STOCHASTIC APPROXIMATION OF FIXED POINTS

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

James L. Sparra

May, 1984

STOCHASTIC APPROXIMATION OF FIXED POINTS

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

James L. Sparra

May, 1984

ABSTRACT

A stochastic approximation algorithm for estimating fixed points of a self map of a compact, convex subset of R^p is presented. Sufficient conditions for the almost sure convergence of the algorithm are given. The algorithm is then applied to the statistical problem of obtaining estimates of the proportions in a mixture density.

TABLE OF CONTENTS

1. Introduction

Let A be a compact, convex subset of R^p and let F be a continuous function from A to A having a finite number of fixed points. Suppose that there is a continuous function $f:A \times R^q \to A$ and an R^q-valued random variable X such that $F(\alpha) = Ef(\alpha,X)$ for each $\alpha \in A$, where $Ef(\alpha,X)$ denotes the expectation of the random variable $f(\alpha,X)$. Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent, identically distributed random variables, each having the same distribution as X. We may assume that the random variables X_i are defined on the same probability space Ω. Given $\alpha^1 \in A$, define a sequence of random variables $\{\alpha^i\}_{i=1}^{\infty}$ by

$$\alpha^{i+1} = \frac{i}{i+1}\alpha^i + \frac{1}{i+1} f(\alpha^i, X_{i+1}). \tag{1.1}$$

In this paper we will give sufficient conditions for the almost sure convergence of (1.1) to a fixed point of F.

Thus we are concerned with the problem of approximating a fixed point of the function F or, equivalently, approximating a zero of the function $F(\alpha) - \alpha$. However, here we are assuming that $F(\alpha)$ is not available for evaluation; rather we may measure only the random variable $f(\alpha,X)$, an unbiased estimate of $F(\alpha)$. (A function such as F is frequently referred to as a "regression function"; intuitively we may think of $f(\alpha,X)$ as $F(\alpha)$ plus a nonnegligible random error term). Thus the application of classical deterministic methods such as fixed point iterations or Newton's method is not appropriate. Obviously the probabalistic element must be taken into account.

The problem of determining a zero of a regression function on R^1 was essentially solved by Robbins and Monro in 1951 by the method which they called stochastic approximation [7]. Their results were subsequently extended by Blum to the case of a regression function defined on R^p [1]. The procedure suggested by Robbins and Monro may be described as follows. Suppose that G is a regression function on R^p, that for each $\alpha \varepsilon R^p$ there is a random variable $g(\alpha)$ such that $Eg(\alpha) = G(\alpha)$. Assume that G has a unique zero $\hat{\alpha}$ and that

$$\inf_{\varepsilon < ||\alpha - \hat{\alpha}||} \langle G(\alpha), \alpha - \hat{\alpha} \rangle > 0 \tag{1.2}$$

for all $\varepsilon > 0$. Let $\{a_i\}_{i=1}^{\infty}$ be a sequence of positive numbers such that

$$\sum_{i=1}^{\infty} a_i = \infty, \quad \sum_{i=1}^{\infty} a_i^2 < \infty \, . \tag{1.3}$$

Given an arbitrary $\alpha^1 \epsilon R^p$ define

$$\alpha^{i+1} = \alpha^i - a_{i+1} g(\alpha^i) \tag{1.4}$$

Then, with some additional assumptions, the sequence given by (1.4) converges to $\hat{\alpha}$ with probability one.

We note that (1.1) is a special case of (1.4), with $g(\alpha) = \alpha - f(\alpha, X)$ and $a_i = \frac{1}{i}$. Thus (1.1) is a Robbins-Monro stochastic approximation algorithm and the problem of its convergence could be treated in the general framework outlined above by taking $G(\alpha) = \alpha - F(\alpha)$. Then it would be necessary to show that F satisfies (1.2). That is, assuming that F has a unique fixed point $\hat{\alpha}$, F must satisfy

$$\inf_{\varepsilon<||\alpha-\hat{\alpha}||} \langle\alpha-F(\alpha), \alpha-\hat{\alpha}\rangle > 0 \tag{1.5}$$

for all $\varepsilon > 0$. However, we are concerned with the convergence of (1.1) in a setting which is more restricted than that of the general stochastic approximation theory. Therefore we have chosen to attack the problem de novo, our results being given in sections 2 and 3 of this paper. There we show, without imposing any additional constraints on the regression function F, that if $A \subseteq R^1$ then (1.1) converges to a fixed point of F with probability one. For $A \subseteq R^p$, $p > 1$, we obtain convergence by requiring that F satisfy a condition that subsumes (1.5) as a special case. Moreover we relax the requirement that F have a unique fixed point. Thus we provide sufficient conditions for the convergence of (1.1) which are weaker than those required by the general stochastic approximation theory.

The approximation problem considered here is a natural generalization of the following deterministic problem. Suppose that G is a continuous self map of a convex, compact subset A of R^p. Then, as is well known, G has a fixed point in A. The problem then is to approximate a fixed point of G. The classical approach is to choose an arbitrary point $\alpha^1 \in A$ and construct a sequence $\{\alpha^i\}_{i=1}^{\infty}$ of successive iterates of α^1 under G,

$$\alpha^{i+1} = G(\alpha^i). \tag{1.6}$$

If the sequence converges then obviously the limit is a fixed point of G. The sequence will converge if, for instance, G is a contractive mapping. For cases where this successive iteration scheme is divergent several authors [2,5] have suggested replacing (1.6) with

$$\alpha^{i+1} = \frac{i}{i+1}\alpha^{i} + \frac{1}{i+1}G(\alpha^{i}). \tag{1.7}$$

Such an iteration process is called, for obvious reasons, a mean value iteration scheme. Of course (1.7) is just (1.1) with $G(\alpha) = F(\alpha) = f(\alpha,X)$. Thus our convergence results are applicable to (1.7) and in section 3 we consider some mean value iteration theorems as corollaries to our results.

