STOCHASTIC APPROXIMATION FOR FUNCTIONALS

D.L. Hanson[1]

State University of New York at Binghamton

and

Ralph P. Russo[2]

University of Iowa

Suppose $\underline{F}$ is a class of distributions containing the discrete distributions and the distribution F_x for each real x. Suppose ϕ is a real valued functional on $\underline{F}$ and define $\theta(x) = \phi(F_x)$ so that $\theta(.)$ is a parameter of the family $\{F_x\}$. Fix α. A stochastic approximation procedure for finding the x for which $\theta(x) = \alpha$ is presented. When $\phi(F)$ is the mean of F, a form of this procedure is just the Robbins-Monro process. When $\phi(F)$ is the p-th quantile of F, a form of this procedure is just the quantile process introduced by the authors in an earlier paper. Some convergence theorems, examples, and generalizations are presented.

1. Introduction.

Suppose that for each real x (or for each x in some interval) there is a distribution F_x from which we can sample at will. Suppose $\underline{F}$ is a collection of distribution functions containing all empirical distribution functions (i.e., all distribution functions of the form $F(t) = \frac{1}{n}\sum_{k=1}^{n} I_{[a_k,\infty)}(t)$) and all of the distribution functions F_x. Let ϕ be a real valued functional on $\underline{F}$ and define $\theta(x) = \phi(F_x)$ so that $\theta(.)$ is some parameter of the family $\{F_x\}$. Our objective

[1] Research supported by National Science Foundation Grants MCS-8100740 and DMS-8402584.

[2] Research supported by National Science Foundation Grant MCS-8100740 and a grant from the Old Gold Foundation of the University of Iowa.
AMS 1980 subject classification. 62L20
Key words and phrases. Stochastic approximation, location parameter.

is to "find" the x, call it x_0, for which $\theta(x)$ takes on some specified value α.

We will approximate x_0 sequentially as follows:

(1.1) For some $n_0 \geq 1$, fixed or random, we observe y_k at x_k (i.e., sample y_k from the distribution F_{x_k}) for $k=1,\ldots,n_0$ where $x_1,\ldots,x_{n_0}$ are fixed or random and obtained in some arbitrary way.

We let R be a fixed positive real number and for $n \geq n_0$ we proceed as follows:

(1.2) We project the points (x_k,y_k) for $k=1,\ldots,n$ onto the line $x = x_n$ along lines having slope R to obtain points (x_n,y_{nk}) where

(1.2a) $$y_{nk} = y_k - R(x_k - x_n).$$

(1.3) Let F_n (not to be confused with F_x for $x = n$) be the "empirical" distribution function assigning probability $1/n$ to y_{nk} for $k=1,\ldots,n$. Project the point $(x_n,\phi(F_n))$ onto the line $y = \alpha$ along a line having slope R to obtain (x_{n+1},α) so that

(1.3a) $$x_{n+1} = x_n - [\phi(F_n) - \alpha]/R.$$

(1.4) Observe y_{n+1} at x_{n+1} and iterate.

Call this the ϕ-process.

The intuition for this procedure is that if a) the quantile curves for the family F_x are nondecreasing in x, and b) $F(t) \leq G(t)$ for all t implies $\phi(F) \geq \phi(G)$, then repeated observations from distributions F_x with x's to the right of x_0 should produce an "empirical" F_n which weights large values too heavily, $\phi(F_n)$ should be too big (i.e., bigger than α), and x_{n+1} should be less than x_n. Conversely, lots of sampling from distributions F_x with x's to the left of x_0 should cause x_{n+1} to be to the right of x_n.

Suppose $\phi(F) = \int x dF(x)$ so that $\theta(x)$ is just the mean of F_x. In the Robbins-Monro procedure (Robbins and Monro (1951)) we obtain x_1 in some arbitrary fashion, and for $n \geq 1$ we observe y_n at x_n and define

$x_{n+1} = x_n - \frac{a}{n}(y_n - \alpha)$. In the ϕ-process we have $\phi(F_n) = \frac{1}{n}\sum_{k=1}^{n}[y_k - R(x_k - x_n)]$ so that $x_{n+1} = x_n - \{\frac{1}{n}\sum_{k=1}^{n}[y_k - R(x_k - x_n)] - \alpha\}/R = \frac{1}{n}\sum_{k=1}^{n}[x_k - (y_k - \alpha)/R]$. This is true for all n so $x_{n+1} = \frac{n-1}{n}x_n + \frac{1}{n}[x_n - (y_n - \alpha)/R] = x_n - \frac{R^{-1}}{n}(y_n - \alpha)$. Thus if we let $R = 1/a$, the ϕ-process reduces to the Robbins-Monro process.

