TESTING FOR THE MAXIMUM MEAN IN A MIXTURE OF NORMALS

H.L. Gray, Southern Methodist University
Wayne A. Woodward, Southern Methodist University
Gary McCartor, Mission Research Corporation

SMU/DS/TR-228

June 1989

.

Testing for the Maximum Mean in a Mixture of Normals[*]

by

H.L. Gray, Southern Methodist University

Wayne A. Woodward, Southern Methodist University

Gary McCartor, Mission Research Corporation

ABSTRACT

We consider the test of the null hypothesis that the largest mean in a mixture of an unknown number of normal components is less than or equal to some threshold. This test is motivated by the problem of assessing whether or not the Soviet Union has been operating in compliance with the Nuclear Test Ban Treaty. In our analysis, the number of normal components is assessed using AIC while the hypothesis test itself is based on asymptotic results given by Beehbodian for a mixture of two normal components. A bootstrap approach is also considered for estimating the standard error of the largest estimated mean. The performance of the tests are examined through the use of simulation.

Key Words: Mixture of Normals, AIC, Bootstrap, Simulation, Nuclear Test
Ban Treaty Verification

1. INTRODUCTION

The mixture of normals model is one that has been studied extensively dating back to Pearson (1894), Charlier (1906), and others. More currently there are major monographs which focus on the subject by Everett and Hand (1981), Titterington, Smith and Makow (1985) and Mc Laughlan and Basford (1988). In addition the contemporary literature continues to explore the theory and application of the mixture of normals. In spite of all of the attention given the mixture of normals, the problem to which we apply the normal mixture here has apparently not been previously considered. In particular, in this paper we develop a test of hypothesis that the maximum mean from a mixture of an unknown number of normal distributions is less than some threshold value. This test is then applied to a problem of national importance, which was in fact the stimulus for the development of the test. More specifically we consider the problem of testing the hypothesis that the Soviet Union has remained in compliance with the Nuclear Test Ban Treaty since 1974, and we show how the test studied in this paper can be used to help answer that question.

2. BACKGROUND

A random variable, X is said to be distributed as a mixture of normals if its probability density function f is given by

$$f(x; \underline{p}, \underline{\mu}, \underline{\sigma}) = \sum_{k=1}^{\ell} \frac{p_k}{\sqrt{2\pi}\,\sigma_k} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2\right] , \tag{1}$$

where

$$\sum_{k=1}^{\ell} p_k = 1$$

$$\underline{p} = (p_1, p_2, \cdots p_\ell)$$

$$\underline{\mu} = (\mu_1, \mu_2, \cdots \mu_\ell)$$

$$\underline{\sigma} = (\sigma_1, \sigma_2, \cdots \sigma_\ell) .$$

The p_k are usually referred to as the mixing proportions, while $\underline{\mu}$ and $\underline{\sigma}$ are the mean and standard deviation vectors. We will sometimes use the notation $\underline{\theta} = (\underline{p}, \underline{\mu}, \underline{\sigma})$ to denote the entire parameter set. When ℓ is given, the maximum likelihood estimates $\hat{\underline{p}}$, $\hat{\underline{\mu}}$ and $\hat{\underline{\sigma}}$ from a sample of size n are the solutions to the equations

$$\frac{\partial L}{\partial p_k} = \sum_{i=1}^{n} \frac{1}{f(x_i; \hat{\underline{p}}, \hat{\underline{\mu}}, \hat{\underline{\sigma}})} \left[f_k(x_i) - f_\ell(x_i)\right] = 0, \quad k=1, 2, \ldots \ell-1,$$

$$\frac{\partial L}{\partial \mu_k} = \sum_{i=1}^{n} \frac{\hat{p}_k f_k(x_i)}{f(x_i, \hat{\underline{p}}, \hat{\underline{\mu}}, \hat{\underline{\sigma}})} (x_i - \hat{\mu}_k) = 0, \quad k=1, 2, \ldots \ell, \tag{2}$$

$$\frac{\partial L}{\partial \sigma_k} = \sum_{i=1}^{n} \frac{\hat{p}_k f_k(x_i)}{f(x_i; \hat{\underline{p}}, \hat{\underline{\mu}}, \hat{\underline{\sigma}})} [\hat{\sigma}_k^2 - (x_i - \hat{\mu}_k)^2] = 0, \quad k=1, 2, \ldots \ell.$$

where L is the log-likelihood function and

$$f_k(x) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_k} \exp\left[-\frac{1}{2}\left(\frac{x-\hat{\mu}_k}{\hat{\sigma}_k}\right)^2\right] . \tag{3}$$

Although in the general problem, the likelihood function is unbounded, this is not the case when $\sigma_i \equiv \sigma$, $i = 1, \ldots, \ell$. Fortunately in the problem we wish to consider, the assumption that $\sigma_i \equiv \sigma$, with σ known, is a reasonable one and one which we will make henceforth. In this event

$\hat{\underline{p}}$ and $\hat{\underline{\mu}}$, are consistent and asymptotically normal. See Redner and

Walker (1984) for a discussion of properties in general, i.e. whether or

not the σ_i are known. The solution of the system (2) is nontrival.

