Z. Wahrscheinlichkeitstheorie verw. Gebiete
56, 329–360 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Consistency and Asymptotic Efficiency of Slope Estimates in Stochastic Approximation Schemes*

Tze Leung Lai and Herbert Robbins

Department of Mathematical Statistics, Columbia University, New York, NY 10027, USA

1. Introduction and Summary

Consider the general regression model

$$y_n = M(x_n) + \varepsilon_n \qquad (n=1,2,\ldots) \tag{1.1}$$

where the errors $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables with mean 0 and variance σ^2. We shall assume that the regression function $M(x)$ is a Borel function with a unique zero at θ such that the following three conditions are satisfied:

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

$$\inf_{\delta \leq |x-\theta| \leq \delta^{-1}} \{M(x)(x-\theta)\} > 0 \text{ for every } 0<\delta<1; \tag{1.3}$$

$$|M(x)| \leq A|x| + D \text{ for some } A, D > 0 \text{ and all } x. \tag{1.4}$$

Adaptive stochastic approximation schemes for choosing the levels of x at which y is to be observed are useful in applications of the following nature. Suppose that in (1.1) x_i is the dosage level of a drug given to the i-th patient who turns up for treatment and that y_i is the response of the patient. Suppose also that an optimal response value y^* is desired. Without loss of generality, we shall (replacing y_i by $y_i - y^*$ if necessary) assume that $y^*=0$. If θ 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 as to approach the unknown θ as rapidly as possible? Calling $\sum_1^n (x_i-\theta)^2$ the *cost* of the design at stage n, we have shown in [10] that in the case where $\beta(=M'(\theta))$ is known, the apparent dilemma of having to choose between a small cost (of interest to current patients) and a good estimate of θ (of interest perhaps to future patients) can be resolved, at least for large n, by using the stochastic approxi-

* Research supported by the National Institutes of Health and the National Science Foundation under grants 1 R01-GM-27976 and NSF-MCS-79-05811

0044-3719/81/0056/0329/$06.40

mation scheme

$$x_{n+1}=x_n-y_n/(n\beta) \quad (n=1,2,\ldots) \tag{1.5}$$

with x_1 = initial guess of θ. In practice, although β is usually unknown, Theorem 1 below shows that the desirable asymptotic properties of the scheme (1.5) still hold if we replace β in (1.5) by a strongly consistent estimator b_n of β. As in [10], we shall call the stochastic approximation scheme

$$x_{n+1}=x_n-y_n/(nb_n) \quad (n=1,2,\ldots) \tag{1.6}$$

adaptive if $\{b_n\}$ is a sequence of positive random variables such that $\lim_{n\to\infty} b_n=\beta$ a.s. The following theorem, proved in [10], gives some asymptotic properties of adaptive stochastic approximation schemes.

Theorem 1. *Let $M(x)$ be a Borel function satisfying* (1.2)–(1.4). *Suppose that $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables with $E\varepsilon=0$ and $0<E\varepsilon^2=\sigma^2<\infty$. Let x_1 be a random variable independent of $\varepsilon_1, \varepsilon_2, \ldots$. Let $\mathfrak{F}_0$ denote the σ-field generated by x_1, and for $k\geqq 1$ let $\mathfrak{F}_k$ denote the σ-field generated by $x_1, \varepsilon_1, \ldots, \varepsilon_k$. Let $\{b_n\}$ be a sequence of positive random variables such that b_n is $\mathfrak{F}_{n-1}$ - measurable for all $n\geqq 1$ and $\lim_{n\to\infty} b_n=\beta$ a.s. For $n=1,2,\ldots$, define inductively y_n by* (1.1) *and x_{n+1} by* (1.6). *Then*

$$n^{\frac{1}{2}}(x_n-\theta)\xrightarrow{\mathfrak{L}} N(0,\sigma^2/\beta^2) \quad \textit{as } n\to\infty; \tag{1.7}$$

(1.8) $\lim_{n\to\infty} x_n=\theta$ *a.s., and in fact*

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

$$\lim_{n\to\infty}\left\{\sum_1^n (x_i-\theta)^2/\log n\right\}=\sigma^2/\beta^2 \quad a.s. \tag{1.9}$$

In ignorance of β, an obvious choice for b_n in the stochastic approximation scheme (1.6) is the usual least squares estimate

$$\hat{\beta}_n=\frac{\sum_1^n (x_i-\bar{x}_n)y_i}{\sum_1^n (x_i-\bar{x}_n)^2}, \tag{1.10}$$

at least in the linear case $M(x)=\beta(x-\theta)$. (Here and in the sequel, we use $\bar{a}_n$ to denote the arithmetic mean of any n numbers $a_1, \ldots, a_n$.) However, since the levels x_i are sequentially determined random variables, it is not obvious that $\hat{\beta}_n$ will in fact converge to β a.s. as $n\to\infty$. In the case where positive upper and lower bounds B and b for β are known, as often occurs in practice, it is natural to truncate $\hat{\beta}_n$ above and below by B and b. For the stochastic approximation scheme (1.6) with $b_n=b\vee(\hat{\beta}_n\wedge B)$, where $\vee$ and $\wedge$ denote maximum and minimum respectively, we establish in Sect. 5 that $\hat{\beta}_n$, and therefore b_n also, indeed converge a.s. to β. This is the content of

Theorem 2. *Let $M(x)$, β, θ, σ; x_1; $\varepsilon_1, \varepsilon_2, \ldots$ be as in Theorem 1, and assume that $M(x)$ is continuously differentiable in some open neighborhood of θ. Let b, B be positive constants such that $b<\beta<B$. Define inductively y_n, x_{n+1}, $\hat{\beta}_n$ by* (1.1), (1.6), (1.10), *and define*

(1.11) $b_n = b \vee (\hat{\beta}_n \wedge B)$ *for* $n \geqq n_0$, *where*

$$n_0 = \inf\left\{k: \sum_{i=1}^{k} (x_i - \bar{x}_k)^2 > 0\right\},$$

setting b_n equal to some constant c between b and B for $n<n_0$. Then

(1.12) $\hat{\beta}_n \to \beta$ *a.s. and therefore* $b_n \to \beta$ *a.s.*

Moreover, (1.7), (1.8), *and* (1.9) *still hold.*

In a recent paper, Anbar ([1], Theorem 2) has considered the strong consistency of $\hat{\beta}_n$ in Theorem 2 under the further assumption that $B<2b$. Although in practice it is often possible to give upper and lower bounds for β, the requirement that the upper bound be less than twice the lower bound is rather restrictive and appears somewhat artificial. However, Anbar's proof of the strong consistency of $\hat{\beta}_n$ depends very heavily on the extra condition that $B<2\beta$ in the assumption $b<\beta<B$. By using a different approach, we are able to remove in Theorem 2 this unduly stringent condition. We prove Theorem 2 by making use of a general theorem on the strong consistency of $\hat{\beta}_n$ in stochastic designs established in Sect. 2 together with some general results, established in Sects. 3 and 4, on the order of magnitude of $x_n - \theta$ and of $\sum_1^n (x_i-\theta)^2$ for general stochastic approximation schemes of the type (1.6), where $\{b_n\}$ is an arbitrary sequence of positive random variables (not necessarily convergent) satisfying some weak conditions on boundedness and measurability.

While Theorem 2 assumes known positive bounds on β so that the least squares estimate $\hat{\beta}_n$ can be truncated below and above by these bounds in the definition of b_n in (1.11), the general results on stochastic approximation and least squares estimation developed in Sects. 2, 3, 4 also enable us to deal with the case in which prior lower and upper bounds for β are not known. We handle this general case by using a "stochastic" truncation of $\hat{\beta}_n$ in the definition of b_n in the following

Theorem 3. *Let $d<1<D$ and c be positive constants. Let $\{f(n)\}$ and $\{g(n)\}$ be two nondecreasing sequences of positive constants such that $\lim_{n\to\infty} f(n) = \lim_{n\to\infty} g(n) = \infty$, $f(1)\,g(1)>1$, and*

$$\sum_1^\infty \frac{1}{j g(j) f^2(j)} = \infty, \tag{1.13}$$

(1.14) $\{g(n)\}$ *is slowly varying (i.e.,* $\lim_{n\to\infty} g([an])/g(n) = 1$ *for all* $a>0$),

$$g(n) = o\left(\left\{\sum_1^n \frac{1}{j g(j) f^2(j)}\right\}^\eta\right) \quad \textit{for some } \eta < 1/4. \tag{1.15}$$

Let $\{m(n)\}$ be a sequence of positive integers such that

$$(1.16)\qquad m(n)\to\infty \quad \text{and} \quad \log m(n)=o\left(\sum_1^n \frac{1}{jf(j)}\right) \quad \text{as} \quad n\to\infty,$$

$$(1.17)\qquad m(1)=1 \quad \text{and} \quad m(n)\leqq n-1 \quad \text{for} \quad n\geqq 2.$$

Suppose that in Theorem 2 we replace b_n in (1.11) by

$$(1.18)\qquad b_n=c \quad \text{if} \quad \sum_1^{m(n)}(x_i-\bar{x}_{m(n)})^2=0,$$
$$=(g(n))^{-1}\vee d\beta_n^*\vee\{\hat{\beta}_n\wedge D\beta_n^*\wedge f(n)\} \text{ otherwise,}$$

where

$$(1.19)\qquad \beta_n^*=\frac{\sum_1^{m(n)}(x_i-x_n)\,y_i}{\sum_1^{m(n)}(x_i-x_n)^2}.$$

Then with probability 1,

$$(1.20)\qquad \lim_{n\to\infty}\hat{\beta}_n=\beta, \quad \text{and} \quad b_n=\hat{\beta}_n \quad \text{for all large } n.$$

Moreover, (1.7), (1.8), *and* (1.9) *still hold.*

A simple example of $\{f(n)\}$, $\{g(n)\}$, and $\{m(n)\}$ satisfying the assumptions of the above theorem is $f(n)=g(n)=a+b(\log n)^\rho$ and $m(n)=[\exp((\log n)^\delta)]\wedge\{(n-1)\vee 1\}$, where $a>1$, $b>0$, $0<\rho<1/7$, and $0<\delta<1-\rho$. The proof of Theorem 3 is given in Sect. 6, where we also discuss the rationale behind the use of $d\beta_n^*$ and $D\beta_n^*$ to truncate the least squares estimate $\hat{\beta}_n$ in the definition of b_n in (1.18).

In Sect. 7, we show that the slope estimate $\hat{\beta}_n$ in the adaptive stochastic approximation scheme of Theorem 2 or Theorem 3 has an asymptotically normal distribution and that, in the linear regression model with normally distributed errors, $\hat{\beta}_n$ is an asymptotically efficient estimate of β. This is the content of

Theorem 4. *Suppose that in Theorem 2 or Theorem 3 we replace* (1.2) *by the stronger assumption that as $x\to\theta$*

$$(1.21)\qquad M(x)=\beta(x-\theta)+0(|x-\theta|^{1+\eta}) \text{ for some } \eta>0 \text{ and } \beta>0.$$

Then

$$(1.22)\qquad (\log n)^{\frac{1}{2}}(\hat{\beta}_n-\beta)\xrightarrow{\mathfrak{L}}N(0,\beta^2),$$

and therefore

$$(1.23)\qquad (\log n)^{\frac{1}{2}}(b_n-\beta)\xrightarrow{\mathfrak{L}}N(0,\beta^2).$$

Moreover, if $M(x)=\beta(x-\theta)$ and the ε_i are i.i.d. $N(0,\sigma^2)$ random variables, then $\hat{\beta}_n$ is an asymptotically efficient estimator of β in the sense that given any $r>0$

and β,

(1.24) $$\limsup_{n\to\infty} P_\beta[(\log n)^{\frac{1}{2}}|T_n-\beta|\leqq r] \leqq \lim_{n\to\infty} P_\beta[(\log n)^{\frac{1}{2}}|\hat{\beta}_n-\beta|\leqq r]$$

for any other estimator $T_n=T_n(x_1, y_1, \ldots, x_n, y_n)$ *such that if* $\{\gamma_n\}$ *is a sequence of constants satisfying* $\gamma_n=0((\log n)^{-\frac{1}{2}})$ *then*

(1.25) $$\lim_{n\to\infty} \{P_{\beta+\gamma_n}[(\log n)^{\frac{1}{2}}|T_n-\beta|\leqq r]-P_\beta[(\log n)^{\frac{1}{2}}|T_n-\beta|\leqq r]\}=0.$$

The notation P_β (or $P_{\beta+\gamma_n}$) in (1.24) and (1.25) signifies that the unknown slope parameter of the linear regression function is equal to β (or $\beta+\gamma_n$). As pointed out by Weiss and Wolfowitz ([13], pp. 19–20), the regularity condition (1.25) on the estimator T_n rules out super-efficiency.

