

Order Selection in a Finite Mixture of Birnbaum-Saunders Distributions

Walaa A. El-Sharkawy
Cairo University

Moshira A. Ismail
Cairo University

Abstract

One of the most significant and difficult problems in a mixture study is the selection of the number of components. In this paper, using a Monte Carlo study, we evaluate and compare the performance of several information criteria for selecting the number of components arising from a mixture of Birnbaum-Saunders distributions. In our comparison, we consider information criteria based on likelihood-based statistics and classification likelihood-based statistics. The performance of information criteria is determined based on the success rate in selecting the number of components. In the simulation study, we investigate the effect of degrees of separation, sample sizes, mixing proportions, and true model complexity on the performance of information criteria. Furthermore, we compare the performance of the proposed information criteria under unpenalized and penalized estimation. Finally, we discuss the performance of the proposed information criteria for a real data set.

Keywords: EM algorithm, finite mixture, Monte Carlo method, order selection, penalized estimation.

1. Introduction

Finite mixture models are a popular statistical modeling technique used to model situations in which data arises from a population consisting of several homogeneous sub-populations. Several authors have applied mixture modeling in various practical issues such as agriculture, botany, economics, industrial engineering, marketing, medicine, reliability, survival analysis, zoology, and so on. Titterton, Smith, and Makov (1985), McLachlan and Basford (1988), Lindsay (1995), McLachlan and Peel (2000), Böhning, Seidel, Alfö, Garel, Patilea, and Walther (2007), Schlattmann (2009), and references therein include detailed description and applications of mixture modeling.

One of the most important tasks in mixture modeling is selecting the appropriate number of components that describe the data, since the number of components g refers to the number of separate groups in the data. Proper selection of the correct number of components is critical because the number of components selected can have a significant effect on the practical interpretations of the modeling results. In the literature, different approaches have been suggested for selecting the number of components in mixture models, in both frequentist and Bayesian settings. These approaches can be classified as, information-based approaches,

penalized distance-based approaches, penalized log-likelihood-based approaches, Bayesian approaches, and testing hypothesis-based approaches (see, [McLachlan and Peel \(2000\)](#) and [Niu \(2014\)](#) for a detailed review).

In this study, we focus on information-based approaches to select the number of components in a Birnbaum-Saunders (BS) mixture model. The information criteria can be categorized into statistics based on the log-likelihood function and statistics based on the classification function. The methodology of using information criteria in order selection in mixture models consists of estimating the models with different number of components using the expectation-maximization (EM) algorithm ([Dempster, Laird, and Rubin 1977](#)) and comparing those estimated models using information criteria, where the model with the lowest information criteria values is chosen as the best model.

[Balakrishnan, Gupta, Kundu, Leiva, and Sanhueza \(2011\)](#) introduced a two-component mixture model based on BS distributions for modeling heterogeneous populations when the subpopulations have skewed distributions, since the BS distribution ([Birnbaum and Saunders 1969](#)) is positively skewed, and studied the characteristics and estimation procedure of the proposed model. [Benites, Maehara, Vilca, and Marmolejo-Ramos \(2017\)](#) extended the work proposed by [Balakrishnan et al. \(2011\)](#) via considering a g -component mixture of BS distributions for modeling multimodel populations to provide flexibility in different data analysis situations. [Benites et al. \(2017\)](#) proved the identifiability for a two-component mixture of BS distributions and suggested in the EM algorithm the k -bumps initialization algorithm to estimate the parameters of the BS distribution mixture model. Also, they implemented bootstrap procedures via real data to test the hypotheses about the number of components g in a mixture of BS distributions. [El-Sharkawy and Ismail \(2020\)](#) proved the identifiability for a g -component mixture of BS distributions, and they discussed parameter estimation and testing homogeneity in the mixture of BS distributions using EM algorithm and EM test, respectively, based on random censoring data. [El-Sharkawy and Ismail \(2021\)](#) proposed the modified likelihood ratio test and the shortcut method of the bootstrap test for testing the number of components g in a mixture of BS distributions under a random right censoring scheme. Nevertheless, to our knowledge there are no studies that use information criteria to select the number of components g in a mixture of BS distributions.

The purpose of the paper is to compare the performance of all the information criteria introduced in the next section for selecting the number of components g in a mixture of BS distributions. We investigate the performance of these criteria through a Monte Carlo simulation study under different conditions for degrees of separation, sample sizes, mixing proportions, and true model complexity. The objective is to evaluate the effect of these factors on the proper selection of the number of components and construct a comparison between criteria based on log-likelihood and classification functions. Also, we evaluate the performance of these criteria based on unpenalized and penalized estimates of the model's parameters.

The paper is organized as follows. Section 2 reviews several information criteria to select the number of components in a finite mixture model considered in our study. Section 3 describes the finite mixture of the BS distributions and the procedure of model fitting using the EM algorithm. Section 4 presents the simulation study design and the results. Section 5 provides an illustrative example of real life. Finally, Section 6 presents some concluding remarks.

2. Information criteria

In this section, we introduce several information criteria for order selection in finite mixture models. The information criteria can be classified into likelihood-based statistics and classification likelihood-based statistics, also referred to as complete data likelihood-based statistics. These criteria have been developed to balance between accuracy and complexity of the model. The general form can be expressed as

$$-2\ell + P, \tag{1}$$

where ℓ is the log likelihood of a fitted model and P is the penalty term that may depend on the number of free model parameters k , sample size n , or both. The first term is interpreted in (1) as a measure of the accuracy of the model, while the second term is interpreted as a measure of the complexity of the model. Selecting the optimal number of components is that corresponding to the minimum value of the criterion in (1). Next to the information criteria, depending on the complete data likelihood function, classification likelihood-based statistics have been developed for finite mixture models. These statistics are used to measure the quality of the classification given by a model.

2.1. Likelihood-based information criteria

The most widely known and used model selection criterion is Akaike information criterion (AIC), proposed by Akaike (1973, 1974). It's derived as an estimate of expected relative Kullback-Leibler (K-L) divergence (Kullback and Leibler 1951) between the true model and the fitted model, based on the maximized empirical log-likelihood function of the expected K-L divergence. The AIC is defined as

$$AIC = -2\ell + 2k, \quad (2)$$

with penalty term $P = 2k$.

Another more popular criterion that is used for model selection is the Bayesian information criterion (BIC) proposed by Schwarz (1978). This criterion has a Bayesian derivation to select a model from a set of candidate models with the largest posterior probability, but it can also be used for model selection in a non-Bayesian context. It is defined as

$$BIC = -2\ell + k \log n, \quad (3)$$

with penalty term $P = k \log n$, which is a complexity penalization as opposed to $2k$ in the AIC. This criterion has the property of consistency which as n grows large, selects the true model with probability one (for more details see, Andrews and Currim (2003), Andrew and Joseph (2012), and Vrieze (2012)).

Various extensions of AIC have been proposed to overcome its lack of performance in model selection applications. A small sample correction of AIC was proposed by Hurvich and Tsai (1989) to overcome that the AIC tends to overfit models in small samples and that led to a criterion called corrected AIC (AICc). AICc is defined as

$$AICc = -2\ell + 2k \left(\frac{n}{n-k-1} \right), \quad (4)$$

with penalty term $P = 2k \left(\frac{n}{n-k-1} \right)$ (also see Sugiura (1978)). Burnham and Anderson (2002) recommended use of AICc if the ratio n/k is small (say < 40). Both AIC and AICc will be nearly identical and have a strong tendency to choose the same model if the ratio n/k is sufficiently large.

