Estimation using ordered data in a mixture of Normals based on evidential analysis

M. Arashi* and M. Emadi†

School of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box 91775-1159, Mashhad, Iran

Abstract

Often, in the essence of practical situations, some distribution functions come up with a finite mixture model for multinomial observations. To do inference for the mixture proportion and the parameters of a mixture of two Normal distributions using ordered data based on evidential analysis, we consider the misleading and weak evidences to obtain the true decision for underlying hypotheses.

Keywords: Statistical evidence, Likelihood ratio, Finite mixture distribution.

1 Introduction

1.1 History of Mixture Models

The finite mixture problem has quite a lengthy history, but in a short view, decomposing a finite mixture of distributions is a very difficult problem, as can be seen by looking at Pearson (1894), Tan and Chang (1972) and Fryer and Robertson (1972). Also, a full-book treatment on the subject of mixture distributions (with particular emphasis on normal mixtures) is provided by Titterington et al. (1985). For more details, see McLachlan and Basford (1988).

Introduction and summary for mixtures of normal distributions dating back to the late 19th century and the writings of Newcomb (1886) and Pearson (1894). The likelihood approach to the fitting of mixture models, in particular normal mixtures, has since been utilized by several authors, including Hosmer (1973a and b, 1974, 1978), O'neill (1978), Ganesalingam and Mclachlan (1978,1979a, 1980a).

In this paper, using evidential analysis we do inference for the mixture proportion of the ordered data arises from mixture of two symmetric Normal distributions.

Suppose the random variable (r.v.) X is distributed as a mixture of two Normal distributions like as $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$ with the following pdf;

$$\begin{aligned} f_X(x;\mu_1,\mu_2,\sigma_1,\sigma_2,\phi) &= \phi N(x;\mu_1,\sigma_1^2)+(1-\phi)N(x;\mu_2,\sigma_2^2) \\ &= \frac{1}{\sqrt{2\pi}}\Big\{\frac{\phi}{\sigma_1}\exp[\frac{-(x-\mu_1)^2}{2\sigma_1^2}]+\frac{(1-\phi)}{\sigma_2}\exp[\frac{-(x-\mu_2)^2}{2\sigma_2^2}]\Big\} \qquad (1)\end{aligned}$$

where $0\leq\phi\leq 1$.

*E-mail: m_arashi_stat@yahoo.com

†E-mail: emadi@math.um.ac.ir

1.2 Ordered

A systematic expos
David (2003) and G
populations has be
and Sarhan and Gr

1.3 Statistica

An important role
is "what the data
confidence intervals
no defined concept
a given body of dat
to it's sequel "can
(1997)). Emadi and
and Emadi (2005)
question is how a
the amount of stati
the probabilities o
iid observation. W
variable X under
observations X_1,
Let η be any measu
Then the probabil
$1-c) = K_1(1-$
of weak evidence
$W_2 = P_2(1-c<\eta$
is a threshold of st
Let λ be the likeli

Through out the p

2 Design

In this section, usi
problem of the mi
For simplicity, sup

where μ and σ ar
Let $X_1, X_2, \cdots, X$

1.2 Ordered Data

A systematic exposition of the theory of order statistics and extremes can be found in books by David (2003) and Galambos (1981). The problem of the distribution of order statistics in normal populations has been extensively considered in the literature. For full details, see Ruben (1954) and Sarhan and Greenberg (1962).