If $0 < p < 1$ and $\phi(F)$ is the right endpoint of the interval of p-th quantiles of F, then $\phi(F_n)$ is the right endpoint of the interval of p-th quantiles of F_n. If $y_{nk} > \phi(F_n)$ $(y_{nk} < \phi(F_n))$, then the projection of (x_n, y_{nk}) onto the line $y = \alpha$ along a line with slope R has x-coordinate $s_k = x_n - (y_{nk} - \alpha)/R = x_k - (y_k - \alpha)/R$ which is less than (greater than) the x-coordinate $x_{n+1} = x_n - [\phi(F_n) - \alpha]/R$ of the corresponding projection onto $y = \alpha$ of $(x_n, \phi(F_n))$. It is easily seen that x_{n+1} is the left endpoint of the interval of (1-p)-th quantiles of the "empirical" distribution function $\frac{1}{n}\sum_{k=1}^{n} I_{[s_k,\infty)}$. Thus, in this case, the ϕ-process reduces to the quantile process presented in Hanson and Russo (1981).

The purpose of this paper is further study of the ϕ-process. We restrict our attention to functionals ϕ which are location parameters, i.e., to functionals ϕ which satisfy $\phi(F(x - C)) = \phi(F(x)) + C$. (See Lemma 2.1.)

In Section 2 we consider the case (i.i.d. errors) where the distributions differ only by a location parameter. (The location parameter is $\phi(F_x)$.) We present an almost sure convergence theorem, a corollary showing some of the generality of the ϕ-process, and give an example. In Section 3 we present an almost sure convergence theorem for the case where the errors are not i.i.d. We present a variation and generalization of the ϕ-process in Section 4. We give an almost sure convergence theorem for this variation and give an example showing the versatility of this process. In the authors' opinion, this is the most interesting section of the paper. The model is fairly complicated, requiring lots of notation, and for that reason is not given here in the introduction. Section 5 contains some remarks and queries.

Because of a shortage of space, and because we feel that the assumptions made in each of the three theorems are stronger than will ultimately be necessary, we present only the proof of the second theorem in this paper.

2. The i.i.d. case.

For notational convenience we will use $\phi(X)$ to mean $\phi(F_X)$, and $\phi(X_1,\ldots,X_n)$ or $\phi(\{X_i : 1 \leq i \leq N\})$ to mean $\phi(\frac{1}{N}\sum_{k=1}^{N} I_{[X_k,\infty)})$. By y_x we shall mean some random variable having distribution F_x. We use the notation "$X \geq Y$ st." to mean "X is stochastically greater than or equal to Y" which means that $P\{X \geq t\} \geq P\{Y \geq t\}$ for all t.

Let α and R be fixed real numbers with $R > 0$, and let $\underline{F}$ be the collection of distribution functions as defined in Section 1. Let $\underline{C}$ be the class of all real valued functionals ϕ on $\underline{F}$ such that

(2.1) there is a unique real number x_ϕ such that $\phi(F_{x_\phi}) = \alpha$;

(2.2) there is a δ in $(0,R)$ such that $0 < [\phi(F_x) - \alpha]/[x - x_\phi] \leq 2R - \delta$ for all $x \neq x_\phi$;

(2.3) $\inf_{|x-x_\phi|\geq t} |\phi(F_x) - \alpha| > 0$ for all $t > 0$;

(2.4) if $X \geq Y + C$ st. then $\phi(F_X) \geq \phi(F_Y) + C$; and

(2.5) if k is a nonnegative integer, M and x are real numbers, $\{Y_i\}$ is an i.i.d. sequence of random variables all having distribution function F_x, and $G_n(t) = \frac{1}{n}\sum_{k=1}^{n} I_{[Y_k,\infty)}(t)$ is the "empirical" distribution function of $Y_1,\ldots,Y_n$, then

(2.5a) $$\phi(\tfrac{k}{k+n} I_{[M,\infty)} + \tfrac{n}{k+n} G_n\} \to \phi(F_x) \text{ a.s. as } n \to \infty.$$

LEMMA 2.1. If ϕ is in $\underline{C}$ and $X = Y + C$ st. then $\phi(X) = \phi(Y) + C$.

Proof. An immediate consequence of (2.4).

Our notational usage allows the alternative $\phi(F_X) = \phi(F_Y) + C$ as a conclusion to Lemma 2.1. Lemma 2.1 says that ϕ is a location parameter.

