Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Mixture Models in View of Evidential Analysis

M. Emadi [a]

[a] Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran

Version of record first published: 02 Jan 2013.

To cite this article: M. Emadi (2013): Mixture Models in View of Evidential Analysis, Communications in Statistics - Simulation and Computation, DOI:10.1080/03610918.2012.677923

To link to this article: http://dx.doi.org/10.1080/03610918.2012.677923

Communications in Statistics—Simulation and Computation®, 00: 1–12, 2013

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2012.677923

Mixture Models in View of Evidential Analysis

M. EMADI

Department of Statistics, Ordered and Spatial Data Center of Excellence,
Ferdowsi University of Mashhad, Mashhad, Iran

In some practical inferential situations, it is needed to mix some finite sort of distributions to fit an adequate model for multi-modal observations. In this article, using evidential analysis, we determine the sample size for supporting hypotheses about the mixture proportion and homogeneity. An Expectation-Maximization algorithm is used to evaluate the probability of strong misleading evidence based on modified likelihood ratio as a measure of support.

Keywords Finite mixture distribution; Modified likelihood ratio; Penalty term; Statistical evidence; Strong misleading evidence

Mathematics Subject Classification 62f03.

1. Introduction

Introduction and summary mixtures of normal distributions have a long history in statistics, dating back to the late 19th century and the writings of Newcomb (1886) and Pearson (1894). Since then, they appear as models in diverse areas of applied research. However, even in the simplest of cases, the two-component normal mixture, one encounters serious theoretical as well as computational difficulties when attempting to perform basic statistical analysis such as parameter estimation and goodness of fit. Inference procedures based on statistical transforms, such as the empirical characteristic function and the empirical moment generating function, were introduced by Press (1972) (parameter estimation) and Heathcote (1972) (goodness of fit) and have been studied by Bernhard and Meintanis (2005) in the empirical characterization for normal mixtures. Following Rao's (1948) article likelihood estimation appears not to have been pursued further until Hasselblad (1966, 1969) addressed the problem, initially for a mixture of g univariate normal distributions with equal variances. The likelihood approach to the fitting of mixture models, in particular normal mixtures, has since been utilized by several authors, including Hosmer (1973a, 1973b, 1974, 1978), O'neill (1978), and Ganesalingam and Mclachlan (1978, 1979, 1980).

Butler (1986) noted that Jeffreys (1932) used essentially the Expectation-Maximization (EM) algorithm in iteratively computing the estimates of the means of two univariate normal population, which had known variances and which were mixed in known proportions.

In this article, we basically engage with the problem of testing in a finite mixture of normal distributions. As a prelude, consider the following finite mixture model with the cdf

Received November 12, 2012; Accepted March 14, 2012

Address correspondence to M. Emadi, Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box 91775-1159, Mashhad, Iran; E-mail: emadi@um.ac.ir

given by

$$F(x) = \sum_{j=1}^{m} p_j \Phi\left(\frac{x_j - \mu_j}{\sigma_j}\right), \tag{1.1}$$

where $\Phi(.)$ is the cdf of standard normal distribution. Further, $\sum_{i=1}^{m} p_i = 1$.

It is well known that if X is distributed according to (1.1), then the mean, variance, skewness, and kurtosis of X, respectively, will be given by

$$\mu = \sum_{j=1}^{m} p_j \mu_j,$$

$$\sigma^2 = \sum_{j=1}^{m} p_j \left(\sigma_j^2 + \mu_j^2\right) - \mu^2,$$

$$\alpha_3 = \frac{1}{\sigma^3} \sum_{j=1}^{m} p_j(\mu_j - \mu)\left[3\sigma_j^3 + (\mu_j - \mu)^2\right],$$

$$\alpha_4 = \frac{1}{\sigma^4} \sum_{j=1}^{m} p_j \left[3\sigma_j^4 + 6(\mu_j - \mu)^2\sigma_j^2 + (\mu_j - \mu)^4\right].$$

Figures 1–3 shows the graph of mixture of two normal distributions with prespecified values $m = 2,\ p_1 = p_2 = p,\ \sigma 1 = \sigma 2 = \sigma,\ \mu_1$, and μ_2.

