CR 151 234

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

(NASA-CR-151234) A STOCHASTIC APPROXIMATION ALGORITHM FOR ESTIMATING MIXTURE PROPORTIONS (Houston Univ.) 7 p HC A02/MF A01 CSCL 12A

N77-19825

Unclas
G3/65 20570

A STOCHASTIC APPROXIMATION
ALGORITHM FOR ESTIMATING
MIXTURE PROPORTIONS
BY JAMES SPARRA
REPORT #60 NOVEMBER 1976

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-15000

A Stochastic Approximation Algorithm for Estimating Mixture Proportions

James Sparra

University of Houston
Department of Mathematics
Houston, Texas

Report #60
NASA Contract NAS-9-15000

November 1976

A Stochastic Approximation Algorithm for Estimating Mixture Proportions

by

James Sparra

1. Summary. A stochastic approximation algorithm for estimating the proportions in a mixture of normal densities is presented. The algorithm is shown to converge to the true proportions in the case of a mixture of two normal densities.

2. Introduction. Let $\Lambda = \{\alpha \in R^m : \alpha_i > 0$ and $\sum_{i=1}^{m} \alpha_i = 1\}$. For each i, $i = 1,\ldots,m$, let μ_i be an element of R^n and Σ_i be a positive definite real symmetric $n \times n$ matrix. Let X be a random variable with values in R^n and with density function.

$$p(\hat{\alpha},x) = \sum_{i=1}^{m} \hat{\alpha}_i p_i(x), \quad \text{for} \quad x \in R^n$$

where $\hat{\alpha} \in \Lambda$ and

$$p_i(x) = (2\Pi)^{-n/2}|\Sigma_i|^{-1/2}\exp\{-\tfrac{1}{2}(x-\mu_i)^T\Sigma_i{}^{-1}(x-\mu_i)\}$$

for each $i = 1,\ldots,m$.

We assume that $\hat{\alpha}$ is not known but that μ_i and Σ_i are known for $i = 1,\ldots,m$. An algorithm for estimating $\hat{\alpha}$ will be presented in part 3 of this paper and in part 4 the algorithm will be shown to converge to $\hat{\alpha}$ in mean square and with probability 1 in the case where $m = 2$.

3. The Algorithm. Let $\{x_k\}_{k=o}^{\infty}$ be a sequence of observations on X. Let $\alpha^o \in \Lambda$. For $n \geq 0$ define α^{n+1} by

$$\alpha_i^{n+1} = \alpha_i^n - c_n(\alpha_i^n - \frac{\alpha_i^n p_i(x_n)}{p_{\alpha^n}(x_n)}),$$

where

$$p_{\alpha^n}(x_n) = \sum_{i=1}^{m} \alpha_i^n p_i(x_n)$$

and $\{c_k\}_{k=o}^{\infty}$ is a sequence of positive numbers such that

$$\sum_{k=o}^{\infty} c_k = \infty \quad \text{and} \quad \sum_{k=o}^{\infty} c_k^2 < \infty .$$

We note that each iterate is in Λ and that, since X is a random variable, each iterate may itself be considered a random variable.

4. Convergence of the Algorithm.

Theorem: If $\hat{\alpha} \in R^2$ then the algorithm described in part 3 converges to $\hat{\alpha}$ in mean square and with probability 1.

Proof: We refer the reader to the algorithm described in [1,pp. 332-333] and to the proof of convergence given in [1,pp. 350-352]. The applicability of the theorem given there is clear if we let $f(\alpha) = E(Z_\alpha)$, for each $\alpha \in \Lambda$, where

$$(Z_\alpha)_i = \alpha_i - \frac{\alpha_i(p_i \circ X)}{p_\alpha \circ X} .$$

In order to show convergence we must show that conditions (A1)-(A3) in [1,pp. 332-333] are satisfied. First we note that

$$f(\alpha) = (\alpha_1 - \alpha_1 g_1(\alpha_1),\ \alpha_2 - \alpha_2 g_2(\alpha_2))$$

where

$$g_1(\alpha_1) = \int_{R^n} \frac{p_1(x)}{\alpha_1 p_1(x) + (1-\alpha_1)p_2(x)}\ p_{\hat{\alpha}}(x)dx$$

and

$$g_2(\alpha_2) = \int_{R^n} \frac{p_2(x)}{(1-\alpha_2)p_1(x) + \alpha_2 p_2(x)}\ p_{\hat{\alpha}}(x)dx.$$