THEOREM 2.1. If ϕ is in $\underline{C}$; the random variables $y_x - \phi(F_x)$ all have distribution function G; n_0 is a positive integer valued random variable and $x_1,\ldots,x_{n_0}$ are arbitrary real valued "random variables"; and x_n for $n > n_0$ is the ϕ-process as defined by (1.1) - (1.3a), then $x_n \to x_\phi$ a.s.

LEMMA 2.2. $\underline{C}$ is closed under finite maxima and finite minima.

COROLLARY TO THEOREM 2.1. Suppose that $I_1,\ldots,I_N$ are finite non-empty subsets of $\underline{C}$ and that $\phi = \max_{1\le k\le N} \min_{\phi^*\varepsilon I_k} \phi^*$ (or $\phi = \min_{1\le k\le N} \max_{\phi^*\varepsilon I_k} \phi^*$). Then $x_\phi = \min_{1\le k\le N} \max_{\phi^*\varepsilon I_k} x_{\phi^*}$ (or $x_\phi = \max_{1\le k\le N} \min_{\phi^*\varepsilon I_k} x_{\phi^*}$) and Theorem 2.1 applies to ϕ.

For $0 < p < 1$ let $m_p(x)$ be the midpoint (or left endpoint or ...) of the interval of p-th quantiles of F_x.

<u>Example 2.1</u>. A drug is used to elevate (lower) blood pressure. Clearly, too great an elevation (lowering) of the blood pressure is potentially dangerous to the patient. In such a situation it is reasonable to assume that the quantile curves $m_p(x)$ are strictly increasing over reasonable dosage levels of the drug, and that the distributions F_x are continuous. Suppose you would like to find a dosage level x_0 such that at least 90% of all patients have a response of at least α (increase in blood pressure, percentage increase in blood pressure or ...) to a dosage level x_0 but, because it might be hazardous to their health, you would like no more than 5% of all patients to have a response of β or higher to the dosage. This last condition might be considered to be a "safety constraint". Let $\phi_1(F_x) = m_{0.1}(x) - \alpha$, $\phi_2(F_x) = m_{0.95}(x) - \beta$, and $\phi = \max\{\phi_1,\phi_2\}$. Let a be such that $m_{0.1}(a) = \alpha$ and b be such that $m_{0.95}(b) = \beta$. Let x_ϕ be such that $\phi(F_{x_\phi}) = 0$. The ϕ-process will sequentially converge to $x_\phi = \min\{a,b\}$ which will be such that either: i) $x_\phi = a$, 90% of the population has a response of α or more when given dosage level x_ϕ of the drug, and no more than 5% of the population has a response of β or more to dosage level x_ϕ of the drug; or ii) $x_\phi = b < a$, (only) 5% of the

population has a response of β or more to dosage level x_ϕ of the drug, less than 90% of the population has response level α or higher to dosage level x_ϕ, but the highest dosage possible is being given to patients without violating the safety constraint.

The same logic would apply to fireworks or industrial explosives where one would want to guarantee, if possible, a high probability of having at least a certain fixed explosive power, but would want to guarantee that there is only a very small probability that the explosive power would exceed some danger threshold.

Proof of Theorem 2.1. Omitted. (See Hanson and Russo (1985)).

Proof of Lemma 2.2. The proof just involves showing that the maximum or minimum satisfies (2.1) through (2.5) if the individual functions do.

Proof of Corollary to Theorem 2.1. Suppose I is a finite subset of $\underline{C}$. It is easy to see that if $\phi = \min_{\phi^* \varepsilon I} \phi^*$ then $x_\phi = \max_{\phi^* \varepsilon I} x_{\phi^*}$, and if $\phi = \max_{\phi^* \varepsilon I} \phi^*$ then $x_\phi = \min_{\phi^* \varepsilon I} x_{\phi^*}$. The corollary follows from these facts, Lemma 2.2, and Theorem 2.1.

3. The non-i.i.d. case.

THEOREM 3.1. Let R and δ be real numbers such that $0 < \delta < R$. Suppose $\underline{F}$ is a class of distribution functions containing all empirical distribution functions. Suppose $F_x \varepsilon \underline{F}$ for each real x and that

(3.1a) $$F_y(t) \leq F_x(t)$$

and

(3.1b) $$F_x(t) \leq F_y(t + (2R - \delta)(y - x))$$

for all real x, y and t such that $x \leq y$.

Suppose ϕ is a real valued functional on $\underline{F}$ which satisfies (2.1), (2.4) and (2.5). Let n_0 be a positive integer valued random variable; let $x_1,\ldots,x_{n_0}$ be arbitrary real "random variables"; and for $n > n_0$ let x_n be the ϕ-process as defined by (1.1) through (1.3a). Then $x_n \to x_\phi$ a.s.