In section 4 we apply our results to the following statistical parameter estimation problem. Suppose that h is a probability density that is a convex combination of densities,

$$h = \sum_{i=1}^{p} \hat{\alpha}_i h_i .$$

Such a density is called a mixture density and the scalars $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ are called the proportions of the mixture. Let X be a random variable whose distribution has density h and let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent random variables having the same distribution as X (This is the mathematical formulation of the process of random sampling from a population of measurements distributed according to h; the X_i's are commonly called sample or observation random variables and particular values of the X_i's are called independent observations on X). The problem we address is that of estimating the proportions vector $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)$, given a sequence of independent observations on X. By recognizing that $\hat{\alpha}$ is a fixed point of a particular regression function we are able to implement (1.1) to provide a stochastic approximation algorithm for estimating $\hat{\alpha}$. Young and Coraluppi [9] and Lawton [3] have presented stochastic approximation algorithms for approximating $\hat{\alpha}$ in mixtures of normal densities.

2. Preliminary Lemmas

In this section we provide the basic tools for establishing sufficient conditions for the convergence of (1.1). For the remainder of the paper Ω will denote the underlying probability space for the random variables X_i and if $\omega \in \Omega$ then the value $X_i(\omega)$ will be denoted by x_i.

Suppose that g is a continuous function from $A \times R^q$ to R^r such that if $\alpha \in A$ then there is a positive number ε such that

$$\sup_{\substack{\beta \in B(\alpha,\varepsilon) \\ x \in R^q}} \{||g(\alpha,x)-g(\beta,x)||\} < \infty \tag{2.1}$$

where $B(\alpha,\varepsilon)$ denotes the open ball about α of radius ε and $||\cdot||$ denotes the Euclidean norm. Let $G(\alpha) \equiv Eg(\alpha,X)$ for $\alpha \in A$ and let ~~$\Omega_0(g)$ denote the set of all $\omega \in \Omega$ such that for each $\alpha \in A$ and for~~ each positive integer k,

$$\frac{1}{n} \sum_{i=1}^{n} g(\alpha,x_i) \to G(\alpha) \quad \text{as} \quad n \to \infty \tag{2.2}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} I_k(x_i) \to \mu_X(B_k) \quad \text{as} \quad n \to \infty \tag{2.3}$$

where B_k denotes the closed ball of radius k about the origin of R^q, I_k denotes the indicator function of B_k, and μ_X denotes the probability measure on R^q induced by the random variable X. The convergence of (1.1), under suitable conditions, will be proved for $\omega \in \Omega_o(f)$. Thus it is necessary to have

Lemma 2.1: $\text{prob}(\Omega_o(g)) = 1$.

Proof: For $\alpha \in A$ let M_α denote the set of all $\omega \in \Omega$ such that (2.2) holds and for a positive integer k let N_k denote the set of all $\omega \in \Omega$ such that (2.3) holds. Let S be a countable dense subset of A and define

$$\Omega_1 \equiv \left(\bigcap_{\alpha \in S} M_\alpha \right) \cap \left(\bigcap_{k=1}^{\infty} N_k \right).$$

That is, Ω_1 is the set of all $\omega \in \Omega$ such that (2.2) and (2.3) hold for all $\alpha \in S$ and all positive integers k. By the Strong Law of Large Numbers (See Loéve [4]) each of the sets M_α and N_k has probability measure one. Hence it follows that $\text{prob}(\Omega_1) = 1$.

Now we will show that $\Omega_1 \subseteq \Omega_o(g)$. Suppose that $\omega \in \Omega_1$ and $\alpha \in A$. Let $\varepsilon > o$ and set

$$b = \sup_{\substack{\beta \in B(\alpha,\varepsilon') \\ x \in R^q}} \{||g(\alpha,x) - g(\beta,x)||\}$$

where ε' is a positive number such that (2.1) is satisfied. Assume that $b \neq o$, noting that (2.2) holds if $b = o$. Take k to be a positive integer such that

$$\mu_x(B_k) > 1 - \frac{\varepsilon}{4b}$$

and take $\beta \in S$ such that

$$||\alpha-\beta|| < \varepsilon', \ ||G(\alpha)-G(\beta)|| < \frac{\varepsilon}{4}, \ ||g(\alpha,x)-g(\beta,x)|| < \frac{\varepsilon}{4}$$

for all $x \in B_k$. Let n_o be a positive integer such that if $n \geq n_o$ then

$$||G(\beta) - \frac{1}{n}\sum_{i=1}^{n} g(\beta,x_i)|| < \frac{\varepsilon}{4}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I_k(x_i) > 1 - \frac{\varepsilon}{4b} .$$

Then if $n \geq n_o$,

$$\begin{aligned}
||G(\alpha) - \frac{1}{n}\sum_{i=1}^{n} g(\alpha,x_i)|| &\leq ||G(\alpha) - G(\beta)|| + ||G(\beta) - \frac{1}{n}\sum_{i=1}^{n} g(\beta,x_i)|| \\
&\quad + ||\frac{1}{n}\sum_{i=1}^{n} g(\beta,x_i) - \frac{1}{n}\sum_{i=1}^{n} g(\alpha,x_i)|| \\
&\leq ||G(\alpha) - G(\beta)|| + ||G(\beta) - \frac{1}{n}\sum_{i=1}^{n} g(\beta,x_i)|| \\
&\quad + ||\frac{1}{n}\sum_{i=1}^{n} I_k(x_i)(g(\beta,x_i) - g(\alpha,x_i))|| \\
&\quad + ||\frac{1}{n}\sum_{i=1}^{n} (1 - I_k(x_i))(g(\beta,x_i) - g(\alpha,x_i))|| \\
&\leq ||G(\alpha) - G(\beta)|| + ||G(\beta) - \frac{1}{n}\sum_{i=1}^{n} g(\beta,x_i)|| \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} I_k(x_i)||g(\beta,x_i) - g(\alpha,x_i)|| \\
&\quad + b\,\frac{1}{n}\sum_{i=1}^{n} (1 - I_k(x_i)) \\
&< \varepsilon.
\end{aligned}$$

Thus for $\omega \in \Omega_1$ (2.2) holds for all $\alpha \in A$. Trivially, for $\omega \in \Omega_1$ (2.3) holds for each positive integer k. Hence $\Omega_1 \subseteq \Omega_o(g)$ and $\text{prob}(\Omega_o(g)) = 1$.

