**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

# ORDER ESTIMATION FOR MULTIVARIATE MIXTURE REGRESSION MODELS

**ABDELAZIZ ALOUI**
LPAIS, FSDM, University Sidi Mohamed Ben Abdellah, Fez, Morocco.

**HASSANIA HAMZAOUI ***
LPAIS, FSDM, University Sidi Mohamed Ben Abdellah, Fez, Morocco.
*Corresponding Author Email: hassania.hamzaoui@usmba.ac.ma

**ABDELAZIZ EL MATOUAT**
LMMPA, ENS, University Sidi Mohamed Ben Abdellah, Fez, Morocco.
University Le Havre Normandie, Le Havre, France.

**Abstract**

In this paper, we study the problem of jointly selecting the number of components and explanatory variables for multivariate mixture regression models. In practice, the selection model using the Akaike Information Criterion $AIC$ is not satisfactory, as it can lead to an overestimation of the number of components and the number of explanatory variables. To improve selection, Naik et al. (2007) developed a new criterion based on Akaike's technique, the Mixture Regression Criterion MRC for the simultaneous determination of the number of components and explanatory variables for univariate mixture regression models. We propose a generalization of the criterion $MRC\ for$ multivariate mixture regression models. The performance of the new criterion is validated on simulated data by comparing it to the Akaike criterion $AIC$ and the Schwarz criterion $BIC$.

**Keywords:** Model Selection, Information Criteria, Criterion $MRC$, Multivariate Mixture Regression Models.

## 1. INTRODUCTION

Mixture regression models combine several regression models to model different sub-populations in relation to the observed data. Each sub-population is modeled by a different regression model whose weight corresponds to the proportion of this sub-population in the total population. Mixture regression models are useful for modeling situations where the relationships between explanatory variables may vary between sub-populations. They are also useful for detecting heterogeneous sub-populations with different relationships between explanatory variables. These models first appeared in the economic literature as "switching regression" (Quandt, 1972; Quandt et al., 1978) [13, 14] and have been widely used to explore the source of heterogeneity when groups of individuals respond differently to a predictor. Over the last decades, mixture regression models have been used in a large number of applications in a wide range of scientific disciplines. These models are commonly used in econometrics, biostatistics, and social sciences.

Selecting the appropriate order in multivariate mixture regression models typically involves determining the number of components in the mixture and the number of explanatory variables in each regression model.

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

The classical model selection criterion $AIC$ (Akaike, 1973) [1] may not yield satisfactory results in selecting both the number of components and explanatory variables, particularly when the sample size is small, it leads to an over estimation of the number of components and the number of variables for mixture regression models. To overcome this problem, Naik et al. [12] developed the Mixture Regression Criterion $MRC$ for the simultaneous determination of the number of components and explanatory variables for univariate mixture regression models, based on the log-likelihood of complete data and asymmetric divergence of Kullback [8] between the true and fitted approximating models. Using the same technique, we derive a new criterion, denoted $MRC_v$ $(vector\ MRC)$, which generalizes the $MRC$ criterion to multivariate mixture regression models.

The paper is organized as follows. We present in Section 2 the multivariate mixture regression model and parameter estimation using the $EM$ algorithm. In Section 3, we derive the criterion $MRC_v$. In Section 4, we highlight the performance of our criterion $MRC_v$ compared to the criteria $AIC$ and $BIC$ on simulated data from multivariate gaussian mixture regression models.

## 1. Multivariate gaussian mixture regression models

Let $y$ be a $m$-dimensional response variable and a $p$-dimensional explanatory variable $x$.

A multivariate gaussian mixture regression model consists of expressing $y$ as a function of $x$ as follows

$$y = \begin{cases} \beta_1' x + \epsilon_1 & with\ the\ probability\ \alpha_1 \\ \beta_2' x + \epsilon_2 & with\ the\ probability\ \alpha_2 \\ \qquad\qquad . & \\ \qquad\qquad . & \\ \qquad\qquad . & \\ \beta_k' x + \epsilon_k & with\ the\ probability\ \alpha_k \end{cases}$$

such that for $j = 1, \dots, k,\ 0 < \alpha_j \leq 1$ and $\sum_{j=1}^{k} \alpha_j = 1,\ \beta_j$ is the $p \times m$ matrix of regression coefficients, and $\epsilon_j$ is the $m-$ dimensional random error, we suppose that $\epsilon_j$ follows a gaussian distribution with mean $0$ and covariance matrix $\Sigma_j$, $\epsilon_j \sim \mathcal{N}_m(0, \Sigma_j)$.

Let $\phi = \{(\alpha_j, \beta_j, \Sigma_j),\ j = 1, \dots, k\}$ be the set of parameters. The conditional density of $y$ given $x$ is

$$f(y; x, \phi) = \sum_{j=1}^{k} \alpha_j f_j\left(y; x, \beta_j, \Sigma_j\right) \qquad (1)$$

$k$ is the number of model components and $f_j$ is the density function of the gaussian distribution of dimension $m$ with mean $\beta_j' x$ and covariance matrix $\Sigma_j$.

$$f_j\left(y; x, \beta_j, \Sigma_j\right) = (2\pi)^{-\frac{m}{2}} \left(\det\left(\Sigma_j\right)\right)^{-\frac{1}{2}} exp\{-\frac{1}{2}(y - \beta_j' x)' \Sigma_j^{-1} (y - \beta_j' x)\}$$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

Let $\{(y_1, x_1), \ldots, (y_n, x_n)\}$ be an independent observed sample from the model (1), the log-likelihood function is

$$L(\phi; Y, X) = \sum_{i=1}^{n} log \left\{ \sum_{j=1}^{k} \alpha_j f_j(y_i; x_i, \beta_j, \sum_j) \right\} \quad (2)$$

where $Y = (y_1, \ldots, y_n)'$ and $X = (x_1, \ldots, x_n)'$.

