The Annals of Statistics
1979, Vol. 7, No. 6, 1196–1221

ADAPTIVE DESIGN AND STOCHASTIC APPROXIMATION[1]

BY T. L. LAI AND HERBERT ROBBINS

Columbia University

When $y = M(x) + \varepsilon$, where M may be nonlinear, adaptive stochastic approximation schemes for the choice of the levels $x_1, x_2, \cdots$ at which $y_1, y_2, \cdots$ are observed lead to asymptotically efficient estimates of the value θ of x for which $M(\theta)$ is equal to some desired value. More importantly, these schemes make the "cost" of the observations, defined at the nth stage to be $\Sigma_1^n(x_i - \theta)^2$, to be of the order of $\log n$ instead of n, an obvious advantage in many applications. A general asymptotic theory is developed which includes these adaptive designs and the classical stochastic approximation schemes as special cases. Motivated by the cost considerations, some improvements are made in the pairwise sampling stochastic approximation scheme of Venter.

1. Introduction. We shall consider the general regression model

$$y_i = M(x_i) + \varepsilon_i \qquad i = 1, 2, \cdots \tag{1.1}$$

where the errors $\varepsilon_1, \varepsilon_2, \cdots$ are i.i.d. random variables with mean zero and variance σ^2. Unless otherwise stated, the above notations and assumptions will be used throughout the sequel. We shall always assume that the regression function $M(x)$ is a Borel function satisfying the following three conditions:

$$M(\theta) = 0 \text{ for a unique } \theta \text{ and } M'(\theta) = \beta \text{ exists and is positive;} \tag{1.2}$$

$$\inf_{\delta \leqslant |x-\theta| \leqslant \delta^{-1}} \{M(x)(x - \theta)\} > 0 \qquad \text{for all} \quad 0 < \delta < 1; \tag{1.3}$$

$$|M(x)| \leqslant c|x| + d \qquad \text{for some} \quad c, d > 0 \qquad \text{and all} \quad x. \tag{1.4}$$

Suppose that in (1.1) x_i is the dosage level of a drug given to the ith patient who turns up for treatment and that y_i is the response of the patient. Suppose the mean response of the patients under treatment should be at some optimal given level h. Without loss of generality, we shall (replacing y_i by $y_i - h$ if necessary) assume that $h = 0$. To achieve this mean response $h = 0$, if the unique (by (1.2)) root θ of the equation $M(\theta) = 0$ were known, then the dosage levels should all be set at θ. Since θ is usually unknown, how can the dosage levels x_i be chosen so that they approach θ rapidly? In the choice of the dosage levels x_i our primary objective here is in the treatment of the patients rather than in finding an efficient design to estimate the ideal dosage level θ. Calling $\Sigma_1^n(x_i - \theta)^2$ the *cost* of the design at stage n, we have announced in [8] that the apparent dilemma of choosing between a small cost and

Received March 1978; revised July 1978.

[1]This research was supported by the National Institutes of Health, the National Science Foundation, and the Office of Naval Research under grants 5R01-GM-16895, NSF MCS 76-09179, and N00014-75-C-0560

AMS 1970 *subject classifications.* Primary 62L20, 62K99; secondary 60F15.

Key words and phrases. Adaptive design, adaptive stochastic approximation, regression, logarithmic cost, asymptotic normality, iterated logarithm, pairwise sampling schemes, least squares.

a good estimate of θ can be resolved by using a suitable adaptive design. In the present paper and its companion papers [10], [11] and [12], we investigate the properties of the adaptive designs announced in [8] and prove the theorems that were stated without proof in [8]. We also consider some other adaptive designs and analyze their performance.

In an adaptive design, the choice of each level x_i will depend on the data so far observed, i.e., x_i is a function of $x_1, y_1, \cdots, x_{i-1}, y_{i-1}$. Consider first the simple case where $M(x)$ is linear so that

$$y_i = \beta(x_i - \theta) + \varepsilon_i \qquad i = 1, 2, \cdots . \tag{1.5}$$

Suppose that β is in fact *known*. To estimate θ at stage n, it is natural to use the least squares estimator

$$\theta_n^* = \bar{x}_n - \beta^{-1}\bar{y}_n (= \theta - \beta^{-1}\bar{\varepsilon}_n). \tag{1.6}$$

(Here and in the sequel, we use the notation $\bar{a}_n$ for the arithmetic mean $n^{-1}\Sigma_1^n a_i$ of any n numbers $a_1, \cdots, a_n$.) The last equality in (1.6) shows that *irrespective of how the levels x_i are chosen*, whether preassigned or sequentially determined,

$$E(\theta_n^* - \theta)^2 = \sigma^2/(n\beta^2), \tag{1.7}$$

and

$$n^{\frac{1}{2}}(\theta_n^* - \theta) \to_{\mathcal{L}} N(0, \sigma^2/\beta^2) \quad \text{as} \quad n \to \infty, \tag{1.8}$$

where $\to_{\mathcal{L}}$ denotes convergence in distribution. In particular, if we use the adaptive design

$$x_{i+1} = \theta_i^* = \bar{x}_i - \beta^{-1}\bar{y}_i \qquad i = 1, 2, \cdots \tag{1.9}$$

and let x_1 (= initial guess of θ) be a random variable with finite second moment, then it follows from (1.7) and (1.9) that the expected cost of the design (1.9) at stage n is of the order of $\log n$, i.e.,

$$E\{\Sigma_1^n (x_i - \theta)^2\} = (\sigma^2/\beta^2)\log n + 0(1). \tag{1.10}$$

While the desirable properties (1.8) and (1.10) for the adaptive design (1.9) have been obtained under the assumption of the linear model (1.5) with β known, the following theorem says that similar properties still hold in the general nonlinear case, provided again that $\beta(= M'(\theta))$ is known.

THEOREM 1. *Let $\varepsilon, \varepsilon_1, \varepsilon_2, \cdots$ be* i.i.d. *random variables with $E\varepsilon = 0$ and $E\varepsilon^2 = \sigma^2$. Let $M(x)$ be a Borel function satisfying* (1.2), (1.3) *and* (1.4). *Let x_1 be a random variable independent of $\varepsilon_1, \varepsilon_2, \cdots$. For $i = 1, 2, \cdots$, define inductively y_i by* (1.1) *and x_{i+1} by* (1.9). *Then*

$$n^{\frac{1}{2}}(x_n - \theta) \to_{\mathcal{L}} N(0, \sigma^2/\beta^2) \quad \textit{as} \quad n \to \infty; \tag{1.11}$$

$$\lim_{n\to\infty} x_n = \theta \text{ a.s., } \textit{and in fact,} \tag{1.12}$$

$$\limsup_{n\to\infty} (n/2 \log\log n)^{\frac{1}{2}} |x_n - \theta| = \sigma/\beta \text{ a.s.};$$

$$\lim_{n\to\infty} \{\Sigma_1^n (x_i - \theta)^2 / \log n\} = \sigma^2/\beta^2 \text{ a.s.} \tag{1.13}$$

Theorem 1, which was stated without proof in [8], is a special case of Theorem 2 below. The almost sure convergence results (1.12) and (1.13) are of particular interest for the kind of applications described above. As time progresses, one obtains more and more subjects for treatment; and the choice of the dosage levels is a continuing process. Theorem 1 says that if one uses the adaptive design (1.9), then with probability 1 the levels x_n will converge to the ideal level θ, the cost $\Sigma_1^n(x_i - \theta)^2$ will eventually grow like $(\sigma^2/\beta^2)\log n$, and that x_n will still be an efficient estimator of θ in the sense of (1.11).

In practice, the slope β of the regression function at the level θ will be unknown. In ignorance of β, if one simply substitutes for β some guess b of its value in the recursion (1.9), then the following analogue of Theorem 1 holds.

THEOREM 2. *Let $\varepsilon, \varepsilon_1, \varepsilon_2, \cdots$ be* i.i.d. *with $E\varepsilon = 0$ and $E\varepsilon^2 = \sigma^2$. Let $M(x)$ be a Borel function satisfying* (1.2), (1.3) *and* (1.4). *Let x_1 be a random variable independent of $\varepsilon_1, \varepsilon_2 \cdots$. For $i = 1, 2, \cdots$, define inductively y_i by* (1.1) *and x_{i+1} by*

$$x_{i+1} = \bar{x}_i - b^{-1}\bar{y}_i, \tag{1.14}$$

where b is a positive constant.

(i) *Let $f(t) = 1/\{t(2 - t)\}$ for $0 < t < 2$. If $b < 2\beta$, then*

$$n^{\frac{1}{2}}(x_n - \theta) \to_{\mathcal{L}} N(0, (\sigma^2/\beta^2)f(b/\beta)); \tag{1.15}$$

$$\limsup_{n\to\infty}(n/2\log\log n)^{\frac{1}{2}}|x_n - \theta| = (\sigma/\beta)f^{\frac{1}{2}}(b/\beta) \text{ a.s.}; \tag{1.16}$$

$$\lim_{n\to\infty}\{\Sigma_1^n(x_i - \theta)^2/\log n\} = (\sigma^2/\beta^2)f(b/\beta) \text{ a.s.} \tag{1.17}$$

(ii) *Assume that $M(x)$ further satisfies*

$$M(x) = \beta(x - \theta) + 0(|x - \theta|^{1+\eta}) \text{ as } x \to \theta \text{ for some } \eta > 0. \tag{1.18}$$

If $b > 2\beta$, then there exists a random variable z such that

$$n^{\beta/b}(x_n - \theta) \to z \text{ a.s.}, \tag{1.19}$$

and, therefore,

$$\{\Sigma_1^n(x_i - \theta)^2/n^{1-(2\beta/b)}\} \to z^2/\{1 - (2\beta/b)\} \text{ a.s.} \tag{1.20}$$

(iii) *Suppose $b = 2\beta$ and that $M(x)$ further satisfies* (1.18). *Then*

$$(n/\log n)^{\frac{1}{2}}(x_n - \theta) \to_{\mathcal{L}} N(0, \sigma^2/b^2), \tag{1.21}$$

and

$$\{\Sigma_1^n(x_i - \theta)^2/(\log n)^2\} \to_{\mathcal{L}} (\sigma^2/b^2)\textstyle\int_0^1 w^2(t)\,dt, \tag{1.22}$$

where $w(t)$, $t \geqslant 0$, is the standard Wiener process.

In Section 2 we shall prove a more general result that contains Theorem 2 as a special case. Theorem 2 says that if in (1.14) $b < 2\beta$ then the cost $\Sigma_1^n(x_i - \theta)^2$ grows like a constant times $\log n$ and that $n^{\frac{1}{2}}(x_n - \theta)$ is asymptotically normal. The factor $f(b/\beta)$ in (1.15)–(1.17) has its minimum value 1 for $b = \beta$, and

$f(b/\beta) = (1 - r^2)^{-1}$ if $b = (1 \pm r)\beta$ $(0 < r < 1)$. Thus, even if our guess b of β has a relative error of 50%, the variance of the asymptotic distribution of $n^{\frac{1}{2}}(x_n - \theta)$ and the asymptotic cost $\Sigma_1^n(x_i - \theta)^2$ for the adaptive design (1.14) only exceed the corresponding minimum values for $b = \beta$ by a factor of $\frac{4}{3}$. On the other hand, if $b \geqslant 2\beta$, the cost is of a much larger order of magnitude and the rate of convergence of x_n to θ is much slower.

The adaptive design (1.14) sticks to an initial guess b of β, and its asymptotic performance is unsatisfactory when b exceeds 2β. Instead of adhering to an initial guess of β, it is natural to consider the possibility of estimating β from the data already observed and using that estimate in the choice of the next x-value. This means replacing the recursion (1.14) by

$$x_{i+1} = \bar{x}_i - b_i^{-1}\bar{y}_i \tag{1.23}$$

where $b_i = b_i(x_1, y_1, \cdots, x_i, y_i)$ is some estimate of β based on the data already observed. Adaptive designs of the type (1.23) will be discussed in [11], where we shall show that *by a suitable choice of* b_i *the desirable asymptotic properties* (1.11), (1.12), *and* (1.13) *still hold for the design* (1.23).

In this paper we consider an alternative way of modifying the adaptive design (1.9) when β is not known. We first note another way of expressing the recursive scheme (1.9), or more generally (1.14), in the following lemma.

