

# Selecting the Quantity of Models in Mixture Regression

Dawei Lang, Wanzhou Ye

College of Science, Shanghai University, Shanghai, China  
Email: wzhy@shu.edu.cn

Received 14 June 2016; accepted 22 July 2016; published 25 July 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Mixture regression is a regression problem with mixed data. Specifically, in the observations, some data are from one model, while others from other models. Only after assuming the quantity of the model is given, EM or other algorithms can be used to solve this problem. We propose an information criterion for mixture regression model in this paper. Compared to ordinary information criterion by data simulations, results show our criterion has better performance on choosing the correct quantity of models.

## Keywords

Mixture Regression, Model Based Clustering, Information Criterion, AIC, BIC

---

## 1. Introduction

Mixture regression is a special situation in regression problem. Rather than getting samples in one distribution, the data of mixture regression are from multiple distributions (the information of which distribution every observation from is unknown), which will make a bad effect in parameter estimation. The mixture regression problem can be described as follows [1]:

$$Y = X' \beta_k + \sigma_k \epsilon_k, k = 1, 2, \dots, g \quad (1)$$

$$\pi_{ik} = P(x_i \in \Omega_k) \quad (2)$$

$X_{n \times p}$  is independent observation matrix with  $n$  observations with  $p$  variables.  $x_i, i = 1, 2, \dots, n$  means  $i$ th observation vector from  $n$  observations. The length of  $x_i$  is  $p$ .  $Y_{1 \times n}$  is response variable from observation data with the length of  $n$ .  $\beta_k$  and  $\sigma_k (k = 1, 2, \dots, g)$  is the unknown parameters (weight) of the variable and scale parameter in different models.  $\epsilon$  is a random error independent from  $X_i$ .  $\pi_{ik}$  is the probability of  $i$ th observation is from the  $k$ th distribution  $\Omega_k$ . (i.e.  $\pi_{ik} = P(x_i \in \Omega_k)$ ). To solve the mixture regression problem,

it need two parts. Firstly, confirming which model every sample is from is required. Secondly, parameters in each model should be estimated. That is the reason to call mixture regression model as model-based clustering [2] [3].

For all the mixture regression problem,  $\pi_{ik}$  is unknown which has:  $\sum_{k=1}^g \pi_{ik} = 1$ .

Furthermore,  $Z_{n \times k}$  is defined as classification matrix of mixture regression. Every element of classification matrix is the estimator of  $\pi: \hat{\pi}$ . And  $z_{ik} \in \{0,1\}$  shows the information of  $i$ th observation is from  $k$ th distribution or not ( $z_{ik}=1$  means  $i$ th observation is from  $k$ th distribution). Classification matrix is one of the most important results in mixture regression problem. If we know the true  $Z$ , we can simply split the data into different linear regression and get the parameter estimation.

Parameter estimation can be obtained by EM algorithm. Fraley *et al.*, [4] [5] state the EM algorithm in ordinary mixture regression model which means every model in it is an ordinary linear regression. EM algorithm of ordinary mixture regression is as follows:

Column vector  $z_i$  in classification matrix  $Z$  can be considered as a multinomial distribution. The probability of this multinomial distribution is  $\tau_1, \tau_2, \dots, \tau_g$ . When  $z_i$  is fixed, probability distribution function(PDF) of  $y_i$  is  $\prod_{k=1}^g f_k(y_i | \theta_k)^{z_{ik}}$ . And complete-data likelihood is:

$$\mathcal{L}(\theta_k, \tau_k, z_{ik} | x) = \sum_{i=1}^n \sum_{k=1}^g z_{ik} \log [\tau_k f_k(y_i | \theta_k)] \quad (3)$$

E-step in mixture regression model can be obtained by:

$$\hat{z}_{ik} \leftarrow \frac{\hat{\tau}_k f_k(y_i | \hat{\theta}_k)}{\sum_{j=1}^g \hat{\tau}_j f_j(y_i | \hat{\theta}_j)} \quad (4)$$