The maximum likelihood estimator for the parameters $\phi$ is

$$\hat{\phi} = \underset{\phi}{argmax}\{L(\phi; Y, X)\} \quad (3)$$

However, no explicit solution is available due to the complex expression of $L(\phi; Y, X)$. We use the $EM$ (Expectation-Maximization) algorithm (Dempster et al, 1977) [4] to solve this problem in which we introduce a latent variable $Z$ of dimension $n \times k$ such that

$$z_{ij} = \begin{cases} 1 & if\ (y_i, x_i)\ arises\ from\ the\ j^{th} component \\ 0 & otherwise \end{cases}$$

We put $z_i = (z_{i1}, \ldots, z_{ik})$, the completed data became $((y_i, x_i, z_i), i = 1, \ldots, n)$. The log-likelihood function for the completed data is

$$L_c(\phi; Y, X, Z) = \log \prod_{i=1}^{n} \prod_{j=1}^{k} \{\alpha_j f_j(y_i; x_i, \beta_j, \Sigma_j)\}^{z_{ij}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \{log\ \alpha_j + log\ f_j(y_i; x_i, \beta_j, \Sigma_j)\} \quad (4)$$

The EM algorithm is an iterative estimation algorithm, starting with an initial parameter value $\phi^{(\circ)}$. For each iteration, we calculate the new parameters $\phi^{(q+1)}$ from those of the previous iteration $\phi^{(q)}$.

**E step:**

Calculate the expectation of $L_c$ conditional on the observed data and the current parameter. We obtain

$$Q(\phi, \phi^{(q)}) = E(L_c / Y, X, \phi^{(q)}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \tau_{ij}^{(q)} \{\log \alpha_j + \log f_j\ (y_i; x_i, \beta_j, \sum_j)\}$$

where $\tau_{ij}^{(q)} = E(z_{ij} / y_i, x_i, \phi^{(q)}) = P(z_{ij} = 1 / y_i, x_i, \phi^{(q)})$

by the Bayes's formula, we have

$$\tau_{ij}^{(q)} = \frac{\alpha_j^{(q)}\ f_j(y_i; x_i, \beta_j^{(q)}, \sum_j^{(q)})}{\sum_{j=1}^{k} \alpha_j^{(q)} f_j(y_i; x_i, \beta_j^{(q)}, \sum_j^{(q)})} \quad (5)$$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

**M step:**

Compute $\phi^{(q+1)}$ maximizing $Q(\phi, \phi^{(q)})$ with respect to $\phi$. We obtain the following updates

$$\alpha_j^{(q+1)} = \sum_{i=1}^{n} \frac{\tau_{ij}^{(q)}}{n}$$

$$\beta_j^{(q+1)} = (\tilde{X}_j^{(q)\prime}\ \tilde{X}_j^{(q)})^{-1}\tilde{X}_j^{(q)\prime}\ \tilde{Y}_j^{(q)}$$

$$\sum_j^{(q+1)} = \frac{\tilde{Y}_j^{(q)\prime}\left(I - \tilde{H}_j^{(q)}\right)\tilde{Y}_j^{(q)}}{tr(W_j^{(q)})}$$

where $W_j^{(q)} = diag\left(\tau_j^{(q)}\right)$, $\tau_j^{(q)} = (\tau_{1j}^{(q)}, \ldots, \tau_{nj}^{(q)})'$, $\tilde{Y}_j^{(q)} = (W_j^{(q)})^{1/2}Y$

$\tilde{X}_j^{(q)} = (W_j^{(q)})^{1/2}X$ and $\tilde{H}_j^{(q)} = \tilde{X}_j^{(q)}(\tilde{X}_j^{(q)\prime}\ \tilde{X}_j^{(q)})^{-1}\tilde{X}_j^{(q)\prime}$

We denote $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$, $\hat{\beta}_1, \ldots, \hat{\beta}_k$, $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_k$ the estimators obtained by the EM algorithm (we fix a stop threshold for the iterations).

## 3. Estimating the number of components and variables in multivariate gaussian mixture regression models

Our objective is the joint selection of the number of components and explanatory variables in multivariate gaussian mixture regression models.

### 3.1. The criterion $MRC$ for univariate gaussian mixture regression models

In a univariate gaussian mixture regression model, the conditional density of the response variable $y$ as a function of the explanatory variable $x$ is written as follows

$$f(y; x, \phi) = \sum_{j=1}^{k} \alpha_j f_j\left(y; x, \beta_j, \sigma_j^2\right)$$

where $\phi = \left\{\left(\alpha_j, \beta_j, \sigma_j^2\right), j = 1, \ldots, k\right\}$ is the set of parameters such that

$0 < \alpha_j \leq 1$ and $\sum_{j=1}^{k}\alpha_j = 1$, $k$ is the number of model components, $\beta_j$ is a $p \times 1$ parameter vector, $x$ is a fixed $p \times 1$ vector of explanatory variables and $f_j$ is the density function of the univariate gaussian distribution with mean $x'\beta_j$ and variance $\sigma_j^2$. The parameters are estimated by the $EM$ algorithm

$$\hat{\alpha}_j = \sum_{i=1}^{n} \frac{\hat{\tau}_{ij}}{n}$$

$$\hat{\beta}_j = (\hat{X}_j'\hat{X}_j)^{-1}\hat{X}_j'\hat{Y}_j$$

$$\hat{\sigma}_j^2 = \frac{\hat{Y}_j'(I - \hat{H}_j)\hat{Y}_j}{tr(\hat{W}_j)}$$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

