MINIMUM DISTANCE ESTIMATION OF MIXTURE MODEL
PARAMETERS - ASYMPTOTIC RESULTS
AND SIMULATION COMPARISONS
WITH MAXIMUM LIKELIHOOD

by

Wayne A. Woodward, William C. Parr
William R. Schucany, Henry L. Gray

Technical Report No. SMU/DS/TR-178
Department of Statistics ONR Contract

June 1983

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas   75275

MINIMUM DISTANCE ESTIMATION OF MIXTURE MODEL PARAMETERS –
ASYMPTOTIC RESULTS AND SIMULATION COMPARISONS
WITH MAXIMUM LIKELIHOOD

Wayne A. Woodward, William C. Parr,
William R. Schucany, and Henry L. Gray

ABSTRACT

The estimation of mixing proportions in the mixture model is discussed with emphasis on the mixture of two normal components with all five parameters unknown. Simulations are presented which compare minimum distance(MD) and maximum likelihood(ML) estimation of the parameters of this mixture-of-normals model. Some practical issues of implementation of these results are also discussed. Simulation results indicate that ML techniques are superior to MD when component distributions actually are normal, while MD techniques provide better estimates than ML under symmetric departures from component normality. Results are presented which establish strong consistency and asymptotic normality of the MD estimator under conditions which include the mixture-of-normals model. Asymptotic variances and relative efficiencies are obtained for further comparison of the MDE and MLE.

MINIMUM DISTANCE ESTIMATION OF MIXTURE MODEL PARAMETERS - ASYMPTOTIC RESULTS AND SIMULATION COMPARISONS WITH MAXIMUM LIKELIHOOD

Wayne A. Woodward, William C. Parr,
William R. Schucany, and Henry L. Gray

1. Introduction

An important problem in aerospace remote sensing is the estimation of the mixing proportions $p_1, p_2, \ldots, p_m$ in the mixture density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \ldots + p_m f_m(x)$$

where m is the number of components(crops) in the mixture and for component $i, f_i(x)$ is a density. The variable of interest, X, is some measurement such as the reflected energy in four bands of the light spectrum as measured by the LANDSAT satellite, certain linear combinations of these readings, or other derived "feature" variables.

Generally, parameter estimation in mixture model applications has been accomplished by assuming that the component distributions are normal and using maximum likelihood(ML) techniques. In a recent report, Woodward, et. al.(1982) have examined the use of minimum distance(MD) estimation based on the Cramér-von Mises distance, as an alternative to maximum likelihood. Both ML and MD estimation

schemes in that paper were based upon the mixture of two univariate normal distributions whose density function is given by

$$f(x) = \frac{p}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{(1-p)}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2},$$

where all 5 parameters μ_1, σ_1, μ_2, σ_2, and p are unknown. It was also assumed that no training data are available, i.e., the only observations are from the mixture distribution. In this setting, motivated by the crop example, p is the parameter of paramount importance while location and scale of the components are nuisance parameters. Woodward, et. al. (1982) compare ML and MD estimation techniques on simulated mixtures of normal, t(4), and chi-square(9) densities with varying amounts of separation. The results indicate that the MDE is more robust than the MLE to symmetric departures from component normality, while neither technique provides satisfactory results when component distributions are skewed.

In this report, we present further simulation results comparing ML and MD estimation of the mixing proportion based on a mixture-of-normals model, when in fact the component distributions are not normal, yet represent symmetric departures from normality. Unless otherwise indicated, reference to the MDE in this report will involve the use of Cramér-von Mises distance. We also present asymptotic results which establish the strong consistency

and asymptotic normality of MD estimators of the parameters in the mixture-of-normals model, and finally provide asymptotic relative efficiencies for comparing the MLE and MDE in this setting.

2. Simulation Results

In this section we report the results of a Monte Carlo study designed to compare the ML and MD estimators based upon a mixture-of-normals when the simulated component distributions are normal and when they are non-normal. These comparisons are made under varying degrees of separation between the two component distributions. All computations were performed on the CDC 6600 at Southern Methodist University.

