AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

The Use Of Genetic Algorithm In Estimating The Parameter Of Finite Mixture Of Linear Regression

Urdak I. Kareem	Fadhaa M. Hashim
Iraqurdak124ibrahim@gmail.com	fali197477@gmail.com
Dept. of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq	Dept. of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

Article Information

Article History:
Received: April, 4, 2021
Accepted: May, 6, 2021
Available Online: June, 25, 2022

Keywords:
Mixture of linear regression, the a robust bi-square (MixBi), MM-Estimator, Gaussian Mixture, RobGA, Classification Error(CE)..

Correspondence:
Urdak I. Kareem
Iraqurdak124ibrahim@gmail.com

Abstract

The estimation of the parameters of linear regression is based on the usual Least Square method, as this method is based on the estimation of several basic assumptions. Therefore, the accuracy of estimating the parameters of the model depends on the validity of these hypotheses. The most successful technique was the robust estimation method which is minimizing maximum likelihood estimator (MM-estimator) that proved its efficiency in this purpose. However, the use of the model becomes unrealistic and one of these assumptions is the uniformity of the variance and the normal distribution of the error. These assumptions are not achievable in the case of studying a specific problem that may include complex data of more than one model. To deal with this type of problem, a mixture of linear regression is used to model such data. In this article, we propose a genetic algorithm-based method combined with (MM-estimator), which is called in this article (RobGA), to improve the accuracy of the estimation in the final stage. We compare the suggested method with robust bi-square (MixBi) in terms of their application to real data representing blood sample. The results showed that RobGA is more efficient in estimating the parameters of the model than the MixBi method with respect to mean square error (MSE) and classification error (CE).

DOI: https://doi.org/10.55562/jrucs.v51i1.536

1. Introduction

Mixtures of linear regression models is used to model the relationship between variables. They comprise multiple unknown homogeneous group variables i.e. modeling data that do not follow a single linear regression model. It is used in many sciences, including biology, economics, marketing and medicine. R.Quandt and J.Ransey are the first to employ mixture of linear regression [31]. The mixture of linear regression model is known as clusterwise linear regression [34]. They defined it as a switching regression model. In addition, the moment-generating function method is used to

estimate its parameters. P.N.Jones and G.J.Mclachlan applied mixture of linear regression models to data by using an expectation-maximization algorithm(EM) to fit these models [30].

Various types of $\varpi -$ functions are discussed by Hennig and an upper limit to the maximal asymptotic bias of MM estimates is presented [17].

T.R.Turner fits a two-component combination of a single variable linear regression of a data set using the algorithm Expectation Maximization [36].

Some researchers proposed the use of the weighted factor for each of the data for the robust of the estimation process for mixture of linear regression models. There are some relevant robust methods for linear cluster [26],[18],[19],[33].

Mueller, et. al. and Berrendero et al. concluded that the maximum limits of the asymptotic bias match a moderately large limit (about 0.3) [5].

Another researchers clarify the sensitivity of the Maximum Likelihood Estimate (MLE) such as the least squares estimate for outliers or Heavy tailed error distributions.[12],[13],[38]

The paper is divided into many sections, section 1 gives a general review of mixture of linear regression model. In section 2 and 3, the method of estimating the parameter of linear regression model is explained. Section 4 describes the application followed by discussing the result by comparing the two methods.

2. Method

To define the linear regression model, we assume (x_i, y_i) that the component is a pair of the y_i response variable and a set of independent variables x_i, It is assumed that the response variable y_i depend on the independent variables x_i modeled by the multiple regression model and that for each pair(x_i, y_i) of observations there is a G latent random indicator. Here we have $P(x) = \pi_j$ for $j = 1,2, \dots, d$ then the mixture of linear regression model can be express as follows[10]:

$$y_i = x_i'\beta_j + \epsilon_{ij} \quad (1)$$

Where :

y_i: is the value of the response variable of i^{th} observations.

x_i' : stands for the transformer p-dimensional vector for i^{th} observations of the independent variables.

β_j: is the regression parameter vector with (p + 1) dimensional for j^{th} of the component.

ϵ_{ij}: is random error Assuming that the error is random within each compound, each of them Identical Independent distribution where $\epsilon_{ij} \sim N(0, \sigma^2{}_i)$.

k: sample size.

In a more general form, the joint probability density function for a mixture of linear regression can be expressed[10]:

$$f(y_i; x_i, \varphi) = \sum_{j=1}^{d} \pi_j \vartheta_j(y_i; x_i, \beta_i, \sigma_i^2) \quad (2)$$

Here π_jIt is the probabilities of mixture models..where $\sum_{j=1}^{d} \pi_j = 1$ for $j = 1, \dots, d$ and $0 < \pi_j < 1$. And $\vartheta_j(y_i; x_i, \beta_i, \sigma_i^2)$ is the density function of normal distribution with mean $x_i'\beta_j$ and variance σ_i^2. Then $\varphi = (\beta_1, \dots, \beta_d, \sigma_1^2, \dots . \sigma_d^2, \pi_1, \dots, \pi_d)$ represents the complete parameter which can be estimated by using maximize log-likelihood[10]:

$$L(\varphi|x) = \sum_{i=1}^{k} \log\left(\sum_{j=1}^{d} \pi_j \vartheta_j(y_i; x_i, \beta_i, \sigma_i^2)\right) \quad (3)$$

This is based on the view of the difficulty that we face in the case of applying the maximize log-likelihood method to solve equation (3). Therefore, several methods will be adopted to find the estimates $\hat{\pi}$, $\hat{\beta}$, $\hat{\sigma}^2$, and to find the solution. Thus, we use two method and comparison between them to get the best performance.

3. Estimation

In order to obtain the mixture of linear regression model, the parameters of the mixture of linear regression model is estimated, as there are several methods for estimating the parameters as the two methods are relied on

3.1. Introduction Robust mixture of linear regression models

As the Log-likelihood function eq.(3) of the greatest potential is infinite, it tends to infinity if one of the observations lies precisely, it is a component on a single component line and the contrast of the corresponding component tends to be zero. There are many studies that dealt with that Log-likelihood function unbounded such as that given by Hathaway R. J. [15,16], Chen, J. et. al. [7], and Yao, W. [40]. Therefore to avoid the log-likelihood function in eq.(3) unbounded, we assume the variance is equally.