Proof. To facilitate the proof we set up a model of our process on a different probability space. The space and random variables n_o, x_k, y_k, u_k, and y_k^a for all real a and positive integers k can be constructed so that:

(3.2) for each k the random variable u_k is uniformly distributed on (0,1) and is independent of $\{n_0;\ x_1,\ldots,x_k;\ u_1,\ldots,u_{k-1}\}$;

(3.3) $y_k^a = F_a^{-1}(u_k) = \min\{t:F_a(t) \geq u_k\}$ for all k, a, and ω;

(3.4) $y_k = y_k^{x_k}$ for all k and ω; and

(3.5) for all $n \geq n_0$ and all ω we have x_{n+1} defined by (1.1) through (1.3a).

The actual construction of our probability space is standard and omitted.

We will prove that

(3.6) $$P\{\limsup |x_n - x_\phi| = +\infty\} = 0,$$

and then that

(3.7) $$P\{\limsup x_n > x_\phi \text{ or } \liminf x_n < x_\phi\} = 0.$$

We assume that $x_\phi = \alpha = 0$.

Proof of (3.6). Let $a > 0$ be fixed and let A be the set of ω's such that $\phi(\{y_k^a(\omega):1 \leq k \leq n\}) \to \phi(F_a)$ and such that $\phi(\{y_k^{-a}(\omega):1 \leq k \leq n\}) \to \phi(F_{-a})$. From (2.5) we see that $P(A) = 1$.

Suppose $\omega \in A$. From (2.1), (2.4) and (3.1a) we have

$\phi(F_{-a}) < \alpha = 0 < \phi(F_a)$. Choose $n_1 = n_1(\omega) > n_0(\omega)$ such that if $n \geq n_1$ then

$$\phi(\{y_k^a: 1 \leq k \leq n\}) > 0 \text{ and } \phi(\{y_k^{-a}: 1 \leq k \leq n\}) < 0. \tag{3.8}$$

Let $M = M(\omega) = \max\{|x_1|, \ldots, |x_{n_1}|, a(2R - \delta)/\delta\}$. Clearly $|x_k| \leq M$ for $k=1,\ldots,n_1$. Suppose $|x_k| \leq M$ for $k=1,\ldots,n$. Because of (3.1b), if $\alpha \leq \beta$ we have for all ω, k and a

$$y_k^\alpha(\omega) \leq y_k^\beta \leq y_k^\alpha + (2R - \delta)(\beta - \alpha). \tag{3.9}$$

Thus

$$x_{n+1} = x_n - R^{-1}\phi(\{y_{nk}: 1 \leq k \leq n\}) = - R^{-1}\phi(\{y_k - Rx_k: 1 \leq k \leq n\})$$

$$\leq -R^{-1}\phi(\{y_k^a - (2R - \delta)(a - x_k) - Rx_k: 1 \leq k \leq n \text{ and } x_k \leq a\}$$

$$\cup\{y_k^a - Rx_k: 1 \leq k \leq n \text{ and } a < x_k\})$$

$$\leq -R^{-1}\phi(\{y_k^a - (R - \delta)M - a(2R - \delta): 1 \leq k \leq n \text{ and } x_k \leq a\}$$

$$\cup\{y_k^a - RM: 1 \leq k \leq n \text{ and } a < x_k\})$$

$$=M - R^{-1}\phi(\{y_k^a + M\delta - a(2R - \delta): 1 \leq k \leq n \text{ and } x_k \leq a\} \cup \{y_k^a: 1 \leq k \leq n \text{ and } a < x_k\})$$

$$\leq M - R^{-1}\phi(\{y_k^a: 1 \leq k \leq n\}) \leq M.$$

A similar argument gives $x_{n+1} \geq -M$ and an application of mathematical induction completes the proof of (3.6).

Proof of (3.7). Let $a > 0$ be arbitrary, let $L = L(\omega) = \limsup_{k\to\infty} |x_k(\omega)|$, and let

$D_a = \{\omega : a(2R - \delta)/\delta \leq L(\omega) < \infty\}$. Let $\varepsilon = \inf_{a \leq |x|} |\phi(F_x)|$. It follows from (2.1), (2.4) and (3.1a) that (2.3) holds. Thus $\varepsilon > 0$ and $\varepsilon = \min\{\phi(F_a), -\phi(F_{-a})\}$. Choose ε' so that $0 < \varepsilon' < \varepsilon/(2R)$. For $\omega \in D_a$ let

$$\nu(\omega) = \min\{\nu \geq 1: |x_k| \leq L + \varepsilon' \text{ for all } k > \nu\}.$$