LEMMA 1. *Let* $\{x_i, i \geqslant 1\}$ *and* $\{y_i, i \geqslant 1\}$ *be two sequences of real numbers. For any constant* c *and positive integer* n, *the following two statements are equivalent*:

$$x_{i+1} = \bar{x}_i - c\bar{y}_i \quad \textit{for all} \quad i = 1, \cdots, n; \tag{1.24}$$

$$x_{i+1} = x_i - cy_i/i \quad \textit{for all} \quad i = 1, \cdots, n. \tag{1.25}$$

Lemma 1 is easily proved by induction on n. Now, the recursion (1.25) with $c > 0$ is a special case of the general *stochastic approximation scheme*

$$x_{i+1} = x_i - c_i y_i \qquad i = 1, 2, \cdots \tag{1.26}$$

introduced by Robbins and Monro [15], where $\{c_i\}$ is an arbitrary sequence of positive constants such that

$$\Sigma_1^\infty c_i = \infty \quad \text{and} \quad \Sigma_1^\infty c_i^2 < \infty. \tag{1.27}$$

For the regression model (1.1), under the assumptions on the errors ε_i and on the regression function $M(x)$ described in the first paragraph, it is known (cf. [1], [15]) that (1.27) is a sufficient condition for the x_n generated by the stochastic approximation scheme (1.26) to converge to θ in mean square and with probability 1. It is also known (cf. [3], [16]) that if $c_i = (ib)^{-1}$, where b is a positive constant $< 2\beta$, then the x_n generated by (1.26) has an asymptotically normal distribution as given by (1.15). Therefore, by the equivalence in Lemma 1, the asymptotic normality (1.15) in Theorem 2 follows immediately. It is also known that if the c_i are of a larger order of magnitude than i^{-1}, then x_n may converge to θ in distribution at a rate much slower than that of (1.15). In particular, Chung [3] has considered

$c_i = i^{-(1-\delta)}$ for certain positive values of $\delta < \frac{1}{2}$ and has shown under some restrictive assumptions that

$$n^{(1-\delta)/2}(x_n - \theta) \to_{\mathcal{L}} N(0, \sigma^2/(2\beta)). \tag{1.28}$$

He has also noted that an asymptotically optimal choice of c_i for the stochastic approximation scheme (1.26) is $c_i = (i\beta)^{-1}$, at least for the linear case $M(x) = \beta(x - \theta)$ (cf. Sections 6 and 7 of [3]).

In ignorance of β, it is natural to try using $c_i = (ib_i)^{-1}$ in (1.26), where $b_i = b_i(x_1, y_1, \cdots, x_i, y_i)$ is some estimate of β based on the data already observed. Of course, we want b_i to be a strongly consistent estimator of β so that hopefully the asymptotic properties (1.11), (1.12) and (1.13) will be preserved. We shall call any adaptive design

$$x_{i+1} = x_i - y_i/(ib_i), \tag{1.29}$$

where b_i is a strongly consistent estimator of $\beta(= M'(\theta))$, an *adaptive stochastic approximation scheme*, and in Section 2 we shall show that adaptive stochastic approximation schemes have the desirable properties (1.11), (1.12) and (1.13) of Theorem 1. More generally, if b_i in (1.29) converges to some positive constant b with probability 1, we shall call (1.29) a *quasi-adaptive stochastic approximation scheme*. An asymptotic theory will be developed in Section 2 for quasi-adaptive stochastic approximation schemes, and these general results not only include Theorem 2, and therefore Theorem 1 as well, as special cases, but also establish the desired asymptotic properties (1.11), (1.12) and (1.13) for adaptive stochastic approximation schemes. In Sections 3 and 4 we shall describe two different methods of constructing adaptive stochastic approximation schemes and apply the results of Section 2 to the analysis of these procedures.

2. Asymptotic properties of quasi-adaptive stochastic approximation schemes. Throughout this section the following notations will be used. Let $\varepsilon, \varepsilon_1, \cdots$ be i.i.d. random variables with $E\varepsilon = 0$. (Although we shall often also assume that $E\varepsilon^2 < \infty$, there are certain places where we can relax this assumption.) Let $M(x)$ be a Borel function satisfying (1.2), (1.3) and (1.4). Let x_1 be a random variable independent of $\varepsilon_1, \varepsilon_2, \cdots$ Let $\mathcal{F}_0$ denote the σ-field generated by x_1, and for $k \geq 1$ let $\mathcal{F}_k$ denote the σ-field generated by $x_1, \varepsilon_1, \cdots, \varepsilon_k$. For $i = 1, 2, \cdots$, let $y_i = M(x_i) + \varepsilon_i$, where $\{x_i\}$ is a stochastic approximation scheme defined by

$$x_{i+1} = x_i - y_i/(ib_i), \tag{2.1}$$

and $\{b_i\}$ is a sequence of positive random variables.

The following representation theorem, stated without proof in [9], is a very useful tool for analyzing quasi-adaptive stochastic approximation schemes.

THEOREM 3. *Let b be a positive constant and let $\{b_n\}$ be a sequence of positive random variables such that* $\lim_{n\to\infty} b_n = b$ a.s.

(i) *For the stochastic approximation scheme* (2.1), *if* $\lim_{n\to\infty} x_n = \theta$ a.s., *then the following representation holds*:

$$x_{n+1} = \theta + \left(n^{-\beta/b}/\tau_n\right)\left\{\Sigma_1^n \delta_k \varepsilon_k + \rho_0\right\}, \tag{2.2}$$

where ρ_0, τ_k *and* δ_k *are random variables having the following properties*:

$$\tau_k > 0 \quad \text{and} \quad \delta_k = -k^{(\beta/b)-1}\tau_k/b_k, \qquad k \geqslant 1; \tag{2.3}$$

$$P\left[\tau_{n+1} - \tau_n = o(\tau_n/n) \quad \text{as} \quad n \to \infty\right] = 1. \tag{2.4}$$

(ii) *Suppose that* b_n *is* $\mathcal{F}_{n-1}$*-measurable for all* $n \geqslant 1$, *and assume either that* $E(|\varepsilon|\log|\varepsilon|) < \infty$ *or that* ε *is symmetric. Then for the stochastic approximation scheme* (2.1), $\lim_{n\to\infty} x_n = \theta$ a.s., *and in the representation* (2.2) *we further obtain that*

$$\tau_k \text{ and therefore } \delta_k \text{ also are } \mathcal{F}_{k-1}\text{-measurable for all } k \geqslant 1. \tag{2.5}$$

(iii) *Assume that* $E|\varepsilon|^r < \infty$ *for some* $r > 1$, *that* $M(x)$ *also satisfies* (1.18), *and that*

$$P\left[b_n - b = o(n^{-\lambda}) \text{ as } n \to \infty\right] = 1 \text{ for some positive constant } \lambda. \tag{2.6}$$

Then for the stochastic approximation scheme (2.1), $\lim_{n\to\infty} x_n = \theta$ a.s., *and in the representation* (2.2), *the random variables* τ_k *and* δ_k *satisfy* (2.3) *and*

$$P\left[\tau_{n+1}/\tau_n = 1 + o(n^{-(1+p)}) \text{ as } n \to \infty\right] = 1 \text{ for some positive constant } p. \tag{2.7}$$

Consequently,

$$\lim_{n\to\infty} \tau_n = \tau \text{ exists and is positive a.s., and } P\left[\tau_n = \tau + o(n^{-p}) \text{ as } n \to \infty\right] = 1. \tag{2.8}$$

Moreover, if b_n *is* $\mathcal{F}_{n-1}$*-measurable for all* $n \geqslant 1$, *then* (2.5) *also holds.*

REMARKS. In the particular case $b_n = b$ for all n, and under the stronger moment condition $E\varepsilon^2 < \infty$, similar representation results have been obtained by Major and Révész [13], Kersting [7], and Gaposhkin and Krasulina [6]. Gaposhkin and Krasulina [6] have obtained the representation (2.2) for this particular case and established the properties (2.3), (2.5) and

$$P\left[\{\tau_n\} \text{ is slowly varying as } n \to \infty\right] = 1. \tag{2.9}$$

(A sequence $\{L(n)\}$ is said to be *slowly varying* as $n \to \infty$ if $L([cn])/L(n) \to 1$ for all $c > 0$. If $\{L(n)\}$ is slowly varying, then the sequence $\{n^\alpha L(n)\}$ is said to be *regularly varying* with exponent α.) By Theorem 4 of [2], (2.4) implies (2.9); in fact, a sequence $\{L(n)\}$ of positive numbers is slowly varying if and only if there exists a sequence $\{c_n\}$ of positive numbers such that $L(n)/c_n \to 1$ and $c_{n+1} - c_n = o(c_n/n)$. The property (2.4) is a very useful tool for studying the limiting behavior of the stochastic approximation scheme (2.1) and of the cost $\Sigma_1^n(x_i - \theta)^2$ for the design. It enables us to reduce the problem to that of the martingale $\Sigma_1^n \varepsilon_i/b_i$ (when b_i is $\mathcal{F}_{i-1}$-measurable for all i) via a partial summation technique (cf. [9]).

The representation considered by Major and Révész [13] is somewhat different from (2.2). They assume that $M(x) = \beta(x - \theta) + U(x)$ where $U(x) = O((x - \theta)^2)$

as $x \to \theta$ (i.e., (1.18) holds with $\eta = 1$) and obtain the representation

$$x_{n+1} = \theta - n^{-\beta/b}\{\Sigma_1^n k^{(\beta/b)-1}(1 + 0(k^{-1}))(\varepsilon_k + U(x_k)) + \rho_0\}$$

for the case where $b_n = b$ for all n and under the assumption $E\varepsilon^2 < \infty$. Again, for this special case $b_n = b$, and under the assumptions $E\varepsilon^2 < \infty$ and (1.18), Kersting [7] recently showed that if $b < 2\beta$ then

$$x_{n+1} = \theta - n^{-\beta/b}\Sigma_1^n k^{(\beta/b)-1}\varepsilon_k + \rho_n$$

where the ρ_n are random variables such that $n^\lambda \rho_n \to 0$ a.s. for some $\frac{1}{2} < \lambda < \beta/b$. The methods of Kersting and of Major and Révész depend very heavily on the assumption that $b_n = b$ for all n. The following proof of Theorem 3 is based on a generalization of the argument of Gaposhkin and Krasulina.

PROOF OF THEOREM 3. Suppose that $\lim_{n\to\infty} x_n = \theta$ a.s. Without loss of generality, we can assume that $\theta = 0$. Therefore, in view of (1.2),

$$M(x_n) = (\beta + \xi_n)x_n, \quad \text{where} \quad \xi_n \to 0 \text{ a.s.} \tag{2.10}$$

Hence by (2.1),

$$x_{n+1} = (1 - n^{-1}d_n)x_n - \varepsilon_n/(nb_n), \quad \text{where} \quad d_n = (\beta + \xi_n)/b_n.$$

It then follows that

$$x_{n+1} = \beta_{m-1,n}x_m - \Sigma_{k=m}^n \beta_{kn}\varepsilon_k/(kb_k), \tag{2.11}$$

where

$$\beta_{kn} = \Pi_{j=k+1}^n(1 - j^{-1}d_j),$$
$$k = 0, 1, \cdots, n-1; \beta_{nn} = 1.$$

Clearly $d_n \to \beta/b$ a.s. Therefore, for almost all ω, if k is sufficiently large, say $k \geqslant k_0(\omega)$, then

$$\beta_{kn} = \gamma_n\gamma_k^{-1} \quad \text{for} \quad n \geqslant k, \tag{2.12}$$

where

$$\gamma_n = \Pi_{j=1}^n\left(\max\{1 - j^{-1}d_j, \tfrac{1}{2}\}\right). \tag{2.13}$$

Since $d_n \to \beta/b$ a.s., it is easy to see from (2.13) that with probability 1, γ_n is regularly varying with exponent $-\beta/b$ (cf. [2]). Let $\tau_n = (n^{\beta/b}\gamma_n)^{-1}$. Then τ_n is slowly varying with probability 1. To show that τ_n satisfies (2.4), we note that with probability 1,

$$\tau_n/\tau_{n+1} = (1 + n^{-1})^{\beta/b}\{1 - n^{-1}(\beta/b + o(1))\} = 1 + o(n^{-1}).$$

From (2.11) and (2.12), we obtain that

$$x_{n+1} = -(n^{\beta/b}/\tau_n)\Sigma_1^n k^{(\beta/b)-1}\tau_k\varepsilon_k/b_k$$
$$+ \{(n^{-\beta/b}/\tau_n)\Sigma_1^{k_0} k^{(\beta/b)-1}\tau_k\varepsilon_k/b_k + \beta_{k_0,n}x_{k_0+1}\}.$$

Hence (2.2) holds.

To prove (ii), suppose that b_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$, and assume either that $E(|\varepsilon|\log|\varepsilon|) < \infty$ or that ε is symmetric. Then by Theorem 4 of [9], $\lim_{n\to\infty} x_n = 0$ a.s. Moreover, since b_n is $\mathcal{F}_{n-1}$-measurable for all n, so is ξ_n (as defined by (2.10)). Hence $d_n(= (\beta + \xi_n)/b_n)$ and therefore γ_n (as defined by (2.13)) also are $\mathcal{F}_{n-1}$-measurable, and (2.5) follows.