In these simulations, the mixing proportion, p, takes on the values .25, .50, and .75. For a given mixture, the component distributions differ from each other only in location and scale. In particular, $f_1(x)$ is taken to be the density associated with a random variable $X=aY$ while $f_2(x)$ is the density for $X=Y+b$ where $a>0$, $b>0$. Thus, a is the ratio of scale parameters for the densities f_1 and f_2, and similarly, b is the difference in location parameters. The random variable Y in our simulations is either normal, Student's t with 2 or 4 degrees of freedom, or double exponential. In our simulations we use $a=1$ and $a=\sqrt{2}$ while b

is selected to provide the desired separation between the
component distributions. The number of modes of the mixture
density depends to a large extent on this separation between
the two component distributions. Although, for sufficient
separation, the mixture model has a characteristic bimodal
shape, the density may by unimodal when there is only
moderate separation between the components, and in this
case, parameter estimation is more difficult than it is in
the bimodal cases. For purposes of quantifying this
separation between the components, a measure of "overlap"
between two distributions was defined by Woodward et.
al.(1982).

For each set of parameter configurations, 500 samples
of size n=100 were generated from the corresponding mixture
distribution. Simulations were based on the IMSL
multiplicative congruential uniform random number generator
GGUBS. Normal component observations were generated using
IMSL subroutine GGNPM which uses the polar method, while
t(n) observations were based on the ratio of independent
chi-square and normal deviates, each obtained using IMSL
routines. Double exponential components were based on ln(U)
where U is uniform(0,1), and randomly assigning either a
positive or negative sign. In all cases, observations from
the basic component distribution under investigation were
simulated and then assigned to either component 1 or
component 2 depending upon whether an independent

uniform(0,1) was less than or greater than p. The observations were then scaled and shifted (with a and b) to provide observations from the appropriate component.

For each sample simulated, both the MDE and MLE were obtained. The iterative procedures discussed by Woodward et. al. (1982) were implemented in such a way that acceptable parameter estimates are obtained for each sample. For example, if the iterative procedure fails to converge in the specified number of iterations, the last value obtained in the iteration is taken to be the estimate if this value is "reasonable" according to preset criteria. In general, if any of the following conditions existed at any step in the iteration,

$$\hat{\sigma}_1 > Y_n - Y_1 \ (= \text{sample range})$$
$$\hat{\sigma}_2 > Y_n - Y_1$$
$$\hat{\mu}_1 < Y_1 - \frac{Y_n - Y_1}{10}$$
$$\hat{\mu}_2 > Y_n + \frac{Y_n - Y_1}{10}$$

iteration is terminated and the corresponding estimate is taken to be the starting value. This did not occur in any of the 500 repititions, for most configurations, but did occur a maximum of 7 times out of 500 for MD estimates of the parameters of a mixture of t(2) components. The extreme observations which occasionally appear in samples from t(2) mixtures, also forced a modification in the first step of the MLE iteration to avoid a division by zero. Although both

estimation procedures provide estimates of all 5 of the parameters, only the results for estimation of p will be tabulated since the mixing proportion is the parameter of primary interest, as previously mentioned. In addition, when dealing with the non-normal mixtures, the remaining parameter estimates often do not have a meaningful interpretation.

In Table 1 we present summary results of the simulations comparing the performance of the MLE and MDE for mixtures of normal components while in Table 2 we display the results for the non-normal components. The results for normal and t(4) components were previously given in Woodward et. al.(1982). Estimates of the bias and MSE based upon the simulations are given by:

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

and

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

where n_s is the number of samples, and $\hat{p}_i$ denotes an estimate of p for the ith sample. It should be noted that nMSE is the quantity actually given in the tables since this facilitates comparison with asymptotic variances in Section 4. Since the MLE and MDE are both asymptotically unbiased (this will be discussed for the MDE in the next section), $n_s \text{MSE}/\sigma^2$ is approximately $\chi^2(500)$. It is easy to

Table 1 - Simulation Results for Mixtures of Normal Components

Sample size = 100

Number of Replications = 500

			Overlap = .10				Overlap = .03			
p	Ratio of Scale Factors(a)		Bias	nMSE	E	MDE Closer	Bias	nMSE	E	MDE Closer
.25	1	MDE	.125	7.80	.55	.38	.026	1.09	.49	.39
		MLE	.052	4.26			.008	.539		
		Start	.084	2.06			.048	.782		
.50	1	MDE	.010	3.86	.83	.41	.001	.420	.91	.46
		MLE	.000	3.21			.000	.382		
		Start	-.005	1.22			.001	.634		
.25	$\sqrt{2}$	MDE	.084	5.30	.42	.32	.027	.956	.51	.38
		MLE	.002	2.25			.006	.489		
		Start	-.004	.894			.014	.510		
.50	$\sqrt{2}$	MDE	.005	2.79	.86	.43	.008	.441	.94	.45
		MLE	-.009	2.41			.009	.416		
		Start	-.089	1.85			-.048	.866		
.75	$\sqrt{2}$	MDE	-.137	8.36	.58	.36	-.024	1.08	.44	.42
		MLE	-.086	4.87			-.002	.470		
		Start	-.158	3.97			-.093	1.56		

