DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

NASA CR-141882

(NASA-CR-141882) MAXIMUM LIKELIHOOD SIGNATURE ESTIMATION (Houston Univ.) 10 p HC $3.25 CSCL 12A
N75-26740
Unclas
G3/64 27270

MAXIMUM LIKELIHOOD SIGNATURE ESTIMATION
BY HOMER F. WALKER
APRIL 1975

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-12777

3801 CULLEN BLVD.
HOUSTON, TEXAS 77004

Maximum Likelihood Signature Estimation

by

Homer F. Walker
Mathematics Department
University of Houston
Houston, Texas

Report #42
NASA Contract NAS-9-12777

Maximum Likelihood Signature Estimation

Abstract

In this outline, we discuss maximum-likelihood estimates, based on an unlabeled sample of observations, of unknown parameters in a mixture of normal distributions. Several "successive approximation" procedures for obtaining such maximum-likelihood estimates are described. These procedures, which are theoretically justified by the local contractibility of certain maps, are designed to take advantage of good initial estimates of the unknown parameters. It is anticipated that they can be profitably applied to the signature extension problem, in which good initial estimates of the unknown parameters are obtained from segments which are geographically near the segments from which the unlabeled samples are taken. Additional problems to which these methods are applicable include: estimation of proportions and adaptive classification (estimation of mean signatures and covariances)

1. Introduction

Let $\{x_k\}_{k=1,\ldots,N} \subseteq \mathbb{R}^n$ be an unlabeled sample of observations from a mixture of m populations, where each population is normally distributed, and let some (possible empty) subset of the signature parameters $\{\alpha_i, \mu_i, \Sigma_i\}_{i=1,\ldots,m}$ be known. (Here, α_i is the *a priori* probability that a sample observation comes from the i^{th} population; μ_i and Σ_i are, respectively, the mean vector and covariance matrix for observation from the i^{th} population in $\mathbb{R}^n$.) A *maximum-likelihood estimate* of the remaining parameters is a choice of those parameters which maximizes the *log-likelihood function*

$$L = \sum_{k=1}^{N} \log p(x_k).$$

In this expression, p denotes the <u>mixture density function</u>, i.e., for $x \in \mathbb{R}^n$,

$$p(x) = \sum_{i=1}^{m} \alpha_i p_i(x),$$

where

$$p_i(x) = \frac{1}{(2\pi)^{n/2}\ |\Sigma_i|^{1/2}}\ e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}$$

Techniques for obtaining maximum-likelihood estimates of this type have been studied by many authors and are considered by a number of them to be superior in general to other methods of estimating the parameters of a mixture of normal distributions. (See, for example, [2] and [6].) Clearly, L is a differentiable function of the signature parameters to be estimated, and there are many approaches to obtaining a maximum of such a function. We discuss several such approaches, each involving "successive approximation" iterative procedures suggested by the particular form of L.

The iterative procedures to be described in the following are based upon manipulating the gradient of L, with respect to the unknown parameters, and incorporating the resulting expressions in fixed-point equations for the unknown parameters. Some of these iterative schemes have been studied by other authors; others are new. In recent preliminary results of Peters and Walker, convergence of the iterates has been established for initial estimates within a sufficiently small neighborhood of the maximum likelihood estimate. This is accomplished by establishing that the appropriate maps are locally contractive at their fixed points. Consequently, these procedures are well-suited for application to the signature extension problem, whenever "reasonable" initial signature estimates (i.e., those satisfying the contractability condition) can be obtained from segments which are geographically near the segments from which the unlabeled samples are taken. We discuss the application of these schemes to the signature extension problem and other problems.

2. The likelihood equations.

The procedures require (for a given unlabeled sample) the calculation of the partial derivatives of the log-likelihood function with respect to the signature parameters. Equating the resulting partial derivatives to zero (the necessary extremum condition), a straight-forward calculation yields the likelihood equations:

$$\left.\begin{aligned}
&(1.a) \quad && \alpha_i = \frac{\alpha_i}{N} \sum_{k=1}^{N} \frac{p_i(x_k)}{p(x_k)} \\
&(1.b) \quad && \mu_i = \left\{ \frac{1}{N} \sum_{k=1}^{N} x_k \frac{p_i(x_k)}{p(x_k)} \right\} \Big/ \left\{ \frac{1}{N} \sum_{k=1}^{N} \frac{p_i(x_k)}{p(x_k)} \right\} \\
&(1.c) \quad && \Sigma_i = \left\{ \frac{1}{N} \sum_{k=1}^{N} (x_k-\mu_i)(x_k-\mu_i)^T \frac{p_i(x_k)}{p(x_k)} \right\} \Big/ \left\{ \frac{1}{N} \sum_{k=1}^{N} \frac{p_i(x_k)}{p(x_k)} \right\}
\end{aligned}\right\} \quad i = 1,\ldots,m$$