Lemma 2.2: Suppose that $\{Z_i\}_{i=1}^{\infty}$ is a sequence in R^p and $Z \in R^p$ is such that

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \to Z \quad \text{as} \quad n \to \infty .$$

Then if $\varepsilon > o$ and $\lambda > o$ there is a positive integer n_o such that if $n \geq n_o$ and $\frac{m}{n} \geq \lambda$ then

$$\left|\left|\frac{1}{m}\sum_{i=n+1}^{n+m} Z_i - Z\right|\right| < \varepsilon.$$

Proof: Let δ be a positive number less than $\frac{\varepsilon\lambda}{\lambda + 2}$ and let n_o be ~~a positive integer~~ such ~~that if~~ ~~$n \geq$~~ n_o then

$$\left|\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - Z\right|\right| < \delta.$$

Suppose that $n \geq n_o$ and $\frac{m}{n} \geq \lambda$. Then, noting that

$$\frac{1}{m}\sum_{i=n+1}^{n+m} Z_i - \frac{1}{n+m}\sum_{i=1}^{n+m} Z_i = \frac{n}{m}\left(-\frac{1}{n}\sum_{i=1}^{n} Z_i + \frac{1}{n+m}\sum_{i=1}^{n+m} Z_i\right),$$

we have

$$||\frac{1}{m}\sum_{i=n+1}^{n+m} Z_i - Z|| \le ||\frac{1}{m}\sum_{i=n+1}^{n+m} Z_i - \frac{1}{n+m}\sum_{i=1}^{n+m} Z_i|| + ||\frac{1}{n+m}\sum_{i=1}^{n+m} Z_i - Z||$$

$$= \frac{n}{m}||\frac{1}{n}\sum_{i=1}^{n} Z_i - \frac{1}{n+m}\sum_{i=1}^{n+m} Z_i|| + ||\frac{1}{n+m}\sum_{i=1}^{n+m} Z_i - Z||$$

$$\le \frac{n}{m}(||\frac{1}{n}\sum_{i=1}^{n} Z_i - Z|| + ||\frac{1}{n+m}\sum_{i=1}^{n+m} Z_i - Z||) + ||\frac{1}{n+m}\sum_{i=1}^{n+m} Z_i - Z||$$

$$< \frac{2\delta}{\lambda} + \delta < \varepsilon,$$

as was to be shown.

The following lemma is the basic tool that will be used in obtaining convergence results for (1.1). Recall that $B(\alpha,\varepsilon)$ denotes the open ball about α of radius ε.

Lemma 2.3: Let $\omega \in \Omega_0(g)$. If $\alpha \in A$ and ε, λ, and δ are positive numbers then there is a positive number δ_0 with $\delta_0 < \delta$ and a positive integer n_0 such that if $n \ge n_0$, $\frac{m}{n} \ge \lambda$, and $\beta_i \in A \cap B(\alpha,\delta_0)$ for $i = n+1, \ldots, n+m$ then

$$||\frac{1}{m}\sum_{i=n+1}^{n+m} g(\beta_i,x_i) - G(\alpha)|| < \varepsilon.$$

Proof: Let $\omega \in \Omega_0(g)$, fix $\alpha \in A$ and let ε, λ, and δ be positive numbers. Set

$$b = \sup_{\substack{\beta\in B(\alpha,\varepsilon') \\ x\in R^q}} \{||g(\alpha,x) - g(\beta,x)||\}$$

where ε' is a positive number such that (2.1) is satisfied (Assume $b \neq o$; otherwise the lemma is trivial). Take k to be a positive integer such that

$$\mu_X(B_k) > 1 - \frac{\varepsilon}{3b} .$$

Let $\delta_o > o$ be such that $\delta_o < \min\{\varepsilon',\delta\}$ and such that if $||\alpha-\beta|| < \delta_o$ and $x \in B_k$ then

$$||g(\alpha,x) - g(\beta,x)|| < \frac{\varepsilon}{3} .$$

By lemma 2.2 there is a positive integer n_o such that if $n \geq n_o$ and $\frac{m}{n} \geq \lambda$ then

$$||\frac{1}{m} \sum_{i=n+1}^{n+m} g(\alpha,x_i) - G(\alpha)|| < \frac{\varepsilon}{3} .$$

and

$$\frac{1}{m} \sum_{i=n+1}^{n+m} I_k(x_i) > 1 - \frac{\varepsilon}{3b} .$$

Then, if $n \geq n_o$, $\frac{m}{n} \geq \lambda$, and $\beta_i \in A \cap B(\alpha,\delta_o)$ for $i = n + 1, \ldots, n + m$,

$$||\frac{1}{m} \sum_{i=n+1}^{n+m} g(\beta_i,x_i) - G(\alpha)|| \leq ||\frac{1}{m} \sum_{i=n+1}^{n+m} g(\beta_i,x_i) - \frac{1}{m} \sum_{i=n+1}^{n+m} g(\alpha,x_i)||$$

$$+ ||\frac{1}{m} \sum_{i=n+1}^{n+m} g(\alpha,x_i) - G(\alpha)||$$

$$\leq \frac{1}{m} \sum_{i=n+1}^{n+m} I_k(x_i)||g(\beta_i,x_i) - g(\alpha,x_i)||$$

$$+ \frac{1}{m} \sum_{i=n+1}^{n+m} (1 - I_k(x_i)) || g(\beta_i, x_i) - g(\alpha, x_i) ||$$

$$+ || \frac{1}{m} \sum_{i=n+1}^{n+m} g(\alpha, x_i) - G(\alpha) ||$$

$$< \varepsilon.$$

As indicated, the preceding lemma will be used several times. In order to make its invocation more concise we introduce the following.