Table 2.   Simulation Results for Mixtures
of Non-normal Components

Sample size = 100

Number of replications = 500

Double Exponential Components

	Ratio of Scale		Overlap = .10				Overlap = .03			
p	Factors(a)		Bias	nMSE	E	MDE Closer	Bias	nMSE	E	MDE Closer
.25	1	MDE	.054	2.96	2.13	.66	.030	.545	1.18	.50
		MLE	.091	6.31			.026	.645		
		Start	.065	1.40			.078	1.04		
.50	1	MDE	.007	1.03	4.04	.69	-.001	.286	1.29	.54
		MLE	.007	4.16			-.001	.368		
		Start	-.004	1.17			.000	.414		
.25	$\sqrt{2}$	MDE	.102	4.42	1.40	.60	.035	.775	1.07	.48
		MLE	.034	6.17			.037	.832		
		Start	.011	.926			.050	.678		
.50	$\sqrt{2}$	MDE	.032	1.50	2.71	.68	.003	.259	1.44	.58
		MLE	.073	4.06			.009	.372		
		Start	-.088	1.86			-.035	.570		
.75	$\sqrt{2}$	MDE	-.037	2.20	2.94	.73	-.026	.344	.94	.44
		MLE	-.067	6.47			-.014	.323		
		Start	-.151	3.31			-.107	1.63		

t(4) Components

			Bias	nMSE	E	MDE Closer	Bias	nMSE	E	MDE Closer
.25	1	MDE	.104	6.18	1.19	.61	.020	.466	1.89	.49
		MLE	.096	7.35			.029	.883		
		Start	.068	1.59			.072	.998		
.50	1	MDE	.004	1.82	3.07	.69	.000	.266	1.64	.53
		MLE	.015	5.59			-.005	.436		
		Start	.006	1.21			-.001	.496		
.25	$\sqrt{2}$	MDE	.098	5.20	.89	.53	.029	.605	1.61	.49
		MLE	.061	4.63			.044	.976		
		Start	-.010	.810			.036	.654		
.50	$\sqrt{2}$	MDE	.022	1.80	2.77	.67	.001	.300	1.85	.55
		MLE	.028	4.99			.010	.554		
		Start	-.072	1.52			-.046	.778		
.75	$\sqrt{2}$	MDE	-.058	3.68	2.13	.65	-.016	.361	1.57	.50
		MLE	-.076	7.84			-.012	.567		
		Start	-.137	3.07			-.108	1.75		

Table 2 -- Continued

t(2) Components

Overlap = .10   Overlap = .03

p	Ratio of Scale Factors(a)		Bias	nMSE	E	MDE Closer	Bias	nMSE	E	MDE Closer
.25	1	MDE	.076	3.42	4.30	.80	.024	.308	10.32	.65
		MLE	.199	14.7			.083	3.18		
		Start	.067	1.85			.096	1.37		
.50	1	MDE	-.001	1.34	9.03	.92	-.005	.264	9.24	.63
		MLE	.024	12.1			-.009	2.44		
		Start	-.004	1.39			-.002	.364		
.25	$\sqrt{2}$	MDE	.118	4.92	2.26	.69	.031	.452	7.70	.69
		MLE	.169	11.1			.106	3.48		
		Start	.006	1.18			.071	.962		
.50	$\sqrt{2}$	MDE	.016	1.52	7.76	.89	-.001	.243	8.27	.68
		MLE	.028	11.8			.029	2.01		
		Start	-.078	2.08			-.032	.508		
.75	$\sqrt{2}$	MDE	-.059	2.99	5.79	.85	-.022	.300	11.40	.63
		MLE	-.186	17.3			-.045	3.42		
		Start	-.137	3.37			-.122	1.96		

show then, that the approximate standard error of a tabled $n\hat{MSE}$ is (.0632)(nMSE). In addition, we also provide the ratio

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

as an empirical relative efficiency measure.

