Saudi Journal of Engineering and Technology (SJEAT)
Scholars Middle East Publishers
Dubai, United Arab Emirates
Website: http://scholarsmepub.com/

ISSN 2415-6272 (Print)
ISSN 2415-6264 (Online)

Exploring the Use of Hypothesis Testing in Determining the Number of Components in Gaussian Mixed Model

Mohd Izhan Mohd Yusoff[1*], Ibrahim Mohamed[2], Mohd. Rizam Abu Bakar[3]
[1]Institute of Mathematical Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
Telekom Research & Development Sdn Bhd., 63000 Cyberjaya, Selangor, Malaysia
[2]Institute of Mathematical Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
[3]Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul, Ehsan, Malaysia

Original Research Article

***Corresponding author**
Mohd Izhan Mohd Yusoff

Article History
Received: 12.12.2017
Accepted: 21.12.2017
Published: 30.12.2017

DOI:
10.21276/sjeat.2017.2.12.5

Abstract: Gaussian Mixed Model (GMM) has seen an increase in terms of usage especially to tackle the issue or problem of fraud activities in telecommunication industry. Like any other methods, GMM has its equal share of problems related to the maximum likelihood estimation and the determination of the number of components in GMM. In this paper we will highlight solutions to the said problems such as Expectation Maximization (EM) algorithm and the methods that are normally used to determine the number of components in GMM, which include the most recent research work done by the authors using Kernel method and Akaike Information Criteria (AIC); and the successful derivation of hypothesis testing in the determination of the number of components in GMM. The said derivation has never been attempted before due to the difficulty and complexity of GMM, as exemplified by the use of EM algorithm in solving its maximum likelihood estimation problem. The performance of the hypothesis testing, which is positive and promising despite using different percentage of overlapping; and the comparison of hypothesis testing to AIC, which produced conflicting results under certain conditions will also be highlighted.

Keywords: Expectation Maximization (EM), Gaussian Mixed Models (GMM), Kernel method, Akaike Information Criteria (AIC), Hypothesis testing.

INTRODUCTION

Gaussian Mixed Model (GMM) is best known in providing a robust speaker representation for the difficult task of speaker identification on *short-time speech spectra*, which is a cosine, transformed of log energy filter outputs from processed magnitude spectrum from a 20 ms short time segment of speech by simulated me-scale filter-bank [1]. Tanigushi *et al.*, [2] presented three approaches to fraud detection in communications networks [1]; they are Neural networks with supervised learning; Probability density estimation methods (GMM); and Bayesian networks. The performance of these methods was validated with data from a real mobile communications networks. The feature vectors used in this application describing the subscriber's behavior were based on toll tickets. For supervised learning approach, the features used were summary statistics over the whole observed time period as no times of fraud were recorded in the data. For the two latter approaches, the features described the daily behavior for every subscriber. To improve the fraud detection system, they recommended the combination of the three presented methods plus the incorporation of rule based systems.

[1] The Communications Fraud Control Association (cfca.org) periodically estimates the extent of worldwide telecommunications fraud. In 1999 this estimate was $12 billion, in 2003 it was between $35 and $ 40 billion, in 2006 it was between $55 and $60 billion, and in 2009 it was between $70 and $78 billion. Becker *et al.*, [3] gave examples of some common varieties of fraud in telecommunication and the one that is of our interest: Intrusion fraud. This occurs when an existing, otherwise legitimate account, typically a business, is compromised in some way by an intruder, who subsequently makes or sells calls on this account. In contrast to subscription calls, the legitimate calls may be interspersed with fraudulent calls, calling for an anomaly detection algorithm. Other examples can be found in Jacobs [4] and McClelland [5].

Likelihood function and log likelihood function for GMM are defined by $L(\mathbf{X} \mid \theta) = \prod_{j=1}^{n} f(\mathbf{x}_j \mid \theta)$ and $l(\mathbf{X} \mid \theta) = \log L(\mathbf{X} \mid \theta) = \sum_{j=1}^{n} \log \left(\sum_{i=1}^{K} a_i \phi(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right)$, respectively where $\mathbf{X} = \left(\mathbf{x}_1^t, \ldots, \mathbf{x}_n^t\right)^t$. Maximum likelihood estimation (m.l.e) aims at finding $\hat{\theta}$ that maximizes $l(\mathbf{X} \mid \theta)$; with respect to θ, see Mardia et al. ([1]979). The expression $\log \left(\sum_{i=1}^{K} a_i \phi(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right)$ in $l(\mathbf{X} \mid \theta)$ is difficult to solve and, to overcome this problem, we use Expectation Maximization (EM) algorithm. The algorithm was first introduced by Dempster *et al*., [6] and, since then, there's a significant increase of its usage especially in finding maximum likelihood for probabilistic models.

A number of authors highlighted the importance of identifying the right number of components in the GMM, say *k* components, and subsequently choosing good initial values of the model parameters μ_i and σ_i^2, $i = 1, 2, \ldots k$, in the EM algorithm. Schlattmann [7] noted on the difficulty to use log-likelihood-ratio statistics to test the number of components and subsequently suggested using a non-parametric bootstrapping approach. Similarly, Wang *et al.* [8] pointed the same concerns and introduced an algorithm called stepwise-split-and-merge EM algorithm to solve the said problem. In addition, Miloslavsky *et al*., [9] investigated the possibility of using the minimization of Kullback-Leiber distance between fitted mixture model and the true density as a method for estimating *k* where the said distance was estimated using cross validation. Zhuang *et al*., [10] viewed the mixture distribution as a contaminated Gaussian density and proposed a recursive algorithm called the Gaussian mixture density decomposition algorithm for identifying each Gaussian component in the mixture. The most recent research work in determining number of components in GMM was done by Yusoff *et al*., [11] via improvement of EM algorithm for GMM. The said improvement involves the following approach. Kernel method, Silverman [12] and

$$\hat{f}(t_k) \approx \sum_{l=-\frac{M}{2}}^{\frac{M}{2}} \exp\left(-\frac{2\pi kl}{M} i\right) \times \exp\left(-\frac{1}{2} h^2 \left(\frac{2\pi l}{\beta - \alpha}\right)^2\right) \times \left(\frac{1}{M} \sum_{k=0}^{M-1} \xi_k \exp\left(\frac{2\pi kl}{M} i\right)\right)$$

, is used in the first step to determine the number of components, say *K* components, and to find Means as initial values to start EM algorithm for GMM. EM algorithm for GMM is executed in the second step to find the final estimates of parameters using *k=1* number of components, Means obtained from the first step, and Variances fixed at 1 as initial values. Log likelihood function and Akaike Information Criteria (AIC) [13] are calculated in the third step using final estimates of parameters from the second step. The second and third steps are repeated in the fourth step for *k=2,….,K* number of components. All (*K*) AICs obtained from the fourth step are compared in the fifth step and the one that gives the minimum value is chosen (which gives the true or correct number of components). It has been shown that the modified EM algorithm has good performance upon investigation via simulation, though as expected; the performance depends on the percentage of overlapping of the Gaussian components [[2]]. Other works can also be found, see for example, Lee *et al*., [14] and Celeux *et al*., [15].