When  $z_{ik}$  is fixed, M-step is finished by maximizing  $\tau_k$  and  $\theta$  by Formula (3). For a normal mixture regression problem,  $f_k$  of E-step can be replaced by PDF of normal distribution function  $\phi_k$ :

$$\Phi_k(y_i | -\mu_k, \Sigma_k) = \frac{\exp\left\{-\frac{1}{2}(y_i - \mu_k)^T \Sigma_k^{-1} (y_i - \mu_k)\right\}}{\sqrt{\det(2\pi\Sigma_k)}} \quad (5)$$

As every observation is independent, covariance matrix can be defined as  $\Sigma_k = \lambda I$ , parameter of E-step can be calculated quickly in M-step by:

$$\hat{\tau}_k \leftarrow \frac{n_k}{n}; \quad \hat{\mu}_k \leftarrow \frac{\sum_{i=1}^n \hat{z}_{ik} y_i}{n_k}; \quad n_k \equiv \sum_{i=1}^n \hat{z}_{ik} \quad (6)$$

Song *et al.*, [6] has finished EM algorithm with robust mixture regression. Q Wu *et al.*, [7] proposed EM algorithm in quantile regression. Furthermore D. Lang *et al.*, [8] explained a fast iteration method for mixture regression problem which can solve mixture regression when random error in different distributions.

Moreover, all the algorithms mentioned below is considering the quantity of models  $g$  is known. However this will not happened in every condition. The number of models  $g$  need to be chosen before the algorithm. When  $X$  is a low dimension matrix, a scatter plot can be drawn for choosing  $g$ . To get the true quantity of models, watching scatter plot and giving a conclusion is not suitable for a high-dimension situation. It was meaningful to discussing how to create a proper method choosing the right quantity of models in a mixture regression problem.

The rest of the paper is organized as follows. Section 2 will discuss the equivalence between mixture regression and ordinary regression when classification matrix is fixed. We extend a method based on information criterion in Section 3. Section 4 is the data simulation of different information criterions. Proof of theorem is in the [Appendix](#) section.

## 2. Equivalence of Linear Regression

Unsupervised learning has its method to choose the quantity of clusters, like GAP statics in K-means [9].

Mixture regression can be regarded as a model based clustering including judging which cluster every observation should be grouped as well as the parameter estimation.

To find a proper method for choosing the quantity of models, we need to find the relationships between mixture regression and other algorithms. In some conditions, such as classification matrix  $Z$  is fixed and random error has the same variance, mixture regression can be written as a linear regression.

**Theorem 1 (Equivalence between Mixture Regression and Linear Regression)** *If the estimator of  $\pi$ , classification  $Z = \hat{\pi}$  is fixed, mixture regression can be written as*

$$Y = \tilde{X}\beta + \tilde{\epsilon} \quad (7)$$

When random error in every model is independent and identically distributed from a normal distribution ( $\sigma_0^2 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_g^2$ ). Random error in mixture regression is from a normal distribution, either.

$$\tilde{\epsilon} \sim N(0, \sigma_0^2)$$

The proof can be found in the [Appendix](#).

After proofing this theorem, we can use the evaluation methodology from regression to solve the quantity choosing in mixture regression.

### 3. Information Criterion for Quantity of Clusters Choosing

#### 3.1. Information Criterion

For a regression problem, Akaike information criterion (AIC) or Bayesian information criterion (BIC) [10] is always used for evaluating a regression model [11]. Information criterion is based on information theory, it shows the information lost in a specify model. A trade-off between goodness of fitting and the complexity of the model is considered in information criterion:

$$\text{AIC} = 2k - 2\ln(L) \quad (8)$$

$$\text{BIC} = \ln(n) \cdot k - 2 \cdot \ln(L). \quad (9)$$

The best model is the one with the minimum AIC (BIC).  $L$  is the likelihood function which states the goodness of fitting (expression (3)).  $k$  is the penalty of the information criterion standing for the number of unknown parameters in the model. In linear regression,  $k$  means the number of dependent variables. As for BIC, the penalty is larger, weight of penalty comes to  $\ln(n)$  from 2.

#### 3.2. Information Criterion in Mixture Regression

In mixture regression, parameters in classification matrix should be considered as part of the estimator variables. Despite these variables, the model will tend to choosing a larger quantity of models which is also an overfitting problem.