To prove (iii), suppose that $E|\varepsilon|^r < \infty$ for some $r > 1$ and that $\{b_n\}$ satisfies (2.6). Then

$$\Sigma_1^n (ib_i)^{-1}\varepsilon_i = \Sigma_1^n (ib)^{-1}\varepsilon_i + \Sigma_1^n (ibb_i)^{-1}(b - b_i)\varepsilon_i.$$

Since $E|\varepsilon|^r < \infty$, $\Sigma_1^n i^{-1}\varepsilon_i$ converges a.s. by the three-series theorem. Moreover, since $E(\Sigma_1^\infty i^{-1-\lambda}|\varepsilon_i|) < \infty$ and so $\Sigma_1^\infty i^{-1-\lambda}|\varepsilon_i| < \infty$ a.s., therefore by (2.6), with probability 1

$$\Sigma_1^\infty (ib_i)^{-1}|b - b_i|\,|\varepsilon_i| = \Sigma_1^\infty o\big(i^{-1-\lambda}|\varepsilon_i|\big) < \infty.$$

Hence $\Sigma_1^n (ib_i)^{-1}\varepsilon_i$ converges a.s. Therefore, using the same argument as in [1] (see also Lemma 5 in Section 3 below), it can be shown that $\lim_{n\to\infty} x_n = 0$ a.s., and so the representation (2.2) holds.

We note that by (2.3),

$$(2.14)\quad \Sigma_1^n \delta_k \varepsilon_k = -b^{-1}\big(\Sigma_1^n k^{(\beta/b)-1}\tau_k\varepsilon_k\big) - \big\{\Sigma_1^n k^{(\beta/b)-1}\tau_k\big(b_k^{-1} - b^{-1}\big)\varepsilon_k\big\}$$
$$= -b^{-1}U_{1n} - U_{2n}, \text{ say.}$$

Take $0 < \rho < \lambda$ and $q > 0$ such that $\beta/b > q$ and $\rho > q$. Then by (2.6) and (2.9), $\tau_n|b_n^{-1} - b^{-1}| = o(n^{-\rho})$ a.s., and therefore with probability 1

$$(2.15)\quad |U_{2n}| \leq n^{(\beta/b)-q}\Sigma_1^n o\big(k^{q-1-\rho}|\varepsilon_k|\big) = 0\big(n^{(\beta/b)-q}\big),$$

since $\rho > q$ implies that $\Sigma_1^\infty k^{q-1-\rho}|\varepsilon_k| < \infty$ a.s. Without loss of generality, we can assume that $r < 2$ and $\beta/b > 1 - r^{-1}$. Let $S_n = \Sigma_1^n \varepsilon_i$. Then $n^{-1/r}S_n \to 0$ a.s., and by (2.4),

$$k^{(\beta/b)-1}\tau_k - (k+1)^{(\beta/b)-1}\tau_{k+1} \sim -(b^{-1}\beta - 1)k^{(\beta/b)-2}\tau_k \text{ a.s.}$$

Therefore, in view of (2.9), with probability 1

$$(2.16)\quad U_{1n} = n^{(\beta/b)-1}\tau_n S_n + \Sigma_1^{n-1}\big\{k^{(\beta/b)-1}\tau_k - (k+1)^{(\beta/b)-1}\tau_{k+1}\big\}S_k$$
$$= 0\big(n^{(\beta/b)-1+r^{-1}}\tau_n\big).$$

Let $0 < \zeta < \min\{q, 1 - r^{-1}\}$. Then from (2.2), (2.9), (2.14), (2.15) and (2.16), it follows that $\lim_{n\to\infty} n^\zeta x_n = 0$ a.s.

Assume that $M(x)$ also satisfies (1.18). Then with η given by (1.18) and ξ_n defined by (2.10), since $n^\zeta x_n \to 0$ a.s., we can write

$$(2.17)\quad \xi_n = \xi_n' n^{-\zeta\eta}, \qquad \text{where} \quad \xi_n' \to 0 \text{ a.s.}$$

Without loss of generality, we shall assume that $\eta \leq 1$, and so $\zeta\eta < 1$. Let

$p = \min\{\zeta\eta, \lambda\}$, where λ is given by (2.6). Then by (2.6) and (2.17),

$$d_n(= (\beta + \xi_n)/b_n) = \beta/b + \xi_n'' n^{-p}, \quad \text{where } \xi_n'' \to 0 \text{ a.s.} \tag{2.18}$$

From (2.13) and (2.18), it is easy to see that (2.7) and (2.8) hold. □

We now make use of the above representation theorem to obtain the following generalization of Theorem 2 to quasi-adaptive stochastic approximation schemes.

THEOREM 4. *Assume that* $E\varepsilon^2 = \sigma^2 < \infty$, *and suppose that* $\{b_n\}$ *is a sequence of positive random variables such that* b_n *is* $\mathcal{F}_{n-1}$*-measurable for all* $n \geq 1$ *and* $\lim_{n\to\infty} b_n = b$ a.s., *where* b *is a positive constant. Let* $\{x_n\}$ *be the quasi-adaptive stochastic approximation scheme defined by* (2.1).

(i) *Let* $b < 2\beta$. *Then* (1.15), (1.16) *and* (1.17) *hold.*

(ii) *Let* $b > 2\beta$. *If* $M(x)$ *further satisfies* (1.18) *and* $\{b_n\}$ *further satisfies* (2.6), *then there exists a random variable* z *such that* (1.19) *and* (1.20) *hold.*

(iii) *Let* $b = 2\beta$. *If* $M(x)$ *further satisfies* (1.18) *and* $\{b_n\}$ *further satisfies* (2.6), *then* (1.21) *and* (1.22) *hold.*

(iv) *Suppose* $b < 2\beta$. *Let* b_n^* *be a sequence of positive random variables* (*not necessarily* $\mathcal{F}_{n-1}$*-measurable*) *such that*

$$P\left[b_n^* - b_n = o\left(n^{-\frac{1}{2}} v_n\right) \text{ as } n \to \infty\right] = 1, \tag{2.19}$$

where $v_1, v_2, \cdots$ *are* i.i.d. *positive random variables such that* $Ev_1^2 < \infty$. *Let* $x_1^* = x_1$ *and for* $i = 1, 2, \cdots$, *define* $y_i^* = M(x_i^*) + \varepsilon_i$ *and* $x_{i+1}^* = x_i^* - y_i^*/(ib_i^*)$. *Then* (1.15) *and* (1.16) *still hold with* x_i^* *in place of* x_i. *If condition* (2.19) *is strengthened to*

$$P\left[b_n^* - b_n = o\left((n \log\log n)^{-\frac{1}{2}} v_n\right) \text{ as } n \to \infty\right] = 1 \tag{2.20}$$

and there exists a positive integer m *such that* b_n^* *is* $\mathcal{F}_{n+m}$*-measurable for all* $n \geq 1$, *then* (1.17) *still holds with* x_i^* *in place of* x_i.

REMARK. The first three parts of this theorem deal with quasi-adaptive stochastic approximation schemes whose estimate b_n of β at the nth stage is based only on $(x_1, y_1, \cdots, x_{n-1}, y_{n-1})$. Although y_n is also observed, it is not used to estimate β, for otherwise b_n would not be $\mathcal{F}_{n-1}$-measurable. The requirement that b_n be $\mathcal{F}_{n-1}$-measurable gives a martingale structure to the sum $\Sigma_1^n \delta_k \varepsilon_k$ in the representation (2.2), and our proof of Theorem 4(i)–(iii) depends on this martingale structure. On the other hand, since y_n has also been observed, it seems artificial not to use it in b_n simply because this would destroy the expedient martingale property. In this connection, Theorem 4(iv) is of particular interest. It implies that given an estimator $b_n^*(= b_n^*(x_1, y_1, \cdots, x_n, y_n))$, if b_n^* is close to $b_n = b_{n-1}^*$ in the sense of (2.20), then the desired conclusions still hold, at least in the important case $b < 2\beta$. We shall see in Section 3 and [10] that most "reasonable" estimators b_n^* for the present problem satisfy the approximation property (2.20) with $b_n = b_{n-1}^*$. Note that Theorem 2 is the special case of Theorem 4 with $b_n = b$ for all n. We shall need the following three lemmas in the proof of Theorem 4.

LEMMA 2. *Let* $z_1, z_2, \cdots$ *be* i.i.d. *random variables with* $E|z_1| < \infty$.

(i) $\lim_{n\to\infty}(\Sigma_1^n i^{-1}z_i)/(\log n) = Ez_1$ a.s.

(ii) *Let* $\{r_n\}$ *be a sequence of random variables such that* $P[r_n = o(n^{-\rho})$ *as* $n \to \infty] = 1$ *for some* $0 < \rho \leqslant 1$. *Then, with probability* 1,

$$(2.21)\qquad \Sigma_1^n r_i z_i = o(n^{1-\rho}) \quad \text{if } \rho < 1, \qquad = o(\log n) \quad \text{if } \rho = 1.$$

PROOF. To prove (i), let $S_n = \Sigma_1^n z_i$. We note that

$$(2.22)\qquad \Sigma_1^n i^{-1}z_i = \Sigma_1^{n-1}(i^{-1} - (i+1)^{-1})S_i + n^{-1}S_n.$$

As $i \to \infty$, $i^{-1} - (i+1)^{-1} \sim i^{-2}$ and $i^{-1}S_i \to Ez_1$ a.s. Hence it follows from (2.22) that $\Sigma_1^n i^{-1}z_i \sim (Ez_1)\log n$ a.s.

To prove (ii), let $S_n' = \Sigma_1^n |z_i|$ and note that

$$\Sigma_1^n i^{-\rho}|z_i| = \Sigma_1^{n-1}(i^{-\rho} - (i+1)^{-\rho})S_i' + n^{-\rho}S_n'.$$

As $i \to \infty$, $i^{-\rho} - (i+1)^{-\rho} \sim \rho i^{-(\rho+1)}$, $i^{-1}S_i' \to E|z_1|$ a.s. and $|r_i| = o(i^{-\rho})$ a.s. Hence (2.21) follows. □

LEMMA 3. *Let* $z_1, z_2, \cdots$ *be* i.i.d. *random variables such that* $Ez_1 = 0$ *and* $Ez_1^2 = \sigma^2 < \infty$. *Let* $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots$ *be an increasing sequence of σ-fields such that* z_i *is* $\mathcal{G}_i$*-measurable and is independent of* $\mathcal{G}_{i-1}$ *for all* $i \geqslant 1$. *Let* $u_1, u_2, \cdots$ *be a sequence of random variables such that* u_i *is* $\mathcal{G}_{i-1}$*-measurable for all* $i \geqslant 1$ *and* $\lim_{n\to\infty}u_n = A$ a.s. *for some constant* A. *Then, redefining the random variables on a new probability space if necessary, there exists a standard Wiener process* $w(t)$, $t \geqslant 0$, *such that*

$$(2.23)\qquad \max_{m\leqslant n}|\Sigma_1^m k^{-\frac{1}{2}}u_k z_k - A\sigma w(\log m)|/(\log n)^{\frac{1}{2}} \to_P 0 \quad \textit{as} \quad n \to \infty.$$

PROOF. Write

$$(2.24)\qquad \Sigma_1^m k^{-\frac{1}{2}}u_k z_k = A\Sigma_1^m k^{-\frac{1}{2}}z_k + \Sigma_1^m k^{-\frac{1}{2}}(u_k - A)z_k.$$

Redefining the random variables on a new probability space if necessary, there exists a standard Wiener process $w(t)$, $t \geqslant 0$, such that

$$(2.25)\qquad \max_{m\leqslant n}|\Sigma_1^m k^{-\frac{1}{2}}z_k - \sigma w(\log m)|/(\log n)^{\frac{1}{2}} \to_P 0 \quad \text{as} \quad n \to \infty,$$

(cf. [5]). Let $\tilde{u}_i = (u_i - A)I_{[|u_i - A| \leqslant 1]}$. Since $\lim_{n\to\infty}u_n = A$ a.s. and (2.25) holds, it remains to show that

$$(2.26)\qquad \max_{m\leqslant n}|\Sigma_1^m k^{-\frac{1}{2}}\tilde{u}_i z_i|/(\log n)^{\frac{1}{2}} \to_P 0.$$

Noting that $E(\Sigma_1^n k^{-\frac{1}{2}}\tilde{u}_i z_i)^2 = \sigma^2\Sigma_1^n k^{-1}E\tilde{u}_i^2 = o(\log n)$, (2.26) follows easily from the martingale inequality. □