In this paper, we intend to use the approach that has never been attempted before that is hypothesis testing in determining the number of components of the GMM which is eventually to be used to identify fraud calls in telecommunication industry. It is organized as follows: Sections 2 presents the theory on GMM and EM algorithm to estimate the parameters of the GMM model. Sections 3 and 4 give the derivation of the said hypothesis testing, and comparison between using AIC and hypothesis testing in determining number of components in GMM, respectively. Concluding remarks are given in Section 5.

GAUSSION MIXED MODEL

GMM is defined by the summation of all *K* components where each component consists of prior probability (weight) *a* and probability density function with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. In other words,

[2] Yusoff *et al*., [11] proposed a new algorithm that has successfully not only detected fraud calls as suspected by the leading telecommunication company in Malaysia, but also identify suspicious calls which can be candidates of fraud call. The new fraud detection algorithm uses the above improvement of EM algorithm for GMM and can be incorporated into decision making or expert like system.

$$\sum_{i=1}^{K} a_i \phi(\mathbf{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \sum_{i=1}^{K} a_i \frac{1}{\sqrt{(2\pi)^d \mid \boldsymbol{\Sigma}_i \mid}} \exp\left(\frac{-(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)}{2} \right) \tag{1}$$

where prior probability (weight) of component i that is a_i satisfy the constraint $\sum_{i=1}^{K} a_i = 1$. The first interest is to estimate the parameters of model (1).

EM ALGORITHM

In a general set-up of EM algorithm as given in Dempster *et al.*, [6], we need to estimate the distribution of X, in sample space χ, but we can only observe X indirectly through Y, in sample space Y. In many cases, there is a mapping $x \to y(x)$ from χ to Y, and x is only known to lie in a subset of χ, denoted by $\chi(y)$, which is determined by the equation $y = y(x)$. The distribution of X is parameterized by a family of distributions $f(x \mid \theta)$, with parameters $\theta \in \Omega$ or x. The distribution of Y, $g(y \mid \theta)$ is

$$g(y \mid \theta) = \int_{\chi(y)} f(x \mid \theta) dx . \tag{2}$$

The EM algorithm aims at finding θ that maximizes equation (2), $g(y \mid \theta)$, given an observed y. Let the function

$$Q(\theta' \mid \theta) = E(\log f(x \mid \theta') \mid y, \theta) \tag{3}$$

be the expected value of $\log f(x \mid \theta')$ given y and θ. The expectation i.e. equation (3) is assumed to exist for all pairs (θ', θ). In particular, it is assumed that $f(x \mid \theta) > 0$ for $\theta \in \Omega$. According to Dempster *et al.*, [6], EM Iteration consists of two steps:

E-Step: Compute $Q\left(\theta \mid \theta^{(p)}\right)$

M-step: Choose $\theta^{(p+1)}$ to be a value of $\theta \in \Omega$ that maximizes $Q\left(\theta \mid \theta^{(p)}\right)$.

For the case of GMM, we define $Q(\theta' \mid \theta) = E\left[\log \prod_{i=1}^{n} a'_{y_i} \phi\left(\mathbf{x}_i \mid \boldsymbol{\mu}'_{y_i}, \boldsymbol{\Sigma}'_{y_i}\right) \middle| \mathbf{X}, \theta \right]$ where $y_i \in \{1,2,\ldots,K\}$, $y_i=k$ if the i^{th} sample was generated by the k^{th} mixture component. It is simplified, by using amongst others Bayes formula, that is $f(\theta \mid x) \propto f(x \mid \theta) P(\theta)$ where $f(\theta \mid x)$ = posterior probability, $f(x \mid \theta)$ = likelihood function, and $P(\theta)$ = prior probability [16, 17] to the following equations:

$$Q(\theta' \mid \theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} p_{i,k} \log a'_k + \sum_{i=1}^{n} \sum_{k=1}^{K} p_{i,k} \log \phi\left(\mathbf{x}_i \mid \boldsymbol{\mu}'_k, \boldsymbol{\Sigma}'_k\right) \tag{4}$$

where

$$p_{ik} = \frac{a_k \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_l a_l \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \tag{5}$$

and

$$\phi(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^d \mid \boldsymbol{\Sigma}_k \mid}} \exp\left(\frac{-(\mathbf{x}_i-\boldsymbol{\mu}_k)^t \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i-\boldsymbol{\mu}_k)}{2} \right). \tag{6}$$

Hence, EM Iteration for GMM is defined by:

E-Step: Equation (5) is calculated (involving equation (6)).

M-Step: The following formulas (derived from Lagrange multipliers, $\frac{\partial Q}{\partial \boldsymbol{\mu}_j} = 0$ and $\frac{\partial Q}{\partial \boldsymbol{\Sigma}_j^{-1}} = 0$, respectively where Q here refers to equation (4)) are calculated:

$$a_j = \frac{1}{n}\sum_i p_{ij} \tag{7}$$

$$\boldsymbol{\mu}_j = \frac{\sum_i p_{ij}\mathbf{x}_i}{\sum_i p_{ij}} \tag{8}$$

$$\boldsymbol{\Sigma}_j = \frac{\sum_i p_{ij}\left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)\left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)^t}{\sum_i p_{ij}}. \tag{9}$$

The above steps (i.e. E-step and M-step involving equations (5) till (9)) are repeated until convergence is met (or achieved).

THE DERIVATION OF THE HYPOTHESIS TESTING

In the previous section, we mentioned improvement of EM algorithm for GMM, which is the most recent research work done by Yusoff *et al.*, [18], where steps 3, 4, and 5 are involved in the determination of the number of components (using Akaike Information Criteria, AIC). The AIC used in the above steps could be replaced by the following hypothesis testing:

$$H_0 : \boldsymbol{\theta} = \left(\begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 \\ \vdots \\ \boldsymbol{\Sigma}_k \end{bmatrix}\right) \text{ versus } H_1 : \boldsymbol{\theta}^* = \left(\begin{bmatrix} a_1^* \\ \vdots \\ a_{k^*}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_1^* \\ \vdots \\ \boldsymbol{\mu}_{k^*}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1^* \\ \vdots \\ \boldsymbol{\Sigma}_{k^*}^* \end{bmatrix}\right)$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ are final estimates of parameters obtained from Step 2, k and k^* are number of parameters where $k,k^*=1,2,\ldots,K$ (and preferably $k^*>k$).