For every observation, 0–1 variable with the number of  $g-1$  can ensure classification among  $g$  models. For example, if  $g=2$ , for the  $i$ th observation,  $z_{i1}$  can complete determinate  $i$ th observation is from which cluster(model). As for the situation of  $g=3$ ,  $z_{i1}, z_{i2}$  are requested to determinate the  $i$ th observation.  $k$  value (number of unknown parameters in the model) in information criterion of mixture regression should be:

$$k = g * p + n(g-1) \quad (10)$$

Akaike information criterion for Mixture regression(AICM) and Bayesian information criterion for mixture (BICM) regression is:

$$\text{AICM} = 2[g * p + n(g-1)] - 2\ln(L) = 2n(g-1) + \sum_{k=1}^g \text{AIC}(\text{model}_g) \quad (11)$$

$$\text{BICM} = \ln(n) \cdot [g * p + n(g-1)] - 2 \cdot \ln(L) = \ln(n)n(g-1) + \sum_{k=1}^g \text{AIC}(\text{model}_g) \quad (12)$$

AICM and BIC can be used for the quantity selecting in mixture regression problem. However, penalty weight for  $g$  in BICM is  $\ln(n)n$ , rather than  $2n$  in AICM which will lead to an underfitting result when  $g$  is larger. We will see the details of this point in next section.

### 4. Data Simulation

In order to validating the rationality of the model, we designed numeric simulations and generated sample data  $(X, Y)$ :

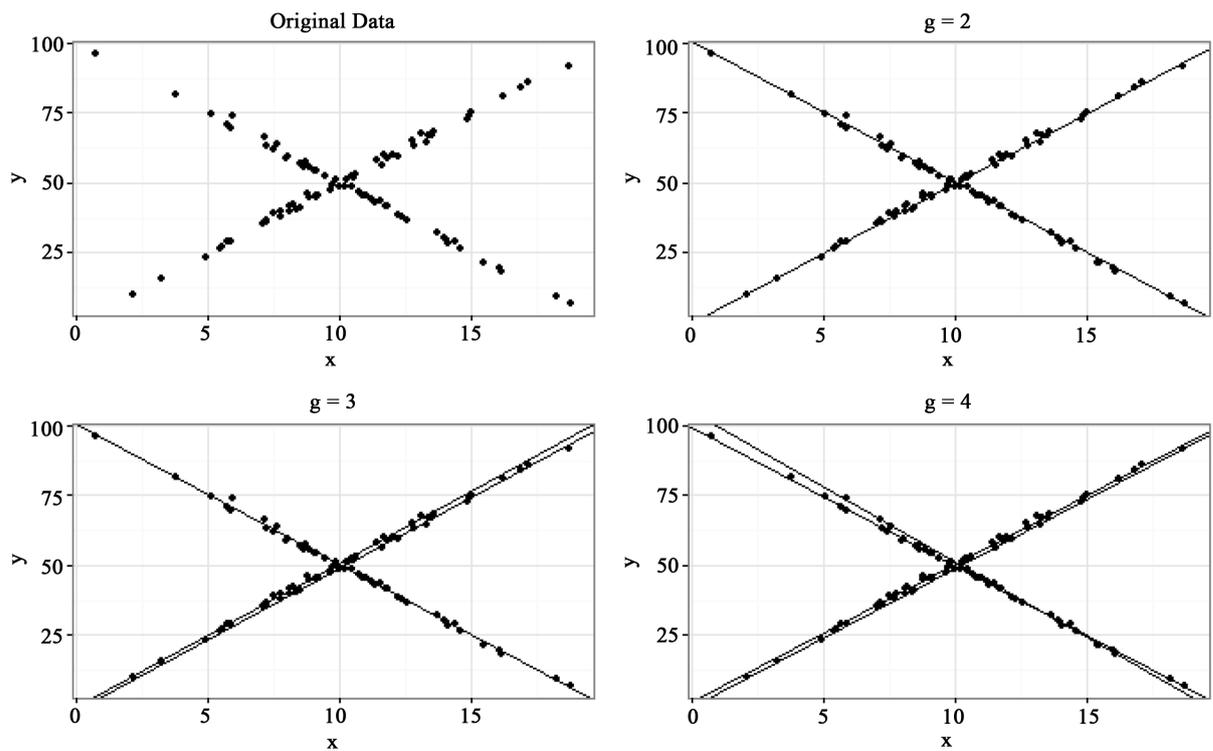
- Simulation I: 100 samples from 2 distributions. ( $n = 100, g = 2$ ).
- Simulation II: 200 samples from 2 distributions. ( $n = 200, g = 2$ ).
- Simulation III: 150 samples from 3 distributions. ( $n = 150, g = 3$ ).

