DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

NASA CR-160266

(NASA-CR-160266) A MODIFICATION OF THE LIKELIHOOD EQUATIONS FOR NORMAL CLASSES WITH FIELD STRUCTURE, PRELIMINARY REPORT (Houston Univ.) 10 p HC A02/MF A01 CSCL 12A

N79-27894

Unclas
G3/64 29202

A MODIFICATION OF THE LIKELIHOOD EQUATIONS FOR NORMAL CLASSES WITH FIELD STRUCTURE PRELIMINARY REPORT

BY
CHARLES PETERS

REPORT 73 JUNE 1979

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-15543

HOUSTON, TEXAS 77004

A MODIFICATION OF THE LIKELIHOOD EQUATIONS FOR NORMAL CLASSES WITH FIELD STRUCTURE. (PRELIMINARY REPORT)

by

Charles Peters
Department of Mathematics
University of Houston
Houston, Texas

Report #73
NAS-9-15543 Mod 2S
June 1979

1. Introduction

A possible objection to the use of UHMLE or CLASSY in conjunction with AMOEBA is that both these algorithms ignore the association of pixels in fields. Indeed AMOEBA is based on the explicit assumption that pixels in the same field represent the same real class [1], while the assumptions underlying the maximum likelihood algorithms imply that the classification of a pixel is independent of the classification of other pixels. In this report a statistical model based on normal mixtures is proposed which takes into account the organization of LANDSAT agricultural data into fields which are homogeneous as to crop type. Likelihood equations for the parameters of the model are derived which may be solved iteratively as in UHMLE.

2. The Model

We assume that the data elements (pixel data vectors) are real n-vectors each from one of the statistical populations $\Pi_1, \ldots, \Pi_m$ with n-variate density functions $p(x|\Pi_\ell)$, $\ell = 1, \ldots, m$. We assume that the data is organized into sets (fields) $F_1, \ldots, F_p$, where F_j has N_j data elements which have been previously ordered in some arbitrary fashion so that the data elements in F_j form a nN_j-dimensional vector denoted by $X_j = \begin{matrix} X_{j1} \\ \vdots \\ X_{jN_j} \end{matrix}$

Define random variables $\{\theta_{jk} \in \{1, \ldots, m\} | j=1, \ldots, p; k=1; \ldots, N_j\}$ by $\theta_{jk}=\ell$ if and only if X_{jk} is from Π_ℓ. We assume that all the observations from F_j are from the same class, sc that we may write $\theta_{jk} = \theta_j$ for all $j = 1, \ldots, p$; $k, \ell = 1, \ldots, N_j$. Finally, we assume that $(X_1, \theta_1), \ldots, (X_p, \theta_p)$ are independent, that the θ_j's are identically distributed, that $\alpha_\ell = \text{Prob}\,[\theta_j = \ell] > 0$ and that $\sum_{\ell=1}^{m} \alpha_\ell = 1$. Under the stated assumptions, the joint density of $X_1, \ldots, X_p$ is $p(x_1, \ldots, x_p) = \prod_{j=1}^{p} \sum_{\ell=1}^{m} \alpha_\ell\, p\,(x_j)$, where $p_\ell(x_j) = p(x_{j1}, \ldots, x_{jNj} | \theta_j = \ell)$ is the joint density of the elements of F_j given that F_j represents class Π_ℓ.

Let $N = N_1 + \quad + N_p$ and for each ℓ let M_ℓ denote the total number of the N observations X_{jk} which come from class Π_ℓ. The following proposition shows that with reasonable restrictions on the field sized N_j the values of $\{M_\ell : \ell = 1, \ldots, m\}$ can be inferred from a knowledge of the parameters α_ℓ. Thus, acreage estimates of the classes can be derived from estimates of the parameters α_ℓ.

Proposition 1: (a) $E(M_\ell) = \alpha_\ell N$

(b) $\frac{M_\ell}{N} \to \alpha_\ell$ in probability as $p \to \infty$ if and only if

$$\lim_{p \to \infty} \frac{1}{N^2} \sum_{j=1}^{p} N_j^2 = 0.$$

(c) If $\sum_{j=1}^{\infty} N_j^2 \Big/ j^2 < \infty$, then $\frac{M_\ell}{N} \to \alpha_\ell$ almost surely.