Notation: Let g, α, ε, λ, and δ be as in lemma 2.3. The positive integer n_0 given by the lemma will be denoted by

$$N(\varepsilon, \lambda)$$

and the positive number δ_0 will be denoted by

$$D(\varepsilon, \delta).$$

Although N and D depend on g and α, this is suppressed in the notation as the relevant g and α will be clear from context. Note that N does not depend on δ and D does not depend on λ.

The following lemma is a simple but useful observation. Recall that α^i denotes the ith iterate of (1.1).

Lemma 2.4: $\alpha^{n+m} = \frac{n}{n+m} \alpha^n + \frac{1}{n+m} \sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i)$.

Proof: By induction on m.

To conclude this section we show that, with probability one, if (1.1) converges then it converges to a fixed point of F.

Theorem 2.1: Let $\omega \in \Omega_0(f)$. If $\alpha^i \to \alpha$ then $F(\alpha) = \alpha$.

Proof: Let $\omega \in \Omega_0(f)$ and suppose that $\alpha^i \to \alpha$. Let $\varepsilon > 0$ and put

$$\delta = \min\{D(\tfrac{\varepsilon}{4}, 1), \tfrac{\varepsilon}{4}\}.$$

Choose $n \geq N(\frac{\varepsilon}{4}, 1)$ such that if $m \geq n$ then

$$||\alpha^m - \alpha|| < \delta$$

Fix $m \geq n$ such that

$$\frac{n}{n+m}||\alpha^n|| < \frac{\varepsilon}{4} \quad \text{and} \quad \frac{n}{n+m}||F(\alpha)|| < \frac{\varepsilon}{4} .$$

Then

$$\begin{aligned}
||F(\alpha) - \alpha|| &\leq ||F(\alpha) - \alpha^{n+m}|| + ||\alpha^{n+m} - \alpha|| \\
&= ||F(\alpha) - \frac{n}{n+m}\alpha^n - \frac{1}{n+m}\sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i)|| + ||\alpha^{n+m} - \alpha|| \\
&\leq ||F(\alpha) - \frac{1}{n+m}\sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i)|| + \frac{n}{n+m}||\alpha^n|| + ||\alpha^{n+m} - \alpha|| \\
&= ||\frac{n}{n+m}F(\alpha) + \frac{m}{n+m}F(\alpha) - \frac{m}{n+m}\cdot\frac{1}{m}\sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i)|| \\
&\qquad + \frac{n}{n+m}||\alpha^n|| + ||\alpha^{n+m} - \alpha|| \\
&\leq \frac{n}{n+m}||F(\alpha)|| + ||F(\alpha) - \frac{1}{m}\sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i)||
\end{aligned}$$

$$+ \frac{n}{n+m}||\alpha^n|| + ||\alpha^{n+m} - \alpha||$$

$$< \varepsilon.$$

Thus $F(\alpha) = \alpha$ and the proof is completed.

3. Convergence Theorems

In this section we establish sufficient conditions for the convergence of (1.1) to a fixed point of F. As might be expected, things are considerably simplified when $A \subseteq R^1$; the lemmas of the previous section alone are sufficient to prove

Theorem 3.1: If $A \subseteq R^1$ then (1.1) converges to a fixed point of F with probability one.

Proof: Let $\omega \in \Omega_0(f)$. Suppose by way of contradiction that (1.1) does not converge, that $\liminf \alpha^i < \limsup \alpha^i$. Take c such that

$$\liminf \alpha^i < c < \limsup \alpha^i$$

and such that $F(c) \neq c$. To be concrete, assume that $F(c) > c$. Let

$$\varepsilon = \tfrac{1}{2}(F(c) - c), \; \delta = \tfrac{1}{2}(\limsup \alpha^i - c)$$

and put

$$\lambda = \frac{D(\varepsilon,\delta)}{2b}$$

where

$$b = \sup_{\substack{\alpha\in A \\ x\in R^q}}\{|f(\alpha,x) - \alpha|\}.$$

Fix n and m to be positive integers such that

$$n \geq \max\{N(\varepsilon,\lambda),\ 1/\lambda\}$$

$$\alpha^n > c + \frac{D(\varepsilon,\delta)}{2},\ \alpha^{n+m} < c$$

and

$$c \leq \alpha^{n+k} \leq c + \frac{D(\varepsilon,\delta)}{2}$$

for $k = 1, \ldots, m - 1$. Then noting that

$$\alpha^i - \alpha^{i-1} = \frac{1}{i}(f(\alpha^{i-1},\ x_i) - \alpha^{i-1})$$

we have

$$\frac{D(\varepsilon,\delta)}{2} < |\alpha^{n+m} - \alpha^n| \leq \sum_{i=n+1}^{n+m} |\alpha^i - \alpha^{i-1}| \leq m(\frac{b}{n}).$$

Thus,

$$\frac{m}{n} > \frac{D(\varepsilon,\delta)}{2b} = \lambda.$$

Then,

$$c > \alpha^{n+m} = \frac{n}{n+m}\alpha^n + \frac{1}{n+m}\sum_{i=n+1}^{n+m} f(\alpha^{i-1},\ x_i)$$

$$= \frac{n}{n+m}\alpha^n + \frac{m}{n+m} \cdot \frac{1}{m}\sum_{i=n+1}^{n+m} f(\alpha^{i-1},\ x_i)$$

$$> \frac{n}{n+m} c + \frac{m}{n+m} c = c,$$

which is of course a contradiction. Thus (1.1) converges for $\omega \in \Omega_0(f)$. Moreover, by theorem 2.1 (1.1) converges to a fixed point of F for

$\omega \in \Omega_0(f)$. Recalling that $\text{prob}(\Omega_0(f)) = 1$, the proof is completed.

For the multidimensional case further assumptions must be made to assure the convergence of (1.1). We assume that there is a continuous concave functional L defined on A such that if $\alpha \in A$ and $F(\alpha) \neq \alpha$ then there is a real number $t \in (o,1]$ such that

$$L((1-t)\alpha + tF(\alpha)) > L(\alpha) \tag{3.1}$$

With this additional hypothesis we can obtain convergence of (1.1) for compact, convex $A \subseteq R^p$, $p \geq 1$. We will show that $L(\alpha^i)$ converges and thence that α^i converges to a fixed point of F.