LEMMA 4. *With the same notations and assumptions as in Lemma* 3, *suppose that* $\{\tau_n\}$ *is a sequence of positive random variables satisfying* (2.7) *and therefore* (2.8) *as well. Then, redefining the random variables on a new probability space if necessary, we* *have as* $n \to \infty$

(2.27) $$\max_{m \leqslant n} |\Sigma_1^m k^{-\frac{1}{2}} \tau_k u_k z_k - A\sigma\tau w(\log m)| / (\log n)^{\frac{1}{2}} \to_P 0.$$

PROOF. Let $S(m) = \Sigma_1^m k^{-\frac{1}{2}} u_k z_k$. By partial summation,

(2.28) $$\Sigma_1^m k^{-\frac{1}{2}} \tau_k u_k z_k = \Sigma_1^{m-1}(\tau_k - \tau_{k+1}) S(k) + (\tau_m - \tau) S(m) + \tau S(m).$$

By (2.7), (2.8) and Lemma 3, for $1 \leqslant m \leqslant n$,

(2.29) $$\Sigma_1^{m-1} |\tau_k - \tau_{k+1}| \, |S(k)| + |\tau_m - \tau| \, |S(m)|$$
$$= \Sigma_1^{m-1} 0(k^{-(1+p)})\{|w(\log k) + \Delta_n(k)|\} + 0(m^{-p})\{|w(\log m) + \Delta_n(m)|\},$$

where $\Delta_n(m)$ are random variables such that

(2.30) $$\max_{m \leqslant n} |\Delta_n(m)| / (\log n)^{\frac{1}{2}} \to_P 0.$$

From (2.28), (2.29), (2.30) and Lemma 3, (2.27) follows immediately. □

PROOF OF THEOREM 4(i). By a partial summation technique, we have shown in [9] that the central limit theorem (1.15) and the law of the iterated logarithm (1.16) follow from the representation in Theorem 3 and certain martingale limit theorems. We now prove the asymptotic behavior (1.17) of the cost $\Sigma_1^n (x_i - \theta)^2$. By Theorem 3(ii),

(2.31) $$\Sigma_1^n (x_i - \theta)^2 = (x_1 - \theta)^2 + \Sigma_{i=1}^{n-1} i^{-2\beta/b} \tau_i^{-2} \{\Sigma_{k=1}^i \delta_k \varepsilon_k + \rho_0\}^2,$$

where ρ_0, τ_k, and δ_k are random variables satisfying (2.3), (2.4) and (2.5). Let $a = 2\beta/b$ (> 1). Define

(2.32) $$S_n = \Sigma_1^n \delta_k \varepsilon_k, \quad \tilde{S}_n = S_n + \rho_0, \quad a(n) = \Sigma_{i=n}^{\infty} i^{-a} \sim n^{-a+1}/(a-1).$$

We note that

(2.33)
$$\Sigma_1^n i^{-a} \tau_i^{-2} \tilde{S}_i^2 = \Sigma_1^n \tau_i^{-2} \tilde{S}_i^2 (a(i) - a(i+1))$$
$$= \Sigma_2^n a(i)(\tau_i^{-2} - \tau_{i-1}^{-2}) \tilde{S}_i^2 + \Sigma_2^n a(i) \tau_{i-1}^{-2} (\tilde{S}_i^2 - \tilde{S}_{i-1}^2) - a(n+1) \tau_n^{-2} \tilde{S}_n^2$$
$$+ a(1) \tau_1^{-2} \tilde{S}_1^2.$$

By (2.4) and (2.9), with probability 1,

(2.34)
$$\Sigma_2^n a(i) |\tau_i^{-2} - \tau_{i-1}^{-2}| \tilde{S}_i^2 = (2 + o(1)) \Sigma_2^n a(i) |\tau_i - \tau_{i-1}| \tau_i^{-3} \tilde{S}_i^2 + 0(1)$$
$$= o(\Sigma_1^n (a(i)/i) \tau_i^{-2} \tilde{S}_i^2) + 0(1) = o(\Sigma_1^n i^{-a} \tau_i^{-2} \tilde{S}_i^2) + 0(1).$$

(We add the 0(1) term in (2.34) because we have not yet shown that $\Sigma_1^n i^{-a} \tau_i^{-2} \tilde{S}_i^2 \to \infty$ with probability 1.) Obviously,

(2.35) $$\Sigma_2^n a(i) \tau_{i-1}^{-2} (\tilde{S}_i^2 - \tilde{S}_{i-1}^2) = \Sigma_2^n a(i) \tau_{i-1}^{-2} (\delta_i^2 \varepsilon_i^2 + 2\delta_i \varepsilon_i S_{i-1} + 2\delta_i \varepsilon_i \rho_0).$$

Since the summands are nonnegative, it follows from (2.3), (2.9) and (2.32) that with probability 1

$$(2.36)\qquad \Sigma_2^n a(i)\tau_{i-1}^{-2}\delta_i^2\varepsilon_i^2 \sim (\Sigma_2^n i^{-1}\varepsilon_i^2)/\{b^2(a-1)\}$$
$$\sim (\sigma^2 \log n)/\{b(2\beta - b)\}.$$

The last relation above follows from Lemma 2(i) since $E\varepsilon^2 = \sigma^2 < \infty$. We note that $a(i)\tau_{i-1}^{-2}\delta_i S_{i-1} \to 0$ a.s. by (1.16), (2.2) and (2.9). Moreover, $\{\Sigma_{i=2}^n a(i)\tau_{i-1}^{-2}\delta_i S_{i-1}\varepsilon_i, \mathcal{F}_n, n \geqslant 2\}$ is a martingale transform. Therefore, using a standard truncation argument like that used in the proof of Lemma 3 (to ensure finite expectations) and the strong law for martingales (cf. [14], page 150), we obtain that with probability 1

$$(2.37)\qquad \Sigma_2^n a(i)\tau_{i-1}^{-2}\delta_i\varepsilon_i S_{i-1} = o(\Sigma_2^n a^2(i)\tau_{i-1}^{-4}\delta_i^2 S_{i-1}^2) + 0(1)$$
$$= o(\Sigma_1^{n-1} i^{-a}\tau_i^{-2}S_i^2) + 0(1) = o(\Sigma_1^{n-1} i^{-a}\tau_i^{-2}\tilde{S}_i^2) + 0(1).$$

The 0(1) term in (2.37) indicates that $\Sigma_2^n a(i)\tau_{i-1}^{-2}\delta_i\varepsilon_i S_{i-1}$ converges a.s. on $[\Sigma_1^\infty i^{-a}\tau_i^{-2}S_i^2 < \infty]$. To see the last relation in (2.37), we note that $S_i^2 \leqslant 2(\tilde{S}_i^2 + \rho_0^2)$ and that $\Sigma_1^\infty i^{-a}\tau_i^{-2} < \infty$ a.s. since $a > 1$ and (2.9) holds. Since

$$\Sigma_2^\infty a^2(i)\tau_{i-1}^{-4}\delta_i^2 = \Sigma_2^\infty 0(i^{-a}\tau_i^{-2}) < \infty \text{ a.s.},$$

it follows from the (local) martingale convergence theorem ([14], page 148) and a standard truncation argument like that used in the proof of Lemma 3 (to ensure finite expectations) that

$$(2.38)\qquad \Sigma_2^n a(i)\tau_{i-1}^{-2}\delta_i\varepsilon_i \quad \text{converges a.s.}$$

In view of the law of the iterated logarithm (1.16) and the representation (2.2), we obtain that

$$(2.39)\qquad (n^{-\beta/b}/\tau_n)\tilde{S}_n = 0(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \text{ a.s.}$$

Therefore, with probability 1,

$$(2.40)\qquad a(n+1)\tau_n^{-2}\tilde{S}_n^2 = o(\log n).$$

From the relations (2.33)–(2.40), the desired conclusion (1.17) for $\Sigma_1^n(x_i - \theta)^2$ follows. □

To better understand the partial summation technique in the preceding proof of (1.17), consider the special case $b = \beta$ and $M(x) = \beta(x - \theta)$. In this case, as indicated in Section 1, $x_{i+1} = \theta - \beta^{-1}\bar{\varepsilon}_i$, and so (1.17) reduces to the following interesting corollary on the fluctuation behavior of sample means.

COROLLARY 1. *Let* $\varepsilon, \varepsilon_1, \cdots$ *be* i.i.d. *random variables with* $E\varepsilon = 0$ *and* $E\varepsilon^2 = \sigma^2$. *Then*

$$\lim_{n\to\infty}(\Sigma_1^n\bar{\varepsilon}_i^2)/(\log n) = \sigma^2 \text{ a.s.}$$

To analyze $\Sigma_1^n\bar{\varepsilon}_i^2 = \Sigma_1^n i^{-2}(\Sigma_1^i\varepsilon_j)^2$, partial summation is the natural method.

PROOF OF THEOREM 4(ii). Since $b > 2\beta$, $\Sigma_1^\infty \delta_i^2 < \infty$ a.s. Hence the martingale transform $\Sigma_1^n \delta_k \varepsilon_k$ converges a.s. as $n \to \infty$. Therefore, by Theorem 3(iii),

$$\tau n^{\beta/b}(x_n - \theta) \to \rho_0 + \Sigma_1^\infty \delta_k \varepsilon_k \text{ a.s.}$$

Hence (1.19) holds with $z = \tau^{-1}\{\rho_0 + \Sigma_1^\infty \delta_k \varepsilon_k\}$. The relation (1.20) is an immediate consequence of (1.19). □

PROOF OF THEOREM 4(iii). We note that since $b = 2\beta$, $\delta_k = -k^{-\frac{1}{2}}\tau_k / b_k$. Let $\tilde{S}_n = \rho_0 + \Sigma_1^n \delta_k \varepsilon_k$ as before. Since $\lim_{k\to\infty} b_k = b$ a.s. and τ_k satisfies (2.7) and (2.8) by Theorem 3(iii), Lemma 4 implies that, redefining the random variables on a new probability space if necessary, there exists a standard Wiener process $w(t)$, $t \geqslant 0$, such that

$$\max_{k \leqslant n} |\tau^{-1}\tilde{S}_k - (\sigma/b) w(\log k)| / (\log n)^{\frac{1}{2}} \to_P 0. \tag{2.41}$$

From (2.2) and (2.41), the asymptotic normality result (1.21) follows immediately. We note that with probability 1,

$$\begin{aligned} \Sigma_1^n k^{-1} w^2(\log k) &= \textstyle\int_1^n t^{-1} w^2(\log t)\, dt + o(\log n) \\ &= \textstyle\int_0^{\log n} w^2(s)\, ds + o(\log n), \qquad \text{setting } s = \log t. \end{aligned} \tag{2.42}$$

Since $\int_0^{\log n} w^2(s)\, ds$ has the same distribution as $(\log n)^2 \int_0^1 w^2(t)\, dt$, it then follows from (2.41) and (2.42) that

$$\left(\tau^{-2}\Sigma_1^{n-1} k^{-1} \tilde{S}_k^2\right) / (\log n)^2 \to_{\mathcal{L}} (\sigma^2/b^2) \textstyle\int_0^1 w^2(t)\, dt. \tag{2.43}$$

The desired conclusion (1.22) then follows from (2.43) and the fact that

$$\Sigma_1^n (x_i - \theta)^2 = (x_1 - \theta)^2 + \Sigma_1^{n-1}\left(k^{-1}/\tau_k^2\right)\tilde{S}_k^2 \sim \tau^{-2}\Sigma_1^{n-1} k^{-1} \tilde{S}_k^2 \text{ a.s.} \qquad \square \tag{2.44}$$

PROOF OF THEOREM 4(iv). Assume that (2.19) holds. We note that

$$\Sigma_1^n (i b_i^*)^{-1} \varepsilon_i = \Sigma_1^n (i b_i)^{-1} \varepsilon_i + \Sigma_1^n (i b_i b_i^*)^{-1}(b_i - b_i^*)\varepsilon_i.$$

Since b_i is $\mathfrak{F}_{i-1}$-measurable and $\Sigma_1^\infty (ib)^{-2} < \infty$ a.s., the martingale transform $\Sigma_1^n (i b_i)^{-1}\varepsilon_i$ converges a.s. Since $E\{\Sigma_1^\infty i^{-\frac{3}{2}} v_i |\varepsilon_i|\} < \infty$ and so $\Sigma_1^\infty i^{-\frac{3}{2}} v_i |\varepsilon_i| < \infty$ a.s., therefore by (2.19), with probability 1,

$$\Sigma_1^\infty (i b_i b_i^*)^{-1} |b_i - b_i^*|\, |\varepsilon_i| = \Sigma_1^\infty o\left(i^{-\frac{3}{2}} v_i |\varepsilon_i|\right) < \infty.$$

Hence $\Sigma_1^n (i b_i^*)^{-1}\varepsilon_i$ converges a.s. Therefore, using the same argument as in [1] (see also Lemma 5 in Section 3 below), it can be shown that $\lim_{n\to\infty} x_n = \theta$ a.s. Hence by Theorem 3(i),

$$x_{n+1}^* = \theta + \left(n^{-\beta/b}/\tau_n^*\right)\{\Sigma_1^n \delta_k^* \varepsilon_k + \rho_0^*\}, \tag{2.45}$$

where τ_k^* and δ_k^* satisfy (2.3) and (2.4).