In order to take advantage of the paired nature of our ML and MD estimates, we counted the proportion of samples for which $\hat{p}_D$ is closer to p than is $\hat{p}_L$, where $\hat{p}_D$ and $\hat{p}_L$ denote the MD and ML estimates respectively. We present this proportion in the tables under the heading "MDE Closer". This provides an estimate of $P\{|\hat{p}_D-p|<|\hat{p}_L-p|\}$. The standard error of the binomial proportions shown in the tables is no greater than $\sqrt{\frac{(.5)(.5)}{500}} = .022$.

Analyzing the results, and as can be seen by inspection, we find that the estimated Bias and MSE associated with the MLE were generally smaller than those for the MDE when the components were actually 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 mixtures of non-normal components, the relationship between MDE and MLE is reversed in that the MDE generally has the smaller estimated Bias and MSE, especially

for t(2) mixtures. The superiority of the MDE is due in part to the heavy tails in these components. 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 the variances were not constrained away from zero. However, simulation results in which equal variances were assumed (which removes the singularity) and also those that used a penalized MLE suggested by Redner(1980) were very similar to those quoted here.

A surprising result which was previously noted by Woodward et. al.(1982) is that the starting values obtained using the procedure outlined in Section 3 produced estimators that were competitive with both the MLE and MDE. For both the normal and non-normal mixtures, the MSEs associated with the starting values were generally lower than those for the MDE and MLE when overlap=.10. However, when overlap=.03, the starting value estimates were generally poorer than those for the MDE and MLE, except for the t(2) mixtures for which the MLEs were the poorest.

3. Asymptotic Distribution Theory for Minimum Cramér-von Mises Distance Estimation

Asymptotic theory for minimum Cramér-von Mises distance estimators for location parameters can be found in Parr and Schucany(1980), and for the general one parameter case in Parr and de Wet(1981). Bolthausen(1977) gives results for the mutiparameter case, but with conditions which are so strict as to rule out scale parameters for unbounded random variables (see his condition III). The purpose of the results in this section is to extend this previous work to cover multiparameter situations including, among others, the problem of normal mixtures.

Assume that at stage n we observe real-valued X_1, $X_2, \ldots, X_n$ iid from a distribution with cdf G and let G_n denote the usual empirical distribution function. Let $\mathcal{F} = \{F_\theta : \theta \varepsilon \Theta \subseteq R^k\}$, the projection model, be a family of continuous distribution functions and assume that $G \varepsilon \mathcal{F}$, i.e., $G = F_{\theta_0}$ for some $\theta_0 \varepsilon \Theta$. Further, assume that there exists an open set $A \subset \Theta$ with $\theta_0 \varepsilon A$. Also consider the following continuity(C) and differentiability(D) conditions:

(C) If $\theta_n \varepsilon \Theta$, $n = 1, 2, \ldots$, then

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} (F_{\theta_n}(x) - F_{\theta_0}(x))^2 dF_{\theta_0}(x) = 0$$

implies $\lim_{n \to \infty} \theta_n = \theta_0$.

(D) There exists a function $\eta: (0,1) \to R^k$ such that

$$\sup_{-\infty < x < \infty} |F_\theta(x) - F_{\theta_0}(x) - (\theta - \theta_0)'\eta(F_{\theta_0}(x))| = o(||\theta - \theta_0||)$$

as $||\theta - \theta_0|| \to 0$, where $||\cdot||$ is the usual Euclidean

norm on R^k, and $\int_0^1 \eta_i^2(u)du < \infty$ for $i = 1,2,\ldots,k$ where

$\eta'(u) = (\eta_1(u), \eta_2(u), \ldots, \eta_k(u))$.

Notes:

1) Condition C is satisfied if, for instance, $F_\theta(x)$ is continuous in θ at θ_0, pointwise in x (use dominated convergence). It can be interpreted as requiring that θ "continuously parametrize $\mathcal{I}$".

2) If condition C is not satisfied, then this implies $\sup_{-\infty < x < \infty} |F_\theta(x) - F_{\theta_0}(x)|$ can be arbitrarily small without having θ approach θ_0. In such a case, the search for any consistent estimator seems hopeless. In particular, in such a situation, any consistent estimating functional must be discontinuous with respect to the sup-norm, and hence highly nonrobust.

3) Condition D is weaker than (implied by) quadratic mean differentiability of $f_\theta^{1/2}$ - the canonical regularity condition for asymptotic normality of the maximum likelihood estimator (see LeCam (1970) and Pollard (1980)).

