SOME STOCHASTIC APPROXIMATION PROCEDURES FOR USE IN PROCESS CONTROL[1]

By John P. Comer, Jr.

Columbia University and the Procter & Gamble Co.

1. Introduction and summary. In this paper we shall consider mathematical models motivated by chemical process control problems in which we wish to control the process to hold some property of the output material as nearly constant as possible. For example, we might control the viscosity of the process output by varying the setting of a valve in the cooling water line to a heat exchanger. We shall suppose that the viscosity is observed at equally spaced times and that these observations form a sequence $Y_1, Y_2, Y_3, \cdots$. The control procedure specifies the valve setting X_1 to be used during the time interval (0, 1]. Then, immediately after observing Y_n, we reset the cooling water valve to the position X_{n+1} in accordance with the control procedure. The value of X_{n+1} may depend on the prior values $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$. In choosing X_{n+1} we hope to hold Y_{n+1} as close to a fixed value Y_0 as possible. More specifically, our objective will be to keep values of the loss $E(Y_n - Y_0)^2$ small.

In this paper it is assumed that time lags attributable to process dynamics are negligible when compared with the interval between observations. Parts of the theory are being generalized to allow for the presence of such time lags and no serious difficulties have as yet been encountered in this type of generalization.

The basic *process* and the general *control procedure* are described in mathematical terms in Sections 2 and 3. The concept of a *system* consisting of the process and the control procedure is then introduced in Section 4.

The specific type of control procedure we shall consider is a generalization of the simple proportional control procedure discussed by Box and Jenkins [1]. This procedure requires that we specify X_1 and then recursively specify $X_{n+1} = X_n - a_n(Y_n - Y_0)$. We may place bounds on X_n as in (8) below. The sequence $\{a_n\}$ is a sequence of positive numbers which may be specified *a priori* as in Section 5 or sequentially as in Section 9.

In Sections 6 and 7 we study the performance of certain processes when the sequence $\{a_n\}$ is specified *a priori*. Most important are cases where a_n converges to a small value a near zero. With the class of processes used in this paper, it turns out that this procedure for choosing X_n is quite similar to the stochastic approximation procedure first presented by Robbins and Monro in 1951 [7] and later generalized by Dvoretzky [5] and other authors. A comprehensive review of work on stochastic approximation is contained in a recent article by Schmetterer [8].

In Section 8 we specialize to the study of certain stationary processes. At-

Received 29 August 1963; revised 16 April 1964.

[1] This paper is based on the author's Ph.D. thesis at Columbia University.

tention is focused on the asymptotic performance of the controlled process as a function of the limit a of the sequence $\{a_n\}$ in our control procedure. Of particular interest is the asymptotic loss given by $\lim_{n\to\infty} E(Y_n - Y_0)^2$.

A procedure for the sequential choice of the sequence $\{a_n\}$ is proposed in Section 9 and its theoretical properties are studied in Section 10. Conditions are given under which both a_n and $E(Y_n - Y_0)^2$ converge. It turns out that this procedure for the sequential choice of a_n is also a stochastic approximation procedure of the Robbins-Monro type. Numerical examples given in Section 11 indicate that the asymptotic loss obtained by using this procedure for sequentially choosing a_n may be nearly as small as the asymptotic loss obtained when the sequence $\{a_n\}$ is chosen *a priori* to converge to the value of a that minimizes the asymptotic loss. However, when a_n is chosen sequentially, we need much less *a priori* knowledge of the process.

2. The process. The basic process, which we shall call a *process of type 1*, is defined in terms of the sequences $\{X_n\}$, $\{Y_n\}$, a function $M(x)$ measurable on the real line and a *stochastic sequence* $\{Z_n\}$. In this paper we shall assume that the process can be described mathematically as follows:

(a) The basic equation defining Y_n is

$$Y_n = M(X_n) + Z_n, \qquad n = 1, 2, \cdots. \tag{1}$$

(b) We shall assume that there exist two known bounds X_* and X^* such that

$$X_* \leqq X_n \leqq X^*, \qquad n = 1, 2, \cdots; \tag{2}$$

but we shall take special note of the effect of removing these bounds on X as we state each theorem.

