General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

AgRISTARS

E83-10402

SR-62-04376 CR-171677

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

NOVEMBER 1982

A COMPARISON OF MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD TECHNIQUES FOR PROPORTION ESTIMATION

Wayne A. Woodward, William R. Schucany
Hildegard Lindsey, H. L. Gray

AUG 1983 RECEIVED NASA STI FACILITY ACCESS DEPT.

(E83-10402) A COMPARISON OF MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD TECHNIQUES FOR PROPORTION ESTIMATION (Southern Methodist Univ.) 29 p HC A03/MF A01 CSCL 05B G3/43

N83-34395

Unclas 00402

Center for Applied Mathematical and Statistical Research
Southern Methodist University
Dallas, Texas 75275

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SR-62-04376		

4. Title and Subtitle	5. Report Date
A Comparison of Minimum Distance and Maximum Likelihood Techniques for Proportion Estimation	November 1982
	6. Performing Organization Code
7. Author(s) Wayne A. Woodward, William R. Schucany, Hildegard Lindsey, H. L. Gray	8. Performing Organization Report No.
	10. Work Unit No.
9. Performing Organization Name and Address Center for Applied Mathematical and Statistical Research Southern Methodist University Dallas, Texas 75275	11. Contract or Grant No. NAS 9-16438
	13. Type of Report and Period Covered Technical Report
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Johnson Space Center Houston, Texas Tech. Monitor: M. Steib	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The estimation of mixing proportions $p_1, p_2, \ldots, p_m$ in the mixture density $f(x) = \Sigma_{i=1}^{m} p_i f_i(x)$ is often encountered in agricultural remote sensing problems in which case the p_i's usually represent crop proportions. In these remote sensing applications, component densities $f_i(x)$ have typically been assumed to be normally distributed, and parameter estimation has been accomplished using maximum likelihood (ML) techniques. In this paper we examine minimum distance (MD) estimation as an alternative to ML where, in this investigation, both procedures are based upon normal components. Results indicate that ML techniques are superior to MD when component distributions actually are normal, while MD estimation provides better estimates than ML under symmetric departures from normality. When component distributions are not symmetric, however, it is seen that neither of these normal based techniques provides satisfactory results.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Mixture model Maximum likelihood Minimum distance Robustness	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
U	U	24	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

A COMPARISON OF MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD TECHNIQUES FOR PROPORTION ESTIMATION

Wayne A. Woodward, William R. Schucany
Hildegard Lindsey and H. L. Gray

Center for Applied Mathematical and Statistical Research
Southern Methodist University
Dallas, Texas 75275

November 1982

A COMPARISON OF MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD TECHNIQUES FOR PROPORTION ESTIMATION

Wayne A. Woodward, William R. Schucany
Hildegard Lindsey and H. L. Gray

Center for Applied Mathematical and Statistical Research
Southern Methodist University
Dallas, Texas 75275

ABSTRACT

The estimation of mixing proportions $p_1, p_2, \ldots, p_m$ in the mixture density $f(x) = \Sigma_{i=1}^{m} p_i f_i(x)$ is often encountered in agricultural remote sensing problems in which case the p_i's usually represent crop proportions. In these remote sensing applications, component densities $f_i(x)$ have typically been assumed to be normally distributed, and parameter estimation has been accomplished using maximum likelihood (ML) techniques. In this paper we examine minimum distance (MD) estimation as an alternative to ML where, in this investigation, both procedures are based upon normal components. Results indicate that ML techniques are superior to MD when component distributions actually are normal, while MD estimation provides better estimates than ML under symmetric departures from normality. When component distributions are not symmetric, however, it is seen that neither of these normal based techniques provides satisfactory results.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	ESTIMATION IN THE MIXTURE OF NORMALS MODEL	2
III.	STARTING VALUES	6
IV.	SIMULATION RESULTS	9
V.	MIXTURES OF ASYMMETRIC DISTRIBUTIONS	17
VI.	CONCLUDING REMARKS	21
	REFERENCES	22

A COMPARISON OF MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD TECHNIQUES FOR PROPORTION ESTIMATION

Wayne A. Woodward, William R. Schucany,
Hildegard Lindsey, and H. L. Gray
Center for Applied Mathematical and Statistical Research
Southern Methodist University

1. Introduction

A common objective in remote sensing is the estimation of the proportions $p_1, p_2, \ldots, p_m$ in the mixture density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \ldots + p_m f_m(x) \tag{1.1}$$

where m is the number of components(crops) in the mixture and for component $i, f_i(x)$ is a (possibly multivariate) density. In past practice this density has been assumed to be (multivariate) normal with X being the reflected energy in four bands of the light spectrum, certain linear combinations of these readings, or other derived "feature" variables. Generally the parameter estimation has been accomplished using maximum likelihood techniques. In this paper we examine the use of minimum distance estimation as an alternative to maximum likelihood and we will compare the performance of the two estimation techniques when dealing with mixtures of normal and of non-normal densities with varying amounts of separation. We will focus on the mixture of two univariate distributions given by

$$f(x) = pf_1(x) + (1-p)f_2(x) \tag{1.2}$$

We are also assuming that only data from the mixture distribution are available. Other sampling schemes in which training samples from the component distributions are also available have been discussed by Hosmer(1973), Redner(1980), and Hall(1981) among others.