#### 4.1. Simulation I

Models from simulation I is:

$$Y = \begin{cases} 5X + \epsilon, & \text{if } g = 1 \\ -5X + 100 + \epsilon, & \text{if } g = 2 \end{cases} \tag{13}$$

where  $\epsilon \sim N(0,1)$  and  $x \sim n(10,4)$ . Every distribution has 50 observations. See **Figure 1** to see the results when  $g = 1, 2, 3, 4$ . We repeated the simulation for 100 times, use Mixreg package in R [12] to got the answer in **Table 1**.



**Figure 1.** Mixture regression when  $n = 50 \times 2$  and  $g = 2, 3, 4$ .

**Table 1.** Simulation I of selecting quantity of models.

Lightaqua		Mean Value of Information Criterionn				Selected			
Lightaqua Rules	Sample Size	$g = 1$	$g = 2$	$g = 3$	$g = 4$	1	2(**)	3	4
AIC	50*2	883.99	290.1	235.36	193.81	0	0	0	100
BIC	50*2	891.8	295.31	243.17	204.23	0	0	0	100
AICM	50*2	883.99	490.1	635.36	793.81	0	98	2	0
BICM	50*2	891.8	755.82	1164.21	1585.78	2	98	0	0
Std.	-	14.68	67.45	18.04	22.75	-	-	-	-

### 4.2. Simulation II

The models in simulation II is same as simulation I. While, the samples in simulation II is 100 for each distribution.

$$Y = \begin{cases} 5X + \epsilon, & \text{if } g = 1 \\ -5X + 100 + \epsilon, & \text{if } g = 2 \end{cases} \quad (14)$$

Figure 2 can be found in Appendix for simulation 2. Table 2 below is results for repeating 100 simulation.

### 4.3. Simulation III

Simulation III has three distributions with 50 samples in each distribution.

$$Y = \begin{cases} 5X + \epsilon, & \text{if } g = 1 \\ -5X + 100 + \epsilon, & \text{if } g = 2 \\ 0.2X + 50 + \epsilon, & \text{if } g = 3 \end{cases} \quad (15)$$

See Figure 3 for simulation III in Appendix, and result is shown in Table 3.

