British Journal of Mathematics & Computer Science
12(1): 1-8, 2016, Article no.BJMCS.16617
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Parameter Estimation for a Mixture of Two Univariate Gaussian Distributions: A Comparative Analysis of The Proposed and Maximum Likelihood Methods

Cliff Richard Kikawa[1*], Michael Yu Shatalov[1], Petrus Hendrik Kloppers[1] and Andrew Mkolesia[1]

[1] *Department of Mathematics and Statistics, Tshwane University of Technology, Pretoria, Republic of South Africa.*

Article Information

DOI: 10.9734/BJMCS/2016/16617
Editor(s):
(1) Morteza Seddighin, Indiana University East Richmond, USA.
Reviewers:
(1) Radosaw Jedynak, Kazimierz Pulaski University of Technology, Poland.
(2) Anonymous, Norfolk State University, USA.
(3) Sofiya Ostrovska, Atilim University, Turkey.
Complete Peer review History: http://sciencedomain.org/review-history/11481

Original Research Article

Received: 09 February 2015
Accepted: 01 July 2015
Published: 21 September 2015

Abstract

Two approaches to parameter estimation for a mixture of two univariate Gaussian distributions are numerically compared. The proposed method (PM) is based on decomposing a continuous function into its odd and even components and estimating them as polynomials, the other is the usual maximum likelihood (ML) method via the expected maximisation (EM) algorithm. An overlapped mixture of two univariate Gaussian distributions is simulated. The PM and ML are used to re-estimate the known mixture model parameters and the measure of performance is the absolute percentage error. The PM produces comparable results to those of to the ML approach. Given that the PM produces good estimates, and knowing that the ML always converges given good initial guess values (IGVs), it is thus recommended that the PM be used symbiotically with the ML to provide IGVs for the EM algorithm.

Keywords: Parameter estimation; univariate gaussian mixture; maximum likelihood; EM algorithm; monte carlo simulation.

2010 Mathematics Subject Classification: 62E05, 62G05.

**Corresponding author: E-mail: richard.kikawa@gmail.com*

1 Introduction

When computing model parameters, several statistical methods that largely depend on the probability density function of the underlying population are employed. In applied statistics, the most frequently encountered model deals with a set of entities on which statistical inferences are made based on a sample drawn from the parent entities. The entities are often times a mixture of several component populations, say K-components with conditional densities $p(y \mid \lambda_i)_{i=1,\ldots,k}$, and prior probabilities $p(\lambda_i)_{i=1,\ldots,k}$, where λ_i denotes the proportion of a given component in the mixture [1]. This scheme is aimed at ascertaining the composition of each individual population in the mixture understudy. The study is conducted on a sample drawn from the parent entities or population. For a great deal in statistical modelling the probability distribution of the constituent K-components is considered known, thereafter the major objective is to estimate the parameters of the probability density associated with the parametric family of the mixture distributions. In this study, the mixture is considered finite since it has known component mixture distributions, (i.e. $K = 2$), of the form

$$f(x, \sigma_1, \sigma_2, \mu_1, \mu_2, \lambda) = \lambda \frac{1}{\sigma_1\sqrt{2\pi}} e^{\{\frac{-(x-\varepsilon\mu_1)^2}{2\sigma_1^2}\}} + (1-\lambda)\frac{1}{\sigma_2\sqrt{2\pi}} e^{\{\frac{-(x-\varepsilon\mu_2)^2}{2\sigma_2^2}\}}. \tag{1.1}$$

A general representation of two normal mixtures with equal mixing proportions of equation 1.1, is of the form

$$f(x, \Psi) = A_1 e^{-\alpha_1(x-\varepsilon\mu_1)^2} + A_2 e^{-\alpha_2(x-\varepsilon\mu_2)^2}, \tag{1.2}$$

where $\Psi = (A_1, A_2, \mu_1, \mu_2, \alpha_1, \alpha_2)$ is the absolute collection of distinct parameters occurring in the mixture and ε is the separation or distance between the absolute mean μ of the overall mixture and the means μ_1 and μ_2 of the individual K-component distributions. Its significance is to emphasize a non-zero separation between the distribution means. The Gaussian distribution has been studied, because it is regarded as the most important distribution in the scientific domain and appears more frequently in the applied statistics [1] [2]. In section 2, we describe the problem of parameter estimation associated with the maximum likelihood (ML) method via the expectation maximization (EM) algorithm.