The expectation maximization algorithm to maximize eq(3) can be as follows.

- **Algorithm1.** This algorithm is based on the initial values of $\{\sigma^{(0)}, \beta_j^{(0)}, \pi_j^{(0)}, j = 1, \dots, d\}$ As in the expectation maximization algorithm, the two steps are repeated, namely step E and step M.
  - **E-step:** In this step, we calculate the classification probabilities[3]:

$$q_{ij}^{(m+1)} = \frac{\pi_j^{(m)} \zeta(y_i; x_i' \beta_j^{(m)}, \sigma^{2(m)})}{\sum_{j=1}^{d} \pi_j^{(m)} \zeta(y_i; x_i' \beta_j^{(m)}, \sigma^{2(m)})} , i = 1, \dots, k, j = 1, \dots, d$$

Here: $\zeta\left(y_i; x_i' \beta_j^{(m)}, \sigma^{2(m)}\right)$ is a normal density.

  - **Step-M: In this step we update the parameter.[3]**

$$\beta_j^{(m+1)} = \left(X W_j^{(m+1)} X'\right)^{-1} X' W_j^{(m+1)} y, \quad (4)$$

$$\pi_j^{(m+1)} = \frac{1}{k} \sum_{i=1}^{k} q_{ij}^{(m+1)},$$

$$\sigma^{2(m+1)} = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{d} q_{ij}^{(m+1)} (y_i - x_i' \beta_j^{(m+1)})^2,$$

3.2.The a Robust bi-square criterion method

The idea of estimating M is not an easy task to use to directly replace the target function in equation (3) with the criteria of the robust. Here, it is proposed to replace the criteria of least squares in equation (4) in the M step of Algorithm 1 with a robust criterion p.Therefore , $\beta_j^{(m+1)}$,j=1,…,d, is the solution.[3]

$$\sum_{i=1}^{k} q_{ij}{}^{(m+1)} x_i \omega \left(\frac{y_i - x_i' \beta_j}{\sigma^{(m)}} \right) = 0, \quad (5)$$

Here: $\omega(.) = p'(.)$ and $\sigma^{(m)}$is the Robust scale estimat of the ϵ_{ij} error
One of the p functions commonly used is the Huber function (Huber ^ 's ω-function) by Huber [21] where:

$$\omega_c(h) = p'(h) = \max\{-c, \min(c, h)\}$$

c: takes the value 1.345. which yields a relative efficiency of 95% when the error density represents a normal distribution.

Another possibility for a ω (.) is the function which is the Tukey function. It is called the Tukey's bisquare function.

$$\omega_c(h) = h\{1 - (1/c)^2\}_+^2$$

It weights the tail contribution of h by a biweigt function .
c: takes the value 4.685 and results in an efficiency ratio of 95%.
If we use L_1, the loss function p (h) = | h | We'll get the median regression. For more details, many researchers have dealt with them[18], [4],[1], [20] and [14]. We note that:

$$\sum_{i=1}^{k} q_{ij}^{(m+1)} x_i \omega\left(\frac{y_i - x_i'\beta_j}{\sigma^{(m)}}\right) \approx \sum_{i=1}^{k} q_{ij}^{(m+1)} x_i W\left(\frac{y_i - x_i'\beta_j}{\sigma^{(m)}}\right)\left(\frac{y_i - x_i'\beta^{(m)}{}_j}{\sigma^{(m)}}\right)$$

$$= \sum_{i=1}^{n} q_{ij}^{*(m+1)} x_i \omega\left(\frac{y_i - x_i'\beta_j}{\sigma^{(m)}}\right),$$

So that $W(h) = \Psi(h)/h$

$$q_{ij}^{*(m+1)} = q_{ij}^{(m+1)} W\left(\frac{y_i - x_i'\beta^{(m)}{}_j}{\sigma^{(m)}}\right).$$

Based on the above approximation, the solution in equation (5) can be approximated by

$$\beta_j^{(m+1)} = \left(\sum_{i=1}^{k} q_{ij}^{*(m+1)} x_i' x_i\right)^{-1} \sum_{i=1}^{k} q_{ij}^{*(m+1)} x_i y_i ,$$

one of the steps of the iterative reweighting algorithm. Where $\beta_j^{(m+1)}$ can be considered an estimation of the weighted least squares With weighters $q_{ij}^{*(m+1)}, i = 1, \dots, k.$

- **Algorithm 2:** Based on initial values of $\{\pi^{(0)}, \beta_j^{(0)}, \sigma^{(0)}, j = 1, \dots, d\}$. The proposed robust for the EM-type algorithm represents the repetition of both the E-step and the M-step.
  - **Step E:** The classification odds are calculated.[3]

$$q_{ij}^{(m+1)} = \frac{\pi_j^{(m)} \psi(y_i; \beta_j^{(m)} x_i', \sigma^{2(m)})}{\sum_{j=1}^{d} \pi_j^{(m)} \psi(y_i; \beta_j^{(m)} x_i', \sigma^{2(m)})}$$

  - **Step M: Update parameters for $\boldsymbol{\beta_j^{(m+1)}}$, $\boldsymbol{\pi_j^{(m+1)}}$, $\boldsymbol{\sigma^{2(m+1)}}$.[3]**

$$\beta_j^{(m+1)} = \left(\sum_{i=1}^{k} q_{ij}^{*(r+1)} x_i' x_i\right)^{-1} \sum_{i=1}^{k} q_{ij}^{*(m+1)} x_i y_i$$

$$= (X W_j^{*(m+1)} X')^{-1} X' W_j^{*(m+1)} y, \quad (6)$$

$$\pi_j^{(m+1)} = \frac{\sum_{i=1}^{k} q_{ij}^{(m+1)},}{k}$$

$$\sigma^{2(m+1)} = \frac{2\sum_{i=1}^{k}\sum_{j=1}^{d} q_{ij}^{(m+1)} (y_i - \beta_j^{(m+1)} x_i')^2 w_{ij}^{(m+1)}}{k} \quad (7)$$