In [12] Venter proposed a modified version of the stochastic approximation scheme (1.6) to obtain successive estimates of β which are consistent. His scheme requires that at the n-th stage ($n=1,2,\ldots$) two observations y_n' and y_n'' be taken, at levels $x_n'=x_n-a_n$ and $x_n''=x_n+a_n$, where $\{a_n\}$ is a sequence of positive constants such that

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

and x_n is the n-th approximation to θ defined by (1.6) with

(1.27) $$y_n=\tfrac{1}{2}(y_n'+y_n'').$$

Assuming that positive constants b and B are known such that $b<\beta<B$, Venter defines the slope estimate b_n in (1.6) by

(1.28) $$b_n=b\vee\left\{B\wedge n^{-1}\sum_{i=1}^{n}(y_i''-y_i')/(2a_i)\right\}.$$

At stage n, $2n$ observations have been taken, and assuming $M(x)$ to be twice continuously differentiable in some open neighborhood of θ, we have shown in [10] that the cost $C_n=\sum_1^n(x_i'-\theta)^2+\sum_1^n(x_i''-\theta)^2$ grows like $\rho n^{1-2\gamma}$ where ρ is a positive constant depending on a and γ. In order that the cost be of logarithmic order, we have to take a_n smaller than (1.26) so that

(1.29) $$\sum_1^n a_i^2=o(\log n).$$

But then b_n as defined in (1.28) fails to be strongly consistent (cf. [10]), and we proposed in [10] the alternative estimator

(1.30) $$\tilde{b}_n=b\vee\left\{B\wedge\tfrac{1}{2}\sum_{i=1}^{n-1}a_i(y_i''-y_i')\bigg/\left(\sum_{i=1}^{n-1}a_i^2\right)\right\},\qquad n\geqq 2,\ (b_1=b).$$

Using $\tilde{b}_n$ as defined in (1.30) and choosing a_n such that (1.29) holds and

$$a_n = 0(n^{-\gamma}) \quad \text{for some } \gamma > 1/4 \text{ but } \sum_1^\infty a_n^2 = \infty, \tag{1.31}$$

we showed in [10] that the cost C_n indeed grows like $(\sigma^2/\beta^2)\log n$ and that

$$\left(\sum_1^n a_j^2\right)^{\frac{1}{2}} (\tilde{b}_n - \beta) \xrightarrow{\mathfrak{L}} N(0, \tfrac{1}{2}\sigma^2). \tag{1.32}$$

In view of (1.29), comparison of (1.32) with (1.22) shows that $\tilde{b}_n$ is asymptotically much less efficient than $\hat{\beta}_n$. Thus the adaptive stochastic approximation scheme of Theorem 2 or Theorem 3 uses an asymptotically more efficient estimator of β than the Venter-type pairwise sampling scheme with $\{a_n\}$ satisfying (1.29) and (1.31). Some simulation studies comparing the performance of these two kinds of adaptive stochastic approximation schemes for moderate sample sizes will be reported elsewhere.

2. Strong Consistency and Asymptotic Normality of $\hat{\beta}_n$ in Stochastic Designs

For the regression model (1.1) with i.i.d. errors $\varepsilon_1, \varepsilon_2, \ldots$ such that $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = \sigma^2 < \infty$, we shall assume throughout this section that the levels x_n are random variables such that

$$x_n \text{ is } \mathfrak{B}_{n-1}\text{-measurable for all } n \geqq 1, \tag{2.1}$$

where $\mathfrak{B}_0 \subset \mathfrak{B}_1 \subset \ldots$ is an increasing sequence of σ-fields such that ε_n is $\mathfrak{B}_n$-measurable and is independent of $\mathfrak{B}_{n-1}$ for all $n \geqq 1$. Defining $\hat{\beta}_n$ as in (1.10) if $\sum_1^n (x_i - \bar{x}_n)^2 > 0$ and setting $\hat{\beta}_n$ equal to some constant b otherwise, we note that

$$\begin{aligned} \hat{\beta}_n &= \frac{\sum_1^n (x_i - \bar{x}_n)(M(x_i) + \varepsilon_i)}{\sum_1^n (x_i - \bar{x}_n)^2} \quad \text{if } \sum_1^n (x_i - \bar{x}_n)^2 > 0, \\ &= b \quad \text{otherwise.} \end{aligned} \tag{2.2}$$

In this section we establish some general results on the a.s. convergence of $\hat{\beta}_n$ and on the asymptotic normality of $\hat{\beta}_n$. These results will be applied in Sects. 5, 6, 7 to prove Theorems 2, 3, 4. We begin with conditions for the asymptotic normality of $\hat{\beta}_n$ in stochastic designs in the following

Theorem 5. *Let $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. with $E\varepsilon = 0$ and $E\varepsilon^2 = \sigma^2 < \infty$. Let $\{x_n\}$ be a sequence of random variables satisfying* (2.1).

(i) *Assume that there exists a constant θ and a sequence of positive constants $\{A_n\}$ such that*

(2.3) $$A_n \to \infty,$$

(2.4) $$\sum_1^n (x_i-\theta)^2/A_n \xrightarrow{P} 1,$$

and

(2.5) $$\max_{1\leq i\leq n} (x_i-\theta)^2/A_n \xrightarrow{P} 0.$$

Then

$$\left\{\sum_1^n (x_i-\theta)\,\varepsilon_i\right\}/A_n^{\frac{1}{2}} \xrightarrow{\mathfrak{L}} N(0,\sigma^2).$$

(ii) *Assume further that*

(2.6) $$(n/A_n)^{\frac{1}{2}}(\bar{x}_n-\theta) \xrightarrow{P} 0.$$

Then $\sum_1^n (x_i-\bar{x}_n)^2/A_n \xrightarrow{P} 1$ *and* $\left\{\sum_1^n (x_i-\bar{x}_n)\,\varepsilon_i\right\}/A_n^{\frac{1}{2}} \xrightarrow{\mathfrak{L}} N(0,\sigma^2)$. *Consequently, if* $M(x)=\alpha+\beta x$ *and* $\hat{\beta}_n$ *is defined by* (2.2), *then*

(2.7) $$A_n^{\frac{1}{2}}(\hat{\beta}_n-\beta) \xrightarrow{\mathfrak{L}} N(0,\sigma^2).$$

(iii) *Let* $M(x)$ *be a Borel function such that, as* $x\to\theta$,

(2.8) $$M(x)=M(\theta)+\beta(x-\theta)+O(|x-\theta|^{1+\eta}) \quad \textit{for some } \eta>0.$$

Assume that (2.3), (2.4), *and* (2.6) *hold and that*

(2.9) $$\sum_1^\infty |x_i-\theta|^{2\rho}<\infty \ \textit{a.s.} \quad \textit{for some } 1<\rho<1+\eta.$$

Then $\lim_{n\to\infty} x_n=\theta$ *a.s. and* (2.5) *also holds. Moreover, defining* $\hat{\beta}_n$ *as in* (2.2), (2.7) *still holds.*

Proof. Let $x_{ni}=(x_i-\theta)/A_n^{\frac{1}{2}}$ and $x'_{ni}=x_{ni}I_{[|x_{ni}|\leq 1]}$. By (2.5),

(2.10) $$P[x_{ni}\neq x'_{ni} \text{ for some } i=1,\ldots,n]\to 0 \quad \text{as} \quad n\to\infty.$$

Moreover, $E|x'_{ni}\varepsilon_i|<\infty$ and $E(x'_{ni}\varepsilon_i\,|\,\mathfrak{B}_{i-1})=0$, and by (2.4) and (2.10),

(2.11) $$\sum_1^n E((x'_{ni}\varepsilon_i)^2\,|\,\mathfrak{B}_{i-1})=\sigma^2\sum_1^n (x'_{ni})^2 \xrightarrow{P} \sigma^2.$$

Using (2.5) and (2.11), it can be shown that

(2.12) $$\sum_1^n E[(x'_{ni}\varepsilon_i)^2 I_{[(x'_{ni}\varepsilon_i)^2>\delta]}\,|\,\mathfrak{B}_{i-1}] \xrightarrow{P} 0 \quad \text{for every } \delta>0.$$

Hence by (2.10) and a theorem of Dvoretzky (see Theorem 2.2 and Sect. 3.1 of [5]), (i) follows. Using (i), (2.2) and the central limit theorem for $\sum_1^n \varepsilon_i$, we obtain (ii).

To prove (iii), first note that $\lim_{n\to\infty} x_n=\theta$ a.s. by (2.9) and hence (2.5) holds by (2.3). Therefore, in view of (ii), (2.2), and (2.8), we need only show that

$$\left\{\sum_1^n (x_i-\bar{x}_n)u_i\right\}\Big/\left\{\sum_1^n (x_i-\bar{x}_n)^2\right\}^{\frac{1}{2}}\to 0 \text{ a.s.} \quad \text{on } \left[\sum_1^n (x_i-\bar{x}_n)^2\to\infty\right], \tag{2.13}$$

where $\{u_n\}$ is a sequence of random variables such that with probability 1

$$|u_n|=O(|x_n-\theta|^{1+\eta}). \tag{2.14}$$

Take $\delta>0$ such that $(1-\delta)(1+\eta)>\rho$, where ρ is given by (2.9). By the Schwarz inequality,

$$\begin{aligned}\left|\sum_1^n (x_i-\bar{x}_n)u_i\right| &\leqq \left\{\sum_1^n (x_i-\bar{x}_n)^2|u_i|^{2\delta}\right\}^{\frac{1}{2}}\left\{\sum_1^n |u_i|^{2(1-\delta)}\right\}^{\frac{1}{2}}\\ &=o\left(\left\{\sum_1^n (x_i-\bar{x}_n)^2\right\}^{\frac{1}{2}}\right) \text{ a.s.} \quad \text{on } \left[\sum_1^n (x_i-\bar{x}_n)^2\to\infty\right],\end{aligned}$$

since $u_n\to 0$ a.s. and $\sum_1^\infty |u_i|^{2(1-\delta)}<\infty$ a.s. by (2.9) and (2.14). ∎

We now consider the question of strong consistency of $\hat{\beta}_n$ for stochastic designs. For multiple regression models with stochastic regressors, Anderson and Taylor [2] have recently considered the strong consistency of least squares estimates under certain conditions on the design. When specialized to the model $M(x)=\alpha+\beta x$, their strong consistency result requires, among other conditions, the condition that

$$\sum_1^n (x_i-\bar{x}_n)^2\gg n \quad \text{a.s.} \tag{2.15}$$

(Here and in the sequel, we shall sometimes use Vinogradov's symbol $\ll$ instead of Landau's 0 notation. Thus, the order relation $c_n\ll d_n$ between two nonnegative sequences c_n and d_n means that there exist positive constants K and m such that $c_n\leqq K d_n$ for all $n\geqq m$.) The following theorem considerably weakens the condition (2.15) to an analogous condition which we shall demonstrate to be in a sense the weakest possible.

Theorem 6. *Let ε, ε_1, ε_2, ... be i.i.d. with $E\varepsilon=0$ and $E\varepsilon^2<\infty$. Let $\{x_n\}$ be a sequence of random variables satisfying* (2.1).

(i) *Assume that*

$$\lim_{n\to\infty}\sum_1^n (x_i-\bar{x}_n)^2/(\log n)=\infty \quad a.s. \tag{2.16}$$

Then as $n\to\infty$,

$$\left\{\sum_1^n (x_i-\bar{x}_n)\varepsilon_i\right\}\Big/\sum_1^n (x_i-\bar{x}_n)^2\to 0 \quad a.s. \tag{2.17}$$

(ii) *Let* $M(x)=\alpha+\beta x$ *and define* $\hat{\beta}_n$ *by* (2.2). *If* (2.16) *holds, then* $\hat{\beta}_n \to \beta$ *a.s.*

(iii) *Let* $M(x)$ *be a Borel function that is continuously differentiable in some open neighborhood of a given point* θ. *Define* $\hat{\beta}_n$ *as in* (2.2). *If* (2.16) *holds and*

$$\lim_{n\to\infty} x_n = \theta \quad a.s., \tag{2.18}$$

then $\hat{\beta}_n \to M'(\theta)$ *a.s.*

We note that in Theorem 6, $\left\{\sum_1^n x_i \varepsilon_i, \mathfrak{B}_n, n \geqq 1\right\}$ is a martingale transform but need not be a martingale, since $E(x_n \varepsilon_n)$ may be undefined. The following lemma lists some properties of such martingale transforms that we use to prove Theorem 6 and other results in the sequel. We note that if a_n is a sufficiently large constant such that $P[|x_n|>a_n] \leqq n^{-2}$ for every n, then $P[x_n = x'_n$ for all large $n]=1$, where $x'_n = x_n I_{[|x_n| \leqq a_n]}$. Moreover, $E|x'_n \varepsilon_n| < \infty$ and so $\sum_1^n x'_i \varepsilon_i$ is a martingale. In view of this truncation argument, parts (i) and (ii) of the following lemma follow from the (local) convergence theorem and the strong law for martingales (cf. [11], pp. 148–150). Part (iii) is a special case of a theorem of Freedman ([6], p. 919), while part (iv) follows from part (iii) and the Kronecker lemma, since $\sum_1^\infty \left\{|u_n| \Big/ \left(1+\sum_1^n |u_i|\right)^\lambda\right\} < \infty$ for $\lambda > 1$.

