INFORMATICA, 1997, Vol. 8, No. 3, 331–343

EFFICIENCY ANALYSIS OF ONE ESTIMATION AND CLUSTERIZATION PROCEDURE OF ONE-DIMENSIONAL GAUSSIAN MIXTURE

Gintautas JAKIMAUSKAS

Institute of Mathematics and Informatics
2600 Vilnius, Akademijos 4, Lithuania
E-mail: gnt@ktl.mii.lt

Abstract. Efficiency of one automatic estimation and clusterization procedure of one-dimensional Gaussian mixture which combines EM algorithm with non-parametric estimation is considered. The paper is based on mathematical methods of statistical estimation of a mixture of Gaussian distributions presented by R. Rudzkis and M. Radavičius (1995). The main result of the implementation of the mathematical methods is completely automatic procedure which can start from no information about unknown parameters and finish with final mixture model (tested for adequacy).

Key words: mixture of Gaussian distributions, EM algorithm.

1. Introduction. Let $\{X_1, X_2, \ldots, X_N\} \stackrel{\text{def}}{=} X^N$ be a sample of i.i.d. one-dimensional random variables with a distribution density f which belongs to the class of Gaussian mixture densities

$$f(x) = \sum_{i=1}^{q} p_i \varphi_i(x) \stackrel{\text{def}}{=} f_q(x, \theta), \tag{1}$$

where $q \in \{1, 2, 3, \ldots\}$, $p_i > 0$, $\sum p_i = 1$, φ_i is a Gaussian distribution density with mean μ_i and variance σ_i^2, $\theta = (p_i, \mu_i, \sigma_i^2, i = 1, 2, \ldots, q)$ is a multidimensional parameter. Let the unknown distribution density $f(\cdot, \bar{\theta}) = f_{\bar{q}}(\cdot, \bar{\theta})$ have $\bar{q}$ components with unknown parameter $\bar{\theta} = (\bar{p}_i, \bar{\mu}_i, \bar{\sigma}_i^2, i = 1, 2, \ldots, \bar{q})$. Number of components $\bar{q}$ may also be unknown. We suppose that all components of distribution density $f(\cdot, \bar{\theta})$ are different: $\bar{\varphi} \neq \bar{\varphi}_j$ if $i \neq j$.

We will consider the problem of estimating the unknown parameter $\bar{\theta}$, including the problem of determining the unknown number of components $\bar{q}$.

If number of components q is known ($q = \bar{q}$), then the maximum likelihood estimate (MLE) $\theta^* = \arg\max_\theta l(\theta)$, where

$$l(\theta) = \sum_{j=1}^{N} \log f_q(X_j, \theta), \tag{2}$$

is an efficient estimate of $\bar{\theta}$. If number of components is unknown, then we have an additional problem – testing of model adequacy.

The most common method for calculating MLE for Gaussian mixtures is so called EM (Expectation Maximization) algorithm. It is an iterative procedure which converges to MLE if starting parameters are sufficiently close to θ^*. For mixture distributions the EM algorithm was proposed independently by Schlesinger (1965), Hasselblad (1966), and Behboodian (1970). By now the properties of the EM algorithm have been studied well enough. On the convergence properties of the EM algorithm see Wu (1983). Also see, e.g., monographs (Everitt and Hand, 1981; Aivazyan *et al.*, 1989; McLacklan and Basford, 1988; Titterington *et al.*, 1985). For further references see Rudzkis and Radavičius (1995). The popularity of the EM algorithm is explained by computational stability and simplicity of implementation on a computer. MLE also can be calculated using common optimization methods (for example, conjugate gradient method, see the book Gill *et al.* (1985).

Nevertheless, many problems arise estimating mixture density (1) using EM algorithm for calculating MLE. This algorithm is not robust if there are even small number of observations from some non-Gaussian mixture component added to mixture model (1). Function (2) has many local maxima, so if we start EM algorithm from some point too far from θ^*, the EM algorithm converges to local maximum, not to the global one. Testing of model adequacy in the case of unknown number of components is also a complicated problem.