2 Estimation Problem

There exist several methods for approximating parameters of a mixture of normal distributions [3], [4] [5]. However, the ML method via the EM algorithm is considered as one of the most robust methods for practical applications [5]. In part, this is because the computed estimates have acceptable sampling and asymptotic properties and as well the likelihood function is admirable for iterative computations for parameter optimisation [1] [6]. However, the ML method leads to implicit equations in the unknown parameters and can not be solved analytically since they do not have closed-form solutions [7], hence an ill-posed problem. Their solutions are mainly obtained through iterative optimization, which requires initial guess values (IGVs) to start the iteration process. The first order necessary conditions of providing IGVs were published by [8] A usually encountered problem is the selection of suitable initial parameter values [9]. From reviewed literature, it appears that no systematic method exists as yet for computing the IGVs. It is more of an art than a science to provide reasonable guesses that will guarantee convergence to the optimal solution [10] [6]. Providing a reasonable guess requires high expertise and also, if the provided guess is far from the optimal value, then high computation time may result [10] [11]. The proposed method (PM) could be used symbiotically with the ML method to provide IGVs. For an informative and detailed presentation of the ML method via the EM algorithm we refer the reader to [1], [12] [13]. In section 3 we provide a simple and elementary mathematical approach on which the formulation of the PM is based.

3 Mathematical Basis

Mathematical notions for decomposing continuous functions that are bounded by real numbers are presented. It is upon these notions that the PM is developed. Consider the following elementary but important trivial notions of symmetry and antisymmetry from the field of functional analysis:

- A function $f : \mathbb{R} \to \mathbb{R}$ is said to be even if $f(-\xi) = f(\xi)$, for all $\xi \epsilon \mathbb{R}$, and odd if $f(-\xi) = -f(\xi)$, for all $\xi \epsilon \mathbb{R}$.
- Every function $f : \mathbb{R} \to \mathbb{R}$ can be decomposed into a sum of its even $f_e(\xi)$ and odd $f_o(\xi)$, components

$$f_e(\xi) = \frac{f(\xi) + f(-\xi)}{2} \text{ and } f_o(\xi) = \frac{f(\xi) - f(-\xi)}{2}. \tag{3.1}$$

Therefore,

$$f(\xi) = f_e(\xi) + f_o(\xi). \tag{3.2}$$

Hence, all functions limited to the domain of real numbers can be expressed as a sum of their even and odd components [14].

3.1 Proposed Method

In this section we state without proof a theorem on which the PM method can closely be linked. For the proof and a detailed discussion of the theorem, the interested reader can refer to [15], [16], [17] [14].

Theorem 1. The Weierstrass approximation theorem states that, any continuous real function defined on a bounded closed interval of real numbers can be uniformly approximated by polynomials.

The formulation of the PM is based on the understanding of functional decomposition (Section 3) and the Weierstrass approximation theorem which was later generalised by [16].

3.1.1 Procedure

Using the decomposition procedure in Equation (3.1), for $\xi = x$, the even component is

$$f_e(x) = \bar{A}_1 e^{-\alpha_1 x^2} + \bar{A}_2 e^{-\alpha_2 x^2} + O(x^2) \tag{3.3}$$

where

$$\left.\begin{array}{l} \bar{A}_1 = A_1 e^{-\alpha_1 \mu_1^2 \varepsilon^2} \\ \bar{A}_2 = A_2 e^{-\alpha_2 \mu_2^2 \varepsilon^2} \\ . \\ . \\ . \end{array}\right\} \tag{3.4}$$

and the odd component

$$f_o(x) = \bar{\beta}_1 x e^{-\alpha_2 x^2} + \bar{\beta}_2 x e^{-\alpha_1 x^2} + O(x^3) \tag{3.5}$$

where

$$\left.\begin{array}{l} \bar{\beta}_1 = A_1 \alpha_1 \mu_1 \varepsilon e^{-\alpha_1 \mu_1^2 \varepsilon^2} \\ \bar{\beta}_2 = A_2 \alpha_2 \mu_2 \varepsilon e^{-\alpha_2 \mu_2^2 \varepsilon^2} \\ . \\ . \\ . \end{array}\right\}. \tag{3.6}$$