Lemma 1. *Let* $z, z_1, z_2, \ldots$ *be i.i.d. random variables such that* $E|z|<\infty$. *Let* $\mathfrak{B}_0 \subset \mathfrak{B}_1 \subset \ldots$ *be an increasing sequence of* σ*-fields such that* z_n *is* $\mathfrak{B}_n$*-measurable and is independent of* $\mathfrak{B}_{n-1}$ *for* $n \geqq 1$. *Let* $\{u_n\}$ *be a sequence of random variables such that* u_n *is* $\mathfrak{B}_{n-1}$*-measurable for all* $n \geqq 1$.

(i) *If* $Ez=0$ *and* $Ez^2<\infty$, *then* $\sum_1^n u_i z_i$ *converges a.s. on* $\left[\sum_1^\infty u_i^2 < \infty\right]$.

(ii) *If* $Ez=0$ *and* $Ez^2<\infty$, *then for every* $\eta > \frac{1}{2}$,

$$\left(\sum_1^n u_i z_i\right) \Big/ \left(\sum_1^n u_i^2\right)^\eta \to 0 \quad a.s. \qquad on \quad \left[\sum_1^\infty u_i^2 = \infty\right].$$

(iii) $\sum_1^\infty |u_i z_i| < \infty$ *a.s.* $\quad$ *on* $\left[\sum_1^\infty |u_i| < \infty\right]$.

(iv) *For every* $\lambda > 1$,

$$\left(\sum_1^n |u_i z_i|\right) \Big/ \left(\sum_1^n |u_i|\right)^\lambda \to 0 \quad a.s. \qquad on \quad \left[\sum_1^\infty |u_i| = \infty\right].$$

Proof of Theorem 6. Simple algebra (cf. [8]) shows that

$$\sum_1^n (x_i - \bar{x}_n)\varepsilon_i = \sum_2^n \left(\frac{i-1}{i}\right)(x_i - \bar{x}_{i-1})(\varepsilon_i - \bar{\varepsilon}_{i-1}), \tag{2.19}$$

and

$$\sum_1^n (x_i - \bar{x}_n)^2 = \sum_2^n \left(\frac{i-1}{i}\right)(x_i - \bar{x}_{i-1})^2. \tag{2.20}$$

Let $d_j = j^{-1}(j-1)(x_j - \bar{x}_{j-1})$, $j \geqq 2$. Then d_j is $\mathfrak{B}_{j-1}$-measurable and therefore $\left\{\sum_2^n d_j \varepsilon_j, \mathfrak{B}_j, j \geqq 2\right\}$ is a martingale transform. Assume that (2.16) holds. Then by (2.16) and (2.20), $\sum_2^n d_j^2 \sim \sum_1^n (x_i - \bar{x}_n)^2 \to \infty$ a.s. Therefore by Lemma 1(ii), with probability 1,

$$\left(\sum_2^n d_j \varepsilon_j\right) \Big/ \sum_1^n (x_i - \bar{x}_n)^2 \to 0. \tag{2.21}$$

We note that by the Schwarz inequality, with probability 1,

$$\left|\sum_2^n d_j \bar{\varepsilon}_{j-1}\right| \leqq \left(\sum_2^n d_j^2\right)^{\frac{1}{2}} \left(\sum_2^n \bar{\varepsilon}_{j-1}^2\right)^{\frac{1}{2}} \sim \left(\sum_1^{n-1} \bar{\varepsilon}_j^2\right)^{\frac{1}{2}} \left\{\sum_1^n (x_i - \bar{x}_n)^2\right\}^{\frac{1}{2}}. \tag{2.22}$$

By Corollary 1 of [10], with probability 1,

$$\sum_1^n \bar{\varepsilon}_j^2 \sim (E\varepsilon^2) \log n. \tag{2.23}$$

From (2.16), (2.22), and (2.23), it then follows that with probability 1

$$\left(\sum_2^n d_j \bar{\varepsilon}_{j-1}\right) \Big/ \sum_1^n (x_i - \bar{x}_n)^2 \to 0. \tag{2.24}$$

From (2.19), (2.21), and (2.24), (i) follows. In view of (2.2), (ii) is an immediate corollary of (i).

To prove (iii), let $g(x) = M(x) - \beta x$. By (2.2) and (2.17), to show that $\hat{\beta}_n \to \beta$ a.s., it suffices to prove that

$$\left\{\sum_1^n (x_i - \bar{x}_n)\, g(x_i)\right\} \Big/ \sum_1^n (x_i - \bar{x}_n)^2 \to 0 \quad \text{a.s.} \tag{2.25}$$

Since $\lim_{n\to\infty} x_n = \theta$ a.s., $g'(\theta) = 0$, and g is continuously differentiable in some neighborhood of θ, we obtain that with probability 1

$$\left|\sum_1^n (x_i - \bar{x}_n)\, g(x_i)\right| \leqq \sum_1^n |x_i - \bar{x}_n|\, |g(x_i) - g(\bar{x}_n)| = o\left(\sum_1^n (x_i - \bar{x}_n)^2\right),$$

and therefore (2.25) follows. ∎

It is natural to ask whether in Theorem 6 the condition (2.16) that $\sum_1^n (x_i - \bar{x}_n)^2$ grows faster than $\log n$ can be weakened to, say, the condition that

$$\sum_1^n (x_i - \bar{x}_n)^2 \gg \log n \quad \text{a.s.} \tag{2.26}$$

The answer to this question turns out to be negative, as is shown by the following

Example. Let $c \neq 0$ be a real constant and let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = 1$. Let $x_1 = 0$ and for $n \geqq 1$ define

$$x_{n+1} = \bar{x}_n + c\bar{\varepsilon}_n. \tag{2.27}$$

By (2.20), (2.23), and (2.27),

$$\sum_1^n (x_i - \bar{x}_n)^2 = c^2 \sum_2^n \left(\frac{i-1}{i}\right) \bar{\varepsilon}_{i-1}^2 \sim c^2 \log n \quad \text{a.s.} \tag{2.28}$$

Therefore (2.26) is satisfied. By (2.19), (2.27), and (2.28),

$$\frac{\sum_1^n (x_i - \bar{x}_n)\varepsilon_i}{\sum_1^n (x_i - \bar{x}_n)^2} = \frac{1}{c} \frac{\sum_2^n \left(\frac{i-1}{i}\right) \bar{\varepsilon}_{i-1}(\varepsilon_i - \bar{\varepsilon}_{i-1})}{\sum_2^n \left(\frac{i-1}{i}\right) \bar{\varepsilon}_{i-1}^2} \to -c^{-1} \quad \text{a.s.,} \tag{2.29}$$

using the strong law for the martingale $\left\{\sum_1^n i^{-1}(i-1)\bar{\varepsilon}_{i-1}\varepsilon_i, n \geqq 1\right\}$. Hence (2.17) fails to hold. Consequently, even in the linear case $M(x) = \alpha + \beta x$, $\hat{\beta}_n$ fails to be strongly consistent; in fact, by (2.2) and (2.29), $\hat{\beta}_n \to \beta - c^{-1}$ a.s.

If the x_n are constants, then the condition $\sum_1^n (x_i - \bar{x}_n)^2 \to \infty$ is both necessary and sufficient for (2.17) to hold (cf. [8]). The above example shows that if the x_n are sequentially determined random variables, then (2.17) may fail to hold even when $\sum_1^n (x_i - \bar{x}_n)^2$ grows like $\log n$. Therefore if $\sum_1^n (x_i - \bar{x}_n)^2$ does not grow faster than $\log n$ (as is the case in adaptive stochastic approximation schemes in view of Theorem 1), additional assumptions on x_n have to be imposed to ensure that (2.17) holds. The following theorem allows very slow growth rates for $\sum_1^n (x_i - \bar{x}_n)^2$, but requires $|\bar{x}_n - \theta|$ to be sufficiently small when compared with $\sum_1^n (x_i - \theta)^2$.

Theorem 7. *Let $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. with $E\varepsilon = 0$ and $E\varepsilon^2 < \infty$. Let $\{x_n\}$ be a sequence of random variables satisfying* (2.1). *Assume that there exists a constant θ such that with probability* 1

$$\sum_1^\infty (x_i - \theta)^2 = \infty, \tag{2.30}$$

$$\limsup_{n \to \infty} n(\bar{x}_n - \theta)^2 \Big/ \sum_1^n (x_i - \theta)^2 < 1, \tag{2.31}$$

$$(n \log \log n)^{\frac{1}{2}} |\bar{x}_n - \theta| = o\left(\sum_1^n (x_i - \theta)^2\right). \tag{2.32}$$

Then (2.17) *holds. Consequently, if* $M(x)$ *is a Borel function that is continuously differentiable in some neighborhood of* θ *and* $\lim_{n\to\infty} x_n=\theta$ *a.s., then* $\hat\beta_n \to M'(\theta)$ *a.s., where* $\hat\beta_n$ *is defined in* (2.2).

Remark. For multiple regression models with stochastic regressors, Drygas [4] has obtained sufficient conditions for the least squares estimate to be consistent. When specialized to the model $M(x)=\alpha+\beta x$, his conditions imply (2.30), (2.31), and

$$n|\bar x_n-\theta| \ll \sum_1^n (x_i-\theta)^2, \tag{2.33}$$

which is much stronger than (2.32).

Proof of Theorem 7. Since (2.31) holds a.s.,

$$\sum_1^n (x_i-\theta)^2 \ll \sum_1^n (x_i-\bar x_n)^2 \quad \text{a.s.} \tag{2.34}$$

Therefore with probability 1

$$\frac{\left|\sum_1^n (x_i-\bar x_n)\,\varepsilon_i\right|}{\sum_1^n (x_i-\bar x_n)^2} \ll \frac{\left|\sum_1^n (x_i-\theta)\,\varepsilon_i\right|}{\sum_1^n (x_i-\theta)^2} + \frac{|\bar x_n-\theta|\left|\sum_1^n \varepsilon_i\right|}{\sum_1^n (x_i-\theta)^2}. \tag{2.35}$$

By (2.30) and Lemma 1(ii),

$$\sum_1^n (x_i-\theta)\,\varepsilon_i \Big/ \sum_1^n (x_i-\theta)^2 \to 0 \quad \text{a.s.} \tag{2.36}$$

By the law of the iterated logarithm, with probability 1,

$$\begin{aligned} |\bar x_n-\theta|\left|\sum_1^n \varepsilon_i\right| &\ll (n\log\log n)^{\frac12}|\bar x_n-\theta| \\ &= o\left(\sum_1^n (x_i-\theta)^2\right), \quad \text{by (2.32).} \end{aligned} \tag{2.37}$$

From (2.35), (2.36), and (2.37), we obtain (2.17). ∎

In view of (1.8) and (1.9), the condition (2.16) on the growth rate of $\sum_1^n (x_i-\bar x_n)^2$ in Theorem 6 fails to hold for the adaptive stochastic approximation schemes of Theorem 1. However, (1.8) and (1.9) imply that the conditions (2.30)–(2.32) of Theorem 7 are satisfied by adaptive stochastic approximation schemes. Therefore, if $\hat\beta_n \to \beta$ a.s. in the stochastic design of Theorem 1, then the random variables x_n defined by the recursion (1.6) with $b_n \sim \hat\beta_n$ satisfy the assumptions of Theorem 7, which in turn implies that $\hat\beta_n \to \beta$ a.s. While this circular argument does not prove that $\hat\beta_n \to \beta$ a.s. in Theorems 2 and 3, we shall establish the desired strong consistency of $\hat{\beta}_n$ in Theorems 2 and 3 by applying the following

Theorem 8. *Let $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. with $E\varepsilon=0$ and $E\varepsilon^2<\infty$. Let $\{x_n\}$ be a sequence of random variables satisfying* (2.1). *Assume that there exist constants $0<\gamma<1/2$ and θ such that with probability* 1

$$x_n-\theta=o(n^{-\gamma}), \tag{2.38}$$

and

$$\sum_1^n (x_i-\theta)^2 \gg \log n. \tag{2.39}$$

Then (2.26) *holds. Let $M(x)$ be a Borel function that is continuously differentiable in some open neighborhood of θ and let $\beta=M'(\theta)$. Define $\hat{\beta}_n$ as in* (2.2). *Then*

$$\left\{\sum_1^n (x_i-\bar{x}_n)\,\varepsilon_i\right\} \Big/ \sum_1^n (x_i-\bar{x}_n)^2 \to 0 \quad \text{and} \quad \hat{\beta}_n \to \beta \quad \text{a.s.}$$

on the event

$$\left[\limsup_{n\to\infty} \sum_1^n (x_i-\theta)^2/(\log n)<\infty\right] \cup \left[\sum_1^n (x_i-\theta)^2/(\log n)\to\infty\right].$$

In fact, there exists an event Ω_0 such that $P(\Omega_0)=1$ and all sample points $\omega\in\Omega_0$ have the following property:

(2.40) *Given $\delta>0$ and $\rho>0$, $\exists$ positive numbers Δ, λ, and N (depending on ω, δ, ρ) such that at ω, for all $n\geqq N$,*

(a) $\sum_1^n (x_i-\theta)^2 \leqq \rho \log n \Rightarrow |\hat{\beta}_n-\beta|<\delta,$