(c) We shall also assume that there exist two fixed bounds β_* and β^* (not necessarily known) such that for any two distinct x, x' in the interval $[X_*, X^*]$ we have

$$0 < \beta_* \leqq (M(x) - M(x'))/(x - x') \leqq \beta^*. \tag{3}$$

(d) The distribution of $\{Z_n\}$ is described in terms of a sequence of conditional distribution functions (not usually known) defined as

$$F_n(z_n \mid z_1, \cdots, z_{n-1}, x_1, \cdots, x_n), \qquad n = 1, 2, \cdots. \tag{4}$$

The desired output of the process is Y_0 where Y_0 is known. The *loss* at any time n is given by the expectation $E(Y_n - Y_0)^2$. We assume that there exists θ in the interval $[X_*, X^*]$ such that

$$M(\theta) = Y_0. \tag{5}$$

3. The control procedure. Any control procedure must specify X_1 within the interval $[X_*, X^*]$ and must define a sequence of measurable functions $\{p_n\}$ which permit us to specify subsequent values of X_n in terms of prior values of X_n and

Y_n as

(6) $$X_n = p_n(X_1, \cdots, X_{n-1}, Y_1, \cdots, Y_{n-1}).$$

The general objective of control will be to reduce the loss $E(Y_n - Y_0)^2$.

4. The system. Once a process and a control procedure are defined, we have a system whose stochastic behavior is completely defined. By using (1), (4), (6) we can define the joint distributions of any finite number of elements of the sequences $\{X_n\}$, $\{Y_n\}$, or $\{Z_n\}$. In this paper we do not require detailed knowledge of these distributions. We need to define them only in order to be able to specify certain restrictions on them.

5. Control procedure of type P1. We now define the first type of control procedure which we shall call a *procedure of type P1*. First specify *a priori* a sequence of positive numbers $\{a_n\}$ such that

(7) $$\lim_{n\to\infty} a_n = a \geqq 0, \qquad \sum_{n=1}^{\infty} a_n = \infty.$$

Then specify a value of X_1 in the interval $[X_*, X^*]$ and proceed to define X_n sequentially as

(8) $$\begin{aligned} X_{n+1} &= X_* && \text{if } X_n - a_n(Y_n - Y_0) < X_* \\ &= X_n - a_n(Y_n - Y_0) && \text{if } X_* \leqq X_n - a_n(Y_n - Y_0) \leqq X^* \\ &= X^* && \text{if } X_n - a_n(Y_n - Y_0) > X^*. \end{aligned}$$

A procedure of type P1 is intuitively appealing because it permits us to decrease X_n when the observed Y_n is above Y_0 and vice versa. When a equals zero the loss always tends to EZ_n^2 and this may be the best choice of a for control of many processes. However, we shall see in the examples in Section 11 that when a is positive we can continue indefinitely to adjust X_n on the basis of recent history of the process and thus reduce $E(Y_n - Y_0)^2$ when the random errors (Z_n) in (1) are highly positively correlated.

Note that when $a = 0$, this procedure is like the conventional Robbins-Monro procedure except that we have omitted the usual requirement that $\sum_{n=1}^{\infty} a_n^2 < \infty$. We can do this because Condition (3) will prove to be an adequate substitute in our theorems.

A procedure of type P1 with $a > 0$ and no bounds on X_n is called proportional control in a recent article by Box and Jenkins [1].

6. Properties of procedures of type P1 when a is small. We shall next consider properties of the system when a is small. We shall require that a be such that

(9) $$0 \leqq a < 1/\beta^*.$$

We also require that the stochastic behavior of the system be such that the following condition holds:

CONDITION C1. There exists a sequence $\{\zeta_m\}$ of non-negative numbers such that

$$\zeta_m \to 0 \quad \text{as} \quad m \to \infty, \tag{10}$$

$$E\{[E(Z_n \mid Z_1, \cdots, Z_{n-m})]^2\} \leqq \zeta_m^2. \tag{11}$$

Some important properties of procedures of type P1 applied to processes of type 1 are summarized in the following theorem:

THEOREM 1. *Suppose we are given a process of type 1 and form a system by applying a control procedure of type P1. Suppose further that the control procedure satisfies* (9) *and that the system satisfies Condition* C1. *Then there exists an upper bound* $\eta(a)$ *such that*:

$$\limsup_{n\to\infty} E(X_n - \theta)^2 \leqq \eta(a), \tag{12}$$

$$\limsup_{n\to\infty} E(Y_n - Z_n)^2 \leqq (\beta^*)^2\eta(a), \tag{13}$$

$$\eta(a) \to 0 \quad \text{as} \quad a \to 0, \tag{14}$$

$$\eta(0) = 0, \tag{15}$$

and a possible version of $\eta(a)$ *is*

$$\eta(a) = \min_{m=0,1,\cdots}\left\{\frac{a\beta_* K^2}{2} + \frac{(\beta^*K + \zeta_0)\zeta_0 am + K\zeta_{m+1}}{\beta_*} + \frac{\beta^*K\zeta_0 a}{\beta_*} + \frac{a\zeta_0^2}{2\beta_*}\right\}, \tag{16}$$

where $K = X^* - X_*$.

It is shown in [2] that this theorem also holds when X_n is unbounded provided we take $K = \zeta_0/\beta_*$. This theorem is based on Theorem 4.1 of [2] as generalized in the appendix of [2]. When all conditions hold except (5), conclusions (12), (14) and (15) of Theorem 1 still hold provided we identify θ as X_* or X^* depending on whether $M(x) > Y_0$ or $M(x) < Y_0$ for all x in the interval $[X_*, X^*]$.

PROOF. To simplify notation we shall set $Y_0 = \theta = 0$ throughout the proof. By setting $m = 0$ in (11) we have

$$EZ_n^2 \leqq \zeta_0^2, \qquad n = 1, 2, \cdots. \tag{17}$$

From (1) and (3) we know that $|Y_n| \leqq \beta^* |X_n| + |Z_n|$ and hence

$$[EY_n^2]^{\frac{1}{2}} \leqq \beta^*[EX_n^2]^{\frac{1}{2}} + [EZ_n^2]^{\frac{1}{2}} \leqq \beta^*K + \zeta_0, \qquad n = 1, 2 \cdots. \tag{18}$$

We now define $\beta_n = M(X_n)/X_n$ so that from (3) we have $\beta_* \leqq \beta_n \leqq \beta^*$ and we can use (8) to obtain

$$|X_{n+1}| \leqq |X_n - a_nY_n| = |(1 - a_n\beta_n)X_n - a_nZ_n|, \quad n = 1, 2, \cdots. \tag{19}$$

Let n_0 be the smallest integer such that $a_n\beta^* < 1$ for all $n \geqq n_0$; the existence of n_0 is assured by (9). We then know that for any $n \geqq n_0$

$$0 < 1 - a_n\beta^* \leqq 1 - a_n\beta_n \leqq 1 - a_n\beta_* < 1. \tag{20}$$

From (17), (18), (19), (20) it follows that for all $n \geqq n_0$ we have

$$\begin{aligned} EX_{n+1}^2 &\leqq (1 - a_n\beta_*)^2 EX_n^2 - 2a_n E(1 - a_n\beta_n)X_nZ_n + a_n^2 EZ_n^2 \\ &\leqq EX_n^2 - 2a_n\beta_*\{EX_n^2 - a_n\beta_*K^2/2 - |EX_nZ_n|/\beta_* \\ &\qquad - |E\beta_nX_nZ_n|\, a_n/\beta_* - a_n\zeta_0^2/2\beta_*\}. \end{aligned} \tag{21}$$