Further, we note that

$$\frac{d^2 g_1(\alpha_1)}{d\alpha_1^2} = \int_{R^n} \frac{p_1(x)[p_1(x) - p_2(x)]^2}{[\alpha_1 p_1(x) + (1-\alpha_1)p_2(x)]^3} \cdot p_{\hat{\alpha}}(x)dx > 0$$

and

$$\frac{d^2 g_2(\alpha_2)}{d\alpha_2^2} = \int_{R^n} \frac{p_2(x)[p_2(x) - p_1(x)]^2}{[(1-\alpha_2)p_1(x) + \alpha_2 p_2(x)]^3} \cdot p_{\hat{\alpha}}(x)dx > 0.$$

Now, $g_1(\hat{\alpha}_1) = 1$ and $g_1(1) = 1$. So, since g_1 has positive second derivative we have that $g_1(\alpha_1) < 1$ if $\alpha_1 \in (\hat{\alpha}_1, 1)$ and $g_1(\alpha_1) > 1$ if $\alpha_1 \in (0, \hat{\alpha}_1)$.

Similarly, $g_2(\hat{\alpha}_2) = 1$ and $g_2(1) = 1$ and $g_2(\alpha_2) < 1$ if $\alpha_2 \in (\hat{\alpha}_2, 1)$ and $g_2(\alpha_2) > 1$ if $\alpha_2 \in (0, \hat{\alpha}_2)$.

We now show that (A1)-(A3) are satisfied: Let $\alpha \in A$. Then

(A1) $f(\alpha) = 0$ iff $g_1(\alpha_1) = 1 = g_2(\alpha_2)$ iff $\alpha = \hat{\alpha}$.

(A2) $(\alpha-\hat{\alpha})^T f(\alpha) = (\alpha_1-\hat{\alpha}_1)(\alpha_1-\alpha_1 g_1(\alpha_1)) + (\alpha_2-\hat{\alpha}_2)(\alpha_2 - \alpha_2 g_2(\alpha_2))$.

If $\alpha_1 > \hat{\alpha}_1$ then $g_1(\alpha_1) < 1$ and $(\alpha_1-\alpha_1 g_1(\alpha_1)) > 0$. Then also $\alpha_2 < \hat{\alpha}_2$ and $g_2(\alpha_2) > 1$ and $(\alpha_2-\alpha_2 g_2(\alpha_2)) < 0$. Thus, if $\alpha_1 > \hat{\alpha}_1$ then $(\alpha-\hat{\alpha})^T f(\alpha) > 0$. Similarly, if $\alpha_1 < \hat{\alpha}_1$ then $(\alpha-\hat{\alpha})^T f(\alpha) > 0$. Thus, A2 is satisfied in any closed, convex subset of A.

(A3)

$$E(||Z_\alpha||^2) = \sum_{i=1}^{2} (\alpha_i^2 - 2 \int_{R^n} \frac{\alpha_i^2 p_i(x)}{p_\alpha(x)} \cdot p_{\hat{\alpha}}(x)dx + \int_{R^n} \left(\frac{\alpha_i p_i(x)}{p_\alpha(x)}\right)^2 \cdot p_{\hat{\alpha}}(x)dx)$$

Now, we note that each term in the ith summand, $i = 1,2$, is less than 1 so that there is an $h > 0$ such that $E(||Z_\alpha||^2) < h$ for all $\alpha \in A$ and A3 is satisfied.

Bibliography

1. C.C. Blaydon, K.S. Fu, and R.L. Kashyap, "Stochastic Approximation", *Adaptive, Learning and Pattern Recognition Systems*, Academic Press, New York and London, 1970, Edited by J.M. Mendel and K.S. Fu.