(b) $|\bar{x}_n-\theta| \leqq \lambda n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} \Rightarrow |\hat{\beta}_n-\beta|<\delta,$

(c) $\sum_1^{[n^{2\gamma}]} (x_i-\theta)^2 \geqq \Delta \log n \Rightarrow |\hat{\beta}_n-\beta|<\delta.$

Proof. Without loss of generality, we shall assume that $\theta=0$. Since

$$|\bar{x}_n| \leqq \left(n^{-1}\sum_1^n x_i^2\right)^{\frac{1}{2}},$$

it follows from (2.39) and the law of the iterated logarithm that with probability 1

$$\lim_{n\to\infty} \left(\bar{x}_n \sum_1^n \varepsilon_i\right) \Big/ \left(\sum_1^n x_i^2\right) = 0. \tag{2.41}$$

From (2.38), we obtain that with probability 1

$$\sum_1^{[n^{2\gamma}]} (x_i-\bar{x}_n)^2 = \sum_1^{[n^{2\gamma}]} x_i^2 + [n^{2\gamma}]\,\bar{x}_n^2 - 2\bar{x}_n \sum_1^{[n^{2\gamma}]} x_i = \sum_1^{[n^{2\gamma}]} x_i^2 + o(1). \tag{2.42}$$

From (2.39) and (2.42), (2.26) follows. Moreover, by (2.39) and Lemma 1(ii), with probability 1,

$$\lim_{n\to\infty} \left(\sum_1^n x_i \varepsilon_i\right) \Big/ \left(\sum_1^n x_i^2\right) = 0. \tag{2.43}$$

Let A denote the event on which (2.21), (2.23), (2.38), (2.39), (2.41), (2.42), and (2.43) all hold. Then $P(A)=1$. Let

(2.44) $B=[\lim_{n\to\infty} v_n=0]$, where

$$v_n=\left\{\sum_1^n (x_i-\bar{x}_n)\, g(x_i)\right\} \Big/ \sum_1^n (x_i-\bar{x}_n)^2, \qquad g(x)=M(x)-\beta x.$$

Then, as in the proof of (2.25), $P(B)=1$. Let $\Omega_0=A\cap B$. We now show that all sample points $\omega\in\Omega_0$ have the property (2.40).

Let $u_n=\left\{\sum_1^n (x_i-\bar{x}_n)\,\varepsilon_i\right\} \Big/ \sum_1^n (x_i-\bar{x}_n)^2$. Let $\omega\in A$. Then by (2.39) and (2.42), there exist positive numbers C and N_1 (depending on ω) such that at ω

$$\sum_1^{[n^{2\gamma}]} (x_i-\bar{x}_n)^2 \geqq \tfrac{1}{2} \sum_1^{[n^{2\gamma}]} x_i^2 \geqq C\log n \qquad \text{for all } n\geqq N_1. \tag{2.45}$$

Let $\delta>0$ and $\rho>0$. For $n\geqq N_1$, if $\sum_1^n x_i^2 \leqq \rho \log n$ at ω, then $\sum_1^n (x_i-\bar{x}_n)^2 \geqq (C/\rho)\sum_1^n x_i^2$ at ω by (2.45). Hence, by (2.41) and (2.43), we can choose $N_2(\geqq N_1)$ such that at ω

$$|u_n|\leqq \delta/2 \text{ whenever } \sum_1^n x_i^2 \leqq \rho \log n \quad \text{and} \quad n\geqq N_2. \tag{2.46}$$

In view of (2.45), for $n\geqq N_1$, if $|\bar{x}_n|\leqq \lambda n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$ at ω and $0<\lambda<C^{\frac{1}{2}}$, then at ω

$$\sum_1^n (x_i-\bar{x}_n)^2 = \sum_1^n x_i^2 - n\bar{x}_n^2 \geqq \sum_1^n x_i^2 - \lambda^2 \log n \geqq \tfrac{1}{2}\sum_1^n x_i^2.$$

Hence by (2.41) and (2.43), we can choose $N_3(\geqq N_1)$ such that at ω

$$|u_n|\leqq \delta/2 \text{ whenever } |\bar{x}_n|\leqq \lambda n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} \quad \text{and} \quad n\geqq N_3. \tag{2.47}$$

Choose $\Delta>0$ such that

$$(8E\varepsilon^2/\Delta)^{\frac{1}{2}}<\delta/4. \tag{2.48}$$

Let $d_j=j^{-1}(j-1)(x_j-\bar{x}_{j-1})$, $j\geqq 2$. For $n\geqq N_1$, if $\sum_1^{[n^{2\gamma}]} x_i^2 \geqq \Delta \log n$ at ω, then by (2.20) and (2.45),

$$\sum_2^n d_j^2 \geqq \tfrac{1}{2}\sum_1^n (x_i-\bar{x}_n)^2 \geqq \tfrac{1}{4}\sum_1^{[n^{2\gamma}]} x_i^2 \geqq \tfrac{1}{4}\Delta \log n \qquad \text{at } \omega. \tag{2.49}$$

In view of (2.19), (2.20), (2.21), and (2.23), there exists $N_4(\geqq N_1)$ such that at ω, for all $n \geqq N_4$,

$$|u_n| = \left|\sum_2^n d_j(\varepsilon_j - \bar{\varepsilon}_{j-1})\right| \Big/ \sum_1^n (x_i - \bar{x}_n)^2 \tag{2.50}$$

$$\leqq \delta/4 + \left|\sum_2^n d_j \bar{\varepsilon}_{j-1}\right| \Big/ \left(\sum_2^n d_j^2\right)$$

$$\leqq \delta/4 + \left(\sum_2^n \bar{\varepsilon}_{j-1}^2\right)^{\frac{1}{2}} \Big/ \left(\sum_2^n d_j^2\right)^{\frac{1}{2}},$$

by the Schwarz inequality,

$$\leqq \delta/4 + (2E\varepsilon^2)^{\frac{1}{2}} (\log n)^{\frac{1}{2}} \Big/ \left(\sum_2^n d_j^2\right)^{\frac{1}{2}}.$$

From (2.48), (2.49), and (2.50), it then follows that at ω

$$|u_n| \leqq \delta/2 \text{ whenever } \sum_1^{[n^{2\gamma}]} x_i^2 \geqq \Delta \log n \quad \text{and} \quad n \geqq N_4. \tag{2.51}$$

By (2.44), $\lim_{n\to\infty} v_n = 0$ on B and $\hat{\beta}_n = \beta + u_n + v_n$. Since $\Omega_0 = A \cap B$, we have therefore established that all sample points $\omega \in \Omega_0$ satisfy (2.40). In view of (2.40a),

$$\hat{\beta}_n \to \beta \quad \text{on} \quad \Omega_0 \cap \left[\limsup_{n\to\infty} \left(\sum_1^n x_i^2\right)/(\log n) < \infty\right].$$

Moreover, by (2.40c), $\hat{\beta}_n \to \beta$ on $\Omega_0 \cap \left[\left(\sum_1^n x_i^2\right)/(\log n) \to \infty\right]$. Likewise, in view of (2.46) and (2.51), $u_n \to 0$ on

$$A \cap \left(\left[\limsup_{n\to\infty} \left(\sum_1^n x_i^2\right)/(\log n) < \infty\right] \cup \left[\left(\sum_1^n x_i^2\right)/(\log n) \to \infty\right]\right). \quad \blacksquare$$

3. Convergence Rates of $x_n - \theta$ for Stochastic Approximation Schemes with Random Coefficients

The main result of this section is the following

Theorem 9. *Let $M(x)$ be a Borel function satisfying* (1.2)–(1.4). *Let ε_n, b_n be random variables such that*

$$b_n > 0 \quad \textit{and} \quad \lim_{n\to\infty} n b_n = \infty \quad \textit{a.s.}, \tag{3.1}$$

and

$$\sum_1^\infty \{\varepsilon_n/(n b_n)\} \quad \textit{converges a.s.} \tag{3.2}$$

For $n = 1, 2, \ldots$ define inductively y_n by (1.1) *and x_{n+1} by* (1.6). *Assume that there exists an increasing sequence $\{\psi(n)\}$ of positive constants such that*

(3.3) $$\lim_{n\to\infty} \psi(n)=\infty,$$

(3.4) $$\limsup_{n\to\infty} \{nb_n(\psi(n+1)-\psi(n))/\psi(n)\}<\beta \quad a.s.,$$

and

(3.5) $$\sum_1^\infty \{\psi(n+1)\,\varepsilon_n/(nb_n)\} \quad \textit{converges a.s.}$$

Then $\lim_{n\to\infty} \psi(n)(x_n-\theta)=0$ *a.s.*

The following two corollaries of Theorem 9 will be used in Sects. 5 and 6 for the proof of Theorems 2 and 3.

Corollary 1. *Let* $M(x)$, β, θ; x_1; $\varepsilon_1, \varepsilon_2, \ldots$; $\mathfrak{F}_0, \mathfrak{F}_1, \ldots$ *be as in Theorem* 1. *Let* $\{b_n\}$ *be a sequence of positive random variables, and assume that there exist constants* $B\geqq b>0$ *and random variables* $b'_n>0$ *and* u_n *such that*

(3.6) $$b'_n \textit{ and } u_n \textit{ are both } \mathfrak{F}_{n-1}\textit{-measurable for all } n\geqq 1,$$

(3.7) $$P[|b_n-b'_n|\leqslant|u_n\varepsilon_n|]=1,$$

(3.8) $$P[b\leqq b_n\leqq B \textit{ for all large } n]=1, \quad \textit{and}$$

(3.9) $$\sum_1^\infty u_n^2<\infty \quad a.s.$$

For $n=1,2,\ldots$, *define inductively* y_n *by* (1.1) *and* x_{n+1} *by* (1.6). *Then for all* $0<\gamma<\min\{\beta/B, 1/2\}$,

(3.10) $$\lim_{n\to\infty} n^\gamma(x_n-\theta)=0 \quad a.s.$$

Proof. Take $0<\gamma<\min\{\beta/B, 1/2\}$ and let $\psi(n)=n^\gamma$. Then $\psi(n+1)-\psi(n)\sim\gamma n^{\gamma-1}$, and (3.3) and (3.4) hold. Since $\sum_1^\infty u_n\varepsilon_n$ converges a.s., (3.7) implies that

(3.11) $$b_n-b'_n\to 0 \quad \text{a.s.}$$

To show that (3.2) holds, first note that

(3.12) $$\sum_1^n \{\varepsilon_j/(jb_j)\}=\sum_1^n \{\varepsilon_j/(jb'_j)\}-\sum_1^n \{\varepsilon_j(b_j-b'_j)/(jb_jb'_j)\}$$
$$=U_{1n}-U_{2n}, \quad \text{say.}$$

Since $\sum_1^\infty (jb'_j)^{-2}<\infty$ a.s. and b'_j is $\mathfrak{F}_{j-1}$-measurable, U_{1n} converges a.s. by Lemma 1(i). By (3.7), (3.8), (3.11), and the Schwarz inequality, with probability 1

(3.13) $$\sum_1^n |\varepsilon_j(b_j-b'_j)/(jb_jb'_j)|\ll\sum_1^n j^{-1}|\varepsilon_j|\,|u_j\varepsilon_j|$$
$$\leqq\left(\sum_1^n j^{-2}\varepsilon_j^2\right)^{\frac12}\left(\sum_1^n u_j^2\varepsilon_j^2\right)^{\frac12}.$$

Since u_j is $\mathfrak{F}_{j-1}$-measurable, the series U_{2n} is absolutely convergent a.s. by (3.9), (3.13), and Lemma 1(iii). Therefore in view of (3.12), (3.2) holds. A similar argument also proves (3.5). ∎

Corollary 2. *Let $M(x)$, β, θ; x_1; $\varepsilon_1, \varepsilon_2, \ldots$; $\mathfrak{F}_0, \mathfrak{F}_1, \ldots$ be as in Theorem 1. Let $\{f(n)\}$ be a nondecreasing sequence of positive constants such that*

$$\lim_{n\to\infty} f(n)=\infty \quad \textit{and} \quad \sum_1^\infty \frac{1}{nf(n)}=\infty. \tag{3.14}$$

Let $\{b_n\}$ be a sequence of positive random variables such that there exist random variables $b'_n>0$ and u_n satisfying (3.6), (3.7), (3.9), *and*

$$P[bn^{-\delta}\leqq b_n \leqq f(n) \textit{ and } b'_n \geqq bn^{-\delta} \textit{ for all large } n]=1, \tag{3.15}$$

where $b>0$ and $0\leqq \delta<\frac{1}{4}$ are constants: For $n=1,2,\ldots$, define inductively y_n by (1.1) *and x_{n+1} by* (1.6). *Then for all $0<\gamma<\beta$,*

$$\lim_{n\to\infty} (x_n-\theta)\exp\left\{\gamma\sum_1^n \frac{1}{jf(j)}\right\}=0 \quad \textit{a.s.} \tag{3.16}$$

Proof. Take $0<\gamma<\beta$ and let $\psi(n)=\exp\left\{\gamma\sum_1^n (jf(j))^{-1}\right\}$. Then $(\psi(n+1)-\psi(n))/\psi(n)\sim\gamma/\{(n+1)f(n+1)\}$ and the sequence $\{\psi(n)\}$ is slowly varying (cf. [3]). Therefore in view of (3.14) and (3.15), the conditions (3.1), (3.3), and (3.4) are satisfied. Moreover, by an argument like that in the proof of Corollary 1, it can be shown that (3.2) and (3.5) hold. ∎

We preface the proof of Theorem 9 by the following

Lemma 2 ([12], p. 182). *Let a_n, c_n, and d_n be real numbers satisfying*

$$a_{n+1}=(1-c_n)a_n+d_n, \qquad n=1,2,\ldots, \tag{3.17}$$

where $c_n\geqq 0$ for all large n, $\lim_{n\to\infty} c_n=0$, $\sum_1^\infty c_n=\infty$ and $\sum_1^\infty d_n$ converges. Then $\lim_{n\to\infty} a_n=0$.