1.1. *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 its sequel "can we give an objective measure of the strength of statistical evidence?" (Royall 1997). Emadi and Arghami (2003), Emadi et al. (2007), Arashi and Emadi (2007), and Doostparast and Emadi (2006) 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 $X_1, X_2, \ldots,$ 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 $M_1 = P_1(\eta \le 1 - c) = K_1(1 - c)$ and $M_2 = P_2(\eta \ge c) = 1 - K_2(c)$, 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, $0.5 \le c < 1$, is a threshold of strong evidence, and K_1 and K_2 are cdfs of η under H_1 and H_2, respectively. We argue that since both misleading and weak evidence are undesirable, and obtaining strong misleading evidence is more important than obtaining

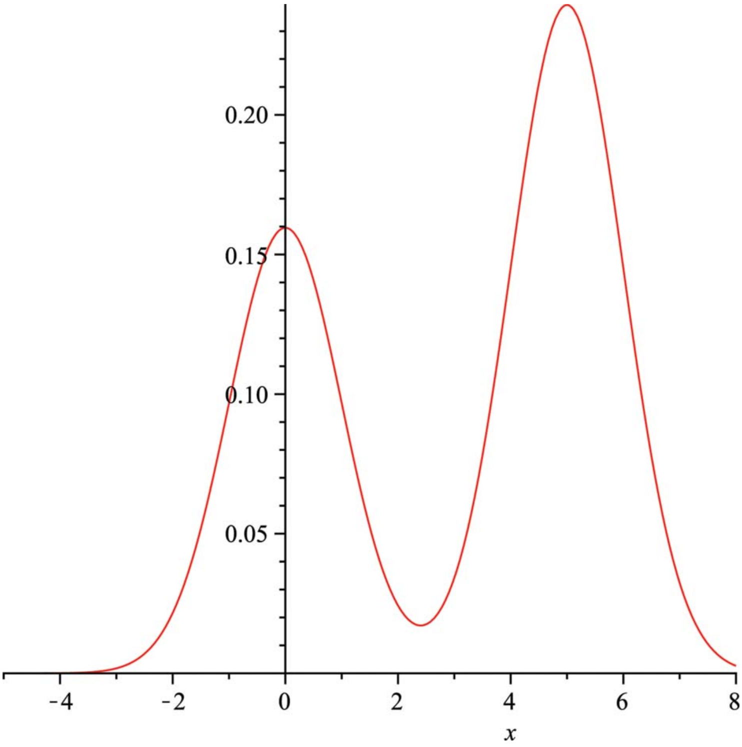

Figure 1. $p = 0.4$, $\sigma = 1$, $\mu_1 = 0$, $\mu_2 = 5$.

just weak evidence, a preexperimental measure desirability of a measure of evidence can be taken to be

$$e(\eta, \gamma) = 1 - \int_{0.5}^{1} \{\gamma[M_1(t) + M_2(t)] + W_1(t) + W_2(t)\}\, dt, \tag{1.2}$$

where $\gamma \geq 1$ and M_i and W_i are, respectively, the probabilities of strong misleading evidence and weak evidence under H_i, $(i = 1, 2)$. The following theorem gives $e(\eta)$ in terms of K_1 and K_2.

Theorem 1.1. *(Emadi et al. 2007)* Under the assumptions of this section, we have

$$e(\eta, \gamma) = \int_0^1 (K_2(t) - K_1(t))dt + (2 - \gamma)\left(\frac{1}{2} + \int_0^{0.5} K_1(t)dt - \int_{0.5}^{1} K_2(t)dt\right).$$

Theorem 1.1 has following corollary.

Corollary 1.1. *Under the assumptions of Theorem 1.1, for* $\gamma = 2$ *we have*

$$\begin{aligned} e(\eta) = e(\eta, 2) &= \int_0^1 [K_2(t) - K_1(t)]dt \\ &= E_{H_1}(\eta) - E_{H_2}(\eta), \end{aligned}$$

where $E_{H_i}(\eta)$ *is the expectation of* η *under* H_i.