Bozdogan (1987) extended the AIC to make it consistent, namely consistent Akaike's information criterion (CAIC) is defined as

$$CAIC = -2\ell + k(\log n + 1), \quad (5)$$

with penalty term $P = k((\log n) + 1)$.

In the context of the mixture-model cluster analysis, Bozdogan (1994) modified the AIC to use it in estimating the number of mixture clusters where the marginal cost per parameter in the AIC definition, the so-called magic number 2 is not adequate for the mixture-model and he increased it to 3 and called it AIC3. AIC3 is defined as

$$AIC3 = -2\ell + 3k, \quad (6)$$

with penalty term $P = 3k$.

Although the derivation of these criteria depends on regularity conditions that do not hold in the framework for mixture models, they are still used to select the order of a finite mixture model (see e.g. Bozdogan, Sclove, and Gupta (1994), Konishi and Kitagawa (2007), McLachlan and Peel (2000), McLachlan and Rathnayake (2014), M and van der Laan M J (2003), Fernández and Arnold (2016) for a theoretical and practical support).

2.2. Classification likelihood-based information criteria

The classification likelihood-based information criteria which are based on the complete data log-likelihood were developed for finite mixture models. Such criteria take into account classification and aim to pick models which can classify the observations in a consistent fashion. A simple relation between the log-likelihood and the complete data log-likelihood which was first noted by Hathaway (1986) can be written as

$$\ell = \ell_c + EN(g), \quad (7)$$

where $EN(g) = -\sum_{j=1}^g \sum_{i=1}^n \hat{\tau}_{ij} \log \hat{\tau}_{ij} \geq 0$ is an entropy function composed of elements $\hat{\tau}_{ij}$ which represent the estimated posterior probabilities that the subject i arises from the j -th mixture component. This entropy can be interpreted as a penalization term which measures the overlap of the mixture components and can be viewed as a measure of the model's ability to provide a relevant data partition. The entropy will be close to zero if the mixture components are well separated, but will take large values if the mixture components are poorly separated (see, Celeux and Soromenho (1996) and McLachlan and Peel (2000)).

Banfield and Raftery (1993) derived an approximate Bayesian criterion based on the complete data log-likelihood called the approximate weight of evidence (AWE) criterion having the form

$$AWE = -2\ell_c + 2k \left(\frac{3}{2} + \log n \right). \quad (8)$$

When the mixture components are well separated that leads to $\ell_c \simeq \ell$, therefore the AWE approximates to the BIC. The penalty term in AWE criterion is complex, which makes it select more parsimonious models than BIC (Fernández and Arnold 2016). The weakness of this criterion is that when the mixture components are not well separated, parameter estimation is biased (see, McLachlan and Peel (2000) for more details).

Biernacki and Govaert (1997) suggested using the relationship that exists between the log-likelihood and the complete data log-likelihood for order selection. This criterion is known as the classification likelihood information criterion (CLC) and is derived directly from equation (7) as

$$CLC = -2\ell + 2EN(g). \quad (9)$$

The performance of CLC works well under equal mixing proportions but it appeared to overestimate the appropriate number of components when no constraints were imposed on the mixing proportions. The explanation for this behavior is that the number of parameters in the mixture model is not penalized by the CLC (see, Biernacki, Celeux, and Govaert (1998)).

Based on the estimated entropy function, Celeux and Soromenho (1996) introduced the normalized entropy criterion (NEC) as follows

$$NEC(g) = \frac{EN(g)}{\ell - \ell^{(1)}}, \quad g > 1, \quad (10)$$

where $\ell^{(1)}$ is the maximum log-likelihood in a one component mixture. The decision of NEC for choosing the number of components in the mixture model between one component and greater than one is not valid, because it is undefined at $g=1$. The improved version of NEC to deal with this problem was proposed by Biernacki, Celeux, and Govaert (1999) for the Gaussian mixture.

To overcome the drawbacks of the CLC and BIC, [Biernacki *et al.* \(1998\)](#) proposed the integrated classification likelihood (ICL) criterion and an approximated version of it using BIC. ICL-BIC was used by [McLachlan and Peel \(2000\)](#) to refer to an approximate form of ICL criterion which is given by

$$ICL - BIC = BIC + 2EN(g). \quad (11)$$

See, for example, [McLachlan and Peel \(2000\)](#), [Oliveira-Brochado and Martins \(2005\)](#), and [Depraetere and Vandebroek \(2014\)](#) for a detailed review of all order selection criteria in the framework of the mixture model.

3. Finite mixture of BS distributions and parameter estimation

3.1. Model formulation

A finite mixture of BS distributions is a weighted sum of g component BS densities as given by

$$f(t; \boldsymbol{\theta}) = \sum_{j=1}^g \pi_j f(t; \alpha_j, \beta_j), \quad t > 0, \quad (12)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_g)$, $\theta_j = (\pi_j, \alpha_j, \beta_j)$, with $\alpha_j > 0$, $\beta_j > 0$ and $\pi_j \geq 0$, $\sum_{j=1}^g \pi_j = 1$ are the shape, scale, and mixing parameters, respectively, and $f(t; \alpha_j, \beta_j)$ is the probability density function (pdf) of the j th component

$$f(t; \alpha_j, \beta_j) = \frac{1}{2\alpha_j\beta_j\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\alpha_j^2} \left(\frac{t}{\beta_j} + \frac{\beta_j}{t} - 2 \right) \right\} \left[\left(\frac{\beta_j}{t} \right)^{1/2} + \left(\frac{\beta_j}{t} \right)^{3/2} \right], \quad j = 1, \dots, g. \quad (13)$$

The BS distribution is a two-parameter, unimodal, and positively skewed distribution used to model fatigue failure of materials subject to cyclic stress. It is also famous as the fatigue life distribution, since it originated from the fatigue of materials but has been expanded to solve problems in areas such as biology, business, economics, reliability, environmental and medical sciences, and many others. For several implementations, the BS distribution was considered as a rival of the log-normal, Weibull, gamma, and inverse Gaussian. A comprehensive review and additional details on the BS distribution, see [Balakrishnan \(2019\)](#) and [Leiva \(2016\)](#).

3.2. Parameter estimation via the EM algorithm

Let $\mathbf{T} = (T_1, T_2, \dots, T_n)$ denote the observed data of size n from a mixture of BS distributions as given in (12). The log-likelihood for the observed-data is given by

$$\ell(\boldsymbol{\theta}|\mathbf{t}) = \sum_{i=1}^n \log \left\{ \sum_{j=1}^g \pi_j f(t_i; \alpha_j, \beta_j) \right\}. \quad (14)$$

To use the EM algorithm for calculating the MLE of the parameters involved in (14), we need to formulate the mixture of BS distributions in the EM framework by introducing allocation indicators $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$, where $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \dots, Z_{ig})$, $i = 1, \dots, n$ and

$$Z_{ij} = \begin{cases} 1 & \text{if observation } i \text{ belongs to component } j, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, log-likelihood for the complete-data (\mathbf{T}, \mathbf{Z}) is given by

$$\ell^c(\boldsymbol{\theta}|\mathbf{t}, \mathbf{z}) = \sum_{i=1}^n \sum_{j=1}^g Z_{ij} \{ \log \pi_j f(t_i; \alpha_j, \beta_j) \}. \quad (15)$$