Let

$$K_M = D_a \cap \{|y_k| + R(|x_k| + L + \varepsilon') \leq M \text{ for } k=1,\ldots,\nu\};$$

note that K_M is increasing in M and that $\bigcup_M K_M = D_a$. Define

$$J = \bigcap_{M=1}^{\infty} \bigcap_{k=0}^{\infty} [\{\omega : \phi(\{\underbrace{-M,\ldots,-M,}_{k \text{ times}}\ y^a_{k+1},\ldots,y^a_{k+n}\}) \to \phi(F_a)\}$$

$$\cap \{\omega : \phi(\{\underbrace{M,\ldots,M,}_{k \text{ times}} y^{-a}_{k+1},\ldots,y^{-a}_{k+n}\}) \to \phi(F_{-a})\}]$$

and, as before, note that (2.5a) implies $P(J) = 1$. We will show that JD_a is empty so that $P(D_a) = 0$. Since $\{\omega : 0 < L(\omega) < \infty\} = \bigcup_{n=1}^{\infty} D_{1/n}$, the proof that $P(D_a) = 0$ will complete the proof of the theorem.

Choose $\omega \in JD_a$, M so that $\omega \in K_M$, and N so that $n \geq N$ implies

$$\phi(\{\underbrace{-M,\ldots,-M,}_{\nu \text{ times}}\ y^a_{\nu+1},\ldots,y^a_{\nu+n}\}) \geq 3\varepsilon/4 \tag{3.10}$$

and

$$\phi(\{\underbrace{M,\ldots,M,}_{\nu \text{ times}}\ y^{-a}_{\nu+1},\ldots,y^{-a}_{\nu+n}\}) \leq -3\varepsilon/4. \tag{3.11}$$

Then for $n \geq N$, using (3.9) and the definitions of ν and D_a,

$$x_{\nu+n+1} = -R^{-1}\phi(\{y_k - Rx_k : 1 \leq k \leq \nu + n\})$$

$$\leq -R^{-1}\phi(\{y_k - Rx_k : 1 \leq k \leq \nu\} \cup \{y^a_{\nu+k} + (R-\delta)x_{\nu+k} - (2R-\delta)a : 1 \leq k \leq n \text{ and } x_{\nu+k} \leq a\}$$

$$\cup \{y^a_{\nu+k} - Rx_{\nu-k} : 1 \leq k \leq n \text{ and } x_{\nu+k} > a\})$$

$$\leq -R^{-1}\phi(\{y_k - Rx_k : 1 \leq k \leq \nu\} \cup \{y^a_{\nu+k} - R(L + \varepsilon') : 1 \leq k \leq n\})$$

$$\leq L + \varepsilon' - R^{-1}\phi(\{\underbrace{-M, \ldots, -M}_{\nu \text{ times}}, y^a_{\nu+1}, \ldots, y^a_{\nu+n}\})$$

$$\leq L + \varepsilon' - R^{-1}(3\varepsilon/4) < L - \varepsilon/(4R).$$

Similarly $x_{\nu+n+1} > -L + \varepsilon/(4R)$ so that $0 < L(\omega) \leq L(\omega) - \varepsilon/(4R)$ giving a contradiction. Thus JD_a is empty.

REMARKS. The same corollary applies to this theorem that applied to Therorem 2.1. Condition (3.1b) need only hold for $x \leq x_\phi \leq y$, but we stated the theorem as we did because x_ϕ is unknown and because weakening the condition complicates the proof slightly.

4. A generalization.

Suppose that for each real x (or for each x in some interval) there is a multivariate distribution $F_x(t_1, \ldots, t_M)$ from which we can sample at will. We will denote the marginals of F_x by $F^{(1)}_x, \ldots, F^{(M)}_x$. For each $i=1,\ldots,M$ let $\underline{F^{(i)}}$ be a class of one-dimensional distribution functions containing the distributions $F^{(i)}_x$ for all x and containing all "empirical distribution functions", and suppose $\phi^{(i)}$ is a real valued functional on $\underline{F^{(i)}}$. Suppose

(4.1) $f(a_1, \ldots, a_M)$ is a real valued function of M real variables such that

(4.1a) f is nondecreasing in each variable and

(4.1b) $f(a_1 + c, \ldots, a_M + c) = f(a_1, \ldots, a_M) + c$ for all real $a_1, \ldots, a_M, c$.

