The Annals of Mathematical Statistics
1972, Vol. 43, No. 4, 1096–1109

THE STOCHASTIC APPROXIMATION APPROACH TO A DISCRIMINATION PROBLEM

BY MARTIN A. HAMILTON

Montana State University

1. Introduction. Let the function M map the real line R into R; let $\{Y(x), x \in R\}$ be a family of independent random variables, where $Y(x)$ has distribution function $F(\cdot \mid x)$ with

$$\int_{-\infty}^{\infty} y\, dF(y \mid x) = M(x) \qquad \text{and} \qquad \int_{-\infty}^{\infty} [y - M(x)]^2\, dF(y \mid x) = \sigma^2(x) < \infty\,,$$

for every $x \in R$.

Let $\alpha \in R$ be a particular value and suppose that there exists a point $\theta \in R$ such that $[M(x) - \alpha](x - \theta) > 0$, for every $x \neq \theta$.

Let $\{Z_i, i = 1, 2, \cdots\}$ be a family of independent, identically distributed random variables, independent of $\{Y(x), x \in R\}$, with distribution function $G(\cdot)$, where $\int_{-\infty}^{\infty} z\, dG(z) = \alpha$ and $\int_{-\infty}^{\infty} (z - \alpha)^2\, dG(z) = \gamma^2 < \infty$.

Suppose that the regression function M and the values θ and α are unknown and that after observing some Z's and $Y(x)$'s at various x values, one wishes to estimate θ.

For ease in presentation, observations corresponding to the Z's will be called "control observations" and observations corresponding to the $Y(x)$'s will be called "test observations."

The discrimination problem described above would be fairly straightforward if the form of the regression function M and the distribution functions $F(\cdot \mid x)$, $x \in R$, and G were known. For example, if M is linear, G is normal and $F(\cdot \mid x)$ is normal with $\sigma^2(x) = \sigma^2$, for every $x \in R$, then a solution to the discrimination problem is given by the well-known Fieller estimate (Fieller [4]).

It is important from a practical point of view, however, to look for a solution to the discrimination problem when little is known a priori about the functions M, G, and F.

For example, suppose a scientist is comparing two drugs, a test drug and a control drug, and that he is interested in designing a biological assay to estimate the number of dose units of the test drug necessary to elicit the same mean response as the standard dose of the control drug. Suppose, further, that the experimenter knows little about the shape of the response function associated with the test drug and about the probability distribution of responses at any one dose level of either drug.

Make the following notational identifications. Let an observed response to the control drug administered at the standard dose level correspond to the random variable Z with mean α. Let the observed response to the test drug at dose level x correspond to $Y(x)$ with mean $M(x)$. Let θ be the dose level of the test drug

Received January 9, 1970; revised November 24, 1971.

that elicits a mean response, $M(\theta)$, equal to α. If $[M(x) - \alpha](x - \theta) > 0$ for every $x \neq \theta$, then this bioassay problem is essentially the discrimination problem. Obviously, neither the Fieller estimate nor any other estimate based on strong assumptions will be appropriate for this bioassay problem.

One solution to the discrimination problem that relies on only weak assumptions about M, G, and F is provided by a simple extension of the Robbins–Monro stochastic approximation process (Robbins and Monro [7]) which is as follows.

Choose a starting value $X_1 \in R$, a sequence of positive constants $\{a_n, n = 1, 2, \cdots\}$, and define the random sequence $\{X_n, n = 2, 3, \cdots\}$ by

$$X_{n+1} = X_n - a_n(Y_n - Z_n)\,, \tag{1.1}$$

where Y_n is a random variable distributed as $Y(X_n)$, $n \geqq 1$, and Z_n is a random variable distributed according to G. Then after n pairs of observations (viz. $(Y_1, Z_1), \cdots, (Y_n, Z_n)$), the estimate of θ is X_{n+1}. In the special case where α is known and $Z_n \equiv \alpha$, $n = 1, 2, \cdots$, this process is identical to the Robbins–Monro process.

This process can be visualized as the stochastic analogue of a numerical approximation process. If the problem were deterministic with $Z \equiv \alpha$ and $Y(x) \equiv M(x)$, for every $x \in R$, then equation (1.1) would become

$$X_{n+1} = X_n - a_n(M(X_n) - \alpha)\,, \qquad n \geqq 1\,, \tag{1.2}$$

which is the definition of a numerical approximation process having the property that $\lim_{n\to\infty} X_n = \theta$ if M and $\{a_n\}$ satisfy some weak conditions (Schmetterer [9] and von Mises and Pollaczek-Geiringer [12]). Equation (1.1) is simply (1.2) with $M(X_n)$ replaced by the unbiased estimate Y_n and α replaced by the unbiased estimate Z_n.

Since all known properties of the original Robbins–Monro process hold for the process of (1.1), the corresponding sequential estimate is an appealing solution to the discrimination problem. It seems, however, that (1.1) does not use all available information at each step in the sequential procedure. Consider, for example, an identical process with (1.1) replaced by

$$X_{n+1} = X_n - a_n(Y_n - n^{-1}\textstyle\sum_{i=1}^{n} Z_i)\,, \qquad n \geqq 1\,. \tag{1.3}$$

Since $n^{-1}\sum_{i=1}^{n} Z_i$ is a better estimate of α than Z_n, one would expect that the estimate of θ provided by the process of (1.1) should have a larger small sample mean square error than the estimate given by the process with equation (1.3). One might expect that these two estimation procedures have quite similar asymptotic properties.

Comment. One could define the X_n-sequence by $X_{n+1} = X_n - a_n(Y_n - \sum_1^n W_{ni} Z_i)$ where the chosen weights $\{W_{ni}; i = 1, \cdots, n\}$ are such that $W_{ni} \geqq 0$ and $\sum_1^n W_{ni} = 1$. Equations (1.1) and (1.3) are special cases of this equation (for (1.1), $W_{nn} = 1$; $W_{ni} = 0$, $i = 1, \cdots, n-1$ and for (1.3), $W_{nn} = 1/n$, $i = 1, \cdots, n$) and they correspond to two extremes from the set of reasonable choices of weights. For

this reason, sequential estimation processes based on (1.1) and (1.3) are especially interesting and attention is restricted to them in this study.

The purpose of this paper is to derive the properties of a stochastic approximation process, called Process II, defined by a generalized version of (1.3) and to compare the properties to those of a process, called Process I, defined by a similarly generalized version of (1.1). Surprisingly, Process I can be better than Process II in both large and small sample situations.

The paper is divided into five sections. These sections contain the introduction, preliminary information including some known asymptotic properties of Process I, the asymptotic properties of Process II, the minimization and comparison of the asymptotic normal variances, and the small sample bias and mean square error of X_n for the case where M is linear.