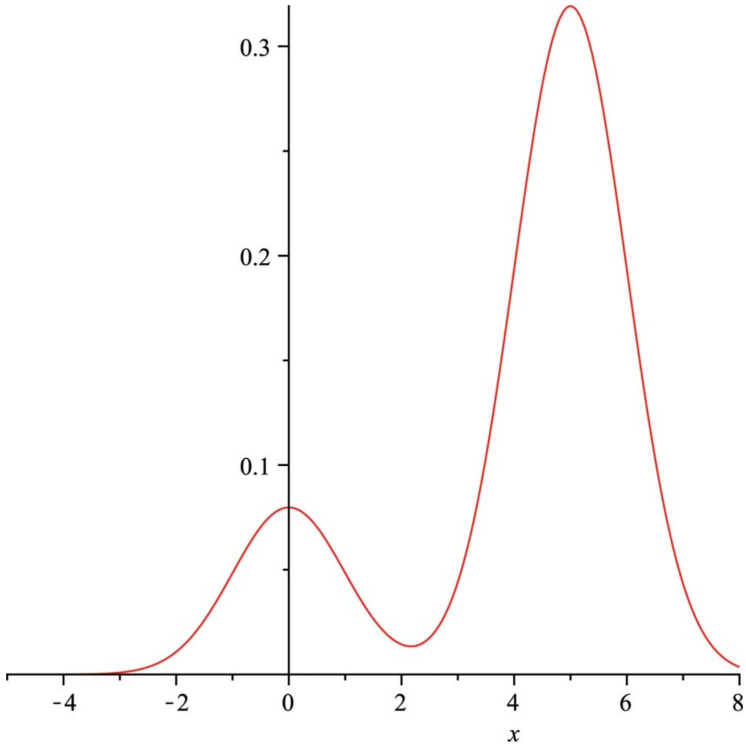

Figure 2. $p = 0.2, \sigma = 1, \mu_1 = 0, \mu_2 = 5$.

It is interesting to note that for $\gamma = 2$, $e(\eta)$ (which was introduced and used by Emadi and Arghami 2003) has another interpretation, this being the area (with unit square) between the curves of $K_1(t)$ and $K_2(t)$.

Let λ be the likelihood ratio for the competing hypotheses $H_1 : \theta = \theta_1$ and $H_2 : \theta = \theta_2$, so that

$$\lambda = \frac{L(\theta_1)}{L(\theta_2)}, \tag{1.3}$$

where $L(\theta_i)$ is the likelihood function when the parameter of interest $\theta = \theta_i$.

Throughout the article, we shall use λ as a measure of support for H_1 against H_2.

2. Determining Sample Size

Assuming the number $m = 2$, normal distributions are taken in a mixture model (1.1). In this case, the density of X is given by

$$g(x) = pf(x, \mu_1, \sigma^2) + (1 - p)f(x, \mu_2, \sigma^2), \tag{2.1}$$

where $f(., \mu, \sigma^2)$ is the density function of $N(\mu, \sigma^2)$ for known σ^2.

In this section, we determine the sample size based on statistical evidences for testing H_1 against H_2. For $m > 2$, the proposed method will be efficient.