The likelihood ratio statistics for testing the above hypothesis is defined by

$$-2\log\lambda = 2\sum_{j=1}^{n}\log\left(\frac{\sum_{i=1}^{k^*} a_i^*\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*\right)}{\sum_{i=1}^{k} a_i\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}\right),$$

see Mardia *et al.*, [19]. It could be written as

$$-2\log\lambda = 2\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k^*} a_i^* \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i^*|}}\exp\left(-\frac{\left(\mathbf{x}_j - \boldsymbol{\mu}_i^*\right)^t \boldsymbol{\Sigma}_i^{*-1}\left(\mathbf{x}_j - \boldsymbol{\mu}_i^*\right)}{2}\right)\right)$$

$$-2\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i|}}\boldsymbol{exp}\left(-\frac{\left(\mathbf{x}_j - \boldsymbol{\mu}_i\right)^t \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}_j - \boldsymbol{\mu}_i\right)}{2}\right)\right). \tag{10}$$

We apply $\log\left(\sum_{i=1}^{h}\phi_i\right) < \log\left(h\left(\prod_{i=1}^{h}(\phi_i)\right)^{-1}\right)$ where $0 < \phi_i < 1, i = 1,2,\ldots, h$ (denoted as Prop1, for details refer to **APPENDIX A**) and $\sum_{i=1}^{k*} a_i^* f_i^* < 1$ [3] to the first term of equation (10) yielding

$$\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k*} a_i^* \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i^*|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i^*)^t\boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j-\boldsymbol{\mu}_i^*)}{2}\right)\right)$$
$$< \sum_{j=1}^{n}\log\left(\left(\prod_{i=1}^{k*}\left(a_i^* \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i^*|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i^*)^t\boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j-\boldsymbol{\mu}_i^*)}{2}\right)\right)\right)^{-1}\right). \qquad (11)$$

The right hand side of equation (11) could be written as

$$\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k*} a_i^* \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i^*|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i^*)^t\boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j-\boldsymbol{\mu}_i^*)}{2}\right)\right)$$
$$< -\sum_{j=1}^{n}\sum_{i=1}^{k*}\left(\log\left(a_i^*\frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i^*|}}\right)\right) + \sum_{j=1}^{n}\sum_{i=1}^{k*}\left(\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i^*)^t\boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j-\boldsymbol{\mu}_i^*)}{2}\right). \qquad (12)$$

We apply $\log\left[\sum_{i=1}^{h}\phi_i\right] > \log\left[\prod_{i=1}^{h}\phi_i\right]$ where $0 < \phi_i < 1, i = 1,2,\ldots, h$ (denoted by Prop2, for details refer to **APPENDIX B**) to the second term of equation (10) yielding

$$\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)}{2}\right)\right)$$
$$> \sum_{j=1}^{n}\log\left(\prod_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)}{2}\right)\right). \qquad (13)$$

The right-hand-side of equation (13) could be written as

$$\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)}{2}\right)\right)$$

[3] Using $0 < a_i^* < 1$, $0 < f_i^* < 1$, $0 < a_i^* f_i^* < a_i^*$ where $i = 1,2,\ldots, k*$ and $\sum_{i=1}^{k*} a_i^* = 1$, we get $\sum_{i=1}^{k*} a_i^* f_i^* < 1$

$$> \sum_{j=1}^{n}\sum_{i=1}^{k}\left(\log\left(a_i \frac{1}{\sqrt{(2\pi)^p \,|\boldsymbol{\Sigma}_i|}}\right)\right) - \sum_{j=1}^{n}\sum_{i=1}^{k}\left(\frac{\left(\mathbf{x}_j-\boldsymbol{\mu}_i\right)^t \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}_j-\boldsymbol{\mu}_i\right)}{2}\right). \quad (14)$$

Applying minus sign to both sides of equation (14), we get

$$-\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p \,|\boldsymbol{\Sigma}_i|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)}{2}\right)\right)$$

$$< -\sum_{j=1}^{n}\sum_{i=1}^{k}\left(\log\left(a_i \frac{1}{\sqrt{(2\pi)^p \,|\boldsymbol{\Sigma}_i|}}\right)\right) + \sum_{j=1}^{n}\sum_{i=1}^{k}\left(\frac{\left(\mathbf{x}_j-\boldsymbol{\mu}_i\right)^t \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}_j-\boldsymbol{\mu}_i\right)}{2}\right). \quad (15)$$

Note that $\sum_{j=1}^{n}\sum_{L=1}^{k*}\left(\mathbf{x}_j-\boldsymbol{\mu}_L^*\right)^t \boldsymbol{\Sigma}_L^{*-1}\left(\mathbf{x}_j-\boldsymbol{\mu}_L^*\right)$ of equation (12) and $\sum_{j=1}^{n}\sum_{L=1}^{k}\left(\mathbf{x}_j-\boldsymbol{\mu}_L\right)^t \boldsymbol{\Sigma}_L^{-1}\left(\mathbf{x}_j-\boldsymbol{\mu}_L\right)$ of equation (15) are greater than $\sum_{L=1}^{k*}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L^*\right)^t \boldsymbol{\Sigma}_L^{*-1}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L^*\right)$ and $\sum_{L=1}^{k}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)^t \boldsymbol{\Sigma}_L^{-1}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)$, respectively. $\sum_{L=1}^{k*}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L^*\right)^t \boldsymbol{\Sigma}_L^{*-1}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L^*\right)$ and $\sum_{L=1}^{k}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)^t \boldsymbol{\Sigma}_L^{-1}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)$ follow χ^2 distribution with *k*p* and *kp* degrees of freedom, respectively (an example is given in **APPENDIX C**). They are derived from the following property Mardia *et al*., [19]:

$$\sum_{i=1}^{n}\left(\mathbf{x}_i-\boldsymbol{\mu}_L\right)^t \boldsymbol{\Sigma}_L^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}_L\right) = ntr\left(\boldsymbol{\Sigma}_L^{-1}\mathbf{S}_L\right) + n\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)^t \boldsymbol{\Sigma}_L^{-1}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)$$