By maximizing $\ell^c(\boldsymbol{\theta}|\mathbf{t}, \mathbf{z})$, the resulting estimates may not guarantee the optimal properties of the likelihood approach (consistency property). Since the maximum likelihood estimates of mixing parameters may be on or near the boundary and as a result, some of the ML estimates of the other parameters become inconsistent, see [Chen, Chen, and Kalbfleisch \(2004\)](#) and [Wong and Li \(2014\)](#). [Chen, Chen, and Kalbfleisch \(2001\)](#) suggested a penalized penalty term on the log-likelihood function dependent on the mixing proportions to push these estimates away from the boundary. Therefore, the penalized log-likelihood for the complete-data (T, Z) is given by

$$p\ell^c(\boldsymbol{\theta}|\mathbf{t}, \mathbf{z}) = \ell^c(\boldsymbol{\theta}|\mathbf{t}, \mathbf{z}) + p(\boldsymbol{\pi}). \quad (16)$$

where $\ell^c(\boldsymbol{\theta}|\mathbf{t}, \mathbf{z})$ is the log-likelihood function defined in (15), $p(\boldsymbol{\pi}) = C \sum_{j=1}^g \log(2\pi_j)$ is the penalty function on mixing proportions such that $p(\boldsymbol{\pi})$ gets its maximum at $\pi_j = 1/g$ and $p(\boldsymbol{\pi}) \rightarrow -\infty$ when $\pi_j \rightarrow 0$ or 1 , $j = 1, \dots, g$, whereas C is a positive constant.

The EM algorithm for ML estimation of the mixture of BS distributions works as follows:

- 1- Start with randomly initial value $\boldsymbol{\theta}^0$ of the parameter $\boldsymbol{\theta}$.
- 2- E-step: At iteration $l \geq 0$, compute

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(l)}) &= E_{\boldsymbol{\theta}^{(l)}} [p\ell^c(\boldsymbol{\theta}|\mathbf{t}, \mathbf{z})|\mathbf{t}] \\ &= \sum_{i=1}^n \sum_{j=1}^g E_{\boldsymbol{\theta}^{(l)}} [Z_{ij}|t_i] \left\{ \log \pi_j + \left(A - \frac{1}{2\alpha_j^2} \left[\frac{t_i}{\beta_j} + \frac{\beta_j}{t_i} - 2 \right] \right. \right. \\ &\quad \left. \left. + \log \left[\left(\frac{\beta_j}{t_i} \right)^{1/2} + \left(\frac{\beta_j}{t_i} \right)^{3/2} \right] - \log \alpha_j \beta_j \right) \right\} + p(\boldsymbol{\pi}), \end{aligned}$$

where A is a constant independent of $\boldsymbol{\theta}$ and

$$E_{\boldsymbol{\theta}^{(l)}} [Z_{ij}|t_i] = \frac{\pi_j^{(l)} [f(t_i; \alpha_j^{(l)}, \beta_j^{(l)})]}{\sum_{m=1}^g \pi_m^{(l)} [f(t_i; \alpha_m^{(l)}, \beta_m^{(l)})]}.$$

- 3- M-step : Maximizing (16) with respect to $\boldsymbol{\theta}_j, j = 1, \dots, g$, gives the following equations:

$$\pi_j^{(l+1)} = \left[\sum_{i=1}^n z_{ij}^{(l)} + C \right] / (n + gC),$$

and

$$\frac{\partial Q}{\partial \alpha_j} = \sum_{i=1}^n z_{ij}^{(l)} \left\{ \alpha_j^{-3} \left[\frac{t_i}{\beta_j} + \frac{\beta_j}{t_i} - 2 \right] - \alpha_j^{-1} \right\}, \quad (17)$$

$$\frac{\partial Q}{\partial \beta_j} = \sum_{i=1}^n z_{ij}^{(l)} \left\{ \frac{t_i}{2\alpha_j^2 \beta_j^2} - \frac{1}{2\alpha_j^2 t_i} - \frac{1}{2\beta_j} + \frac{1}{t_i + \beta_j} \right\}, \quad (18)$$

where $z_{ij}^{(l)} = E_{\boldsymbol{\theta}^{(l)}} [Z_{ij}|t_i]$.

We obtain $\alpha_j^{(l+1)}$ and $\beta_j^{(l+1)}$ by setting (17) and (18) to zero and solving them for α_j and β_j , $j = 1, 2, \dots, g$.

- 4- Iteratively update the estimate $\boldsymbol{\theta}^{(l)}$ with the new estimate $\boldsymbol{\theta}^{(l+1)}$ by alternating between the E- and M- steps until $p\ell^c(\boldsymbol{\theta}^{(l+1)}|\mathbf{t}, \mathbf{z}) - p\ell^c(\boldsymbol{\theta}^{(l)}|\mathbf{t}, \mathbf{z}) < \epsilon$, for some $\epsilon > 0$.

To avoid getting stuck at local maxima while running the EM algorithm in our numerical illustrations, we provide a random initialization technique to improve the chances of achieving the global maximum. This technique can be summarized as follows:

- 1- Generate random initial positions. This is could by drawing random initial values for π_j , denoted by π_j^0 , $j = 1, \dots, g$ from a Dirichlet distribution with all parameters equal to 1. Then, based on the π_j^0 values, the ordered sample is separated into g sets and

for each set the modified moment estimates of α and β presented by Ng, Kundu, and Balakrishnan (2003) are calculated as the initial values for the parameters α_j and β_j , for $j = 1, \dots, g$.

- 2- Run the EM algorithm with few iterations when $C = 0$ and $C = 1$.
- 3- Repeat the Steps 1 and 2, r times, and select the solution with the highest likelihood among those r trials.

4. Simulation study

In this section, we explain the procedure and results of a Monte Carlo simulation study for evaluating and comparing the performance of the information criteria based on the log-likelihood function and classification function, to select the number of components in a mixture of BS distributions using unpenalized and penalized estimation.

4.1. Simulation procedure and design

The simulation procedure involves three stages: (i) Generate data sets from the mixture of BS distributions. (ii) Use the EM algorithm to estimate the parameters of the mixture of BS distributions for different component numbers and then apply each information criterion to choose the best number of components. (iii) Evaluation and comparison of the performances of these information criteria, through studying some design factors such as the degree of separation between the components, the sample size n , the mixing proportion π , and the true number of components that may affect the choice of the number of components in mixture models. Comparison between different information criteria is based on the results of the proportion of selecting the number of components correctly, which are provided by the simulation study. All the computations have been implemented by using R software.

In particular, we randomly generate data sets for various sample sizes namely $n = 25, 50, 100$, and 250 from the mixture of BS distributions when the true number of components $g = 2$ and 3. The penalty constant C is set to be 0 and 1. We consider different scenarios for models of mixtures of BS distributions by varying the mixing proportions and degrees of separation between the components. The Mahalanobis distance is used to measure the distance between the two components of the mixture model and is defined as $\Delta = |\mu_1 - \mu_2|/\sigma$, where σ is the common standard deviation, and μ_1 and μ_2 are the means for the first and second components in the mixture model, respectively. Table 1 displays the full sets of parameters for each scenario, where models M1, M2, M3, and M8, M9, M10 are used to study the effect of degrees of separation using Δ when the true number of components $g = 2$ and 3, respectively. While, M4, M5, and M6, M7 are used to study the effect of mixing proportions when the true number of components $g = 2$ but for $\Delta = 2$ and 4, respectively. The simulation procedure is repeated 5000 times.