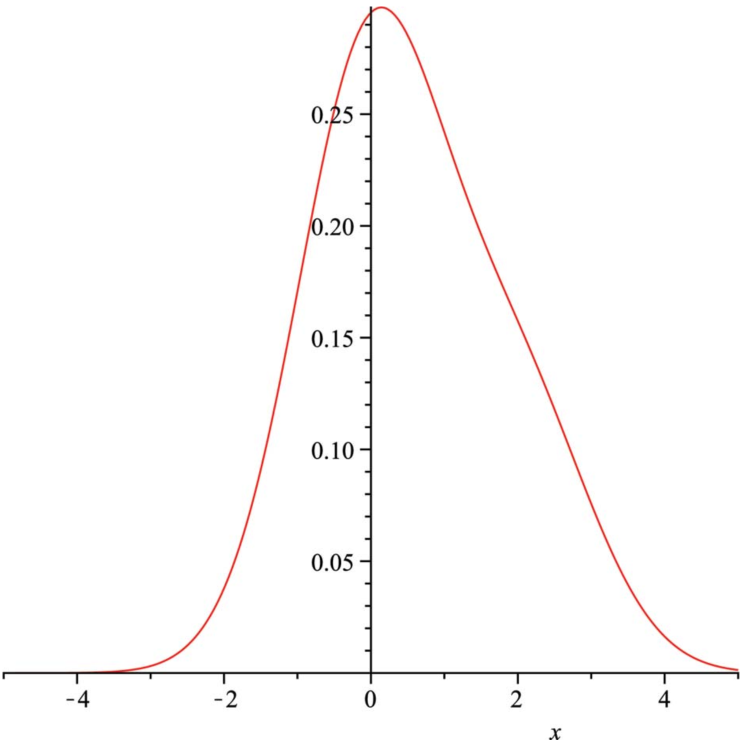

Figure 3. $p = 0.3, \sigma = 1, \mu_1 = 0, \mu_2 = 2.$

Our primary target is to define a criterion for which we can determine the sample size for testing the mixture proportion.

From the fact that probabilities of strong misleading evidence $P_{H_1}(\lambda > k)$ and $P_{H_2}(\lambda < \frac{1}{k})$ have the same influence on determining the sample size, in this situation we use the following formula to determine the sample size

$$n^* = \min\left\{n \in \mathbb{N} : \frac{1}{2}\left[P_{H_1}(\lambda > k) + P_{H_2}\left(\lambda < \frac{1}{k}\right)\right] > \epsilon\right\},$$

where $\epsilon \in (0, 1)$.

It is clear that as ϵ or k gets larger, the sample size increases.

2.1. *Testing Ratio*

In this part, we determine the sample size for the following group of hypotheses:

$$\begin{cases} H_1 : p = p_1 \\ H_2 : p = p_2 \end{cases} \tag{2.2}$$

In order to find the sample size, we need to compute the probability of strong misleading evidence; that it can be done trough finding the values $P_{H_i}(\lambda > k), i = 1, 2.$

From (1.3) together with (2.1), for a random sample of size n from (2.1) we can obtain

$$\lambda = \prod_{i=1}^{n} \frac{p_1 + (1-p_1)t_i}{p_2 + (1-p_2)t_i}, \quad t_i = \frac{f(x_i, \mu_2, \sigma^2)}{f(x_i, \mu_1, \sigma^2)}. \tag{2.3}$$

Then by making use of (2.3), the strong misleading evidence under H_i is given by

$$P_{H_i}(\lambda > k) = P_{H_i}\left[\sum_{i=1}^{n} \ln\left(\frac{p_1 + (1-p_1)t_i}{p_2 + (1-p_2)t_i} > \ln k\right)\right]. \tag{2.4}$$

The expression in (2.4) cannot be computed directly. In this situation, one may use the central limit theorem to calculate the probability of strong misleading evidence as in the following:

$$P_{H_i}(\lambda > k) = P_{H_i}\left[\frac{\sum_{i=1}^{n} \ln\left(\frac{p_1+(1-p_1)t_i}{p_2+(1-p_2)t_i}\right) - \mu_i^*}{\sigma^*} > \frac{\ln k - \mu_i^*}{\sigma^*}\right], \tag{2.5}$$

where μ_i^* and σ^* are means and standard deviation of new random variable ln(λ) under H_i, respectively. At this stage, it is sufficient to calculate μ_i^* and σ^*.

Now let

$$H(T) = \ln\left(\frac{p_1 + (1-p_1)T}{p_2 + (1-p_2)T}\right).$$

Using Taylor's expansion, we can conclude

$$\mu_i^* = \mathrm{E}[H(T_i)] \; = \; H(\mathrm{E}(t_i)) + \frac{1}{2}\mathrm{Var}(t_i)H''(\mathrm{E}(t_i)), \tag{2.6}$$

$$\sigma^{*2} = \mathrm{Var}[H(T_i)] \; = \; \mathrm{Var}(t_i)[H'(\mathrm{E}(t_i))]^2, \tag{2.7}$$

where by some algebraic computations, we have

$$\mathrm{E}(t_i) = p + (1-p)\exp\left[\left(\frac{\mu_2 - \mu_1}{\sigma}\right)^2\right],$$

$$\mathrm{Var}(t_i) = p\exp\left[\left(\frac{\mu_2 - \mu_1}{\sigma}\right)^2\right] + (1-p)\exp\left[3\left(\frac{\mu_2 - \mu_1}{\sigma}\right)^2\right]$$
$$-p^2 - (1-p)^2\exp\left[2\left(\frac{\mu_2 - \mu_1}{\sigma}\right)^2\right] - 2p(1-p)\exp\left[\left(\frac{\mu_2 - \mu_1}{\sigma}\right)^2\right].$$

In sequel, we tabulate some values of sample size for different k (Table 1).