From (8) and (18) we know that $[E(X_{n+1} - X_n)^2]^{\frac{1}{2}} \leqq a_n[EY_n^2]^{\frac{1}{2}} \leqq a_n(\beta^*K + \zeta_0)$ for all $n \geqq 1$. Thus for any $m \geqq 0$ and any $n \geqq m + n_0$ we can use these results, (11), and the fact that X_{n-m} is completely determined by X_1 and $Z_1, \cdots, Z_{n-m-1}$ to show that

$$\begin{aligned} |EX_nZ_n| &\leqq |E(X_n - X_{n-m})Z_n| + |EX_{n-m}Z_n| \\ &\leqq \sum_{i=n-m}^{n-1} |E(X_{i+1} - X_i)Z_n| + |E[X_{n-m}E(Z_n \mid Z_{n-m-1}, \cdots, Z_1)]| \\ &\leqq (\beta^*K + \zeta_0)\zeta_0 \sum_{i=n-m}^{n-1} a_i + K\zeta_{m+1}. \end{aligned} \tag{22}$$

We also note that $|E\beta_nX_nZ_n| \leqq [E\beta_n^2X_n^2EZ_n^2]^{\frac{1}{2}} \leqq \beta^*K\zeta_0$.

To establish (12) we must show that for any $\epsilon > 0$ there exists $n(\epsilon)$ such that for all $n \geqq n(\epsilon)$ we have

$$EX_n^2 \leqq \eta(a) + \epsilon. \tag{23}$$

If $a > 0$, let m be the smallest integer for which the minimum in (16) is attained. We can now choose $n_1(\epsilon) \geqq n_0$ such that for all $n \geqq n_1(\epsilon)$ (21) can be written as

$$EX_{n+1}^2 \leqq EX_n^2 - 2a_n\beta_*\{EX_n^2 - \eta(a) - \epsilon/2\}. \tag{24}$$

If $EX_n^2 < \eta(a) + \epsilon$ for any $n \geqq n_1(\epsilon)$, the inequality also holds for all larger n. If $EX_n^2 \geqq \eta(a) + \epsilon$ for any $n \geqq n_1(\epsilon)$, we know that $EX_{n+1}^2 \leqq EX_n^2 - a_n\beta_*\epsilon$. Because $EX_n^2 \leqq K^2$ at $n = n_1(\epsilon)$ and $\sum_{n=1}^{\infty} a_n = \infty$, we can always choose $n(\epsilon)$ so large that (23) holds for some n between $n_1(\epsilon)$ and $n(\epsilon)$ and hence (23) holds for all $n \geqq n(\epsilon)$ as required. If $a = 0$, $\eta(a) = 0$ and we choose $m(\epsilon)$ to make $K\zeta_{m+1}/\beta_* < \epsilon/4$ for all $m \geqq m(\epsilon)$. Then we can choose $n_1(\epsilon)$ and $n(\epsilon)$ as before. With the definition of $\eta(a)$ in (16) we have now established (12), (14), (15). To establish (13) we need only use (1) to show that

$$E(Y_n - Z_n)^2 = E[M(X_n)]^2 \leqq (\beta^*)^2 EX_n^2, \qquad n = 1, 2, \cdots. \tag{25}$$

7. Specialized conditions that imply Condition C1. Condition C1 basically requires that the random variables Z_n and Z_{n+m} become increasingly independent as m increases. The usual conditions attached to the Robbins-Monro procedure for finding the root θ of the equation $M(x) = Y_0$ when observations of $M(x)$ are subject to measuring error Z_n are that for all $n = 1, 2, \cdots$ we have

$$E[Z_n \mid Z_1, \cdots, Z_{n-1}, X_1, \cdots, X_n] = E(Z_n \mid X_n) = 0, \tag{26}$$

(27) $$E[Z_n^2 \mid Z_1, \cdots, Z_{n-1}, X_1, \cdots, X_n] = E(Z_n^2 \mid X_n) \leqq \zeta_0^2.$$

These two conditions imply Condition C1 with $\zeta_m = 0$ for $m \geqq 1$.

However, Conditions (26) and (27) are unrealistic in the process control context where Z_n and Z_{n+m} are quite likely to be correlated in some rather arbitrary way unless m is quite large. It may be more realistic in the process control context to start by restricting the conditional distributions in (4) so that for all $n = 1, 2, \cdots$ we have

(28) $$F_n(z_n \mid z_1, \cdots, z_{n-1}, x_1, \cdots, x_n) = F_n(z_n \mid z_1, \cdots, z_{n-1}).$$

Under this assumption the Z_n random variables can be correlated and we can impose Condition C1 on the process alone because the distribution of the sequence $\{Z_n\}$ is independent of the control procedure used.