1.3 Statistical Evidence

An important role of statistical analysis in science is interpreting observed data as evidence, that is "what the data say?". Although standard statistical methods (hypothesis testing, estimation, confidence intervals) are routinely used for this purpose, the theory behind those methods contains no defined concept of evidence and no answer to the basic question "when is it correct to say that a given body of data represent evidence supporting one statistical hypothesis against another?" or to it's sequel "can we give an objective measure of the strength of statistical evidence?" (Royall (1997)). Emadi and Arghami (2003), Emadi et al (2006), Arashi and Emadi (2005) and Doostparast and Emadi (2005) have studied some measures of support for statistical hypotheses. An interesting question is how a number of observations verify the mixture of normal distributions , in terms of the amount of statistical evidence they provide about the unknown parameter(s). This article uses the probabilities of observing strong misleading evidence and weak evidence for the numbers of iid observation. We assume that f_i is the probability density function of a continuous random variable X under simple hypotheses H_i, $(i = 1, 2)$. Suppose we can observe the sequence of iid observations X_1, X_2, ..., where each is distributed as X.
Let η be any measure of support for one hypothesis against another with values in the unit interval. Then the probabilities of observing strong misleading evidence under H_1, H_2 are $M_1 = P_1(\eta \leq 1-c) = K_1(1-c)$ and $M_2 = P_2(\eta \geq c) = 1 - K_2(c)$, respectively, and the probabilities of weak evidence under H_1 and H_2 are $W_1 = P_1(1-c < \eta < c) = K_1(c) - K_1(1-c)$ and $W_2 = P_2(1-c < \eta < c) = K_2(c) - K_2(1-c)$, respectively (see, Royall (2000)). Here c, $0.5 \leq c < 1$, is a threshold of strong evidence, and K_1 and K_2 are cdf's of η under H_1 and H_2, respectively.
Let λ be the likelihood ratio for the competing hypotheses $H_1 : \theta = \theta_1$ and $H_2 : \theta = \theta_2$ so that

$$\lambda = \frac{L_1}{L_2} = \prod_{i=1}^{n} \frac{f_1(X_i)}{f_2(X_i)}. \tag{2}$$

Through out the paper we shall use $\eta = \lambda/(\lambda + 1)$ as a measure of support for H_1 against H_2.

2 Design

In this section, using ordered data achieved from the model in (1), we study the evidential inference problem of the mixture proportion based on the parameters of normal.
For simplicity, suppose the parameters in (1) have changed due to

$$\mu_1 = \mu, \quad \mu_2 = \mu + \alpha, \quad \sigma_1 = \sigma, \quad \sigma_2 = \beta\sigma \tag{3}$$

where μ and σ are known, α and β are unknown.
Let $X_1, X_2, \cdots, X_n$ be n r.v.s obtained from the model in (1) by the parameters in (4). Denote

the ordered data by $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$. The pdf of ordered data is as

$$f_{X_{(1)},X_{(2)},\cdots,X_{(n)}}(\underline{x};\mu,\sigma,\alpha,\beta,\phi) = n!\prod_{i=1}^{n}[\phi N(x_i,\mu,\sigma^2)+(1-\phi)N(x_i,\alpha+\mu,\beta^2\sigma^2)] \quad (4)$$

Following, two groups of hypotheses are considered for the mixture proportion in the underlying model. By the statistical evidences in ordered data for each hypothesis in one group, the competition between H_1 and H_2 has analyzed.

Process is repeated r=10000 times for n=50, 100, 200 r.vs obtained from the mixture model in (5). Values of M1, M2, W1 and W2 are found using (3) based on ordered data, in order to achieve the underlying goal. We have used the packages Maple 9.5 and Minitab 14 to do numeric computations.

$$Group1 = \begin{cases} H_1 : \phi = 0.5 \\ H_2 : \phi = 0.75 \end{cases} \quad (5)$$

$$Group2 = \begin{cases} H_1 : \phi = 0.5 \\ H_2 : \phi = 0.25 \end{cases} \quad (6)$$

2.1 Simulation

The scheme of sampling for the simulation is much important to find the correct results. It changes due to the hypothesis as follows.

1) For hypothesis $\phi = 0.5$, we take one random sample from discrete uniform in [0,1]($U[0,1]$); if it obtains 0, then we will take one random sample from $N(\mu,\sigma)$, and it will be from $N(\mu+\alpha,\beta^2\sigma^2)$ if it is 1.

2) For hypothesis $\phi = 0.25$, we take one random sample from discrete uniform in $U[0,3]$; if it obtains 0, then we will take one random sample from $N(\mu,\sigma)$, and it will be from $N(\mu+\alpha,\beta^2\sigma^2)$ if it is not.