Where $j = 1,2,\ldots,d$

$w_{ij}^{(m+1)}$ represents a diagonal matrix of degree k × k with diagonal elements $\left\{q_{ij}^{*(m+1)}, i = 1,2,..,k\right\}$ and

$$w_{ij}^{(m+1)} = \min\left[1 - \left\{1 - \left(\frac{y_i - \beta_j^{(m+1)} x_i'}{1.56\,\sigma^{(m)}}\right)^2\right\}^3, 1\right] \left(\frac{\sigma^{(m)}}{y_i - \beta_j^{(m+1)} x_i'}\right)^2$$

Equation (7) is the proposed robust scale estimate, which broadens the idea of the M (M-estimate) of a scale. We note that equation (7) is similar to the conventional non- robust scale estimate for mixture regression estimation of the adjustment factor "2", and the weighted $w_{ij}^{(m+1)}$ the bisquare weight. Some robust scale estimate can be applied to obtain the weighted $w_{ij}^{(m+1)}$.

The above proposed method can be easily extended to the non-uniformity of variance. The algorithm of robust EM types can be implemented over a constrained parameter space.

$$\Omega_c = \left\{\varphi \in \Omega: \frac{\sigma_r}{\sigma_j} \geq C > 0, 1 \leq r \neq j \leq d\right\} \quad (8)$$

Where c belong to(0,1] $\varphi = (\pi_1, \beta_1', \sigma_1, \ldots\ldots, \pi_{d-1}, \beta_{d-1}', \sigma_{d-1}, \beta_d', \sigma_r)'$ and Ω represents the unconstrained parameter space In equation (1), if x includes only the term intercept 1, the normal mixture model is uniform, So the robust estimation process can also be used to estimate the robust for location parameters.

Initial values: There are many ways to find the initial values of $\{\pi_j^{(0)}, \beta_j^{(0)}, \sigma^{(0)}, j = 1,\ldots, d\}$. One of the methods is to use the trimmed Likelihood estimate method (Neykov,et al.2007)[29]. it is robust to both low leverage and high leverage outliers under certain general conditions. Another possible method is to initially split the data or a subset of the data randomly into (d) groups. For each group, the robust regression method is used as an estimate MM (Yohai ,1987)[41] , To estimate the parameters of the component regression, a similar idea of segmentation was used to find the initial values of finite mixture models. In addition, the robust linear clustering method can be applied To find the values of the initial regression parameter Many researchers have covered this type of studies and from them [18],[19],[12]. However, this does not mean technically that the linear clustering methods do not produce consistent estimates of the regression component . However, in cases it is close enough to provide good initial values, because the proposed algorithm does not require consistent initial values.

Convergence of Algorithm 2: In the estimation equation (4), if we replace q_{ij} with, G_{ij} which is an index The laten component indicator with the proposed algorithm 2 can be considered As an Es algorithm which has been suggested by Elashoff and Ryan to estimate equations with missing

data. Therefore the proposed convergence characteristic of algorithm 2 can be demonstrated similarly to the ES algorithm of Elashoff and Ryan [9].

3.2.1. Asymptotic results

For the simplicity of the explanation and for proof, we assume that the measure parameter used in equation (5) constant Suppose $\varphi = (\beta_1', \dots, \beta_d', \pi_1, \dots, \pi_d)'$ and $\widehat{\varphi}_k$ The estimate found by algorithm 2 represents the proposed robust EM types. Note that the estimate $\widehat{\varphi}_k$ solves the following estimation equation[3]:

$$\sum_{i=1}^{k} q_{ij}(\varphi) x_i \omega\left(\frac{y_i - \beta_j x_i'}{\sigma}\right) = 0, \tag{9}$$

$$\pi_j = \sum_{i=1}^{k} \frac{q_{ij}(\varphi)}{k}, j = 1, \dots, d \tag{10}$$

Where :

$$q_{ij}(\varphi) = \frac{\pi_j \ \phi(y_i; \beta_j \ x_i', \sigma^2)}{\sum_{j=1}^{d} \pi_j \ \phi(y_i; \beta_j \ x_i', \sigma^2)} \tag{11}$$

Let $G_i = (x_i', y_i)'$ and

$$\Phi(z_i, \theta) = \{q_{i1} x_i \omega\left(\frac{y_i - B_1 x_i'}{\sigma}\right), \dots\dots, q_{id} x_i \omega\left(\frac{y_i - B_d x_i'}{\sigma}\right), q_{i1} - \pi_1, \dots, q_{i,d-1} - \pi_{d-1}\}' \tag{12}$$

$q_{ij} = q_{ij}(\varphi)$ It is defined in equation (11) and therefore the proposed estimate $\widehat{\varphi}_k$ solves the equation:

$$S_k(\varphi) = \frac{\sum_{i=1}^{k} \Phi(G_i, \varphi)}{k} = 0$$

3.3. Genetic Algorithm Via Gaussian Mixture and MM-Estimator (RobGA)

3.3.1. Mixture of normal distribution

The normal distribution is one of the most important probability distributions for its representation in many phenomena, and it is appropriate for them, whether the values used in the phenomena are very small and with small probabilities or large phenomena. Finite Mixture of Multivariate Normal Distribution which distributed $\mathcal{N}_n(\mu, \Sigma)$ with $F_k(y; \mu_d, \Sigma_d)$ the mixture density is given as follows[11]:

$$F(y, \varphi) = \ \pi_1 F_k(y; \mu_1, \Sigma_1) + \cdots + \ \pi_k F_k(y; \mu_d, \Sigma_d) \tag{13}$$

And if n=1 the Mixture of univariate normal distributions

$$g(y, \varphi) = \ \pi_1 g_k(y; \mu_1, \sigma_1^2) + \cdots + \ \pi_d g_k(y; \mu_d, \sigma_d^2) \tag{14}$$

As the mixed model is used extensively for continuous, univariate data $y_1, \dots, y_n$ is assumed to be independent observations and has a symmetric distribution i.i.d which satisfies the random variable Y For d the normal distributions are mixture univariate, the intensity of this distribution, which was mentioned in equation (14). The normal distributions mixture with univariate determinants in

general Teicher defines a model in terms of 3d-1 parameters and distinct model parameters $\varphi = (\mu_1, \dots . \mu_d, \sigma_1^2, \dots, \sigma_d^2, \pi_1, \dots, \pi_d)$ Normal mixture is the easiest extension to deal with continuous multivariate observations $y_1, \dots ., y_n$where $y_j$represents a vector of dimension u. the different elements are $y_{i1}, \dots, y_{iu}$ of y_i to measure u for the unit pulled from the population, a mixture model is frequently used for multivariate data, which assumes $y = (y_1, \dots, y_k)$representing observations that are i.i.d which validates a multivariate random variable Y with dimensions u it arises of d from multivariate normal distributions. Here the density is a multivariate normal distribution which is given in equation(13) [35].