Proof of Theorem 9. Without loss of generality, we shall assume that $\theta=0$. By (1.1) and (1.6),

$$x_{n+1}=x_n-\frac{M(x_n)}{nb_n}-\frac{\varepsilon_n}{nb_n}. \tag{3.18}$$

By (3.4), $(nb_n)^{-1}\gg(\psi(n+1)-\psi(n))/\psi(n)$ a.s. Since $\psi(n)\uparrow\infty$, $\sum_1^\infty\{\psi(n+1)-\psi(n)\}/\psi(n)=\infty$ (cf. [7], p. 290), and therefore

$$\sum_1^\infty (nb_n)^{-1}=\infty \quad \text{and} \quad \lim_{n\to\infty}(nb_n)^{-1}=0 \quad \text{a.s.} \tag{3.19}$$

The second relation in (3.19) follows from (3.1). In view of (3.2), (3.18), and (3.19), it follows from Lemma 5 of [10] that $\lim_{n\to\infty} x_n=0$ a.s., and therefore by

(1.2),

$$M(x_n)=(\beta+\xi_n)\,x_n, \quad \text{where} \lim_{n\to\infty}\xi_n=0 \quad \text{a.s.} \tag{3.20}$$

By (3.4), there exists a random variable z such that $0<z<\frac{1}{2}$ and

$$P[(\psi(n+1)-\psi(n))/\psi(n)<\beta(1-2z)/(nb_n) \text{ for all large } n]=1. \tag{3.21}$$

From (3.18) and (3.20), it follows that

$$\psi(n+1)\,x_{n+1}=\left\{\frac{\psi(n+1)}{\psi(n)}\left(1-\frac{\beta+\xi_n}{nb_n}\right)\right\}\psi(n)\,x_n-\frac{\psi(n+1)}{nb_n}\,\varepsilon_n. \tag{3.22}$$

Using (3.21), the monotonicity of $\psi(n)$, and the fact that $\xi_n\to 0$ a.s., we obtain that with probability 1, for all sufficiently large n,

$$1-\frac{\beta+z}{nb_n}\leqq\frac{\psi(n+1)}{\psi(n)}\left(1-\frac{\beta+\xi_n}{nb_n}\right)\leqq 1-\frac{\beta z}{nb_n}. \tag{3.23}$$

Therefore with probability 1,

(3.24) $(\beta+z)/(nb_n)\geqq c_n\geqq\beta z/(nb_n)$ for all sufficiently large n, where

$$1-c_n=\frac{\psi(n+1)}{\psi(n)}\left(1-\frac{\beta+\xi_n}{nb_n}\right).$$

In view of (3.5), (3.19), (3.22), and (3.24), we can apply Lemma 2 with $a_n=\psi(n)\,x_n$, and therefore $\lim_{n\to\infty}\psi(n)\,x_n=0$ a.s. ∎

4. Order of Magnitude of $\sum_1^n (x_i-\theta)^2$ for Stochastic Approximation Schemes with Random Coefficients

In this section we first prove the following theorem and then derive two corollaries which will be applied to prove Theorems 2 and 3.

Theorem 10. *Let $M(x)$, β, θ, σ; x_1; $\varepsilon_1, \varepsilon_2, \ldots$; $\mathfrak{F}_0, \mathfrak{F}_1, \ldots$ be as in Theorem* 1. *Let $\{b_n\}$ be a sequence of positive random variables such that b_n is $\mathfrak{F}_{n-1}$-measurable for all $n\geqq 1$. Assume that there exist two nondecreasing sequences $\{f(n)\}$ and $\{g(n)\}$ of positive constants satisfying* (1.13) *and* (1.14) *such that*

(4.1) $P[(g(n))^{-1}\leqq b_n\leqq f(n)$ *for all large* $n]=1$, *and*

(4.2) $|\log g(n)|^{\lambda}=o\left(\sum_1^n \frac{1}{jg(j)f^2(j)}\right)$ *for some* $\lambda>1$.

For $n=1,2,\ldots$, define inductively y_n by (1.1) *and x_{n+1} by* (1.6). *Then with probability* 1

(4.3) $$\sum_1^n (x_i-\theta)^2 \gg \sum_1^n \frac{1}{jg(j)f^2(j)}.$$

We preface the proof of Theorem 10 by the following two lemmas.

Lemma 3. *Let $z, z_1, z_2, \ldots$ be i.i.d. random variables such that $E|z|<\infty$. Let $\{a_n\}$ be an ultimately nonincreasing sequence of constants such that $\sum_1^\infty a_n=\infty$. Then*

$$\left(\sum_1^n a_j z_j\right) \Big/ \left(\sum_1^n a_j\right) \to Ez \quad a.s.$$

Proof. Note that $\sum_1^n a_j z_j = a_n S_n + \sum_1^{n-1} (a_j-a_{j+1}) S_j$, where $S_j=\sum_1^j z_i$, and apply the strong law of large numbers. ∎

Lemma 4. *Let $\{\varphi(n)\}$ be a nondecreasing and slowly varying sequence of positive constants. Let*

(4.4) $$\gamma_n=\prod_{j=J}^n \left(1-\frac{\varphi(j)}{j}\right), \qquad n \geqq J,$$

where J is so chosen that $\varphi(j)/j<1$ for all $j\geqq J$.

(i) *If $\varphi(n)=c>1$ for all $n\geqq J$, then $\gamma_n \sim b n^{-c}$ for some $b>0$, and therefore*

$$\sum_{k=n}^\infty \gamma_k \sim \frac{b}{(c-1)\,n^{c-1}} \sim n\gamma_n/\{\varphi(n)-1\}.$$

(ii) *If $\lim_{n\to\infty} \varphi(n)>1$, then $\gamma_n \ll n^{-a}$ for some $a>1$, and $\{n\gamma_n\}$ is ultimately decreasing.*

(iii) *If $\lim_{n\to\infty} \varphi(n)=\infty$, then*

(4.5) $$\sum_{k=n}^\infty \gamma_k \sim n\gamma_n/\varphi(n),$$

and for every $\delta>0$, there exist $\xi>0$ and k_0 such that

(4.6) $$(k\gamma_k)/(n\gamma_n)\leqq\delta \quad \text{if } k\geqq k_0 \text{ and } J\leqq n\leqq k-[\xi k/\varphi(k)].$$

Proof. To prove that $\{n\gamma_n\}$ is ultimately decreasing in (ii), we note that

$$\frac{(n+1)\,\gamma_{n+1}}{n\gamma_n} = \left(1+\frac{1}{n}\right)\left(1-\frac{\varphi(n+1)}{n+1}\right)<1 \qquad \text{for all large } n,$$

since $\lim_{n\to\infty} \varphi(n)>1$. The other parts of (i) and (ii) are obvious. To prove (iii), we note that for $k>n\geqq J$,

(4.7) $$\begin{aligned}\gamma_k/\gamma_n &= \exp\left\{\sum_{j=n+1}^k \log\left(1-\frac{\varphi(j)}{j}\right)\right\} \leqq \exp\left\{-\sum_{j=n+1}^k \varphi(j)/j\right\} \\ &\leqq \exp\left\{-\varphi(n+1)\int_{n+1}^{k+1} x^{-1}\,dx\right\} = \left(\frac{n+1}{k+1}\right)^{\varphi(n+1)}.\end{aligned}$$

Therefore

$$(4.8)\qquad \sum_{k=n}^{\infty} \gamma_k/\gamma_n \leqq (n+1)^{\varphi(n+1)} \int_n^{\infty} x^{-\varphi(n+1)}\,dx$$
$$= \frac{n}{\varphi(n+1)-1}\left(\frac{n+1}{n}\right)^{\varphi(n+1)} \sim n/\varphi(n).$$

Moreover, from (ii) and (4.7), it follows that there exists n_0 such that for $n_0 \leqq n \leqq k - [\xi k/\varphi(k)]$,

$$\frac{k\gamma_k}{n\gamma_n} \leqq \frac{k\gamma_k}{n(k)\,\gamma_{n(k)}}, \qquad \text{where } n(k) = k - [\xi k/\varphi(k)],$$
$$\leqq 2\left\{\frac{n(k)+1}{k+1}\right\}^{\varphi(n(k)+1)-1}$$
$$= 2\left\{1 - \frac{[\xi k/\varphi(k)]}{k+1}\right\}^{\varphi(n(k)+1)-1}$$
$$\to 2e^{-\xi} \text{ (as } k\to\infty) \text{ for every fixed } \xi > 0.$$

The last relation above follows from the fact that φ is nondecreasing and slowly varying. Choosing $\xi > \log(2/\delta)$ and noting that $\lim_{k\to\infty} k\gamma_k = 0$, (4.6) follows.

To complete the proof of (iii), given $\varepsilon > 0$, we have for all large n

$$(4.9)\qquad \sum_{k=n}^{\infty} \gamma_k/\gamma_n \geqq \sum_{k=n}^{2n} \exp\left\{-(1+\varepsilon)\,\varphi(n)\int_n^k x^{-1}\,dx\right\}$$
$$= \sum_{k=n}^{2n} (n/k)^{(1+\varepsilon)\varphi(n)} \sim n/\{(1+\varepsilon)\,\varphi(n)\}.$$

From (4.8) and (4.9), (4.5) follows. ∎

Proof of Theorem 10. Without loss of generality we shall assume that $\theta = 0$. Since (4.1) holds and $\sum_1^{\infty} (g(n)/n)^2 < \infty$ by (1.14), it follows from Lemma 1(i) that $\sum_1^{\infty} \varepsilon_n/(nb_n)$ converges a.s. Hence by Lemma 5 of [10], $\lim_{n\to\infty} x_n = 0$ a.s., and therefore (3.20) holds.

From (3.18) and (3.20), it follows that with probability 1

$$(4.10)\qquad x_{n+1}^2 = \left(1 - \frac{\beta+\xi_n}{nb_n}\right)^2 x_n^2 + \left(\frac{\varepsilon_n}{nb_n}\right)^2 - 2\left(1 - \frac{\beta+\xi_n}{nb_n}\right)\frac{x_n\varepsilon_n}{nb_n}.$$

Define

$$(4.11)\qquad \varphi(n) = 3\beta\, g(n) \quad \text{if } \lim_{n\to\infty} g(n) = \infty,$$
$$= c\,(>2) \quad \text{if } \lim_{n\to\infty} g(n) < \infty,$$
$$\text{where } c = 2 + 3\beta \lim_{n\to\infty} g(n).$$

By (4.1) and (4.11),

$$b_n \gg (\varphi(n))^{-1}. \tag{4.12}$$

Let $v_n = 1-(\beta+\xi_n)/(nb_n)$, and note that

$$\lim_{n\to\infty} v_n = 1 \text{ a.s. and } v_n \text{ is } \mathfrak{F}_{n-1}\text{-measurable for } n \geqq 1. \tag{4.13}$$

From (4.1), (4.10), and the fact that $\xi_n \to 0$ a.s., we obtain that with probability 1

$$x_{n+1}^2 \geqq \left(1-\frac{\varphi(n)}{n}\right) x_n^2 + \left(\frac{\varepsilon_n}{nf(n)}\right)^2 - \left(\frac{2v_n x_n}{nb_n}\right)\varepsilon_n \tag{4.14}$$

for all large n, say $n \geqq m(=m(\omega))$.