2.2. *Testing Homogeneity*

In this part, we consider the following homogeneity hypotheses:

$$\begin{cases} H_1 : p(1-p)(\mu_1 - \mu_2) = 0 \\ H_2 : p(1-p)(\mu_1 - \mu_2) \neq 0 \end{cases} \tag{2.8}$$

Table 1
Estimated sample size

k	p_1	p_2	μ_1	μ_2	σ	n^*
3	0.1	0.7	1	1.5	1	54
	0.1	0.7	1	1.5	2	200
	0.2	0.9	1	1.5	2	147
	0.2	0.9	1	3	5	60
	0.5	0.8	1	1.5	1	258
8	0.1	0.7	1	1.5	1	85
	0.1	0.7	1	1.5	2	314
	0.2	0.9	1	1.5	2	231
	0.2	0.9	1	3	5	94
	0.5	0.8	1	1.5	1	406
32	0.1	0.7	1	1.5	1	128
	0.1	0.7	1	1.5	2	467
	0.2	0.9	1	1.5	2	343
	0.2	0.9	1	3	5	139
	0.5	0.8	1	1.5	1	616

The hypotheses in (2.8) represents whether we can consider a mixture model as the underlying sampling model for the population or not. The hypothesis $H_1 : p(1-p)(\mu_1-\mu_2) = 0$ shows that the population model is homogeneous and it has not any sub-population. This occurs whenever $p \in \{0, 1\}$ or $\mu_1 = \mu_2$. On the other hand, the hypothesis $H_2 : p(1-p)(\mu_1-\mu_2) \neq 0$ shows that the relevant model for the underlying population consists of a mixture of two normal distributions.

To test the null hypothesis in (2.8) under evidential perspective, we use the likelihood ratio given by (1.3). It is fully known that under some regularity conditions, $-2\ln\lambda$ is asymptotically distributed according to a chi-square. The main problem here is that the mixture models do not meet the regularity conditions. Thus, it is not possible to apply chi-square distribution for testing.

To overcome this difficulty, Chen (1998) suggested to use a modified likelihood ratio by adding a penalty term $T(p)$ to the likelihood ratio, with the following properties:

$$\lim_{p\to\ 0\ \vee\ 1} T(p) = -\infty, \qquad \arg\max_{p\in[0,1]} T(p) = 0.5. \tag{2.9}$$

Using this modified likelihood ratio, we obtain fitted ps far from zero and unit. He also demonstrated that if σ be known, then the asymptotic distribution of the modified likelihood ratio would be a mixture with equal weights as $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. More interesting, this asymptotic distribution does not dependent on the form of penalty term.

Chen and Kalbfleisch (2001) then suggested the following penalty term:

$$T(p) = c\ln[4p(1-p)], \quad c > 0. \tag{2.10}$$

Afterward, considered the following penalty term:

$$T(p) = c\ln[1-|1-2p|], \quad c > 0. \tag{2.11}$$

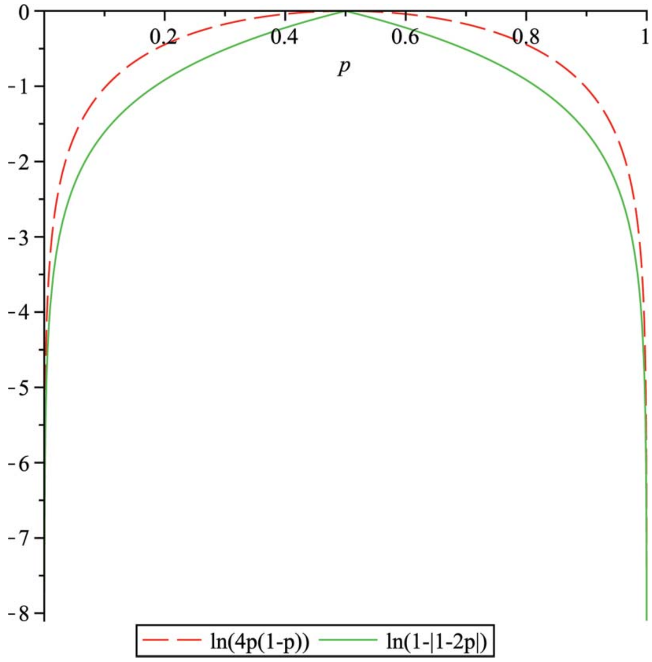

Figure 4. Graph of penalty term.

The proposed penalty terms (2.10) and (2.11) meet the conditions given by (2.9). Figure 4 displays the graph of penalty terms (2.10) and (2.11) for $c = 1$. As it is seen, as p gets closer to zero or unit, the penalty term loads the likelihood ratio more.

The expressions (2.12) and (2.13) propose the estimates of p using EM algorithm (see the Appendix) based on the penalty terms (2.10) and (2.11), respectively.

$$p^{(t+1)} = \frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{2c+n}, \tag{2.12}$$

$$p^{(t+1)} = \begin{cases} \min\left[\frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{c+n}, 0.5\right], & \frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{n} < 0.5 \\ 0.5, & \frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{n} = 0.5 \\ \max\left[\frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{c+n}, 0.5\right], & \frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{n} > 0.5 \end{cases}, \tag{2.13}$$

where T_{ij} is defined by (A.1) in the Appendix.

Now we need to define a measure of support to determine the sample size. In this case, we first modify the log-likelihood function by adding the penalty term (2.10) to the log likelihood function, which yields

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(x_i, \boldsymbol{\theta}) + c \ln[4p(1-p)].$$

Estimating the unknown parameter $\boldsymbol{\theta}$ gives the following modified log-likelihood ratio

$$\lambda_n = 2\left\{ \sup_{\boldsymbol{\theta}\in H_1} l_n(\hat{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in H_o} l_n(\hat{\boldsymbol{\theta}}) \right\}. \tag{2.14}$$

Chen and Chen (2003) derived an upper bound for λ_n, and then demonstrated that its asymptotic distribution is chi-square with two degrees of freedom. We again emphasize that in mixture models the regularity conditions do not satisfy, therefore, we cannot use chi-square distribution.

In order to evaluate the strong misleading evidence to determine the sample size, we need to compute the following probability:

$$P_{H_1}(\lambda > k) = P_{H_1}\left(\frac{\sup_{\Theta_1} l_n(\boldsymbol{\theta})}{\sup_{\Theta_2} l_n(\boldsymbol{\theta})} > k' \right), \tag{2.15}$$

where $\boldsymbol{\Theta}_i$ is the parameter space under H_i, $i = 1, 2$. But the probability in (2.15) cannot be computed since the distribution of λ is unknown. In this case, we apply the modified log-likelihood ratio given by (2.14) to evaluate the strong misleading evidence as follows:

$$P_{H_1}(\lambda_n > k) = P_{H_1}\left(\frac{\sup_{\Theta_1} l_n(\hat{\boldsymbol{\theta}})}{\sup_{\Theta_2} l_n(\hat{\boldsymbol{\theta}})} > k' \right). \tag{2.16}$$

Larger value $P_{H_1}(\lambda_n > k)$ supports H_1 more.

3. Conclusion

In this article, we consider the problem of determining the sample size for testing problem under evidential analysis. In fact we looked for a number of samples in which we can have strong misleading evidence on supporting one hypothesis in mixture models.

As the parameter ratio plays deterministic role in fitting mixture models to a dataset (see Figs. 1–3 for its importance), we considered hypotheses about its true value and homogeneity as well.

We evaluated the strong misleading evidence for supporting two sets of hypotheses to determine the sample size using Taylor's expansion and EM algorithm. For the purpose of probability evaluation, we applied modified likelihood ratio as a measure of support, since its asymptotic distribution could not be applied in mixture models.