In equation (13) $F_k(y; \mu_d, \Sigma_d)$ represents the density of a multivariate normal distribution with mean μ_d and the covariance-variance matrix Σ_d. Multivariate mixture of normal distributions is with covariance-covariance matrices $\Sigma_1, \dots, \Sigma_d$represent high parameters:

$$d(u + u(u + 1)/2 + 1) - 1$$

3.3.2. Parameter Estimation for Mixtures of Normal

Assuming that the data set y includes k of observations that are i.i.d of a random variable distributed according to the mixture of normal distributions and thus the univariate datay = $(\mathcal{Y}_1, \dots ., \mathcal{Y}_k)$, while the multivariate data is $y = (y_1, \dots, y_k)$.

The component parameters and weighted distribution $\pi = (\pi_1, \dots, \pi_d)$are estimated, the mixture distribution of the constant m is fundamental, and based on the y data, univariate mixture models include:

$\mu = (\mu_1, \dots, \mu_d)$: the averages of the component.
$\sigma^2 = (\sigma_1^2, \dots ., \sigma_k^2)$: represents the variances of the component.
The multivariate mixture models consist of:
$\mu = (\mu_1, \dots, \mu_d)$: the vectors of the mean of the component.
$\Sigma = (\Sigma_1, \dots, \Sigma_d)$: Covariance matrix - component covariance.

The estimation of the normal mixture models is based on the moment method by Charlier & Wicksell [6]. Here it is suggested the method of estimated moments by (Day 1969)[8] and found to be inefficient compared to estimating the worst univariate and multivariate potentials of normal mixture models.

The method of estimated moments has been suggested by Lindsay [25] and. Univariate by Lindsay & Basak [23] for the mixture nature of multivariate. The estimation of the greatest potential of two univariate normal distributions is used with $\sigma_1^2 = \sigma_2^2$. In 1970, Wolfe proposed a frequency diagram to calculate the maximum potential estimates of the mixture multivariate normal which is basically considered an alternative to the Expectation Maximization algorithm[11].

3.3.3. Expectation Maximization algorithm with Mixtures of Normal

The EM prediction maximization algorithm is an iterative method for finding the local maximum potential and applies this technique to any mixture models. To calculate the maximum likelihood estimate, univariate and multivariate mixture models.
Univariate mixture models are made by calculating step M[11]:

$$\mu_j^{(m)} = \frac{\sum_{i=1}^{k} q_{ij}^{(m)} x_j}{\sum_{i=1}^{k} q_{ij}^{(m)}}$$

$$(\sigma_j^2)^{(m)} = \frac{\sum_{i=1}^{k} q_{ij}^{(m)} (x_i - \mu_j^{(m)})^{m+1)}}{\sum_{i=1}^{k} q_{ij}^{(m)}}$$

Where$q_{ij}^{(m)}$ is defined in

$$q_{ij}^{(m)} = \frac{\pi_j^m \vartheta_j(x_i, \beta_j, \sigma_j^2)}{\sum_{j=1}^{d} \pi_j^m \vartheta_j(x_i, \beta_j, \sigma_j^2)}$$

The multivariate mixture models are calculated by calculating step M:

$$\mu_j^{(m)} = \frac{\sum_{i=1}^{k} q_{ij}^{(m)} x_j}{\sum_{i=1}^{k} q_{ij}^{(m)}}$$

$$\Sigma_j^{(m)} = \frac{\sum_{i=1}^{k} q_{ij}^{(m)} (x_i - \mu_j^{(m)})^T (x_i - \mu_j^{(m)})}{\sum_{i=1}^{k} q_{ij}^{(m)}}$$

Since there is a certain difficulty in estimating the Maximum likelihood that the Expectation Maximization algorithm has breaks down point when $(\sigma_j^2)^{(m)}$ is (numerically) zero or $\Sigma_j^{(m)}$is singular or nearly singular which occurs when $q_{ij}^{(m)}$ is near from zero or more observations in the next iteration $q_{ij}^{(m+1)}$ is calculated. These Difficulties arise in particular if the EM algorithm is applied to mixture normal Click Diagnosis (Overfitting) for the number of component[11].

3.3.4. MM-estimates

For the estimation, robust regression is the estimation of MM, which is the most general and most popular method of robust regression. Thus, a linear regression model is in eq.(1). The residuals ϵ_j represent identical independent distribution with zero mean and variance σ^2 under the central model. And the independent of the independent variables of x_i. The residuals is given through $s_i = s_i(\hat{\beta}) = y_i - \hat{\beta}x_i'$ with an estimate of M of regression, which is defined as follows:

$$\hat{\beta} = \text{argmin}\, \delta\left(\frac{s_j(\beta)}{\sigma}\right) \tag{15}$$

Where $s_i(\beta) = y_i - \beta x_i'$.
The function δ must have the following properties:
$\delta(s) = \delta(-s)$ Be symmetric.·
$\delta(0) = 0$·
$\delta(s)$ nondecreasing function of $|s|$.·
$\delta(s) > 0\ where \delta(s) < \delta(\infty)$. ·
If δ finite, then it is assumed that $\delta(\infty) = 1$

The estimate is determined by equation (15). It is called the estimator M descending M-estimate of regression. Here that ϖ is proportional to the derivative to δ′that is, $\varpi = \delta'$i.e. it is differentiable about the function δ. as a required scale to find the scale σ equation, it can be either external or real-time estimator. To reduce a function in equation (15), the derivative of a function with respect to the coefficients β is taken and the partial derivative is equalized to 0, wherein a system of equations for estimating the coefficients is obtained [2]

$$\sum_{i=1}^{k} \varpi\, x_i \left(\frac{s_i(\hat{\beta})}{\sigma}\right) = 0 \tag{16}$$

it is usually chosen $\varpi'(0) = 1$

$$\frac{1}{k}\sum_{i=1}^{k} X(\frac{\epsilon_i}{\sigma}) = a$$

a: Represent tuning constant and $X(\epsilon)$ It fulfills the same properties as it does δ .