If we further suppose that the sequence $\{Z_n\}$ is stationary and Gaussian, it is sufficient for Condition C1 to require this sequence to be regular and to have no deterministic component. Here we use the terminology of Doob [4].

8. Properties of procedures of type P1 when the process is stationary. In this paper we shall define a stationary process, which we shall call a *process of type 2*, to be a *process of type 1* as described above restricted to cases in which (28) and Condition C1 hold and in which the sequence $\{Z_n\}$ satisfies the following condition:

CONDITION C2. The Z_n random variables form a stationary sequence which can be decomposed so that $Z_n = Z_n^{(1)} + Z_n^{(2)}$ for all $n = 1, 2, \cdots$. We require that $EZ_n^{(1)} = EZ_n^{(2)} = 0$ and that the $Z_n^{(2)}$ random variables be independently and identically distributed and that $Z_n^{(2)}$ be independent of $Z_m^{(1)}$ for all $m \leqq n$. Furthermore the distribution function for $Z_n^{(2)}$ is to be absolutely continuous with a probability density function $f(z)$ such that $0 < f(z) \leqq f$ for all z and for some bound f. We note that Condition C2 implies that $EZ_n = 0$ and that Condition C1 implies that $EZ_n^2 \leqq \zeta_0^2$. Therefore the following covariance sequence $\{R(m)\}$ exists for any stationary process of type 2:

(29) $$R(m) = EZ_nZ_{n+m}, \qquad m = 0, 1, 2, \cdots.$$

Before we state a theorem covering the properties of stationary processes of type 2 we define the following sequence of random variables $\{V_n\}$ generated by any system.

(30) $$V_n = -\mathrm{SGN}[(Y_n - Y_0)(Y_{n+1} - Y_0)].$$

We shall need the sequence $\{V_n\}$ and its properties when we consider procedures of type P2 in Sections 9 and 10.

The important properties of procedures of type P1 applied to stationary processes of type 2 are summarized in the following theorem:

THEOREM 2. *Suppose we are given a stationary process of type 2 and form a system by applying a control procedure of type* P1. *Then the following limits exist and are continuous functions of a for all $a \geqq 0$.*

(31) $\lim_{n\to\infty} E(X_n - \theta) = \mu_x(a), \quad \lim_{n\to\infty} E(X_n - \theta)^2 = B(a),$

(32) $\lim_{n\to\infty} E(Y_n - Y_0) = \mu_y(a), \quad \lim_{n\to\infty} E(Y_n - Y_0)^2 = T(a),$

$\lim_{n\to\infty} E(Y_n - Y_0)(Y_{n+1} - Y_0) = D(a),$

(33) $\lim_{n\to\infty} EV_n = \mu(a).$

Furthermore, even if X_n *is unbounded, this theorem holds provided we require that* $a < 2/\beta^*$.

For simplicity we have placed restrictions on stationary processes of type 2 which are not necessary for all parts of Theorem 2. However, these conditions permit us to apply simultaneously several theorems in Sections 4, 5, 6, 7 of [2] to establish all parts of Theorem 2. Proofs for all these contributing theorems are included in the main reference [2].

Note that the limits established in Theorem 2 depend on the sequence $\{a_n\}$ only through the limit a and that $\lim_{a\to 0} B(a) = B(0) = 0$ by Theorem 1. The number $T(a)$ represents the asymptotic loss for this process when controlled by a procedure of type P1 with a specified value of a.

We can appraise the asymptotic performance of a procedure of type P1 applied to a particular stationary process of type 2 in terms of two different efficiencies. First suppose that we know the entire stochastic structure of the stationary process. We can then calculate the general least squares predictor $\hat{Z}_n = E(Z_n \mid Z_1, \cdots, Z_{n-1})$ and can choose X_n to make $\hat{Z}_n = -M(X_n)$. Suppose this procedure leads to an asymptotic loss $\hat{T}$. Now consider the asymptotic loss $T(a)$ under a procedure of type P1. Let the value of a which minimizes $T(a)$ be a_M and let this minimum loss be T_M. The asymptotic efficiency of procedures of type P1 relative to the ideal procedure can then be defined as $\mathcal{E}_M = \hat{T}/T_M$. This is the sacrifice we make to achieve simplicity in the control procedure.