This paper is based on mathematical methods of statistical estimation of a mixture of Gaussian distributions presented by Rudzkis and Radavičius (1995). These methods include: adding background cluster to model (1), methods for selection of initial values for the EM algorithm, methods for joining components and methods for testing of model adequacy. The main idea of these methods is combining EM algorithm with non-parametric estimation.

These mathematical methods were implemented on a computer using Borland Pascal compiler. There are two implementations – one for one-dimensional

case (briefly described below), second for multidimensional case. The implementation includes solutions of some other computational problems (for example, reducing calculation time not loosing accuracy too much). The main result of the implementation (made with J. Sušinskas in close cooperation with R. Rudzkis and M. Radavičius) was completely automatic procedure which can start from no information about unknown parameters and finish with final mixture model (tested for adequacy). Step-by-step procedures are also available.

In this section we list mathematical methods described in Rudzkis and Radavičius (1995) (we use slightly different notation because we consider only one-dimensional case) and briefly describe implementation of mathematical methods. Results of efficiency analysis are given in next section. In Jakimauskas and Sušinskas (1996) one can find much more detailed description, examples and some additional results of efficiency analysis.

Sample classification. Problems of estimating the unknown parameter and sample classification are closely related. For given distribution density $f_q(x,\theta)$ assignment of each $X \in X^N$ to some of the distinct classes (i.e., subsets of X^N) $K_i, i = 1,2,\ldots,q$, can be done using non-random classification (for example, Bayes rule of classification) or randomized classification. Bayes rule of classification assigns an observation $X \in X^N$ to the ith class if $i = \arg\max_{k=1,2,\ldots,q} p_k\varphi_k(X)$. Randomized classification assigns observations $X \in X^N$ (randomly and independently each of other) to the ith class with probability

$$\pi_i(X,\theta) = \frac{p_i\varphi_i(X)}{f_q(X,\theta)},\ i = 1,2,\ldots,q. \tag{3}$$

Any rule of classification does not allow to get precise estimates of true classes because mixture density components of (1) overlap each other. Minimal mean classification error is achieved using Bayes rule of classification. Note that distribution densities which correspond to classes obtained using Bayes rule of classification in model (1) are truncated Gaussian densities and differ from true density components.

EM algorithm. EM algorithm is the most common method for calculating the estimate θ^*. Suppose an initial parameter value θ^0 is given (selection of θ^0 will be discussed later). Next parameter value θ^1 is defined by the following

equalities:

$$p_i^1 = \frac{1}{N}\sum_{j=1}^{N} \pi_i(X_j, \theta^0), \quad i = 1, 2, \ldots, q, \tag{4a}$$

$$\mu_i^1 = \frac{1}{N}\sum_{j=1}^{N} \frac{\pi_i(X_j, \theta^0)}{p_i^0} X_j, \quad i = 1, 2, \ldots, q, \tag{4b}$$

$$(\sigma_i^2)^1 = \frac{1}{N}\sum_{j=1}^{N} \frac{\pi_i(X_j, \theta^0)}{p_i^0} (X_j - \mu_i^0)^2, \quad i = 1, 2, \ldots, q. \tag{4c}$$

Equalities (4) are the formulae of one iteration of EM algorithm. Next iteration begins with assigning obtained value of θ^1 to θ^0 and ends with recalculating θ^1 using formulae (4) with changed classification rule (3). EM algorithm usually ends after some predefined number of iterations. It is known that θ values in EM algorithm converge to some local maximum of function (2). To achieve global maximum of (2) it is necessary to select initial parameter value θ^0 sufficiently close to θ^*.

Background cluster. The procedure of EM algorithm is more stable if we use an additional background (or noise) cluster for temporary calculations. We extend class definition in (1) by adding some non-Gaussian component. Then sample elements which can be hardly assigned to any of Gaussian clusters are assigned to additional noise cluster. At the end we return to model (1) by deleting noise cluster or replacing it by additional Gaussian component.