4) Usually, $\eta_i(u) = \left.\frac{\partial F_\theta(x)}{\partial \theta_i}\right|_{x=F_\theta^{-1}(u)}$ and condition D simply

states the uniform validity of the first order Taylor approximation to $F_\theta(x)$. If k=1 and θ is a location parameter, a sufficient condition to imply D) is that F_θ possess a uniformly continuous density.

Before continuing define the kxk symmetric matrices A and B by

$$A = \{a_{ij}\} \ , \quad B = \{b_{ij}\}$$

with $a_{ij} = \int\limits_0^1 \eta_i(u)\eta_j(u)\,du$

and $b_{ij} = \int\limits_0^1 \int\limits_0^1 \{\min(u,v) - uv\}\, \eta_i(u)\eta_j(v)\,du\,dv$

and assume A to be of full rank. We can now state and outline the proof of the following strong consistency and asymptotic normality results.

<u>Theorem 1</u>: Let θ_n be a minimum distance estimator of θ for all n=1, 2, ... Then, if condition C holds, $\theta_n \to \theta_0$ with probability one.

<u>Proof</u>: Clearly, $\int (G_n - F_{\theta_0})^2 dF_{\theta_0} \to 0$ with probability one, and hence also $\inf\limits_{\theta \varepsilon \Theta} \int (G_n - F_\theta)^2 dF_\theta \to 0$ with probability one. Now,

$$\sup_\theta \left| \int (G_n - F_\theta)^2 dF_\theta - \int (F - F_\theta)^2 dF_\theta \right| \leq 4 \sup_{-\infty < t < \infty} |G_n(t) - F_{\theta_0}(t)| \to 0$$

with probability one. Hence,

$$\int (F_{\theta_0} - F_{\theta_n})^2 dF_{\theta_n} = \int (F_{\theta_n} - F_{\theta_0})^2 dF_{\theta_0} \to 0$$

with probability one, and strong consistency of θ_n follows from the assumption.

__Theorem 2__: Assume conditions C and D and that A is of full rank. Then, if $f_\theta(x)$ is continuous in θ at θ_0 for every x,

$$\sqrt{n}\,(\theta_n - \theta_0) \overset{\mathcal{L}}{\longrightarrow} N(\underline{0},\ A^{-1}BA^{-1}).$$

__Proof.__ (Sketched)

Set

$$K_n(\xi) = n\int (G_n - F_{\theta_0 + \xi/\sqrt{n}})^2 dF_{\theta_0 + \xi/\sqrt{n}} \quad \text{for } \xi \in R^k.$$

Then we have

$$K_n(\xi) = n\int (G_n - F_{\theta_0} - (F_{\theta_0 + \xi/\sqrt{n}} - F_{\theta_0}))^2 dF_{\theta_0}$$

$$+ n\int (G_n - F_{\theta_0} - (F_{\theta_0 + \xi/\sqrt{n}} - F_{\theta_0}))^2 d[F_{\theta_0 + \xi/\sqrt{n}} - F_{\theta_0}]$$

$$= o_p(1) + \int_0^1 (U_n(t) - \xi'\eta(t) - R_n(t))^2 dt,$$

uniformly in ξ for $\xi'\xi \leq C$, for any $C < \infty$, where

$\sup\limits_{0 < t < 1} |R_n(t)| \to 0$ with probability one, also uniformly in ξ for $\xi'\xi \leq C$. Here, $U_n(t) = \sqrt{n}(G_n(F_{\theta_0}^{-1}(t)) - t)$, $0 < t < 1$.

By an extension of the argument of Pyke (1970, p. 29-30) to the present context, we obtain that the limiting law of the random variable minimizing $K_n(\xi)$ over ξ is also that of the value minimizing

$$\int_0^1 (B(t) - \xi'\eta(t))^2 dt,$$

where B is a Brownian bridge. The result then follows immediately.

It can be shown that the mixture of normals model satisfies the conditions of both Theorem 1 and Theorem 2.