In general we will not know the entire stochastic structure of the stationary process and will have to apply a procedure of type P1 with a value of a that differs from a_M. The asymptotic efficiency of the procedure with the actual choice of a relative to the optimum procedure of type P1 with $a = a_M$ can therefore be defined as $\mathcal{E}(a) = T_M/T(a)$.

In the main reference [2] other criteria that a good control procedure should satisfy are discussed. These include performance at small values of n and performance in the presence of various types of superimposed disturbances. It is also noted that procedures of type P1 are more likely to provide effective control if the Z_n random variables tend to be positively correlated.

9. Control procedures of type P2. When knowledge of the process is incomplete, it will be impossible to choose the optimum value of a under a procedure of type P1. In the main reference [2] the possibility of selecting a reasonably good value of a on the basis of partial knowledge of the process is discussed. However, it seems reasonable to hope that the sequence $\{a_n\}$ might be chosen stochastically in such a way as to converge to a reasonably good value of a dictated by the actual performance of the system.

We now define a procedure of type P2 to provide a method for the stochastic choice of $\{a_n\}$. First specify two bounds a_*, a^* such that $0 < a_* < a^*$. Also specify *a priori* a sequence of positive numbers $\{\lambda_n\}$ such that

$$\lim_{n\to\infty} \lambda_n = 0, \qquad \sum_{n=1}^{\infty} \lambda_n = \infty. \tag{34}$$

Then specify X_1 in the interval $[X_*, X^*]$ and a_1 in the interval $[a_*, a^*]$. Next define a_n sequentially as

$$\begin{aligned} a_{n+1} &= a_* && \text{if } a_n \exp(-\lambda_n V_n) < a_*, \\ &= a_n \exp(-\lambda_n V_n) && \text{if } a_* \leqq a_n \exp(-\lambda_n V_n) \leqq a^*, \\ &= a^* && \text{if } a_n \exp(-\lambda_n V_n) > a^*, \end{aligned} \tag{35}$$

where V_n was defined in (30). Once a_n has been calculated we can proceed to define X_{n+1} sequentially as in (8).

Procedures of type P2 are motivated by the fact that we probably need more correction (larger a_n) if successive values of $(Y_n - Y_0)$ have the same sign. A similar idea was employed by Kesten [6] to accelerate convergence of stochastic approximation procedures. There is a close analogy between a procedure of type P2 applied to the stochastic choice of a_n and a procedure of type P1 with $a = 0$ applied to the stochastic choice of X_n. The analogy between (8) and (35) is clearer if we transform (35) to obtain $\log a_{n+1} = \log a_n - \lambda_n V_n$. We expect that procedures of type P2 will force a_n to converge to the value of a at which $\lim_{n\to\infty} EV_n = \mu(a) = 0$. We shall define such a value of a to be a_0.

10. Properties of procedures of type P2 when the process is stationary. To timplify notation we shall restrict ourselves to a special case of the main theorems in [2] concerning procedures of type P2. We shall require that the process be of type 2 and that the sequence $\{Z_n\}$ satisfy the following condition:

CONDITION C3. The sequence $\{Z_n\}$ is a stationary Gaussian sequence which is regular and has no deterministic component (see Doob [4]). This means we can express Z_n in the form

$$Z_n = \sum_{i=0}^{\infty} c_i \xi_{n-i}, \qquad n = 1, 2, \cdots, \tag{36}$$

where $\{\xi_{n-i}\}$ is a sequence of independent $N(0, 1)$ random variables, $c_0 > 0$, $\sum_{i=0}^{\infty} |c_i|^2 = \zeta_0^2 < \infty$ and the equation $\sum_{i=0}^{\infty} c_i z^i = 0$ has no roots z_j such that $|z_j| < 1$. Using this notation, we then require that

$$\sum_{m=0}^{\infty} \left[\sum_{i=m}^{\infty} c_i^2 \right]^{\alpha} < \infty \quad \text{for some } \alpha \text{ such that } 0 < \alpha < 1. \tag{37}$$

Note that Condition C3 implies both Conditions C1 and C2.