where $n\mathbf{S}_L = \sum_{i=1}^{n}\left(\mathbf{x}_i-\overline{\mathbf{x}}_L\right)\left(\mathbf{x}_i-\overline{\mathbf{x}}_L\right)^t$ and $\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right)^t \boldsymbol{\Sigma}_L^{-1}\left(\overline{\mathbf{x}}_L-\boldsymbol{\mu}_L\right) \sim \chi_p^2$. The lower limit (it is found by Prop1 swapping places with Prop2 that is Prop 2 and Prop 1 are applied to first and second term of equation (10), respectively) has the same distribution as the upper limit hence $-2\log\lambda = 2\sum_{j=1}^{n}\log\left(\frac{\sum_{i=1}^{k^*} a_i^*\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*\right)}{\sum_{i=1}^{k} a_i\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}\right)$ satisfy χ^2 distribution with $(k^*+k)p$ degrees of freedom. H_0 is accepted when $-2\log\lambda = 2\sum_{j=1}^{n}\log\left(\frac{\sum_{i=1}^{k^*} a_i^*\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*\right)}{\sum_{i=1}^{k} a_i\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}\right)$ is less than $\chi^2_{\alpha,(k+k^*)p}$.

If $-2\log\lambda = 2\sum_{j=1}^{n}\log\left(\dfrac{\sum_{i=1}^{k^*} a_i^*\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*\right)}{\sum_{i=1}^{k} a_i\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}\right) < 0$ due to the following properties $\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k} a_i\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)\right) < 0$ and $\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k*} a_i^*\phi\left(\mathbf{x}_j \mid \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*\right)\right) < 0$, we swap places between $\boldsymbol{\theta} = \left(\begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 \\ \vdots \\ \boldsymbol{\Sigma}_k \end{bmatrix}\right)$ and $\boldsymbol{\theta}^* = \left(\begin{bmatrix} a_1^* \\ \vdots \\ a_{k^*}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_1^* \\ \vdots \\ \boldsymbol{\mu}_{k^*}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1^* \\ \vdots \\ \boldsymbol{\Sigma}_{k^*}^* \end{bmatrix}\right)$ and test $H_0 : \boldsymbol{\theta}^* = \left(\begin{bmatrix} a_1^* \\ \vdots \\ a_{k^*}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_1^* \\ \vdots \\ \boldsymbol{\mu}_{k^*}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1^* \\ \vdots \\ \boldsymbol{\Sigma}_{k^*}^* \end{bmatrix}\right)$ versus $H_1 : \boldsymbol{\theta} = \left(\begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 \\ \vdots \\ \boldsymbol{\Sigma}_k \end{bmatrix}\right)$ where H_0 is accepted when $-2\log\lambda$ (which is now greater than 0) is less than $\chi^2_{\alpha,(k+k^*)p}$.

THE PERFORMANCE OF THE HYPOTHESIS TESTING

The performance of the hypothesis testing is described by using Box and Muller Transformation [20] to generate simulation data for $\sum_{i=1}^{K} a_i\phi(\mathbf{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \sum_{i=1}^{K} a_i \dfrac{1}{\sqrt{(2\pi)^d \mid \boldsymbol{\Sigma}_i \mid}} \exp\left(\dfrac{-(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}{2}\right)$ focusing on two components with the following properties: a_1 is chosen from 0.1,…,0.9, a_2 is derived from the following formula $a_2 = 1 - a_1$, μ_1 is fixed at 0.0; μ_2 is chosen from 0.25,0.50,0.75,…,3.0, σ_{11} and σ_{22} are chosen from $\frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{10}, 1, 2, 3, \cdots, 10$. A total of 25 samples, each with 1000 observations, are generated for each model.

Note that

$$Percentage = \left(\frac{\text{Number of hypothesis testings that accepts 2 number of components}}{\text{Total number of hypothesis testings}}\right)\%$$

is calculated and *Range* is assigned to each model where Range equals to 1 represents percentage of overlapping between 0% and 25%, 2 represents percentage of overlapping between 25% and 50%, 3 represents percentage of overlapping between 50% and 75% and 4 represents percentage of overlapping between 75% and 100% [18].

Example of an output is given in Figure-1, which represents ($a_1 = 0.2$, $\mu_1 = 0.0$, $\sigma_1^2 = (0.707)^2$) and ($a_2 = 0.8$, $\mu_2 = 0.25$, $\sigma_2^2 = 1.0$).The Range for the given example equals to 3 and H_0 is accepted when $k^* = 3$ where p is greater than 0.05. Note that $-2\log\lambda$ is given in the first bracket and probability value p is given in the second.

Akaike Information Criteria

no_of_components	AIC	Min
1	2759.241221463931	2759.241221463931
2	2742.695779311172	2742.695779311172
3	2745.7438644121394	2742.695779311172

Hypothesis Testing

H_0:theta,k=1
H_0:theta,k=1 versus H_1:theta*,k*=2 (22.54544215275928)(p=5.028078782631841E-5)
H_0:theta,k=1 versus H_1:theta*,k*=3 (25.497357051791823)(p=3.9955836008321434E-5)

H_0:theta,k=2
H_0:theta,k=2 versus H_1:theta*,k*=3 (2.951914899032545)(p=0.7073991861415829)

Fig-1: Results of AIC and hypothesis testing for $\sum_{i=1}^{2} a_i \phi\left(x, \mu_i, \sigma_i^2\right)$

where ($a_1 = 0.2$ ***,*** $\mu_1 = 0.0$ ***,*** $\sigma_1^2 = (0.707)^2$ ***) and (*** $a_2 = 0.8$ ***,*** $\mu_2 = 0.25$ ***,*** $\sigma_2^2 = 1.0$ ***).***

The results are displayed in Figure-2 where the values used can be found in Table-1. The value under the column titled "(Freq/Total)%" of Table-1 that corresponds to Percentage equals to 100 decreases not lower than 50 as the Range increases.