2. Preliminaries. Let the product Borel space, field, and measure generated jointly by the two independent families of random variables $\{Y(x), x \in R\}$ and $\{Z_i, i = 1, 2, \cdots\}$ be denoted by Ω, $\mathscr{F}$, and P respectively and let the typical element of Ω be denoted by ω.

The term "step" will henceforth be used to indicate the taking of additional observations to adjust the estimate of θ. For example, if the process of (1.3) is being employed, the experimenter observes (Y_n, Z_n) and changes his estimate from X_n to X_{n+1} "at the nth step." The "size of the nth step" is $|X_{n+1} - X_n|$.

Equations (1.1) and (1.3) provide estimation procedures in which pairs of observations (Y_i, Z_i) are taken sequentially. In practice, the experimenter may want to take more than one control observation (or possibly no control observation) at selected steps in the process. For example, it may be nearly as easy for him to take k control observations as one control observation at a time. Or he may feel that after a few steps of the process, few control observations need be taken. The following processes will allow the experimenter to vary the number of control observations taken from step to step.

Process I. Choose a starting value $X_1 \in R$, a sequence of positive numbers $\{a_n, n = 1, 2, \cdots\}$ and a strictly increasing sequence of positive integers $\{b_n, n = 1, 2, \cdots\}$. Observe the sets of random variables $(Y_i, Z_{b_{i-1}+1}, \cdots, Z_{b_i})$, $i = 1, 2, \cdots$, sequentially, where Y_i is distributed as $Y(X_i)$ and

$$X_{n+1} = X_n - a_n(Y_n - \tilde{Z}_n), \tag{2.1}$$

where

$$\tilde{Z}_n = (b_n - b_{n-1})^{-1} \sum_{i=b_{n-1}+1}^{b_n} Z_i .$$

At the kth step of experimentation the estimate of θ is X_{k+1}. Notice that at least one control observation must be taken at each step.

Process II. Choose a starting value $X_1 \in R$, a sequence of positive numbers $\{a_n, n = 1, 2, \cdots\}$ and a non-decreasing sequence of positive integers $\{b_n, n = 1, 2, \cdots\}$. Observe the sets $(Y_i, Z_{b_{i-1}+1}, \cdots, Z_{b_i})$, $i = 1, 2, \cdots$, sequentially,

where Y_i is distributed as $Y(X_i)$ and

$$X_{n+1} = X_n - a_n(Y_n - \bar{Z}_n), \qquad \text{where} \quad \bar{Z}_n = b_n^{-1} \sum_{i=1}^{b_n} Z_i . \tag{2.2}$$

At the kth step of experimentation (i.e., after observing $(Y_1, \cdots, Y_k, Z_1, \cdots, Z_{b_k})$) the estimate of θ is X_{k+1}. Notice that if $b_k = b_{k-1}$, then only the test observation Y_k is taken at the kth step in the process.

The properties of Process I, which is a trivial extension of the Robbins-Monro procedure, are known. The properties of Process II have not previously been studied.

The following conditions will be referred to by number in subsequent sections.

Conditions on the distribution functions $F(\cdot \mid x)$, $x \in R$.

F 1. $\sigma^2(x) \leqq \sigma^2 < \infty$ for every $x \in R$.

F 2. $\lim_{x \to \theta} \sigma^2(x) = \sigma^2(\theta)$.

F 3. $\lim_{R \to \infty} \lim_{\zeta \to 0} \sup_{|x-\theta|<\zeta} \int_{\{|y - M(x)| > R\}} [y - M(x)]^2 \, dF(y \mid x) = 0$.

F 4. $F([w + M(x)] \mid x) = F_0(w)$, for every $x \in R$ and $w \in R$, where $F_0(\cdot)$ does not depend on x; $\int_{-\infty}^{\infty} w \, dF_0(w) = 0$; and $\int_{-\infty}^{\infty} w^2 \, dF_0(w) = \sigma^2(\theta) < \infty$. This condition means that $\{[Y(x) - M(x)]\}$ are independently and identically distributed.

Conditions on the regression function M.

M 1. $M(\cdot)$ is Borel measurable and for some $K_1, K_2 > 0$,

$$|M(x) - \alpha| \leqq K_2|x - \theta| + K_1, \qquad \text{for every} \quad x \in R .$$

M 2. For every $\delta \in (0, 1)$,

$$\inf_{\delta < |x-\theta| < \delta^{-1}} \{|M(x) - \alpha|\} > 0 .$$

M 3. There exist $\beta > 0$ and $s = 1$ or 2 such that for every $x \in R$,

$$M(x) = \alpha + \beta(x - \theta) + \delta(x, \theta), \qquad \text{where} \quad \delta(x, \theta) = o(|x - \theta|^s)$$

as $x \to \theta$.

M 4. $M(\cdot)$ is Borel measurable and for some $K_3, K_4 > 0$,

$$K_3|x - \theta| \leqq |M(x) - \alpha| \leqq K_4|x - \theta| , \qquad \text{for every} \quad x \in R .$$

Conditions on the sequence of positive numbers $\{a_n\}$.

A 1. (a) $\sum_1^{\infty} a_n = \infty$ (b) $\sum_1^{\infty} a_n^2 < \infty$.

A 2. There exist $K_0 > 0$ and $\Psi \in (\frac{1}{2}, 1]$ such that

$$a_n = K_0 n^{-\Psi}[1 + o(1)], \qquad n = 1, 2, \cdots .$$

It is easy to show that A 2 implies A 1.

Conditions on the non-decreasing sequence of positive integers $\{b_n\}$.

B 1. $\sum_1^{\infty} a_n b_n^{-\frac{1}{2}} < \infty$.

B 2. There exist $K_5 > 0$ and $\lambda > 2 - 2\Psi$, where Ψ is defined in condition A 2, such that $b_n = K_5 n^{\lambda}[1 + o(1)]$.

The following well-known lemma is used repeatedly in Section 3.

LEMMA 2.1. *Let k_0 be a fixed positive integer. Then*

(i) $\lim_{n\to\infty}[(a+1)n^{-(a+1)}\sum_{k=k_0}^{n}k^a]=1$, *if* $a>-1$,

(ii) *there exists a positive constant $c(k_0)$ such that*

$$\sum_{k=k_0}^{n}k^{-1}\leqq c(k_0)+\log_e n, \quad \textit{for every} \quad n>1;$$

and

(iii) $\sum_{k=k_0}^{n}k^a\leqq(-a-1)^{-1}[(k_0-1)^{1+a}-n^{1+a}]$, *if* $a<-1$,

The following theorem is a special case of Theorem 3 of Venter [10], which is a generalization of a theorem of Dvoretzky [3]. The details will not be given here.