In the following, we will assume that a solution of any subset of the likelihood equations is a maximum likelihood estimate of the corresponding signature parameters. For example, if a set of mean vectors and covariance matrices is given, then a maximum-likelihood estimate of the a priori probabilities is a solution $\{\alpha_i\}_{i=1,\ldots,m}$ of the equations (1.a).

3. The natural iterative procedure.

The likelihood equations, as given, suggest the following iterative procedure: Beginning with some initial estimate, obtain successive approximations of the unknown parameters by inserting the preceding approximations in the expressions on the right-hand sides of the appropriate equations (1.a), (1.b), (1.c). Such a scheme for obtaining maximum-likelihood estimates has been investigated by several authors.

Empirical studies in [2], [3], and [4] suggest that this scheme is convergent, even if all the parameters are unknown, and that convergence appears to be particularly fast when the populations are "widely separated".

Unfortunately, the likelihood equations may have several solutions, and the iterates may converge to a solution which is not a maximum likelihood estimate if care is not taken in the choice of an initial estimate. No theoretical evidence of convergence is given in [2], [3], or [4].

Coberly and Peters [1] have proved that, if the unknown parameters are the *a priori* probabilities, then the scheme is *locally convergent*, i.e., convergent for an initial estimate which is sufficiently near a maximum-likelihood estimate. They also report on numerical studies in which the computational feasibility of this procedure is demonstrated. Recent results of Walker state that the scheme is locally convergent when the unknown parameters are the means, whenever there are only two populations (i. e., $m = 2$) or whenever the populations are "widely separated". The local convergence results are all achieved by showing that the expressions on the right-hand sides of the appropriate likelihood equations are locally contractive functions of the unknown parameters (in some vector norm) near a maximum-likelihood estimate.

4. A modified iterative procedure.

We will now describe a modification of the iterative procedure just given for which more extensive local convergence results have been obtained. Because these results are not yet sufficiently complete to allow the covariance matrices to be unknown parameters, we will give the fixed-point equations for the *a priori* probabilities and the mean vectors only. These equations are

$$\left.\begin{aligned} &(2.a) \qquad \alpha_i = (1-\epsilon)\alpha_i + \epsilon \frac{\alpha_i}{N}\sum_{k=1}^{N}\frac{p_i(x_k)}{p(x_k)} \\ &(2.b) \qquad \mu_i = (1-\epsilon)\mu_i + \epsilon\left\{\frac{1}{N}\sum_{k=1}^{N} x_k\frac{p_i(x_k)}{p(x_k)}\right\}\Bigg/\left\{\frac{1}{N}\sum_{k=1}^{N}\frac{p_i(x_k)}{p(x_k)}\right\} \end{aligned}\right\} i = 1, \ldots, m$$

where the scalar parameter ϵ is to be determined. Clearly, these equations are satisfied if and only if the equations (1.a) and (1.b) are satisfied.

If any or all the *a priori* probabilities and the mean vectors are unknown, the the equations (2.a) and (2.b) suggest an iterative scheme analogous to that associated with equations (1.a), (1.b), and (1.c). Recent preliminary results of Peters and Walker state that this scheme is locally convergent when $\epsilon \le \frac{2}{mn+1}$. If only the means are unknown, then the scheme is locally convergent for $\epsilon \le \frac{2}{m}$. As before, these local convergence results are obtained by showing that the appropriate maps in the equations (2.a) and (2.b) satisfy local contractibility conditions near a maximum-likelihood estimate.

This iterative scheme appears to be new, and we feel that it holds considerable promise. It is as easy to implement in practice as the scheme described in the preceding section, and it should converge just as rapidly.