COMPARISON BETWEEN USING AIC AND HYPOTHESIS TESTING IN DETERMINING NUMBER OF COMPONENTS IN GMM

Akaike Information Criteria (AIC) used in the improvement of EM algorithm for GMM developed by Yusoff *et al.*, [18] is defined by $AIC = 2\,pmtr - 2\,Log(L)$ where $pmtr$ is the number of parameters and $Log(L)$ is the maximized log-likelihood function. Let

$$AIC_k = 2\Omega k - 2\sum_{j=1}^{n} \log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p |\mathbf{\Sigma}_i|}} \exp\left(-\frac{(\mathbf{x}_j - \boldsymbol{\mu}_i)^t \mathbf{\Sigma}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i)}{2}\right)\right) \tag{16}$$

and

$$AIC_{k*} = 2\Omega k * -2\sum_{j=1}^{n} \log\left(\sum_{i=1}^{k^*} a_i^* \frac{1}{\sqrt{(2\pi)^p |\mathbf{\Sigma}_i^*|}} \exp\left(-\frac{(\mathbf{x}_j - \boldsymbol{\mu}_i^*)^t \mathbf{\Sigma}_i^{*-1} (\mathbf{x}_j - \boldsymbol{\mu}_i^*)}{2}\right)\right) \tag{17}$$

where the second term of AIC_k and AIC_{k*} is taken from equation (10), and $2\Omega = 6$ if $p = 1$. In this section, we present two cases. They are:

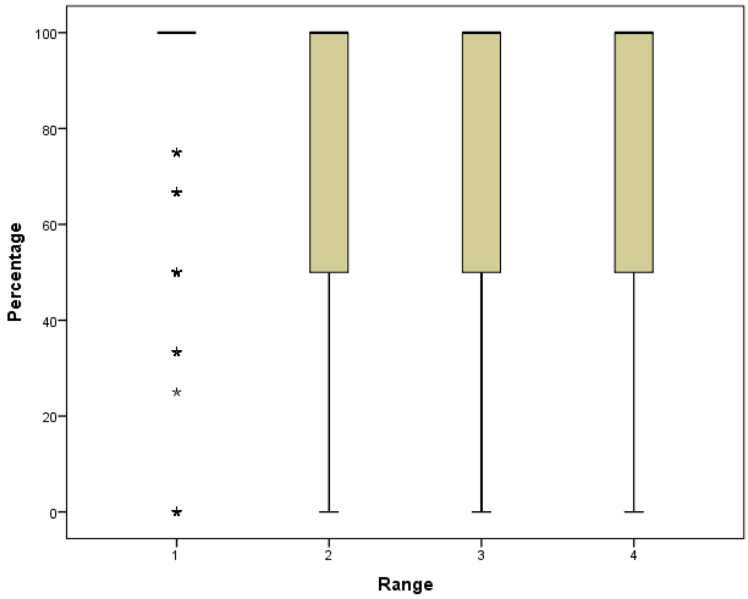

Fig-2: Percentage is plotted against Range in the Box plot

Table-1: Frequency table for Range equals to (a) 1, (b) 2, (c) 3 and (d) 4

		Frequency	(Freq/Total)%
Valid	0	21	1.2
	25	1	0.1
	33.3	24	1.4
	50	197	11.6
	66.7	131	7.7
	75	15	0.9
	100	1313	77.1
	Total	1702	100

(a)

		Frequency	(Freq/Total)%
Valid	0	29	4.9
	33.3	12	2
	50	121	20.3
	66.7	37	6.2
	75	7	1.2
	100	391	65.5
	Total	597	100

(b)

		Frequency	(Freq/Total)%
Valid	0	13	9.5
	50	36	26.3
	66.7	10	7.3
	75	1	0.7
	100	77	56.2
	Total	137	100

(c)

		Frequency	(Freq/Total)%
Valid	0	6	20.7
	50	6	20.7
	66.7	2	6.9
	100	15	51.7
	Total	29	100

(d)

Case 1: Let $AIC_{k*} > AIC_k$ (according to Step 5, AIC_k is minimum therefore it is chosen) where $k^*>k$. Using equations (16) and (17), we get

$$2\Omega k^* - 2\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k^*} a_i^* \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i^*|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i^*)^t\,\boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j-\boldsymbol{\mu}_i^*)}{2}\right)\right)$$

$$> 2\Omega k - 2\sum_{j=1}^{n}\log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p\,|\boldsymbol{\Sigma}_i|}}\exp\left(-\frac{(\mathbf{x}_j-\boldsymbol{\mu}_i)^t\,\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)}{2}\right)\right) \qquad (18)$$

Equation (18) could be written as

$$2\Omega(k^* - k) > \beta \tag{19}$$

where

$$\beta = 2\sum_{j=1}^{n} \log\left(\sum_{i=1}^{k^*} a_i^* \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}_i^*|}} \exp\left(-\frac{(\mathbf{x}_j - \boldsymbol{\mu}_i^*)^t \boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j - \boldsymbol{\mu}_i^*)}{2}\right)\right)$$
$$- 2\sum_{j=1}^{n} \log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{(\mathbf{x}_j - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_i)}{2}\right)\right)$$

Let $\gamma \sim \chi^2_{\alpha,(k+k^*)p}$. If $2\Omega(k^* - k) > \beta > \gamma$, we reject H_0, contradicting the AIC results. If $2\Omega(k^* - k) > \beta$ and $\gamma > \beta$, we accept H_0.

Case 2: Let $AIC_{k^*} < AIC_k$ (according to Step 5, AIC_{k^*} is minimum therefore it is chosen) where *k**>*k*. By repeating the process in Case 1, that is using equations (16) and (17), we get

$$2\Omega(k^* - k) < \beta \tag{20}$$

where

$$\beta = 2\sum_{j=1}^{n} \log\left(\sum_{i=1}^{k^*} a_i^* \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}_i^*|}} \exp\left(-\frac{(\mathbf{x}_j - \boldsymbol{\mu}_i^*)^t \boldsymbol{\Sigma}_i^{*-1}(\mathbf{x}_j - \boldsymbol{\mu}_i^*)}{2}\right)\right)$$
$$- 2\sum_{j=1}^{n} \log\left(\sum_{i=1}^{k} a_i \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{(\mathbf{x}_j - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_i)}{2}\right)\right)$$

Let $\gamma \sim \chi^2_{\alpha,(k+k^*)p}$. If $2\Omega(k^* - k) < \beta$ and $\beta > \gamma$, we reject H_0. If $2\Omega(k^* - k) < \beta < \gamma$, we accept H_0, contradicting the AIC results.

REAL TELECOMMUNICATION DATA

The call detail record, which was supplied by Telekom Malaysia Berhad (henceforth, TM), consists of calls made by customers that fell victim to fraud activities. We performed several steps on the original data in order to have the data in a desired format i.e. group the real data according to Service No, find the country that matches with the country code as well as dialed digits and sort the real data according to Seize Time. Seize Time refers to the time when the call was made, the duration of the calls is in the following format: hour (hh), minute (mm) and second (ss), and the customers are labeled A, B, C,… to ensure confidentiality.