With $\varphi(n)$ defined by (4.11), define γ_n and J as in (4.4). Choosing $m>J$ in (4.14), we obtain by iterating (4.14) that with probability 1

$$\begin{aligned} x_{k+1}^2 \geqq \frac{\gamma_k}{\gamma_{m-1}} x_m^2 &+ \sum_{j=m}^{k} \left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{\varepsilon_j}{jf(j)}\right)^2 \\ &-2\sum_{j=m}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{v_j x_j}{jb_j}\right)\varepsilon_j \quad \text{for } k \geqq m. \end{aligned} \tag{4.15}$$

From (4.15), Lemma 4(ii), and the fact that $x_n \to 0$ a.s., it follows that with probability 1

$$\begin{aligned} \sum_{k=J}^{n} x_k^2 \geqq \sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{\varepsilon_j}{jf(j)}\right)^2 \\ -2\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{v_j x_j}{jb_j}\right)\varepsilon_j + O(1). \end{aligned} \tag{4.16}$$

Let $\Gamma_j = \sum_{k=j}^{\infty} \gamma_k$. By Lemma 4,

$$\Gamma_j \sim j\gamma_j/\{\varphi(j)-1\} \sim \Gamma_{j+1}. \tag{4.17}$$

Interchanging the order of summation gives

$$\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{v_j x_j}{jb_j}\right)\varepsilon_j = \sum_{j=J}^{n}\left(\frac{\Gamma_j-\Gamma_{n+1}}{\gamma_j}\right)\left(\frac{v_j x_j}{jb_j}\right)\varepsilon_j. \tag{4.18}$$

Applying Lemma 1 to the martingale transform $\left\{\sum_{j=J}^{n}(\Gamma_j v_j x_j \varepsilon_j)/(j\gamma_j b_j),\ \mathfrak{F}_n,\ n \geqq J\right\}$, we obtain that with probability 1

$$(4.19)\qquad \sum_{j=J}^{n}(\Gamma_j v_j x_j \varepsilon_j)/(j\gamma_j b_j)$$
$$= o\left(\left\{\sum_{j=J}^{n}(\Gamma_j v_j x_j)^2/(j\gamma_j b_j)^2\right\}\right) + O(1)$$
$$= o\left(\sum_{j=J}^{n} x_j^2\right) + O(1), \qquad \text{by (4.12), (4.13), and (4.17),}$$

where the $O(1)$ term indicates that $\sum_{j=J}^{n}(\Gamma_j v_j x_j \varepsilon_j)/(j\gamma_j b_j)$ converges a.s. on the event $\left[\sum_{j=J}^{\infty}(\Gamma_j v_j x_j)^2/(j\gamma_j b_j)^2 < \infty\right]$.

We now show that with probability 1

$$(4.20)\qquad \Gamma_{n+1}\left\{\sum_{j=J}^{n}\frac{v_j x_j \varepsilon_j}{j b_j \gamma_j}\right\} = o\left(\left\{\sum_{J}^{n} x_j^2\right\}^{\frac{1}{2}}\left\{\sum_{1}^{n}\frac{1}{j g(j) f^2(j)}\right\}^{\frac{1}{2}}\right) + o\left(\sum_{J}^{n} x_j^2\right).$$

Let $V_j = 1 + \sum_{k=J}^{j}(v_k x_k/b_k)^2$. By Lemma 1(i), for every $\frac{1}{2} < \eta < 1$,

$$(4.21)\qquad \sum_{j=J}^{n}\left\{\left(\frac{v_j x_j \varepsilon_j}{j b_j \gamma_j}\right)\Big/\left(\frac{V_j^{\frac{1}{2}}(\log V_j)^{\eta}}{j\gamma_j}\right)\right\} \quad \text{converges a.s.}$$

Since $(j\gamma_j)^{-1}\uparrow\infty$ by Lemma 4, we obtain by (4.21) and the Kronecker lemma that with probability 1

$$(4.22)\qquad \sum_{j=J}^{n}(v_j x_j \varepsilon_j)/(j b_j \gamma_j) = o(V_n^{\frac{1}{2}}(\log V_n)^{\eta}/(n\gamma_n)).$$

From (4.12) and (4.13) it follows that

$$(4.23)\qquad V_n \ll \varphi^2(n)\left(1 + \sum_{j=J}^{n} x_j^2\right) \quad \text{a.s.}$$

By choosing $\eta(>\frac{1}{2})$ sufficiently close to $\frac{1}{2}$, we obtain from (4.2), (4.11), and (4.23) that with probability 1

$$(4.24)\qquad (\log V_n)^{\eta} \ll \left\{\log\left(1 + \sum_{J}^{n} x_j^2\right)\right\}^{\eta} + \{\log\varphi(n)\}^{\eta}$$
$$= \left\{\log\left(1 + \sum_{J}^{n} x_j^2\right)\right\}^{\eta} + o\left(\left\{\sum_{1}^{n}\frac{1}{j g(j) f^2(j)}\right\}^{\frac{1}{2}}\right).$$

From (4.17), (4.22), (4.23), and (4.24), (4.20) follows.

We note that

$$(4.25)\qquad \sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{\varepsilon_j}{j f(j)}\right)^2 = \sum_{j=J}^{n}\left(\frac{\Gamma_j - \Gamma_{n+1}}{\gamma_j}\right)\left(\frac{\varepsilon_j}{j f(j)}\right)^2.$$

By Lemma 4 and (4.11), with probability 1,

$$\text{(4.26)} \qquad \sum_{j=J}^{n} \left(\frac{\Gamma_j}{\gamma_j}\right)\left(\frac{\varepsilon_j}{jf(j)}\right)^2 \gg \sum_{j=J}^{n} \frac{\varepsilon_j^2}{jf^2(j)\,\varphi(j)} \gg \sum_{j=J}^{n} \frac{\varepsilon_j^2}{jf^2(j)\,g(j)}$$
$$\sim \sum_{j=J}^{n} \frac{\sigma^2}{jf^2(j)\,g(j)}, \quad \text{by Lemma 3.}$$

We now show that with probability 1

$$\text{(4.27)} \qquad \Gamma_{n+1}\left\{\sum_{j=J}^{n} \frac{\varepsilon_j^2}{j^2 f^2(j)\,\gamma_j}\right\} = o\left(\sum_{1}^{n} \frac{1}{jf^2(j)\,g(j)}\right).$$

To prove (4.27), first assume that $\lim_{n\to\infty} g(n) < \infty$. Then by (4.11), $\varphi(n) \equiv c > 2$, and a straightforward application of Lemma 4(i) and the strong law shows that the left hand side of (4.27) is $O(1)$ a.s. Now assume that $\lim_{n\to\infty} g(n) = \infty$. We first note that by Lemma 3

$$\text{(4.28)} \qquad \sum_{1}^{n} \frac{\varepsilon_j^2}{jf^2(j)} = \sigma^2(1+o(1)) \sum_{1}^{n} \frac{1}{jf^2(j)} \quad \text{a.s.}$$

By Lemma 4(iii), given $0 < \delta < 1$, we can choose $\xi > 0$ and k_0 such that (4.6) holds. By (4.6), (4.17), and (4.28), with probability 1,

$$\text{(4.29)} \qquad \Gamma_{n+1} \sum_{j=J}^{n-[\xi n/\varphi(n)]} \frac{\varepsilon_j^2}{j^2 f^2(j)\,\gamma_j}$$
$$\leqq \frac{\delta\sigma^2(1+o(1))}{\varphi(n)} \sum_{j=J}^{n-[\xi n/\varphi(n)]} \frac{1}{jf^2(j)}$$
$$\leqq \frac{\delta\sigma^2(1+o(1))}{3\beta} \sum_{j=J}^{n} \frac{1}{jf^2(j)\,g(j)}, \quad \text{by (4.11).}$$

By (4.17) and (4.28), with probability 1,

$$\text{(4.30)} \qquad \Gamma_{n+1} \sum_{j=n-[\xi n/\varphi(n)]}^{n} \frac{\varepsilon_j^2}{j^2 f^2(j)\,\gamma_j}$$
$$\leqq \frac{1+o(1)}{\varphi(n)} \left\{ \sum_{j=n-[\xi n/\varphi(n)]}^{n} \frac{\sigma^2}{jf^2(j)} + o\left(\sum_{1}^{n} \frac{1}{jf^2(j)}\right)\right\},$$
$$\ll \frac{g^2(n)}{\varphi(n)} \sum_{j=n-[\xi n/\varphi(n)]}^{n} j^{-1} + o\left(\frac{1}{\varphi(n)} \sum_{1}^{n} \frac{1}{jf^2(j)}\right),$$

since $f(n)\,g(n) \geqq 1$ for all large n by (4.1). Choosing δ arbitrarily small, we obtain (4.27) from (4.29) and (4.30).

From (4.16), (4.18), (4.19), (4.20), (4.25), (4.26), and (4.27), it follows that with probability 1

$$\sum_{j=J}^{n} x_j^2 \gg \sum_{1}^{n} \frac{1}{jf^2(j)\,g(j)} + o\left(\sum_{j=J}^{n} x_j^2\right)$$
$$+ o\left(\left\{1 + \sum_{j=J}^{n} x_j^2\right\}^{\frac{1}{2}} \left\{\sum_{1}^{n} \frac{1}{jf^2(j)\,g(j)}\right\}^{\frac{1}{2}}\right).$$

Hence (4.3) follows. ∎

Since b_n in Theorems 2 and 3 is $\mathfrak{F}_n$-measurable but not $\mathfrak{F}_{n-1}$-measurable, Theorem 10 is not directly applicable. However, we can modify the preceding proof to handle this case as well in the following

Corollary 3. *Suppose that in Theorem* 10 *we replace the assumption that* b_n *is* $\mathfrak{F}_{n-1}$*-measurable for all n by the condition*

(4.31) b_n *is* $\mathfrak{F}_n$*-measurable for all n and there exist random variables* $b'_n>0$ *and* u_n *such that* (3.6), (3.7), (3.9) *hold and* $b'_n \gg (g(n))^{-1}$ *a.s.*

Moreover, suppose that we replace the assumption (4.2) *by the stronger condition* (1.15). *Then* (4.3) *still holds.*

Proof. An argument as in the proof of Corollary 1 shows that $\sum_1^\infty \varepsilon_n/(nb_n)$ converges a.s., and therefore $\lim_{n\to\infty} x_n=0$ a.s. and (3.20) still holds. Since b_n is $\mathfrak{F}_n$-measurable, x_n and therefore ξ_n (as defined in (3.20)) are $\mathfrak{F}_{n-1}$-measurable. As in the proof of Theorem 10, let $v_n=1-(\beta+\xi_n)/(nb_n)$. By (1.14), (3.7), (4.1), and (4.31), letting $v'_n=1-(\beta+\xi_n)/(nb'_n)$, we obtain that with probability 1

$$|v_n-v'_n| \ll g^2(n)\,|b_n-b'_n|/n=o(|u_n\varepsilon_n|). \tag{4.32}$$

Moreover, $\lim_{n\to\infty} v_n=1=\lim_{n\to\infty} v'_n$ a.s. and v'_n is $\mathfrak{F}_{n-1}$-measurable for all $n\geqq 1$. Define $\varphi(n)$ by (4.11) and γ_n and J by (4.4).

From (3.7), (4.1), (4.31), and (4.32), it follows that with probability 1

$$\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{v_j x_j}{jb_j}\right)\varepsilon_j=\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{v'_j x_j}{jb'_j}\right)\varepsilon_j +O\left(\left\{g^2(n)\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{j\gamma_j}\right)|u_j x_j|\,\varepsilon_j^2\right\}\right). \tag{4.33}$$

Since x_j, v'_j, and b'_j are all $\mathfrak{F}_{j-1}$-measurable, the same argument as that in the proof of Theorem 10 shows that with probability 1

$$\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{\gamma_j}\right)\left(\frac{v'_j x_j}{jb'_j}\right)\varepsilon_j =o\left(\sum_{j=J}^{n}x_j^2\right)+o\left(\left\{\sum_{J}^{n}x_j^2\right\}^{\frac{1}{2}}\left\{\sum_{1}^{n}\frac{1}{jg(j)f^2(j)}\right\}^{\frac{1}{2}}\right). \tag{4.34}$$

Interchanging the order of summation and using (4.17), we obtain that with probability 1

$$g^2(n)\sum_{k=J}^{n}\sum_{j=J}^{k}\left(\frac{\gamma_k}{j\gamma_j}\right)|u_j x_j|\,\varepsilon_j^2 \leqq g^2(n)\sum_{j=J}^{n}\left(\frac{\Gamma_j}{j\gamma_j}\right)|u_j x_j|\,\varepsilon_j^2 \tag{4.35}$$

$$\begin{aligned}
&\ll g^2(n) \sum_{j=J}^{n} |u_j \varepsilon_j|\,|x_j \varepsilon_j| \\
&\leqq g^2(n) \left(\sum_{j=J}^{n} u_j^2 \varepsilon_j^2\right)^{\frac{1}{2}} \left(\sum_{j=J}^{n} x_j^2 \varepsilon_j^2\right)^{\frac{1}{2}} \\
&\ll g^2(n) \left(\sum_{j=J}^{n} x_j^2 \varepsilon_j^2\right)^{\frac{1}{2}}, \qquad \text{by (3.9) and Lemma 1(iii),} \\
&= o\left(\left\{\sum_{J}^{n} x_j^2\right\}^{1-2\eta} \left\{\sum_{1}^{n} \frac{1}{j g(j) f^2(j)}\right\}^{2\eta}\right).
\end{aligned}$$

The last relation above follows from (1.15) and Lemma 1(iii) and (iv), noting that $1-2\eta>1/2$. The rest of the proof is similar to that of Theorem 10. ∎

Letting $g(n)=b^{-1}$ and $f(n)=B$ in Corollary 3, we obtain the following corollary which will be used in the proof of Theorem 2.