However it can be obtained via the EM (Expectation Maximization) algorithm,

a result discussed by Redner and Walker (1984) as well as others.

The resulting iterative solution is given by the following equations

$$\hat{p}_k^{(m)} = \frac{\hat{p}_k^{(m-1)}}{n} \sum_{i=1}^{n} \frac{f_k^{(m-1)}(x_i)}{f^{(m-1)}(x_i)}$$

$$\hat{\mu}_k^{(m)} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \frac{f_k^{(m-1)}(x_i)}{f^{(m-1)}(x_i)}}{\sum_{i=1}^{n} \frac{f_k^{(m-1)}(x_i)}{f^{(m-1)}(x_i)}} \qquad (4)$$

where m denotes the m-th iterate, k=1, 2, ... ℓ while $f^{(m)}$ and $f_k^{(m)}$

represent the m-th iterate of the mixture density given in (1) and the kth

component density in (3) respectively.

One feature of the equations in (4) is that they require starting

values. These can be obtained in a number of ways. In this paper the

starting values were obtained using a simple clustering algorithm. The

algorithm begins by considering all combinations of $p_1^{(0)}$, $p_2^{(0)}$,..., $p_\ell^{(0)}$

such that $p_1^{(0)} + p_2^{(0)}+ ...$, $p_\ell^{(0)} = 1$ and each $p_i^{(0)}$ is a multiple

of $\frac{1}{10}$, with $\ell \leq 10$. Each combination $p_1^{(0)}$, $p_2^{(0)}$,...,$p_\ell^{(0)}$ partitions the

sample. Essentially, the smallest $np_1^{(0)}$ data values fall into cluster 1,

the next $np_2^{(0)}$ observation are placed into cluster 2, etc. For each

combination of $p_1^{(0)}$,$p_2^{(0)}$,...,$p_\ell^{(0)}$, the sum of within-cluster sample variances

is obtained as a measure of within-cluster variability. The starting

values $\hat{p}_1(0)$,..., $\hat{p}_\ell(0)$ are taken to be the combination of $p_i^{(0)}$'s resulting

3

in the minimum sum of within-cluster variability. The starting values for $\mu_1, \ldots, \mu_\ell$ are then the sample means of clusters associated with $\hat{p}_1^{(0)}, \ldots, \hat{p}_\ell^{(0)}$.

3. ESTIMATING ℓ

The estimation of ℓ is nontrival and in general cannot be effectively accomplished by simple inspection of the histogram even when the sample size is large. This is due to the fact that the μ_i have to be reasonably well separated for the true mixture density to exhibit multiple modes. For example a necessary condition for the mixture of two normals to be bimodal is that $|\mu_1 - \mu_2| > 2 \min(\sigma_1, \sigma_2)$. Figure 3.1, shows the density functions for mixtures of two normals for various parameter configurations. Figure 3.2 shows the shape for a particular mixture of 3 normals. These figures make the difficulty of estimating ℓ clear and show the need for a quantitative method for estimating ℓ.

At first glance one might be tempted to use the likelihood principle to estimate ℓ. However since the dimension of the parameter space changes with ℓ, the maximum likelihood method does not apply. Fortunately we can use Akaike's Information Criteria (AIC), which is a generalization of the likelihood method that employs the likelihood function accompanied by a penalty function (see Akaike, 1974). For the mixture of normals density function given by (1), the AIC for each value of ℓ can be calculated as follows (see Redner, Kitagawa and Coberly, 1981):

$$
AIC(\ell) = -\frac{2}{n} \sum_{i=1}^{n} \ell n \left\{ \sum_{k=1}^{\ell} \frac{\hat{p}_k}{\sqrt{2\pi}\,\sigma} \exp \left[-\frac{1}{2} \left(\frac{x_i - \hat{\mu}_k}{\sigma} \right)^2 \right] \right\} + \frac{2\ell - 1}{n}, \tag{5}
$$

4

where $\hat{p}_k$ and $\hat{\mu}_k$ are determined by (4). The AIC estimate for ℓ is then given by

$$\hat{\ell} = \{\ell; \text{ AIC} = \text{minimum}, \ell = 1, 2, \ldots, M\} , \tag{6}$$

where M is a preassigned positive integer.