Let $Z^*(n) = \Sigma_1^n \varepsilon_k / b_k^*$ and $Z(n) = \Sigma_1^n \varepsilon_k / b_k$. Then by (2.19), with probability 1,

$$|Z^*(n) - Z(n)| \leqslant \Sigma_1^n o\left(k^{-\frac{1}{2}} v_k |\varepsilon_k|\right) = o(n^{\frac{1}{2}}), \tag{2.46}$$

since $E\varepsilon^2 < \infty$ and $Ev_1^2 < \infty$ (see Lemma 2(ii)). Set $Z^*(0) = 0$ and $Z^*(t) = Z^*(n)$

for $n - 1 < t \leq n$, $n = 1, 2, \cdots$. From (2.46) and Theorem 2 of [9], it then follows that, redefining the random variables on a new probability space if necessary, there exist standard Wiener processes $w(t)$ and $w^*(t)$, $t \geq 0$, such that

(2.47) $$\max_{0 \leq t \leq 1} |r^{-\frac{1}{2}} Z^*(rt) - (\sigma/b) w(t)| \to_P 0 \quad \text{as} \quad r \to \infty,$$

and

(2.48) $$\lim_{t\to\infty} |Z^*(t) - (\sigma/b) w^*(t)| / (t \log \log t)^{\frac{1}{2}} = 0 \text{ a.s.}$$

Let $\alpha = \beta/b - 1$. We note that

(2.49) $$\Sigma_1^n \delta_k^* \varepsilon_k = \Sigma_1^{n-1} \big(k^\alpha \tau_k^* - (k+1)^\alpha \tau_{k+1}^*\big) Z^*(k) + n^\alpha \tau_n Z^*(n),$$

and

(2.50) $$k^\alpha \tau_k^* - (k+1)^\alpha \tau_{k+1}^* \sim -\alpha k^{\alpha - 1} \tau_k^* \text{ a.s.}$$

by (2.4). From (2.47)–(2.50) it is not hard to show that (1.15) and (1.16) still hold with x_i^* in place of x_i (see the proof of Theorem 7 of [9]).

Now assume that the stronger condition (2.20) holds in place of (2.19), and that b_n^* is $\mathcal{F}_{n+m}$-measurable for all $n \geq 1$. We shall show that (1.17) still holds with x_i^* in place of x_i. Let $a = 2\beta/b$ and $a(n) = \Sigma_{i=n}^\infty i^{-a}$ as before and set $S_n^* = \Sigma_1^n \delta_k^* \varepsilon_k$, $\tilde{S}_n^* = S_n^* + \rho_0^*$. Clearly the relations (2.33)–(2.36) and (2.39) – (2.40) still hold with τ_i^*, δ_i^*, S_i^* and $\tilde{S}_i^*$ in place of τ_i, δ_i, S_i and $\tilde{S}_i$. Hence we need only show that in analogy with (2.37) and (2.38), with probability 1,

(2.51) $$\Sigma_2^n a(i) \tau_{i-1}^{*-2} \delta_i^* S_{i-1}^* \varepsilon_i = o\big(\Sigma_1^{n-1} i^{-a} \tau_i^{*-2} S_i^{*2}\big) + o(\log n),$$

and

(2.52) $$\Sigma_2^n a(i) \tau_{i-1}^{*-2} \delta_i^* \varepsilon_i \quad \text{converges.}$$

To prove (2.51) and (2.52), we note that τ_n^* and therefore S_n^* also are $\mathcal{F}_{n+m}$-measurable, since b_n^* is $\mathcal{F}_{n+m}$-measurable for all $n \geq m$ (see the proof of Theorem 3). By (2.4), with probability 1,

(2.53)
$$\tau_i^* = \tau_{i-m-1}^*\big(1 + o(i^{-1})\big) \quad \text{and} \quad \tau_{i-1}^* = \tau_{i-m-1}^*\big(1 + o(i^{-1})\big) \quad \text{as} \quad i \to \infty.$$

As in (2.39), we have

(2.54) $$S_n^* = 0\Big(\tau_n^* n^{(\beta/b) - \frac{1}{2}} (\log \log n)^{\frac{1}{2}}\Big) \text{ a.s.,}$$

and therefore with probability 1

(2.55) $$\Sigma_2^\infty a(i) i^{-1} \tau_{i-m-1}^{*-1} \big(i^{(\beta/b)-1}/b_i^*\big) |S_{i-1}^*| \, |\varepsilon_i| = \Sigma_2^\infty 0\Big(i^{-\frac{3}{2}} (\log \log i)^{\frac{1}{2}}\Big) \cdot |\varepsilon_i| < \infty.$$

By (2.20), with probability 1,

(2.56) $$b_n^* = b_n + o\Big((n \log \log n)^{-\frac{1}{2}} v_n\Big),$$

and therefore using (2.54),

$$(2.57)\quad \Sigma_2^n a(i)\tau_i^{*-1} i^{(\beta/b)-1}|b_i^* - b_i|\,|S_i^*|\,|\varepsilon_i| = \Sigma_2^n o(i^{-1}v_i|\varepsilon_i|) = o(\log n), \qquad \text{by Lemma 2.}$$

For $i > m + 1$, let $U_i = -\tau_{i-m-1}^* \Sigma_{j=i-m}^{i-1} j^{(\beta/b)-1}\varepsilon_j/b_j$ and write

$$(2.58)\qquad S_{i-1}^* = S_{i-m-1}^* + U_i + R_i.$$

From (2.53) and (2.56), it follows that with probability 1

$$(2.59)\qquad R_i = o\Big(\Big\{ i^{(\beta/b)-\frac{3}{2}}(\log\log i)^{-\frac{1}{2}}\tau_i^*\Sigma_{j=i-m}^{i-1}(|\varepsilon_j| + |\varepsilon_j v_j|)\Big\}\Big),$$

and therefore

$$(2.60)\qquad \Sigma_2^\infty a(i)\tau_i^{*-1} i^{(\beta/b)-1}|R_i|\,|\varepsilon_i| = \Sigma_{i=2}^\infty o\Big(i^{-\frac{3}{2}}|\varepsilon_i|\Sigma_{j=i-m}^{i-1}\{|\varepsilon_j| + |\varepsilon_j v_j|\}\Big) < \infty.$$

From (2.53) and (2.55)–(2.59), to prove (2.51), it suffices to show that with probability 1

$$(2.61)\quad \Sigma_{i=m+2}^n a(i)\tau_{i-m-1}^{*-1}(i^{(\beta/b)-1}/b_i)(S_{i-m-1}^* + U_i)\varepsilon_i = o(\Sigma_1^{n-1} i^{-a}\tau_i^{*-2}S_i^{*2}) + o(\log n).$$

Since τ_{i-m-1}^*, S_{i-m-1}^*, b_i and U_i are $\mathcal{F}_{i-1}$-measurable, the left-hand side of (2.61) forms a martingale transform, so using the strong law for martingales as in (2.37),

$$(2.62)\quad \Sigma_{i=m+2}^n a(i)\tau_{i-m-1}^{*-1}(i^{(\beta/b)-1}/b_i)(S_{i-m-1}^* + U_i)\varepsilon_i = o\big(\Sigma_{m+2}^n a^2(i)\tau_i^{*-2} i^{2\beta/b-2}(S_{i-m-1}^* + U_i)^2\big) + 0(1) \text{ a.s.}$$

Since $\lim n^{-\frac{1}{2}}\varepsilon_n = 0$ a.s. and $\lim n^{-\frac{1}{2}}v_n = 0$ a.s., we obtain from (2.54), (2.58) and (2.59) that with probability 1

$$S_{i-1}^{*2} - (S_{i-m-1}^* + U_i)^2 = 2S_{i-1}^* R_i - R_i^2 = o(\tau_i^{*2} i^{2(\beta/b-1)+1}),$$

and therefore

$$(2.63)\qquad \Sigma_{i=m+2}^n a^2(i)\tau_i^{*-2} i^{2\beta/b-2}|S_{i-1}^{*2} - (S_{i-m-1}^* + U_i)^2| = o(\log n).$$

From (2.62) and (2.63), (2.61) follows as desired.

Making use of the fact that $\beta/b > \frac{1}{2}$ and an argument as in (2.53), (2.55), (2.56) and (2.57), to show that (2.52) holds with probability 1 we need only prove that

$$(2.64)\qquad \Sigma_{i=m+2}^n a(i)\tau_{i-m-1}^{*-1} i^{(\beta/b)-1}\varepsilon_i/b_i \qquad \text{converges a.s.}$$

Since τ_{i-m-1}^* and b_i are $\mathcal{F}_{i-1}$-measurable, (2.64) follows easily from the (local) martingale convergence theorem as in (2.38). □

In the preceding proof of Theorem 4 (see also the proof of Theorem 7 of [9]), the representation given by Theorem 3 is the only property of the stochastic approximation scheme (2.1) that we have used. This suggests the following more general theorem which we shall need in Section 3.

THEOREM 5. *Assume that* $E\varepsilon^2 = \sigma^2 < \infty$ *and suppose that* $\{b_n\}$ *is a sequence of positive random variables such that* b_n *is* $\mathcal{F}_{n-1}$*-measurable for all* $n \geqslant 1$ *and* $\lim_{n\to\infty} b_n = b$ a.s., *where* b *is a positive constant. Let* $\{\tau_n\}$ *be a sequence of positive random variables such that* (2.4) *holds and* τ_n *is* $\mathcal{F}_{n-1}$*-measurable for all* $n \geqslant 1$. *Let* X_n *and* ρ_n *be random variables such that for* $n \geqslant 1$

$$X_{n+1} = \theta - \left(n^{-\beta/b}/\tau_n\right)\left\{\Sigma_1^n k^{(\beta/b)-1}\tau_k \varepsilon_k / b_k + \rho_n\right\}, \tag{2.65}$$

where $\beta > 0$ *and* θ *are constants.*

(i) *Let* $b < 2\beta$ *and assume that*

$$\rho_n = o\left(\tau_n n^{(\beta/b)-\frac{1}{2}}\right) \text{ a.s.} \tag{2.66}$$

Then (1.15), (1.16) *and* (1.17) *still hold with* X_n *in place of* x_n.

(ii) *Let* $b > 2\beta$ *and assume that*

$$\rho_n \textit{ converges a.s. to some random variable } \rho_0. \tag{2.67}$$

Suppose furthermore that τ_n *converges* a.s. *to a positive random variable* τ *as* $n \to \infty$. *Then there exists a random variable* z *such that* (1.19) *and* (1.20) *hold with* X_n *in place of* x_n.

(iii) *Let* $b = 2\beta$ *and assume that*

$$\rho_n = o\left((\log n)^{\frac{1}{2}}\right) \text{ a.s.} \tag{2.68}$$

Suppose furthermore that $\{\tau_n\}$ *satisfies the stronger assumption* (2.7) *instead of* (2.4). *Then* (1.21) *and* (1.22) *still hold with* X_n *in place of* x_n.

(iv) *Let* $b < 2\beta$. *Let* τ_n^* *and* b_n^* *be positive random variables (not necessarily* $\mathcal{F}_{n-1}$*-measurable) such that* τ_n^* *satisfies* (2.4) (*with* τ_n^* *in place of* τ_n) *and* b_n^* *satisfies the approximation property* (2.19) *for some sequence* $\{v_n\}$ *of positive* i.i.d. *random variables with* $Ev_1^2 < \infty$. *Let* $\{\rho_n\}$ *be a sequence of random variables such that* (2.66) *holds with* τ_n^* *in place of* τ_n. *Suppose for* $n \geqslant 1$ *that*

$$X_{n+1}^* = \theta - \left(n^{-\beta/b}/\tau_n^*\right)\left\{\Sigma_1^n k^{(\beta/b)-1}\tau_k^* \varepsilon_k / b_k^* + \rho_n\right\}. \tag{2.69}$$

Then (1.15) *and* (1.16) *still hold with* X_n^* *in place of* x_n. *If condition* (2.19) *is strengthened to* (2.20) *and there exists a positive integer* m *such that* b_n^* *and* τ_n^* *are* $\mathcal{F}_{n+m}$*-measurable for all* $n \geqslant 1$, *then* (1.17) *still holds with* X_i^* *in place of* x_i.