Lemma 3.1: Let $\omega \in \Omega_0(f)$. If $F(\alpha) \neq \alpha$ and $\delta > o$ then there are positive numbers δ_1, δ_2, and γ, with $\delta_2 < \delta_1 < \delta$ and a positive integer n_0 such that if $n \geq n_0$ and $||\alpha^n - \alpha|| < \delta_2$ then there is a positive integer m such that $||\alpha^{n+m} - \alpha|| > \delta_1$ and if m is the least such positive integer then $L(\alpha^{n+m}) \geq L(\alpha) + \gamma$.

Proof: For $\alpha,\beta \in R^p$ with $\alpha \neq \Theta$, $\beta \neq \Theta$ define

$$c(\alpha,\beta) \equiv \frac{\langle\alpha,\beta\rangle}{||\alpha||\cdot||\beta||}$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product. Let $\omega \in \Omega_0(f)$, $\alpha \in A$ such that $F(\alpha) \neq \alpha$, and $\delta > o$. Let $t \in (o,1]$ such that (3.1) is satisfied. Set

$$\alpha_t = (1-t)\alpha + tF(\alpha).$$

By the concavity of L there is an $\varepsilon > o$ such that if $\beta \in A$ and $o < ||\beta - \alpha|| \leq ||\alpha_t - \alpha||$ and $c(\beta - \alpha, F(\alpha) - \alpha) \geq 1 - \varepsilon$ then $L(\beta) > L(\alpha)$; and since $c(\cdot, F(\alpha) - \alpha)$ is continuous except at Θ there is an $\varepsilon' > o$ such that if $||F(\alpha) - \beta|| < \varepsilon'$ then $c(\beta - \alpha, F(\alpha) - \alpha) > 1 - \frac{\varepsilon}{2}$. Moreover ε' may be taken so that if $||F(\alpha) - \beta|| < \varepsilon'$ then $||\beta - \alpha|| > ||\alpha_t - \alpha||$. By theorem 2.1 there is a $\delta'' > o$ such that $\mathrm{card}(\{i \mid ||\alpha^i - \alpha|| \geq \delta''\})$ is infinite. Let

$$\delta''' = \min\{\delta, \delta'', ||\alpha_t - \alpha||/2\}$$

and put

$$\delta_1 = D(\varepsilon', \delta''')$$

Again by the continuity of $c(\cdot, F(\alpha) - \alpha)$ there is a $\delta' > o$ such that if $\alpha', \beta' \in R^p$ and

$$\min\{||\alpha'||, ||\beta'||\} \geq \frac{\delta_1^{\,2}}{2b + \delta_1}$$

where

$$b = \sup_{\substack{\beta \in A \\ x \in R^q}} \{||f(\beta, x) - \beta||\}$$

and if $||\alpha' - \beta'|| < \delta'$ then $|c(\alpha', F(\alpha) - \alpha) - c(\beta', F(\alpha) - \alpha)| < \frac{\varepsilon}{2}$. Let

$$\delta_2 = \min\{\delta', \delta_1/2\}.$$

Put

$$M = \{\beta \in A \mid \delta_1 \leq ||\beta - \alpha|| \leq ||\alpha_t - \alpha|| \text{ and } c(\beta - \alpha, F(\alpha) - \alpha) \geq 1 - \varepsilon\}$$

Let

$$\gamma = \inf_{\beta\in M}\{L(\beta) - L(\alpha)\}$$

Note that $\gamma > o$. Choose n_o such that

$$n_o > \max\{N(\varepsilon', \delta_1/2b), b/\delta_1\}$$

Now suppose that $n \geq n_o$ and $||\alpha^n - \alpha|| < \delta_2$. Since $\delta_2 < \delta_1 < \delta''$ there is a least positive integer m such that $||\alpha^{n+m} - \alpha|| > \delta_1$. Then we have that

$$\frac{\delta_1}{2} < ||\alpha^{n+m} - \alpha^n|| \leq \sum_{i=n+1}^{n+m} ||\alpha^i - \alpha^{i-1}|| \leq m(\frac{b}{n}).$$

So, $\frac{m}{n} > \frac{\delta_1}{2b}$, $||\alpha^{i-1} - \alpha|| < D(\varepsilon', \delta''')$ for $i = n + 1, \ldots, n + m$ and $n > N(\varepsilon', \delta_1/2b)$. Hence by lemma 2.3,

$$||\frac{1}{m} \sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i) - F(\alpha)|| < \varepsilon' \tag{3.2}$$

Set

$$s = \sum_{i=n+1}^{n+m} f(\alpha^{i-1}, x_i).$$

Then from (3.2) it follows that

$$|c(\frac{1}{m} s - \alpha, F(\alpha) - \alpha) - 1| < \varepsilon/2 \tag{3.3}$$

and

$$||\frac{1}{m} s - \alpha|| > ||\alpha_t - \alpha|| > \delta_1. \tag{3.4}$$

Now

$$||\alpha^{n+m} - \alpha|| > \delta_1 > \frac{\delta_1^2}{2b + \delta_1},$$

and since $m > (n\delta_1)/2b$ we have from (3.4)

$$||\frac{m}{n + m}(\frac{1}{m} s - \alpha)|| > \frac{\delta_1}{2b + \delta_1}||\frac{1}{m} s - \alpha|| > \frac{\delta_1^2}{2b + \delta_1}$$

Then noting that

$$||(\alpha^{n+m} - \alpha) - \frac{m}{n + m}(\frac{1}{m} s - \alpha)||$$

$$= ||(\frac{n}{n + m} \alpha^n + \frac{1}{n + m} s - \alpha) - (\frac{1}{m + n} s - \frac{m}{n + m} \alpha)||$$

$$= \frac{n}{n + m}||\alpha^n - \alpha|| < \delta_2 \le \delta'$$

we have

$$|c(\alpha^{n+m} - \alpha, F(\alpha) - \alpha) - c(\frac{m}{n + m}(\frac{1}{m} s - \alpha), F(\alpha) - \alpha)| < \frac{\varepsilon}{2}. \quad (3.5)$$