THEOREM 2.1. *Assume Process* I *and conditions* F 1, M 1, M 2, *and* A 1. *Then* $\lim_{n\to\infty}X_n=\theta$ a.s. *and* $\lim_{n\to\infty}E(X_n-\theta)^2=0$.

The method of proof due to Sacks [8] can be used directly to prove Theorem 2.2 below.

THEOREM 2.2. *Suppose Process* I *under conditions* F 1, F 2, F 3, M 3, M 4, *and* B 2 *with* $\lambda=1$ *and* $K_5\geqq 1$. *Let* $a_n=K_0n^{-1}$ *for every* $n\geqq 1$, *where* $K_0>(2\beta)^{-1}$. *Then* $n^{\frac{1}{2}}(X_n-\theta)$ *is asymptotically normally distributed with mean* 0 *and variance* $K_0^2(\sigma^2(\theta)+K_5^{-1}\gamma^2)(2K_0\beta-1)^{-1}$.

3. Asymptotic properties of Process II.

THEOREM 3.1. *Suppose Process* II *under conditions* F 1, M 1, M 2, A 1, *and* B 1. *Then* $\lim_{n\to\infty}X_n=\theta$ a.s. *and* $\lim_{n\to\infty}E(X_n-\theta)^2=0$.

PROOF. It is not difficult to demonstrate that conditions F 1, M 1, M 2, A 1, and B 1 satisfy the hypothesis of Theorem 3 of Venter [10] (Hamilton [5]). The method of Blum [1] can also be used to prove $\lim_{n\to\infty}X_n=\theta$ a.s.

COROLLARY 3.1. *Theorem* 3.1 *holds if* A 1 *and* B 1 *are replaced by* A 2 *and* B 2.

PROOF. It is easy to show that A 2 and B 2 imply A 1 and B 1.

Comment. Similar consistency conditions for Process II are given by Hamilton [5] for the cases where the b_n-sequence is random and/or the a_n-sequence is random. For Process I with random a_n, see Venter [11].

THEOREM 3.2. *Suppose Process* II *under conditions* M 3, M 4, *and* B 2 *with* $\Psi=1$. *Let* $K_0>(2K_3)^{-1}$ *and* $a_n=K_0n^{-1}$, $n=1,2,\cdots$.

(i) *Suppose* $\lambda<1$ (*see* B 2), $s=2$ (*see* M 3), F 1, F 2, *and* F 3. *Then* $n^{\lambda/2}(X_n-\theta)$ *has an asymptotic normal distribution with mean* 0 *and variance* $[2K_5^{-1}\gamma^2K_0\beta^{-1}]\times[2K_0\beta-\lambda]^{-1}$.

(ii) *Suppose* $\lambda=1$, $s=2$, *and* F 4. *Then* $n^{\frac{1}{2}}(X_n-\theta)$ *has an asymptotic normal distribution with mean* 0 *and variance* $K_0^2[2K_0\beta-1]^{-1}[\sigma^2(\theta)+2K_5^{-1}\gamma^2K_0^{-1}\beta^{-1}]$.

(iii) *Suppose* $\lambda>1$, $s=1$, F 1, F 2, *and* F 3. *Then* $n^{\frac{1}{2}}(X_n-\theta)$ *has an asymptotic normal distribution with mean* 0 *and variance* $K_0^2\sigma^2(\theta)[2K_0\beta-1]^{-1}$.

PROOF. Without loss of generality, let $\theta = 0$. Define $b_0 = 0$. Henceforth, an unsubscripted K will be used generically for a positive constant and $\delta(X_k, 0)$ of M 3 will be written δ_k. Define, for $a > 0$; $k = 0, 1, \cdots, n$; $n = 1, 2, \cdots$,

$$\beta_{kn} = \prod_{i=k+1}^{n}(1 - i^{-1}a), \qquad k = 0, 1, \cdots, n-1 \tag{3.1}$$
$$= 1, \qquad k = n.$$

Following the method of Sacks [8] (2.2) may be written

$$X_{n+1} = X_1\beta_{0n} - \beta^{-1}\sum_{k=1}^{n} ak^{-1}\beta_{kn}\delta_k - \beta^{-1}\sum_{k=1}^{n} ak^{-1}\beta_{kn}[Y_k - M(X_k)] + \beta^{-1}\sum_{k=1}^{n} ak^{-1}\beta_{kn}(\bar{Z}_k - \alpha), \tag{3.2}$$

where β_{kn} is defined by (3.1) with $a = K_0\beta > \frac{1}{2}$.

Let $h_n = (\sum_{k=1}^{n} a^2k^{-2}\beta_{kn}^2)^{-\frac{1}{2}}$ and $g_n = \min\{1, n^{-(1-\lambda)/2}\}$, $\lambda > 0$, $n = 1, 2, \cdots$.

According to Sacks [8],

$$h_n = n^{\frac{1}{2}}(2a - 1)^{\frac{1}{2}}a^{-1}[1 + o(1)] \quad \text{as} \quad n \to \infty \tag{3.3}$$

and

$$\beta_{kn} = k^a n^{-a}[1 + o(1)] \quad \text{as} \quad k \to \infty, \ k \leqq n. \tag{3.4}$$

To find the asymptotic distribution of $n^{\frac{1}{2}}g_nX_n$, it is sufficient to derive the asymptotic distribution $g_nh_nX_n$. For, if $g_nh_nX_n$ is asymptotically distributed as $N(0, \sigma_*^2)$ then $n^{\frac{1}{2}}g_nX_n$ is asymptotically distributed as $N(0, a^2(2a - 1)^{-1}\sigma_*^2)$.

The following four lemmas form the body of the proof.

LEMMA 3.1. *Under* (i), (ii), *and* (iii), $\lim_{n\to\infty} g_nh_nX_1\beta_{0n} = 0$.

LEMMA 3.2. *Under* (i), (ii), *and* (iii), $\lim_{n\to\infty} g_nh_n\sum_{k=1}^{n} ak^{-1}\beta_{kn}\delta_k = 0$ *in probability.*

LEMMA 3.3. $g_nh_n\sum_{k=1}^{n} ak^{-1}\beta_{kn}[Y_k - M(X_k)]$ *is asymptotically distributed as degenerate at* 0 *if* (i) *and as* $N(0, \sigma^2(0))$ *if* (ii) *or* (iii).

LEMMA 3.4. $g_nh_n\sum_{k=1}^{n} ak^{-1}\beta_{kn}(\bar{Z}_k - \alpha)$ *is asymptotically distributed as* $N(0, 2\gamma^2(2a - 1)a^{-1}K_5^{-1}(2a - \lambda)^{-1})$ *if* (i); *as* $N(0, 2\gamma^2a^{-1}K_5^{-1})$ *if* (ii); *and as degenerate at* 0 *if* (iii).

PROOF OF LEMMA 3.1. The proof follows directly from Sacks ([8], Lemma 2).