- **Estimated-S**

To solve the lower breaking point of the M estimation, Rousseews and Leroy assume that the estimator S finds the least dispersion of the residuals. Here, S-estimates of regression are the parameter values $\hat{\beta}_r$ that minimize the M-estimate of scale $\hat{\sigma}_r = \hat{\sigma}_r\left(s(\hat{\beta}_r)\right)$ of the associated residuals,

$$\hat{\beta}_r = \text{argmin}\hat{\sigma}_r(s(\beta)).$$

The maximal breakdown point $((1-\delta)/2)$ which is score of S is reaching on use $a = (1-\delta/k)/2$ The standard MM estimate has three major design problems with a high $\delta/_k$ ratio:

1. Bias in r scale.
2. Loss of efficiency of the estimated parameters.
3. Failure to maintain test levels for the required value[22].

3.3.5. Genetic Algorithm

The genetic algorithm (GA) is an advanced idea of natural selection and genetics, as it represents an efficient and useful random search method, inspired by Darwin's theory of survival of the fittest. As the genetic algorithm is one of the important sections of the random search methods that are used to solve problems related to improvement, the genetic algorithm is considered a smart structure and in addition it is easy to implement.

The idea of a genetic algorithm (GA) is to randomly create some solutions to the problem, then examine these solutions and compare them with some criteria designed by the algorithm designer, The best solutions are the remaining ones. As for the less effective solutions, they are neglected according to the rule (survival of the fittest). The next step is to mix the remaining solutions (the most effective solutions) to produce new solutions similar to what happens in living organisms by mixing their genes as the resulting new organism will have its features from a mixture of characteristics it parents, and that these solutions result from intermarriage, which are also subject to examination and review to know the extent of its effectiveness and the extent of its proximity to the optimal solution, Thus the mating and selection processes are carried out until the process reaches a specified number of iterations (according to the algorithm's designer's discretion) or the resulting solutions or one of them reach a high efficiency ratio. Thus, genetic algorithms (GA) succeeded in generating hypotheses by replicating mutations and recombining most parts of existing hypotheses. The genetic algorithm can process any target function with or without constraints, whether linear or non-linear, and the solution space may be in one or more dimensions. As the working stages of the genetic algorithm are as follows:

1. **Start**: It consists in creating a random population of chromosomes, in the sense of obtaining possible solutions to the problem.
2. **Fitness:** It represents a measure of quality that can determine the best subset of genes and that each gene contains a specific meaning. It is found by converting the target function into a function suitable for solution in the genetic algorithm.

3. **New Population:** This stage is represented by the generation of a new individual by repeating the following stages until the individual is completed, which includes:

- **Selection**: Two parental chromosomes are selected from the primary population based on the matching function (which has the best match values).
- **Crossover**: Represented by the implementation of one of the crossing processes in order to obtain offspring and is between two chromosomes.
- **Mutation**: With the possibility of a mutation, the mutation of the new ancestor is carried out in a specific location on the chromosome and occurs between genes in a single chromosome.
- **Replacement**: Sons are replaced in the new population and become fathers in the later population.
- When the stop criterion is met, a test is performed, in this case the genetic algorithm is stopped and the good solution returned from the last generation.
- Return to stage 2, as each repetition of this process is called generation[31,42].

3.3.6. Applying RobGA

1. Using the mixture distribution based on EM algorithm, we classify the values of the response variable, which is assumed to be distributed normally, into Y_1, Y_2.
2. We specify the values of the $G_i$variable for each observation.

$$G_i = \{0,1\} \; \forall \; i$$

Depends on the result of applying mixture distribution.

3. We classify the X matrix based on the values of the variable G_i to be in the form X_1, X_2, which corresponds to the response variable Y_1, Y_2.
4. We apply the MM estimator to observations Y_2, X_2 and Y_1, X_1 to obtain $\hat{\beta}_1$, $\hat{\beta}_2$.
5. π_1 , π_2 are calculated according to the equations

$$\pi_1 = \frac{\sum_{i=1}^{k} G_{i1}}{k} \quad , \pi_2 = \frac{\sum_{i=1}^{k} G_i}{k},$$

Where:

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

And so we get $\hat{\pi}, \hat{\beta}_1, \hat{\beta}_2$.

6. The potentials are optimized using a genetic algorithm, with an appropriate initial value for each component.

4. Application

In order to study the performance of the two methods, we analyze a real data set were collected from the defective components of blood leading to blood diseases from Azadi Teaching Hospital in Kirkuk, the sample size (n = 176) was collected, which was three variables that included the response variable and the rest are independent variables, A comparison between the two methods for estimating the parameters is conducted. In addition to assessing the performance of the estimators were done in order to find out which of the methods is of high efficiency, depending on the criterion of Average Mean squares error (Wackerly, et. al., 2008)[37] It is symbolized by an abbreviation(AMSE) which represents the sum of the average squares of error for each observation divided by the frequencies R.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{k} \text{ , for } i = 1, 2, \dots, k$$

We use Classification Error (CE) which suggested by (Kassambara, 2017)[22] If the Classification Error is close to one, it means the Classification is appropriate.

$$CE = \frac{\text{Number of truely classified items in the sample}}{k}$$

Then in the real data were analyzed by relying on the programming language R . Where the variables included:
Y: Hemoglobin of blood.·
X_1: Blood Viscosity.·
X_2: Mean Corpuscular Hemoglobin(MCH).
At beginning, the normal distribution of the response variable was tested to see if the response variable was normally distributed.
Whereas:
Null hypothesis: The data is distributed normally.
Alternative hypothesis: Data is not distributed normally.
Where One Sample Kolmogorov-Smirnov Test used to see if the data undergo a normal distribution. Then if the significant value is greater than 0.05 (P-value> 0.05), then null hypothesis is accepted which states that the data are normally distributed and the alternative hypothesis is rejected[23].

mean	**sd**
8.6075668	**3.2997511**
One-sample Kolmogorov-Smirnov test	
p-value = 0.0008451<0.05	

We notice that the significant value of (0.0008451) is less than 5%, which means we accept the alternative hypothesis that says it is not distributed normally and we reject the null hypothesis. Then, the response variable Y data was classified into a number of compounds with a maximum number (5) and the Bayesian Information Criterion (BIC) was calculated[24]

$$BIC = -2 \times \log(L) + \nu \log k$$

Here :
log(L): Represents log-likelihood function that is used to estimate the model.
ν: is the number of parameters that are estimated in the model.