3) For hypothesis $\phi = 0.75$, we take one random sample from discrete uniform in $U[0,3]$; if it obtains 0, then we will take one random sample from $N(\mu+\alpha,\beta^2\sigma^2)$, and it will be from $N(\mu,\sigma^2)$ if it is not.

For study the behavior of variation of mixture model with respect to(w.r.t) location and scale parameters in each normal distribution, we use simulation for some parameters α and β for the two groups of hypotheses. The graphical results for each group of parameters are as the same in the special case $\alpha = 5$ with $\beta = 1, 2, 5$ in Group1 of hypothesis.

2.2 Concluding Remarks

(1) For fix parameter α, the value of M1 is increasing w.r.t β. Furthermore, M1 is quite low, and the process of increasing behave in a low manner.

(2) Like as (1), under the hypothesis $\phi = 0.75$ the probability of weak evidences or strong misleading evidence of H_1 highly increasing w.r.t β.

(3) The probability of weak evidences under H_1 and H_2 sharply increasing w.r.t β.

(4) Smaller scale parameter causes smaller strong misleading and weak evidences.

(5) Emadi et al (2005) introduced optimality criteria (OC) as:

$$OC = 2(M1+M2)+(W1+W2).$$

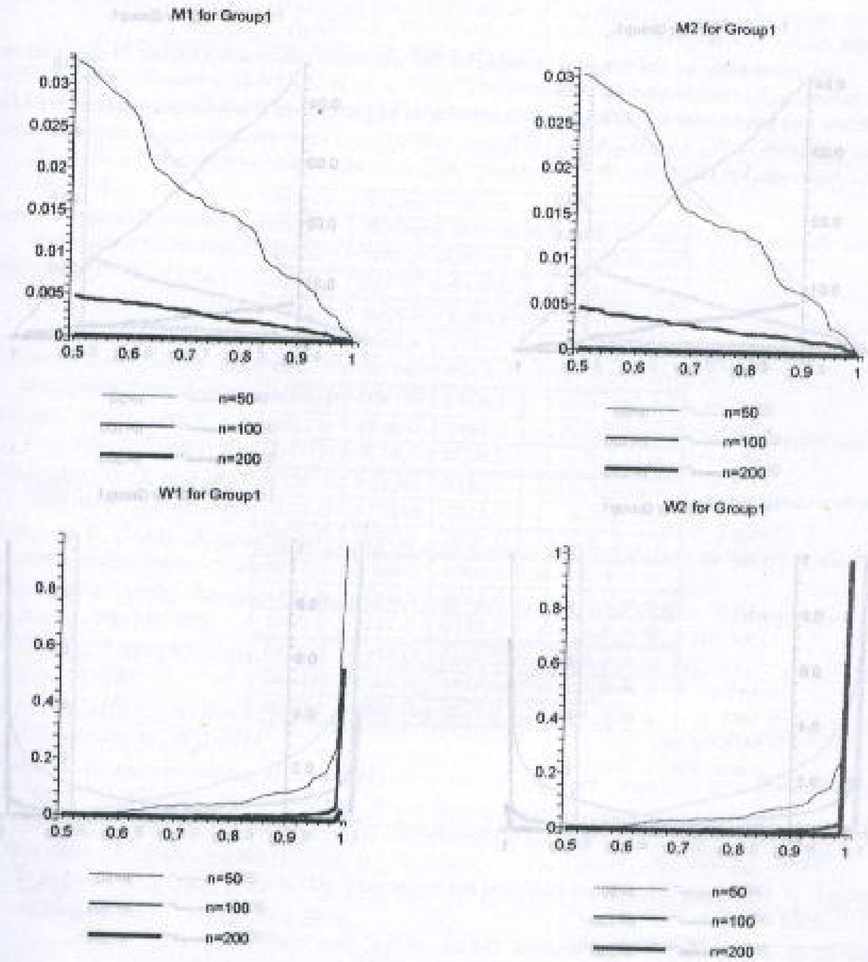

Figure 1: **alpha=5,beta=1**

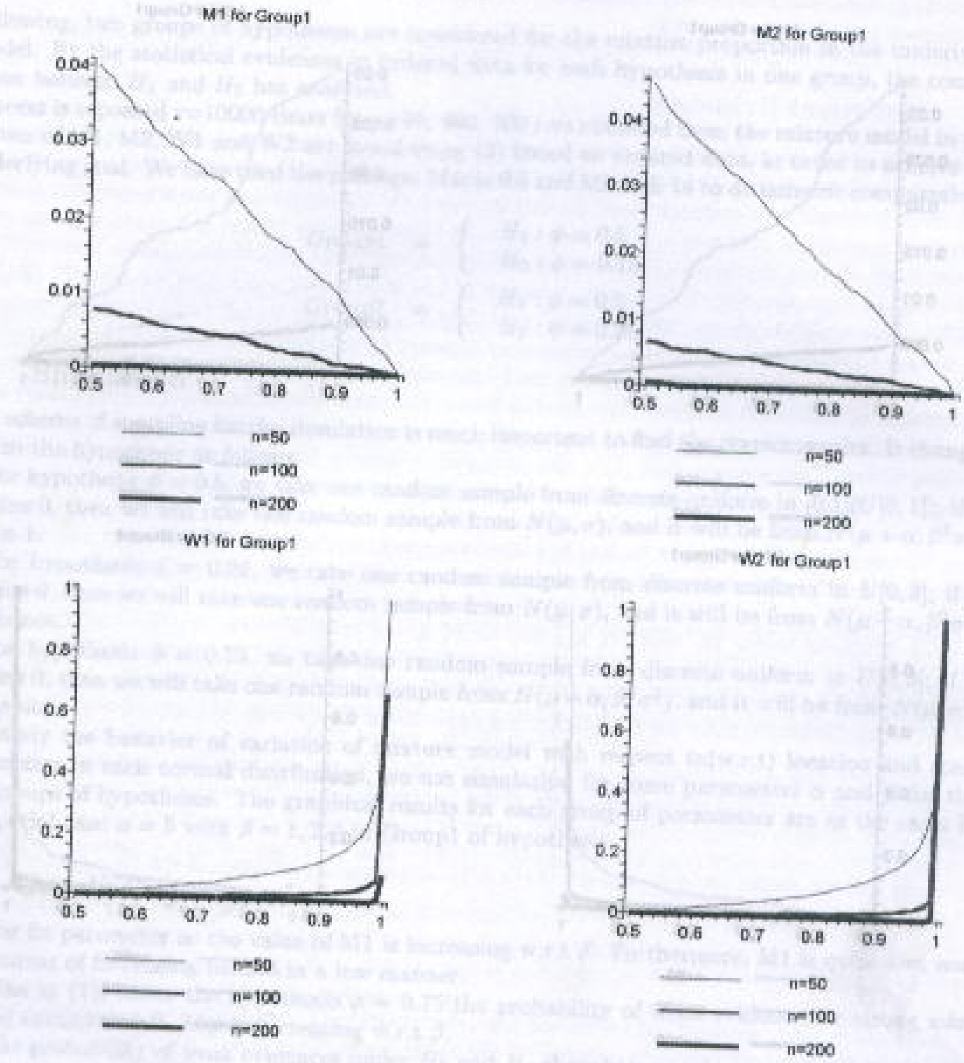

Figure 2: **alpha=5,beta=2**

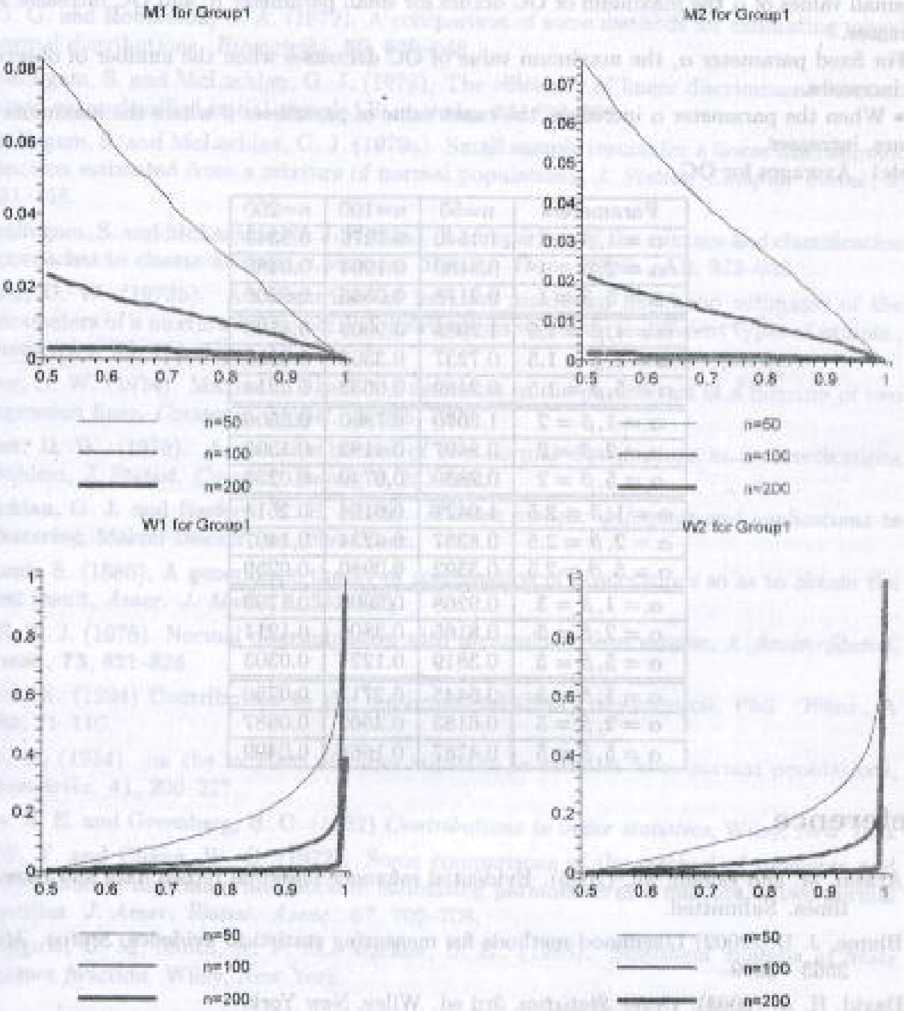

Figure 3: **alpha=5,beta=5**

Smaller value of OC shows that, evidential analysis has smaller faults.
In the Table1 averages of OC are computed for the hypotheses of Group1. From Table1, we could find that

• Plot of OC w.r.t the parameter β is unimodal, has a maximum point, which varies as α changes; for small values of α the maximum of OC occurs for small parameter β; and OC increases when β increases.

•• For fixed parameter α, the maximum value of OC decreases when the number of observations (n) increases.