PROOF OF LEMMA 3.2. Choose $t > 0$. By M 3, there exists $\zeta > 0$ such that for every $x \in (-\zeta, \zeta)$, $|\delta(x, 0) \cdot x^{-s}| \leqq t^2$. Since M 4 implies M 1 and F 4 implies F 1, Corollary 3.1 holds and $\lim_{n\to\infty} X_n^s = 0$ a.s. Thus there exists $N > a - 1$ such that $\Pr\{|X_n^s| \leqq \zeta, \text{ for every } n \geqq N\} \geqq 1 - t$.

Let

$$V(n, N) = g_nh_n\sum_{k=N}^{n} ak^{-1}\beta_{kn}\delta_k, \qquad n \geqq N. \tag{3.5}$$

Then, as shown by Sacks ([8], page 383), for every $n > N$.

$$\Pr\{|V(n, N)| > t\} \leqq t[g_nh_n\sum_{k=N}^{n} ak^{-1}\beta_{kn}E(|X_k^s|)] + t.$$

Clearly,

$$E(X_{n+1}^2) = E\{X_n^2[1 - K_0n^{-1}(M(X_n) - \alpha)X_n^{-1}]^2\} + O(n^{-2}) + 2d_n, \tag{3.6}$$

where $d_n = E\{X_n(\bar{Z}_n - \alpha)K_0n^{-1}[1 - K_0n^{-1}X_n^{-1}(M(X_n) - \alpha)]\}$.

By M4, for every $n > K_0K_3$, $|1 - K_0n^{-1}X_n^{-1}(M(X_n) - \alpha)| \leqq (1 - K_3K_0n^{-1})$; and therefore, $|d_n| \leqq E\{|X_n(\bar{Z}_n - \alpha)|K_0n^{-1}(1 - K_3K_0n^{-1})\}$ or

$$|d_n| \leqq K_0n^{-1}\gamma n^{-\lambda/2}\{E(X_n^2)\}^{\frac{1}{2}}, \qquad \text{for every} \quad n > K_0K_3. \tag{3.7}$$

By Corollary 3.1, $\lim_{n\to\infty} E(X_n^2) = 0$; and therefore,

$$|d_n| = o(n^{-1-\lambda/2}) \quad \text{as} \quad n \to \infty. \tag{3.8}$$

If $s = 2$ and $\lambda \leqq 1$, then by M4, (3.6), and (3.8)

$$E(|X_{n+1}^s|) \leqq E(X_n^2)[1 - 2K_0n^{-1}K_3 + K_0^2n^{-2}K_4^2] + o(n^{-1-\lambda/2});$$

and by Venter ([10] Lemma 2),

$$E(|X_{n+1}^2|) = O(n^{-\lambda/2}), \qquad \lambda \leqq 1. \tag{3.9}$$

If $s = 1$ and $\lambda > 1$, then by (3.6), (3.7), and M4

$$\{E(|X_{n+1}^s|)\}^2 \leqq E(X_{n+1}^2) \leqq E(X_n^2)[1 - n^{-1}(2K_0K_3 + K_0^2n^{-1}K_4^2)] + O(n^{-2}) + [E(X_n^2)]^{\frac{1}{2}}O(n^{-1-\lambda/2}).$$

By Corollary 3.1, $\{E(X_n^2)\}^{\frac{1}{2}} \to 0$ as $n \to \infty$. The sequence of positive numbers $\{E(X_n^2), n = 1, 2, \cdots\}$ satisfies Lemma 2 of Burkholder [2] (with $g = \lambda/2$, $r = \frac{1}{2}$, $p = 1$, and $c = 2K_0K_3$); and therefore, $E(X_n^2) = O(n^{-1})$ or

$$E(|X_n|) = O(n^{-\frac{1}{2}}), \qquad \lambda > 1. \tag{3.10}$$

By Lemma 2.1, (3.3), (3.4), (3.9), and (3.10), $g_nh_n\sum_{k=N}^{n} ak^{-1}\beta_{kn}E(|X_k|^s) = O(1)$ under (i), (ii), and (iii) so that $\lim_{n\to\infty} V(n, N) = 0$ in probability. It is not difficult to show that $\lim_{n\to\infty} g_nh_n\sum_{k=1}^{N-1} ak^{-1}\beta_{kn}\delta_k = 0$ in probability and therefore Lemma 3.2 is true.

PROOF OF LEMMA 3.3. The proof follows directly by imitating the proof of Sacks [8] and appealing to a slightly extended form of Sacks' Lemma 6.

PROOF OF LEMMA 3.4. Let

$$\begin{aligned} T_n &= g_nh_n\sum_{k=1}^{n} ak^{-1}\beta_{kn}b_k^{-1}[\sum_{j=1}^{b_k}(Z_j - \alpha)] \\ &= g_nh_n\sum_{j=1}^{b_n}(Z_j - \alpha)\sum_{k=c_j}^{n} ak^{-1}\beta_{kn}b_k^{-1}, \end{aligned}$$

where $c_j = \max\{k: b_k < j, k = 0, 1, 2, \cdots\} + 1$.
Let $T_{jn} = h_ng_n(Z_j - \alpha)(\sum_{k=c_j}^{n} ak^{-1}\beta_{kn}b_k^{-1})$, $j = 1, \cdots, b_n$.

Then $T_n = \sum_{j=1}^{b_n} T_{jn}$ and $\{T_{jn}, j = 1, \cdots, b_n\}$ is a set of independent random variables with $E(T_{jn}) = 0$ and

$$E(T_{jn}^2) = g_n^2h_n^2(\sum_{k=c_j}^{n} ak^{-1}\beta_{kn}b_k^{-1})^2\gamma^2.$$

Also, since for $j \leqq k \leqq n$,

$$\beta_{jn} = \beta_{jk}\beta_{kn},$$

$$\operatorname{Var}(T_n) = \gamma^2g_n^2h_n^2[\sum_{k=1}^{n}(ak^{-1}\beta_{kn}b_k^{-1})^2b_k + 2\sum_{k=2}^{n} a^2k^{-2}\beta_{kn}^2kb_k^{-1}\sum_{j=1}^{k-1}\beta_{jk}j^{-1}]. \tag{3.11}$$

By B 2, (3.4), and (3.3), given $\xi > 0$, there exists $N > a$ such that for every $n > N$,

$$(1 - \xi)(kn^{-1})^a \leqq \beta_{kn} \leqq (1 + \xi)(kn^{-1})^a ; \tag{3.12}$$

for every $n > k > N$,

$$(1 - \xi)(2a - 1)a^{-2}n \leqq h_n^2 \leqq (1 + \xi)(2a - 1)a^{-2}n ; \tag{3.13}$$

for every $k \geqq N$,

$$(1 - \xi)K_5 k^{\lambda} \leqq b_k \leqq (1 + \xi)K_5 k^{\lambda} ; \tag{3.14}$$

and, since $a > \frac{1}{2}$, by Lemma 2.1 there exists $N_1 > N$ such that for every $k > N_1$,

$$(1 - \xi)k^a a^{-1} \leqq \textstyle\sum_{j=N}^{k-1} j^{a-1} \leqq (1 + \xi)k^a a^{-1} . \tag{3.15}$$