A mathematical software, *Maple15*, is used for the functional decomposition. Equations (3.3 & 3.5) represent polynomial-exponential series in x. When higher accuracy is required in both theoretical and practical problems, higher order terms could be included, though this comes at a cost of heavy

computation. Consider the following representation of Equation (3.3) having neglected the terms $\bar{A}_3 x^2 e^{-\alpha_2 x^2}$ and $\bar{A}_4 x^2 e^{-\alpha_1 x^2}$

$$f_e(x) = \bar{A}_1 e^{-\alpha_1 x^2} + \bar{A}_2 e^{-\alpha_2 x^2}. \tag{3.7}$$

Let $y = x^2$, hence $f(\sqrt{y}) = f(x)$, which for easy presentation is written as $\bar{f}(y) = f(\sqrt{y}) = f(x)$ in the subsequent equations. Equation (3.3) can now be written as

$$\bar{f}(y) = \bar{A}_1 e^{-\alpha_1 y} + \bar{A}_2 e^{-\alpha_2 y}. \tag{3.8}$$

It is observed that, Equation (3.8) represents a general solution of a second order differential equation that could be written as

$$\frac{d^2\bar{f}}{dy^2} + \beta_1 \frac{d\bar{f}}{dy} + \beta_2 \bar{f} = 0. \tag{3.9}$$

β_1 and β_2 can now be estimated from a dataset $\{\bar{f}(y_i), y_i\}$, $i = 1, ..., n$. Firstly $\frac{d\bar{f}}{dy}$ and $\frac{d^2\bar{f}}{dy^2}$ are calculated numerically and the ordinary least squares (OLS) method is used to calculate β_1 and β_2. Having estimated β_1 and β_2, the relationship between the exponents α_1 and α_2 of Equation (3.8) and the parameters β_1 and β_2 of Equation (3.9) can be written as: $\beta_1 = \alpha_1 + \alpha_2$ and $\beta_2 = \alpha_1\alpha_2$ and solved in terms of β_1 and β_2, where $\alpha_1 and \alpha_2$ are exponents as defined in Equation (1.2)

$$\alpha_1, \alpha_2 = \frac{\beta_1}{2} \pm \sqrt{\frac{\beta_1^2}{4} - \beta_2}. \tag{3.10}$$

Substitute the estimated values of α_1 and α_2, into Equation (3.8) and let

$$Y_1 = e^{-\alpha_1 y} \text{ and } Y_2 = e^{-\alpha_2 y}. \tag{3.11}$$

Substituting Equations (3.11) into Equation (3.8) yields

$$\bar{f}(y) = \bar{A}_1 Y_1 + \bar{A}_2 Y_2. \tag{3.12}$$

If a dataset $\{x_j, f(x_j)\}$ where $j = 1, ..., n$ is available, then a new dataset $\{y_j = x_i^2, f(y_j) = f(\sqrt{x_i})\}$ where $j = 1, ..., n$ could be constructed. Equation (3.12) can as well be solved using the OLS methods to obtain parameter estimates for $\bar{A}_1$ and $\bar{A}_2$, since $Y_{1j} = e^{-\alpha_1 y_j}$ and $Y_{2j} = e^{-\alpha_2 y_j}$. Substituting for $\bar{A}_1$ and $\bar{A}_2$, in the first two expressions of Equation (3.4), two equations are obtained containing four unknown parameters. Therefore, the parameters cannot be estimated explicitly at this stage. Let us write the first two expressions of Equation (3.4) as

$$\bar{A}_1 = A_1 e^{-\alpha_1 \mu_1^2 \varepsilon^2}, \tag{3.13}$$

and

$$\bar{A}_2 = A_2 e^{-\alpha_2 \mu_2^2 \varepsilon^2}. \tag{3.14}$$

On estimating the odd component Equation (3.5) of the mixture model, we first divide it by x so that higher order x values in it can appear just like those in the even component. This is aimed at estimating it exactly in the same way as the even component

$$\frac{f_o(x)}{x} = \bar{\beta}_1 e^{-\alpha_1 x^2} + \bar{\beta}_2 e^{-\alpha_2 x^2} + O(x^2) \tag{3.15}$$