Two methods of adding background cluster given below are based on non-parametric estimate $\widehat{f}$ supposed to be sufficiently close to the true density $f(\cdot, \bar{\theta})$.

The first method is adding an uniform density component only on the set $\{x : \bar{f}(x) > \varepsilon_u\}$, where ε_u is some small positive value. We get the modified model

$$f_q(x, \theta) = p_0 u_0 \mathbf{1}_{\{x : \widehat{f}(x) > \varepsilon_u\}}(x) + \sum_{i=1}^{q} p_i \varphi_i(x), \tag{5}$$

where $p_0 \geqslant 0, \sum_{i=0}^{q} p_i = 1, u_0 = 1/\mathrm{Vol}\{x : \widehat{f}(x) > \varepsilon_u\}$.

The second method is adding density component only on the set $\{x : \widehat{f}(x) > f_q(x, \theta)\}$, where $f_q(x, \theta)$ is defined by (1). We get the model

$$f_q(x, \theta) = C \cdot \left(\left(\widehat{f}(x) - \sum_{i=1}^{q} p_i \varphi_i(x) \right)_+ + \sum_{i=1}^{q} p_i \varphi_i(x) \right), \tag{6}$$

where C is a norming constant, $(\cdot)_+$ denotes positive value. To get the modified EM algorithm (in both cases) we define non-Gaussian component $p_0\varphi_0$ and add formulae with indices $i = 0$ in (3) and (4a).

Non-parametric density estimate. One of the most simple and widely used non-parametric density estimates is kernel estimate with variable bandwidth $h = h(x, X^N)$

$$\hat{f}(x) = \sum_{j=1}^{N} \frac{1}{Nh} W\left(\frac{x - X_j}{h}\right). \tag{7}$$

Selection of kernel function W depends on information about unknown density. Because class (1) is fairly wide, selection of simple and easy to calculate non-negative kernel function is the best choice. Such kernel function is, for example, $W(x) = (3/4)(1 - x^2)_+$. Moreover, this kernel function is optimal in certain sense (see Rudzkis and Radavičius (1995), p. 43).

More complicated problem is selecting bandwidth h. Its values can be calculated using k-nearest-neighbor approach. Given k value, for each $X \in X^N$ we can find interval that contains at least k nearest data points (including itself). Half length of this interval can be treated as bandwidth value for this X. Moreover, exact k value is not required, because the estimate (7) does not change much when k is changed slightly.

Selection of initial parameters. Convergence of EM algorithm to global maximum θ^* depends on the good choice of the initial parameter θ^0. This is a complicated problem, especially when true number of components is unknown. This problem can be solved by combining separate steps, each one supposed to be an attempt to make a more adequate model. For example, an algorithm of finding q components can be divided into algorithm of finding one next component repeated until required number of components is reached. Note that algorithm of finding next component usually may contain many other procedures (for example, EM algorithm) that try to make estimated parameter more precise before the end of this algorithm.

Key problem is finding next component. When new component is found, new parameter θ may be made more precise using EM algorithm or other methods. We refer to this problem as to problem of selection of initial parameters. Note that we use modified mixture model (5) or (6) with background cluster. Method using second derivative of non-parametric estimate is given in Rudzkis and Radavičius (1995).

Let we have θ^0 with $q \geqslant 0$ components and we will find parameter θ^1 with $q+1$ component. Let $p_i^1 = p_i^0, \mu_i^1 = \mu_i^0, (\sigma_i^2)^1 = (\sigma_i^2)^0, i = 1, 2, \ldots, q$. Then we define mean μ_{q+1}^1 using one of equalities

$$\mu_{q+1}^1 = \arg \max_{X \in X^N} (\widehat{f}(X) - f_q(X, \theta^0)), \tag{8a}$$

$$\mu_{q+1}^1 = \arg \max_{X \in X^N} ((\widehat{f}(X))^{1/2} - (f_q(X, \theta^0))^{1/2}), \tag{8b}$$

$$\mu_{q+1}^1 = \arg \max_{X \in X^N} (\widehat{f}(X)/f_q(X, \theta^0)), \tag{8c}$$

Selection of equality (8b) is based on the asymptotic equal distribution variance at each point under very general conditions to be hold.