Suppose α is a fixed real number and that

(4.2) there is a unique x (call it x_0) such that

$$f[\phi^{(1)}(F^{(1)}_x), \ldots, \phi^{(M)}(F^{(M)}_x)] = \alpha \text{ when } x = x_0.$$

Suppose A is a fixed real number and we "know" that

$$(4.3) \qquad A < x_0.$$

As before, we wish to estimate x_0 sequentially. We do so as follows:

(4.4) For some $n_0 \geq 1$, fixed or random, we observe $y_k = (y_k^{(1)}, \ldots, y_k^{(M)})$ at x_k (i.e., we sample y_k from the distribution F_{x_k}) for $k=1,\ldots,n_0$ where $x_1,\ldots,x_{n_0}$ are fixed or random and obtained in some arbitrary way.

Let R be a fixed real number. For $n \geq n_0$ we proceed as follows:

(4.5) For each $i=1,\ldots,M$ we project the points $(x_k, y_k^{(i)})$ for $k=1,\ldots,n$ onto the line $x = x_n$ along lines having slope R to obtain points $(x_n, y_{nk}^{(i)})$ where

$$(4.5a) \qquad y_{nk}^{(i)} = y_k^{(i)} - R(x_k - x_n).$$

(4.6) Let $F_n^{(i)}$ be the "empirical" distribution function $\frac{1}{n}\sum_{k=1}^{n} I_{[y_{nk}^{(i)},\infty)}$.

(4.7) Project the point $(x_n, f[\phi^{(1)}(F_n^{(1)}),\ldots,\phi^{(M)}(F_n^{(M)})])$ onto the

| line $y = \alpha$ along a line having slope R to obtain (x^*_{n+1}, α) and then define

$$(4.8) \qquad x_{n+1} = \max\{A, x^*_{n+1}\} = \max\{A, x_n - R^{-1}[f(\phi^{(1)}(F_n^{(1)}),\ldots,\phi^{(M)}(F_n^{(M)})) - \alpha]\}.$$

(4.9) Observe y_{n+1} at x_{n+1} and iterate.

Call this the (ϕ,f)-process.

THEOREM 4.1. Assume the model of this section and assume that:

(4.10) $F_x^{(i)}(t) \geq F_y^{(i)}(t)$ for all real x, y and t such that $x \leq y$ and all $i=1,\ldots,M$;

(4.11) there is a $\delta > 0$ (we assume $0 < \delta < R$) such that $[\phi^{(i)}(F_y^{(i)}) - \phi^{(i)}(F_x^{(i)})]/[y - x] \leq 2R - \delta$ for all real x and y such that $x < y$ and all $i=1,\ldots,M$;

(4.12) for each $i=1,\ldots,M$, if $\phi^{(i)}(F) \leq \phi^{(i)}(G)$ and $0 < \alpha < 1$, then

(4.12a) $$\phi^{(i)}(F) \leq \phi^{(i)}(\alpha F + (1-\alpha)G) \leq \phi^{(i)}(G);$$

(4.13) $X \geq Y + C$ st. implies $\phi^{(i)}(F_X) \geq \phi^{(i)}(F_Y) + C$ for each $i=1,\ldots,M$; and

(4.14) if k and i are fixed nonnegative integers such that $1 \leq i \leq M$, d and x are real numbers, $\{Y_j\}$ is an i.i.d. sequence of random variables having distribution function $F_x^{(i)}$, and G_n is the empirical distribution function of $Y_1,\ldots,Y_n$, then

(4.14a) $$\phi^{(i)}\left(\frac{k}{k+n} I_{[d,\infty)} + \frac{n}{k+n} G_n\right) \to \phi^{(i)}(F_x^{(i)}) \text{ a.s. as } n \to \infty.$$

Then $x_n \to x_0$ a.s. as $n \to \infty$.

REMARK 4.1. The class of ϕ's which satisfy (4.12a) is very restricted. In particular, $\phi(F) = \int x\, dF(x)$ and $\phi(F) = m_\alpha(F)$ = the α-th quantile of F both satisfy (4.12a), but not much else does. (See Leurgans (1981) for some comments on this.) However, if f_1 and f_2 are in the class of functions available to us, then so are $\max\{f_1,f_2\}$ and $\min\{f_1,f_2\}$. If $\alpha \in [0,1]$ then in most cases $\alpha f_1 + (1-\alpha)f_2$ will also be in the class; in particular, this is true if the quantile curves are strictly increasing (i.e., if $x < y$ implies $F_x^{(i)}(t) < F_y^{(i)}(t)$ for all i and t). Thus we can get quite a few "generalized functionals" when we look at $f[\phi^{(1)}(F_n^{(1)}),\ldots,\phi^{(M)}(F_N^{(M)})]$.