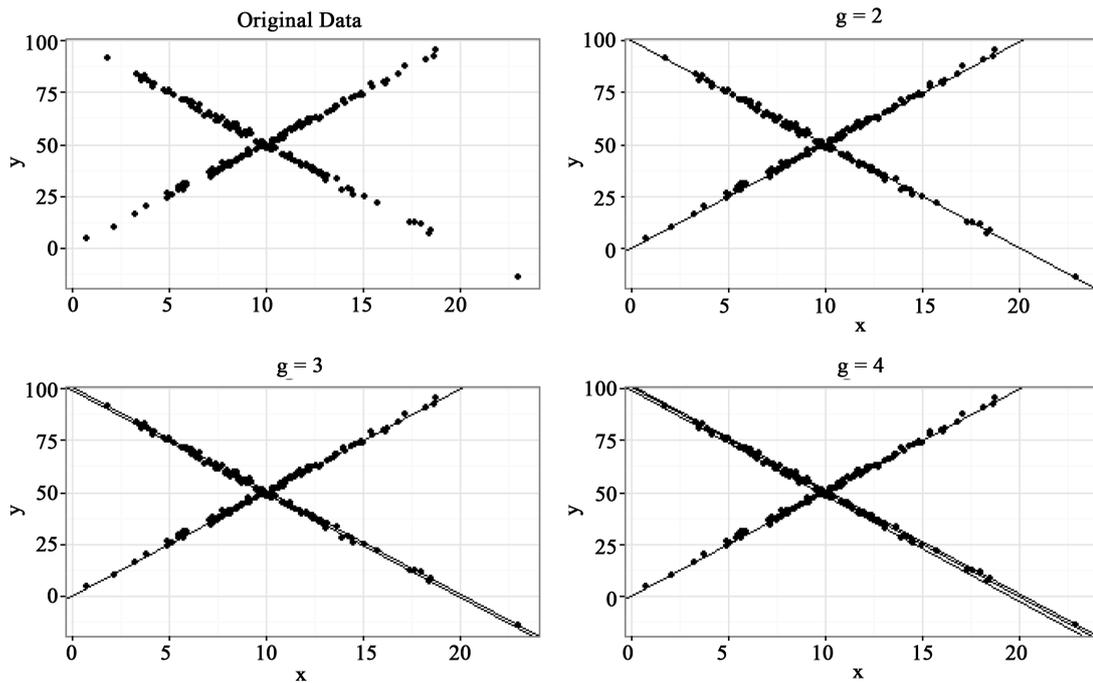


Figure 2. Mixture regression when  $n = 100 \times 2$  and  $g = 2, 3, 4$ .

Table 2. Simulation II of selecting quantity of models.

Lightaqua		Mean Value of Information Criterion				Selected			
Lightaqua Rules	Sample Size	$g = 1$	$g = 2$	$g = 3$	$g = 4$	1	2(**)	3	4
AIC	100*2	1766.66	565.28	484.74	413.171	0	0	2	98
BIC	100*2	1776.55	571.87	494.64	426.36	0	0	2	98
AICM	100*2	1766.66	965.28	1284.74	1613.17	0	100	0	0
BICM	100*2	1776.55	1631.54	2613.96	3605.35	0	100	0	0
Std.	-	20.51	18.10	30.88	34.94	-	-	-	-

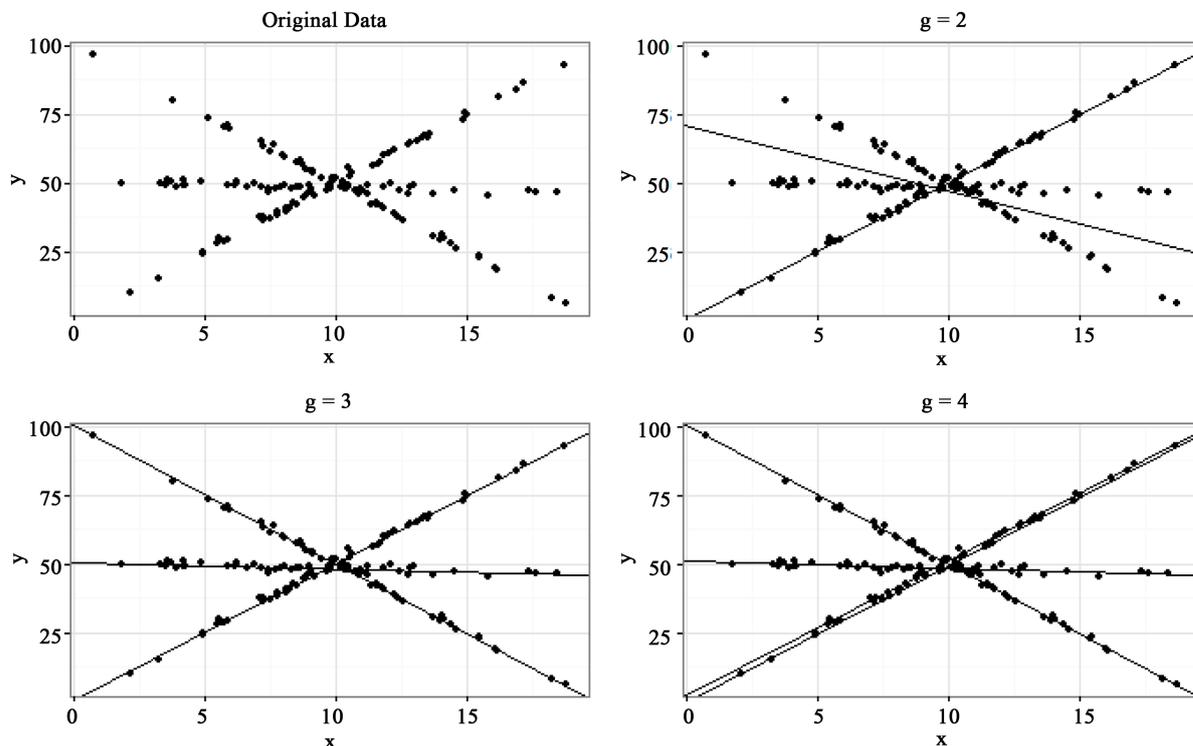


Figure 3. Mixture regression when  $n = 50 \times 3$  and  $g = 2, 3, 4$ .