where $\widehat{W}_j = diag(\hat{\tau}_j)$, $\hat{\tau}_j = (\hat{\tau}_{1j}, \dots, \hat{\tau}_{nj})'$, $\hat{Y}_j = (\widehat{W}_j)^{1/2} Y$, $\hat{X}_j = (\widehat{W}_j)^{1/2} X$

and $\hat{H}_j = \hat{X}_j(\hat{X}_j'\hat{X}_j)^{-1}\hat{X}_j'$

For the simultaneous determination of the unknown number of components $k^\circ$ and the unknown number $p^\circ$ of explanatory variables for a univariate gaussian mixture regression model, Naik et al. [12] developed the Mixture Regression Criterion $MRC$ based on the Kulback asymmetric divergence and the log-likelihood of complete data

$$MRC(k,p) = \sum_{j=1}^{k} \hat{n}_j \log(\hat{\sigma}_j^2) + \sum_{j=1}^{k} \frac{\hat{n}_j(\hat{n}_j + p_j)}{(\hat{n}_j - p_j - 2)} - 2\sum_{j=1}^{k} \hat{n}_j \log(\hat{\alpha}_j)$$

with $p = (p_1, \dots, p_k)$, $\hat{n}_j = tr(\widehat{W}_j)$ and $p_j = tr(\hat{H}_j)$

$$(\hat{k}, \hat{p}) = \underset{1 \le k \le K,\ 1 \le p_{1,\dots,}p_k \le P}{argmin} \{MRC(k,p)\}$$

where $K$ and $P$ are sufficiently large such that $K \geq k^\circ$ and $P \geq \max\{p_j^\circ, j = 1, \dots, k^\circ\}$.

The criterion $MRC$ is composed of three terms:

The first term measures the lack of fit, the second term imposes a penalty on the regression parameters and the third term penalizes the number of components.

We propose to extend their technique to multivariate gaussian mixture regression models by developing a more general criterion denoted $MRC_v$.

### 3.2. Extension of the criterion $MRC$ for multivariate gaussian mixture regression models

Suppose that the density associated with the true model is

$$f^\circ(y; x^\circ, \phi^\circ) = \sum_{j=1}^{k^\circ} \alpha_j^\circ f_j^\circ(y; x^\circ, \beta_j^\circ, \Sigma_j^\circ) \qquad (6)$$

where $y \in \mathbb{R}^m$, $x^\circ \in \mathbb{R}^{p^\circ}$, $\phi^\circ = \{(\alpha_j^\circ, \beta_j^\circ, \Sigma_j^\circ),\ j = 1, \dots, k^\circ\}$ is the set of true model parameters such that ($0 < \alpha_j^\circ \leq 1$ and $\sum_{j=1}^{k^\circ} \alpha_j^\circ = 1$), $k^\circ$ is the number of components of the true model and $f_j^\circ$ is the density function of the gaussian distribution of dimension $m$ with mean $(\beta_j^\circ)' x^\circ$ and covariance matrix $\Sigma_j^\circ$.

We assume that the class of candidate models, as defined by (1), includes the true model of order $(k^\circ, p^\circ)$. Under this assumption, the columns of $X$ can be rearranged so that $X^\circ \beta_j^\circ = X\beta_j^*$ with $\beta_j^* = ((\beta_j^\circ)', (\beta_j^1)')'$ and $\beta_j^1$ is the null matrix of dimension $(p - p^\circ) \times m$,

for $j = 1, \dots, k^\circ$ when $p \geq p^\circ$. (Edward, J. et al., 1994) [6]

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

The candidate model has been fitted using the observed sample $Y$ and the estimated parameters $\hat{\theta} = (\hat{\phi}, \hat{\tau})$ where $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_k)$ are obtained by the $EM$ algorithm. Let $Y^* = (y_1^*, \dots, y_n^*)'$ be a sample associated with the true model and independent of $Y$, and we measure the goodness of fit by

$$I\left(\theta^\circ, \hat{\theta}(Y)\right) = E_{Y^*_{|\theta^\circ}}\left\{L_c^\circ(\phi^\circ; Z^*, Y^*, X) - L_c(\hat{\phi}(Y); \hat{\tau}(Y), Y^*, X)\right\} \quad (7)$$

where $Z^*$ is an $n \times k^\circ$ matrix such that

$$z_{ij}^* = \begin{cases} 1 & if\ (y_i^*, x_i)\ arises\ from\ the\ j^{th} component \\ 0 & otherwise \end{cases}$$

$E_{Y^*_{|\theta^\circ}}$ denotes the expectation under the true model and $\theta^\circ = (\phi^\circ, \tau^\circ)$ where $\tau_{ij}^\circ = E\left(z_{ij}^*/y_i^*\right)$

We base the order estimation on the minimization of

$$E_{Y_{|\theta^\circ}}\left\{I(\theta^\circ, \hat{\theta}\ (Y))\right\} = E_{Y_{|\theta^\circ}}\left\{E_{Y^*_{|\theta^\circ}}\{L_c^\circ(\phi^\circ; Z^*, Y^*, X)\}\right\} - E_{Y_{|\theta^\circ}}\left\{E_{Y^*_{|\theta^\circ}}\{L_c(\hat{\phi}(Y); \hat{\tau}(Y), Y^*, X)\}\right\}$$