In addition to Condition C3 we shall need some conditions on the function $\mu(a)$ analogous to Conditions (5) and (3) on $M(x)$. We therefore require that the process satisfy the following condition:

Condition C4. When procedures of type P1 are applied to generate $\mu(a)$, there exists a unique a_0 in the interval $[a_*, a^*]$ such that

$$\mu(a_0) = 0, \tag{38}$$

and there exist two bounds b_*, b^* such that for any a in $[a_*, a^*]$ that is distinct from a_0 we have

$$0 < b_* \leqq \mu(a)/(a - a_0) \leqq b^*. \tag{39}$$

The important properties of procedures of type P2 applied to our special class of stationary processes are summarized in the following theorem:

Theorem 3. *Suppose we are given a stationary process of type* 2 *and form a system by applying a control procedure of type* P2. *Suppose further that the process satisfies Conditions* C3 *and* C4. *It then follows that*

$$\lim_{n\to\infty} E(a_n - a_0)^2 = 0, \tag{40}$$

$$\lim_{n\to\infty} E(Y_n - Y_0)^2 = T(a_0), \tag{41}$$

where $T(a_0)$ is the asymptotic loss that would result if we applied a procedure of type P1 *with $a = a_0$ to the same process. Furthermore, even if X_n is unbounded, this theorem holds provided $a^* < 2/\beta^*$.*

Proof of a generalized version of Theorem 3 is included in Section 10 of [2]. A procedure of type P2 can be viewed as an application of the Robbins-Monro procedure to achieve convergence of a_n to a_0. The effectiveness of procedures of type P2 can be appraised in terms of the efficiency $\mathcal{E}(a_0) = T_M/T(a_0)$.

11. Numerical examples. We shall illustrate the performance of procedures of type P1 and P2 by applying them to a simple class of stationary processes. The processes will all be such that $M(x) = x$ and such that the value of X_n is unbounded. We note here that recent studies in which more complex systems have been simulated on a computer indicate that moderate departures from linearity in $M(x)$ do not seriously affect the performance of procedures of types P1 or P2.

All the conditions for Theorem 3 will be satisfied in the examples and we shall take $c_i = c^i(1 - c^2)^{1/2}$ in (36) with $0 < c < 1$. This means that in (29) we have $R(m) = c^m$. Using the notation of Theorem 2, we can say that $\mu_x(a) = \mu_y(a) = 0$; and for any given c we can calculate $T(a)$, $D(a)$, and $\mu(a)$ as functions of a. In particular we know that asymptotically the sequence $\{Y_n\}$ is Gaussian and stationary. Therefore, $\mu(a) = 0$ if and only if $D(a) = 0$ and we can calculate a_0 as the value of a at which $D(a) = 0$. $T(a)$ and $D(a)$ can be expressed concisely in terms of $r = (1 - a)$ for $-1 < r \leqq 1$ as

$$T(a) = 2(1 - c)/(1 + r)(1 - rc), \tag{42}$$

$$D(a) = (1 - c)[r(1 + c) - (1 - c)]/(1 + r)(1 - rc). \tag{43}$$

Furthermore we have

$$\begin{aligned} T_M &= 1 && \text{if } 0 < c \leqq \tfrac{1}{3}, \\ &= 8c(1 - c)/(1 + c)^2 && \text{if } \tfrac{1}{3} \leqq c < 1, \end{aligned} \tag{44}$$

$$(45) \qquad a_M = 0 \qquad \text{if } 0 < c \leqq \tfrac{1}{3},$$
$$= (3c - 1)/2c \qquad \text{if } \tfrac{1}{3} \leqq c < 1,$$