EXAMPLE

Figure-3, produced from using real telecommunication data as mentioned in the previous section, shows hypothesis testing results support those of AIC (note that $-2\log\lambda$ is given in the first bracket whereas probability value is given in the second). Case 1 and 2 can be found in the hypothesis testing when ($k = 2$, $k^* = 3$) and ($k = 1$, $k^* = 2,3$), respectively. The probability value α is fixed at 0.05. Figure-4 shows, especially when k and k^* equal to 2 and 3, respectively, hypothesis testing results do not support those of AIC. Case 2 can be found in all of the hypothesis testing.

CONCLUSION

In the previous sections, we gave a brief introduction to GMM and EM algorithm; and we showed the effects of fraud activities to telecommunication industry. We also mentioned when would we determine the number of components in GMM and gave several examples that are normally used in the determination of the number of components in GMM, including the use of AIC in the determination process.

We successfully derived hypothesis testing, which we believe could be used as an alternative method to AIC in the determination of the number of components in GMM. The performance of the hypothesis testing is generally positive although different percentage of overlapping was used.

Next we compared hypothesis testing to AIC using mathematical derivation and real telecommunication data where AIC and hypothesis testing produced different (or conflicting) results under certain conditions. Hypothesis testing results depend on log-likelihood function and the choice of the probability value α that gives $\chi^2_{\alpha,(k+k^*)p}$. AIC results on the other hand depend on log-likelihood function only as shown in Figure-5 and 6. Hypothesis testing results are similar to those of AIC if α is set at different value (i.e. other than 0.05).

Further research on the behavior of the hypothesis testing especially when it conflicts with AIC is required that will include the use of the power of a test.

APPENDIX A

The formula $0 < \left(\prod_{i=1}^{n} \phi_i\right)\left(\sum_{j=1}^{n} \phi_j\right) < n$ is derived from using $0 < \left(\prod_{i=1}^{n} \phi_i\right) < 1$, $0 < \left(\sum_{j=1}^{n} \phi_j\right) < n$ and $0 < \left(\prod_{i=1}^{n} \phi_i\right)\left(\sum_{j=1}^{n} \phi_j\right) < \left(\sum_{j=1}^{n} \phi_j\right) < n$. Apply logarithm to $0 < \left(\prod_{i=1}^{n} \phi_i\right)\left(\sum_{j=1}^{n} \phi_j\right) < n$, we get $\log\left(\left(\prod_{i=1}^{n} \phi_i\right)\left(\sum_{j=1}^{n} \phi_j\right)\right) < \log(n)$ or $\log\left(\left(\sum_{j=1}^{n} \phi_j\right)\right) < \log(n) - \log\left(\left(\prod_{i=1}^{n} \phi_i\right)\right)$. An example is given in Figure A.1.

Akaike Information Criteria

no_of_components	AIC	Min
1	92.957704551177	92.957704551177
2	78.93876969062063	78.93876969062063
3	80.78840789298813	78.93876969062063

Hypothesis Testing

H_0:theta,k=1
H_0:theta,k=1 versus H_1:theta*,k*=2 (20.01893486055637)(Prob=1.6832767836781848E-4)
H_0:theta,k=1 versus H_1:theta*,k*=3 (24.16929665818887)(Prob=7.386963782137005E-5)

H_0:theta,k=2
H_0:theta,k=2 versus H_1:theta*,k*=3 (4.150361797632499)(Prob=0.5279773066298088)

Fig-3: Results of AIC and hypothesis testing for Customer C

Akaike Information Criteria

no_of_components	AIC	Min
1	47.53326292383591	47.53326292383591
2	35.70610331847866	35.70610331847866
3	33.64022457875683	33.64022457875683

Hypothesis Testing

H_0:theta,k=1
H_0:theta,k=1 versus H_1:theta*,k*=2 (17.827159605357252)(Prob=4.7764602611496796E-4)
H_0:theta,k=1 versus H_1:theta*,k*=3 (25.89303834507908)(Prob=3.325553903926139E-5)

H_0:theta,k=2
H_0:theta,k=2 versus H_1:theta*,k*=3 (8.065878739721828)(Prob=0.15264184961471652)