2. Estimation in the Mixture of Normals Model

In this section we will assume that $f_1(x)$ and $f_2(x)$ in (1.2) are normal densities with mean and variance μ_1, σ_1^2 and μ_2, σ_2^2 respectively where it is assumed that all five parameters $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$, and p are unknown. Techniques for estimating these parameters will be discussed.

(a) Maximum Likelihood

Several recent articles have dealt with the problem of obtaining the maximum likelihood estimates of μ_1, σ_1^2, μ_2, σ_2^2, and p (Hasselblad(1966), Day(1969), Wolfe(1970), Hosmer(1975), Fowlkes(1979), Lennington and Rassbach(1979), and Redner(1980).) Since the likelihood function

$$L = f(x_1)f(x_2) \ldots f(x_n) \tag{2.1}$$

where n is the sample size, is not a bounded function in this case (see Day(1969)), the objective in the maximum likelihood approach is to find a local maximum of L. This maximum is usually found by setting the partial derivatives of log(L) with respect to each of the 5 parameters equal to zero and solving the resulting set of equations, called the

likelihood equations. Since closed form solutions of these equations do not exist, they must be solved using iterative techniques. Hasselblad(1966) and Wolfe(1969) suggested that these equations be solved by taking advantage of their fixed point form. Redner(1980) and Redner and Walker(1982) have pointed out that this fixed point technique is essentially an application of the EM algorithm (see Dempster, Laird and Rubin(1977)) with the only difference being that using the EM algorithm, the estimates of σ_1^2 and σ_2^2 at step k involve the updated k^{th} step estimates of μ_1 and μ_2

Fowlkes(1979), on the other hand, maximized the likelihood function directly by utilizing a quasi-Newton method for minimizing -log(L) and found that good starting values were crucial for acceptable performance. Hosmer(1975) stated that using the likelihood equations, starting values were not a serious problem in his experience. In order to determine which of the two techniques seemed preferable in our simulation studies we replicated simulations performed by Fowlkes in which various sets of poor starting values were used to initiate the minimization procedure. We simulated realizations from the mixture utilized by Fowlkes and estimated the parameters using both direct maximization and the EM algorithm. The results of our simulations indicate that the EM algorithm approach is preferable and hence we have used this technique for obtaining MLEs in our simulations.

(b) Minimum Distance

Although ML estimation procedures are known to have certain optimality properties, their sensitivity to violations of the underlying assumptions is also recognized. The development of estimation procedures which perform well even under moderate deviations from assumptions has been a topic of major interest in recent literature. One of these robust procedures which has received recent attention is that of minimum distance(MD) estimation introduced by Wolfowitz(1957). Parr and Schucany(1980), for example, have shown that MD techniques provide robust estimators of the location parameter of a symmetric distribution. Minimum distance estimation has been used for parameter estimation in the mixture model by Choi and Bulgren(1968) and MacDonald(1971) with some success although, to our knowledge, the question of sensitivity to assumptions in this setting has not been addressed. These previous authors assumed that the parameters of the component distributions were known and that only the mixing proportion(s) was to be estimated.

In order to briefly describe minimum distance estimation, we let $X_1, X_2, \ldots, X_n$ denote a random sample from a population with distribution function F and let F_n denote the empirical distribution function, i.e. $F_n(x)=k/n$ where k is the number of observations less than or equal to x. Further, let $\mathcal{H} = \{H_\theta : \theta \in \Omega\}$ denote a family of distributions depending on the possibly vector valued

ORIGINAL PAGE IS
OF POOR QUALITY

parameter θ. The MD estimate of θ is that value of θ for which the distance between F_n and H_θ is minimized. It is not necessary that $F \in \mathcal{H}$. Of course, when a mixture of two normals is used as the projection family, H_θ becomes

$$H_\theta(x) = p \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{y-\mu_1}{\sigma_1}\right)^2} dy + (1-p) \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2} dy.$$

Certain considerations become obvious at this point. First, we must define what we mean by the "distance" between two distributions. Several such distance measures have appeared in the literature. The reader is referred to the article by Parr and Schucany(1980) for a discussion of these measures. For our purposes we have chosen the Cramér-von Mises distance, W^2, between distribution functions G_1 and G_2 which is given by

$$W^2 = \int_{-\infty}^{\infty} [G_1(x) - G_2(x)]^2 dG_2(x).$$

In our setting a computing formula for the Cramér-von Mises distance between F_n and H_θ is given by

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} [H_\theta(Y_i) - \frac{i-.5}{n}]^2 ,$$

where Y_i is the ith order statistic. The similarity between W_n^2 and the sum of squared differences between the empirical distribution function F_n and H_θ used by Choi and Bulgren(1968) should be noted.

Another consideration involves the minimization procedure to be employed in minimizing W_n^2. Parr and

Schucany used the IMSL quasi-Newton algorithm ZXMIN. Our comparisons have shown ,however, that the IMSL routine ZXSSQ which uses Marquardt's(1963) method for minimizing a sum of squares was significantly faster, usually taking no more than half the time required by ZXMIN. In the simulation studies reported in the next section we have used the Marquardt minimization procedure when calculating the MDE. It should be noted that minimization is subject to the constraints $\sigma_1^2 \geq 0$, $\sigma_2^2 \geq 0$, and $0 \leq p \leq 1$. Another finding which deserves mention before proceeding is that similar to the technique we have chosen for calculating the MLE, the MDE has the desirable property that it is relatively insensitive to starting values.