Variance of $(q+1)$th component may be calculated using one of the following methods. The first one is the empirical variance of the noise cluster. The second one is empirical variance of a part (say, 1/5) of the noise cluster points nearest to just selected mean. Probability of $(q+1)$th component may be calculated using one of the following methods. The first one is assigning $p_{q+1}^1 = 1/(q+2)$ (we have $q+1$ Gaussian components and one noise cluster) and proportionally recalculating other probabilities, so that sum of all probabilities will be equal to 1. The second method is evaluating $p_{q+1}\varphi_{q+1}$ (for details see Rudzkis and Radavičius (1995), p. 41) and recalculating component probabilities using formula (4a). Methods of finding mean, variance and probability may be changed at consecutive attempts to find a new component.

Testing model adequacy. If number of mixture components is unknown we must have some criteria for testing the adequacy of sample data to the model with estimated parameters. Parametric criteria (based most often on likelihood ratio criterion) are often used for testing the model adequacy. However, the application of these criteria in our case faces some theoretical problems. In order to check the model adequacy, it is more reasonable to employ non-parametric criteria.

Denote (for details see Rudzkis and Radavičius (1995), p. 49)

$$\psi_3 = \frac{1}{N} \sum_{j=1}^{N} \frac{\widehat{f}(X_j)}{f_q(X_j, \theta)} - 1. \tag{9}$$

We reject hypothesis of adequacy of sample data to the model with obtained θ if value of ψ_3 is not sufficiently small. Given the significance level α, it

is desirable to choose the rejection level as close as possible to the quantile $u_a : P\{\psi_3 > u_a\} = \alpha$. We recommend criterion (9), because we can calculate sufficiently accurate distribution characteristics of ψ_3.

This non-parametric criterion is not sufficient to make a decision that a model is adequate. Additional criteria are based on behavior of probability of the noise cluster p_0 (if we use model (5)) and on value of ML function $l(\theta)$. We reject the hypothesis of model adequacy if new value of p_0 is significantly less than its previous value. Similarly, we reject this hypothesis if value of $l(\theta)$ with new parameter θ is significantly greater than value of $l(\theta)$ with previous parameter. In both cases we make a decision that the new parameter made significant improvement and we must make more attempts to improve the model.

Implementation of mathematical methods. Main procedures of the implementation of mathematical methods are Find Next procedure and Auto procedure. The first one finds parameters of new component. The second one finds Gaussian mixture components automatically starting from any specified parameter value, including the case when starting parameter is unknown. Other procedures are: Refinement (implementation of EM algorithm), Join (joins components that differ insignificantly), Optimization (general optimization procedure using conjugate gradient method (see Gill *et al.* (1985)), No Noise (deletes noise cluster), Clusterization (assigns cluster numbers to each sample element).

Find Next procedure finds parameters of new component and returns indicator *done*, indicating whether all components are found or not. We can keep fixed any number of parameters. Also we can set minimal and maximal number of components $q_{\min}$ and $q_{\max}$. Default model with background cluster is given by (5). We can change it to the model (6).

Auto procedure finds Gaussian mixture components automatically starting from any parameter value. To reduce calculation time for most calculations we use special grouped sample. At the beginning (if we start from $q > 0$) we apply Refinement procedure and (optionally) Optimization procedure. After this we have more precise parameters to start main loop. Main loop of finding mixture components consists of Find Next procedure repeated until indicator *done* becomes *true*. After this we have maximum number of components. Remaining procedures can only decrease this number.