Dividing Equation (3.5) by x, the odd terms are converted into even, see Equation (3.15) and in this way both the even and odd components can be compared. Only the first two terms of Equation (3.15) are considered. Let $y = x^2$ then $f(x) = f(\sqrt{y})$ and, $\bar{f}(y) = f(\sqrt{y})$, as before in the estimation of the even component. Neglecting the terms $\bar{\beta}_3 x^2 e^{-\alpha_2 x^2}$ and $\bar{\beta}_4 x^2 e^{-\alpha_1 x^2}$, Equation (3.15) can be written as

$$\bar{f}(y) = \bar{B}_1 e^{-\alpha_1 y} + \bar{B}_2 e^{-\alpha_2 y}. \tag{3.16}$$

Computing the exponential functions as in Equation (3.11) yields

$$Z_1 = e^{-\alpha_1 y} \text{ and } Z_2 = e^{-\alpha_2 y}. \tag{3.17}$$

Let $\bar{f}(y) = \bar{f}(z)$ and substitute Equation (3.17) into Equation (3.16) to obtain

$$\bar{f}(z) = \bar{B}_1 Z_1 + \bar{B}_2 Z_2. \tag{3.18}$$

Again if a dataset $\{x_j, f(x_j)\}$ is available, then Equation (3.18) can as well be solved using OLS methods to obtain parameter estimates for $\bar{B}_1$ and $\bar{B}_2$. The estimated parameters are then substituted into Equation (3.6) to obtain

$$\bar{\beta}_1 = A_1\alpha_1\mu_1\varepsilon e^{-\alpha_1\mu_1^2\varepsilon^2}, \tag{3.19}$$

and

$$\bar{\beta}_2 = A_2\alpha_2\mu_2\varepsilon e^{-\alpha_2\mu_2^2\varepsilon^2}. \tag{3.20}$$

Recall that at some stage, the odd function was divided by an x factor so that it could approximate an even function. In this way the equations from the odd function can be compared to those from the even component. Dividing Equation (3.14) by Equation (3.20) and simplifying, we are able to estimate μ_2 as

$$\hat{\mu}_2 = \frac{\bar{\beta}_2}{\bar{A}_2\alpha_2\varepsilon}, \text{for } \varepsilon \neq 0. \tag{3.21}$$

On making appropriate algebraic substitutions the rest of the unknown parameters; μ_1, A_1, and A_2 are obtained from the following equations

$$\hat{A}_2 = -\frac{\bar{A}_2^3\alpha_2}{\bar{\beta}_2^2 - \bar{A}_2^2\alpha_2}, \tag{3.22}$$

$$\hat{\mu}_1 = \frac{\bar{\beta}_1}{\bar{A}_1\alpha_1\varepsilon}, \tag{3.23}$$

and

$$\hat{A}_1 = -\frac{\bar{A}_1^3\alpha_1}{\bar{\beta}_1^2 - \bar{A}_1^2\alpha_1}. \tag{3.24}$$

After estimating the general unknown parameters A_1, A_2, μ_1 and μ_2, of Equation (1.2), it can be concluded that the parameters of the mixture model have been completely estimated. It is also observed that the two means of the original problem, Equation (1.1) are estimated from Equations (3.21 & 3.23); the standard deviations are computed from the relationships of Equations (3.22 & 3.24) to the original problem, Equation (1.1). At this point, it is important to caution the reader that, two values for each of the standard deviations (σ_1 and σ_2) could be obtained from the latter relationships, therefore, an average of the values is considered as the appropriate estimate for the respective parameters. Hence the estimated parameters $\hat{A}_1$, $\hat{A}_2$, $\hat{\mu}_1$ and $\hat{\mu}_2$, can be used as the IGVs for the unknown parameters in the EM algorithm.