4. Asymptotic Relative Efficiencies

Theorem 2 of the previous section indicates that for the mixture-of-normals model, we have

$$\sqrt{n}\,(\theta_n - \theta_0) \xrightarrow{\mathscr{L}} N(\underline{0}, A^{-1}BA^{-1}) \ ,$$

where $\theta_0 = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p)$ and θ_n is the vector of corresponding MD estimators using Cramér-von Mises distance. Likewise, it is well known that

$$\sqrt{n}\,(\hat{\theta}_L - \theta_0) \xrightarrow{\mathscr{L}} N(\underline{0}, I^{-1}(\theta_0)) \ ,$$

where $\hat{\theta}_L$ is the MLE of θ_0 and $I(\theta_0)$ is Fisher's information matrix. We will employ the usual terminology and refer to $A^{-1}BA^{-1}$ and $I(\theta_0)$ as asymptotic variance - covariance matrices and to their diagonal elements as asymptotic variances of the corresponding estimators. In this section we will present computed asymptotic variances for the MDE of p, which is denoted by $\hat{p}_D$, and compare these with the asymptotic variances associated with the MLE, denoted by $\hat{p}_L$.

The components of the matrix A were evaluated using the expression

$$\int_{-\infty}^{\infty} \xi_i(x)\,\xi_j(x)\,f_\theta(x)\,dx \ ,$$

where $F_\theta(x)$ and $f_\theta(x)$ denote the distribution function and density function respectively for the mixture, θ_i is the ith

component of θ, and

$$\xi_i(x) = \frac{\partial F_\theta(x)}{\partial \theta_i} \quad .$$

This integral was evaluated using IMSL subroutine DCADRE which employs Romberg extrapolation to perform numerical integration of an integral over a finite interval. In our implementation, we used DCADRE to evaluate the integral

$$\int_U^L \xi_i(x) \; \xi_j(x) \; f_\theta(x)dx,$$

where $L=\min(-10\sigma_1+\mu_1, -10\sigma_2+\mu_2)$ and $U=\max(10\sigma_1+\mu_1, 10\sigma_2+\mu_2)$ with maximum allowable absolute error specified as 1.0×10^{-15} and relative error of 1.0×10^{-12}. The double integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F_\theta(\min(x,y) - F_\theta(x)F_\theta(y)\}\xi_i(x)\xi_j(y)f_\theta(x)f_\theta(y)dxdy$$

involved in calculating the elements of the matrix B is approximated by using IMSL subroutine DBLIN to perform a Romberg integration of the integral
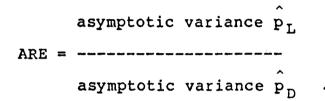
$$\int_L^U \int_L^U \{F_\theta(\min(x,y) - F_\theta(x)F_\theta(y)\}\xi_i(x)\xi_j(y)f_\theta(x)f_\theta(y)dxdy$$

with maximum allowable absolute errror specified as 1.0×10^{-9} .

The calculation of the information matrix for the

mixture-of-normals model is discussed by Behboodian(1972).
We have followed Behboodian's procedure and used
Gauss-Hermite quadrature to approximate the integrals
involved. Using 48-point quadrature we obtain good agreement
with Behboodian's tabled results.

In Table 3 we display the asymptotic variances for $\hat{p}_D$
and $\hat{p}_L$ along with asymptotic relative efficiency (ARE)
calculated as

$$ARE = \frac{\text{asymptotic variance } \hat{p}_L}{\text{asymptotic variance } \hat{p}_D} \ .$$

These values are calculated for each of the parameter
configurations employed in Table 1 for the normal mixtures.
As in Table 1, the asymptotic results indicate that the MDE
compares more favorably with the MLE when p=.5 while its
relative performance is not as good for p=.25 or p=.75.

Table 3 - Asymptotic Relative Efficiencies

p	Ratio of Scale Factors(a)		Overlap = .10		Overlap = .03	
			Asymptotic Variance	ARE	Asymptotic Variance	ARE
		MDE	13.60 (7.80)*	.42 (.55)	.471 (1.09)	.69 (.49)
.25	1	MLE	5.67 (4.26)		.323 (.539)	
	•	MDE	4.54 (3.86)	.65 (.83)	.398 (.420)	.89 (.91)
.50	1	MLE	2.95 (3.21)		.355 (.382)	
		MDE	18.77 (5.30)	.32 (.42)	.511 (.956)	.65 (.51)
.25	$\sqrt{2}$	MLE	5.96 (2.25)		.330 (.489)	
		MDE	3.49 (2.79)	.68 (.86)	.395 (.441)	.89 (.94)
.50	$\sqrt{2}$	MLE	2.39 (2.41)		.353 (.416)	
		MDE	5.51 (8.36)	.58 (.58)	.420 (1.08)	.73 (.44)
.75	$\sqrt{2}$	MLE	3.18 (4.87)		.305 (.470)	

*Associated Monte Carlo results from Table 1 are given in parentheses.