Table 1: Classification of response variable data Y and calculation of the Bayesian Information Criterion((BIC)

No. of Comp.	**1**	**2**	**3**	**4**	**5**
BIC	**930.0415**	**879.2192****	**891.5817**	**906.8432**	**919.2678**

It has been shown through Table1 that the lowest value in the Bayesian Information Criterion (BIC) was achieved when the number of components (2) was (879.2192), This means that the response variable data follow the mixture distribution with the number of components (2). Each variable was separately tested to see if it was normally distributed.

- **Component 1**

mean	**sd**
6.39374685	**1.28036381**
One-sample Kolmogorov-Smirnov test	
p-value = 0.9335>0.05	

We notice in the first component that the Significant value of (0.9335) is greater than 5%, and thus we accept the null hypothesis that says the data are normally distributed.

- **Component 2**

mean	**sd**
12.4817517	**1.8676443**
One-sample Kolmogorov-Smirnov test	
p-value = 0.7512>0.05	

In the second component, we notice that the significant value of (0.7512) is greater than 5%, and thus we accept the null hypothesis that says the data are normally distributed.

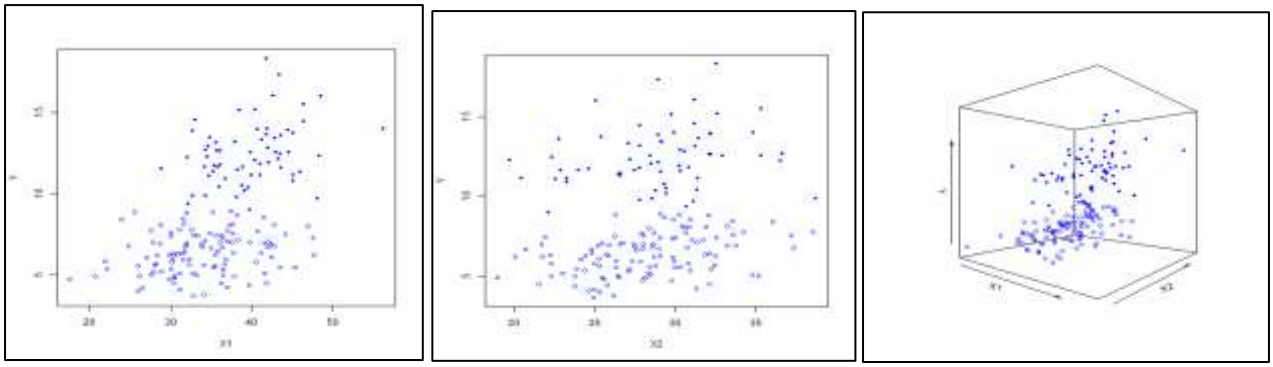

Figure 1: The graph above is for the real data after classifying it into two components.
Hollow circle: Represent component1.
Full circle: Represent component2.

The above figure shows the real data after classifying it into two components, which is the application of the mixed normal distribution to the Y response variable data, which resulted in two components. On the left side the figure represents the response variable with the independent variable X_1, the second figure is the response variable with the independent variableX_2, and the third figure stands for the response variable with the variablesX_1 and X_2

Table2: Estimation of the parameters of the two component mixture of linear regression model in the case of real data with sample size (n = 176)

	π_1	π_2	σ_1	σ_2	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	MSE	CE
Bi-square	0.497	0.502	1.014	2.168	-1.016	0.083	0.156	-4.298	0.312	0.122	7.031	0.929
RobGA	0.371	0.628	1.886	1.127	5.502	0.129	0.057	0.647	0.079	0.106	5.002	0.988

The above table represents the results of real data for estimating the parameters of the mixture of linear regression model with a sample size (n = 176), The results show that the proposed method (RobGA) has the lowest mean squares of error (MSE) of (5.002734) and thus is more effective than the other of the method.

MixBi **RobGA**

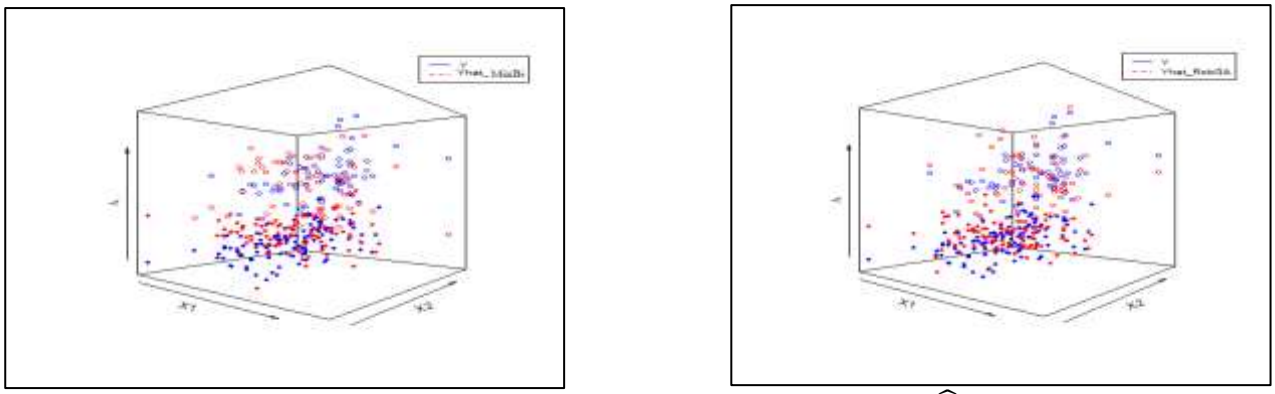

Figure 2: Represents the diffusion points of the real data for $Y_i$and $\widehat{Y}_i$ for the two real and estimated compounds with a sample size (n = 176).