4.2. Results

In each table, the quantities corresponding to the highest proportion of selected number of components are shown in boldface. Tables 2-4 display the results of each information criterion for the competing models $g^* = 1, 2, 3$, and 4 when the true number of components $g = 2$ for M1, M2, and M3, respectively. The overall findings in these tables show that the degree of separation and the sample size significantly affect the performance of information criteria, since the performance of all information criteria increases by increasing both the values of the degree of separation and the sample size or either of them, holding all other factors fixed. It is also noted from these tables that all the information criteria perform well as they are able to select the correct model (a mixture of BS distributions with $g^* = 2$) with varying percentages, regardless of the values of Δ and n . Furthermore, we observe that the classification based

Table 1: Parameter settings for mixture of BS distributions when $g = 2$ and 3

Model	parameter	Δ
M1	$0.5BS(0.25,0.5)+0.5BS(0.5,2)$	2
M2	$0.5BS(0.25,1)+0.5BS(0.25,2.5)$	3
M3	$0.5BS(0.2,1.5)+0.5BS(0.2,4)$	4
M4	$0.3BS(0.25,1)+0.7BS(0.25,2.5)$	2
M5	$0.1BS(0.25,1)+0.9BS(0.25,2.5)$	2
M6	$0.3BS(0.2,1.5)+0.7BS(0.2,4)$	4
M7	$0.1BS(0.2,1.5)+0.9BS(0.2,4)$	4
M8	$1/3BS(0.25,0.5)+1/3BS(0.5,2)+1/3BS(0.3,5)$	2
M9	$1/3BS(0.25,1)+1/3BS(0.25,2.5)+1/3BS(0.2,5.2)$	3
M10	$1/3BS(0.2,1.5)+1/3BS(0.2,4)+1/3BS(0.1,7.2)$	4

criteria perform better than the log-likelihood based criteria when the values of Δ and n are small. However, the performances of the classification based criteria and log-likelihood based criteria converge as the values of Δ and n increase. A possible explanation for this result is that the percentages of selecting the mixture of BS distribution with $g^* = 1$ in the classification based criteria have been omitted since these criteria can not be numerically defined for the mixture model with $g = 1$. The effect of the unpenalized and penalized estimation on the performance of information criteria is also observed from these tables, as it is clear that all the information criteria have higher percentages of correct fitting when $C = 1$ than when $C = 0$, and these percentages converge when the values of Δ and n increase. Comparing the results of all criteria, it is observed that AWE and ICL-BIC perform the best in tables 2-4, even when both values of Δ and n are small, with percentages of correct fitting $\geq 85\%$, while the other criteria perform quite satisfactory.

The effect of the mixing proportions on the performance of information criteria can be observed for $\pi = 0.5, 0.3$, and 0.1 in Tables 2, 5, and 6 when $\Delta = 2$ and in Tables 4, 7, and 8 when $\Delta = 4$, respectively. It is observed that the performance of the criteria decreases when the mixing proportions within a model are less similar. It is noted from Tables 2 and 4-8, that the classification based criteria select the correct model always with varying percentages, where percentages decrease as π decreases, increase as the values of Δ and n increase, and are higher for $C = 1$ than for $C = 0$, holding all other factors fixed in each case. On the other hand, the log-likelihood based criteria don't perform well for lower mixing proportions π and small Δ . Looking across the Tables 5-8 to compare the criteria, we note that AWE and ICL-BIC show the best performance even when both values of Δ and n are small with the percentages of correct fitting $> 70\%$. While the CLC and NEC show much less performance compared to AWE and ICL-BIC when the mixing proportion π decreases especially when the values of Δ and n are small. The likelihood-based information criteria reveal worse performance at small sample sizes and the performance becomes more satisfactory as the values of Δ and n increase.

Tables 9-11 show the results of each information criterion for the competing models $g^* = 1, 2, 3, 4$, and 5 when the true number of components $g = 3$ for M8, M9, and M10, respectively. In Table 9, poor performance is observed from all information criteria as the correct model (a mixture of BS distributions with $g^* = 3$) is not selected even when $n = 250$. When both the values of Δ and n increase, all information criteria show improvement for selecting the correct model. It is observed from Table 10, that for $\Delta = 3$ and $n = 50$, only CLC and NEC select the correct model with percentages reaching to 43% . However, all the information criteria select the correct model with percentages reaching to 63% and 89% when $n = 100$ and 250 , respectively. From Table 11, where $\Delta = 4$, only the classification criteria select the correct model with percentages reaching to 46% when $n = 25$, whereas when $n = 50$ all the criteria select the correct model with percentages reaching to 59% except CAIC and BIC and the

Table 3: Results of the Monte Carlo study for the true number of components $g=2$, according to M2

n	g^*	Method	AIC	AIC3	CAIC	AICc	BIC	CLC	AWE	NEC	ICL-BIC
25	1	$C=0$	0.17	0.40	0.47	0.39	0.43	—	—	—	—
		$C=1$	0.23	0.46	0.47	0.43	0.46	—	—	—	—
	2	$C=0$	0.56	0.47	0.50	0.52	0.46	0.61	0.93	0.75	0.94
		$C=1$	0.72	0.51	0.51	0.54	0.51	0.67	0.98	0.80	0.97
	3	$C=0$	0.12	0.06	0.02	0.04	0.05	0.16	0.02	0.12	0.05
		$C=1$	0.04	0.02	0.01	0.02	0.02	0.14	0.01	0.10	0.03
	4	$C=0$	0.15	0.07	0.01	0.05	0.06	0.23	0.05	0.13	0.01
		$C=1$	0.01	0.01	0.01	0.01	0.01	0.19	0.01	0.10	0.00
50	1	$C=0$	0.04	0.15	0.23	0.08	0.29	—	—	—	—
		$C=1$	0.05	0.15	0.23	0.08	0.29	—	—	—	—
	2	$C=0$	0.78	0.79	0.67	0.84	0.67	0.87	0.98	0.88	0.97
		$C=1$	0.91	0.83	0.72	0.89	0.69	0.90	0.99	0.93	0.98
	3	$C=0$	0.09	0.04	0.06	0.05	0.03	0.06	0.01	0.06	0.02
		$C=1$	0.03	0.01	0.04	0.02	0.01	0.05	0.01	0.04	0.01
	4	$C=0$	0.09	0.02	0.04	0.03	0.01	0.07	0.01	0.06	0.01
		$C=1$	0.01	0.01	0.01	0.01	0.01	0.05	0.00	0.03	0.01
100	1	$C=0$	0.00	0.01	0.00	0.00	0.04	—	—	—	—
		$C=1$	0.00	0.01	0.00	0.00	0.01	—	—	—	—
	2	$C=0$	0.86	0.95	0.95	0.91	0.94	0.97	0.99	0.97	0.99
		$C=1$	0.98	0.98	0.98	0.98	0.98	0.99	1.00	0.98	0.99
	3	$C=0$	0.08	0.03	0.05	0.06	0.01	0.02	0.01	0.02	0.01
		$C=1$	0.02	0.01	0.02	0.02	0.01	0.01	0.00	0.02	0.01
	4	$C=0$	0.06	0.01	0.00	0.03	0.01	0.01	0.00	0.01	0.00
		$C=1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
250	1	$C=0$	0.00	0.00	0.00	0.00	0.00	—	—	—	—
		$C=1$	0.00	0.00	0.00	0.00	0.00	—	—	—	—
	2	$C=0$	0.94	0.99	1.00	0.94	1.00	1.00	1.00	0.99	1.00
		$C=1$	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
	3	$C=0$	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
		$C=1$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	$C=0$	0.03	0.01	0.00	0.03	0.00	0.00	0.00	0.01	0.00
		$C=1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00