Write $\operatorname{Var}(T_n) = \gamma^2[T_n^{(1)} + 2(T_n^{(2)} + T_n^{(3)} + T_n^{(4)} + T_n^{(5)})]$, where

$$T_n^{(1)} = g_n^2 h_n^2 \textstyle\sum_{k=1}^{n} a^2 k^{-2} \beta_{kn}^2 b_k^{-1} ;$$

$$T_n^{(2)} = \textstyle\sum_{k=2}^{N} \sum_{j=1}^{k-1} D(j, k, n) ; \qquad T_n^{(3)} = \sum_{k=N+1}^{n} \sum_{j=1}^{N-1} D(j, k, n) ;$$

$$T_n^{(4)} = \textstyle\sum_{N+1}^{N_1} \sum_{j=N}^{k-1} D(j, k, n) ; \qquad T_n^{(5)} = \sum_{k=N_1+1}^{n} \sum_{j=N}^{k-1} D(j, k, n) ;$$

and $D(j, k, n) = (g_n h_n a k^{-1} \beta_{kn})^2 k b_k^{-1} \beta_{jk} j^{-1}$.

Using (3.12)—(3.15), Lemma 2.1, and Lemma 3 of Sacks [8], it is easy to show that $|T^{(1)}|$, $|T^{(2)}|$, $|T^{(3)}|$, and $|T^{(4)}|$ approach zero as $n \to \infty$ and that $T_n^{(5)} = K_5^{-1} a^{-1}(2a - 1)(2a - \lambda)^{-1}[1 + o(1)]$ if $\lambda \leqq 1$ and $T_n^{(5)} = o(1)$ if $\lambda > 1$ as $n \to \infty$.

Combining these results,

$$\begin{aligned} \lim_{n\to\infty} \operatorname{Var}(T_n) &= 2\gamma^2 K_5^{-1} a^{-1}(2a - 1)(2a - \lambda)^{-1} , && \text{if } \lambda < 1 \\ &= 2\gamma^2 K_5^{-1} a^{-1} , && \text{if } \lambda = 1 \\ &= 0 , && \text{if } \lambda > 1 . \end{aligned} \tag{3.16}$$

Choose $\zeta > 0$ and let $\Phi_{jn}(\zeta) = 1$ if $|T_{jn}| > \zeta$ and $\Phi_{jn}(\zeta) = 0$ otherwise. Then

$$\textstyle\sum_{j=b_n}^{n} E\{T_{jn}^2 \Phi_{jn}(\zeta)\} = \sum_{j=b_n}^{n} \int_{R_{jn}} T_{jn}^2(\omega)\, dP(\omega) ,$$

where $R_{jn} = \{\omega : |T_{jn}(\omega)| > \zeta\}$; i.e.,

$$R_{jn} \subset \{\omega : |Z_j - \alpha| > \zeta h_n^{-1} g_n^{-1}[\textstyle\sum_{k=c_j}^{n} a k^{-1} b_k^{-1} |\beta_{kn}|]^{-1}\} .$$

It is not difficult to show that for some $K > 0$ and n sufficiently large,

$$\zeta g_n^{-1} h_n^{-1}[\textstyle\sum_{k=c_j}^{n} a k^{-1} b_k^{-1} |\beta_{kn}|]^{-1} \geqq \zeta K n^{-\min(\lambda,1)/2} .$$

Let $R_n = \{\omega : |Z(\omega) - \alpha| > \zeta K n^{-\min(\lambda,1)/2}\}$. Then

$$\textstyle\sum_{j=b_n}^{n} E[T_{jn}^2 \Phi_{jn}(\zeta)] \leqq \sum_{j=b_n}^{n} h_n^2 g_n^2 (\sum_{k=c_j}^{n} a k^{-1} b_k^{-1} \beta_{kn})^2 \int_{R_n} (Z - \alpha)^2\, dP . \tag{3.17}$$

It is easy to show that the right-hand side of (3.17) approaches zero as $n \to \infty$ by using the argument that established (3.16).

By the Lindeberg-Feller Central Limit Theorem for double sequences, T_n has an asymptotic normal distribution with mean zero and variance $\lim_{n\to\infty} \operatorname{Var}(T_n)$ as given by (3.16).

The proof of Lemma 3.4 is complete.

PROOF OF THEOREM 3.2. Under the hypothesis of Theorem 3.2 (i), by Slutsky's Theorem, Lemmas 3.1, 3.2, 3.3, and 3.4, and equation (3.2), $g_n h_n X_n$ has an asymptotic normal distribution with mean zero and variance $2\beta^{-2}\gamma^2 K_5^{-1}a^{-1}(2a-1)(2a-\lambda)^{-1}$ or $n^{\lambda/2}X_n$ is asymptotically distributed as $N(0, \gamma^2 K_5^{-1} 2K_0\beta^{-1}(2K_0\beta - \lambda)^{-1})$.

Under the hypothesis of Theorem 3.2 (ii), the two families of random variables $\{\sum_{k=1}^n ak^{-1}\beta_{kn}[Y_k - M(X_k)]\}$ and $\{\sum_{k=1}^n ak^{-1}\beta_{kn}(\bar{Z}_k - \alpha)\}$ are independent, and by Lemmas 3.3 and 3.4 the asymptotic distribution of $g_n h_n \sum_{k=1}^n ak^{-1}\beta_{kn} \times \{[Y_k - M(X_k)] - [\bar{Z}_k - \alpha]\}$ is $N(0, \sigma^2(0) + 2\gamma^2 a^{-1}K_5^{-1})$. The conclusion of Theorem 3.2 (ii) follows directly.

Similarly, under the hypothesis of Theorem 3.2 (iii), $g_n h_n X_n$ is asymptotically distributed as $N(0, \sigma^2(0))$ and the conclusion of Theorem 3.2 (iii) follows directly.

The proof of Theorem 3.2 is completed.

COROLLARY 3.2. *Theorem* 3.2 *remains true if* $K_3 = 0$ *and* $a_n = K_0 n^{-1}$, $n = 1, 2, \cdots$, *where* $K_0 > (2\beta)^{-1}$.

PROOF. Corollary 3.2 follows from Theorem 3.2 by an argument suggested by Hodges and Lehmann [6]. The details will not be given here.