Considering only the terms that depend on the candidate model, the order estimation is obtained by minimizing $E_{Y_{|\theta^\circ}}\left\{E_{Y^*_{|\theta^\circ}}\{-L_c(\hat{\phi}\ (Y); \hat{\tau}(Y), Y^*, X)\}\right\}$. We have

$$L_c\left(\hat{\phi}(Y); \hat{\tau}(Y), Y^*, X\right) = \sum_{j=1}^{k}\sum_{i=1}^{n} \hat{\tau}_{ij}\left\{log\ \hat{\alpha}_j + log\ f_j\left(y_i^*; x_i, \hat{\beta}_j, \hat{\Sigma}_j\right)\right\}$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n} \hat{\tau}_{ij}\left\{log\ \hat{\alpha}_j - \frac{m}{2} log(2\pi)\right\}$$

$$-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n} \hat{\tau}_{ij}\ log(\det(\hat{\Sigma}_j))$$

$$-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n} \hat{\tau}_{ij}\ (y_i^* - \hat{\beta}_j' x)' \hat{\Sigma}_j^{-1} (y_i^* - \hat{\beta}_j' x)$$

Then $L_c\left(\hat{\phi}(Y), \hat{\tau}(Y), Y^*, X\right) = \sum_{j=1}^{k} tr\left(\widehat{W}_j\right) log\ \hat{\alpha}_j - \frac{nm}{2} log(2\pi)$

$$-\frac{1}{2}\sum_{j=1}^{k} tr(\widehat{W}_j)\ log(\det(\hat{\Sigma}_j))$$

$$-\frac{1}{2}\sum_{j=1}^{k} tr\left[\widehat{W}_j^{1/2}\left(Y^* - X\hat{\beta}_j\right)\hat{\Sigma}_j^{-1}\left(Y^* - X\hat{\beta}_j\right)'\widehat{W}_j^{1/2}\right] \quad (8)$$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

So

$$-2E_{Y^*_{|\theta^\circ}}\{L_c(\hat{\Phi}(Y),\hat{\tau}(Y),Y^*,X)\} = -2\sum_{j=1}^{k^\circ} tr\left(\widehat{W}_j\right) log\ \hat{\alpha}_j + nm\ log(2\pi) + \sum_{j=1}^{k^\circ} tr\left(\widehat{W}_j\right) log(det(\hat{\Sigma}_j)) + E_{Y^*_{|\theta^\circ}}\left\{\sum_{j=1}^{k} tr\left[\widehat{W}_j^{1/2}(Y^*-X\hat{\beta}_j)\hat{\Sigma}_j^{-1}(Y^*-X\hat{\beta}_j)'\widehat{W}_j^{1/2}\right]\right\} \quad (9)$$

We put $A = E_{Y^*_{|\theta^\circ}}\left\{\sum_{j=1}^{k} tr\left[\widehat{W}_j^{1/2}\left(Y^*-X\hat{\beta}_j\right)\hat{\Sigma}_j^{-1}\left(Y^*-X\hat{\beta}_j\right)'\widehat{W}_j^{1/2}\right]\right\}$

and $U_j^\circ = Y^* - X^\circ\beta_j^\circ = Y^* - X\beta_j^*$

Then $A = E_{Y^*_{|\theta^\circ}}\left\{\sum_{j=1}^{k} tr\left[\widehat{W}_j^{1/2}(X\beta_j^* + U_j^\circ - X\hat{\beta}_j)\right]\hat{\Sigma}_j^{-1}\left[\widehat{W}_j^{1/2}(X\beta_j^* + U_j^\circ - X\hat{\beta}_j)\right]'\right\}$

$$= E_{Y^*_{|\theta^\circ}}\left[\sum_{j=1}^{k} tr\{\hat{X}_j(\beta_j^*-\hat{\beta}_j)\hat{\Sigma}_j^{-1}(\beta_j^*-\hat{\beta}_j)'\hat{X}_j'\}\right] + E_{Y^*_{|\theta^\circ}}\left[\sum_{j=1}^{k} tr\left\{\widehat{W}_j^{1/2}U_j^\circ\hat{\Sigma}_j^{-1}(U_j^\circ)'\widehat{W}_j^{1/2}\right\}\right]$$

$$= E_{Y^*_{|\theta^\circ}}\left[\sum_{j=1}^{k} tr\{\hat{X}_j(\beta_j^*-\hat{\beta}_j)\hat{\Sigma}_j^{-1}(\beta_j^*-\hat{\beta}_j)'\hat{X}_j'\}\right] + E_{Y^*_{|\theta^\circ}}\left[\sum_{j=1}^{k} tr\{\hat{\Sigma}_j^{-1}(U_j^\circ)'\widehat{W}_j(U_j^\circ)\}\right]$$

$$= \sum_{j=1}^{k^\circ} tr\{\hat{X}_j(\beta_j^*-\hat{\beta}_j)\hat{\Sigma}_j^{-1}(\beta_j^*-\hat{\beta}_j)'\hat{X}_j'\} + \sum_{j=1}^{k^\circ} tr\{\hat{\Sigma}_j^{-1}tr(\widehat{W}_j)\Sigma_j^\circ\} \quad (10)$$

So

$$E_{Y_{|\theta^\circ}}\left\{-2E_{Y^*_{|\theta^\circ}}\{L_c(\hat{\Phi}(Y),\hat{\tau}(Y),Y^*,X)\}\right\} = E_{Y_{|\theta^\circ}}\left\{-2\sum_{j=1}^{k^\circ} tr\left(\widehat{W}_j\right)\hat{\alpha}_j + nm\log(2\pi)\right\} + E_{Y_{|\theta^\circ}}\left\{\sum_{j=1}^{k^\circ} tr\left(\widehat{W}_j\right) log(det(\hat{\Sigma}_j))\right\} + E_{Y_{|\theta^\circ}}\left\{\sum_{j=1}^{k^\circ} tr\{\hat{X}_j(\beta_j^*-\hat{\beta}_j)\hat{\Sigma}_j^{-1}(\beta_j^*-\hat{\beta}_j)'\hat{X}_j'\}\right\} + E_{Y_{|\theta^\circ}}\left\{\sum_{j=1}^{k^\circ} tr\left(\widehat{W}_j\right) tr\left(\hat{\Sigma}_j^{-1}\Sigma_j^\circ\right)\right\} \quad (11)$$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

