gamma''_n=\beta''_{j_1 n}$. Thus $\beta''_{mn}=\gamma''_n/\gamma''_m$ for every $n\geqq m\geqq j_1$.

Notice that γ''_n is a random variable measurable with respect to $\mathscr{F}(X_1,Z_1,\ldots,Z_{n-1})$.

Since $m=m(n)=o((\log n)^{1/2+\varepsilon})$ for every $\varepsilon>0$ and since $a_1>\frac{1}{2}$, $\beta'_{m-1,n}=o((\log n)^{a_1(1/2+\varepsilon)}n^{-a_1})$, it is sufficient to consider only the second term on the right-hand side of (9).

Now apply Lemma 3 with $\xi_k=k^{-1/2+\varepsilon}Z_k$ where $\varepsilon>0$ is to be determined later. By (Z1) there exist positive constants c_1, c_2, c_3 and c_4 such that

$$c_1 n^{2\varepsilon}\leqq V_n^2\leqq c_2 n^{2\varepsilon}$$

and

$$c_3(\log\log n)^{1/2}\leqq U_n\leqq c_4(\log\log n)^{1/2}$$

for all n sufficiently large. Let $\{\kappa_n\}$ be a sequence of positive numbers such that $\kappa_n\to 0$ as $n\to\infty$ and $\sum_n \log\log n)^{1+\delta/2}\kappa_n^{-(2+\delta)}n^{-1-\delta/2}<\infty$ for every $\delta>0$. By (Z1)(i),

$$\mathbf{P}(|Z_n|>\kappa_n(n/\log\log n)^{1/2})\leqq c_1(\log\log n)^{1+\delta/2}\kappa_n^{-(2+\delta)}n^{-1-\delta/2}.$$

Therefore by the Borel–Cantelli Lemma with probability one

$$|Z_n|\leqq\kappa_n(n/\log\log n)^{1/2}$$

for all n except perhaps for a finite number of n's. We thus assume without loss of generality that this inequality hold a.s. for all $n\geqq 3$. Hence

$$|\xi_n| \leq \kappa_n n^\varepsilon (\log\log n)^{-1/2} \leq \kappa_n' V_n / U_n \quad \text{a.s.}$$

where $\kappa_n' \to 0$ as $n \to \infty$.

Thus the conditions of Lemma 3 are satisfied and therefore

$$\sum_{k=1}^{n} k^{-1/2+\varepsilon} Z_k = o(\eta_n n^\varepsilon (\log\log n)^{1/2}) \quad \text{a.s.}$$

where $\{\eta_n\}$ is any sequence such that $\eta_n \to \infty$ as $n \to \infty$.

To complete the proof let $\alpha_k = \gamma_k''^{-1} k^{-1/2-\varepsilon}$ for $k \geq j_1$. Now $\gamma_k''^{-1} \geq c_5 k^{a_1}$ hence $\alpha_k > c_5 k^{a_1 - 1/2 - \varepsilon}$. Since $a_1 > \frac{1}{2}$ we can choose ε to be so small such that $a_1 - \frac{1}{2} - \varepsilon > 0$ and thus $\alpha_k \to \infty$ as $k \to \infty$. Therefore by Kronecker's Lemma

$$\alpha_n^{-1} \sum_{k=m}^{n} \alpha_k k^{-1/2+\varepsilon} Z_k = o(\eta_n n^\varepsilon (\log\log n)^{1/2}) \text{ a.s.}$$

for all $m > j_1$. Or alternatively

$$\gamma_n'' \sum_{k=m}^{n} k^{-1} \gamma_k''^{-1} Z_k = o(\eta_n n^{-1/2} (\log\log n)^{1/2}) \quad \text{a.s.}$$

which completes the proof.

Corollary. *Let X_n and η_n be as in Lemma 4. Then $\bar{X}_n = o(\eta_n n^{-1/2}(\log\log n)^{1/2})$ and $n\bar{X}_n^2 = o(\eta_n \log\log n)$ a.s. as $n \to \infty$, where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$.*

Proof Obvious

The consistency of $b_{m,n}$ can now be proved.