: It represents the real values of the response variable Y_i.
: Is estimated values of the response variable $\widehat{Y}_i$.

Hollow circle: is component1.

Full circle: is component2.

component1: is Male.

Component2: is Female.

There is a scheme for the two methods used to estimate the parameters of the mixture of linear regression model (MixBi, RobGA) in the case of real data with a sample size (k = 176). The hollow circles for the estimated values of the response variable are relative of hollow circles to the true values of the response variable. As well as near full circle of values the Estimated response variable are from the filled circuits to the true values of the response variable. We observe from the above figures, each of the methods used is bi-square where the elements of the components are close together due to the fact that the amount of classification error is Few, this leads to difficulty in differentiating between two components. In (RobGA) mode it appears that the ease of differentiating between two components is more efficient than the bi-square method The amount of classification error is high. It was found that the RobGA method had given rating clear observations to their totals as shown in the drawing. It indicates how accurately the classification is how close the estimated values of $\widehat{Y}_i$ are to the true values of Y_i.

5. Conclusion

In this paper, two methods were compared to calculate the mixture of linear regression model. The results showed that the proposed method represents (RobGA) the best for estimating the mixture of linear regression model through dependence on the comparison benchmark, The average mean squares error(MSE) is smaller than value of the Classification Error (CE) (close to zero), This makes it more difficult it is to differentiate between the components. If the value of the CE is large (close to one), it becomes easy to differentiate between the two components.

The predicted model according to RobGA will be then:

$$\hat{y}_i = 0.371(5.502 + 0.129x_{i1} + 0.057\, x_{i2}) + 0.629(0.647 + 0.079\, x_{i1} + 0.106\, x_{i2})$$

The above estimated model can be used to predict the value of the response variable when we have new data.

References

[1] Andrews, D. F. ,(1974), "A robust method for multiple linear regression. Technometrics", Vol.16, No. 4, pp. 523-531.

[2] Alma, Ö. G., (2011), "Comparison of robust regression methods in linear regression", Int. J. Contemp. Math. Sciences,Vol. 6, No.9, pp. 409-421.

[3] Bai, X., Yao, W., & Boyer, J. E. ,(2012), "Robust fitting of mixture regression models. Computational Statistics & Data Analysis", Vol. 56, No. 7, pp. 2347-2359.

[4] Beaton, A. E., & Tukey, J. W. ,(1974), "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data", Technometrics,Vol.16, No. 2, pp. 147-185.

[5] Berrendero, J. R., Mendes, B. V., & Tyler, D. E. (2007). "On the maximum bias functions of MM-estimates and constrained M-estimates of regression". Annals of statistics,Vol. 35,No. 1, pp. 13-40.

[6] Charlier, C. V. L. and S. D. Wicksell, (1924), "On the dissection of frequency functions", Arkiv for Matematik, Astronomi och Fysik18, 1–64.

[7] Chen, J., Tan, X., Zhang, R., (2008). Inference for normal mixture in mean and variance. Statistica Sinica Vol. 18, No. 2, 443–465.

[8] Day, N. E., (1969), "Estimating the components of a mixture of normal distributions", Biometrika,Vol. 56, No. 3, pp. 463-474.

[9] Elashoff, M. and Ryan, L., (2004), "An EM Algorithm for Estimating Equations", Journal of Computational and Graphical Statistics, Vol. 13, No. 1, 48-65.

[10] Faria, S., & Soromenho, G., (2010), "Fitting mixtures of linear regressions", Journal of Statistical Computation and Simulation,Vol. 80, No. 2, pp. 201-225.
[11] Frühwirth-Schnatter, S.(2006), "Finite mixture and Markov switching models", Springer Series in Statistics.
[12] García-Escudero, L. A., Gordaliza, A., San Martin, R., Van Aelst , S., & Zamar, R. (2009), "Robust linear clustering", Journal of the Royal Statistical Society: Series B (Statistical Methodology),Vol.71, No.1, pp. 301-318.
[13] García-Escudero, L. A., Gordaliza, A., Mayo-Íscar, A., & San Martín, R. (2010), "Robust clusterwise linear regression through trimming", Computational Statistics & Data Analysis, Vol. 54, No. 12, pp. 3057-3069.
[14] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), Robust Statistics: The Approach Based on Influence Functions, Wiley, New York.
[15] Hathaway, R.J., 1985. "A constrained formulation of maximum-likelihood estimation for normal mixture distributions". Annals of Statistics 13, 795–800.
[16] Hathaway, R.J., 1986. "A constrained EM algorithm for univariate mixtures". Journal of Statistical Computation and Simulation 23, 211–230.
[17] Henning, C. (1995). Efficient high-breakdown-point estimators in robust regression: Which function to choose?. Statistics & Risk Modeling, 13(3), 221-242.
[18] Hennig, C. (2002). Fixed point clusters for linear regression: computation and comparison. Journal of Classification, Vol. 19, No. 2, 249-276.
[19] Hennig, C. (2003). "Clusters, Outliers, and Regression: Fixed Point Clusters". Journal of Multivariate Analysis, Vol. 86, No. 1, 183-212.
[20] Holland, P. W., & Welsch, R. E., (1977),"Robust regression using iteratively reweighted least-squares", Communications in Statistics-theory and Methods,Vol. 6, No.9, pp. 813-827.
[21] Huber, P. J., (1981), Robust Statistics, Wiley, New York.
[22] Kassambara, A. (2017), Machine Learning Essentials: Practical Guide in R, 1st Edition, STHDA. Publishing.
[23] Lilliefors, H. W. (1967), "On the Kolmogorov-Smirnov test for normality with mean and variance unknown", Journal of the American statistical Association, Vol.62, No. 318, pp. 399-402.
[24] Lindsay, B. G., & Basak, P., (1993)," Multivariate normal mixtures: a fast consistent method of moments", Journal of the American Statistical Association, Vol. 88, No. 422,pp. 468-476.
[25] Lindsay, B. G., (1989), "Moment matrices: Applications in mixtures", The Annals of Statistics Vol. 17, No. 2, 722–740.
[26] Markatou, M. (2000). "Mixture Models, Robustness, and the Weighted Likelihood Methodology". Biometrics, Vol. 56, No. 2, pp. 483-486.
[27] McLachlan, G.J., Peel, D. (2000), "Finite mixture models", New York.
[28] Mueller, C. H. and Garlipp, T. (2005). Simple consistent cluster methods based on redescending M-estimators with an application to edge identification in images. Journal of Multivariate Analysis Vol. 92, No. 2, 359-385.
[29] Neykov, N., Filzmoser, P., Dimova, R., and Neytchev, P., (2007), "Robust fitting of mixtures using the Trimmed Likelihood Estimator", Computational Statistics & Data Analysis, Vol. 52, No. 1, 299-308.
[30] P.N. Jones and G.J. McLachlan, (1992), "Fitting finite mixture models in a regression context", Aust. J. Statist. Vol. 34, No. 2, pp. 233–240.
[31] R. Quandt and J. Ramsey, (1978), "Estimating mixtures of normal distributions and switching regression", J. Amer. Statist. Assoc. Vol. 73, No. 364, pp. 730–738.
[32] Shahzad, W., Rehman, Q., & Ahmed, E. , (2017) , " Missing data imputation using genetic algorithm for supervised learning", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 8, No. 3.