3. Starting Values

In order for the estimators discussed in the previous chapter to be used in practice, starting values for the iterative procedures must be provided. We have chosen to obtain starting values in this two component univariate setting using a partitioning technique which is very easy to implement. In the discussion to follow we will assume, without loss of generality, that $\mu_1 < \mu_2$. This technique involves first obtaining the initial estimate of p, denoted by p_0, and then estimating the remaining four parameters given p_0. Under the current implementation, only the 9 values .1,.2,...,.9 are allowed as possible

values for p_0. For each allowable value of p_0, the sample is divided into two subsamples :

$$Y_1, Y_2, \ldots, Y_{n_1}$$
$$Y_{n_1+1}, Y_{n_1+2}, \ldots, Y_n$$

where Y_i is the ith order statistic and n_1 is np_0 rounded to the nearest integer. The value for p_0 is that value of p for which $p(1-p)(m_1-m_2)^2$ is maximized, where m_j is the sample median of the jth subsample. The criterion used here is a robust counterpart to the classical cluster analysis procedure of selecting the clusters for which the within cluster sum-of-squares is minimized. It is easy to show, however, that the within cluster sum-of-squares is minimized in the two cluster case when $p(1-p)(\bar{x}_1-\bar{x}_2)^2$ is maximized, where $\bar{x}_j$ is the sample mean of cluster j and and $p=n_1/n$ with n_1 the number of sample values placed in cluster 1. Such a clustering is based upon a cut-point, c , for which all sample values below c are assigned to the cluster associated with population 1. It must be observed, however, that due to the overlap between the two mixture distributions, some sample points assigned to cluster 1 may be from population 2 and some observations from population 1 may be in cluster 2. The effect of this truncation of the right tail in population 1 is that the sample mean from cluster 1 is likely to underestimate μ_1 while μ_2 is likely to be overestimated. In addition σ_1^2 and σ_2^2 are likely to be underestimated by s_1^2 and s_2^2. If we

assume that the overlap between the two populations is not too severe, then the sample values in cluster 1 to the left of m_1 are relatively pure observations from population 1 in which case m_1 is a "good" estimate of the population mean in the case of symmetric distributions. This reasoning also indicates that m_1 and m_2 should provide better estimates of μ_1 and μ_2 than would $\bar{x}_1$ and $\bar{x}_2$. In order to estimate the variances of the component distributions we again will depend upon the fact that the values to the left of m_1 and to the right of m_2 are "pure" samples from populations 1 and 2 respectively. Thus, we will use only this portion of the data for estimation of the sample variances. We have used the fact that the interquartile range of a standard normal distribution is .6745, to estimate σ_1^2 by

$$\sigma_1^2(0) = \left(\frac{m_1 - r_1^{(.25)}}{.6745} \right)^2 ,$$

where $r_j^{(q)}$ is the qth percentile from the jth cluster, j=1,2. Similarly, $\sigma_{2(0)}^2 = [(r_2^{(.75)} - m_2)/.6745]^2$.

In the next section we will discuss the results of a major simulation investigation comparing ML and MD estimation. In these simulations the iterative techniques were initiated by the starting values as discussed in the previous paragraph. A preliminary simulation investigated the performance of the starting values described here. In this preliminary study we compared the convergence

initiated from these starting values with that when the iterative procedures are started at the true parameter values. The convergence from these two starts was almost always to the same parameter estimates, a result which held for both the MLE and MDE. For this reason and results to be shown in Section 4, we believe this starting value procedure to be adequate.

4. Simulation Results

In the previous two sections we have discussed ML and MD estimators for the parameters of the mixture of two distributions. In this section we report the results of simulations designed to compare these two estimators when the component distributions are normal and when they are non-normal. In addition we have made our comparisons under varying degrees of separation between the two distributions. All computations were performed on the CDC 6600 at Southern Methodist University.

In our comparison of the MDE and MLE we have begun by comparing their performance when the normality assumption is valid, i.e., when the component distributions actually are normal. We should mention that because of the optimality properties of the MLE we would expect that the MLE would be superior in this situation. Since in practice the validity of the normality assumption is subject to question, we are also very interested in the performance of the MDE and MLE when the component distributions are

not normal. To this end we have simulated mixtures in which the component distributions are distributed as a Student's t with 4 degrees of freedom. We simulated 500 samples of size n=100 from mixtures of normal and of t(4) components for each of the following parameter configurations:

Mixing proportion

.25

.50

.75

Variances

$\sigma_1^2 = \sigma_2^2$

$\sigma_1^2 = 2\sigma_2^2$

The nature of the mixture model also depends on the amount of separation between the two component distributions. While, for sufficient separation, the mixture model has a characteristic bimodal shape, Behboodian(1970) has shown, for example, that a sufficient condition for the mixture density (of two normal components) to be unimodal is that $|\mu_1-\mu_2| \leq 2\min(\sigma_1,\sigma_2)$. Of course, in this situation, parameter estimation is difficult.