Proof: (a) Write $M_\ell = \sum_{j=1}^{p} \sum_{k=1}^{N_j} \chi_{\underline{\ell}}\ (\theta_{jk})$

$$= \sum_{j=1}^{p} N_j\ \chi_\ell\ (\theta_j)$$

where $\chi_\ell(r) = \begin{cases} 1 & r = \ell \\ 0 & r \neq \ell \end{cases}$

Then $E(M_\ell) = \sum_{j=1}^{p} N_j \, E(\chi_\ell(\theta_j)) = \sum_{j=1}^{p} N_j \, \alpha_\ell = N\alpha_\ell$.

(b) Since $\frac{M_\ell}{N} - \alpha_\ell = \frac{M_\ell}{N} - \left(E \frac{M_\ell}{N}\right)$ is bounded, it converges to zero in probability iff var $\frac{M_\ell}{N} \to 0$ as $p \to \infty$. Since the terms $N_j \chi_\ell(\theta_j)$ are independent,

$$\text{var}\left(\frac{M_\ell}{N}\right) = \frac{1}{N^2} \sum_{j=1}^{p} N_j^2 \ \text{var}\,(\chi_\ell(\theta_j)) = \frac{1}{N^2} \sum_{j=1}^{p} N_j^2 \, \alpha_\ell(1-\alpha_\ell).$$

The conclusion follows.

(c) The assertion follows immediately from Kolmogorov's version of the strong law of large numbers [3].

3. Maximum Likelihood Estimation of the Parameters

In this section we suppose that the class conditional densities $p(x|\Pi_\ell)$ of the data elements X_{jk} are n-variate normal $N(x:\mu_\ell,\Sigma_\ell)$ and that $\{X_{jk}: k=1, \ldots, N_1\}$ are class conditionally independent; i.e., that

$$P_\ell(x_j) = \prod_{k=1}^{N_j} N(x_{jk};\, \mu_\ell,\, \Sigma_\ell).$$

for $j=1, \ldots, p$. In this case the joint density of $X_1, \ldots, X_p$,

$$p(x_1, \ldots, x_p) = \prod_{j=1}^{p} \sum_{\ell=1}^{m} \alpha_\ell \prod_{k=1}^{N_j} N(x_{jk}; \mu_\ell, \Sigma_\ell),$$

is parametrized by $\{(\alpha_\ell, \mu_\ell, \Sigma_\ell)|\ell=1, \ldots, m\}$ where $\alpha_\ell \geq 0$, $\Sigma\, \alpha_\ell = 1$, $\mu_\ell \epsilon R^n$, and Σ_ℓ is a real n x n positive definite symmetric matrix. Whenever a density is evaluated using estimates of its parameters, we denote it, e.g., by $\hat{p}(x_1, \ldots, x_p)$. By a <u>maximum likelihood estimate</u> (MLE) of the parameters $\{(\alpha_\ell, \mu_\ell, \Sigma_\ell)\}$ we mean an element $\{(\hat{\alpha}_\ell, \hat{\mu}_\ell, \hat{\Sigma}_\ell)|\ell=1, \ldots, m\}$ of the parameter set which locally maximizes $\hat{p}(X_1, \ldots, X_p)$. By arguments similar to those used in [2], the following necessary conditions for a MLE are derived.

1) $$\frac{1}{p} \sum_{j=1}^{p} \frac{\hat{p}_\ell(X_j)}{\hat{p}(X_j)} \leq 1$$ with equality when $\alpha_\ell > 0$

2) $$\hat{\mu}_\ell = \sum_{j=1}^{p} \frac{N_j \hat{p}_\ell(X_j)}{\hat{p}(X_j)} \overline{X}_j \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{p}_\ell(X_j)}{\hat{p}(X_j)}$$

3) $$\hat{\Sigma}_\ell = \sum_{j=1}^{p} \frac{\hat{p}_\ell(X_j)}{\hat{p}(X_j)} \sum_{k=1}^{N_j} (X_{jk}-\hat{\mu}_\ell)(X_{jk}-\hat{\mu}_\ell)^T \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{p}_\ell(X_j)}{\hat{p}(X_j)}$$

In equation (2) $\overline{X}_j = \frac{1}{N_j} \sum_{k=1}^{N_j} X_{jk}$ is the mean of the jth field observations.