Appendix

In this section, we propose the estimates of parameters of a mixture of two normal distributions using EM algorithm.

Suppose $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample from the model (2.1) and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ is a latent variable such that

$$X_i|(Z_i = 1) \sim N(\mu_1, \sigma_1^2), \quad \text{and} \quad X_i|(Z_i = 2) \sim N(\mu_2, \sigma_2^2),$$

$P(Z_i = 1) = p_1$, and $P(Z_i = 2) = p_2 = 1 - p_1$, $i = 1, \ldots, n$, and $\boldsymbol{\theta} = (p_1, p_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ is the unknown vector parameter.

Then the likelihood functions are given by

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}) &= \prod_{i=1}^{n} \prod_{j=1}^{2} I(Z_i = j) p_j f(x_i, \mu_j, \sigma_j) \\ &= \exp\left\{ \sum_{i=1}^{n} \sum_{j=1}^{2} I(Z_i = j) \left[\log(p_j) - \frac{\log(\sigma_j^2) + (x_i - \mu_j)^2 - \log(2\pi)}{2\sigma_j^2} \right] \right\}, \end{aligned}$$

where $I(.)$ is the indicator function.

Now assume that $\boldsymbol{\theta}^{(t)}$ be the initial estimate of $\boldsymbol{\theta}$ and let

$$\begin{aligned} T_{j,i}^{(t)} &= P(Z_i = j | X_i = x_i; \boldsymbol{\theta}^{(t)}) \\ &= \frac{p_j^{(t)} f(x_i, \mu_j^{(t)}), \sigma_j^{(t)}}{p_1^{(t)} f(x_i, \mu_1^{(t)}), \sigma_1^{(t)} + p_2^{(t)} f(x_i, \mu_2^{(t)}), \sigma_2^{(t)}}, \end{aligned} \tag{A.1}$$

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) &= E_{\boldsymbol{Z}|\boldsymbol{X}} \left[\log L(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}) \right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{2} T_{j,i}^{(t)} \left[\log(p_j) - \frac{\log(\sigma_j^2) + (x_i - \mu_j)^2 - \log(2\pi)}{2\sigma_j^2} \right]. \end{aligned} \tag{A.2}$$

At this stage, we evaluate the parameters for which the term $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$ maximizes. In other words,

$$\begin{aligned} p^{(t+1)} &= \arg \max_{p} [Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})] \\ &= \arg \max_{p} \left\{ \left[\sum_{i=1}^{n} T_{1,i}^{(t)} \right] \log(p_1) + \left[\sum_{i=1}^{n} T_{2,i}^{(t)} \right] \log(p_2) \right\}. \end{aligned}$$

Thus, for the estimates of p_j, $j = 1, 2$, we obtain

$$p_j^{(t+1)} = \frac{\sum_{i=1}^{n} T_{j,i}^{(t)}}{\sum_{i=1}^{n} \left(T_{1,i}^{(t)} + T_{2,i}^{(t)} \right)} = \frac{1}{n} \sum_{i=1}^{n} T_{j,i}^{(t)}. \tag{A.3}$$

Also for the estimates of (μ_1, σ_1^2), we have

$$\begin{aligned} \left(\mu_1^{(t+1)}, \sigma_1^{(t+1)} \right) &= \arg \max_{(\mu_1, \sigma_1^2)} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) \\ &= \arg \max_{(\mu_1, \sigma_1^2)} \left[\sum_{i=1}^{n} T_{1,i}^{(t)} \left\{ - \frac{\log(\sigma_1^2) + (x_i - \mu_1)^2}{2\sigma_1^2} \right\} \right], \end{aligned}$$

resulting in

$$\mu_1^{(t+1)} = \frac{\sum_{i=1}^{n} x_i T_{1,i}^{(t)}}{\sum_{i=1}^{n} T_{1,i}^{(t)}},$$

$$\sigma_1^{(t+1)} = \frac{\sum_{i=1}^{n} \left(x_i - \mu_1^{(t+1)}\right)^2 T_{1,i}^{(t)}}{\sum_{i=1}^{n} T_{1,i}^{(t)}}. \tag{A.4}$$