Then noting (3.3) and (3.5) we have

$$|c(\alpha^{n+m} - \alpha, F(\alpha) - \alpha) - 1| \le |c(\alpha^{n+m} - \alpha, F(\alpha) - \alpha) - c(\frac{1}{m} s - \alpha, F(\alpha) - \alpha)|$$

$$+ |c(\frac{1}{m} s - \alpha, F(\alpha) - \alpha) - 1|$$

$$= |c(\alpha^{n+m} - \alpha, F(\alpha) - \alpha) - c(\frac{m}{n+m}(\frac{1}{m} s - \alpha), F(\alpha) - \alpha)|$$

$$+ |c(\frac{1}{m} s - \alpha, F(\alpha) - \alpha) - 1|$$

$$< \varepsilon,$$

and

$$||\alpha^{n+m} - \alpha|| \le ||\alpha^{n+m} - \alpha^{n+m-1}|| + ||\alpha^{n+m-1} - \alpha||$$

$$\le \frac{1}{n+m}||f(\alpha^{n+m-1}, x_{n+m}) - \alpha^{n+m-1}|| + \delta_1$$

$$\le \frac{b}{n} + \delta_1 \le 2\delta_1 \le ||\alpha_t - \alpha||$$

Thus $\alpha^{n+m} \in M$. Hence $L(\alpha^{n+m}) \ge L(\alpha) + \gamma$ and the lemma is proven.

Lemma 3.2: If $\omega \in \Omega_o(f)$ then $L(\alpha^i)$ converges.

Proof: Let $\omega \in \Omega_o(f)$ and assume by way of contradiction that $L(\alpha^i)$ does not converge. Let c be a real number such that

$$\liminf L(\alpha^i) < c < \limsup L(\alpha^i)$$

and $L^{-1}(c)$ contains no fixed points of F. For $\alpha \in L^{-1}(c)$ let $\delta(\alpha)$ be a positive number such that if $||\beta - \alpha|| < \delta(\alpha)$ then

$$L(\beta) \ge \frac{1}{2}(c + \liminf L(\alpha^i))$$

and let $\delta_1(\alpha)$, $\delta_2(\alpha)$, and $n_o(\alpha)$ be as in lemma 3.1 with $\delta_2(\alpha) < \delta_1(\alpha) < \delta(\alpha)$. Since $L^{-1}(c)$ is compact, there are points $\beta_1, \ldots, \beta_r$ in $L^{-1}(c)$ such that the open balls $B(\beta_j, \delta_2(\beta_j))$, $j = 1, \ldots, r$, cover $L^{-1}(c)$ and since

$$||\alpha^i - \alpha^{i-1}|| \to o \quad \text{as} \quad i \to \infty$$

there is a positive integer n_1 such that if $n \geq n_1$, $L(\alpha^n) \geq c$, and $L(\alpha^{n+1}) \leq c$ then $\alpha^n \in B(\beta_j, \delta_2(\beta_j))$ for some j, $i \leq j \leq r$. Fix n and m to be positive integers such that

$$n \geq \max\{n_1, n_0(\beta_1), \ldots, n_0(\beta_r)\},$$

$$L(\alpha^n) \geq c,\ L(\alpha^{n+m}) < \frac{1}{2}(c + \liminf L(\alpha^i))$$

and

$$\frac{1}{2}(c + \liminf L(\alpha^i)) \leq L(\alpha^{n+k}) \leq c \tag{3.5}$$

for $k = 1, \ldots, m - 1$. Then

$$\alpha^n \in B(\beta_{j_0}, \delta_2(\beta_{j_0}))$$

for some j_0. Let k_0 be the least positive integer such that

$$\alpha^{n+k_0} \notin B(\beta_{j_0}, \delta_1(\beta_{j_0})).$$

Then, by lemma 3.1,

$$L(\alpha^{n+k_0}) > c. \tag{3.6}$$

But $k_0 < m$, since for $k < k_0$

$$\alpha^{n+k} \in B(\beta_{j_0}, \delta_1(\beta_{j_0}))$$

and hence

$$L(\alpha^{n+k}) \geq \frac{1}{2}(c + \liminf L(\alpha^i))$$

for $k = 1, \ldots, k_0$. But this, together with (3.6), contradicts (3.5). Thus $L(\ ^i)$ converges for $\omega \in \Omega_0(f)$.

Theorem 3.2: If $\omega \in \Omega_0(f)$ then (1.1) converges to a fixed point of F.

Proof: Let $\omega \in \Omega_0(f)$. Set

$$c = \lim_{i\to\infty} L(\alpha^i).$$

Then the cluster points of $\{\alpha^i\}_{i=1}^{\infty}$ all lie in $L^{-1}(c)$. But all of these cluster points are fixed points of F, since by Lemma 3.1 if α is a cluster point and $F(\alpha) \neq \alpha$ then

$$\limsup L(\alpha^i) > L(\alpha) = c.$$

Hence, since F has only a finite number of fixed points and

$$||\alpha^i - \alpha^{i-1}|| \to 0 \quad \text{as} \quad i \to \infty$$

$\{\alpha^i\}_{i=1}^{\infty}$ has exactly one cluster point. Thus, if $\omega \in \Omega_0(f)$ then (1.1) converges to a fixed point of F.

Corollary: Suppose that F has a unique fixed point $\hat{\alpha}$ and that

$$<F(\alpha) - \alpha, \hat{\alpha} - \alpha> > 0$$

for $\alpha \neq \hat{\alpha}$. Then (1.1) converges to $\hat{\alpha}$ with probability one.

Proof: Take $L(\alpha) = -||\alpha - \hat{\alpha}||^2$ in the theorem.