$$(46) \qquad T(a_0) = (1 - c)(1 + c)^2/(1 + c^2),$$

$$(47) \qquad a_0 = 2c/(1 + c).$$

We also know that the general least squares predictor in this case has an asymptotic loss $\hat{T}$ given by

$$(48) \qquad \hat{T} = 1 - c^2.$$

The derivation of these and other illustrative examples is covered in the main reference [2]. The algebraic derivation is similar to that presented by Cox [3] in an article on exponentially weighted moving average predictors. The connection between the illustrative examples with $M(x) = x$ and the prediction problem requires that we identify $(-X_n)$ with the predicted value $\hat{Z}_n$ of Z_n so that Y_n in (1) becomes the prediction error $Y_n = Z_n - (-X_n) = Z_n - \hat{Z}_n$.

To illustrate the performance of procedures of types P1 and P2 for different levels of positive correlation among the Z_n random variables, some results have been calculated and presented in Table 1. The smallest value of c is taken to be an arbitrarily small positive value (0^+) in order to represent the case where the Z_n random variables are independent. The more general version of Theorem 3 in the main reference [2] permits the use of $c = 0$. Note that in our example $a = .5$ is good for a wide range of values of c. If the slope of $M(x)$ is β this corresponds to a choice of $a = .5/\beta$. However, $M(x)$ must be linear with known slope β before we can take advantage of this choice of a. Table 1 illustrates the fact that the Robbins-Monro version of a procedure of type P1 (i.e. with $a = 0$) is minimax in the sense that it minimizes the maximum loss when we know nothing of the

TABLE 1

Asymptotic loss and efficiency for various procedures

Procedure		c					
		0^+	.1	.3	.5	.7	.9
1. Gen. least sq.	$\hat{T}$	1.00	.99	.91	.75	.51	.19
2. P1 with $a = a_M$:	a_M	0	0	0	.50	.79	.94
	T_M	1.00	1.00	1.00	.88	.58	.20
	$\mathcal{E}_M$	1.00	.99	.91	.85	.88	.95
3. P1 with $a = 0$	$T(0)$	1.00	1.00	1.00	1.00	1.00	1.00
	$\mathcal{E}(0)$	1.00	1.00	1.00	.88	.58	.20
4. P1 with $a = .5$	$T(.5)$	1.33	1.26	1.10	.88	.62	.24
	$\mathcal{E}(.5)$	.75	.79	.91	1.00	.93	.83
5. P2	a_0	0	.18	.46	.67	.82	.95
	$T(a_0)$	1.00	1.08	1.09	.90	.58	.20
	$\mathcal{E}(a_0)$	1.00	.93	.92	.98	1.00	1.00

correlation structure of $\{Z_n\}$ *a priori*. Any other choice of a will result in a higher loss when c is close enough to zero. Also $a = 0$ seems to be the best choice of a for control of processes that do not have a rather highly positively correlated Z_n sequence. However, it is also pointed out in [2] that a small value of a may be a poor choice if the process is not truly stationary. When the correlation structure of the Z_n sequence is unknown, procedures of type P2 seem to lead to an asymptotic loss $T(a_0)$ that compares quite well with T_M for all values of c and is usually even better than the choice of $a = .5$.

13. Acknowledgments. I wish to express my gratitude to Professor C. L. Mallows, who was my advisor while I was working on this dissertation, for the help and encouragement I received from him.

REFERENCES

[1] Box, G. E. P. and Jenkins, G. M. (1962). Some statistical aspects of adaptive optimization and control. *J. Roy. Statist. Soc. Ser. B* **24** 297–343.
[2] Comer, J. P. (1962). Some stochastic approximation procedures for use in process control. Doctoral dissertation, Columbia Univ.
[3] Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. *J. Royal Statist. Soc. Ser. B* **23** 414–422.
[4] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.
[5] Dvoretzky, A. (1956). On stochastic approximation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 39–55. Univ. of California Press.
[6] Kesten, H. (1958). Accelerated stochastic approximation. *Ann. Math. Statist.* **29** 41–59.
[7] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.
[8] Schmetterer, L. (1961). Stochastic approximation. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 587–609. Univ. of California Press.