••• When the parameter α increases, the exact value of parameter β where the maximum of OC occurs, increases.

Table1. Averages for OC

Parameters	n=50	n=100	n=200
$\alpha=1,\beta=1$	1.1540	0.7276	0.3343
$\alpha=2,\beta=1$	0.5180	0.1904	0.0489
$\alpha=5,\beta=1$	0.2178	0.0566	0.0206
$\alpha=1,\beta=1.5$	1.2988	0.9009	0.4708
$\alpha=2,\beta=1.5$	0.7237	0.3306	0.0985
$\alpha=5,\beta=1.5$	0.2489	0.0632	0.0214
$\alpha=1,\beta=2$	1.2070	0.7800	0.3606
$\alpha=2,\beta=2$	0.8497	0.4192	0.1393
$\alpha=5,\beta=2$	0.2830	0.0740	0.0225
$\alpha=1,\beta=2.5$	1.0427	0.6194	0.2514
$\alpha=2,\beta=2.5$	0.8387	0.4234	0.1407
$\alpha=5,\beta=2.5$	0.3392	0.0980	0.0259
$\alpha=1,\beta=3$	0.9268	0.5008	0.1799
$\alpha=2,\beta=3$	0.8165	0.3806	0.1211
$\alpha=5,\beta=3$	0.3819	0.1227	0.0303
$\alpha=1,\beta=5$	0.6445	0.2713	0.0750
$\alpha=2,\beta=5$	0.6183	0.2502	0.0687
$\alpha=5,\beta=5$	0.4787	0.1664	0.0409