Fig-4: Results of AIC and hypothesis testing for Customer D

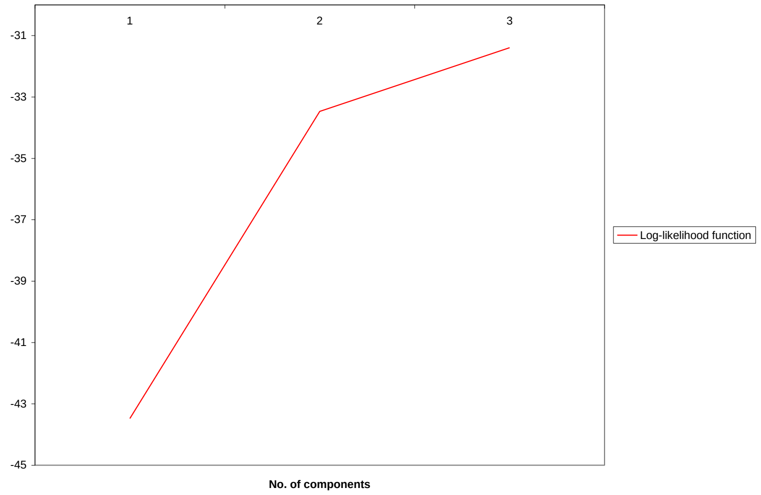

Fig-5: Log-likelihood function against number of components for Customer C

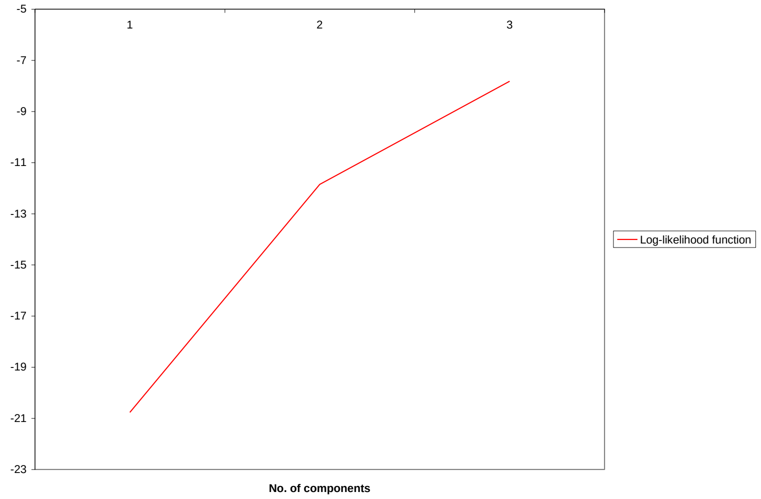

Fig-6: Log-likelihood function against number of components for Customer D

APPENDIX B

The formula $\left(\frac{n}{n}\right)\left(\prod_{i=1}^{n}\phi_i\right) < \left(\sum_{j=1}^{n}\phi_j\right)$ or $\left(\prod_{i=1}^{n}\phi_i\right) < \left(\sum_{j=1}^{n}\phi_j\right)$ is derived from using $0 < \left(\prod_{i=1}^{n}\phi_i\right) < 1$, $0 < \left(\sum_{j=1}^{n}\phi_j\right) < n$ and $\left(\frac{1}{n}\right)\left(\prod_{i=1}^{n}\phi_i\right) < \left(\prod_{i=1}^{n}\phi_i\right) < \phi_j$ or $\left(\frac{1}{n}\right)\left(\prod_{i=1}^{n}\phi_i\right) < \phi_j$ where $j = 1,2,\ldots,n$. An example is given in Figure B.1.

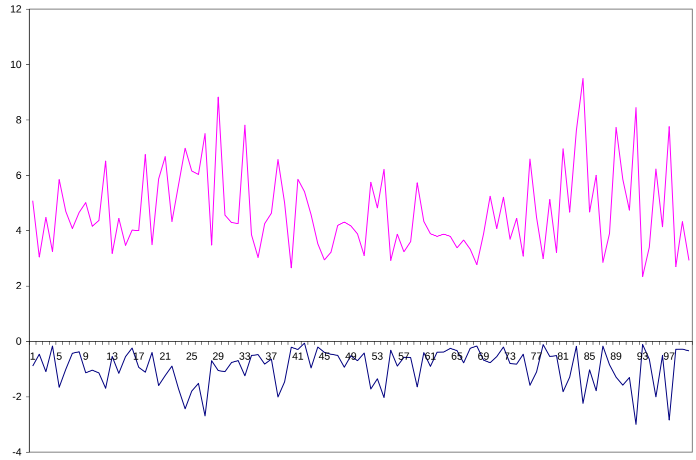

Fig-A.1: The first (i.e. upper) and second (i.e. lower) lines represent $\log(2) - \log\left(\left(\prod_{i=1}^{2} \phi_i\right)\right)$ **and** $\log\left(\left(\sum_{j=1}^{2} \phi_j\right)\right)$**, respectively. X-axis represents 100 samples (generated by using random numbers)**

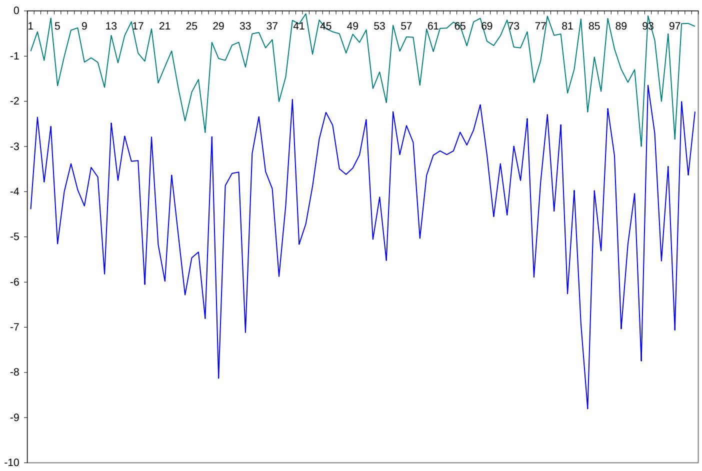

Fig-B.1: The first (i.e. upper) and second (i.e. lower) lines represent $\log\left(\left(\sum_{j=1}^{2} \phi_j\right)\right)$ **and** $\log\left(\left(\prod_{i=1}^{2} \phi_i\right)\right)$**, respectively. X-axis represents 100 samples (generated by using random numbers).**

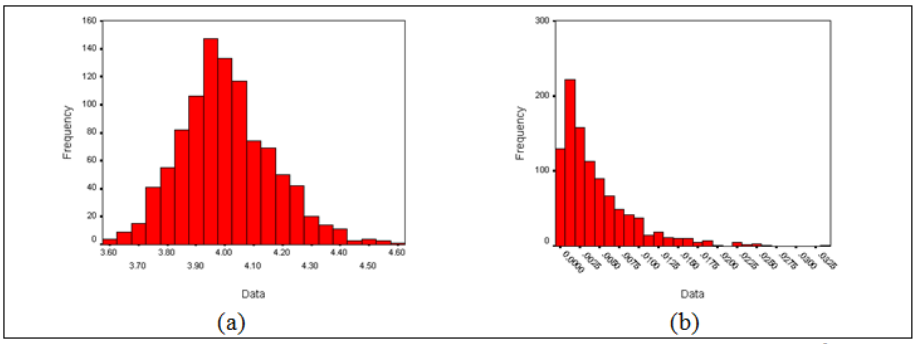

Fig-C.1: Simulation data is displayed in the histogram for (a) $\sum_{L=1}^{2}\left(\frac{\bar{x} - \mu_L}{\sigma_L}\right)^2$ **and (b)** $\sum_{L=1}^{2}\left(\frac{\bar{x}_L - \mu_L}{\sigma_L}\right)^2$ **.**

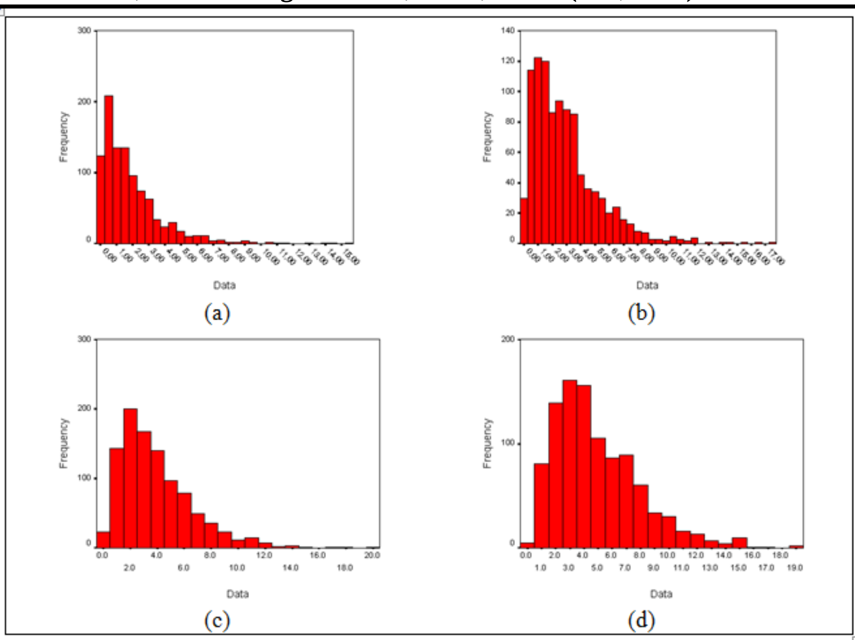

Fig-C.2: χ^2-distribution with (a) 2, (b) 3, (c) 4 and (d) 5 degrees of freedom.