Comment. The proof of Theorem 3.2 contains information about rates of convergence. Preceding (3.10) is $E(X_n - \theta)^2 = O(n^{-1})$, if $\lambda > 1$, and using (3.6), (3.7), Lemma 2 of Venter [10], and an inductive proof, it can be shown that for any $\zeta > 0$, $E(X_n - \theta)^2 = O(n^{-\lambda(1-\zeta)})$, if $\lambda \leqq 1$.

4. Optimal design. Using Theorems 2.2 and 3.2 something can be said about optimal choices for λ, K_0, and K_5.

Assume Process II with $a_n = K_0 n^{-1}$ and let $N_\lambda(n)$ be the total number of observations taken in the first n steps when $b_n \sim K_5 n^\lambda$. Then, for large n, $N_\lambda(n) \sim n + K_5 n^\lambda$. Let $m(n, \lambda) = (n + K_5 n^\lambda)(1 + K_5)^{-1}$. Then $N_1(m(n, \lambda)) \sim N_\lambda(n)$.

Let $V_\lambda{}^*(n)$ be the asymptotic normal variance of $n^{\min\{1,\lambda\}/2}(X_n - \theta)$ for Process II according to Theorem 3.2. Then the variance of the normal distribution one would use to place confidence intervals around $(X_n - \theta)$ when n is large is given by $V_\lambda(n) = n^{-\min\{1,\lambda\}} V_\lambda{}^*(n)$.

Let $\lambda_0 < 1$ and $\lambda_1 > 1$. Then $V_{\lambda_0}(n)V_1^{-1}(m(n, \lambda_0)) \to \infty$; as $n \to \infty$; and $V_{\lambda_1}(n)V_1^{-1}(m(n, \lambda_1)) \to \infty$ as $n \to \infty$. Thus, regardless of the values of K_0, K_5, γ^2, β, and $\sigma^2(\theta)$, the confidence interval about $(X_n - \theta)$ for $\lambda \neq 1$ is wider than for $\lambda = 1$ when the total number of observations is a large fixed number. (The choice $\lambda = 1$ is also best if one assigns a finite cost to each observation and minimizes the length of the confidence interval with respect to λ subject to a large, fixed total cost, even if the cost of a control observation is different from the cost of a test observation.) On the basis of these asymptotic considerations it seems advisable to choose the sequence $\{b_n\}$ subject to $b_n \sim K_5 n$. Henceforth, let $\lambda = 1$ and drop the subscript from $V_1(n)$.

Consider Process II with $a_n = K_0 n^{-1}$ and $b_n \sim K_5 n$. Let $N_1(n)$ be fixed and

large and write $N_1(n)$ simply as N. Let $T = \gamma^{-2}\sigma^2(\theta)$. Let

$$V = \min_{K_0, K_5}\{V(n);\ N \text{ fixed},\ K_5 > 0,\ \beta K_0 > \tfrac{1}{2}\}\,.$$

Let $\hat{K}_0$ and $\hat{K}_5$ be the values of K_0 and K_5 at which V is attained. Then

$$\hat{K}_0 = (2\beta)^{-1}[1 + (1 + 4T^{-1}\hat{K}_5^{-1})^{\frac{1}{2}}]\,; \qquad \hat{K}_5 = (2/T\beta\hat{K}_0)^{\frac{1}{2}}\,. \tag{4.1}$$

Approximate solutions for (4.1) are easy to compute using the Newton–Raphson method.

Consider Process I with $a_n = K_0 n^{-1}$ and $b_n \sim K_5 n$, $K_5 \geqq 1$. Referring to Theorem 2.2 let $U(n) = K_0^2(2K_0\beta - 1)^{-1}\gamma^2[T + K_5^{-1}]n^{-1}$ and

$$U = \min_{K_0, K_5}\{U(n);\ N \text{ fixed: } K_5 \geqq 1;\ K_0\beta > \tfrac{1}{2}\}\,.$$

Let $\tilde{K}_0$ and $\tilde{K}_5$ be the values of K_0 and K_5 at which U is attained. Then

$$\tilde{K}_0 = \beta^{-1}\,; \qquad \tilde{K}_5 = \max\{1,\ T^{-\frac{1}{2}}\}\,. \tag{4.2}$$

The quantities

$$N\gamma^{-2}\beta^2 V = (\hat{K}_0\beta)^2(2\hat{K}_0\beta - 1)^{-1}(1 + \hat{K}_5)[T + 2(\hat{K}_5\hat{K}_0\beta)^{-1}] \tag{4.3}$$

and

$$N\gamma^{-2}\beta^2 U = (\tilde{K}_0\beta)^2(2\tilde{K}_0\beta - 1)^{-1}(1 + \tilde{K}_5)[T + \tilde{K}_5^{-1}] \tag{4.4}$$

are standardized forms of the minimum variance of the asymptotic (in n) distribution of $(1 + K_5)^{\frac{1}{2}}n^{\frac{1}{2}}(X_n - \theta) \sim N^{\frac{1}{2}}(X_n - \theta)$ for Process II and Process I, respectively. The right-hand sides of (4.3) and (4.4) are easy to compute and conveniently illustrate the relationship between the minimum variance and the important quantities K_5, $K_0\beta$, and $\sigma^2(\theta)\cdot\gamma^{-2} = T$. Some typical values of $\hat{K}_5$, $\beta\hat{K}_0$, $\tilde{K}_5$, $\beta\tilde{K}_0$, $N\gamma^{-2}\beta^2 V$, and $N\gamma^{-2}\beta^2 U$ are presented in Table 1.

Notice that bounds in terms of $K_5 T$ can be placed on $U(n)V^{-1}(n)$; e.g., $\frac{1}{4} < (K_5 T+1)(K_5 T+4)^{-1} < U(n)V^{-1}(n) < (1+K_5^{-1}T^{-1})$ and $\lim_{K_5 T\to\infty}[U(n)V^{-1}(n)] = 1$.

Table 1 suggests that if K_0 and K_5 can be chosen optimally, then Process II is asymptotically better than Process I if and only if T is large. It should be pointed out, however, that Process I as defined and discussed above is constrained so

TABLE 1

Optimum values of K_0 and K_5 and the corresponding standardized asymptotic normal variance

	$T = \sigma^2(\theta)/\gamma^2$								
	.01	.10	.25	.5	1	2	4	10	100
Process II									
$\hat{K}_0\beta$	4.38	2.43	2.00	1.75	1.57	1.42	1.31	1.20	1.07
$\hat{K}_5$	6.76	2.87	2.00	1.51	1.13	.84	.62	.41	.14
$N\gamma^{-2}\beta^2 V$	1.49	2.29	3.00	3.86	5.22	7.41	11.09	20.40	129.77
Process I									
$\tilde{K}_0\beta$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\tilde{K}_5$	10.00	3.16	2.00	1.41	1.00	1.00	1.00	1.00	1.00
$N\gamma^{-2}\beta^2 U$	1.21	1.73	2.25	2.91	4.00	6.00	10.00	22.00	202.00

that at any stage in the sequential procedure, the total number of test observations is less than or equal to the total number of control observations. Clearly, such a procedure will have poor properties when the test observations are much more variable than the control observations. If the definition of Process I is altered in the obvious way to allow m test observations to be taken at X_k in the kth step, $k = 1, 2, \cdots$, then it can be shown that $N\gamma^{-2}\beta^2$ times the minimum variance of the asymptotic distribution of the resulting estimate is less than $\gamma^{-2}\beta^2 V$, even for large values of T. It can also be shown that altering the definition of Process II in a similar manner produces no change (other than notational) in the discussion of the properties of Process II.