Table 3. Simulation III of selecting quantity of models.

Lightaqua		Mean Value of Information Criterion					Selected				
Lightaqua Rules	Sample Size	$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 5$	1	2	3(**)	4	5
AIC	50*3	1264.38	908.35	429.91	368.35	326.90	0	0	0	1	99
BIC	50*3	1273.41	914.38	438.95	380.40	341.95	0	0	0	3	97
AICM	50*3	1264.38	1208.35	1029.91	1268.35	1526.9	1	2	97	0	0
BICM	50*3	1273.41	1665.97	1942.14	2635.18	3348.33	100	0	0	0	0
Std.	-	22.09	64.34	97.55	23.38	24.18	-	-	-	-	-

### 5. Conclusion

According to the results in three simulations, we can see AICM and BICM show a good result in small  $g$  ( $g = 2$ ) which choose the true quantity of models at a rate over 98%. While, ordinary AIC and BIC cannot point out the right quantity even once. In large samples, AICM and BICM perform well in simulation II. In small samples, simulation I, AICM tends to overfit the quantity and BICM tend to underfit the quantity in low probability of 2%. Simulation III shows an interesting results when  $g = 3$ ; BICM is too underfitting, which means the weight of penalty is too large for selecting the quantity. AICM choose correctly for 97 times among 100 times. That validates the information we gave in Section 3.

### References

[1] McLachlan, G. and Peel, D. (2004) Finite Mixture Models. John Wiley & Sons, Hoboken.  
 [2] Fraley, C. and Raftery, A.E. (2002) Model-Based Clustering, Discriminant Analysis, and Density Estimation. *American Statistical Association*, **97**, 611-631. <http://dx.doi.org/10.1198/016214502760047131>

- 
- [3] Ingrassia, S., Minotti, S.C. and Punzoa, A. (2014) Model-Based Clustering via Linear Cluster-Weighted Models. *Computational Statistics and Data Analysis*, **71**, 159-182. <http://dx.doi.org/10.1016/j.csda.2013.02.012>
  - [4] Fraley, C. and Raftery, A.E. (1998) How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis. *The Computer Journal*, **41**, 578-588. <http://dx.doi.org/10.1093/comjnl/41.8.578>
  - [5] Fraley, C. and Raftery, A.E. (2002) Model-Based Clustering, Discriminant Analysis, and Density Estimation. *American Statistical Association*, **97**, 611-631. <http://dx.doi.org/10.1198/016214502760047131>
  - [6] Song, W.X., Yao, W.X. and Xing, Y.R. (2014) Robust Mixture Regression Model Fitting by Laplace Distribution. *Computational Statistics and Data Analysis*, **71**, 128-137. <http://dx.doi.org/10.1016/j.csda.2013.06.022>
  - [7] Wu, Q. and Yao, W. (2016) Mixtures of Quantile Regressions. *Computational Statistics & Data Analysis*, **93**, 162-176. <http://dx.doi.org/10.1016/j.csda.2014.04.014>
  - [8] Lang, D.W. and Ye, W.Z. (2015) A Fast Iteration Method for Mixture Regression Problem. *Journal of Applied Mathematics and Physics*, **3**. <http://dx.doi.org/10.4236/jamp.2015.39136>
  - [9] Tibshirani, R., Walther, G. and Hastie, T. (2002) Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63**, 411-423.
  - [10] Aho, K., Derryberry, D. and Peterson, T. (2014) Model Selection for Ecologists: The Worldviews of AIC and BIC. *Ecology*, **95**, 631-636. <http://dx.doi.org/10.1890/13-1452.1>
  - [11] Naik, P.A., Shi, P. and Tsai, C.-L. (2007) Extending the Akaike Information Criterion to Mixture Regression Models. *Journal of the American Statistical Association*, **102**.
  - [12] Mixreg, R.T. Functions to Fit Mixtures of Regressions. R Package Version 0.0-5.