After the main loop we refine parameters and perform special joining procedure which joins components that differ insignificantly (using Akaike's information criterion). From this point we use another special grouped sample with bigger length. We apply Refinement procedure and try to delete small components. Final refinement of parameters and deleting noise is done using initial sample. Auto procedure ends with Clusterization procedure, which assigns cluster number to each $X \in X^N$.

2. Efficiency analysis. In this section we present some results of testing efficiency. We have tested:

A) how much close components with equal probabilities and standard deviations but different means can be effectively separated,

B) how much close components with equal probabilities and means but different standard deviations can be effectively separated,

C) how much small component with the same standard deviation, but with different mean and small probability can be detected.

The main goal of the tests was to determine values of the parameter at which corresponding problem can be effectively solved. We assume that a problem is effectively solved if all components are found for at least 95 per cent of random realizations.

A special program was written for these tests. All test examples start with no information about theoretical parameters.

Efficiency analysis is based on comparing values l_{Auto} of maximum likelihood function (2) corresponding to parameters obtained by Auto procedure with analogous values l_{MLE} corresponding to parameters obtained using 20 iterations of Refinement procedure starting from known theoretical parameters (these parameters are considered to be sufficiently near from MLE estimate). This comparison is done only for those realizations which estimated number of components is equal to the theoretical ones.

For each mixture model we give empirical means and standard deviations for l_{Auto} and l_{MLE} and for difference $l_{\text{Auto}} - l_{\text{MLE}}$. Also we give ratio of empirical standard deviations corresponding to $l_{\text{Auto}} - l_{\text{MLE}}$ and l_{MLE}. As seen below, test results show that the difference $l_{\text{Auto}} - l_{\text{MLE}}$ is sufficiently small and considered clusterization procedures may be treated as efficient procedures.

We need two special rules that work in full automatic way to determine that all components are found. The first one is to assign correct component

order to estimated parameters so that each partial theoretical density match corresponding partial estimated density. The second one is to make a decision whether each partial theoretical density is close enough to the estimated one. We say that all components are found if number of components in estimated model is not less than in theoretical model and each partial estimated density is close enough to the theoretical one.

Assignment of component order to estimated parameters is done in the following way. Let $f(\cdot, \bar{\theta})$ be theoretical density function and $f_q(x, \theta)$, $q \geqslant \bar{q}$ be an estimated density function. Let $j = j(i), i = 1, 2, \ldots, \bar{q}$, be some function that assigns different numbers from the set $\{1, 2, \ldots, q\}$ for different i. We select such function $j(i)$ that

$$\sum_{i=1}^{\bar{q}} ||p_{j(i)}\varphi_{j(i)} - \bar{p}_i\bar{\varphi}_i||^2 / ||\bar{p}_i\bar{\varphi}_i||^2 \tag{10}$$

has minimal value. After this estimated components are reordered according to obtained function $j(i)$.

Decision whether estimated component is close enough to theoretical one is made comparing distance between estimated density and theoretical density with distance between theoretical density and the same theoretical density shifted by the value of standard deviation. To be more precise, we compare $||p_i, \varphi(\cdot, \mu_i, \sigma_i^2) - \bar{p}_i\varphi(\cdot, \bar{\mu}_i, \bar{\sigma}_i^2)||$ with $||\bar{p}_i\varphi(\cdot, \bar{\mu}_i + \bar{\sigma}_i, \bar{\sigma}_i^2) - \bar{p}_i\varphi(\cdot, \bar{\mu}_i, \bar{\sigma}_i^2)||$ for $i = 1, 2, \ldots, \bar{q}$, and make a decision that all components are found if for all $i = 1, 2, \ldots, \bar{q}$, ratios of these values are less than 1.0. Recall that we suppose that estimated density components are already reordered and number of estimated density components is not less that number of theoretical density components. Of course, selection of such shift value and selection of entire method leaves a bit of discusssion.