Table 5: Results of the Monte Carlo study for the true number of components $g=2$, according to M4

n	Method	AIC	AIC3	CAIC	AICc	BIC	CLC	AWE	NEC	ICL-BIC	
25	1	$C=0$	0.24	0.50	0.71	0.50	0.54	—	—	—	—
		$C=1$	0.34	0.56	0.73	0.54	0.60	—	—	—	—
	2	$C=0$	0.38	0.29	0.15	0.34	0.27	0.43	0.83	0.52	0.77
		$C=1$	0.49	0.32	0.16	0.35	0.28	0.48	0.90	0.59	0.84
	3	$C=0$	0.21	0.12	0.08	0.10	0.11	0.30	0.11	0.29	0.18
		$C=1$	0.13	0.09	0.08	0.08	0.08	0.28	0.07	0.25	0.14
	4	$C=0$	0.17	0.09	0.06	0.06	0.08	0.27	0.06	0.19	0.05
		$C=1$	0.04	0.03	0.03	0.03	0.04	0.24	0.03	0.16	0.02
50	1	$C=0$	0.10	0.27	0.61	0.16	0.45	—	—	—	—
		$C=1$	0.13	0.29	0.62	0.19	0.46	—	—	—	—
	2	$C=0$	0.55	0.54	0.31	0.61	0.44	0.64	0.93	0.62	0.83
		$C=1$	0.72	0.61	0.31	0.70	0.46	0.72	0.94	0.71	0.89
	3	$C=0$	0.23	0.15	0.07	0.19	0.10	0.21	0.06	0.21	0.13
		$C=1$	0.13	0.09	0.06	0.10	0.07	0.17	0.05	0.17	0.10
	4	$C=0$	0.12	0.04	0.01	0.04	0.01	0.15	0.01	0.17	0.04
		$C=1$	0.02	0.01	0.01	0.01	0.01	0.11	0.01	0.12	0.01
100	1	$C=0$	0.01	0.04	0.27	0.01	0.16	—	—	—	—
		$C=1$	0.01	0.04	0.27	0.02	0.16	—	—	—	—
	2	$C=0$	0.68	0.79	0.66	0.74	0.75	0.84	0.96	0.76	0.92
		$C=1$	0.85	0.84	0.66	0.85	0.75	0.88	0.97	0.81	0.95
	3	$C=0$	0.20	0.14	0.07	0.19	0.08	0.09	0.04	0.13	0.05
		$C=1$	0.13	0.11	0.07	0.12	0.09	0.08	0.03	0.11	0.04
	4	$C=0$	0.11	0.03	0.00	0.06	0.01	0.07	0.00	0.11	0.03
		$C=1$	0.01	0.01	0.00	0.01	0.00	0.04	0.00	0.08	0.01
250	1	$C=0$	0.00	0.00	0.01	0.00	0.00	—	—	—	—
		$C=1$	0.00	0.00	0.01	0.00	0.01	—	—	—	—
	2	$C=0$	0.77	0.86	0.91	0.78	0.91	0.93	0.95	0.85	0.94
		$C=1$	0.91	0.92	0.93	0.91	0.93	0.95	0.99	0.84	0.97
	3	$C=0$	0.15	0.11	0.07	0.14	0.08	0.02	0.02	0.08	0.02
		$C=1$	0.08	0.07	0.06	0.08	0.06	0.01	0.01	0.10	0.01
	4	$C=0$	0.08	0.03	0.01	0.08	0.01	0.05	0.03	0.07	0.04
		$C=1$	0.01	0.01	0.00	0.01	0.00	0.04	0.00	0.06	0.02

Table 6: Results of the Monte Carlo study for the true number of components $g=2$, according to M5

n	Method	AIC	AIC3	CAIC	AICc	BIC	CLC	AWE	NEC	ICL-BIC	
25	1	$C=0$	0.42	0.65	0.78	0.66	0.68	—	—	—	—
		$C=1$	0.63	0.79	0.87	0.78	0.81	—	—	—	—
	2	$C=0$	0.27	0.17	0.08	0.18	0.15	0.39	0.75	0.37	0.71
		$C=1$	0.25	0.12	0.04	0.13	0.10	0.36	0.83	0.31	0.80
	3	$C=0$	0.19	0.12	0.09	0.11	0.12	0.27	0.20	0.28	0.20
		$C=1$	0.08	0.06	0.06	0.06	0.06	0.29	0.15	0.29	0.15
	4	$C=0$	0.12	0.06	0.05	0.05	0.05	0.34	0.05	0.35	0.09
		$C=1$	0.04	0.03	0.03	0.03	0.03	0.35	0.02	0.40	0.05
50	1	$C=0$	0.41	0.57	0.83	0.43	0.73	—	—	—	—
		$C=1$	0.49	0.70	0.88	0.58	0.81	—	—	—	—
	2	$C=0$	0.31	0.30	0.10	0.41	0.18	0.53	0.89	0.40	0.73
		$C=1$	0.40	0.22	0.05	0.33	0.12	0.48	0.90	0.32	0.78
	3	$C=0$	0.18	0.10	0.05	0.12	0.06	0.27	0.09	0.29	0.20
		$C=1$	0.08	0.05	0.05	0.07	0.05	0.27	0.08	0.30	0.17
	4	$C=0$	0.10	0.03	0.02	0.04	0.03	0.20	0.02	0.31	0.07
		$C=1$	0.03	0.03	0.02	0.02	0.02	0.25	0.02	0.38	0.05
100	1	$C=0$	0.15	0.32	0.74	0.18	0.60	—	—	—	—
		$C=1$	0.24	0.46	0.80	0.27	0.71	—	—	—	—
	2	$C=0$	0.55	0.54	0.20	0.57	0.33	0.60	0.85	0.45	0.73
		$C=1$	0.64	0.46	0.14	0.62	0.23	0.54	0.85	0.32	0.72
	3	$C=0$	0.20	0.10	0.04	0.17	0.05	0.25	0.14	0.25	0.21
		$C=1$	0.10	0.06	0.05	0.09	0.05	0.29	0.14	0.33	0.22
	4	$C=0$	0.10	0.04	0.02	0.08	0.02	0.15	0.01	0.30	0.06
		$C=1$	0.02	0.02	0.01	0.02	0.01	0.17	0.01	0.35	0.06
250	1	$C=0$	0.00	0.03	0.31	0.01	0.25	—	—	—	—
		$C=1$	0.01	0.04	0.37	0.02	0.20	—	—	—	—
	2	$C=0$	0.56	0.72	0.67	0.59	0.74	0.62	0.82	0.47	0.73
		$C=1$	0.80	0.86	0.63	0.81	0.76	0.67	0.85	0.37	0.75
	3	$C=0$	0.31	0.22	0.02	0.30	0.01	0.29	0.18	0.26	0.23
		$C=1$	0.18	0.09	0.00	0.17	0.04	0.21	0.13	0.38	0.18
	4	$C=0$	0.13	0.03	0.00	0.10	0.00	0.09	0.00	0.27	0.04
		$C=1$	0.01	0.01	0.00	0.00	0.00	0.12	0.02	0.25	0.07