5. Concluding Remarks

We believe that the results of this paper provide further evidence that the use of the MDE should be considered in crop proportion estimation procedures developed by NASA. Our results, again, and more conclusively than before, 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.

Woodward et. al. (1983) have investigated basing the MD estimation procedure on a mixture of Weibull components in order to allow for possible asymmetry in the component distributions. Their results indicate that this approach provides a viable alternative to the normal-based procedures discussed here. Research is also proceeding on the case of multiple (>2) components in the mixture.

The results of Section 4 indicate that the MDE does not perform as well as would be hoped when the data actually do arise from a mixture-of-normals model. We are currently examining the use of the Hellinger metric in this regard due the results of Beran(1977) concerning the full asymptotic relative efficiency of minimum Hellinger distance estimators.

REFERENCES

1. Behboodian, J.(1972). "Information Matrix for a Mixture of Two Normal Distributions," J. Statist. Comput. Simul. 1, 295-314.

2. Beran, R.(1977). "Minimum Hellinger Distance Estimates for Parametric Models," Annals of Statistics 5, 445-463.

3. Bolthausen, E.(1977). "Convergence in Distribution of Minimum-Distance Estimators," Metrika, 24, 215-228.

4. Day, N. E.(1969). "Estimating the Components of a Mixture of Normal Distributions," Biometrika, 56, 463-474.

5. LeCam, L.(1970)."On the Assumptions Used to Prove Asymptotic Normality of Maximum Likelihood Estimates," Annals Math. Statist., 41, 802-828.

6. Parr, W.C. and De Wet,T.(1981). "On Minimum Cramer-von Mises-norm Parameter Estimation," Communications in Statistics, A10, 1149-1166.

7. Parr, W. C. and Schucany W. R.(1980)."Minimum Distance and Robust Estimation, "J. Amer. Statist. Assoc., 75, 616-624.

8. Pollard, D.(1980)."The Minimum Distance Method of Testing," Metrika, 27, 43-70.

9. Pyke, R.(1970)."Asymptotic Results on Rank Statistics,"
   Nonparametric Techniques in Statistical
   Inference, ed. by M.L.Puri, Cambridge University
   Press, London.

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

11. Woodward, W.A., Gunst, R.F., Lindsey, H., and Gray,
    H.L.(1983). "Proportion Estimation in Mixtures of
    Asymmetric Distributions," NASA Technical Report
    SR-63-04409, May 1983.

12. Woodward, W.A., Schucany, W.R., Lindsey, H. and Gray,
    H.L.(1982). "A Comparison of Minimum Distance and
    Maximum likelihood Techniques for Proportion
    Estimation," NASA technical report SR-62-04376, November
    1982.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SMU/DS/TR/178	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Minimum Distance Estimation of Mixture Model Parameters - Asymptotic Results and Simulation Comparisons with Maximum Likelihood		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER SMU/DS/TR-178
7. AUTHOR(s) Wayne A. Woodward, William C. Parr, William R. Schucany, Henry L. Gray		8. CONTRACT OR GRANT NUMBER(s) N00014-82-k-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Southern Methodist University Dallas, TX 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-479
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE June 1983
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS*(II different from Controlling Office)*		15. SECURITY CLASS. *(of this report)*
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

   This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, II different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side II necessary and identify by block number)*

   Mixture model; robustness; maximum likelihood; minimum distance; asymptotic relative efficiency; simulation; iterative routines, EM algorithm

20. ABSTRACT *(Continue on reverse side II necessary and identify by block number)*

   The estimation of mixing proportions in the mixture model is discussed with emphasis on the mixture of two normal components with all five parameters unknown. Simulations are presented which compare minimum distance (MD) and maximum likelihood (ML) estimation of the parameters of this mixture-of-normals model. Some practical issues of implementation of these results are also discussed. Simulation results indicate that ML techniques are superior to MD when component distributions actually are normal, while MD techniques provide

DD FORM
1 JAN 73 1473   EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

better estimates that ML under symmetric departures from component normality. Results are presented which establish strong consistency and asymptotic normality of the MD estimator under conditions which include the mixture-of-normals model. Asymptotic variances and relative efficiencies are obtained for further comparison of the MDE and MLE.