Remark. In the process of parameter estimation small values of x are used. In this case it is possible to neglect terms like $x^2e^{-\alpha x^2}$ in comparison with $e^{-\alpha x^2}$.

Remark. The origin, $x = 0$, in the vicinity of the maximum of the original distribution is considered. Stated otherwise, the working interval is $0 < x \ll 1$.

Assumption 1. *For our computations it is assumed that the value of $\varepsilon = 1$. In otherwords the respective means of the distributions have a unity separation from their absolute mean.*

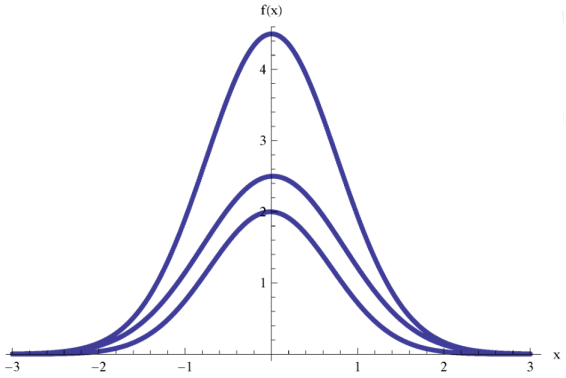

Fig. 1. Two-component Gaussian mixture models. The upper most curve is the mixture model, lower curve has, $\mu_1 = -0.01$ and middle curve has $\mu_2 = 0.2$

4 A Monte Carlo Simulation Study

Our interest in this study is to compare the estimates from the proposed method with those of the maximum-likelihood method and make appropriate recommendations based on the accuracy of the results. We perform a simulation study of two univariate normal mixture distribution based on two assumptions; one is that the distributions have equal mixing proportions and two, that the mixing distributions had different standard deviations and means. The mixture distribution function is expressed as

$$\lambda N(\mu_1, \sigma_1) + (1 - \lambda)N(\mu_2, \sigma_2), \tag{4.1}$$

where N emphasizes the normal distribution and that $\lambda + 1 - \lambda = 1$, lambda considered known as stated in the assumptions. For our simulation it is taken as 0.5. Since the separation between components is determined by varying their parameters, one normal component is $0.5N(-0.01, 0.403)$ and the other $0.5N(0.2, 0.448)$. The values of the standard deviations $\sigma_1 = 0.403$ and $\sigma_2 = 0.448$, and means $\mu_1 = -0.01$ and $\mu_2 = 0.2$ are chosen in such a way that an overlapped mixture is obtained.

In our simulation study, a fixed sample of size 599 was generated using *Mathematica 8* basing on the parameters indicated above. The synthetic data exhibited the required behaviour of an overlapped normal mixture distribution, see Figure 1. In section 5, we present the results and their detailed discussion.

5 Results and Discussion

Using the synthetic sample data obtained in section 4, the PM and ML methods were applied to compute the known parameters of a mixture model. The results of the two methods were compared and the measure of accuracy of their estimates was the absolute percentage error (APE), see Table 1. A smaller value of the APE of a method relative to the other indicated a higher accuracy of that respective method in re-estimating the considered parameter. From our results it can be observed that the PM produces better estimates of the standard deviations and for the second mean, μ_2 as compared to the ML approach, see Table 1.

Table 1: Exact parameter values for the Gaussian mixture; estimated values for both the proposed (PM) and maximum likelihood (ML) methods and their respective absolute percentage error (APE).