4. TESTING THE HYPOTHESIS THAT MAXIMUM $\mu_k \leq$ T.

Let μ_{max} = maximum $\{\mu_i, i=1, 2, \ldots, \ell\}$, where ℓ is unknown. For a given threshold, T, we wish to test the hypothesis H_0: $\mu_{max} \leq$ T against H_A: $\mu_{max} >$ T. If $\hat{\underline{\theta}}$ is the MLE obtained from (4), then $\hat{\underline{\theta}}$ is asymptotically a $2\ell-1$ variate normal with covariance matrix

$$V = n^{-1}R^{-1},$$

where

$$R = (r_{ij}) \tag{7}$$

$$r_{ij} = E\left[\frac{\partial}{\partial\theta_i} \ell nf(x) \cdot \frac{\partial}{\partial\theta_j} \ell nf(x) \right] .$$

(e.g. see Everitt and Hand, 1981).

In order to obtain a test of hypotheses we estimate R, which is a nontrivial task in general. It has however been accomplished by Behboodian (1972) for the mixture of two normals. Even though the method of Behboodian could be extended to $\ell > 2$, it would undoubtably take a very large sample to obtain good estimates of the r_{ij}. Suppose, however, the μ_i are sufficiently separated that it is reasonable to assume that $E[\max \hat{\mu}_i] = \mu_{max}$ and $E[\text{2nd largest } \hat{\mu}_i] = \mu_{m2}$, where μ_{m2} is the 2nd largest mean. In this setting, we denote $\max \hat{\mu}_i = \hat{\mu}_{max}$ and second largest $\hat{\mu}_i = \hat{\mu}_{m2}$.

5

We will employ a test statistic of the form

$$Z = \frac{\hat{\mu}_{max} - T}{\widehat{SE}(\hat{\mu}_{max})} \tag{8}$$

where $\widehat{SE}(\hat{\mu}_{max})$ is the estimated standard error of the largest estimated mean. Thus, the problem is to estimate the variability of $\hat{\mu}_{max}$. We assume that the variance of $\hat{\mu}_{max}$ is affected only slightly by data values associated with components whose means are less than μ_{m2}. Therefore, since our only interest is in determining whether or not $\hat{\mu}_{max}$ is less than a given threshold, it seems intuitively reasonable to assess the variability of $\hat{\mu}_{max}$ by applying Behboodian's (1972) results to the two component mixture involving the components with estimated means $\hat{\mu}_{m2}$ and $\hat{\mu}_{max}$. In Section 6 we will argue that in the test ban compliance application considered here, these are reasonable assumptions.

The procedure utilized is as follows:

(a) Select ℓ with AIC and obtain MLE estimates of $(\underline{p}, \underline{\mu})$.

(b) Consider the two component mixture consisting of the largest two estimated means obtained in (a). Estimate the 3 parameters μ_{m2}, μ_{max} and p^* (mixing proportion) by $\hat{\mu}_{m2}$, $\hat{\mu}_{max}$, and $\hat{p}^* = (\hat{p}_{m2}/(\hat{p}_{m2} + \hat{p}_{max})$ respectively where components 1 and 2 correspond to μ_{m2} and μ_{max} respectively and $\hat{p}_{m2}$ and $\hat{p}_{max}$ are ML estimates of the proportions in the components associated with μ_{m2} and μ_{max} respectively. Note that we also approximate the sample size n_2 from this two component mixture to be $n_2 = n(\hat{p}_{m2} + \hat{p}_{max})$.

(c) For this two component mixture, the information matrix, R, given in (7) is estimated by $\hat{R}$ obtained by applying Behboodian's (1972) results to the two component mixture with $\theta_1 = \hat{\mu}_{m2}$, $\theta_2 = \hat{\mu}_{max}$ and

6

$\theta_3 = \hat{p}^*$. In particular, the elements are found using Gaussian 48 point quadrature to approximate integrals of the form

$$M_{mk}(i,j) = \int_{-\infty}^{\infty} (\frac{x-\mu_i}{\sigma})^m \frac{(x-\mu_j)^k}{\sigma} \frac{f_i(x)f_j(x)}{f(x)} \, dx \; .$$

The components $r(i,j)$ in (7) are found from $M_{mk}(i,j)$ as given by Behboodian with obvious simplification for the fact that in our case σ is known. Thus, we obtain

$$\hat{SE}(\hat{\mu}_{max}) = \sqrt{\hat{v}(2,2)/n_2}$$

where $\hat{v}(2,2)$ is the element in the second row and second column of $\hat{V} = \hat{R}^{-1}$. The hypothesis is then tested by treating Z in (8) as a standard normal random variable. We will refer to the test procedure described in this section as the Upper Two-Component Test (UTCT).