For purposes of quantifying this separation between the components, we will define a measure of "overlap" between two distributions. Without loss of generality we

assume that population 1 is centered to the left of population 2. We define "overlap" to be the probability of misclassification using the rule:

Classify an observation x as:

population 1 if $x < x_c$

population 2 if $x \geq x_c$,

where x_c is the unique point between μ_1 and μ_2 such that

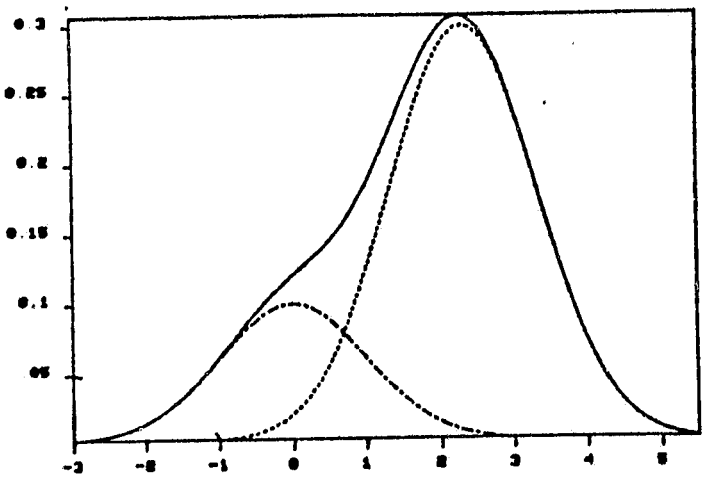
$$pf_1(x_c) = (1-p)f_2(x_c).$$

We have based our current study on "overlaps" of .03 and .10. In Figure 1 we display the mixture densities associated with normal components and $\sigma_1^2 = \sigma_2^2$. For each mixture, the scaled components $pf_1(x)$ and $(1-p)f_2(x)$ are also shown. Note that the densities for p=.75 are not displayed here since when $\sigma_1^2 = \sigma_2^2$, it follows that $f^p(x) = f^{1-p}(\mu_1+\mu_2-x)$ where $f^h(x)$ denotes the mixture density associated with a mixing proportion of h. Thus the shapes of the densities at p=.75 can be inferred from those at p=.25. Likewise, parameter estimation for p=.75 is not included in the results of the simulations when $\sigma_1^2 = \sigma_2^2$.

Although both estmation procedures provide estimates of all 5 of the parameters, only the results for the estimation of p will be shown since the mixing proportion is the parameter of primary interest. In addition, when dealing with the non-normal mixtures, the remaining parameter

ORIGINAL PAGE IS
OF POOR QUALITY

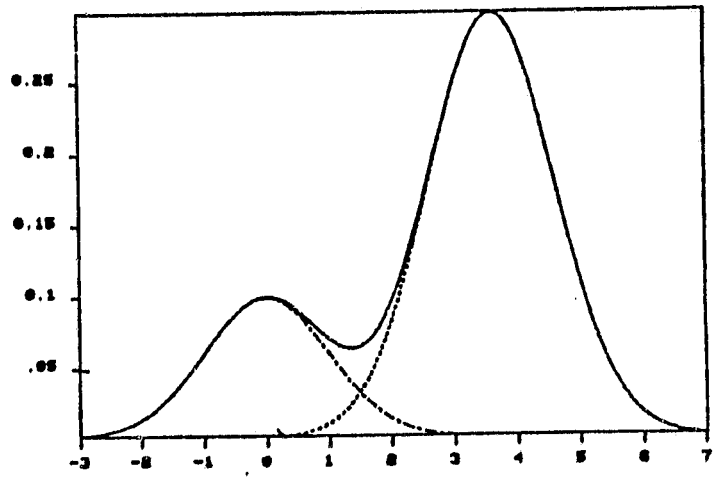
FIGURE 1 - Mixture Densities Associated with Normal Components and $\sigma_1^2 = \sigma_2^2 = 1$