Reference

[1] Arashi, M. and Emadi, M., (2005). Evidential inference based on record data and inter-record times. Submitted.

[2] Blume, J. D. (2002) Likelihood methods for measuring statistical evidence, *Statist. Med.* **21**, 2563 - 2599

[3] David, H. A. (2003). *Order Statistics*, 3rd ed. Wiley, New York.

[4] Doostparast, M. and Emadi, M. (2005). Statistical evidence methodology for model acceptance based on record values, Submitted.

[5] Emadi, M. and Arghami, N. R. (2003) Some measures of support for statistical hypotheses. *J. Stat. Theory Appl.*, **2**, 165–176

[6] Emadi, M., Ahmadi, J. and Arghami, N. R. (2006) Comparing of record data and random observation based on statistical evidence *To be appear in Statistical Papers*.

[7] Galambos, J. (1987). *The asymptotic theory of extreme order statistics*, 2nd edi., Kreiger, Florida.

[8] Evans, M., Hastings, N. and Peacock, B. (2000). *Statistical distribution*, Third Edition, John Wiley and Sons, New York.

[9] Fryer, J. G. and Roberston, C. A. (1972). A comparison of some methods for estimating mixed normal distributions. *Biometrika*, **59**, 639–648.

[10] Ganesalingam, S. and McLachlan, G. J. (1978). The efficiency of linear discriminant function based on unclassified initial sample, *Biometrika*, **65**, 658–662.

[11] Ganesalingam, S. and McLachlan, G. J. (1979a). Small sample results for a linear discriminant function estimated from a mixture of normal populations, *J. Statist. Comput. Simul.*, **9**, 151–158.

[12] Ganesalingam, S. and McLachlan, G. J. (1980a). A comparison of the mixture and classification approaches to cluster analysis, *Commun. Statist.- Theor. Meth.*, **A9**, 923–933.

[13] Hosmer, D. W. (1973b). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample., *Biometrics*, **29**, 761–770.

[14] Hosmer, D. W. (1974). Maximum likelihood estimates of the parameters of a mixture of two regression lines, *Commun. Statist.*, **3**, 995–1006.

[15] Hosmer, D. W. (1978). A use of mixtures of two normal distributions in a classifications problem, *J. Statist. Comput. Simul.*, **6**, 281–294.

[16] McLachlan, G. J. and Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*, Marcel Dekker Inc., New York.

[17] Newcomb, S. (1886). A generalized theory of combination of observations so as to obtain the best result, *Amer. J. Math.*, **8**, 343–366.

[18] O'Neill, T. J. (1978). Normal discrimination with unclassified observations, *J. Amer. Statist. Assoc.*, **73**, 821–826.

[19] Pearson, K. (1894) Contribution to the mathematical theory of evolution, *Phil. Trans.*, A **185**, 71–110.

[20] Ruben, H. (1954). on the moment of order statistics in samples from normal populations, *Biometrika*, **41**, 200–227.

[21] Sarhan, A. E. and Greenberg, B. G. (1962) *Contributions to order statistics*, Wiley, New York

[22] Tan, W. Y. and Chang, W. C. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. *J. Amer. Statist. Assoc.*, **67**, 702–708.

[23] Titterington, D. M, Smith, A. F. M., Markov, U. E., (1985). *Statistical analysis of finite mixture function*. Wiley, New York.