APPENDIX C

The characteristics of the hypothesis testing as mentioned in the previous section are best described by performing the following processes for $\sum_{i=1}^{2} a_i \phi\left(x, \mu_i, \sigma_i^2\right)$ where ($a_1 = 0.4$, $\mu_1 = 0.0$, $\sigma_1^2 = 1.0$) and ($a_2 = 0.6$, $\mu_2 = 2.0$, $\sigma_2^2 = 0.25$) (Yusoff et al., 2013) and repeat them 1000 times:

We generate 1000 simulation data using Box and Muller Transformation (Box et al., 1958) and calculate $\sum_{L=1}^{2}\left(\frac{\bar{x} - \mu_L}{\sigma_L}\right)^2$ and $\sum_{L=1}^{2}\left(\frac{\bar{x}_L - \mu_L}{\sigma_L}\right)^2$. The one thousand (1000) simulation data is then plotted as shown in Figure C.1. Figure C.2 shows several χ^2 distributions for comparison purposes. Note that the histogram displayed in Figure C.1 (B) is similar in terms of shape to Figure C.2 (A).

CONFLICT OF INTEREST

Mohd Izhan Mohd Yusoff, the main author of the paper titled "EXPLORING THE USE OF HYPOTHESIS TESTING IN DETERMINING THE NUMBER OF COMPONENTS IN GAUSSIAN MIXED MODEL", worked in Telekom Research & Development Sdn Bhd (a subsidiary of Telekom Malaysia Berhad) for almost 20 years, the same organization that is sponsoring his PhD study at the Institute of Mathematical Sciences, University of Malaya. The title of his study was proposed by two supervisors, who are the co-authors of the said paper, namely Ibrahim Mohamed, and Mohd. Rizam Abu Bakar. Due to the affiliation of the first author to Telekom Malaysia Berhad, we are able to obtain the relevant data sets to illustrate the development of the theory considered in this paper. In other words, Telekom Malaysia Berhad, as mentioned in the said paper, is merely supplying the customer’s call detail record and does not provide any financial contribution towards the project. The data analysis and the manuscript preparation have been carried out independently. In conclusion, there is no potential conflict of interest in the study.

ACKNOWLEDGEMENTS

The authors would like to express their thanks to Telekom Malaysia Berhad for their cooperation this study. This research is financially supported by Fundamental Research Grant Scheme (No. FP012-2013A).

REFERENCES

1. Reynolds, D. A., & Rose, R. C. (1995). Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. *IEEE Transactions on speech and audio processing, 3*(1).

2. Tanigushi, M., Haft, M., Hollmen, J., & Tresp, V. (1998). Fraud detection in communications networks using neural and probabilistic methods. *Proceeding of the 1998 IEEE International Conference in Acoustics, Speech and Signal Processing (ICASSP'98), 2*, page 1241-44.
3. Becker, R. A., Volinsky, C., & Wilks, A. R. (2010). Fraud detection in Telecommunications: History and Lessons Learned. *Technometrics, 52*(1).
4. Jacobs, R. (2002). Telecommunications Fraud: The single biggest cause of revenue loss for telecommunication providers. White Paper. © Dimension Data.
5. McClelland, S. (2003). Plumbing network leaks: telecom fraud, perhaps the ultimate in virtual white collar crime, is alive, well, and on the increase. Commentary - Industry Overview, Telecommunications International,
6. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from incomplete data via the EM algorithm. *Journal Royal Statistics Society, 39*(1), 1-21.
7. Schlattmann, P. (2003). Estimating the number of components in a finite mixture model: the special case of homogeneity. *Computational Statistics and Data Analysis, 41*, 441-451.
8. Wang, H. X., Luo, B., Zhang, Q. B., & Wei, S. (2004). Estimation for the number of components in a mixture model using stepwise split-and-merge EM algorithm. *Pattern Recognition Letters, 25,* 1799-1809.
9. Miloslavsky, M., & Van Der Laan, M. J. (2003). Fittting of mixture with unspecified number of components using cross validation distance estimate. *Computational Statistics and Data Analysis, 41,* 413-428.
10. Zhuang, X., Huang, Y., Palaniappan, K., & Zhao, Y. (1996). Gaussian Mixture Density Modelling, Decomposition, and Applications. *IEEE Transactions on Image Processing, 5,* 9.
11. Yusoff, M. I. M., Mohamed, I., & Bakar, M. R. A. (2013). Improved Expectation Maximization Algorithm for Gaussian Mixed Model Using the Kernel Method. *Mathematical Problems in Engineering, 2013,* Article ID 757240, 9 Pages, http://dx.doi.org/10.1155/2013/757240.
12. Silverman, B. W. (1986). Density estimation for statistics and data analysis. London: Chapman and Hall Ltd.
13. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatics Control, 19*(6).
14. Lee, Y., Lee, K. Y., & Lee, J. (2006). The estimating optimal number of Gaussian Mixtures based on incremental k-means for Speaker Identification. *International Journal of Information Technology, 12*(7), 13-21.
15. Celeux, G., & Soromenho, G. (1996). An Entropy Criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195-212.
16. Tsay, R. S. (2005). Analysis of Financial Time Series: Financial Econometrics. John Wiley and Sons.
17. Bilmes, J. A. (1998). A Gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, University of Berkeley, ICSI-TR-97-021.
18. Yusoff, M. I. M., Mohamed, I., & Bakar, M. R. A. (2013). Fraud detection in Telecommunication using Gaussian Mixed Model. 3rd International Conference in Research & Innovation in Information Systems, Article ID 6716681, Pages 27-32,
19. Mardia, K. ., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. London: Academic Press Inc. (London) Ltd.
20. Box, G. E. P. & Muller, M. E. (1958). A note on the generating of random normal deviates. *Ann. Math. Statistic, 29*(2), 610-611.