PROOF. We shall only consider Part (i) of the theorem, since the argument for the other parts is similar. In the proof of Theorem 4(i), we have actually established that Theorem 5(i) holds for the special case $\rho_n = 0$ for all n. This in turn obviously implies that the central limit theorem (1.15) and the law of the iterated logarithm (1.16) also hold for the more general case where $\{\rho_n\}$ satisfies (2.66). Let

$$S_n = \Sigma_1^n k^{(\beta/b)-1}\tau_k \varepsilon_k / b_k.$$

For the special case $\rho_n = 0$ for all n, the relation (1.17) can be written as

$$\Sigma_1^n S_j^2 / \left(j^{2\beta/b}\tau_j^2\right) \sim \sigma^2(\log n)/\{b(2\beta - b)\} \text{ a.s.} \tag{2.70}$$

We now show that (2.70) implies that (1.17) also holds in the more general case where $\{\rho_n\}$ satisfies (2.66). By (2.66) there exist positive random variables η_j such that

(2.71) $$\eta_j \to \infty \text{ a.s.} \quad \text{and} \quad \eta_j(|\rho_j| + 1) = o\left(\tau_j j^{(\beta/b)-\frac{1}{2}}\right) \text{ a.s.},$$

and therefore

(2.72) $$\Sigma_1^n \eta_j^2(|\rho_j| + 1)^2 / \left(j^{2\beta/b}\tau_j^2\right) = o(\log n) \text{ a.s.}$$

From (2.72), it follows that with probability 1

(2.73) $$\begin{aligned} \Sigma_1^n S_j^2 / \left(j^{2\beta/b}\tau_j^2\right) &= \Sigma_1^n S_j^2 I_{[|S_j| > \eta_j(|\rho_j|+1)]} / \left(j^{2\beta/b}\tau_j^2\right) + o(\log n) \\ &= (1 + o(1))\Sigma_1^n (S_j + \rho_j)^2 I_{[|S_j| > \eta_j(|\rho_j|+1)]} / \left(j^{2\beta/b}\tau_j^2\right) \\ &\quad + o(\log n) \qquad (\text{since } \eta_j \to \infty \text{ a.s.}) \\ &= (1 + o(1))\Sigma_1^n (S_j + \rho_j)^2 / \left(j^{2\beta/b}\tau_j^2\right) + o(\log n). \end{aligned}$$

From (2.70) and (2.73), it then follows that (1.17) also holds under the assumption (2.66). □

3. Venter's design and some modifications. In [17] Venter proposed a modification of the Robbins-Monro stochastic approximation scheme (1.26) to obtain successive estimates of the unknown slope β which have the desired property of converging to β with probability 1. Venter's design requires that at the mth stage ($m = 1, 2, \cdots$) two observations y_m' and y_m'' be taken, at levels $x_m' = x_m - a_m$ and $x_m'' = x_m + a_m$, where $\{a_m\}$ is a sequence of positive constants such that

(3.1) $$a_m \sim am^{-\gamma} \text{ for some constants } a > 0 \text{ and } \tfrac{1}{4} < \gamma < \tfrac{1}{2},$$

and x_m is the mth approximation to θ, defined recursively by

(3.2) $$\begin{aligned} x_1 &= \text{initial guess of } \theta, \\ x_{i+1} &= x_i - y_i / (ib_i). \end{aligned}$$

The quantity y_i in (3.2) estimates the (unobserved) response at the level x_i, and is defined by

(3.3) $$y_i = \tfrac{1}{2}(y_i' + y_i'').$$

Assuming that positive constants b and B are known such that

(3.4) $$b < \beta < B,$$

Venter defines the slope estimate b_i in (3.2) by

(3.5) $$b_i = b \vee \left\{B \wedge i^{-1}\Sigma_{j=1}^i (y_j'' - y_j') / (2a_j)\right\},$$

where the symbols $\vee$ and $\wedge$ denote maximum and minimum respectively.

We note that for $j = 1, 2, \cdots$,

(3.6) $$y_j' = M(x_j - a_j) + \varepsilon_j', \qquad y_j'' = M(x_j + a_j) + \varepsilon_j'',$$

where the errors $\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_1'', \varepsilon_2'', \cdots$ are i.i.d. with mean 0 and variance σ^2. In particular, for the linear case $M(x) = \beta(x - \theta)$, (3.6) implies that

$$y_j'' - y_j' = 2\beta a_j + (\varepsilon_j'' - \varepsilon_j'), \tag{3.7}$$

which depends only on β but not on θ. Since $\gamma < \frac{1}{2}$, $\Sigma_1^\infty j^{-2(1-\gamma)} < \infty$, and therefore in view of (3.1) it easily follows that

$$m^{-1}\Sigma_1^m(\varepsilon_j'' - \varepsilon_j')/a_j \to 0 \text{ a.s.} \tag{3.8}$$

Hence b_i defined by (3.5) is strongly consistent, at least in the linear case. This argument was extended by Venter [17] to general regression functions $M(x)$ which satisfy (1.2)–(1.4) and

(3.9)

$$\sup_{|y|<A}|M(x + y) - M(x)| \leqslant A^* \quad \text{for some } A, A^* > 0 \quad \text{and all } x;$$

(3.10) $M(x)$ is k-times continuously differentiable in some neighborhood of θ,

where $k \geqslant 2$ is an integer satisfying $k\beta > \gamma B$.

At stage m, Venter's scheme has taken $n = 2m$ observations. Let $\tilde{\theta}_n = x_{m+1}$ be the estimate of θ and $\tilde{\beta}_n = b_m$ be the estimate of β given by (3.2) and (3.5) respectively, and let

$$C_n = \Sigma_1^m(x_i' - \theta)^2 + \Sigma_1^m(x_i'' - \theta)^2 \tag{3.11}$$

be the cost of these n observations. The following theorem, which can be proved by using Theorem 5, shows that although $\tilde{\theta}_n$ approaches θ at the asymptotically optimal rate given in (1.11) and (1.12) of Theorem 1, the cost C_n incurred by Venter's scheme is of a much larger order of magnitude than the logarithmic cost in (1.13) of Theorem 1. As to the conditions of the theorem, we are able to relax Venter's assumptions on $M(x)$ and also to remove the assumption (3.4) on prior knowledge of bounds for β.

THEOREM 6. *Assume that $M(x)$ satisfies* (1.2)–(1.4) *and* (3.10) *with $k = 2$. Let a_m be a sequence of positive constants satisfying* (3.1) *and let $\xi_m > \zeta_m$ be two sequences of positive constants such that*

$$\limsup_{m\to\infty}\zeta_m < \beta < \liminf_{m\to\infty}\xi_m, \tag{3.12a}$$

$$\Sigma_1^\infty(i\zeta_i)^{-2} < \infty, \tag{3.12b}$$

and

$$\Sigma_1^\infty(i\xi_i)^{-1} = \infty. \tag{3.12c}$$

Let $\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_1'', \varepsilon_2'', \cdots$ be i.i.d. *with mean* 0 *and variance σ^2.*

(i) *Define x_i, y_i, y_i', y_i'' by* (3.2), (3.3) *and* (3.6), *and b_i by*

$$b_i = \zeta_i \vee \{\xi_i \wedge (i-1)^{-1}\Sigma_{j=1}^{i-1}(y_j'' - y_j')/(2a_j)\}, \quad i \geqslant 2, b_1 = \zeta_1. \tag{3.13}$$

Let $n = 2m$, $\tilde{\theta}_n = x_{m+1}$, $\tilde{\beta}_n = b_m$ *and define* C_n *as in* (3.11). *Then as* $n \to \infty$,

(3.14) $$\tilde{\theta}_n \to \theta \text{ and } \tilde{\beta}_n \to \beta \text{ a.s.},$$

(3.15) $$\limsup (n/2 \log\log n)^{\frac{1}{2}} |\tilde{\theta}_n - \theta| = \sigma/\beta \text{ a.s.},$$

(3.16) $$n^{\frac{1}{2}}(\tilde{\theta}_n - \theta) \to_{\mathcal{L}} N(0, \sigma^2/\beta^2),$$

(3.17) $$C_n/n^{1-2\gamma} \to 4^\gamma a^2/(1-2\gamma) \text{ a.s.},$$

(3.18) $$n^{\frac{1}{2}-\gamma}(\tilde{\beta}_n - \beta) \to_{\mathcal{L}} N(0, \sigma^2/\{4^\gamma a^2(1+2\gamma)\}),$$

where $a > 0$ *and* $\frac{1}{4} < \gamma < \frac{1}{2}$ *are given by the condition* (3.1) *on the sequence* $\{a_m\}$.

(ii) *Assume in place of* (3.12*b*) *the stronger condition*

(3.19) $$\Sigma_1^\infty i^{-2+\gamma} \zeta_i^{-2} < \infty.$$

Suppose that in (i) *we replace* b_i *as defined in* (3.13) *by*

(3.20) $$b_i^* = \zeta_i \vee \{\xi_i \wedge i^{-1}\Sigma_{j=1}^i (y_j'' - y_j')/(2a_j)\}, \qquad i \geq 1.$$

Then the relations (3.14)–(3.18) *still hold.*

REMARKS. (a) Venter [17] proved (3.14), (3.16) and (3.18) in Theorem 6(ii) for the special case $\zeta_m = b < \beta < B = \xi_m$ under the more restrictive smoothness conditions (3.9) and (3.10) with $k \geq 2$ such that $k\beta > \gamma B$. The asymptotic behavior (3.17) of the cost C_n, however, has not been considered in the literature.

(b) Dropping Venter's assumption (3.4) on prior knowledge of bounds for β, we can choose $\zeta_m \to 0$ and $\xi_m \to \infty$ such that (3.12c) and (3.19) hold. Obviously the condition (3.12a) is then also satisfied.

(c) For the case $M''(\theta) \neq 0$, Venter has shown that the constant γ in condition (3.1) has to be chosen $> \frac{1}{4}$ and that (3.16) actually fails to hold if $\gamma = \frac{1}{4}$ (see Theorem 3 of [17]).

(d) Fabian [4] has proved (3.14) and (3.16) in Theorem 6(i) for the special case $\zeta_m = c_1 m^{-\alpha}$ and $\xi_m = c_2 \log(m+1)$ with $0 < c_1 < c_2$ and $0 < \alpha < \frac{1}{2}$. His proof is simpler than that of Venter and does not require Venter's smoothness conditions and the assumption (3.4) on prior bounds for β. However, his method requires that the last summand $(y_i'' - y_i')/(2a_i)$ be dropped in b_i^*, and therefore he considers b_i instead of b_i^*. His argument depends heavily on the fact that b_i is $\mathcal{F}_{i-1}$-measurable, where

(3.21) $$\mathcal{F}_m = \mathcal{B}(x_1, \varepsilon_1', \varepsilon_1'', \cdots, \varepsilon_m', \varepsilon_m'') \qquad \mathcal{F}_0 = \mathcal{B}(x_1).$$

Our argument is different from those of Venter and Fabian and works for both b_n and b_n^*.

The following lemma will be used in the proofs of Theorems 6 and 7.

LEMMA 5. *Let $M(x)$ be a Borel function satisfying* (1.4), *and assume that* (1.3) *holds for some real number θ. Let x_n, u_n, v_n, t_n, and t'_n be random variables such that*

$$x_{n+1} = x_n - v_n\{M(x_n + t_n) + M(x_n + t'_n)\} + u_n, \quad n = 1, 2, \cdots, \tag{3.22}$$

$$\lim_{n\to\infty} t_n = \lim_{n\to\infty} t'_n = 0 \text{ a.s.}, \tag{3.23}$$

$$v_n \geqslant 0 \textit{ and } \lim_{n\to\infty} v_n = 0 \text{ a.s.}, \tag{3.24}$$

and

$$\Sigma_1^N u_n \textit{ converges a.s. as } N \to \infty. \tag{3.25}$$

Then x_n converges a.s. *to some random variable. If, furthermore,*

$$\Sigma_1^\infty v_n = \infty \text{ a.s.}, \tag{3.26}$$

then $\lim_{n\to\infty} x_n = \theta$ a.s.

PROOF. From (3.22) and (3.25), it follows that

$$x_{N+1} + \Sigma_1^N v_n\{M(x_n + t_n) + M(x_n + t'_n)\} \text{ converges a.s. as } N \to \infty. \tag{3.27}$$

When $t_n = t'_n = 0$ for all n, the rest of the proof is exactly like that of Blum in Lemma 3 and Theorem 1 of [1]. An obvious modification of Blum's argument extends to the more general case where t_n and t'_n satisfy (3.22). □

We now proceed to prove Theorem 6. We shall only prove Part (ii) of the theorem in detail. The proof of Part (i) is similar and is, in fact, simpler, and we shall comment on it after the proof of Part (ii).