Corollary 4. *With the same notations and assumptions as in Corollary* 1, *assume that* b_n *is* $\mathfrak{F}_n$*-measurable for all* $n \geqq 1$. *Then with probability* 1

$$\sum_{1}^{n} (x_i-\theta)^2 \gg \log n. \tag{4.36}$$

5. Proof of Theorem 2

To prove Theorem 2, we apply Theorem 8, Corollaries 1 and 4, and the following

Lemma 5 ([9], pp. 3066–3067). *With the same notations and assumptions as in Corollary* 1, *there exists an event* Ω_1 *such that* $P(\Omega_1)=1$ *and all sample points* $\omega \in \Omega_1$ *have the following property:*

(5.1) $\exists\, C>0$ *and positive integers* M, k *(depending on* ω*) such that at* ω, *for all* $l \geqq M$ *and* $m \geqq l^k$,

$$3\beta/2 \geqq b_n (\geqq b)\ \forall\, l \leqq n \leqq m \Rightarrow |\bar{x}_n| \leqq C n^{-\frac{1}{2}} (\log\log n)^{\frac{1}{2}}\ \forall\, l^k \leqq n \leqq m + m^{\frac{1}{2}}.$$

Proof of Theorem 2. We first prove that $\hat{\beta}_n \to \beta$ a.s. Note that b_n is $\mathfrak{F}_n$-measurable and x_n is $\mathfrak{F}_{n-1}$-measurable, where $\mathfrak{F}_k$ denotes the σ-field generated by $x_1, \varepsilon_1, \ldots, \varepsilon_k$. Define

$$\begin{aligned}
b'_n &= b \vee \left\{ \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)\, y_i + (x_n - \bar{x}_n)\, M(x_n)}{\sum_{1}^{n} (x_i - \bar{x}_n)^2} \wedge B \right\}, \qquad n \geqq n_0, \\
&= c, \qquad n < n_0,
\end{aligned} \tag{5.2}$$

Then b'_n is $\mathfrak{F}_{n-1}$-measurable, $b'_n = b_n$ if $n < n_0$, and for $n \geqq n_0$

(5.3) $$|b_n - b'_n| \leqq (|x_n - \bar{x}_n|\,|\varepsilon_n|) \Big/ \sum_1^n (x_i - \bar{x}_n)^2$$
$$\leqq 2(|x_n - \bar{x}_{n-1}|\,|\varepsilon_n|) \Big/ \sum_1^n (x_i - \bar{x}_{i-1})^2.$$

The last relation above follows from (2.20) and the fact that $x_n - \bar{x}_n = (1 - n^{-1}) \cdot (x_n - \bar{x}_{n-1})$. Letting

(5.4) $$u_n = |x_n - \bar{x}_{n-1}| \Big/ \sum_1^n (x_i - \bar{x}_{i-1})^2, \quad n \geqq n_0,$$
$$= 0, \quad n < n_0,$$

we note that $\sum_1^\infty u_k^2 < \infty$. Therefore the assumptions of Corollaries 1 and 4 are satisfied. By Corollaries 1 and 4, the assumptions (2.38) and (2.39) of Theorem 8 are satisfied by $\{x_n\}$ (with $0 < \gamma < \min\{\beta/B, 1/2\}$). Without loss of generality we shall assume that $\theta = 0$. By Theorem 8, $\sum_1^n (x_i - \bar{x}_n)^2 \gg \log n$ a.s. and

(5.5) $$\hat{\beta}_n \to \beta \quad \text{a.s. on } \left[\limsup_{n\to\infty} \left(\sum_1^n x_i^2\right)/(\log n) < \infty\right].$$

It therefore remains to prove that

(5.6) $$\hat{\beta}_n \to \beta \quad \text{a.s. on } \left[\limsup_{n\to\infty} \left(\sum_1^n x_i^2\right)/(\log n) = \infty\right].$$

By Theorem 8, there exists an event Ω_0 such that $P(\Omega_0) = 1$ and all sample points $\omega \in \Omega_0$ have the property (2.40). Moreover, by Lemma 5, there exists an event Ω_1 such that $P(\Omega_1) = 1$ and all sample points $\omega \in \Omega_1$ have the property (5.1). Let $D = \left[\limsup_{n\to\infty} \left(\sum_1^n x_i^2\right) \Big/ (\log n) = \infty\right]$. To prove (5.6), it suffices to show that

(5.7) $$\hat{\beta}_n \to \beta \quad \text{on } D \cap \Omega_0 \cap \Omega_1.$$

Let $\delta > 0$ such that $b < \beta - \delta$ and $\beta + \delta < \min\{3\beta/2, B\}$. Therefore, by the definition of b_n in (1.11), for $n \geqq n_0$,

(5.8) $$|\hat{\beta}_n - \beta| < \delta \Rightarrow b_n = \hat{\beta}_n \quad \text{and} \quad b_n \leqq 3\beta/2.$$

Let $\omega \in D \cap \Omega_0 \cap \Omega_1$. In view of (2.40), (5.1), and (5.8), there exist $C > 0$, $\Delta > 0$, and positive integers $\tilde{M}$, $k (\geqq 2)$ such that at ω, for all $n \geqq \tilde{M}$, $l \geqq \tilde{M}$, and $m \geqq l^k$,

(5.9) $$|\hat{\beta}_n - \beta| < \delta \;\forall\, l \leqq n \leqq m \Rightarrow |\bar{x}_n| \leqq C n^{-\frac{1}{2}} (\log\log n)^{\frac{1}{2}} \;\forall\, l^k \leqq n \leqq m + m^{\frac{1}{2}},$$

(5.10) $$|\bar{x}_n| \leqq C n^{-\frac{1}{2}} (\log\log n)^{\frac{1}{2}} \Rightarrow |\hat{\beta}_n - \beta| < \delta,$$

(5.11) $$\sum_1^{[n^{2\gamma}]} x_i^2 \geqq \Delta \log n \Rightarrow |\hat{\beta}_n - \beta| < \delta.$$

Since $\omega \in D$, we can choose an integer $L > \tilde{M}$ such that at ω

$$\sum_{1}^{[L^{2\gamma}]} x_i^2 > k\Delta \log L,$$

and therefore at ω, for $L \leqq n \leqq L^k$,

$$\sum_{1}^{[n^{2\gamma}]} x_i^2 > k\Delta \log L \geqq \Delta \log n. \tag{5.12}$$

Hence by (5.11) and (5.12), $|\hat{\beta}_n - \beta| < \delta$ at ω for all $L \leqq n \leqq L^k$. In view of (5.9), this in turn implies that for $\nu = L^k + 1$,

$$|\bar{x}_\nu| \leqq C\nu^{-\frac{1}{2}}(\log\log \nu)^{\frac{1}{2}} \quad \text{at } \omega,$$

and therefore $|\hat{\beta}_\nu - \beta| < \delta$ at ω by (5.10). Proceeding inductively in this way, we then obtain that at ω, for $j = 1, 2, \ldots$,

$$\begin{aligned} &|\hat{\beta}_n - \beta| < \delta \quad \text{for } L \leqq n \leqq L^k + j \\ &\quad \Rightarrow |\bar{x}_n| \leqq Cn^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}} \quad \text{for } n = L^k + j + 1, \text{ by (5.9)}, \\ &\quad \Rightarrow |\hat{\beta}_n - \beta| < \delta \quad \text{for } n = L^k + j + 1, \text{ by (5.10)}. \end{aligned}$$

Thus, the above inductive argument shows that $|\hat{\beta}_n - \beta| < \delta$ at ω for all $n \geqq L$. Since δ is arbitrary, we have established (5.7). Hence $\hat{\beta}_n \to \beta$ a.s. and $b_n \to \beta$ a.s.

Since $b_n \to \beta$ a.s., we obtain by Theorem 4 of [9] that

$$|x_n| \ll n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}} \quad \text{a.s.} \tag{5.13}$$

Since $\sum_{1}^{n} (x_i - \bar{x}_n)^2 \gg \log n$ a.s., (5.4) and (5.13) imply that

$$|u_n| \ll (\log\log n)^{\frac{1}{2}}/(n^{\frac{1}{2}} \log n) \quad \text{a.s.} \tag{5.14}$$

By (5.3) and (5.14), $b_n - b'_n = o((n \log\log n)^{-\frac{1}{2}} |\varepsilon_n|)$ a.s. Hence, by Theorem 4(iv) of [10], (1.7)–(1.9) still hold. ∎

6. Removal of the Assumption of Known Prior Bounds on β and Proof of Theorem 3

The proof of Theorem 3 makes use of Corollaries 2 and 3 to show that the following two properties hold for the stochastic approximation scheme $\{x_i\}$ of Theorem 3: As $m \to \infty$,

$$x_m - \theta = o\left(\exp\left\{-\tfrac{1}{2}\beta \sum_{1}^{m} \frac{1}{jf(j)}\right\}\right) \quad \text{a.s.,} \tag{6.1}$$

and

$$\sum_{1}^{m} (x_i - \theta)^2 \to \infty \quad \text{a.s.} \tag{6.2}$$

Since x_i is $\mathfrak{F}_{i-1}$-measurable, (6.2) and Lemma 1(ii) imply that

$$\left\{\sum_1^m (x_i-\theta)\,\varepsilon_i\right\} \Big/ \left\{\sum_1^m (x_i-\theta)^2\right\} \to 0 \quad \text{a.s.} \tag{6.3}$$

Moreover, in view of (1.2) and (6.1),

$$\left\{\sum_1^m (x_i-\theta)\,M(x_i)\right\} \Big/ \left\{\sum_1^m (x_i-\theta)^2\right\} \to \beta \quad \text{a.s.} \tag{6.4}$$

From (1.1), (6.3), and (6.4), it follows that

$$\left\{\sum_1^m (x_i-\theta)\,y_i\right\} \Big/ \left\{\sum_1^m (x_i-\theta)^2\right\} \to \beta \quad \text{a.s.} \tag{6.5}$$

Now in ignorance of θ, if we replace θ in (6.5) by some random variable which is very close to θ relative to $\{x_1, \ldots, x_m\}$, it is plausible that (6.5) may still hold. This suggests that at stage n we replace θ in (6.5) by x_n and set $m=m(n)$, where $\{m(n)\}$ is a sequence of positive integers such that $m(n)\to\infty$ and $m(n)=o(n)$. This is the rationale behind the "preliminary estimator" β_n^* defined in (1.19). Moreover, (6.1) suggests that if $m(n)$ is chosen so that (1.16) holds, then $\beta_n^*\to\beta$ a.s., as is shown in the following

Lemma 6. *Let $M(x)$, β, θ; $\varepsilon_1, \varepsilon_2, \ldots$; $\mathfrak{F}_0, \mathfrak{F}_1, \ldots$ be the same as in Theorem* 1. *Let $\{f(n)\}$ be a nondecreasing sequence of positive constants satisfying* (3.14), *and let $\{m(n)\}$ be a sequence of positive integers satisfying* (1.16). *Let $\{x_n\}$ be a sequence of random variables such that* (6.1) *and* (6.2) *hold and x_n is $\mathfrak{F}_{n-1}$-measurable for all $n\geqq 1$. Define y_n by* (1.1) *and β_n^* by* (1.19). *Then*

$$\sum_1^{m(n)} (x_i-x_n)^2 \to \infty \quad \textit{and} \quad \beta_n^* \to \beta \quad \textit{a.s.}$$

Proof. By (1.16) and (6.1),

$$m(n)(x_n-\theta)\to 0 \quad \text{a.s.,} \tag{6.6}$$

and therefore

$$\sum_1^{m(n)} (x_i-x_n)^2 = \sum_1^{m(n)} (x_i-\theta)^2 + o(1) \quad \text{a.s.} \tag{6.7}$$

Since $M(x_m)\to 0$ a.s. and $m^{-1}\sum_1^m |\varepsilon_i| \to E|\varepsilon_1|$ a.s., (6.6) implies that

$$|x_n-\theta| \sum_1^{m(n)} |y_i| \leqq |x_n-\theta| \sum_1^{m(n)} (|M(x_i)|+|\varepsilon_i|) \to 0 \quad \text{a.s.} \tag{6.8}$$

By (6.2), (6.5), (6.7), and (6.8), it follows that

$$\beta_n^* = \frac{\sum_1^{m(n)} (x_i-\theta)\,y_i - (x_n-\theta)\sum_1^{m(n)} y_i}{\sum_1^{m(n)} (x_i-\theta)^2 + o(1)} \to \beta \quad \text{a.s.} \quad \blacksquare \tag{6.9}$$

Proof of Theorem 3. We first show that b_n as defined in (1.18) satisfies the assumptions of Corollaries 2 and 3. In view of (1.17), β_n^* is $\mathfrak{F}_{n-1}$-measurable. Moreover, x_n is $\mathfrak{F}_{n-1}$-measurable and b_n is $\mathfrak{F}_n$-measurable. For every $n \geqq 1$, if $\sum_1^{m(n)} (x_i - \bar{x}_{m(n)})^2 = 0$, set $b_n' = c$ and $u_n = 0$, otherwise let

$$(6.10) \quad b_n' = (g(n))^{-1} \vee d\beta_n^* \vee \left\{ \frac{\sum_1^{n-1} (x_i - \bar{x}_n)\, y_i + (x_n - \bar{x}_n)\, M(x_n)}{\sum_1^{n} (x_i - \bar{x}_n)^2} \wedge D\beta_n^* \wedge f(n) \right\},$$

$$u_n = |x_n - \bar{x}_{n-1}| \Big/ \sum_1^n (x_i - \bar{x}_{i-1})^2.$$

Then b_n' and u_n are $\mathfrak{F}_{n-1}$-measurable, $\sum_1^\infty u_n^2 < \infty$, and by the same argument as in (5.3), $|b_n - b_n'| \leqq 2u_n |\varepsilon_n|$. Clearly $\{f(n)\}$ satisfies (3.14). Therefore the assumptions of Corollaries 2 and 3 are satisfied. By Corollaries 2 and 3, (6.1) and (6.2) hold. Hence by Lemma 6, $\beta_n^* \to \beta$ a.s.