Table 9: Results of the Monte Carlo study for the true number of components $g=3$, according to M8

n	Method	AIC	AIC3	CAIC	AICc	BIC	CLC	AWE	NEC	ICL-BIC	
25	1	$C=0$	0.25	0.48	0.64	0.56	0.52	—	—	—	—
		$C=1$	0.61	0.76	0.81	0.76	0.78	—	—	—	—
	2	$C=0$	0.16	0.14	0.09	0.19	0.13	0.36	0.78	0.36	0.66
		$C=1$	0.15	0.10	0.06	0.10	0.08	0.38	0.87	0.32	0.81
	3	$C=0$	0.19	0.17	0.14	0.15	0.16	0.34	0.13	0.36	0.13
		$C=1$	0.08	0.07	0.07	0.07	0.07	0.27	0.07	0.21	0.07
	4	$C=0$	0.21	0.12	0.09	0.08	0.11	0.19	0.07	0.27	0.15
		$C=1$	0.12	0.05	0.05	0.06	0.05	0.20	0.05	0.25	0.11
	5	$C=0$	0.19	0.09	0.04	0.02	0.08	0.11	0.02	0.16	0.06
		$C=1$	0.04	0.02	0.01	0.01	0.02	0.15	0.01	0.12	0.01
50	1	$C=0$	0.10	0.33	0.65	0.23	0.52	—	—	—	—
		$C=1$	0.28	0.45	0.68	0.35	0.57	—	—	—	—
	2	$C=0$	0.30	0.26	0.16	0.31	0.23	0.42	0.88	0.57	0.76
		$C=1$	0.41	0.36	0.18	0.43	0.26	0.51	0.86	0.61	0.80
	3	$C=0$	0.24	0.21	0.12	0.25	0.16	0.27	0.10	0.30	0.20
		$C=1$	0.19	0.12	0.09	0.15	0.10	0.27	0.12	0.29	0.18
	4	$C=0$	0.16	0.14	0.05	0.16	0.07	0.18	0.02	0.09	0.04
		$C=1$	0.10	0.06	0.04	0.06	0.05	0.15	0.02	0.08	0.02
	5	$C=0$	0.20	0.06	0.02	0.05	0.02	0.13	0.00	0.04	0.00
		$C=1$	0.02	0.01	0.01	0.01	0.01	0.07	0.00	0.02	0.00
100	1	$C=0$	0.00	0.00	0.02	0.00	0.00	—	—	—	—
		$C=1$	0.00	0.00	0.02	0.00	0.01	—	—	—	—
	2	$C=0$	0.49	0.53	0.80	0.50	0.76	0.68	0.95	0.73	0.84
		$C=1$	0.57	0.72	0.84	0.63	0.81	0.70	0.95	0.74	0.85
	3	$C=0$	0.27	0.39	0.15	0.36	0.21	0.22	0.05	0.20	0.14
		$C=1$	0.38	0.26	0.13	0.33	0.16	0.21	0.05	0.22	0.14
	4	$C=0$	0.11	0.04	0.01	0.07	0.01	0.09	0.00	0.06	0.01
		$C=1$	0.02	0.01	0.00	0.02	0.00	0.08	0.00	0.04	0.01
	5	$C=0$	0.13	0.04	0.02	0.07	0.02	0.01	0.00	0.01	0.01
		$C=1$	0.03	0.01	0.01	0.02	0.02	0.01	0.00	0.00	0.00
250	1	$C=0$	0.00	0.00	0.00	0.00	0.00	—	—	—	—
		$C=1$	0.00	0.00	0.00	0.00	0.00	—	—	—	—
	2	$C=0$	0.70	0.72	0.56	0.64	0.56	0.88	0.94	0.90	0.92
		$C=1$	0.72	0.75	0.70	0.73	0.57	0.91	0.95	0.91	0.95
	3	$C=0$	0.13	0.23	0.44	0.21	0.40	0.10	0.06	0.08	0.08
		$C=1$	0.26	0.24	0.29	0.25	0.42	0.08	0.05	0.07	0.05
	4	$C=0$	0.09	0.03	0.00	0.09	0.04	0.02	0.00	0.02	0.00
		$C=1$	0.02	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.00
	5	$C=0$	0.08	0.02	0.00	0.06	0.00	0.00	0.00	0.00	0.00
		$C=1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00

Table 10: Results of the Monte Carlo study for the true number of components $g=3$, according to M9

n	Method	AIC	AIC3	CAIC	AICc	BIC	CLC	AWE	NEC	ICL-BIC	
25	1	$C=0$	0.04	0.16	0.19	0.18	0.19	—	—	—	—
		$C=1$	0.08	0.20	0.17	0.19	0.25	—	—	—	—
	2	$C=0$	0.26	0.30	0.34	0.43	0.31	0.43	0.68	0.43	0.62
		$C=1$	0.55	0.48	0.39	0.53	0.45	0.50	0.75	0.51	0.75
	3	$C=0$	0.26	0.24	0.27	0.23	0.22	0.27	0.19	0.27	0.25
		$C=1$	0.22	0.19	0.33	0.18	0.18	0.29	0.17	0.25	0.20
	4	$C=0$	0.20	0.16	0.13	0.12	0.15	0.27	0.10	0.22	0.09
		$C=1$	0.09	0.09	0.07	0.07	0.08	0.14	0.06	0.17	0.04
	5	$C=0$	0.24	0.14	0.07	0.04	0.13	0.03	0.03	0.08	0.04
		$C=1$	0.06	0.04	0.04	0.03	0.04	0.07	0.02	0.07	0.01
50	1	$C=0$	0.00	0.04	0.21	0.01	0.10	—	—	—	—
		$C=1$	0.00	0.03	0.18	0.02	0.08	—	—	—	—
	2	$C=0$	0.38	0.57	0.58	0.58	0.63	0.20	0.85	0.22	0.67
		$C=1$	0.68	0.71	0.64	0.72	0.71	0.25	0.86	0.29	0.75
	3	$C=0$	0.28	0.23	0.14	0.26	0.18	0.36	0.12	0.43	0.23
		$C=1$	0.25	0.19	0.14	0.21	0.16	0.37	0.11	0.41	0.19
	4	$C=0$	0.13	0.06	0.04	0.07	0.04	0.32	0.03	0.27	0.10
		$C=1$	0.04	0.04	0.02	0.03	0.03	0.28	0.03	0.24	0.06
	5	$C=0$	0.21	0.10	0.03	0.08	0.05	0.12	0.00	0.08	0.00
		$C=1$	0.03	0.03	0.02	0.02	0.02	0.10	0.00	0.06	0.00
100	1	$C=0$	0.00	0.05	0.03	0.01	0.03	—	—	—	—
		$C=1$	0.03	0.12	0.04	0.05	0.04	—	—	—	—
	2	$C=0$	0.23	0.23	0.45	0.23	0.27	0.34	0.38	0.07	0.34
		$C=1$	0.34	0.24	0.37	0.34	0.38	0.34	0.35	0.30	0.36
	3	$C=0$	0.41	0.52	0.46	0.49	0.52	0.44	0.62	0.60	0.61
		$C=1$	0.51	0.57	0.57	0.50	0.54	0.50	0.65	0.64	0.64
	4	$C=0$	0.19	0.13	0.05	0.17	0.17	0.20	0.00	0.28	0.05
		$C=1$	0.10	0.06	0.01	0.08	0.04	0.15	0.00	0.05	0.04
	5	$C=0$	0.17	0.07	0.01	0.10	0.01	0.02	0.00	0.05	0.00
		$C=1$	0.02	0.01	0.01	0.03	0.00	0.01	0.00	0.01	0.00
250	1	$C=0$	0.00	0.00	0.01	0.00	0.01	—	—	—	—
		$C=1$	0.00	0.00	0.03	0.00	0.02	—	—	—	—
	2	$C=0$	0.20	0.30	0.20	0.20	0.20	0.11	0.19	0.09	0.22
		$C=1$	0.10	0.20	0.20	0.10	0.20	0.08	0.11	0.10	0.16
	3	$C=0$	0.60	0.60	0.73	0.63	0.73	0.79	0.81	0.76	0.75
		$C=1$	0.79	0.74	0.74	0.79	0.74	0.85	0.89	0.81	0.83
	4	$C=0$	0.13	0.07	0.05	0.12	0.05	0.10	0.00	0.14	0.02
		$C=1$	0.09	0.04	0.03	0.09	0.04	0.07	0.00	0.06	0.01
	5	$C=0$	0.07	0.03	0.01	0.05	0.01	0.00	0.00	0.01	0.00
		$C=1$	0.02	0.02	0.00	0.02	0.00	0.00	0.00	0.03	0.00