$\hat{\tau}_j$ is a consistent estimator of $\tau_j^\circ = (\tau_{1j}^\circ, \dots, \tau_{nj}^\circ)$ where $\tau_{ij}^\circ = E(z_{ij}^\circ / y_i, x_i)$ [15, 9, 12]

Consequently, we can consider $W_j^\circ = diag(\tau_j^\circ)$ and $H_j^\circ$ in the third and fourth terms of (11) instead of $\widehat{W}_j$ and $\widehat{H}_j$ respectively, where $H_j^\circ = X_j^\circ \left(X_j^{\circ\prime} X_j^{\circ\prime}\right)^{-1} X_j^{\circ\prime}$ and $X_j^\circ = \left(W_j^\circ\right)^{1/2} X$ and to simplify further we assume that in the true model, the classes are disjoint such that the diagonal elements of $W_j^\circ$ are equal to either $1$ or $0$.

Furthermore, $\hat{\Sigma}_j$ is asymptotically independent of $\hat{\beta}_j$ and $n_j^\circ \hat{\Sigma}_j$ is asymptotically distributed as $Wishart_m(\Sigma_j^\circ, n_j^\circ - p_j^\circ)$, and $E_{Y_{|\theta^\circ}}\left(\hat{\Sigma}_j^{-1}\right) \approx d_j^\circ \left(\Sigma_j^\circ\right)^{-1}$ where $d_j^\circ = \frac{n_j^\circ}{n_j^\circ - \left(p_j^\circ + m + 1\right)}$ (Anderson, pp. 270, 290) [2]. Consequently

$$E_{Y_{|\theta^\circ}} \left\{ \sum_{j=1}^{k^\circ} tr\{X_j^\circ (\beta_j^* - \hat{\beta}_j) \hat{\Sigma}_j^{-1} (\beta_j^* - \hat{\beta}_j)' (X_j^\circ)'\} \right\}$$

$$= \sum_{j=1}^{k^\circ} \left\{ tr \left[ E_{Y_{|\theta^\circ}} \left(\hat{\Sigma}_j^{-1}\right) E_{Y_{|\theta^\circ}} \left\{ \left(\beta_j^* - \hat{\beta}_j\right)' (X_j^\circ)' \, X_j^\circ \left(\beta_j^* - \hat{\beta}_j\right) \right\} \right] \right\}$$

$$\approx \sum_{j=1}^{k^\circ} \left\{ tr \left[ d_j^\circ \left(\Sigma_j^\circ\right)^{-1} E_{Y_{|\theta^\circ}} \left\{ \left(\beta_j^* - \hat{\beta}_j\right)' (X_j^\circ)' \, X_j^\circ \left(\beta_j^* - \hat{\beta}_j\right) \right\} \right] \right\}$$

$$= \sum_{j=1}^{k^\circ} \left\{ d_j^\circ \, tr \left[ \; E_{Y_{|\theta^\circ}} \left\{ \left(\Sigma_j^\circ\right)^{-1} \left(\beta_j^* - \hat{\beta}_j\right)' (X_j^\circ)' \, X_j^\circ \left(\beta_j^* - \hat{\beta}_j\right) \right\} \right] \right\}$$

$$= \sum_{j=1}^{k^\circ} \left\{ d_j^\circ \, E_{Y_{|\theta^\circ}} \left[ vec\left(\hat{\beta}_j - \beta_j^*\right)' \left\{ \left(\Sigma_j^\circ\right)^{-1} \otimes (X_j^\circ)' X_j^\circ \right\} vec\left(\hat{\beta}_j - \beta_j^*\right) \right] \right\}$$

$$= \sum_{j=1}^{k^\circ} \left\{ d_j^\circ p_j^\circ m \right\} \qquad (12)$$

Because $vec\left(\hat{\beta}_j - \beta_j^*\right)' \left\{ \left(\Sigma_j^\circ\right)^{-1} \otimes (X_j^\circ)' X_j^\circ \right\} vec\left(\hat{\beta}_j - \beta_j^*\right) \sim \chi^2_{p_j^\circ m}$ (Edward et al., 1994) [6]

where $vec\left(\hat{\beta}_j - \beta_j^*\right)$ is the vector of dimension $p_j^\circ m$ obtained by stacking the columns of the matrix $\left(\hat{\beta}_j - \beta_j^*\right)$ and $\otimes$ is the Kronecker product.

We have

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

$$E_{Y_{|\theta^{\circ}}}\left\{\sum_{j=1}^{k^{\circ}} tr\left(W_j^{\circ}\right) tr\left(\hat{\Sigma}_j^{-1}\Sigma_j^{\circ}\right)\right\} \approx \sum_{j=1}^{k^{\circ}} tr\left(W_j^{\circ}\right) d_j^{\circ} tr\left[\left(\Sigma_j^{\circ}\right)^{-1}\Sigma_j^{\circ}\right]$$

$$= \sum_{j=1}^{k^{\circ}} tr\left(W_j^{\circ}\right) d_j^{\circ} m \qquad (13)$$

We replace the results (12) and (13) in (11) after we replace $k^{\circ}$ and $W_j^{\circ}$ by $k$ and $\widehat{W_j}$ respectively. (Naik et al., 2007) [12].