We note that if f is independent of x then $f(\alpha, x) = F(\alpha)$ and we obtain deterministic mean value iteration theorems as corollaries to theorem 3.1 and 3.2. In particular as an immediate corollary to theorem 3.1 we have the following theorem of Franks and Marzec [2]:

Theorem 3.3: Let $G(x)$ be a continuous self map of $[o,1]$ having a finite number of fixed points. Then the iterative scheme given by

$$\text{(i)} \quad x_{n+1} = G(\overline{x}_n)$$

$$\text{(ii)} \quad \overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\text{(iii)} \quad \overline{x}_1 = x_1 \in [o,1]$$

converges to a fixed point of G.

The following special case of a more general theorem of Outlaw [5] is an easy corollary to theorem 3.2:

Theorem 3.4: Let $G(x)$ be a nonexpansive (i.e. $||G(x) - G(y)|| \leq ||x - y||$) self map of a convex, compact subset E of R^p having a unique fixed point x_o. Then the iteration scheme

$$x_{n+1} = \frac{n}{n+1} x_n + \frac{1}{n+1} G(x_n), \; x_1 \in E$$

converges to x_o.

Proof: Let $L(x) = -||x - x_o||^2$ in theorem 3.2.

4. Proportions Estimation

We now address the problem of obtaining estimates for the proportions in a mixture of density functions. Let X be a q-dimensional random variable whose density h is a mixture, i.e.

$$h = \sum_{i=1}^{p} \hat{\alpha}_i h_i$$

where

$$\hat{\alpha}_i > o, \quad \sum_{i=1}^{p} \hat{\alpha}_i = 1,$$

and each h_i is a probability density function. We assume that the h_i's are known and are linearly independent. Then, given a sequence $\{x_i\}_{i=1}^{\infty}$ of independent observations on X, the problem is to estimate the proportions vector $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)$.

Let A be the set of points α in R^p satisfying the constraints

$$\alpha_i \geq o, \quad \sum_{i=1}^{p} \alpha_i = 1.$$

Define $f: A \times R^q \rightarrow A$ by

$$f_i(\alpha, x) = \frac{\alpha_i h_i(x)}{\sum_{j=1}^{p} \alpha_j h_j(x)}$$

and define F by

$$F(\alpha) = Ef(\alpha, X).$$

Then F is a self map of A and $\hat{\alpha}$ is a fixed point of F. These considerations suggest that we may apply the procedure (1.1) to approximate $\hat{\alpha}$.

Define a functional L on A by

$$L(\alpha) \equiv E\log\left(\sum_{i=1}^{p} \alpha_i h_i(X)\right),$$

the expectation of the log-likelihood function. Note that L is concave over A. Properties of the log-likelihood function relevant to the proportions estimation problem have been studied by other authors [6,8] and we make use of their results here.

Theorem 4.1 (H. Walker): $L(F(\alpha)) \geq L(\alpha)$, with equality if and only if $F(\alpha) = \alpha$.

The proof of the theorem requires the following two lemmas. To simplify notation we make the convention that if $\alpha \in A$ then $F(\alpha)$ will be denoted by $\overline{\alpha}$ and the function $f_i(\alpha, \cdot)$ will be denoted by $f_i(\alpha)$; and, when appearing in the argument of the expectation operator E, $f_i(\alpha)$ and h_i will denote $f_i(\alpha, X)$ and $h_i(X)$, respectively.

Lemma 4.1: $\sum_{i=1}^{p} E(f_i(\alpha)\log(\overline{\alpha}_i h_i)) \geq \sum_{i=1}^{p} E(f_i(\alpha)\log(\alpha_i h_i))$, with equality if and only if $\alpha = \overline{\alpha}$.

Proof: The inequality in the lemma holds if and only if

$$\sum_{i=1}^{p} E(f_i(\alpha)\log \frac{\overline{\alpha}_i}{\alpha_i}) \geq o \tag{4.1}$$

Set

$$b_i = E\left(\frac{h_i}{\sum_{j=1}^{p}\alpha_j h_j}\right)$$

Then $\overline{\alpha}_i = \alpha_i b_i$ and (4.1) becomes

$$\sum_{i=1}^{p} \alpha_i b_i \log b_i \geq o.$$

Set $g(b) = b\log b$. Since g is strictly convex we have

$$\sum_{i=1}^{p}\alpha_i g(b_i) \geq g(\sum_{i=1}^{p}\alpha_i b_i),$$

with equality if and only if

$$b_j = \sum_{i=1}^{p}\alpha_i b_i$$

for $\alpha_j \neq o$. Now,

$$\sum_{i=1}^{p}\alpha_i b_i = \sum_{i=1}^{p}\overline{\alpha}_i = 1.$$

Thus,

$$\sum_{i=1}^{p} \alpha_i g(b_i) \geq g(1) = o$$

with equality if and only if $b_j = 1$ for $\alpha_j \neq o$; that is, if and only if $\overline{\alpha}_j = \alpha_j b_j = \alpha_j$ for all α_j. This completes the proof.

Lemma 4.2: $E(\sum_{i=1}^{p} f_i(\alpha)\log f_i(\overline{\alpha})) \leq E(\sum_{i=1}^{p} f_i(\alpha)\log f_i(\alpha))$.

Proof: Recalling that $g(b) = b\log b$ is convex and that

$$\sum_{i=1}^{p} f_i(\alpha) = \sum_{i=1}^{p} f_i(\overline{\alpha}) = 1,$$

we have that

$$\sum_{i=1}^{p} f_i(\overline{\alpha}) \frac{f_i(\alpha)}{f_i(\overline{\alpha})} \log \frac{f_i(\alpha)}{f_i(\overline{\alpha})} \geq o.$$

That is,

$$\sum_{i=1}^{p} f_i(\alpha)\log f_i(\alpha) \geq \sum_{i=1}^{p} f_i(\alpha)\log f_i(\overline{\alpha}) \tag{4.2}$$

The proof is completed upon taking expectations of both sides of (4.2).