Unless m and n are small, it will be a more practical method of obtaining maximum likelihood estimates than Newton's method or the method of scoring, as described by Kale [5]. Actually, Newton's method and the method of scoring should require fewer iterations for convergence than this scheme. However, the computational effort in these methods may be considerable because the inverse of an $(n+1)m \times (n+1)m$ matrix must be calculated at each iteration. The modified versions of Newton's method and the method of scoring given in [5] will require the same number of iterations as this method. However, there is additional computation involved at each iteration for these modified methods.

5. Applications to signature extension and other problems.

The iterative procedures described in the preceding sections appear well-suited for application to the signature extension problem. This problem has been characterized as that of developing a computationally useful method of "extending signatures" from one sample segment to geographically nearby sample segments. In this context, "extending signatures" means modifying a given set of signature estimates in order to obtain a set which is more useful for the purposes at hand, e.g., classification or estimation of proportions.

Although incomplete, the results given here are encouraging. The numerical studies reported in [1], [2], [3], and [4] demonstrate the computational feasibility of the procedure described in Section 3. The procedure discussed in Section 4 appears to be no more difficult to implement. All results, both empirical and theoretical, obtained so far lead one to believe that the iterative schemes of Sections 3 and 4 will converge to a maximum-likelihood estimate of the signatures whenever "reasonably good" initial signature estimates are provided. (Initial signature estimates which lie within a "radius of local contractibility" of a maximum-likelihood estimate, as suggested by the results of Sections 3 and 4, can certainly be considered "reasonably good".) It is our hope that, in practice, initial signature estimates obtained from "geographically nearly sample segments" will prove to be "reasonably good" in this sense.

In addition, we anticipate that these iterative procedures will be profitably applied to other problems of remote sensing. The iterative scheme for the equations (1.a) is shown in [1] to be a viable approach to the problem of estimation of proportions. Even more reliable proportion estimates should result when the remaining equations (1.b) and (1.c) are also utilized to provide maximum-likelihood estimates of all the signature parameters. (The equations (2.a), (2.b), and their analogues for the covariance matrices can, of course, be used to the same end.) Also, an effective solution to the signature extension problem would appear to be applicable to the adaptive classification problem. Indeed, this problem, that of continually updating population statistics on the basis of incoming samples, is clearly seen to be closely related to the signature extension problem from both a mathematical and a statistical point of view.

6. Future areas of work.

Despite the encouraging results obtained so far concerning the iterative procedures described in the preceding sections, considerable research remains to be done. Generally speaking, the major theoretical problem is to determine the precise circumstances under which these iterative procedures can be expected to converge to maximum-likelihood estimates. More specifically, the local convergence results given here must be extended, hopefully to allow any subset (including all) of the signature parameters to be unknown. In addition, it is necessary to determine quantitatively how near an initial signature estimate must lie to a maximum-likelihood estimate in order for the iterates to converge to a maximum likelihood estimate.

In the absence of more extensive theoretical results, it will be necessary to run many numerical trials, varying the unknown parameter sets, the true population signatures, the starting values and other factors, in order to determine empirically when the iterates can be expected to converge to a maximum-likelihood estimate. Whether or not further theoretical results are obtained, numerical procedures need to be studied with an eye toward optimizing computational efficiency. For example, allowing the covariance matrices to vary arbitrarily in these procedures will require the calculation of their determinants and inverses at each iteration. Hence, one might study iterative schemes in which the covariance matrices are assumed to vary in a particularly simple way, e.g., by multiplication on the left and right by diagonal matrices.

BIBLIOGRAPHY

1. W.A. Coberly and B.C. Peters, "The numerical evaluation of the maximum likelihood estimate of mixture proportions," to appear.

2. N.E. Day, "Estimating the components of a mixture of normal distributions," Biometrika 56 (1969), pp. 463-474.

3. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, 1973.

4. V. Hasselblad, "Estimation of parameters for a mixture of normal distributions," Technometrics 8 (1966), pp. 431-446.

5. B.K. Kale, "On the solution of the likelihood equations by iteration processes. The multiparametric case," Biometrika 49 (1962), pp. 479-486.

6. J.H. Wolfe, "Pattern clustering by multivariate mixture analysis," Multivariate Behavorial Research, 5 (1970), pp. 329-350.