[33] Shen, H., Yang, J., and Wang, S. (2004). Outlier Detecting in Fuzzy Switching Regression Models. Arti_cial Intelligence: Methodology, Systems, and Applications, Part of Lecture Notes in Computer Science Book Series, (LNAI Vol. 3192/2004), 208-215.

[34] Späth H (1979) Algorithm 39 clusterwise linear regression. Computing 22(4):367–373.

[35] Sperrin M, Jaki T, Wit E (2010) Probabilistic relabeling strategies for the label switching problem in Bayesian mixture models. Stat. Comput., Vol. 20, No. 3, 20:357–366.

[36] Teicher, H., (1963), "Identifiability of finite mixtures". The Annals of Mathematical Statistics Vol. 34, No. 4, 1265–1269.

[37] Turner, T. R., (2000)," Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions", Journal of the Royal Statistical Society: Series C (Applied Statistics),Vol. 49, No. 3,pp. 371-384.

[38] Wackerly, Dennis; Mendenhall, William; Scheaffer, Richard L. (2008). Mathematical Statistics with Applications, 7th Edition. Belmont, CA, USA: Thomson Higher Education. ISBN 978-0-495-38508-0.

[39] Wei, Y., (2012), "Robust mixture regression models using t-distribution", Master Report advisor, Dr. Yao, W. Department of Statistics, Kansas State University.

[40] Wolfe, J. H. (1970)," Pattern clustering by multivariate mixture analysis", Mul- tivariate Behavioral Research, Vol. 5, No. 3, 329–350.

[41] Yao, W., (2010). "A profile likelihood method for normal mixture with unequal variance". Journal of Statistical Planning and Inference Vol. 140, No. 7, 2089–2098.

[42] Yohai, V. J. ,(1987)," High Breakdown Point and High Efficiency Estimates for Regression". The Annals of Statistics, Vol. 15, No. 2, 642-656.

مجلة كلية الرافدين الجامعة للعلوم (2022)؛ العدد 51؛ 237- 252

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)
مجلة كلية الرافدين الجامعة للعلوم
Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain University College for Sciences

استخدام الخوارزمية الجينية في تقدير معالم الانحدار الخطي مختلط المحدود

ئورده ك ابراهيم كريم	د.فضاء مزهر هاشم
Iraqurdak124ibrahim@gmail.com	fali197477@gmail.com
قسم الإحصاء - كلية الإدارة والاقتصاد - جامعة بغداد، العراق	

المستخلص

لتقدير معالم الانحدار الخطي يتم الاستناد على طريقة المربعات الصغرى الاعتيادية (OLS) حيث يستند هذا الاسلوب في التقدير على عدة فرضيات الاساسية وعليه فان مدى دقة تقدير معالم النموذج في الواقع المطبق تتوقف على مدى صحة هذه الفرضيات فاذا كانت احدى هذه الفرضيات غير دقيقة يصبح استعمال النموذج أمراً غير واقعي وبالتالي يؤدي الى نتائج غير دقيقة ومن احدى هذه الفرضيات هي تجانس التباين والتوزيع الخطأ في بعض الاحيان لا يتم الاعتماد على مثل هذه الافتراضات في حالة دراسة مشكلة معينة قد تتضمن بيانات معقدة بما في ذلك بيانات مكونة اكثر من نموذج انحدار واحد وللتعامل مع هذا النوع من المشاكل يتم استعمال نموذج الانحدار الخطي المختلط في هذه المقالة نوضح الية التعامل في حالة حساسية تقدير المربعات الصغرى الاعتيادية اتجاه القيم المتطرفة حيث تم استعمال طريقتين لغرض تقدير المعلمات تضمنت طريقة معيار ثنائي التربيع الطريقة المقترحة تحسين مقدرات MM باستعمال الخوارزمية الجينية GA حيث تم تطبيقها على البيانات الحقيقية بهدف معرفة اي من الطريقتين ذات كفاءة عالية.

معلومات البحث

تواريخ البحث:
تاريخ تقديم البحث: 2021/4/4
تاريخ قبول البحث: 2021/5/6
تاريخ رفع البحث على الموقع: 2022/6/25

الكلمات المفتاحية:
المربعات الصغرى الاعتيادية، القيم الشاذة، الانحدار الخطي المختلط، معيار ثنائي التربيع الحصين، تحسين مقدرات MM باستعمال الخوارزمية الجينية GA.

للمراسلة:
ئورده ك ابراهيم كريم
Iraqurdak124ibrahim@gmail.com

DOI: https://doi.org/10.55562/jrucs.v51i1.536