PROOF OF THEOREM 6(ii). We shall first prove that

$$\lim_{n\to\infty} x_n = \theta \text{ a.s.} \tag{3.28}$$

In view of (3.2), (3.3) and (3.6), $\{x_n\}$ satisfies (3.22) with $t_n = -t'_n = a_n$, $v_n = \frac{1}{2}(nb_n^*)^{-1}$, and $u_n = -\frac{1}{2}(\varepsilon'_n + \varepsilon''_n)/(nb_n^*)$. Therefore, by Lemma 5, to prove (3.28) it suffices to show that

$$\Sigma_1^N(\varepsilon'_n + \varepsilon''_n)/(nb_n^*) \text{ converges a.s.} \tag{3.29}$$

Since b_i^* is $\mathcal{F}_i$-measurable, x_{i+1} is also $\mathcal{F}_i$-measurable by (3.2). Therefore, although b_i^* is not $\mathcal{F}_{i-1}$-measurable, a slight modification of it gives the $\mathcal{F}_{i-1}$-measurable random variable

$$b'_i = \zeta_i \vee \left\{\xi_i \wedge \frac{1}{i}\left[\frac{M(x_i + a_i) - M(x_i - a_i)}{2a_i} + \Sigma_{j=1}^{i-1}\frac{y''_j - y'_j}{2a_j}\right]\right\}. \tag{3.30}$$

We note that with probability 1

$$\begin{aligned}(nb_n^*)^{-1} - (nb'_n)^{-1} &= (nb_n^* b'_n)^{-1}(b'_n - b_n^*) \\ &= 0(|\varepsilon''_n - \varepsilon'_n|/(n^2\zeta_n^2 a_n)).\end{aligned} \tag{3.31}$$

Since $a_n \sim an^{-\gamma}$, it follows from (3.19) that

$$\Sigma_1^\infty E\{(|\varepsilon_n''|^2 + |\varepsilon_n'|^2)/(n^2\zeta_n^2 a_n)\} < \infty, \tag{3.32}$$

and therefore $\Sigma_1^\infty\{|\varepsilon_n' + \varepsilon_n''|\,|\varepsilon_n' - \varepsilon_n''|/(n^2\zeta_n^2 a_n)\} < \infty$ a.s. Hence, by (3.31),

$$\Sigma_1^N(\varepsilon_n' + \varepsilon_n'')\{(nb_n^*)^{-1} - (nb_n')^{-1}\} \text{ converges a.s. as } N \to \infty. \tag{3.33}$$

Since b_i' is $\mathcal{F}_{i-1}$-measurable, $\{\Sigma_1^n(\varepsilon_i' + \varepsilon_i'')/(ib_i'), \mathcal{F}_n, n \geqslant 1\}$ is a martingale. Moreover, since $b_n' \geqslant \zeta_n$ and $\Sigma_1^\infty(n\zeta_n)^{-2} < \infty$, $\Sigma_1^\infty E\{(\varepsilon_n' + \varepsilon_n'')^2/(nb_n')^2\} < \infty$. Therefore, by the martingale convergence theorem,

$$\Sigma_1^N(\varepsilon_n' + \varepsilon_n'')/(nb_n') \text{ converges a.s. as } N \to \infty. \tag{3.34}$$

From (3.33) and (3.34), the desired conclusion (3.29) follows and so (3.28) holds.

Using (3.10) (with $k = 1$) and (3.28), we obtain that, with probability 1,

$$M(x_j + a_j) - M(x_j - a_j) = 2a_j(\beta + o(1)) \qquad \text{as } j \to \infty. \tag{3.35}$$

From (3.6), (3.8) and (3.35), it follows easily that $b_m^* \to \beta$ a.s. Hence (3.14) holds.

Let $\varepsilon_i = \frac{1}{2}(\varepsilon_i' + \varepsilon_i'')$. Then $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \frac{1}{2}\sigma^2$. Since (3.10) holds with $k = 2$ and $x_i \to \theta$ a.s., we can apply Taylor's expansion to two terms to obtain that

$$\tfrac{1}{2}\{M(x_i + a_i) + M(x_i - a_i)\} = (\beta + \eta_i)x_i + \omega_i, \tag{3.36}$$

where η_i and ω_i are $\mathcal{B}(x_i)$-measurable random variables such that

$$\eta_i \to 0 \qquad \text{and} \qquad \omega_i = 0(a_i^2) \text{ a.s.} \tag{3.37}$$

We note that by (3.3) and (3.36),

$$\begin{aligned} y_i &= \tfrac{1}{2}\{M(x_i + a_i) + M(x_i - a_i)\} + \varepsilon_i \\ &= (\beta + \eta_i)x_i + (\varepsilon_i + \omega_i). \end{aligned} \tag{3.38}$$

Therefore, using exactly the same argument as in the proof of Theorem 3, and noting that $b_n^* \to \beta$ a.s., it can be shown that

$$\begin{aligned} x_{n+1} &= \theta - (n\tau_n^*)^{-1}\{\Sigma_1^n\tau_k^*(\varepsilon_k + \omega_k)/b_k^* + \rho_0\} \\ &= \theta - (n\tau_n^*)^{-1}\{\Sigma_1^n\tau_k^*\varepsilon_k/b_k^* + \rho_n\}, \end{aligned} \tag{3.39}$$

where ρ_0 and τ_n^* are random variables such that τ_n^* satisfies (2.4) and is positive and $\mathcal{F}_n$-measurable (since b_n^* is $\mathcal{F}_n$-measurable), and

$$\begin{aligned} \rho_n &= \rho_0 + \Sigma_1^n\tau_k^*\omega_k/b_k^* \\ &= 0(\tau_n^* n^{1-2\gamma}) \text{ a.s.} \qquad \text{by (3.37).} \end{aligned} \tag{3.40}$$

Since $1 - 2\gamma < \frac{1}{2}$, (3.40) implies that $\{\rho_n\}$ satisfies (2.66) (with $b = \beta$). Moreover, b_n' is $\mathcal{F}_{n-1}$-measurable and, with probability 1,

$$b_n^* - b_n' = n^{-1}(\varepsilon_n'' - \varepsilon_n')/(2a_n) = 0(n^{-(1-\gamma)}(1 + |\varepsilon_n'| + |\varepsilon_n''|)) \text{ as } n \to \infty. \tag{3.41}$$

Hence $\{x_m\}$ admits a representation of the type in Theorem 5(iv). Therefore, by

Theorem 5(iv), as $m \to \infty$

(3.42) $$m^{\frac{1}{2}}(x_m - \theta) \to_{\mathcal{L}} N\left(0, \tfrac{1}{2}\sigma^2/\beta^2\right),$$

(3.43) $$\limsup (m/2 \log\log m)^{\frac{1}{2}}|x_m - \theta| = \sigma/\left(2^{\frac{1}{2}}\beta\right) \text{ a.s.},$$

(3.44) $$\Sigma_1^m (x_i - \theta)^2/\log m \to \tfrac{1}{2}\sigma^2/\beta^2 \text{ a.s.}$$

From (3.42) and (3.43), (3.15) and (3.16) follow immediately. Since $C_n = 2\{\Sigma_1^m(x_i - \theta)^2 + \Sigma_1^m a_i^2\}$ and (3.44) holds, while

$$\Sigma_1^m a_i^2 \sim a^2\left(\tfrac{1}{2}n\right)^{1-2\gamma}/(1 - 2\gamma),$$

(3.17) follows.

Using (3.10) (with $k = 2$) together with (3.43) and the fact that $a_j > j^{-\frac{1}{2}}(\log\log j)^{\frac{1}{2}}$ for all large j, we can sharpen (3.35) to

(3.45)
$$M(x_j + a_j) - M(x_j - a_j) = 2a_j M'(x_j + r_j a_j) \quad (\text{where } |r_j| \leqslant 1)$$
$$= 2a_j\{\beta + 0(r_j a_j + x_j - \theta)\} = 2a_j(\beta + 0(a_j)) \text{ a.s.}$$

Hence, with probability 1,

(3.46) $$m^{-1}\Sigma_{j=1}^m (M(x_j + a_j) - M(x_j - a_j))/(2a_j) - \beta$$
$$= m^{-1}\Sigma_{j=1}^m 0(a_j) = 0(m^{-\gamma}) = o\left(m^{-(\frac{1}{2}-\gamma)}\right) \qquad \text{since } \gamma > \tfrac{1}{4}.$$

Since $E(\varepsilon_1'' - \varepsilon_1')^2 = 2\sigma^2$ and $\Sigma_1^m a_i^{-2} \sim a^{-2} m^{1+2\gamma}/(1 + 2\gamma)$, we obtain by the Feller-Lindeberg central limit theorem that

(3.47) $$(2m)^{\frac{1}{2}-\gamma} m^{-1}\Sigma_{j=1}^m (\varepsilon_j'' - \varepsilon_j')/(2a_j) \to_{\mathcal{L}} N\left(0, \sigma^2/\{4^\gamma a^2(1 + 2\gamma)\}\right).$$

From (3.46) and (3.47), (3.18) follows immediately. ▯

PROOF OF THEOREM 6(i). Here b_i is $\mathcal{F}_{i-1}$-measurable and, therefore, the convergence of $\Sigma_1^N(\varepsilon_n' + \varepsilon_n'')/(nb_n)$ follows from the fact that $\Sigma_1^\infty(n\zeta_n)^{-2} < \infty$ and the martingale convergence theorem. The rest of the proof is the same as that of Theorem 6(ii), except that Theorem 5(i) can be used instead of Theorem 5(iv). ▯

A close examination of the preceding proof suggests that in order to reduce the cost C_n to the desired logarithmic order of magnitude we should choose the sequence $\{a_m\}$ such that

(3.48) $$\Sigma_1^m a_i^2 = o(\log m) \qquad \text{as} \quad n \to \infty.$$

This means that γ in condition (3.1) has to be chosen $\geqslant \frac{1}{2}$ instead. However, even with $\gamma = \frac{1}{2}$ and a_m modified to be of the form $m^{-\frac{1}{2}}(1 + \log m)^{-\delta}$ ($\delta > 0$) so that (3.48) holds, the relation (3.8) is no longer true, and such an a_m is too small for the b_i (or b_i^*) defined by (3.13) (or (3.20)) to be strongly consistent. In order to be able to choose a_m satisfying (3.48) instead of (3.1), we shall use another estimator b_m of

β at stage m. Let $\{a_m\}$ be any sequence of positive constants such that

(3.49) $$a_j = 0(j^{-\gamma}) \quad \text{for some} \quad \gamma > \tfrac{1}{4} \quad \text{but} \quad \Sigma_1^\infty a_j^2 = \infty.$$

Then, by the strong law,

(3.50) $$\Sigma_1^m a_j(\varepsilon_j'' - \varepsilon_j')/\left(\Sigma_1^m a_j^2\right) \to 0 \text{ a.s.}$$

Considering the linear case $M(x) = \beta(x - \theta)$, we note, in view of (3.7) and (3.50), that

(3.51) $$\tfrac{1}{2}\Sigma_1^m a_j(y_j'' - y_j')/\left(\Sigma_1^m a_j^2\right) \to \beta \text{ a.s.}$$

Thus, at least in the linear case, (3.51) gives a strongly consistent estimate of β under the minimal assumption that $\Sigma_1^m a_j^2 \to \infty$, no matter how slow the convergence may be. In the following theorem, we shall prove that this modification of Venter's scheme yields the desired growth rates for both C_n and $n^{\frac{1}{2}}(\tilde{\theta}_n - \theta)$, even in the nonlinear case.