5. A BOOTSTRAP APPROACH

The methodology in the previous section involved several approximations. Among other things, the covariance estimate may be poor unless the sample size is quite large and the actual critical region may not be robust to the dispersion of the μ_2. An alternative method for obtaining an estimate of $SE[\hat{\mu}_{max}]$ is to use bootstrap techniques which are based on the use of resampling (see Efron and Gong, 1983 and Efron and Tibshirani, 1986). Although it is computationally intensive, the bootstrap eliminates the need to estimate the covariance matrix. Moreover, there need be no assumptions about the dispersion of the μ_i except that μ_{max} and μ_{m2} be sufficiently separated so that $E[\hat{\mu}_{max}] = \mu_{max}$.

7

Two different techniques for resampling are used in practice. The nonparametric bootstrap is based on repeatedly resampling from the actual sample. On the other hand, the parametric bootstrap is based on taking the observed sample, estimating the parameters of the assumed model, and then repeatedly generating samples from this estimated parametric model. In either case, variability of a particular estimator, such as $\hat{\mu}_{max}$, can be estimated by the sample variance of the estimator under consideration across the bootstrap samples. In our setting, we implemented the parametric bootstrap in the following manner. As mentioned in Section 2, we assume that our observed sample is from a mixture of normal components with known variance but with unknown number of components. For the observed sample, we then estimate the model parameters including the number of components. Then, 99 samples of the same size as the observed sample are generated from the estimated model. For each of these generated samples, we estimate the model parameters assuming that the number of components is known and is equal to the number of components estimated in the original observed sample. The sample variance of the largest sample means across the 99 samples is then obtained, and we denote this estimate as $\hat{SE}_B(\hat{\mu}_{max})$. The test statistic, corresponding to (8), is

$$
Z = \frac{\hat{\mu}_{max} - T}{\hat{SE}_B(\hat{\mu}_{max})} \quad , \tag{10}
$$

where we employ the same rejection rule as before and $\hat{\mu}_{max}$ is as in (8), i.e. it is the estimate of the largest mean based on the original observed sample.

6. APPLICATIONS AND SIMULATIONS

In 1974 the Soviet Union and the United States negotiated a treaty referred to as the Nuclear Test Ban Treaty (U.S. Congress, Office of Technology Assessment, 1988). This treaty restricted all nuclear testing to underground explosions whose yields do not exceed 150 kilotons. Although this treaty was never ratified by the United States Senate, it is in force under international law by executive agreement. Currently it appears that most, if not all, countries are abiding by such a limit. Since 1974 the Soviets have carried out numerous nuclear tests. Such tests are of course monitored seismically by the United States and estimated yields are obtained. For each blast the logarithm of the estimated yield is obtained. This log-yield data can be reasonably modeled as coming from a mixture of normal components. The assumption that the data follows a mixture model seems reasonable when one considers the fact that nuclear tests are made for purposes of weapons development. Thus, it is not unreasonable to expect that more than one explosion would be made at roughly each of several theoretical yield levels associated with the weapons being developed. Also, since the levels of testing associated with different weapons are likely to differ significantly, one may expect the components to be reasonably well separated. Unfortunately, the actual yield estimates are in fact classified and cannot be given here. Figure 6.1, however, shows histograms for three different sets of 80 simulated log yields which are for the most part representative of the nature of the actual data. The first two of these sets, (a) and (b), are obtained from underlying models representing compliance, i.e. $\mu_{max} \leq 2.176$ where in this case our variable of interest is log yield and $2.176 = \log 150$. Specifically, these two sets were each simulated from a mixture of three

components with means 1.699, 2.0, and 2.176 respectively. In the simulations, 16 samples were generated from the first component, 22 from the second component and 42 from the third component. The third set, (c), is for a hypothetical case which is not in compliance with the treaty. This sample came from an underlying mixture of 4 components whose means were 1.699, 2.0, 2.176 and 2.3. The number of observations from the four components were 16, 35, 20 and 9 respectively. Visual examination of these histograms does not provide a clear answer concerning the compliance or noncompliance. However, applying the Z-test given in (8) with $T = 2.176$ results in a rejection of the null hypothesis of compliance for the noncompliance set (c) and failure to reject the compliance null hypothesis for sets (a) and (b) at the $\alpha = .05$ level of significance. Specifically, the results for the test outlined in Section 4 for these three data sets are summarized in Table 6.1.