$p = .25$

$\mu_1 = 0,\ \mu_2 = 2.32$

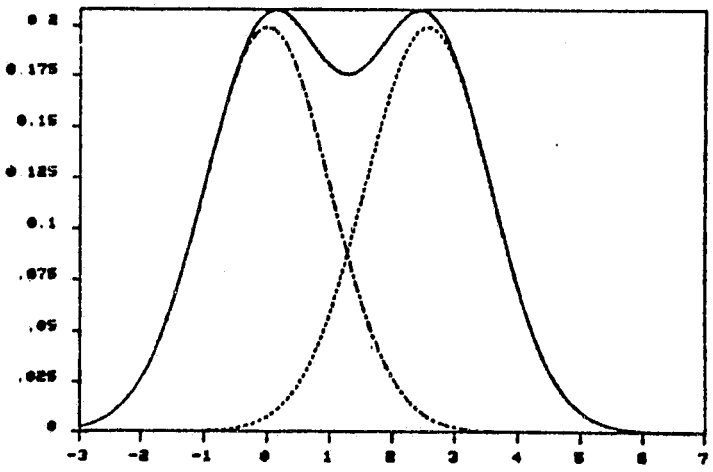
overlap = .10



$p = .25$

$\mu_1 = 0,\ \mu_2 = 3.6$

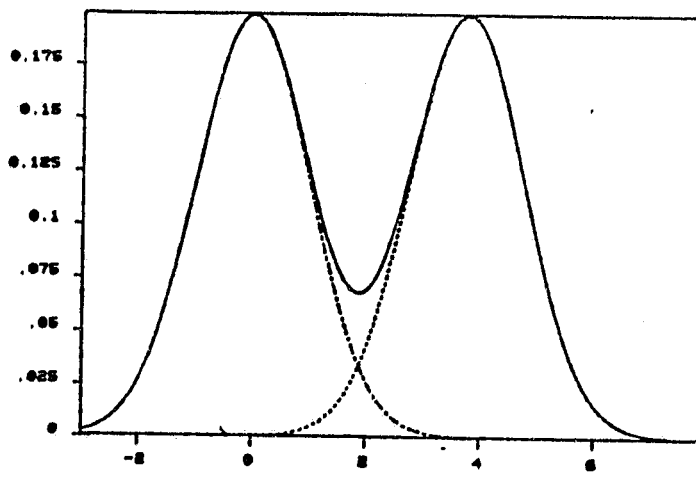
overlap = .03



$p = .5$

$\mu_1 = 0,\ \mu_2 = 2.56$

overlap = .10



$p = .5$

$\mu_1 = 0,\ \mu_2 = 3.76$

overlap = .03

——— mixture -·-·-·-·- $pf_1(x)$ ····· $(1-p)f_2(x)$

estimates often do not have a meaningful interpretation. In these simulations we have used the procedure discussed in the previous section to obtain starting values. It should be noted that although we refer to mixtures of t(4) distributions here, they are actually mixtures of distributions associated with the random variable $T'=aT+b$, where T has a t(4) distribution. These modifications are made in order to obtain the desired separation and variance ratios.

In Table 1 we show the results of the simulation comparing the performance of the MLE and MDE. In particular, let $\hat{p}_i$ denote the estimate of p for the ith sample. Then based upon the simulations, estimates of the bias and MSE are given by:

$$\hat{\text{bias}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)$$

$$\hat{\text{MSE}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)^2 ,$$

where n_s is the number of samples. It should be noted that nMSE is the quantity actually given in the table. In addition, we provide the ratio

$$E = \frac{\text{MSE(MLE)}}{\text{MSE(MDE)}}$$

as an efficiency measure.

Upon viewing the results, it can be seen, as expected, that the bias and MSE associated with the MLE were generally smaller than those for the MDE when the components were

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 1

Simulation Results Comparing MLE and MDE

Sample Size = 100
Number of replications = 500

NORMAL

	Overlap = .10					Overlap = .03				
	Bias		nMSE*		E	Bias		nMSE		E
$\sigma_1^2 = \sigma_2^2$	MLE	MDE	MLE	MDE		MLE	MDE	MLE	MDE	
p = .25	.052	.125	4.26	7.80	.55	.008	.026	.54	1.09	.50
p = .50	.000	.010	3.21	3.86	.83	.000	.001	.38	.42	.90
$\sigma_1^2 = 2\sigma_2^2$										
p = .25	.002	.084	2.25	5.30	.42	.006	.027	.49	.96	.51
p = .50	-.009	.005	2.41	2.79	.86	.009	.008	.42	.44	.95
p = .75	-.086	-.137	4.87	8.36	.58	-.002	-.024	.47	1.08	.44

t(4)

	Overlap = .10					Overlap = .03				
	Bias		nMSE		E	Bias		nMSE		E
$\sigma_1^2 = \sigma_2^2$	MLE	MDE	MLE	MDE		MLE	MDE	MLE	MDE	
p = .25	.096	.104	7.35	6.18	1.19	.029	.020	.88	.44	2.00
p = .50	.015	.004	5.59	1.82	3.07	-.005	.000	.47	.27	1.74
$\sigma_1^2 = 2\sigma_2^2$										
p = .25	.061	.098	4.63	5.20	.89	.044	.029	.95	.61	1.56
p = .50	.028	.022	4.49	1.80	2.49	.010	.001	.55	.30	1.83
p = .75	-.076	-.058	7.84	3.68	2.13	-.012	-.016	.57	.36	1.58