## Appendix

### Proof of theorem 1

*Proof.* Linear regression has the form of:

$$Y = X\beta + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

To proof this theorem, mixture regression need to be written as the form above. And when every random error has the same variance, random error in mixture regression is also a normal distribution.

$$Y = \tilde{X}\beta + \tilde{\epsilon}; \quad \tilde{\epsilon} \sim N(0, \tilde{\sigma}^2)$$

In mixture regression problem, *ith* observation  $\left\{ x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T, y_i \right\}$  can be written as:

$$y_i = \begin{cases} \beta_{11}x_{i1} + \beta_{21}x_{i2} + \dots + \beta_{p1}x_{ip} + \epsilon_1, & \text{if } z_{i1} = 1 \\ \beta_{12}x_{i1} + \beta_{22}x_{i2} + \dots + \beta_{p2}x_{ip} + \epsilon_2, & \text{if } z_{i2} = 1 \\ \vdots \\ \beta_{1g}x_{i1} + \beta_{2g}x_{i2} + \dots + \beta_{pg}x_{ip} + \epsilon_g, & \text{if } z_{ig} = 1 \end{cases}$$

We have:

$$\begin{aligned} y_i &= \beta_{11}x_{i1}z_{i1} + \beta_{21}x_{i2}z_{i1} + \dots + \beta_{p1}x_{ip}z_{i1} + \epsilon_{i1}z_{i1} + \beta_{12}x_{i1}z_{i2} + \beta_{22}x_{i2}z_{i2} \\ &\quad + \dots + \beta_{p2}x_{ip}z_{i2} + \epsilon_{i2}z_{i2} + \dots + \beta_{1k}x_{i1}z_{ik} + \beta_{2k}x_{i2}z_{ik} + \dots + \beta_{pk}x_{ip}z_{ik} \\ &\quad + \epsilon_{i3}z_{i3} + \dots + \beta_{1g}x_{i1}z_{ig} + \beta_{2g}x_{i2}z_{ig} + \dots + \beta_{pg}x_{ip}z_{ig} + \epsilon_{ig}z_{ig} \\ y_i &= \beta_{11}x_{i1}z_{i1} + \beta_{12}x_{i1}z_{i2} + \dots + \beta_{1k}x_{i1}z_{ik} + \dots + \beta_{1g}x_{i1}z_{ig} + \dots + \beta_{21}x_{i2}z_{i1} \\ &\quad + \beta_{22}x_{i2}z_{i2} + \dots + \beta_{2k}x_{i2}z_{ik} + \dots + \beta_{2g}x_{i2}z_{ig} + \dots + \beta_{p1}x_{ip}z_{i1} + \beta_{p2}x_{ip}z_{i2} \\ &\quad + \dots + \beta_{pk}x_{ip}z_{ik} + \dots + \beta_{pg}x_{ip}z_{ig} + \epsilon_{i1}z_{i1} + \epsilon_{i2}z_{i2} + \dots + \epsilon_{ik}z_{ik} + \dots + \epsilon_{ig}z_{ig} \\ y_i &= \tilde{X}_i\beta + \tilde{\epsilon}_i \end{aligned}$$

Because *ith* observation can be written as a product of vectors, population of observation can be written as  $Y = \tilde{X}\beta + \tilde{\epsilon}$ . Where  $\tilde{X}$  has:

$$\begin{aligned} \tilde{X} &= \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ip} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n1} & \dots & x_{np} \end{pmatrix}_{n \times p} \cdot \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{pmatrix}_{p \times 2p} \times (Z \ Z \ \dots \ Z)_{n \times gp} \\ &= \begin{pmatrix} x_{11} & \dots & x_{11} & x_{12} & \dots & x_{12} & \dots & x_{1p} & \dots & x_{1p} \\ x_{21} & \dots & x_{21} & x_{22} & \dots & x_{22} & \dots & x_{2p} & \dots & x_{2p} \\ \vdots & \vdots \\ x_{i1} & \dots & x_{i1} & x_{i2} & \dots & x_{i2} & \dots & x_{ip} & \dots & x_{ip} \\ \vdots & \vdots \\ x_{n1} & \dots & x_{n1} & x_{n2} & \dots & x_{n2} & \dots & x_{np} & \dots & x_{np} \end{pmatrix}_{n \times 2p} \times \begin{pmatrix} z_{11} & \dots & z_{1g} & z_{11} & \dots & z_{1g} & \dots & z_{11} & \dots & z_{1g} \\ z_{21} & \dots & z_{2g} & z_{21} & \dots & z_{2g} & \dots & z_{21} & \dots & z_{2g} \\ \vdots & \vdots \\ z_{i1} & \dots & z_{ig} & z_{i1} & \dots & z_{ig} & \dots & z_{i1} & \dots & z_{ig} \\ \vdots & \vdots \\ z_{n1} & \dots & z_{ng} & z_{n1} & \dots & z_{ng} & \dots & z_{n1} & \dots & z_{ng} \end{pmatrix}_{n \times 2p} \\ &= \begin{pmatrix} x_{11}z_{11} & \dots & x_{11}z_{1g} & x_{12}z_{11} & \dots & x_{12}z_{1g} & \dots & x_{1p}z_{11} & \dots & x_{1p}z_{1g} \\ x_{21}z_{21} & \dots & x_{21}z_{2g} & x_{22}z_{21} & \dots & x_{22}z_{2g} & \dots & x_{2p}z_{21} & \dots & x_{2p}z_{2g} \\ \vdots & \vdots \\ x_{i1}z_{i1} & \dots & x_{i1}z_{ig} & x_{i2}z_{i1} & \dots & x_{i2}z_{ig} & \dots & x_{ip}z_{i1} & \dots & x_{ip}z_{ig} \\ \vdots & \vdots \\ x_{n1}z_{n1} & \dots & x_{n1}z_{ng} & x_{n2}z_{n1} & \dots & x_{n2}z_{ng} & \dots & x_{np}z_{n1} & \dots & x_{np}z_{ng} \end{pmatrix}; \end{aligned}$$

$$\beta = (\beta_1, \beta_2, \dots, \beta_g)^T = (\beta_{11}, \beta_{21}, \dots, \beta_{p1}, \dots, \beta_{1g}, \dots, \beta_{pg})$$

For the observation  $\{x_i, y_i\}$  is samed as *ith* single observation above. In this way, a mixture regression can be written as  $Y = \tilde{X}\beta + \tilde{\epsilon}$ . As for the distribution of random error  $\tilde{\epsilon} = \epsilon_1 z_{i1} + \epsilon_2 z_{i2} + \dots + \epsilon_k z_{ik} + \dots + \epsilon_g z_{ig}$  has:

$$\epsilon_i \sim N(0, \sigma_i^2); \quad \sigma_0^2 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_g^2 \quad (16)$$

$z_{ik} \in \{0, 1\}$  is from a multivariate distribution, probability of  $z_{ik} = 1$  is  $p_k$ :

$$\sum_{k=1}^g z_{ik} = 1; \quad \sum_{k=1}^g p_k = 1 \quad (17)$$

In the distribution of variable  $\tilde{\epsilon}$ , for any  $k$  has:

$$\begin{aligned} F_{\tilde{\epsilon}}(k) &= P(\tilde{\epsilon} < k) = P(\epsilon_1 z_{i1} + \epsilon_2 z_{i2} + \dots + \epsilon_k z_{ik} + \dots + \epsilon_g z_{ig} < k) \\ &= p_1 P(\epsilon_1 < k) + p_2 P(\epsilon_2 < k) + \dots + p_g P(\epsilon_g < k) \\ &= p_1 \Phi(k/\sigma_1) + p_2 \Phi(k/\sigma_1) + \dots + p_g \Phi(k/\sigma_g) \\ &= \sum_{k=1}^g p_g \Phi(k/\sigma_0) = \Phi(k/\sigma_0) \sum_{k=1}^g p_g = \Phi(k/\sigma_0) \end{aligned}$$

so  $\tilde{\epsilon} \sim N(0, \sigma_0^2)$



**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc  
 A wide selection of journals (inclusive of 9 subjects, more than 200 journals)  
 Providing a 24-hour high-quality service  
 User-friendly online submission system  
 Fair and swift peer-review system  
 Efficient typesetting and proofreading procedure  
 Display of the result of downloads and visits, as well as the number of cited articles  
 Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>