The results in Table 6.1 are rather anecdotal in nature and do not provide conclusive evidence that the test procedure described in Section 4 performs acceptably. In order to further examine the test, we performed an extensive simulation study using the IBM 3081D computer at Southern Methodist University. Since the testing procedure involves approximations, we first examined the extent to which the observed significance levels agreed with the nominal levels. We simulated samples from mixtures of normals whose component variances were equal and assumed to be known and for which the mixing proportions were approximately equal. The component means took on the values $2.176-(k-1)d\sigma$, $k=1,2,\ldots,\ell$ where ℓ is the number of components, σ is the common component variance and d is a multiplier specifying the separation among the components. Note that $\mu_{max} = 2.176$ in all cases considered here so that these are situations in which the null

TABLE 6.1

Compliance Hypothesis Test Results for the
Samples in Figure 6.1 using the Upper Two-Component Test

	(a)	(b)	(c)
Estimated # of Components	3	2	3
$\hat{\mu}_{max}$	2.257	2.143	2.263
$\hat{\mu}_{m2}$	2.044	1.764	2.009
$\hat{p}_{max}$	.328	.732	.349
$\hat{p}_{m2}$	.410	.268	.437
$\hat{SE}(\hat{\mu}_{max})$	.058	.020	.046
Z (Reject H_0 if Z>1.645)	1.408	−1.648	1.901

hypothesis of compliance is true. We considered the cases in which the number of components, ℓ, was 2, 3 and 4 and in which the multiplier d took on the values 1.5, 2, 2.5 and 5. For each of the 12 resulting combinations we generated 2000 samples of length n = 80. Each sample was analyzed using the UTCT procedure described in Section 4 where the maximum number of components considered by AIC is M=5. In Table 6.2 we show the actual proportion of the 2000 samples for which the null hypothesis of compliance, H_0: $\mu_{max} \leq 2.176$, is rejected when using a nominal α = .05 test. There it can be seen that there is reasonably close agreement with a tendency for the observed significance level to be slightly higher than nominal levels.

Obviously, one would expect that AIC performance would improve as the separation among component means increases. It should be noted that Behboodian (1970) has shown in the case of a two component mixture that the mixture density function is bimodal if and only if $|\mu_1 - \mu_2| > 2\sigma$. In Table 6.3 we examine the ability of AIC to select the correct number of components for the cases in which d = 2 and d = 2.5. There it can be seen that when d = 2, AIC had a definite tendency to underestimate the number of components in the model, as would be expected. Although performance is better when d = 2.5, the table shows that AIC's estimate of the number of components is often incorrect. However, as we see in the results shown here, the performance of the test does not appear to be overly sensitive to the estimation of the number of components.

The use of the bootstrap has been examined for a subset of the configurations considered in Table 6.2 Because of computer time considerations, we restricted our simulations to 200 samples of size n = 80 for the cases in which ℓ = 2,3 and 4 with d = 2 and 2.5. For each

11

TABLE 6.2

Proportion of Times the Hypothesis
H_0: $\mu_{max} \leq 2.176$ is Rejected
using the Upper Two Component Test

# of repetitions/cell = 2000
# of data points = 80

Number of Components

	2	3	4
1.5σ	.052	.044	.050
2σ	.051	.073	.060
2.5σ	.061	.066	.071
5σ	.061	.064	.069

Separation Among Means

TABLE 6.3

AIC Performance

True Number of Components

		2		3		4	
		Separation		Separation		Separation	
		2σ	2.5σ	2σ	2.5σ	2σ	2.5σ
Estimated	1	7	0	0	0	0	0
Number of	2	1954	1928	1091	140	24	1
Components	3	38	72	895	1840	1757	845
	4	1	0	13	19	214	1141
	5	0	0	1	1	5	14

configuration in Table 6.4(a) we show the average of $\hat{SE}(\hat{\mu}_{max})$ and

$\hat{SE}_B(\hat{\mu}_{max})$ across the 200 samples. Additionally, in Table 6.4(b) we show

the observed significance levels for the tests given in Sections 4 and 5.