Then

$$E_{Y_{|\theta^{\circ}}}\left\{E_{Y^*_{|\theta^{\circ}}}\left\{-2L_c\left(\hat{\phi}(Y), \hat{\tau}(Y), Y^*, X\right)\right\}\right\} \simeq E_{Y_{|\theta^{\circ}}}\left\{-2\sum_{j=1}^{k} \hat{n}_j \, log\, \hat{\alpha}_j + nm\, log(2\pi) + \sum_{j=1}^{k} \hat{n}_j \, log\left(det\left(\hat{\Sigma}_j\right)\right)\right\} + \sum_{j=1}^{k} \hat{d}_j \, mp_j + \sum_{j=1}^{k} \hat{d}_j m \, \hat{n}_j \qquad (14)$$

Finally, we base the model order estimation on the minimization of

$$-2\sum_{j=1}^{k} \hat{n}_j \log \hat{\alpha}_j + nm \log(2\pi) + \sum_{j=1}^{k} \hat{n}_j \log\left(det\left(\hat{\Sigma}_j\right)\right) + \sum_{j=1}^{k} \hat{d}_j mp_j + \sum_{j=1}^{k} \hat{d}_j m \, \hat{n}_j$$

Ignoring the term does not depend on the order$(k, p)$, we obtain the criterion for estimating the order of the multivariate gaussian regression mixture model

$$MRC_v(k,p) = \sum_{j=1}^{k} \hat{n}_j \, log\left(det\left(\hat{\Sigma}_j\right)\right) + \sum_{j=1}^{k} \hat{d}_j m\left(p_j + \hat{n}_j\right) - 2\sum_{j=1}^{k} \hat{n}_j \, log\left(\hat{\alpha}_j\right) \qquad (15)$$

where $p = (p_1, \ldots, p_k)$, $\hat{n}_j = tr\left(\widehat{W}_j\right)$, $p_j = tr\left(\widehat{H}_j\right)$ $and$ $\hat{d}_j = \frac{\hat{n}_j}{\hat{n}_j - \left(m + p_j + 1\right)}$

and we have

$$\left(\hat{k}, \hat{p}\right) = \underset{1 \le k \le K,\ 1 \le p_1, \ldots, p_k \le P}{argmin} \{MRC_v(k,p)\}$$

where $K$ and $P$ are sufficiently large such that $K \geq k^{\circ}$ and $P \geq \max\{p_j^{\circ},\ j = 1, \ldots, k^{\circ}\}$.

### 3.3. Remarks

• $m = 1$ (univariate case)

$$MRC_v(k,p) = \sum_{j=1}^{k} \hat{n}_j \, log\left(\hat{\sigma}_j^2\right) + \sum_{j=1}^{k} \hat{n}_j \frac{\left(\hat{n}_j + p_j\right)}{\left(\hat{n}_j - p_j - 2\right)} - 2\sum_{j=1}^{k} \hat{n}_j \, log\left(\hat{\alpha}_j\right) = \mathrm{MRC}\,(k,p)$$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

where $p = (p_1, \dots, p_k)$

This is the criterion defined by Naik et al. [12]

• $k = 1$ and $m = 1$ (case of a single component and univariate regression)

$$MRC_v(1,p) = nlog(\hat{\sigma}^2) + \frac{n(n+p)}{n-p-2} = AIC_c(p)$$

We obtain the corrected Akaike criterion proposed by Hurvich et al. [7]

• $k = 1$ (case of a single component and multivariate regression)

$$MRC_v(1,p) = nlog(\hat{\Sigma}) + dm(p+n) = AIC_c(p)$$

where $d = \frac{n}{n-(m+p+1)}$

We obtain the corrected Akaike criterion proposed by Edward et al. [6]

## 4. Simulation

We apply the criterion $MRC_v$ to simulated data from multivariate gaussian mixture regression models with the same covariates across components ($p_1 = \cdots = p_k = p$) and compare the results obtained with those of the criteria $AIC$ and $BIC$.

**Selection procedure:**

To simultaneously determine the number of components and explanatory variables for multivariate gaussian mixture regression models, we use the following procedure.

For the given $\{(k,p): k = 1, \dots, K, p = 1, \dots, P\}$

1. We use the K-means algorithm (MacQueen et al., 1967) [10] to classify the observations of a candidate matrix $X$ into $k$ groups so that the initial probabilities $\tau_{ij}^{(\circ)}$ can be estimated to initialize the $EM$ algorithm.
2. We apply the $EM$ algorithm to estimate the parameters of the multivariate mixture regression model
3. We calculate the $MRC_v$, $AIC$, and $BIC$

$$MRC_v(k,p) = \sum_{j=1}^{k} \hat{n}_j \log\left(det\left(\hat{\Sigma}_j\right)\right) + \sum_{j=1}^{k} \hat{d}_j m\left(p+\hat{n}_j\right) - 2\sum_{j=1}^{k} \hat{n}_j \log\left(\hat{\alpha}_j\right)$$

$$AIC(k,p) = -2L\left(\hat{\phi}, Y, X\right) + 2u(k,p)$$

$$BIC(k,p) = -2L\left(\hat{\phi}, Y, X\right) + \log(n)u(k,p)$$

$u(k,p)$ is the number of free parameters of multivariate regression mixture model $u(k,p) = (k-1) + kpm + km(m+1)/2$