Since $U(n)V^{-1}(n) \gtreqless 1$ as $K_0\beta \gtreqless 2$, the choice of K_0 has a strong influence on the large sample precision of Process I relative to Process II. Process I will yield larger step sizes than Process II on the average; and therefore, if K_0 is small with respect to $2\beta^{-1}$ then Process II produces step sizes that are too small for rapid convergence of X_n to θ while if K_0 is large with respect to $2\beta^{-1}$ then Process I produces step sizes that are too large.

The implication of the last two paragraphs is that one can do worse by taking sequential steps toward the mean of the control observations (Process II) rather than toward the most recent control observation (Process I). In the next section, it is shown that this unexpected result, based on large sample theory, remains true in a simplified small sample situation.

5. Small sample m.s.e. **when M is linear.** Suppose that $M(x) = \alpha + \beta x$, $\beta > 0$, $x \in R$. Let $\theta = 0$, $a_n = K_0 n^{-1}$, and $b_n = \langle K_5 n \rangle$ if $K_5 \geqq 1$ and $b_n = 1 + \langle K_5 n \rangle$ if $0 < K_5 < 1$, where $\langle x \rangle$ is "the greatest integer $\leqq x$." Let $E_{\mathrm{I}}(X_{n+1}^2)$ and $E_{\mathrm{I}}(X_{n+1})$ denote the mean-square error (m.s.e.) and bias at the nth step of estimation for Process I and let $E_{\mathrm{II}}(X_{n+1}^2)$ and $E_{\mathrm{II}}(X_{n+1})$ be the m.s.e. and bias at the nth step for Process II. Let $a = K_0\beta$ and β_{kn}, $k = 0, 1, \cdots, n$ be defined by (3.1). Finally, let $\sigma^2(x) = \sigma^2(0)$, for every $x \in R$, and $T = \sigma^2(0)\gamma^{-2}$. Then

$$E_{\mathrm{II}}(X_{n+1}^2) = (1 - an^{-1})E_{\mathrm{II}}(X_n^2) + \beta^{-2}a^2n^{-2}(\sigma^2(0) + b_n^{-1}\gamma^2) + 2\beta^{-1}an^{-1}(1 - an^{-1})b_{n-1}b_n^{-1}d_n , \tag{5.1}$$

where

$$d_n = b_n b_{n-1}^{-1} E_{\mathrm{II}}[X_n(\bar{Z}_n - \alpha)] .$$

Now

$$d_n = b_{n-1}^{-1}b_{n-2}[1 - a(n-1)^{-1}]d_{n-1} + \beta^{-1}a(n-1)^{-1}\gamma^2 b_{n-1}^{-1} ;$$

$$d_n = \gamma^2\beta^{-1}b_{n-1}^{-1}\textstyle\sum_{i=1}^{n-1} ai^{-1}\beta_{i,n-1} . \tag{5.2}$$

Substituting (5.2) into (5.1), solving the difference equation (5.1), and standardizing the m.s.e.,

$$N\beta^2\gamma^{-2}E_{\mathrm{II}}(X_{n+1}^2) = N\beta^2\gamma^{-2}X_1^2\beta_{0n}^2 + (T + K_5^{-1})Na^2\beta_{1n}^2 + TNH(n) + K_5^{-1}NG(n) , \quad \text{if } K_5 \text{ is a positive integer;} \tag{5.3}$$

$$N\beta^2\gamma^{-2}E_{\mathrm{II}}(X_{n+1}^2) = N\beta^2\gamma^{-2}X_1^2\beta_{0n}^2 + (T + 1)Na^2\beta_{1n}^2 + TNH(n) + NQ(n) , \quad \text{if } K_5 \in (0, 1) ,$$

where

$$H(n) = \sum_{i=2}^{n} a^2 i^{-2} \beta_{in}^2 ,$$

$$G(n) = \sum_{i=2}^{n} [i^{-1} + 2a^{-1}(1 - ai^{-1}) \sum_{j=1}^{i-1} aj^{-1} \beta_{j,i-1}] a^2 i^{-2} \beta_{in}^2 ,$$

$$Q(n) = \sum_{i=2}^{n} [1 + \langle K_5 i \rangle]^{-1} [1 + 2ia^{-1}(1 - ai^{-1}) \sum_{k=1}^{i-1} ak^{-1} \beta_{k,i-1}] a^2 i^{-2} \beta_{in}^2 .$$

Similarly, it can be shown that

$$N\beta^2\gamma^{-2} E_{\mathrm{I}}(X_{n+1}^2) = N\beta^2\gamma^{-2} X_1^2 \beta_{0n}^2 + (T + K_5^{-1}) Na^2 \beta_{1n}^2 + TNH(n) + K_5^{-1} NH(n) , \quad \text{if } K_5 \text{ is a positive integer.} \tag{5.4}$$

Remember that Process I is well-defined only for $K_5 \geqq 1$.

Since $E_{\mathrm{I}}(X_{n+1}) = E_{\mathrm{II}}(X_{n+1}) = X_1 \beta_{0n}$, equations (5.3) and (5.4), in essence, express the mean-square error as the bias squared plus the variance. The squared-bias term decreases rapidly for increasing n if $2a - 1$ is not small; in fact, $n\beta_{0n}^2 = O(n^{1-2a})$ (Sacks [8], page 374). Hodges and Lehmann ([6] Table 1) list values of $n\beta_{0n}^2$ for $a = .2(.2).8$, 1.2 and $n = 5(5)30$.

If $N = n + \langle nK_5 \rangle$ and $a > \frac{1}{2}$ then

$$\lim_{n\to\infty} \beta^2\gamma^{-2} NE_{\mathrm{II}}(X_{n+1}^2) = (T + 2a^{-1}K_5^{-1}) a^2 (2a - 1)^{-1} (1 + K_5) , \qquad K_5 > 0 ;$$

$$\lim_{n\to\infty} \beta^2\gamma^{-2} NE_{\mathrm{I}}(X_{n+1}^2) = (T + K_5^{-1}) a^2 (2a - 1)^{-1} (1 + K_5) , \qquad K_5 \geqq 1 ;$$

and therefore, for each process, the m.s.e. approaches the asymptotic normal variance as $N \to \infty$.