While there is reasonable agreement between the results for the two

procedures, it appears from the tables that in general $\hat{SE}(\hat{\mu}_{max}) \leq$

$\hat{SE}_B(\hat{\mu}_{max})$ which has the effect that the bootstrap-based test produces

fewer rejections of the null hypothesis, lower than the nominal level for

all 6 configurations examined.

The power of the UTCT has also been examined via computer simulation.

We simulated samples from 4-component models for which $\mu_{max} = 2.301$, 2.398,

2.477, i.e. models for which the null hypothesis H_0: $\mu_{max} \leq 2.176$ is not

true. For each of the models considered, the lower three means are 1.301,

1.903 and 2.176. We considered samples sizes of 22, 44 and 88, and in each

case there were a small number of observations associated with the

component with mean μ_{max} while the remaining observations were

approximately equally divided among the three other components. A total of

1600 samples for each configuration were simulated, and in Table 6.5 we

show the estimated power, i.e. the proportion of samples for which the null

hypothesis was rejected. There it can be seen that this test has

substantial power against several of the alternatives considered. The test

has also shown to be more powerful than other tests under consideration for

compliance testing. Additionally, it performs well when the model

assumptions are not met. For a discussion of these results see Gray and

McCartor (1986).

TABLE 6.4

Comparison of Hypothesis Tests Based on Upper Two-Component Test
(UTCT) and the Test Based on the Bootstrap

(a) Average $\widehat{SE}$ ($\hat{\mu}_{max}$) Across Samples

Number of Components

		2		3		4	
		UTCT	Bootstrap	UTCT	Bootstrap	UTCT	Bootstrap
Separation Among Means	2σ	.0381	.0402	.0380	.0427	.0403	.0462
	2.5σ	.0316	.0326	.0396	.0469	.0402	.0462

(b) Proportion of Times the Hypothesis H_0: $\mu_{max} \leqq 2.176$
is Rejected (Nominal = .05)

Number of Components

		2		3		4	
		UTCT	Bootstrap	UTCT	Bootstrap	UTCT	Bootstrap
Separation Among Means	2σ	.055	.045	.055	.040	.050	.025
	2.5σ	.050	.045	.060	.030	.060	.045

TABLE 6.5

Proportion of Samples for which the Null Hypothesis
H_0: $\mu_{max} \leq 2.176$ is Rejected using the UTCT

Repetitions/cell = 1600

μ_{max}	% of observations with mean μ_{max}	Sample Size		
		22	44	88
2.301	4.5%	.11	.15	.21
2.301	13.6%	.28	.40	.58
2.398	4.5%	.22	.37	.52
2.398	13.6%	.66	.87	.98
2.477	4.5%	.40	.61	.82

7. CONCLUDING REMARKS

In this paper we have developed a test for the hypothesis that the largest mean of a mixture of an unknown number of normals is less than or equal to some threshold value, and we have examined its properties through computer simulations. We discussed the use of Akaike's (1974) AIC criteria for determining the number of components in the mixture. Other techniques are available for this determination including BIC suggested by Akaike (1977). Hannan (1980) showed that BIC provides a consistent alternative to AIC in the case of ARMA model identification. BIC has a more severe penalty for adding additional parameters and thus often selects a smaller number of components than does AIC in our case, causing $\hat{\mu}_{max}$ to tend to be smaller than that obtained using AIC. The impact of this on the UTCT is that it causes the test to be more conservative in the sense that H_0: $\mu_{max} \leq T$ is not rejected as often. We observed this effect in simulations related to those shown in Table 6.2. In particular, the proportion of rejections using BIC was much lower than the nominal level when separation was very slight, while it was closer to the nominal levels than the AIC-based results shown in Table 6.2 for the larger separations. McLachlan and Basford (1988) also mention other techniques for identifying the number of components. It is possible that the UTCT and bootstrap procedures suggested here could be improved with better estimation of the number of components.
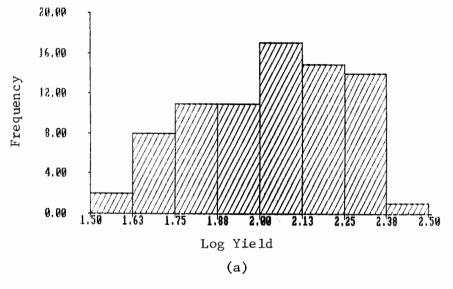
Preliminary investigation into the problem of testing the hypothesis $H_0 : \mu_{max} \leq T$ when the component variances are not assumed to be known and equal have shown that the test performs poorly in this case. This seems to primarily be due to the fact that the identification of the number of components is difficult in this setting. For example, we considered the
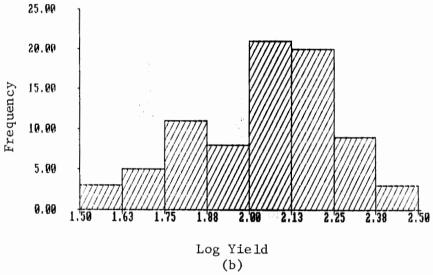
13