Theorem 2. *Let X_n be given by (4) and (5) and $b_{m,n}$ by (3) with $m = m(n) = o((\log n)^{1/2+\varepsilon})$ as $n \to \infty$. If Conditions (M1) through (M4), (Z1) and (Z4) are satisfied and $\alpha_2 < 2K$ then $b_{m,n} \to \alpha$ a.s. where α is defined in (M3).*

Proof. By (M3)

$$b_{m,n} = \alpha + \frac{\sum_{i=m}^{n} (X_i - \bar{X}_{m,n})\delta(X_i)}{\sum_{i=m}^{n} (X_i - \bar{X}_{m,n})^2} + \frac{\sum_{i=m}^{n} (X_i - \bar{X}_{m,n}) Z_i}{\sum_{i=m}^{n} (X_i - \bar{X}_{m,n})^2}. \tag{10}$$

As in the proof of Lemma 2, $\delta(x)$ can be written as

$$\delta(x) = x\psi(x) \quad \text{where} \quad \psi(x) \to 0 \quad \text{as} \quad x \to 0.$$

By Lemma 2 and the corollary to Lemma 4, $\sum_{i=m}^{n}(X_i-\bar{X}_{m,n})^2$ goes to ∞ at least as $\log n$. Thus by the Toeplitz Lemma (Loève (1963) page 238) it follows that the second term of the right-hand side of (10) goes to zero a.s. $n\to\infty$.

By the same argument given in the proof of Lemma 2 it follows that

$$\sum_{i=m}^{n} X_i Z_i = o((\log n)^{1/2+\varepsilon}) \quad \text{a.s.}$$

for every $\varepsilon>0$.

Now it follows from Lemma 3 that

$$\sum_{i=m}^{n} Z_i = o(\eta_n n^{1/2}(\log\log n)^{1/2}) \quad \text{a.s.}$$

Hence by the Corollary, $\bar{X}_{m,n}\sum_{i=m}^{n} Z_i = o(\eta_n \log\log n)$ as $n\to\infty$ and η_n is as in Lemma 4. Hence the last term on the right-hand side of (10) converges to zero a.s. as $n\to\infty$. This completes the proof.

The following theorem states that the process X_n is asymptotically efficient. The proof is omitted since it follows exactly the proof of Sacks (1958).

Theorem 3. *Let X_n be given by* (4) *and* (5) *with $m(n)=o((\log n)^{1/2+\varepsilon})$ as $n\to\infty$. If conditions* (M1) *through* (M4) *and* (Z1) *through* (Z4) *are satisfied and $\alpha_2<2K$ then $n^{1/2}(X_n-\theta)$ converges in law to a normal random variable with mean zero and variance σ^2/α^2.*

Finally it should be noted that once the convergence of $b_{m,n}$ to α is proved it follows from Gaposhkin and Krasulina (1974) that X_n obeys the law of the iterated logarithm, i.e.

$$\limsup_{n\to\infty}\,[2n(\log\log n)^{-1}]^{1/2}(X_n-\theta)=\sigma/\alpha \qquad \text{a.s.}$$

References

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

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

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

[4] Dvoretzky, A. (1956). On stochastic approximation. *Proc. Third Berkeley Symp.* 1, 39–56.

[5] Gaposhkin, V.F. and Krasulina, T.P. (1974). On the law of the iterated logarithm in stochastic approximation processes. *Theor. Probability Appl.* 19, 844–850.

[6] Loève, M. (1963) *Probability Theory*, 3rd. ed. Van Nostrand, Princeton.

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

[8] Robbins, H. and D. Siegmund (1971). A convergence theorem for non-negative almost supermartingales and some applications. *Optimizing Methods in Statistics*. Academic Press, New York, 233–257.

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

[10] Sakrison, D.J. (1965). Efficient recursive estimation; application to estimating the parameters of a convariance function. *Int. J. Eng. Science* 3, 461–483.

[11] Stout, W.F. (1970). A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 15, 279–290.

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