Since $\beta_n^* \to \beta$ a.s., (1.18) and (6.10) imply that with probability 1

$$(6.11) \quad \begin{aligned} (c \wedge \tfrac{1}{2} d\beta) &\leqq b_n \leqq (c \vee 2D\beta), \\ (c \wedge \tfrac{1}{2} d\beta) &\leqq b_n' \leqq (c \vee 2D\beta) \end{aligned}$$

for all large n, so the assumptions of Corollaries 1 and 4 are satisfied. We can then apply Theorem 8 and repeat the argument used in the proof of Theorem 2 to show that $\hat{\beta}_n \to \beta$ a.s. and that (1.7)–(1.9) still hold. ∎

By making use of Lemma 6 together with Theorems 1 and 5, we also obtain the following

Corollary 5. *Suppose that in Theorem 3 we replace b_n in (1.18) by*

$$(6.12) \quad \begin{aligned} b_n &= (g(n))^{-1} \vee (\beta_n^* \wedge f(n)) \quad \textit{if } \sum_1^{m(n)} (x_i - x_n)^2 > 0, \\ &= c \quad \textit{otherwise}. \end{aligned}$$

Then $\beta_n^* \to \beta$ a.s. and (1.7)–(1.9) hold. If $M(x)$ also satisfies (1.21), then

$$(6.13) \quad (\log m(n))^{\frac{1}{2}} (\beta_n^* - \beta) \xrightarrow{\mathfrak{L}} N(0, \beta^2).$$

Proof. Note that b_n in (6.12) is $\mathfrak{F}_{n-1}$-measurable. Therefore by Corollary 2 (with $b_n = b_n'$ and $u_n = 0$) and Theorem 10, (6.1) and (6.2) hold. Hence by Lemma 6, $\beta_n^* \to \beta$ a.s., and therefore $b_n \to \beta$ a.s. This then implies (1.7)–(1.9) by Theorem 1.

Now assume that $M(x)$ also satisfies (1.21). By (1.21), (6.7), and (6.9), with probability 1,

$$\beta_n^* - \beta = \left\{ \sum_1^{m(n)} O(|x_i - \theta|^{2+\eta}) + \sum_1^{m(n)} (x_i - \theta)\,\varepsilon_i \right.$$
$$\left. - (x_n - \theta) \sum_1^{m(n)} y_i + o(1) \right\} \Big/ \left\{ \sum_1^{m(n)} (x_i - \theta)^2 + o(1) \right\}$$
$$= O\left(\frac{1}{\sum_1^{m(n)} (x_i - \theta)^2} \right) + \frac{\sum_1^{m(n)} (x_i - \theta)\,\varepsilon_i}{\sum_1^{m(n)} (x_i - \theta)^2 + o(1)},$$

by (1.8) and (6.8). Hence (6.13) follows from (1.9) and Theorem 5 (i). ∎

Remark. Since $(\log n)^{-1} = o((\log m(n))^{-1})$ by (1.16), although b_n in Corollary 5 is a strongly consistent estimator of β, comparison of (6.13) with (1.23) shows that it is asymptotically much less efficient than its refinement in Theorem 3.

7. Asymptotic Efficiency of $\hat{\beta}_n$ in Stochastic Designs and Proof of Theorem 4

For the linear model $y_n = \alpha + \beta x_n + \varepsilon_n$ with i.i.d. normal errors ε_n, the asymptotic efficiency of the least squares estimate $\hat{\beta}_n$ of β in the stochastic approximation schemes of Theorems 2 and 3 follows from the more general

Lemma 7. *Let $\varepsilon, \varepsilon_1, \ldots,$ be i.i.d. normal $N(0, \sigma^2)$ random variables. Let $\{x_n\}$ be a sequence of random variables satisfying* (2.1), *and assume that* (2.3)–(2.6) *hold for some constants $A_n > 0$ and θ. Let*

$$(7.1) \qquad y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1, 2, \ldots,$$

where α, β are unknown parameters. Suppose that the distribution of x_1 does not depend on β and that for $n \geq 2$ the conditional distribution of x_n given $x_1, y_1, \ldots, x_{n-1}, y_{n-1}$ does not depend on β. Define $\hat{\beta}_n$ by (1.10) *if $\sum_1^n (x_i - \bar{x}_n)^2 > 0$, and set $\hat{\beta}_n$ equal to some constant b otherwise. Then $\hat{\beta}_n$ is an asymptotically efficient estimator of β in the sense that given any $r > 0$ and β,*

$$(7.2) \qquad \limsup_{n \to \infty} P_\beta[A_n^{\frac{1}{2}} |T_n - \beta| \leq r] \leq \lim_{n \to \infty} P_\beta[A_n^{\frac{1}{2}} |\hat{\beta}_n - \beta| \leq r]$$

for any other estimator $T_n = T_n(x_1, y_1, \ldots, x_n, y_n)$ such that if $\{\theta_n\}$ is a sequence of constants satisfying $|\theta_n| \ll A_n^{-\frac{1}{2}}$ then

$$(7.3) \qquad \lim_{n \to \infty} \{P_{\beta+\theta_n}[A_n^{\frac{1}{2}} |T_n - \beta| \leq r] - P_\beta[A_n^{\frac{1}{2}} |T_n - \beta| \leq r]\} = 0.$$

Proof. We first assume that θ and $h = \alpha + \beta\theta$ are both known so that we can reparametrize the model as

$$(7.4) \qquad y_i = h + \beta(x_i - \theta) + \varepsilon_i.$$

Moreover, since the distribution of x_1 and the conditional distribution of x_n given $x_1, y_1, \ldots, x_{n-1}, y_{n-1}$ do not depend on β and since given x_i the conditional distribution of y_i is $N(h+\beta(x_i-\theta), \sigma^2)$, the joint density f_n of $x_1, y_1, \ldots, x_n, y_n$ with respect to some σ-finite measure μ_n is of the form

$$(7.5)\qquad \begin{aligned} &f_n(x_1, y_1, \ldots, x_n, y_n \mid \beta) \\ &= (2\pi\sigma^2)^{-n/2} g_n(x_1, y_1, \ldots, x_{n-1}, y_{n-1}, x_n) \\ &\quad \cdot \exp\left\{-\sum_1^n (y_i - h - \beta(x_i-\theta))^2/(2\sigma^2)\right\}, \end{aligned}$$

where g_n does not depend on β. We shall also assume that σ, g_n, and μ_n are known so that β is the only unknown parameter in the model (7.4)-(7.5).

Consider the estimator

$$(7.6)\qquad \begin{aligned} z_n &= \left\{\sum_1^n (x_i-\theta)(y_i-h)\right\} \Big/ \sum_1^n (x_i-\theta)^2 \quad \text{if } \sum_1^n (x_i-\theta)^2 > 0, \\ &= b \quad \text{otherwise.} \end{aligned}$$

We note that $z_n = \beta + \left\{\sum_1^n (x_i-\theta)\,\varepsilon_i \Big/ \sum_1^n (x_i-\theta)^2\right\}$ in the event $\left[\sum_1^n (x_i-\theta)^2 > 0\right]$. Therefore by Theorem 5(i), for all $r>0$,

$$(7.7)\qquad P_\beta[A_n^{\frac{1}{2}}|z_n-\beta| \leq r] \to (2\pi\sigma^2)^{-\frac{1}{2}} \int_{-r}^{r} \exp\{-\tfrac{1}{2}(x/\sigma)^2\}\,dx$$

uniformly in $\beta \in (-\infty, \infty)$.

Since the function

$$\begin{aligned} \psi(d) &= \int_{d-rA_n^{-1/2}}^{d+rA_n^{-1/2}} f_n(x_1, y_1, \ldots, x_n, y_n \mid \beta)\,d\beta \\ &= g_n(x_1, \ldots, y_{n-1}, x_n) \int_{d-rA_n^{-1/2}}^{d+rA_n^{-1/2}} \\ &\quad \cdot \exp\left\{-\sum_1^n [y_i - h - \beta(x_i-\theta)]^2/(2\sigma^2)\right\} d\beta \end{aligned}$$

has its maximum at $d=z_n$, z_n is the maximum probability estimator of β with respect to the interval $(-r, r)$ (cf. [13]). Therefore in view of the uniform convergence in (7.7), it follows from a theorem of Weiss and Wolfowitz (cf. Theorem 3.1 of [13]) that

$$(7.8)\qquad \limsup_{n\to\infty} P_\beta[A_n^{\frac{1}{2}}|T_n-\beta| \leq r] \leq \lim_{n\to\infty} P_\beta[A_n^{\frac{1}{2}}|z_n-\beta| \leq r]$$

for any other estimator $T_n = T_n(x_1, y_1, \ldots, x_n, y_n)$ such that (7.3) holds for all sequences $\{\theta_n\}$ satisfying $|\theta_n| \ll A_n^{-\frac{1}{2}}$.

Now we drop the assumption that θ, h, σ, g_n are known and consider the estimator $\hat{\beta}_n$. By Theorem 5(ii), for any fixed θ, h, σ, g_n, (7.7) still holds with z_n replaced by $\hat{\beta}_n$. Hence the desired conclusion follows. ∎

Proof of Theorem 4. For the stochastic approximation scheme $\{x_n\}$ of Theorem 2 or of Theorem 3, since $\sum_1^n (x_i-\theta)^2 \sim (\sigma^2/\beta^2)\log n$ a.s. and $|x_n-\theta| \ll n^{-\frac{1}{2}} \cdot (\log\log n)^{\frac{1}{2}}$ a.s. by Theorems 2 and 3, the assumptions (2.3)–(2.6) and (2.9) of Theorem 5 are satisfied with $A_n=(\sigma^2/\beta^2)\log n$. Hence the asymptotic normality of $\hat{\beta}_n$ in (1.22) follows from Theorem 5. Moreover, in view of (1.6) with $b_n=b_n(x_1, y_1, \ldots, x_n, y_n)$, the conditional distribution of x_{n+1} given $x_1, y_1, \ldots, x_n, y_n$ is degenerate at the point $x_n - y_n/(nb_n)$ and does not depend on β. Therefore Lemma 7 is applicable and gives the asymptotic efficiency of $\hat{\beta}_n$ in the linear case $M(x)=\beta(x-\theta)$ with normal errors ε_i. ∎

Acknowledgment. We wish to thank Aryeh Dvoretzky for valuable suggestions and helpful comments on the paper. In particular, the example in Sect. 2 was suggested by him.

References

1. Anbar, D.: A stochastic Newton-Raphson method. J. Statist. Planning & Inference **2**, 153–163 (1978)
2. Anderson, T.W., Taylor, J.: Strong consistency of least squares estimates in dynamic models. Ann. Statist. **7**, 484–489 (1979)
3. Bojanic, R., Seneta, E.: A unified theory of regularly varying sequences. Math. Z. **134**, 91–106 (1973)
4. Drygas, H.: Weak and strong consistency of the least squares estimators in regression models. Z. Wahrscheinlichkeitstheorie verw. Gebiete **34**, 119–127 (1976)
5. Dvoretzky, A.: Asymptotic normality for sums of dependent random variables. Proc. 6th Berkeley Symp. Math. Statist. Probability **2**, 513–535. Berkeley-Los Angeles: Univ. California Press 1972
6. Freedman, D.: Another note on the Borel-Cantelli lemma and the strong law with the Poisson approximation as a by-product. Ann. Probability **1**, 910–925 (1973)
7. Knopp, K.: Theory and Applications of Infinite Series. London-Glasgow: Blackie & Son 1951
8. Lai, T.L., Robbins, H.: Strong consistency of least-squares estimates in regression models. Proc. Nat. Acad. Sci. U.S.A. **74**, 2667–2669 (1977)
9. Lai, T.L., Robbins, H.: Local convergence theorems for adaptive stochastic approximation schemes. Proc. Nat. Acad. Sci. U.S.A. **76**, 3065–3067 (1979)
10. Lai, T.L., Robbins, H.: Adaptive design and stochastic approximation. Ann. Statist. **7**, 1196–1221 (1979)
11. Neveu, J.: Mathematical Foundations of the Calculus of Probability. San Francisco: Holden-Day 1965
12. Venter, J.: An extension of the Robbins-Monro procedure. Ann. Math. Statist. **38**, 181–190 (1967)
13. Weiss, L., Wolfowitz, J.: Maximum Probability Estimators and Related Topics. Berlin-Heidelberg-New York: Springer 1974

Received September 17, 1979; in revised form May 9, 1980