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

4. The orders estimated by these criteria are:

$$(\hat{k},\hat{p}) = \underset{1\le k\le K,1\le p\le P}{argmin}\ \{MRC_v(k,p)\}$$

$$(\hat{k},\hat{p}) = \underset{1\le k\le K,1\le p\le P}{argmin}\ \{AIC(k,p)\}$$

$$(\hat{k},\hat{p}) = \underset{1\le k\le K,1\le p\le P}{argmin}\ \{BIC(k,p)\}$$

- **Example 1**: We consider a multivariate gaussian mixture regression model with dimension $(m = 2)$, and the true number of components $(k^{\circ} = 2)$, each component has a multivariate regression model with three explanatory variables $(p^{\circ} = 3)$, the response variable of each component is generated from $Y_j = X_j^{\circ}\beta_j^{\circ} + \epsilon_j^{\circ}$, $(j = 1, 2)$, where:
  - The elements of the $n_1 \times 3$ matrix $X_1^{\circ}$ and the $n_2 \times 3$ matrix $X_2^{\circ}$ are generated from the uniform distributions $U(0,5)$ and $U(5,10)$ respectively
  - The true regression parameters are $\beta_1^{\circ} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\beta_2^{\circ} = \begin{pmatrix} 6 & 6 \\ 7 & 7 \\ 8 & 8 \end{pmatrix}$
  - The rows of the error matrix $\epsilon_j^{\circ}$ are assumed to be independent with identical $\mathcal{N}_2(0, I_2)$ distributions ($j$ = 1,2)

We consider $100$ samples generated from the true model, of size $n = n_1 + n_2 = 30$ and $n = n_1 + n_2 = 60$.

We consider seven candidate regression models $(P = 7)$ stored in a $n_j \times 7$ matrix $X_j$ in each of the two components. The first three columns of $X_j$ $(j = 1, 2)$ are the same as $X_j^{\circ}$ while the last four columns are generated from $U(0,5)$ and $U(10, 15)$ for $X_1$ and $X_2$ respectively. The observed matrix $X$ is obtained by stacking $X_1$ and $X_2$ one on top of the other. Similarly, using the same method with $Y_1$ and $Y_2$, we construct the observed dependent variable $Y$.

We consider five sets of candidate components $(K = 5)$. Hence, we have $35$ possibilities, seven nested regression models in each of the five sets of candidate components. For each sample, we calculate the selection criteria $MRC_v$, $AIC$, and $BIC$ according to the selection procedure described below. The results are summarized in Tables $1$ and $2$.

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

**Table 1: Order frequencies estimated on 100 samples ($k° = 2,\ p° = 3$), $n_1 = n_2 = 15$**

| Criteria | $MRC_v$ | | | | | | | $BIC$ | | | | | | | $AIC$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | $p° = 3$ | 4 | 5 | 6 | 7 | 1 | 2 | $p° = 3$ | 4 | 5 | 6 | 7 | 1 | 2 | $p° = 3$ | 4 | 5 | 6 | 7 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $k° = 2$ | 0 | 0 | **97** | 2 | 0 | 0 | 0 | 0 | 0 | **71** | 1 | 0 | 0 | 0 | 0 | 0 | **19** | 2 | 4 | 4 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 5 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 2 | 15 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 18 | 0 | 0 | 0 | 0 | 3 | 16 | 29 |

**Table 2: Order frequencies estimated on 100 samples ($k° = 2,\ p° = 3$), $n_1 = n_2 = 30$**

| Criteria | $MRC_v$ | | | | | | | $BIC$ | | | | | | | $AIC$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | $p° = 3$ | 4 | 5 | 6 | 7 | 1 | 2 | $p° = 3$ | 4 | 5 | 6 | 7 | 1 | 2 | $p° = 3$ | 4 | 5 | 6 | 7 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $k° = 2$ | 0 | 0 | **99** | 0 | 0 | 0 | 0 | 0 | 0 | **100** | 0 | 0 | 0 | 0 | 0 | 0 | **85** | 8 | 4 | 1 | 1 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

- **Example 2**: The true model is a multivariate gaussian mixture regression model with dimension $(m = 2)$, and three components $(k° = 3)$, each component has a multivariate regression model with four explanatory variables $(p° = 4)$, the response variable of each component is generated from $Y_j = X_j^{\circ}\beta_j^{\circ} + \epsilon_j^{\circ}$, $(j = 1, 2, 3)$ where:
  - The elements of the $n_1 \times 4$ matrix $X_1^{\circ}$, the $n_2 \times 4$ matrix $X_2^{\circ}$ and the $n_3 \times 4$ matrix $X_3^{\circ}$ are generated from the uniform distributions $U(0,5)$, $U(5,10)$ and $U(10,15)$ respectively
  - The true regression parameters are $\beta_1^{\circ} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\beta_2^{\circ} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 4 \end{pmatrix}$ and $\beta_3^{\circ} = \begin{pmatrix} 5 & 5 \\ 6 & 6 \\ 7 & 7 \\ 8 & 8 \end{pmatrix}$
  - The rows of the error matrix $\epsilon_j^{\circ}$ are assumed to be independent, with identical $\mathcal{N}_2(0, I_2)$ distributions $(j = 1, 2, 3)$

We consider 100 samples generated from the true model, of size $n = n_1 + n_2 + n_3 = 30$ and $n = n_1 + n_2 + n_3 = 90$. We consider seven candidate regression models $(P = 7)$ stored in a $n_j \times 7$ matrix $X_j$ in each of the three components. The first four columns

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

of $X_j$ $(j = 1, 2, 3)$ are the same as $X_j^\circ$ while the last three columns are generated from $U(0,5)$, $U(5,10)$ and $U(10,15)$ for $X_1$, $X_2$ and $X_3$ respectively. The observed matrix $X$ is obtained by stacking $X_1$, $X_2$, and $X_3$ one on top of the other. Similarly, using the same method with $Y_1$, $Y_2$, and $Y_3$, we construct the observed dependent variable $Y$. We consider five sets of candidate components $(K = 5)$. The results are summarized in Tables 3 and 4.