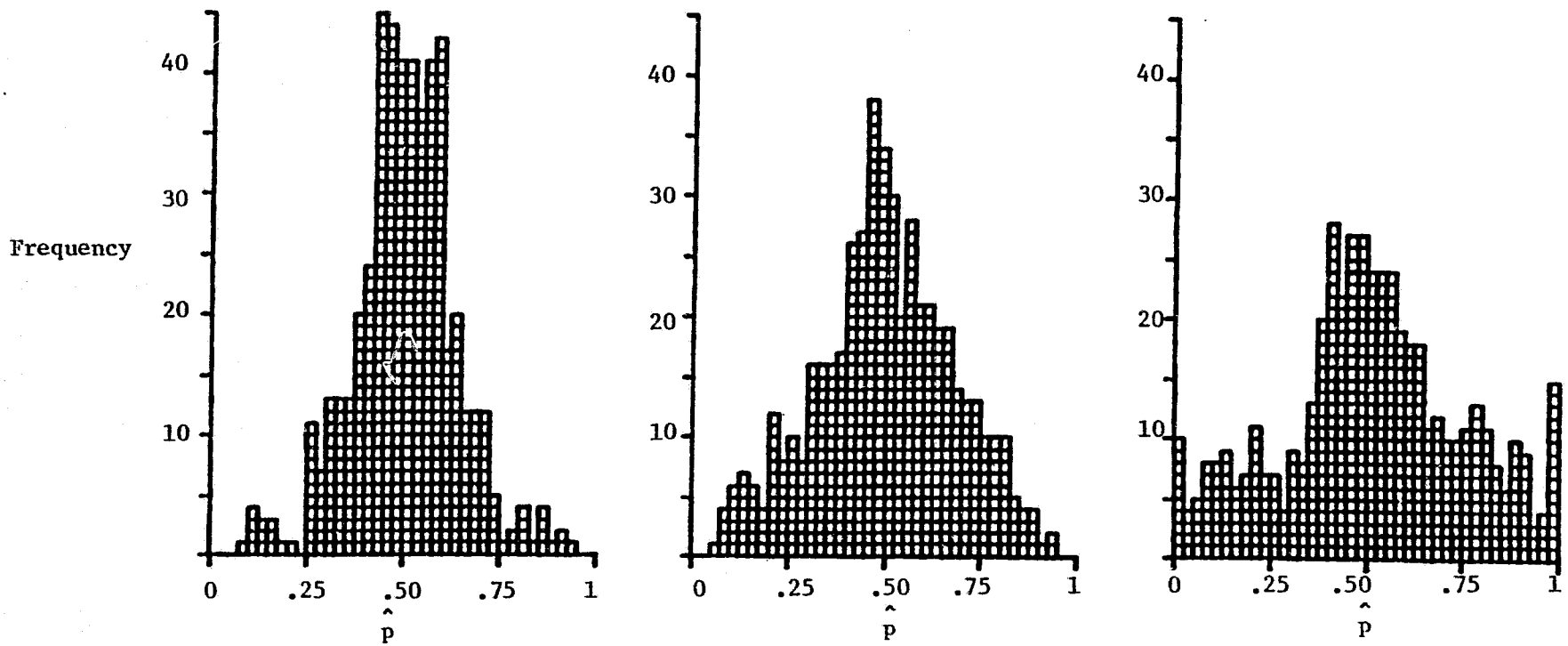
*nMSE = n times the MSE where n = sample size

normally distributed. This relationship between the estimators held for both overlaps. The MLE and MDE were quite similar at p=.5 while for p=.25 and p=.75 the superiority of the MLE is more pronounced.

For the t(4) mixtures the relationship between MDE and MLE is reversed in that the MDE generally has the smaller bias and MSE. The superiority of the MDE in this case is due in part to the heavy tails in the t(4) mixture. The MLE often interpreted an extreme observation as being the only sample value from one of the populations with all remaining observations belonging to the other. Due to the well known singularities associated with a zero variance estimate for a component distribution, Day(1969), we were concerned that the observed behavior of the MLE was due to the fact that we did not constrain the variances away from zero. However, simulation results in which equal variances were assumed (which removes the singularity) and also those which used a penalized MLE suggested by Redner(1980) were very similar to those quoted here.

Although the MSE is a widely used measure among statisticians for assessing the performance of an estimator, the practical implications, for example, of an estimator having an MSE three times larger than that for another estimator, may not be immediately apparent. Recall that each MSE quoted in Table 1 is based upon 500 estimates of p. In order to provide a better appreciation for the practical impact of differences in MSE, in Figure 2 we display histograms of the 500 estimates of p associated with three

Figure 2. Histograms of Estimates of p Based Upon 500 Samples of Size 100 from Mixtures in which p = .5



MSE = .0182 MSE = .0321 MSE = .0559

ORIGINAL PAGE IS OF POOR QUALITY

different MSEs in the table. The true value of p in each case is p=.5. It is obvious that as the MSE increases, the performance of the estimator deteriorates. Notice that the MSE for Figure 2(c) is approximately three times greater than the MSE associated with Figure 2(a), while the MSE for Figure 2(b) is aprroximately twice that for Figure 2(a). Thus, from these histograms, an intuitive feel for efficiency ratios of E=2 and E=3 can be obtained.

A very surprising result is that the starting values obtained using the procedure outlined in Section 3 produced estimators which were competitive with both the MLE and MDE. In fact, for both the normal and t(4) mixtures, the MSEs associated with the starting values were lower than those for the MDE and MLE for every parameter configuration associated with an overlap of .10. At an overlap of .03, however, the starting values estimates were generally poorer than those for the MDE and MLE.

5. Mixtures of Asymmetric Distributions

The simulation results of the previous section focus on the performance of the MLE and MDE under deviations from the assumption of normality. However, the t(4) distribution is symmetric, and recent studies have indicated that there is often a substantial asymmetry in the component distributions for variables of interest in agricultural remote sensing. A Monte Carlo examination of the performance of the MDE and MLE, assuming normal components, when in fact the component

distributions were asymmetric, was performed, and the results of this examination will be discussed in this section.