5. Application

We consider a data set consisting of 245 observations of enzymatic activity in the blood, which involves the metabolism of carcinogenic substances. The data have been first analyzed by Bechtel, Bonaiti-Pellie, Poisson, Magnette, and Bechtel (1993), which observed that these enzyme data are suited for a mixture of two skewed distributions. Also, it was recently studied by Balakrishnan *et al.* (2011) and Benites *et al.* (2017), who considered a two-component mixture of BS distributions for modeling these data. Here we consider the mixture of BS distributions with different number of components for modeling the enzyme data and use the proposed information criteria to select the appropriate number of components in the mixture of BS distributions that fit these data correctly. The resulting fits from the information criteria for the enzyme data based on unpenalized and penalized estimates are given in Table 12. According to the results, all the proposed information criteria can successfully estimate the correct number of components of the enzyme data set, under both the unpenalized and penalized estimates of the model parameters.

Table 12: Comparison results of the log-likelihood based and the classification based criteria in the selection of the number of components for the enzyme data

	$C = 0$				$C = 1$			
	g^*				g^*			
	1	2	3	4	1	2	3	4
-2ℓ	211.0142	108.4054	103.28848	100.73532	213.7868	111.30792	111.0313	117.18782
AIC	215.0141	118.4054	119.2885	122.7353	217.7867	121.3079	127.0313	139.1878
AIC3	217.0141	123.4054	127.2885	133.7353	219.7867	126.3079	135.0313	150.1878
AICc	215.0637	118.6565	119.8986	123.8684	217.8363	121.559	127.6415	140.3209
CAIC	224.0167	141.0671	155.2985	172.2492	226.7893	143.8142	163.0414	188.7017
BIC	222.0167	136.0671	147.2985	161.2492	224.7893	138.8142	155.0414	177.7017
CLC	—	114.7880	170.1996	133.1538	—	117.6984	177.9424	149.6063
AWE	—	184.0037	282.2197	287.1814	—	187.7110	289.9626	303.6339
NEC	—	0.060781	0.621125	0.29397	—	0.060863	0.651169	0.335598
ICL-BIC	—	142.2943	214.2097	193.6676	—	145.2047	221.9525	210.1201

6. Conclusion

In this paper, a Monte Carlo simulation study is conducted to evaluate and compare the performance of the log-likelihood based information criteria and classification based information criteria for selecting the number of components in a mixture of BS distributions, considering two different types of estimation, unpenalized and penalized. Based on the simulation analysis, all the information criteria show both good and poor performance in detecting the correct number of components. Therefore, in general, we can conclude that no particular information criterion is always the best. According to the comparison results, it is observed that the classification based information criteria have the best overall performance in selecting the number of components, especially both AWE and ICL.BIC show better performance than any other criterion. It is also found that the information criteria based on penalized estimates have better performance compared to those based on unpenalized estimates. We also investigate the influence of degrees of separation, sample sizes, mixing proportions, and true model complexity on the performance of information criteria. The investigation show that the degree of separation is the most influential factor on the performance of information criteria and the sample size has less influence as compared to the degree of separation. While the influence of mixture proportions on the performance of information criteria is small as compared to the degree of separation and sample size. Also, it is observed that the performance of all information criteria gets worse with increasing the true model complexity. Finally, an example of enzyme data set is used to illustrate the performance of the proposed information criteria for selecting the appropriate number of components in the mixture of BS distributions that fit these data correctly.

Acknowledgements

The authors would like to thank the anonymous referees for their constructive comments and suggestions which helped to improve this manuscript significantly.

References

- Akaike H (1973). *Information Theory and an Extension of the Maximum likelihood principle*. In Second International Symposium on Information Theory, eds. B.N. Petrov and F. Csaki, pp. 267–281. Akademiai Kiado, Budapest.
- Akaike H (1974). “A New Look at the Statistical Model Identification.” *IEEE Transactions on Automatic Control*, **19**(6), 716–723. doi:<https://doi.org/10.1109/TAC.1974.1100705>.
- Andrew AN, Joseph EC (2012). “The Bayesian Information Criterion: Background, Derivation, and Applications.” *WIREs Computational Statistics*, **4**(2), 199–203. doi:<https://doi.org/10.1002/wics.199>.
- Andrews RL, Currim IS (2003). “A Comparison of Segment Retention Criteria for Finite Mixture Logit Models.” *Journal of Marketing Research*, **40**(2), 235–243. URL <http://www.jstor.org/stable/30038851>.
- Balakrishnan Nand Kundu D (2019). “Birnbbaum-Saunders Distribution: A Review of Models, Analysis, and Applications.” *Applied Stochastic Models in Business and Industry*, **35**(1), 4–49. doi:<https://doi.org/10.1002/asmb.2348>.
- Balakrishnan N, Gupta RC, Kundu D, Leiva V, Sanhueza A (2011). “On Some Mixture Models Based on the Birnbbaum-Saunders Distribution and Associated Inference.” *Journal of Statistical Planning and Inference*, **141**(7), 2175–2190. doi:<https://doi.org/10.1016/j.jspi.2010.12.005>.
- Banfield JD, Raftery AE (1993). “Model-based Gaussian and Non-Gaussian Clustering.” *Biometrics*, **49**(3), 803–821. doi:<https://doi.org/10.2307/2532201>.
- Bechtel Y, Bonaiti-Pellie C, Poisson N, Magnette J, Bechtel P (1993). “A Population and Family Study of N-Acetyltransferase Using Caffeine Urinary Metabolites.” *Clinical Pharmacology and Therapeutics*, **54**(2), 134–141. doi:<https://doi.org/10.1038/clpt.1993.124>.
- Benites L, Maehara R, Vilca F, Marmolejo-Ramos F (2017). “Finite Mixture of Birnbbaum-Saunders Distributions Using the k Bumps Algorithm.” URL <https://arxiv.org/abs/1708.00476>.
- Biernacki C, Celeux G, Govaert G (1998). “Assessing a Mixture Model for Clustering with the Integrated Classification Likelihood.” *Inria*, **RR-3521**. URL <https://hal.inria.fr/inria-00073163>.
- Biernacki C, Celeux G, Govaert G (1999). “An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Mode.” *Pattern Recognition Letters*, **20**(3), 267–272. doi:[https://doi.org/10.1016/S0167-8655\(98\)00144-5](https://doi.org/10.1016/S0167-8655(98)00144-5).
- Biernacki C, Govaert G (1997). “Using the Classification Likelihood to Choose the Number of Clusters.” *Computing Science and Statistics*, **29**(2), 451–457.
- Birnbbaum Z, Saunders S (1969). “A New Family of Life Distributions.” *Journal of Applied Probability*, **6**(2), 319–327. URL <https://www.jstor.org/stable/3212003>.