Proof of Theorem 4.1:

$$L(\overline{\alpha}) - L(\alpha) = E(\sum_{i=1}^{p} f_i(\alpha)(\log(\overline{\alpha}_i h_i) - \log f_i(\overline{\alpha})))$$

$$- E(\sum_{i=1}^{p} f_i(\alpha)(\log(\alpha_i h_i) - \log f_i(\alpha)))$$

$$= \sum_{i=1}^{p} E(f_i(\alpha)\log(\overline{\alpha}_i h_i)) - \sum_{i=1}^{p} E(f_i(\alpha)\log(\alpha_i h_i))$$

$$+ E(\sum_{i=1}^{p} f_i(\alpha)\log f_i(\alpha)) - E(\sum_{i=1}^{p} f_i(\alpha)\log f_i(\overline{\alpha}))$$

The theorem follows from lemmas 4.1 and 4.2.

Theorem 4.2 (Peters, Coberly): α maximizes L on A if and only if

$$E\left(\frac{h_i}{\sum_{j=1}^{p} \alpha_j h_j}\right) \le 1$$

for $i = 1, \ldots, p$, with equality if $\alpha_i > 0$.

Proof: L is concave and A is a convex, compact set. Hence α maximizes L on A if and only if

$$\langle \nabla L(\alpha), \beta - \alpha \rangle \le 0 \qquad (4.3)$$

for all $\beta \in A$, with equality if α is in the relative interior of A.

Note that the preceding statement still holds when A is replaced by a "face" of A (that is, a set of points satisfying the constraint(s) $\alpha_j = o$ for one or more fixed indices j, in addition to the constraints imposed on A). Since A contains the standard basis vectors for R^p, (4.3) is equivalent to

$$\frac{\partial L}{\partial \alpha_i}(\alpha) \le \langle \nabla L(\alpha), \alpha \rangle, \; i = 1, \ldots, p$$

Now

$$\langle \nabla L(\alpha), \alpha \rangle = E(\sum_{i=1}^{p} f_i(\alpha)) = 1.$$

Thus, α maximizes L on A if and only if

$$E\left(\frac{h_i}{\sum_{j=1}^{p} \alpha_j h_j}\right) = \frac{\partial L}{\partial \alpha_i}(\alpha) \le 1$$

for $i = 1, \ldots, p$, with equality if $\alpha_i > o$.

Clearly $\hat{\alpha}$ is the unique maximizer of L on A; moreover $\hat{\alpha}$ is the unique fixed point of F in the relative interior of A. Theorem 4.2 also holds if A is replaced by a face of A. Thus F has at most one fixed point in the relative interior of each face of A. Hence F has finitely many fixed points in A.

Theorem 4.3: If α^1 is in the relative interior of A then (1.1) converges to $\hat{\alpha}$ with probability one.

Proof: Let $\omega \in \Omega_o(f)$. In theorem 4.1 and the comments following theorem 4.2 we have shown that the hypotheses of theorem 3.2 are satisfied. Thus (1.1) converges to a fixed point of F. Assume by way of contradiction that $\alpha^i \to \alpha \neq \hat{\alpha}$. Note that $\alpha_j^i \neq o$ for $j = 1, \ldots, p$ and $i = 1, 2, \ldots$. Define g on $A \times R^q$ by

$$g_i(\beta, x) = \frac{h_i(x)}{\sum_{j=1}^{p} \beta_j h_j(x)} .$$

Then since α is a fixed point of F and $\alpha \neq \hat{\alpha}$ we have by theorem 4.2 that $Eg_k(\alpha, X) > 1$ for some k such that $\alpha_k = o$. Let $\varepsilon = Eg_k(\alpha, X) - 1$ and let n be the least positive integer greater that $N(\varepsilon, 1)$ such that $\alpha_k^{n+i} < D(\varepsilon, 1)$ for all i (here N and D are functions of g). Fix m to be the least positive integer such that $\alpha_k^{n+m} < \frac{1}{n} \alpha_k^n$. Note that $m > n$. Then,

$$\alpha_k^{n+m} = \frac{n}{n+m} \alpha_k^n + \frac{m}{n+m}\left(\frac{1}{m} \sum_{i=n+1}^{n+m} f_k(\alpha^{i-1}, x_i)\right)$$

$$= \frac{n}{n+m} \alpha_k^n + \frac{m}{n+m}\left(\frac{1}{m} \sum_{i=n+1}^{n+m} \alpha_k^{i-1} g_k(\alpha^{i-1}, x_i)\right)$$

$$> \frac{n}{n+m} \alpha_k^{n+m} + \frac{m}{n+m} \alpha_k^{n+m} = \alpha_k^{n+m} ,$$

a contradiction. Thus $\alpha = \hat{\alpha}$.

Bibliography

1. J.R. Blum, "Multidimensional stochastic approximation procedures", Ann. Math. Stat. 25 (1954), pp. 737-744.

2. R. L. Franks and R. P. Marzec, "A theorem on mean-value iterations", Proc. Amer. Math. Soc. 30 (1971), pp. 324-326.

3. W. M. Lawton, "Multidimensional Stochastic approximation using locally contractive functions", Rept. 46 (1975), Contract NAS-9-12777, Dept. of Math., U. of Houston.

4. M. Loéve, Probability Theory, D. Van Nostrand Co., New York, 1963.

5. C. L. Outlaw, "Mean value iteration of nonexpansive mappings in a Banach space", Pacific J. Math. 30 (1969), pp. 747-750.

6. B. C. Peters and W. A. Coberly, "The numerical evaluation of the maximum-likelihood estimate of mixture proportions," Commun. Stat - Theor. Meth. A5 (12) (1976), pp. 1127 - 1135.

7. H. Robbins and S. Monro, "A stochastic approximation method", Ann. Math. Stat. 22 (1951), pp. 400-407.

8. H. Walker, private communication.

9. T. Y. Young and G. Coraluppi, "Stochastic estimation of normal density functions using an information criterion", IEEE Trans. Information Theory IT-16 (1970), pp. 258-263.