Parameters	Exact	PM	ML	APE (PM)	APE (ML)
μ_1	-0.01	0.184184	0.021	1941	310
μ_2	0.200	0.212534	0.049	6.3	75.5
σ_1	0.403	0.543903	0.011	35	97
σ_2	0.448	0.589939	0.199	31.6	55.6

6 Conclusion

Considering the estimates, and the accuracy of the two approaches, it is reasonable to conclude that, the PM produced better results in comparison to the standard ML method. Nevertheless, it has to be noted that the results of the ML method largely depend on the initial approximations. The large error observed might have resulted from the provided starting approximations being far from the neighbourhood of the required solution and could not guarantee convergence of the method to the "exact" optimal values. Results presented in Table 1 should not lead to an instant or "sweeping" conclusion that the PM is a better estimation approach than the ML method in the general case. [18] observes that, the EM algorithm will always converge, but the results depend on the starting approximations of the unknown parameters and the possibility of multiple solutions is always present.

7 Recommendation

Since considerable research has shown that the ML approach via the EM algorithm always converges to the optimal solutions given, "good" IGVs, we thus recommend that the PM approach be used symbiotically with the ML to provide the necessary initial approximations required to commence the iteration process in the EM algorithm.

Table 1 summarises the results obtained by applying the PM and ML estimation approaches on the simulated data. Figure 1 shows the component mixture of the two univariate Gaussian distributions. This kind of mixture is what is commonly referred to as the overlapped mixture.

Acknowledgement

The authors would like to thank the Directorate of Research and Innovation, Tshwane University of Technology for funding this research under the Postdoctoral research fund 2014/2015. We also extend our sincere appreciation to the editorial board and the three anonymous reviewers whose critics and suggestions greatly improved on the presentation of the manuscript.

Competing Interests

The authors declare that no competing interests exist.

References

[1] Davenport JW, Bezdek JC, Hathaway RJ. Parameter estimation for finite mixture distributions. Computers and Mathematics with Applications. 1988;15(10):819-828.

[2] Cox DR, Hinkley DV. Theoretical Statistics. Chapman Hall; 1974.

[3] Day NE. Estimating the components of a mixture of normal distributions. Biometrika. 1969;56:463-474.

[4] Quandt RE, Ramsey JB. Estimating mixtures of normal distributions and switching regressions. Journal of American Statistical Association. 1978;73:730-738.

[5] Redner RA, Walker HF. Mixture densities, maximum likelihood, and the EM algorithm. SIAM Rev. 1984;26:195-239.

[6] Kikawa CR. Methods to solve transcendental least-squares problems and their statistical inferences. Tshwane University of Technology, Republic of South Africa; 2013 (Unpublished Doctoral thesis).

[7] Sanjey KS, Umesh K, Dinesh K. Bayesian estimation of the gamma parameter and reliability function under asymmetric loss function. REVSTAT-Statistical Journal. 2011;9(3):247-260.

[8] Wolfe JH. Pattern clustering by multivariate mixture analysis. Multivariate Behavioural Research. 1970;5:329-350.

[9] Kurimo M. Training mixture density hidden Markov models with self-organizing maps and learning vector-quantization. Technical Report A43, Helsinki University of Technology, Department of Computer Science and Engineering, Helsinki University of Technology: Finland; 1997.

[10] Damodar GN. Basic Econometrics. McGraw-Hill/Irwin., 4 edition; 2004.

[11] Kloppers PH, Kikawa CR, Shatalov MY. A new method for least squares identification of parameters of the transcendental equations.International Journal of the Physical Sciences. 2012;7:5218-5223.

[12] Hathaway R. Constrained maximum-likelihood estimation for a mixture of m univariate normal distributions. Statistical Technical Report No.92, University of South Carolina; 1983.

[13] Morgane B. Finite mixture distributions; 2006.

[14] Einsiedler M, Ward T. Functional analysis notes: Draft July 2; 2012.
Avaliable: http://www.math.ethz.ch/ einsiedl/FA-lecture.pdf (Last accessed in June 2014).

[15] Stone M. The generalised Weierstrass approximation theorem. Mathematics Magazine. 1948a;21:587-605.

[16] Stone M. The generalised Weierstrass approximation theorem. Mathematics Magazine. 1948b;21:167-184.

[17] Rudin W. Principles of mathematical analysis. McGraw-Hill, New York; 1976.

[18] Titterington DM. Statistical analysis of finite mixture distributions. Wiley, Chichester; 1986.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/11481