**Table 3: Order frequencies estimated on 100 samples ($k^\circ = 3$, $p^\circ = 4$), $n_1 = n_2 = n_3 = 10$**

| Criteria | $MRC_v$ | | | | | | | $BIC$ | | | | | | | $AIC$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | $p^\circ=4$ | 5 | 6 | 7 | 1 | 2 | 3 | $p^\circ=4$ | 5 | 6 | 7 | 1 | 2 | 3 | $p^\circ=4$ | 5 | 6 | 7 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 2 | 6 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $k^\circ = 3$ | 0 | 0 | 0 | **90** | 0 | 0 | 0 | 0 | 0 | 0 | **61** | 7 | 1 | 13 | 0 | 0 | 0 | **14** | 2 | 6 | 18 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 6 | 12 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 6 | 0 | 0 | 0 | 1 | 5 | 13 | 22 |

**Table 4: Order frequencies estimated on 100 samples ($k^\circ = 3$, $p^\circ = 4$), $n_1 = n_2 = n_3 = 30$**

| Criteria | $MRC_v$ | | | | | | | $BIC$ | | | | | | | $AIC$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | $p^\circ=4$ | 5 | 6 | 7 | 1 | 2 | 3 | $p^\circ=4$ | 5 | 6 | 7 | 1 | 2 | 3 | $p^\circ=4$ | 5 | 6 | 7 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $k^\circ = 3$ | 0 | 0 | 0 | **94** | 6 | 0 | 0 | 0 | 0 | 0 | **94** | 6 | 0 | 0 | 0 | 0 | 0 | **83** | 11 | 3 | 3 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

In view of the results, whether for mixture models of two or three distributions, the criterion $MRC_v$ is clearly better than both the criteria $AIC$ and $BIC$ for small samples, in Table 3. for example, with $n = 30$, the percentage of good selection is $90\%$ with $MRC_v$, $61\%$ and $14\%$ with $BIC$ and $AIC$ respectively. For large samples, the criteria $MRC_v$ and $BIC$ have the same performance, in Table 2. for example, with $n = 60$, the percentage of good selection is $99\%$ with $MRC_v$ and $100\%$ with $BIC$. We also observe that the criterion $AIC$ is much less performant than the criteria $BIC$ and $MRC_v$ for both small and large samples.

## 5. CONCLUSION

In this paper, we derived the criterion $MRC_v$ for the joint selection of the number of components and variables for multivariate gaussian mixture regression models, which is based on the log-likelihood and the Kullback asymmetric divergence. We compared our

**Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)**
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 11-2024
DOI: 10.5281/zenodo.14207903

criterion $MRC_v$ with the criteria $AIC$ and $BIC$ on simulated data from multivariate gaussian mixture regression models for small and large samples. The numerical results show that the criterion $MRC_v$ outperforms the criteria $AIC$ and $BIC$ for small samples and has the same performance as the criterion $BIC$ for large samples. Therefore, we suggest using in an application the criterion $MRC_v$ against the criteria $AIC$ and $BIC$.

**References**

1) Akaike, H. (1973). *Information theory and an extension of the maximum likelihood Principle*. In second international symposium of Information Theory, ed. B. N. Petrov and F. Csaki, Akademia Kiado, Budapest, 267-281.

2) Anderson, T.W. (2003). *An introduction to multivariate statistical analysis*. Wiley, New York.

3) Burnham, K. P., and Anderson, D. R. (2002). *Model selection and inference: A practical information-theoretic approach (2nd ed.).* Springer-Verlag, New York.

4) Dempster, A.P., Laird, N.M., and Rubin, D.P. (1977). *Maximum likelihood from incomplete data via the EM algorithm.* In Journal of the Royal Statistical Society Series B, 39(**1**):1-38.

5) Depraetere, N., and Vandebroek, M. (2013). *Order selection in finite mixtures of linear regressions.* Journal of multivariate analysis, 55(**3**):871-911.

6) Edward, J.B., and Tsai, C.L. (1994). *Model Selection for Multivariate Regression in Small Samples*. Biometrics, 50(**1**):226-231.

7) Hurvich, C.M., and Tsai, C.L. (1989). *Regression and Time Series Model Selection in Small Samples.* Biometrika, 76(**2**):297-307.

8) Kullback, S. (1968). *Information theory and statistics*. Dover, New York.

9) Leroux, B. (1992). *Consistent Estimation of a Mixing Distribution*. The Annals of Statistics, 20(**3**):1350 –1360.

10) MacQueen, J.B. (1967). *Some Methods for Classification and Analysis of Multivariate Observations*. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, (**1**): 281–297.

11) McLachlan, G., and Peel, D. (2000). *Finite mixture models*. Wiley, New York.

12) Naik, P.A., Shi, P. & Tsai, C.L. (2007). *Extending the Akaike information criterion to mixture regression models*. Journal of American Statistical Association, 102(**477**): 244-254.

13) Quandt, R.E. (1972). *A new approach to estimating switching regressions*. Journal of American Statistical Association, 67(**338**): 306-310.

14) Quandt, R.E., and Ramsey, J.B. (1978). *Estimating mixtures of normal distributions and switching regressions*. Journal of American Statistical Association, 73(**364**): 730-752.

15) Redner, R.A., and Walker, H.F. (1984). *Mixture Densities, Maximum Likelihood and the EM Algorithm*. SIAM Review, 26(**2**):195–239.

16) Schwarz, G. (1978). *Estimating the dimension of a model*. The Annals of Statistics, 6(**2**):461-464.

17) Shan, A., and Yang, F. (2021). *Bayesian inference for finite mixture regression model based on non-iterative algorithm*. Mathematics, 9(**6**):590.

18) Yu, C., and Wang, X. (2019). *A new model selection procedure for finite mixture regression models.* Communications in Statistics - Theory and Methods, 49(**18**):4347-4366.