- Böhning D, Seidel W, Alfö M, Garel B, Patilea V, Walther G (2007). “Advances in Mixture Models.” *Computational Statistics & Data Analysis*, **51**, 5205–5210. doi:<https://doi.org/10.1016/j.csda.2006.10.025>.
- Bozdogan H (1987). “Model Selection and Akaike’s Information Criterion (AIC): The General Theory and Its Analytical Extensions.” *Psychometrika*, **52**, 345–370. doi:<https://doi.org/10.1007/BF02294361>.
- Bozdogan H (1994). *Mixture-Model Cluster Analysis Using Model Selection Criteria and a New Informational Measure of Complexity*. Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach. Springer, Dordrecht. doi:https://doi.org/10.1007/978-94-011-0800-3_3.
- Bozdogan H, Sclove SL, Gupta AK (eds.) (1994). *Information-Based Validity Functionals for Mixture Analysis*. Springer. doi:https://doi.org/10.1007/978-94-011-0800-3_5.
- Burnham KP, Anderson DR (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach, 2nd edn*. Berlin, Heidelberg: Springer.
- Celeux G, Soromenho G (1996). “An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model.” *Journal of Classification*, **13**, 195–212. URL <https://doi.org/10.1007/BF01246098>.
- Chen H, Chen J, Kalbfleisch JD (2001). “A Modified Likelihood Ratio Test for Homogeneity in Finite Mixture Models.” *Journal of the Royal Statistical Society, Series B*, **63**(1), 19–29. URL <https://www.jstor.org/stable/2680631>.
- Chen H, Chen J, Kalbfleisch JD (2004). “Testing for a Finite Mixture Model with Two Components.” *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **66**(1), 95–115. URL <https://www.jstor.org/stable/3647629>.
- Dempster AR, Laird NM, Rubin DB (1977). “Maximum Likelihood from Incomplete Data via EM Algorithm.” *Journal of the Royal Statistical Society, Series B*, **39**, 1–38. URL <https://www.jstor.org/stable/2984875>.
- Depraetere N, Vandebroek M (2014). “Order Selection in Finite Mixtures of Linear Regressions.” *Statistical Papers*, **55**, 871–911. doi:<https://doi.org/10.1007/s00362-013-0534-x>.
- El-Sharkawy WA, Ismail MA (2020). “Mixture of Birnbaum-Saunders Distributions: Identifiability, Estimation and Testing Homogeneity with Randomly Censored Data.” *American Journal of Mathematical and Management Sciences*. doi:<https://doi.org/10.1080/01966324.2020.1837041>.
- El-Sharkawy WA, Ismail MA (2021). “Testing the Number of Components in a Birnbaum-Saunders Mixture Model under a Random Censoring Scheme.” *Statistics, Optimization & Information Computing*, **9**, 157–175. doi:<https://doi.org/10.19139/soic-2310-5070-919>.
- Fernández D, Arnold R (2016). “Model Selection for Mixture-Based Clustering for Ordinal Data.” *Australian & New Zealand Journal of Statistics*, **58**(4), 437–472. doi:<https://doi.org/10.1111/anzs.12179>.
- Hathaway RJ (1986). “Another Interpretation of the EM Algorithm for Mixture Distributions.” *Statistics & Probability Letters*, **4**(2), 53–56. doi:[https://doi.org/10.1016/0167-7152\(86\)90016-7](https://doi.org/10.1016/0167-7152(86)90016-7).
- Hurvich CM, Tsai CL (1989). “Regression and Time Series Model Selection in Small Samples.” *Biometrika*, **76**, 297–307. doi:<https://doi.org/10.1093/biomet/76.2.297>.

- Konishi S, Kitagawa G (2007). *Information Criteria and Statistical Modeling*. Springer, New York.
- Kullback S, Leibler RA (1951). “On Information and Sufficiency.” *Annals of Mathematical Statistics*, **22**, 76–86. URL <https://doi.org/10.1214/aoms/1177729694>.
- Leiva V (2016). *The Birnbaum-Saunders Distribution*. Amsterdam, Elsevier.
- Lindsay BG (1995). “Mixture Models: Theory, Geometry and Applications.” *INSF-CBMS Regional Conference Series in Probability and Statistics*, **5**, I–163. URL <http://www.jstor.org/stable/4153184>.
- M M, van der Laan M J (2003). “Fitting of Mixtures with Unspecified Number of Components Using Cross Validation Distance Estimate.” *Computational Statistics & Data Analysis*, **41**(3), 413–428. URL [https://doi.org/10.1016/S0167-9473\(02\)00166-4](https://doi.org/10.1016/S0167-9473(02)00166-4).
- McLachlan G, Peel D (2000). *Finite Mixture Models*. Wiley, New York.
- McLachlan GJ, Basford KE (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.
- McLachlan GJ, Rathnayake S (2014). “On the Number of Components in a Gaussian Mixture Model.” *WIREs Data Mining and Knowledge Discovery*, **4**, 341–355. doi:<https://doi.org/10.1002/widm.1135>.
- Ng HKT, Kundu D, Balakrishnan N (2003). “Modified Moment Estimation for the Two-parameter Birnbaum-Saunders Distribution.” *Computational Statistics & Data Analysis*, **3**(43), 283–298. doi:[https://doi.org/10.1016/S0167-9473\(02\)00254-2](https://doi.org/10.1016/S0167-9473(02)00254-2).
- Niu X (2014). *Homogeneity Test in Finite Mixture Models Using EM-Test*. PhD thesis, University of Alberta, Canada. URL <https://doi.org/10.7939/R3X63BC3Z>.
- Oliveira-Brochado A, Martins FV (2005). *Assessing the Number of Components in Mixture Models: A Review*. FEP Working Papers 194, Universidade do Porto, Faculdade de Economia do Porto. URL <https://EconPapers.repec.org/RePEc:por:fepwps:194>.
- Schlattmann P (2009). *Medical Applications of Finite Mixture Models*. Springer.
- Schwarz G (1978). “Estimating the Dimension of a Model.” *The Annals of Statistics*, **6**, 461–464. doi:<https://doi.org/10.1214/aos/1176344136>.
- Sugiura N (1978). “Further Analysis of the Data by Akaike’s Information Criterion and the Finite Corrections.” *Communications in Statistics, Theory and Methods*, **A7**, 13–26. doi:<https://doi.org/10.1080/03610927808827599>.
- Titterton DM, Smith AF, Makov UE (1985). *Statistical Analysis of Finite Mixture Distributions (Vol. 7)*. New York: Wiley.
- Usami S (2014). “Performance of Information Criteria for Model Selection in a Latent Growth Curve Mixture Model.” *Japanese Society of Computational Statistics*, **27**, 17–48. doi:https://doi.org/10.5183/jjscs.1309001_207.
- Vrieze SI (2012). “Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion and the Bayesian Information Criterion.” *Psychological Methods*, **17**(2), 228–243. doi:<https://doi.org/10.1037/a0027127>.
- Wong TST, Li WK (2014). “Test for Homogeneity in Gamma Mixture Models Using Likelihood Ratio.” *Computational Statistics & Data Analysis*, **70**, 127–137. doi:<https://doi.org/10.1016/j.csda.2013.09.001>.

Yu Y, Harvill JL (2014). “Bootstrap Likelihood Ratio Test for Weibull Mixture Models Fitted to Grouped Data.” *Communications in Statistics - Theory and Methods*, **48**(18), 4550–4568. doi:<https://doi.org/10.1080/03610926.2018.1494838>.

Affiliation:

Walaa A. El-Sharkawy
Department of Mathematics
Faculty of Science,
Cairo University, Egypt
E-mail: shwalaa@sci.cu.edu.eg

Moshira A. Ismail
Department of Statistics,
Faculty of Economics and Political Science,
Cairo University, Egypt
E-mail: moshiraahmed@feps.edu.eg