Example 4.1. Suppose an experimental drug is being tested to reduce blood pressure. Suppose it has two bad side effects and that if one gives a patient dose x (or x per unit body weight) of the drug then one can observe $y = (y^{(1)},y^{(2)},y^{(3)},y^{(4)})$ where $y^{(1)} = y^{(2)}$ is either blood pressure reduction or percentage blood pressure reduction, and where $y^{(3)}$ and $y^{(4)}$ are (nondecreasing) measurements of (or measurements related to) the bad side effects. It might be desirable to have a reduction of at least $\alpha\%$ in the blood pressure of 50% of all patients, and a reduction of at least $\beta\%$ in the blood pressure readings of 90% of all patients; it might be desirable - or even mandatory - that $y^{(3)}$ be greater than γ in at most 1% of all patients, and $y^{(4)}$ be greater than δ in at most 1% of all patients. Let $F_x^{(i)}$ be the distribution of $y^{(i)}$ under dosage x for i=1,2,3,4. Let $\phi^{(1)}(F) = m_{0.5}(F) - \alpha$,

$\phi^{(2)}(F) = m_{0.1}(F) - \beta$, $\phi^{(3)}(F) = m_{0.99}(F) - \gamma$, and $\phi^{(4)}(F) = m_{0.99}(F) - \delta$.

Suppose we let $f(a,b,c,d) = \max\{\min\{a,b\},c,d\}$ and let x_0 be the x for which

$$f[\phi^{(1)}(F_x^{(1)}),\phi^{(2)}(F_x^{(2)}),\phi^{(3)}(F_x^{(3)}),\phi^{(4)}(F_x^{(4)})] = 0.$$

Then, assuming that the quantiles of $F_x^{(i)}$ are nondecreasing in x, and assuming that $x^{(i)}$ is the unique zero of $\phi^{(i)}(F_x^{(i)})$, $x_0 = \min\{\max\{x^{(1)},x^{(2)}\},x^{(3)},x^{(4)}\}$. Either 1) $x_0 = \max\{x^{(1)},x^{(2)}\}$ so that both conditions and both constraints are satisfied and either 1a) $m_{0.5}(F_{x_0}^{(1)}) = \alpha$ and $m_{0.1}(F_{x_0}^{(2)}) \geq \beta$ or 1b)

$m_{0.1}(F_{x_0}^{(2)}) = \beta$ and $m_{0.5}(F_{x_0}^{(1)}) > \alpha$; or 2) $x_0 = \min\{x^{(3)},x^{(4)}\} < \max\{x^{(1)},x^{(2)}\}$

so that

$m_{0.99}(F_{x_0}^{(3)}) = \gamma$ or $m_{0.99}(F_{x_0}^{(4)}) = \delta, m_{0.5}(F_{x_0}^{(1)}) < \alpha$ and/or $m_{0.1}(F_{x_0}^{(2)}) < \beta$, but we

are giving the maximum dosage allowable without violating one of the constraints.

Proof of Theorem 4.1. Omitted. (See Hanson and Russo (1985)).

5. Remarks.

It should be noted that a general theorem - - - where a vector of observations is observed at each x, a vector of "empirical" distribution functions is obtained, (possibly) a different functional is applied to each component of the vector of empiricals, and one is interested in $f[\phi^{(1)}(F_x^{(1)}),\ldots,\phi^{(M)}(F_x^{(M)})]$ - - - could have been proved assuming i.i.d. errors and (basically) the technical assumptions used for Theorem 2.1, or proved assuming non-i.i.d. errors and the technical assumptions used for Theorem 3.1. We feel that the assumption of "i.i.d. errors" is not reasonable, but find it interesting that under the assumptions of Theorem 2.1 we can deal with ϕ's of the form $\phi = \max \min \phi^*$ as an immediate corrolary to our main theorem.

The slope condition (3.1b) on quantile curves seems too severe, as does the combination of the slope condition (4.11) and the semi-convexity condition (4.12a). We chose to state our general theorem in one of the settings and not the other. We believe that none of these conditions are necessary, but they must be replaced by something else, not simply omitted. It is possible, however, that there is no "nice" set of conditions guaranteeing almost sure convergence for a fairly general version of the ϕ-process. As mentioned in the introduction, the ϕ-process specializes both to the Robbins-Monro process and to the quantile process of Hanson and Russo (1981); drastically different methods of proof and sets of assumptions have been used to obtain almost sure convergence in these two special cases.