For purposes of our examination, we simulated mixtures of $\chi^2(9)$ distributions with p=.5. In these simulations the two distributions differed from each other only by a location shift. Actually the component distribution to the left is $\chi^2(9)$ while that to the right is that of a "shifted" $\chi^2(9)$ with origin no longer at 0. This shift was varied to provide overlaps of .01, .05, and .10. Since our estimation procedures involve a normality assumption, we used the means and variances of the two component $\chi^2(9)$ distributions and the true mixing proportions as our starting values. The problem of obtaining starting values from the data in this case is being examined. In Table 2 we display the results of this simulation. Only when the two component distributions were widely separated (overlap=.01) do the two procedures provide reasonable results. However, when the two chi-square distributions are not widely separated, both estimators tend to seriously underestimate p. In Figure 3 we display the three mixture distributions on which these simulations were based. We see there that it is no surprise that the estimate of p is less than .5, especially for p=.10. Both estimation procedures view this as a mixture of normals, and therefore make the reasonable interpretation that the density to the left has a smaller variance and a mixing proportion less than .5. These results point out the impact which skewed distributions can have on the proportion estimation in the

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 2

Simulation Results
Mixtures of $\chi^2(9)$ Components

Sample Size = 100
Number of replications = 200
p = .5

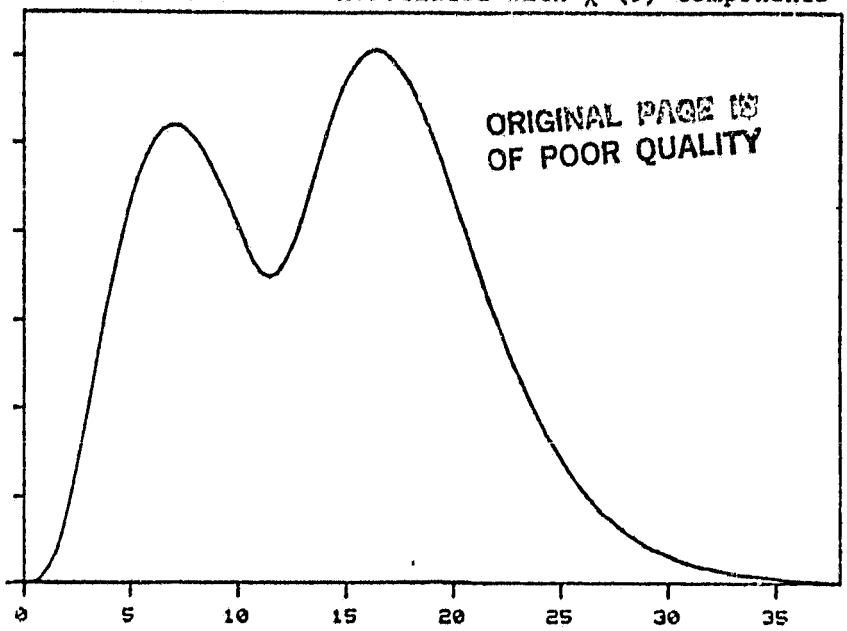
		MLE			MDE		
		$\hat{p}$	Bias	nMSE	$\hat{p}$	Bias	nMSE
	.10	.28	-.22	6.8	.28	-.22	6.6
Overlap	.05	.35	-.15	2.7	.37	-.13	2.3
	.01	.47	-.03	.4	.45	-.05	.5

FIGURE 3

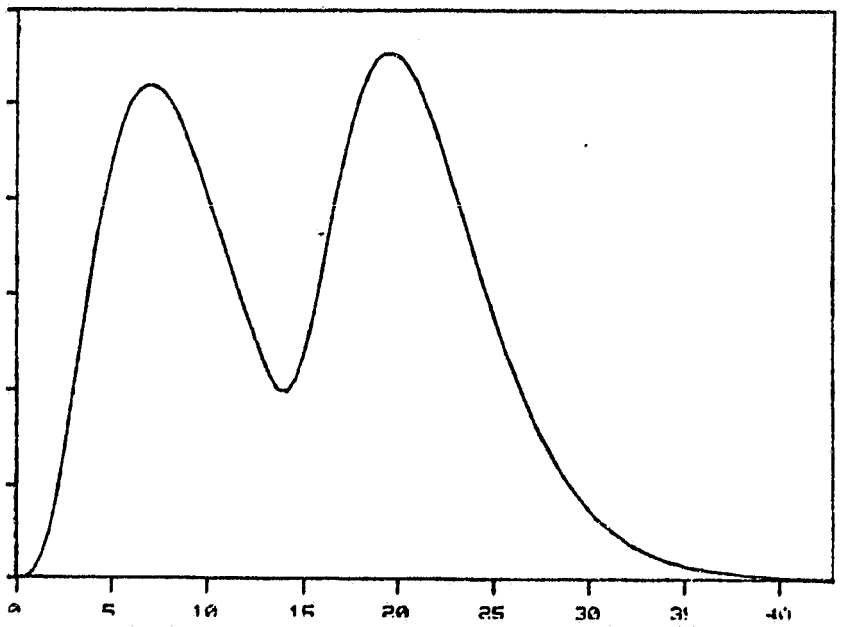
Mixture Densities Associated with $\chi^2(9)$ Components