case in which the component variances were unknown but assumed to be equal, and our simulations typically produced several samples for which AIC picked too many components. If one of these "extra" components is associated with the largest data values, this may cause the estimate of μ_{max} to be excessively large. Also, if more components are selected than are actually in the data, this will have a tendency to cause the component variance estimate to be small. These two effects caused the percentage of rejections of the null hypothesis we obtained to be considerably larger than nominal levels in the cases examined in Table 6.2. By placing constraints on the unknown variance, these results improved. More work is definitely required for the case in which the component variance is unknown.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEEE Transaction on Automatic Control 19, 716-723.

Akaike, H. (1977), "On Entropy Maximization Principle", Applications of Statistics, Ed. P. R. Krisnaiah, Amsterdam: North-Holland, 27-41.

Behboodian, J. (1970), "On the Modes of a Mixture of Two Normal Distributions," Technometrics 12, 131-139.

Behoodian, J. (1972), "Information Matrix for a Mixture of Two Normal Distributions," Journal of Statistical Computation and Simulation, 1, 295-314.

Charlier, C.V.L. (1906), "Researches into the theory of Probability," Lunds Universitets Arskrift, Ny folyd, Afd, 2.1, No. 5.

Efron, B. and Gong, G. (1983), "A Leisurely Look at the Bootstrap, the Jackknife and Cross-Validation," American Statistician, 37, 36-48.

Efron, B. and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy," Statistical Science 1, 54-77.

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

Gray, H. L. and McCartor, G. (1986), MRC Final Draft Report for ACDA/DARPA, MRC-R-967, Mission Research Corporation.

Hannan, E. J. (1980), "The Estimation of the Order of an ARMA Process," Annals of Statistics 8, 1071-1081.

McLachlan, G. J. and Basford, K. E. (1988), Mixture Models, New York: Marcel Dekker, Inc.

Pearson, K. (1894), "Contributions to the Mathematical Theory of Evolution," Philosophical Transactions of the Royal Society of London, Ser A, 71-110.

Redner, R. A. and Walker, H. F. (1984), "Mixture Densities, Maximum Likelihood and the EM Algorithm," SIAM Review 26, 195-239.

Redner, R. A., Kitagawa, G. and Coberly, W. A. (1981), "The Akaike Information Criterion and its Application to Mixture Proportion Estimation", NASA Technical Report No. SR-TI-04207.

Titterington, D. M., Smith, A.F.M and Makov, U.E. (1985), Statistical Analysis of Finite Mixture Distributions, New York: John Wiley and Sons.

U.S. Congress, Office of Technology Assessment, (1988) Seismic Verification of Nuclear Testing Treaties, OTA-ISC-361. Washington, DC: U.S. Government Printing Office.
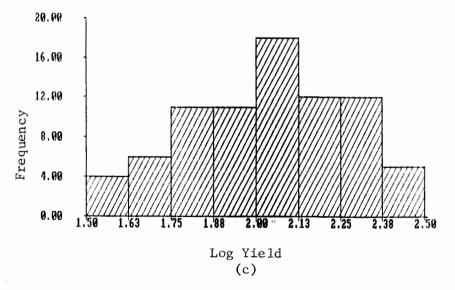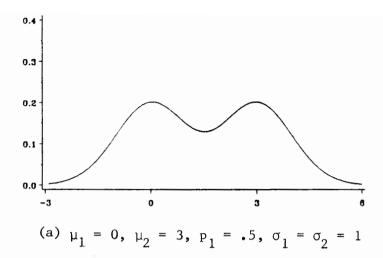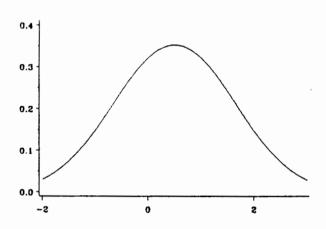
15