For purposes of comparing Process I to Process II and comparing the small sample mean-square errors to their asymptotic values, it is helpful to plot the normalized m.s.e.'s of equations (5.3) and (5.4) as functions of $\log a$. Figure 1 provides typical examples of such plots.

For the asymptotic case, $N = \infty$, the graphs display some of the properties derived in Section 4. The asymptotic curves (see (c) and (f) of Figure 1) are U-shaped and are independent of the choice of X_1. When $K_5 \geqq 1$, the asymptotic curve corresponding to Process II is located above and to the right of the asymptotic curve for Process I and the curves intersect at $\log a = .301$ ($K_0 = 2\beta^{-1}$). The asymptotic curve for Process I always reaches its minimum at $\log a = 0$ ($K_0 = \beta^{-1}$).

For $\log a \geqq 0$, the small sample curves closely resemble the asymptotic curves. In Figure 1 compare (a) and (b) to (c) and compare (d) and (e) to (f).

One would not expect the small sample curves to resemble the asymptotic curves if $\log a < 0$ since the asymptotic curves go to ∞ as $\log a \downarrow -.301$ and are not defined for $\log a < -.301$. Figure 1 shows that the shape of the small sample curves for values of $\log a$ below 0 depend heavily on N and X_1. The influence of increasing N is illustrated by (a) and (b) of Figure 1; the influence of the initial guess X_1 is illustrated by (d) and (e) of Figure 1.

It seems generally true that in small sample situations, the curves will cross at a point to the left of $\log a = .301$.

It is appropriate to conclude that the asymptotic curve for $\log a > 0$ is a good

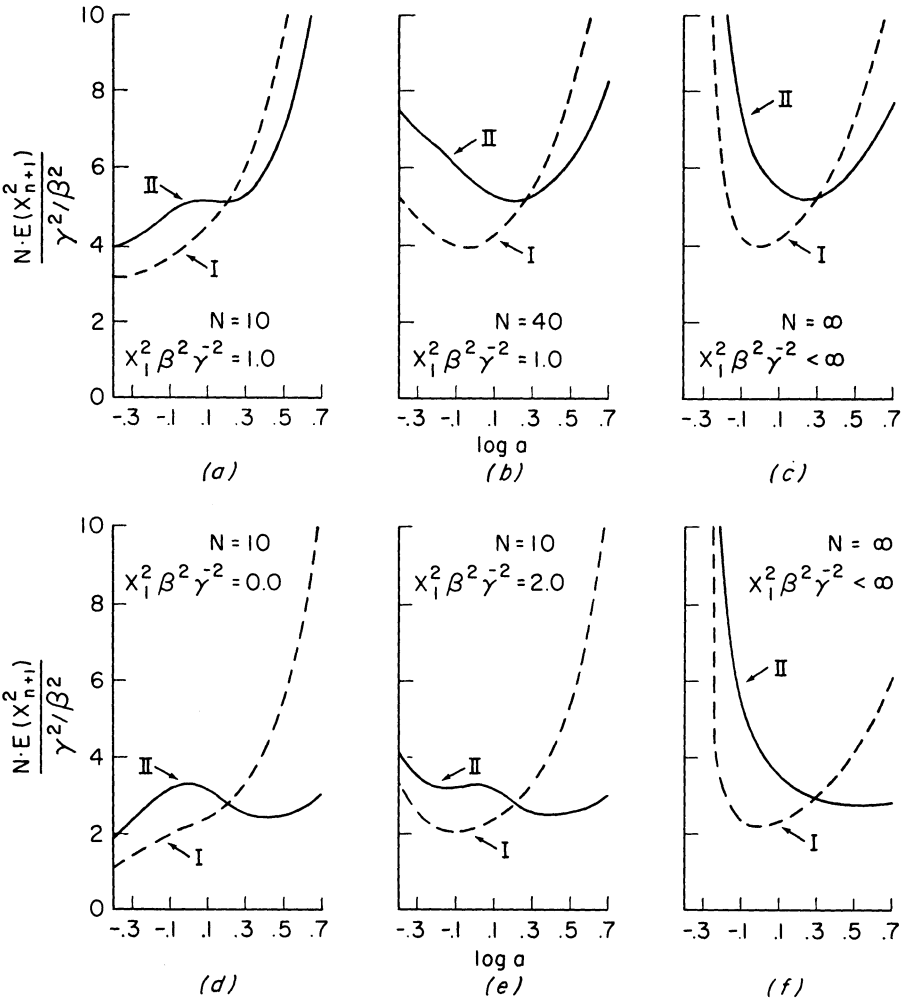

FIG. 1. Normalized mean-square errors (for linear $M(\cdot)$ whenever N is finite) plotted against $\log a$, $a = K_0\beta$, at $\theta = 0$; $K_5 = 1$; $T = 1.0$ in (a), (b), (c); and $T = 0.1$ in (d), (e), (f).

representation of the small sample situation; the only qualification being that Process II may be slightly better relative to Process I in the small sample case than is suggested by the asymptotic theory.

Acknowledgments. The author wishes to thank Rupert G. Miller, Jr. for suggesting this problem and for his many useful comments.

REFERENCES

[1] BLUM, J. R. (1954). Approx. methods which converge w.p. 1. *Ann. Math. Statist.* **25** 382–386.

[2] BURKHOLDER, D. L. (1956). On a class of stochastic approximation procedures. *Ann. Math. Statist.* **27** 1044–1059.

[3] DVORETZKY, A. (1956). On stochastic approximation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 39–55. Univ. of California Press.

[4] FIELLER, E. C. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B* **16** 175–185.

[5] HAMILTON, M. A. (1968). The Robbins-Monro stochastic approximation approach to a discrimination problem. Technical Report No. 8, Department of Statistics, Stanford Univ.

[6] HODGES, J. S. and LEHMANN, E. L. (1956). Two approximations to the Robbins-Monro process. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 95–104. Univ. of California Press.

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

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

[9] SCHMETTERER, L. (1961). Stochastic approximation. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 587–609. Univ. of California Press.

[10] VENTER, J.H. (1966). On Dvoretzky stochastic approximation theorems. *Ann. Math. Statist.* **37** 1534–1544.

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

[12] VON MISES, R. and POLLACZEK-GEIRINGER, H. (1929). Praktische Verfahren der Gleichungsauflosung. *Z. Angew, Math. Mech.* **9** 58–77.

Note added in proof. This paper is based on the author's doctoral dissertation submitted at Stanford University, April, 1968. Part of the paper was written while the author was employed by the National Cancer Institute.