We could have generalized our first formulation, the one in Sections 1, 2 and 3 where we have only one component to the observation at x, only one functional ϕ, and no f. We could have modified (1.2) so as to project the points (x_k,y_k) for each $k=1,\ldots,n$ onto the line $x = x_n$ along lines having slope $R_{n,1} > 0$ to obtain points (x_k,y_{nk}) where $y_{nk} = y_k - R_{n,1}(x_k - x_n)$. We could have modified (1.3) so as to let F_n be the weighted empirical distribution assigning weight a_{nk} to y_{nk} so that $F_n = \sum_{k=1}^{n} a_{nk} I_{[y_{nk},\infty)}$ where $a_{nk} > 0$ for all n and k and $\sum_{k=1}^{n} a_{nk} = 1$. Finally, we could have modified (1.3a) so as to project $(x_n,\phi(F_n))$ onto the line $y = \alpha$ along a line having slope $R_{n,2}$ to obtain $x_{n+1} = x_n - [\phi(F_n) - \alpha]/R_{n,2}$. These modifications allow: i) a different projection slope in obtaining (x_k,y_{nk})'s from the slope used to obtain x_{n+1} so that, for example, we might use $R_{n,1} = R_1$ and $R_{n,2} = R_2$ with $R_1 \neq R_2$ ($R_1 = 0$ might be desirable), ii) the projection slopes to become steeper as $n \to \infty$ (so that the various slope assumptions made in the theorem statement might be omitted), and iii) weights a_{nk} which favor the more recently obtained observations. In addition, any or all of $R_{n,1}$, $R_{n,2}$, and the a_{nk}'s might be random. In particular, one might want to set $R_{n,1} = R_{n,2} = R_n$ and then choose the sequence R_n so as to estimate the slope of the curve $y = \phi(f_x)$ at $x = x_0$.

Consider a general Robbins-Monro process

$$x^*_{n+1} = x^*_n - c_n(y_n - \alpha) \tag{5.1}$$

where we add the restrictions

$$0 < c_n \leq c_1 \text{ for all } n. \tag{5.2}$$

Consider also a generalized ϕ-process as just defined with $R_{n,1} = R_{n,2} = R = 1/c_1$ for all n and with the a_{nk}'s defined recursively by $a_{nn} = c_n/c_1$ for $n \geq 1$ and $a_{nk} = (1 - a_{nn})a_{n-1,k}$ for $k=1,\ldots,n-1$ and $n=2,3,\ldots$. Then, in our ϕ-process, for $n \geq 1$

$$x_{n+1} = x_n - [\phi(F_n) - \alpha]/R = x_n - R^{-1}[\sum_{k=1}^{n} a_{nk}(y_k - R(x_k - x_n)) - \alpha]$$

$$= -R^{-1}[\sum_{k=1}^{n} a_{nk}(y_k - Rx_k) - \alpha]$$

$$= -R^{-1}\{(1 - a_{nn})[\sum_{k=1}^{n-1} a_{nk}(y_k - Rx_k) - \alpha] + a_{nn}(y_n - Rx_n - \alpha)\}$$

$$= (1 - a_{nn})x_n - R^{-1}a_{nn}(y_n - \alpha) + a_{nn}x_n = x_n - c_n(y_n - \alpha)$$

so more general weights will give a Robbins-Monro process of type (5.1) subject only to (5.2).

In cases where $x_n \to x_0$, our early observations are "more biased" than our more recent ones. Our intuition tells us that we might be able to speed up our initial approach to x_0 by weighting more recent observations more heavily than older observations. This amounts to making a_{nk} an increasing (or at least nondecreasing) function of k for each fixed n. One natural way to do this is to simply "throw out" observations when they become too old. We do this by taking a nondecreasing sequence k_n of integers subject to $0 \leq k_n \leq n$ and $k_n \to \infty$ and defining

$$(5.13) \qquad a_{nk} = \begin{cases} 0 \text{ if } 1 \leq k \leq k_n \\ (n - k_n)^{-1} \text{ if } k_n < k \leq n. \end{cases}$$

Very simple arguments give examples in which this procedure is beneficial. Note that we are not talking about asymptotics or rates of convergence, so this does not contradict the known optimality results for the Robbins-Monro process.

REFERENCES

Hanson, D.L. and Russo, R.P. (1981). A new stochastic approximation procedure using quantile curves. Z. Wahrscheinlichkeitstheorie verw. Gebiete **56** 145-162.

Hanson, D.L. and Russo, R.P. (1985). Non-standard stochastic approximation for location parameters. Technical Report No. 109. Department of Statistics and Actuarial Science. University of Iowa. Iowa City, Iowa.

Leurgans, Sue. (1981). The Cauchy mean value property and linear functions of order statistics. Ann. Statist. **9** 905-908.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. Ann. Math. Statist. **25** 400-407.