In a similar fashion, we get

$$\mu_2^{(t+1)} = \frac{\sum_{i=1}^{n} x_i T_{2,i}^{(t)}}{\sum_{i=1}^{n} T_{2,i}^{(t)}},$$

$$\sigma_2^{(t+1)} = \frac{\sum_{i=1}^{n} \left(x_i - \mu_2^{(t+1)}\right)^2 T_{2,i}^{(t)}}{\sum_{i=1}^{n} T_{2,i}^{(t)}}. \tag{A.5}$$

Acknowledgments

The author thanks Dr. M. Arashi (Assistant Professor, School of Mathematics, Shahrood University of Technology, Shahrood, Iran) for his comments on the manuscript. Also, the author is grateful to an anonymous referee for his/her helpful comments and suggestions on the initial version of this article. This research was supported by a grant from Ferdowsi Universirt of Mashhad; (No.[MS88119EMA]).

References

Arashi, M., Emadi, M. (2007). Evidential inference based on record data and inter-record times. *Statistical Papers* 49:291–301.

Bernhard, K., Meintanis, S. G. (2005). Test for normal mixtures based on the empirical characteristic function. *Journal of Computational Statistics and Data Analysis* 49:227–242.

Butler, R. W. (1986). Predictive likelihood inference with applications (with discussion). *Journal of the Royal Statistical Society: Series B*48:1–38.

Chen, H., Chen, J. (2003). Tests for homogeneity in normal mixture with presence of a structural parameter. *Statistica Sinica* 13:351–365.

Chen, H., Kalbfleisch, J. D. (2001). The likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society: Series B* 63:19–29.

Chen, H., Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with structural parameter. *Journal of Statistical Planning and Inference* 129:93–107.

Chen, J. (1998). Penalized likelihood ratio test for finite mixture model with multinomial observations. *Canadian Journal of Statistics* 26:583–599.

Doostparast, M., Emadi, M. (2006). Statistical evidence methodology for model acceptance based on record values. *Journal of the Korean Statistical Society* 35:167–177.

Emadi, M., Ahmadi, J., Arghami, N. R. (2007). Comparing of record data and random observation based on statistical evidence. *Statistical Papers* 48:1–21.

Emadi, M., Arghami, N. R. (2003). Some measures of support for statistical hypotheses. *Journal of Statistical Theory and Application* 2:165–176.

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

Ganesalingam, S., McLachlan, G. J. (1979). Small sample results for a linear discriminant function estimated from a mixture of normal populations. *Journal of Statistical Computational and Simulation* 9:151–158.

Ganesalingam, S., McLachlan, G. J. (1980). A comparison of the mixture and classification approaches to cluster analysis. *Communications in Statistics - Theory and Methods* A9:923–933.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Trigonometrics* 8:431–444.

Hasselblad, V. (1969). Estimation of finite mixtures of distribution from the exponential family. *Journal of the American Statistical Association* 64:1459–1471.

Heathcote, C. R. (1972). A test of goodness of fit for symmetric random variables. *The Australian Journal of Statistics* 14:172–181.

Hosmer, D. W. (1973a). On MLE of the parameters of a mixture of two normal distributions when the sample size is small. *Communications in Statistics* 1:217–227.

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.

Hosmer, D. W. (1974). Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics* 3:995–1006.

Hosmer, D. W. (1978). A use of mixtures of two normal distributions in a classifications problem. *Journal of Statistical Computational and Simulation* 6:281–294.

Jeffreys, H. (1932). An alternative to the rejection of observations. *Proceedings of the Royal Society* LondonA137:78–87.

Newcomb, S. (1886). A generalized theory of combination of observations so as to obtain the best result. *The American Journal of Mathematics* 8:343–366.

O'Neill, T. J. (1978). Normal discrimination with unclassified observations. *Journal of the American Statistical Association* 73:821–826.

Pearson, K. (1894). Contribution to the mathematical theory of evolution. *Philosophical Transactions* A185:71–110.

Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association* 67:842–846.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society: Series B* 10:159–203.

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. New York: Chapman & Hall.

Royall, R. (2000). On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association* 95:760–780.