THEOREM 7. (i) *Suppose that in Theorem* 6(i) *we replace the assumption* (3.1) *on the sequence* $\{a_m\}$ *by the weaker assumption* (3.49) *and also replace the definition* (3.13) *of* b_i *by*

(3.52) $$b_i = \zeta_i \vee \left\{\xi_i \wedge \tfrac{1}{2}\Sigma_{j=1}^{i-1} a_j(y_j'' - y_j')/\left(\Sigma_{j=1}^{i-1} a_j^2\right)\right\}, \qquad i \geqslant 2, b_1 = \zeta_1.$$

Then the relations (3.14), (3.15) *and* (3.16) *still hold. Moreover, instead of* (3.18), *we have*

(3.53) $$\left(\Sigma_{j=1}^m a_j^2\right)^{\frac{1}{2}}(b_m - \beta) \to_{\mathcal{L}} N\left(0, \tfrac{1}{2}\sigma^2\right).$$

If $\{a_m\}$ *further satisfies* (3.48), *then instead of* (3.17), *we have*

(3.54) $$C_n/\log n \to \sigma^2/\beta^2 \text{ a.s.}$$

(ii) *Let* $\{a_n\}$ *be a sequence of positive constants such that*

(3.55) $$a_m \sim am^{-\frac{1}{2}}(\log m)^{-\delta} \quad \textit{for some} \quad a > 0 \quad \textit{and} \quad 0 < \delta \leqslant \tfrac{1}{2}.$$

Then conditions (3.48) *and* (3.49) *are satisfied. Suppose that in Theorem* 6(i) *we replace the assumption* (3.1) *by* (3.55) *and also replace* b_i *as defined in* (3.13) *by*

(3.56) $$b_i^* = \zeta_i \vee \left\{\xi_i \wedge \tfrac{1}{2}\Sigma_{j=1}^{i} a_j(y_j'' - y_j')/\left(\Sigma_{j=1}^{i} a_j^2\right)\right\}.$$

Moreover, in place of (3.12b), *we assume that*

(3.57) $$\Sigma_2^\infty i^{-3/2}\zeta_i^{-2}(\log i)^{\delta - 1} < \infty.$$

Then (3.14), (3.15), (3.16), (3.53) *and* (3.54) *still hold.*

REMARK. While b_i, as defined by (3.52), is $\mathcal{F}_{i-1}$-measurable, b_i^* in (3.56) is not. Let $a_m \sim am^{-\gamma}(\log m)^{-\delta}$, where a and γ are positive constants and $\delta \geqslant 0$. Then

(3.48) fails to hold if $\gamma < \frac{1}{2}$. Obviously,

$$(3.49) \text{ holds} \Leftrightarrow \tfrac{1}{4} < \gamma \leqslant \tfrac{1}{2} \quad \text{and} \quad \delta \leqslant \tfrac{1}{2} \quad \text{when} \quad \gamma = \tfrac{1}{2}.$$

It is easy to see that the conditions (3.48) and (3.49) are both satisfied for the case $\gamma = \frac{1}{2}$ and $0 < \delta \leqslant \frac{1}{2}$.

PROOF OF THEOREM 7(i). As in Theorem 6(i), it is easy to show, using Lemma 5, that $x_m \to \theta$ a.s. Therefore (3.35) again holds. By (3.6), (3.35) and (3.50), we still have $b_m \to \beta$ a.s. Hence, as in Theorem 6(i), (3.15), (3.16), (3.43) and (3.44) still hold. Therefore, if condition (3.48) is also satisfied, then by (3.44), with probability 1,

$$C_n = 2\{\Sigma_1^m (x_i - \theta)^2 + \Sigma_1^m a_i^2\} \sim (\sigma^2/\beta^2)\log m,$$

and so (3.54) holds.

We note that, as in (3.45), with probability 1,

$$(3.58) \qquad M(x_j + a_j) - M(x_j - a_j) = 2a_j\{\beta + 0(a_j) + 0(|x_j - \theta|)\}.$$

By (3.49), $a_j = o(j^{-\rho})$ for some $\rho > \frac{1}{4}$. Therefore, by the Schwarz inequality,

$$(3.59) \qquad \Sigma_1^m a_j^3 = \Sigma_1^m o\big(a_j j^{-2\rho}\big) \leqslant \big(\Sigma_1^m o(a_j^2)\big)^{\frac{1}{2}}\big(\Sigma_1^m j^{-4\rho}\big)^{\frac{1}{2}} = o\Big(\big(\Sigma_1^m a_j^2\big)^{\frac{1}{2}}\Big).$$

Moreover, by (3.43) and the Schwarz inequality, with probability 1,

$$(3.60) \qquad \Sigma_1^m a_j^2 |x_j - \theta| = \Sigma_1^m o\big(a_j j^{-\frac{1}{2}-\rho}(\log\log j)^{\frac{1}{2}}\big) = o\Big(\big(\Sigma_1^m a_j^2\big)^{\frac{1}{2}}\Big).$$

From (3.58)–(3.60), it follows that, with probability 1,

$$(3.61) \quad \big\{\tfrac{1}{2}\Sigma_{j=1}^m a_j(M(x_j + a_j) - M(x_j - a_j))/\big(\Sigma_{j=1}^m a_j^2\big)\big\} - \beta = o\Big(\big(\Sigma_{j=1}^m a_j^2\big)^{-\frac{1}{2}}\Big).$$

By the Feller-Lindeberg central limit theorem,

$$(3.62) \qquad \big(\Sigma_{j=1}^m a_j^2\big)^{-\frac{1}{2}} \Sigma_{j=1}^m \tfrac{1}{2} a_j(\varepsilon_j'' - \varepsilon_j') \to N\big(0, \tfrac{1}{2}\sigma^2\big).$$

From (3.61) and (3.62), (3.53) follows immediately. □

PROOF OF THEOREM 7(ii). Define

(3.63)

$$\tilde{b}_i = \zeta_i \vee \big\{\xi_i \wedge \tfrac{1}{2}\big[a_i(M(x_i + a_i) - M(x_i - a_i)) + \Sigma_{j=1}^{i-1} a_j(y_j'' - y_j')\big]/\big(\Sigma_{j=1}^i a_j^2\big)\big\}.$$

Then $\tilde{b}_i$ is $\mathcal{F}_{i-1}$-measurable and, by (3.55),

$$(3.64) \qquad (nb_n^*)^{-1} - (n\tilde{b}_n)^{-1} = 0\Big(|\varepsilon_n'' - \varepsilon_n'|/\big\{n^{\frac{3}{2}}\zeta_n^2(\log n)^{1-\delta}\big\}\Big) \text{ a.s.}$$

Making use of (3.57) and using the same argument as in (3.32)–(3.34), it then follows that $\Sigma_1^N(\varepsilon_n'' + \varepsilon_n')/(nb_n^*)$ converges a.s. as $N \to \infty$. Hence, by Lemma 5, $x_m \to \theta$ a.s., and so, as in the proof of Theorem 7(i), $b_m^* \to \beta$ a.s. Therefore, by (3.55), with probability 1, for all large i

$$b_i^* - \tilde{b}_i = \tfrac{1}{2}a_i(\varepsilon_i'' - \varepsilon_i')/\big(\Sigma_{j=1}^i a_j^2\big) = o\big(i^{-\frac{1}{2}}(\log i)^{-\delta}(1 + |\varepsilon_i'| + |\varepsilon_i''|)\big).$$

By using the same argument as in the proof of Theorem 6(ii), it follows that (3.15), (3.16), (3.43) and (3.44) still hold. From (3.44) and (3.48), (3.54) follows. Moreover, (3.53) can be obtained by the same argument as in Theorem 7(i). □

4. Adaptive stochastic approximation schemes using least squares estimates of β. In the designs of Section 3, the main reason for choosing two levels x_m' and x_m'' (instead of x_m) at the mth stage is to be able to estimate β in a consistent way by using the differences $y_j'' - y_j'$ $(j = 1, 2, \cdots)$. As pointed out in (3.7), for the linear case $M(x) = \beta(x - \theta)$, these differences $y_j'' - y_j'$ depend only on β and not on θ. It is natural to ask whether consistent estimates b_n of β can be found for the adaptive Robbins-Monro scheme

$$x_{n+1} = x_n - y_n/(nb_n). \tag{4.1}$$

An obvious choice for b_n is the usual least squares estimate

$$\hat{\beta}_n = \Sigma_1^n(x_i - \bar{x}_n)(y_i - \bar{y}_n)/\Sigma_1^n(x_i - \bar{x}_n)^2, \tag{4.2}$$

where we set $\hat{\beta}_n$ equal to some positive constant b when $\Sigma_1^n(x_i - \bar{x}_n)^2 = 0$. In the linear case $M(x) = \beta(x - \theta)$, the strong consistency of $\hat{\beta}_n$ is equivalent to

$$\Sigma_1^n(x_i - \bar{x}_n)\varepsilon_i/\Sigma_1^n(x_i - \bar{x}_n)^2 \to 0 \text{ a.s.} \tag{4.3}$$

By Theorem 4, if $\hat{\beta}_n$ is indeed strongly consistent and $\sigma \neq 0$, then with probability 1, $n^{\frac{1}{2}}(\bar{x}_n - \theta) = 0((\log\log n)^{\frac{1}{2}})$, $\Sigma_1^n(x_i - \theta)^2 \sim (\sigma^2/\beta^2)\log n$, and therefore

$$\Sigma_1^n(x_i - \bar{x}_n)\varepsilon_i = \Sigma_1^n(x_i - \theta)\varepsilon_i + o\big(\Sigma_1^n(x_i - \theta)^2\big), \tag{4.4}$$

$$\Sigma_1^n(x_i - \bar{x}_n)^2 = \Sigma_1^n(x_i - \theta)^2 - n(\bar{x}_n - \theta)^2 \sim \Sigma_1^n(x_i - \theta)^2. \tag{4.5}$$

Thus if $\hat{\beta}_n$ is indeed strongly consistent and $\sigma \neq 0$, then the sequence $\{x_n\}$ will behave nicely in the sense of (4.4) and (4.5), which by the strong law for martingales in turn imply (4.3) and hence the strong consistency of $\hat{\beta}_n$. This suggests that a natural way of proving the strong consistency of $\hat{\beta}_n$ is to show that the design levels x_n satisfy (4.4) and (4.5). In the case where upper and lower bounds B and b (> 0) for β are known, we are able to elaborate this idea to prove the strong consistency of

$$b_n = b \vee (B \wedge \hat{\beta}_n) \tag{4.6}$$

and thereby to obtain from Theorem 4 the following

THEOREM 8. *Let $\varepsilon, \varepsilon_1, \cdots$ be* i.i.d. *random variables with $E\varepsilon = 0$ and $0 < E\varepsilon^2 = \sigma^2 < \infty$. Let $M(x)$ be a Borel function satisfying* (1.2)–(1.4), *and assume that $M(x)$ is continuously differentiable in some open neighborhood of θ. Let b, B be positive constants such that $b < \beta < B$. Let x_1 be a random variable independent of $\varepsilon_1, \varepsilon_2, \cdots$, and define inductively y_n, x_n, $\hat{\beta}_n$ and b_n by* (1.1), (4.1), (4.2) *and* (4.6). *Then* $\lim_{n\to\infty} b_n = \beta$ a.s., *and* (1.11), (1.12), (1.13) *still hold. If $M(x)$ further satisfies* (1.18), *then*

$$(\log n)^{\frac{1}{2}}(b_n - \beta) \to_{\mathfrak{L}} N(0, \beta^2). \tag{4.7}$$

The details of the proof of Theorem 8 are given in [10]. It is interesting to compare the asymptotic distribution (4.7) of b_n with the corresponding result (3.53) for the pairwise sampling scheme of Theorem 7. For this pairwise sampling scheme, which also satisfies (1.11)–(1.13), the relation (3.53) says that the consistent estimator b_m defined by (3.52) is asymptotically normal with variance $\frac{1}{2}\sigma^2/(\Sigma_1^m a_i^2)$, which is of a larger order of magnitude than $(\log m)^{-1}$ in view of (3.48). Thus the adaptive stochastic approximation scheme of Theorem 8 uses an asymptotically more efficient estimator of β than the pairwise sampling scheme of Theorem 7. Some simulation studies comparing the performance of these two kinds of adaptive stochastic approximation procedures for moderate sample sizes will be described in [12]. While Theorem 8 assumes that prior upper and lower bounds for β are known, we are able to remove this assumption by a modification of b_n in (4.6) and of the argument used. The details are given in [10].

REFERENCES

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

[2] Bojanic, R. and Seneta, E. (1973). A unified theory of regularly varying sequences. *Math. Z.* **134** 91–106.

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

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

[5] Freedman, D. (1971). *Brownian Motion and Diffusion*. Holden-Day, San Francisco.

[6] 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.

[7] Kersting, G. (1977). Almost sure approximation of the Robbins-Monro process by sums of independent random variables. *Ann. Probability* **5** 954–965.

[8] Lai, T. L. and Robbins, H. (1978a). Adaptive design in regression and control. *Proc. Nat. Acad. Sci. U.S.A.* **75** 586–587.

[9] Lai, T. L. and Robbins, H. (1978b). Limit theorems for weighted sums and stochastic approximation processes. *Proc. Nat. Acad. Sci. U.S.A.* **75** 1068–1070.

[10] Lai, T. L. and Robbins, H. (1978c). Consistency and asymptotic efficiency of slope estimates in adaptive stochastic approximation. Unpublished manuscript.

[11] Lai, T. L. and Robbins, H. (1978d). Iterated least squares and multiperiod control. Unpublished manuscript.

[12] Lai, T. L. and Robbins, H. (1978e). Adaptive design and estimation of the root of a regression function. Unpublished manuscript.

[13] Major, P. and Révész, P. (1973). A limit theorem for the Robbins-Monro approximation. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27** 79–86.

[14] Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.

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

[16] Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 375–405.

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

Department of Mathematical Statistics
Columbia University
Mathematics Building
New York, New York 10027