ORIGINAL PAGE IS
OF POOR QUALITY

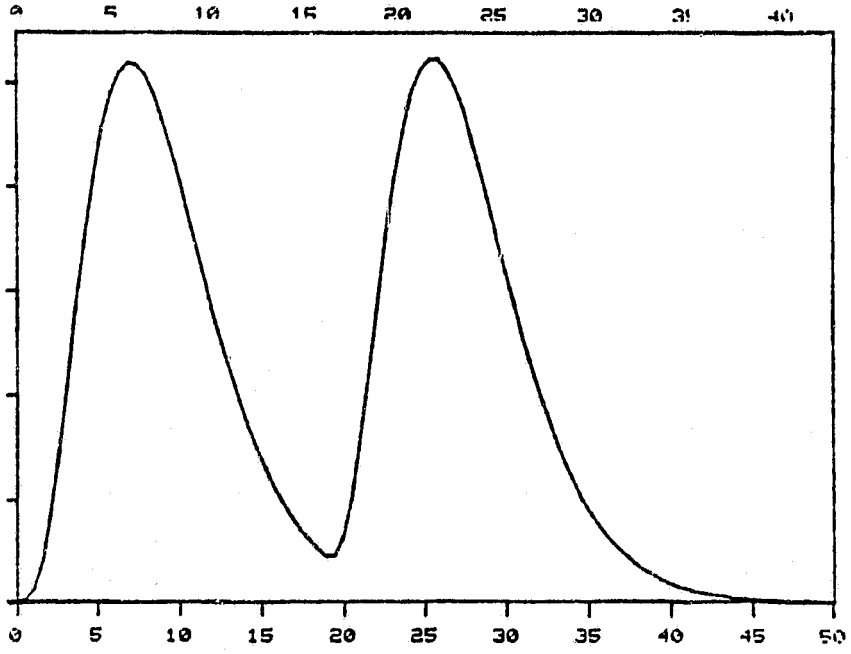
Overlap .10



Overlap .05



Overlap .01



mixture model when normal mixtures are assumed.

Current investigation into this area centers around modifying the estimation procedures by assuming that the underlying component distributions belong to some family of distributions whose members can be either symmetric or asymmetric depending on parameter configurations. At the present time, the Weibull distribution is being examined concerning its usefulness.

6. Concluding Results

We believe that the results of the preceding sections are of sufficient substance to motivate further research in the area of MD estimation in the mixture model. Our results indicate that the MDE is indeed more robust than the MLE in the sense that it is less sensitive to symmetric departures from the underlying assumption of normality of component distributions. Several areas for future investigation have already been identified in addition to the asymmetric components problem discussed in Section 5.

First, simulations similar to the ones presented here should be performed without the assumption of only two populations in the mixture. The performance of the MDE and MLE should be compared when the number of populations is known and larger than two. In addition the applicability of the MDE to the problem of estimating the number of populations also warrants investigation. We plan to examine these possibilities.

Second, the problem of applying the MDE to the multivariate setting is of interest. Preliminary indications are that such an extension will be possible.

Third, the choice of distance measure in the MDE is a topic of interest. Our results are not meant to imply that W is optimal.

Finally, the MDE and MLE must ultimately be compared on real data. Several related practical considerations have not yet been investigated. For example, when applying these estimators to LANDSAT data, the number of iterations allowed must be small due to time constraints. In the simulations described here, these constraints were not imposed and iteration was allowed to continue until convergence was obtained. The performance of the MDE and MLE under convergence restrictions should be examined.

REFERENCES

1. Behboodian, J.(1970). "On the Modes of a Mixture of Two Normal Distributions," Technometrics 12, 131-139.
2. Choi,K. and Bulgren,W. G.(1968). "An Estimation Procedure for Mixtures of Distributions," JRSS-B 30, 444-460.
3. Cohen, A.C.(1967)."Estimation in Mixtures of Two Normal Distributions," Technometrics 9, 15-28.
4. Day, N. E.(1969). "Estimating the Components of a

Mixture of Normal Distributions," Biometrika 56, 463-474.

5. Dempster, A.P., Laird, N.M., and Rubin, D.B.(1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm," JRSS-B 39, 1-38.

6. Everitt, B.S. and Hand, D.J.(1981). Finite Mixture Distributions, Chapman and Hall, London.

7. Fowlkes, E. B.(1979)."Some Methods for Studying the Mixture of two Normal (Lognormal) Distributions," JASA 74, 561-575.

8. Hall,P. (1981). "On the Nonparametric estimation of Mixture Proportions," JRSS-B 43 ,147-156.

9. Hasselblad, V. A.(1966)."Estimation of Parameters for a Mixture of Normal Distributions," Technometrics 8, 431-446.

10. Hosmer, D. W.(1973)."A Comparison of Iterative Maximum Likelihood Estimates of the Parameters of Two Normal Distributions Under Three Different Types of Samples," Biometrika 29, 761-770.

11. Lennington, R. K. and Rassbach, M. E.(1979). "CLASSY-An Adaptive Maximum Likelihood Clustering Algorithm," Proc. Tech. Sessions, The LACIE Symposium, vol. II, 671-690.

12. MacDonald, P.D.M.(1971). "Comment on 'An Estimation Procedure for Mixtures of Distributions' by Choi and Bulgren," JRSS-B 33, 326-329.

13. Parr, W. C. and Schucany W. R.(1980)."Minimum Distance and Robust Estimation,"JASA 75, 616-624.

14. Pearson,K.(1894)."Contribution to the Mathematical Theory of Evolution," Phil trans A 185, 71-110.

15. Redner, R.A.(1980)."Maximum Likelihood Estimation for Mixture Models," JSC-16832, October 1980.

16. Redner, R.A. and Walker, H.F.(1982). "Mixture Densities, Maximum Likelihood, and the EM Algorithm," unpublished manuscript.

17. Wolfe, J.H.(1980). "Pattern Clustering by Multivariate Mixture Analysis," Multivariate Behavioral Research 5. 320-350.

18. Wolfowitz, J.(1957)."The Minimum Distance Method," Annals of Math. Statist. 28, 75-88.