By multiplying (1) by $\hat{\alpha}_\ell$ we obtain

4) $$\hat{\alpha}_\ell = \frac{1}{p} \sum_{j=1}^{p} \frac{\alpha_\ell \hat{p}_\ell(X_j)}{\hat{p}(X_j)}$$

which, together with (2) and (3) suggests an iterative procedure for solution of the likelihood equations (2) - (4) analogous to that used in UHMLE [2]. However, the likelihood equations can be considerably simplified by observing that the sequence $(\overline{X}_1, S_1), \ldots, (\overline{X}_p, S_p)$, is a sufficient statistic for the model, where S_j is the sample scatter matrix of the jth field:

$$S_j = \sum_{k=1}^{N_j} (X_{jk} - \overline{X}_j)(X_{jk} - \overline{X}_j)^T.$$

Equation (3) may be rewritten

$$5) \qquad \hat{\Sigma}_\ell = \sum_{j=1}^{p} \frac{\hat{p}_\ell(X_j)}{\hat{p}(X_j)} S_j \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{p}_\ell(X_j)}{\hat{p}(X_j)}$$

$$+ \sum_{j=1}^{p} \frac{N_j \hat{p}_\ell(X_j)}{\hat{p}(X_j)} (\overline{X}_j - \hat{\mu}_\ell)(\overline{X}_j - \hat{\mu}_\ell)^T \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{p}_\ell(X_j)}{\hat{p}(X_j)}$$

The sufficiency of $\{(\overline{X}_j, S_j)\}_{j=1}^{p}$ implies that

$$\frac{\hat{p}_\ell(X_j)}{\hat{p}(X_j)} = \frac{\hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}$$

where $\hat{q}_\ell(\overline{X}_j, S_j)$ is the estimated joint density of $\overline{X}_j$ and S_j given that F_j represents class ℓ and $\hat{q}(\overline{X}_j, S_j) = \sum_{\ell=1}^{m} \hat{\alpha}_\ell \hat{q}_\ell(\overline{X}_j, S_j)$. The joint density

$\hat{q}_\ell(\overline{X}_j, S_j)$ may be expressed as

$$q_\ell(\overline{X}_j, S_j) = N_n\,(\overline{X}_j;\ \hat{\mu}_\ell, \frac{1}{N_j}\ \hat{\Sigma}_\ell)\ W_n(S_j;\ N_j-1\,,\ \hat{\Sigma}_\ell)$$

where $N_n\,(\overline{X}_j;\ \hat{\mu}_\ell, \frac{1}{N_j}\ \hat{\Sigma}_\ell)$ is the n-variate normal density of $\overline{X}_j$ and $W_n(S_j;\ N_{j-1}, \hat{\Sigma}_\ell)$ is the Wishart density of S_j with N_j-1 degrees of freedom [3]. Thus the likelihood equations may be written as

$$6)\qquad \hat{\alpha}_\ell = \frac{1}{p} \sum_{j=1}^{p} \frac{\hat{\alpha}_\ell \hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}$$

$$7)\qquad \hat{\mu}_\ell = \sum_{j=1}^{p} \frac{N_j \hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}\ \overline{X}_j \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}$$

$$8)\qquad \hat{\Sigma}_\ell = \sum_{j=1}^{p} \frac{\hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}\ S_j \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}$$

$$+ \sum_{j=1}^{p} \frac{N_j \hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}\ (\overline{X}_j - \hat{\mu}_\ell)(\overline{X}_j - \hat{\mu}_\ell)^T \Bigg/ \sum_{j=1}^{p} \frac{N_j \hat{q}_\ell(\overline{X}_j, S_j)}{\hat{q}(\overline{X}_j, S_j)}$$

Equations (6) - (8) are to be used as the basis of the iteration procedure. Indeed when each $N_j = 1$ they reduce to the likelihood equations employed in UHMLE.

4. Concluding Remarks.

The questions of the existence of a consistent MLE as $p \to \infty$ and the local convergence of the iterative procedure will be addressed in a future report. We remark that the standard consistency results of Cramer, Chanda, and Wald (see [2] for references) are not directly applicable since the $(\bar{X}_j, S_j)$ are not identically distributed. Numerical results will also be reported at a later date.

REFERENCES

1. J. Bryant, "On the clustering of multidimensional pictorial data", to appear in Pattern Recognition.

2. B. C. Peters, Jr. and H. F. Walker, "An iterative procedure for obtaining maximum likelihood estimates of the parameters for a mixture of normal distributions", SIAM J. Appl. Math., B, 35, 2, Sept. 1978.

3. S. S. Wilks, Mathematical Statistics, Wiley and Sons, New York, 1962.