Figure 6.1 Histograms for Three Simulated Data Sets

(a) $\mu_1 = 0$, $\mu_2 = 3$, $p_1 = .5$, $\sigma_1 = \sigma_2 = 1$

(b) $\mu_1 = 0$, $\mu_2 = 1$, $p_1 = .5$, $\sigma_1 = \sigma_2 = 1$

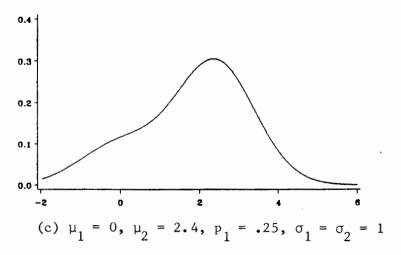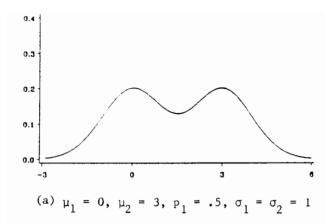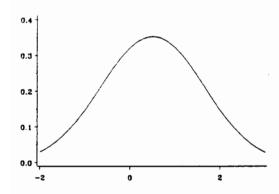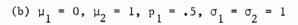(c) $\mu_1 = 0$, $\mu_2 = 2.4$, $p_1 = .25$, $\sigma_1 = \sigma_2 = 1$

Figure 3.1   Density Functions for Mixtures of
Two Normal Components

(a) $\mu_1 = 0$, $\mu_2 = 3$, $p_1 = .5$, $\sigma_1 = \sigma_2 = 1$

(b) $\mu_1 = 0$, $\mu_2 = 1$, $p_1 = .5$, $\sigma_1 = \sigma_2 = 1$

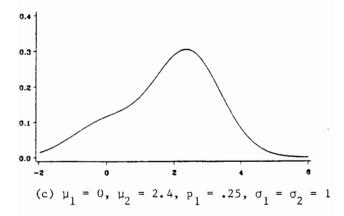(c) $\mu_1 = 0$, $\mu_2 = 2.4$, $p_1 = .25$, $\sigma_1 = \sigma_2 = 1$

Figure 3.1   Density Functions for Mixtures of
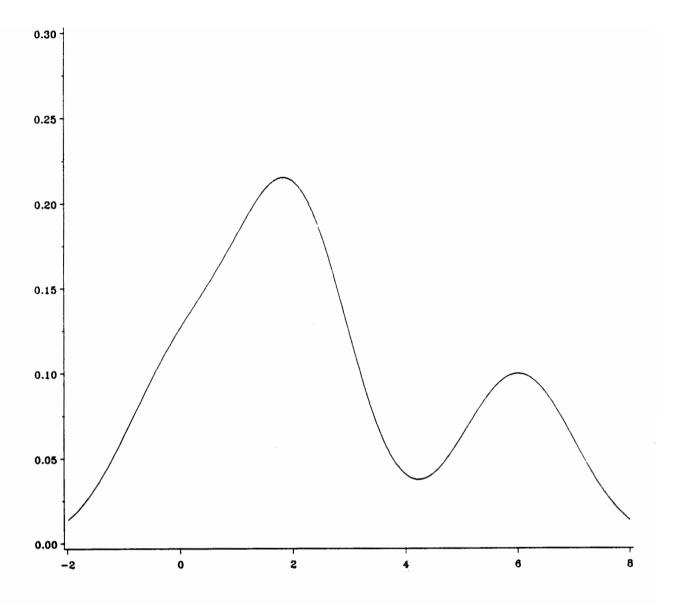Two Normal Components

Figure 3.2   Density Function for a Mixture of
Three Normal Components with

$$\mu_1 = 0, \ \mu_2 = 2, \ \mu_3 = 6$$

$$P_1 = .25, \ P_2 = .5$$

$$\sigma_1 = \sigma_2 = \sigma_3 = 1$$

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SMU/DS/TR-228	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* "Testing for Maximum Mean in a Mixture of Normals"		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H.L. Gray, Wayne A. Woodward, Gary McCartor		8. CONTRACT OR GRANT NUMBER(s) F19628-88-K-0042
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Department of Statistical Science Dallas, TX 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS*(if 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, if different from Report)*

18. SUPPLEMENTARY NOTES

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

Mixture of Normals, AIC, Bootstrap, Simulation, Nuclear Test Ban Treaty Verification

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

We consider the test of the null hypothesis that the largest mean in a mixture of an unknown number of normal components is less than or equal to some threshold. This test is motivated by the problem of assessing whether or not the Soviet Union has been operating in compliance with the Nuclear Test Ban Treaty. In our analysis, the number of normal components is assessed using AIC while the hypothesis test itself is based on asymptotic results given by Beehbodian for a mixture of two normal components. A bootstrap approach is also considered for estimating the standard error of the largest estimated mean. The performance of