We considered different number of mixture models for each problem. For each mixture model 100 random realizations of length $N = 1000$ were simulated. For each realization we performed Auto procedure starting from $q = 0$, and (independently) 20 iterations of EM algorithm using Refinement procedure starting from theoretical parameters. The second procedure was used to get the approximate distribution of difference between maximum likelihood function values for estimated parameter and the MLE.

A. Close means. We consider four mixture models with two components. In all cases $p_1 = p_2 = 0.5, \sigma_1 = \sigma_2 = 1.0$. Means are the following (of course,

only difference between means is significant):

$$\text{(A1)}: \quad \mu_1 = -1.0,\ \mu_2 = 1.0,$$
$$\text{(A2)}: \quad \mu_1 = -1.0,\ \mu_2 = 1.1,$$
$$\text{(A3)}: \quad \mu_1 = -1.0,\ \mu_2 = 1.2,$$
$$\text{(A4)}: \quad \mu_1 = -1.0,\ \mu_2 = 1.5,$$

B. Close standard deviations. We consider three mixture distributions. In all cases $p_1 = p_2 = 0.5$, $\mu_1 = \mu_2 = 0.0$. Standard deviations are the following:

$$\text{(B1)}: \quad \sigma_1 = 1.0,\ \sigma_2 = 0.35,$$
$$\text{(B2)}: \quad \sigma_1 = 1.0,\ \sigma_2 = 0.3,$$
$$\text{(B3)}: \quad \sigma_1 = 1.0,\ \sigma_2 = 0.25.$$

C. Component with small probability. We consider two mixture distributions. In all cases $\mu_1 = -1.0, \mu_2 = 2.0, \sigma_1 = \sigma_2 = 1.0$. Probabilities of components are the following:

$$\text{(C1)}: \quad p_1 = 0.9,\ p_2 = 0.1,$$
$$\text{(C2)}: \quad p_1 = 0.8,\ p_2 = 0.2.$$

Table 1 shows summary information about number of realizations with final number of clusters q after Auto procedure and number of realizations with all clusters found according to the selected decision rule.

Table 1. Summary information (total 100 realizations)

Model	Number of realizations with final q				Number of realiz. with all cl. found
	$q=1$	$q=2$	$q=3$	$q=4$	
(A1)	10	79	10	1	67
(A2)	5	83	11	1	76
(A3)	0	83	16	1	90
(A4)	0	78	17	5	95
(B1)	1	71	24	4	79
(B2)	1	78	16	5	90
(B3)	0	89	8	3	98
(C1)	0	97	3	0	93
(C2)	0	94	5	1	98

The results show that models (A4), (B3) and (C2) may be considered as limit models for listed problems.

Table 2 shows empirical means of l_{Auto}, l_{MLE} and $l_{\text{Auto}} - l_{\text{MLE}}$. Table 3 shows empirical standard deviations of l_{Auto}, l_{MLE}, $l_{\text{Auto}} - l_{\text{MLE}}$ and ratio of empirical standard deviations corresponding to $l_{\text{Auto}} - l_{\text{MLE}}$ and l_{MLE}.

Table 2. Empirical means of values of ML function .

Model	l_{Auto}	l_{MLE}	$l_{\text{Auto}} - l_{\text{MLE}}$
(A1)	−1.756601	−1.756400	−0.000201
(A2)	−1.778872	−1.778773	−0.000099
(A3)	−1.801570	−1.801627	0.000056
(A4)	−1.862134	−1.862265	0.000131
(B1)	−1.070186	−1.070208	0.000022
(B2)	−1.026589	−1.026578	−0.000012
(B3)	−0.978246	−0.978248	0.000002
(C1)	−1.658685	−1.658689	0.000005
(C2)	−1.794119	−1.794157	0.000037

Table 3. Empirical standard deviations of values of ML function

Model	l_{Auto}	$l_{\text{MLE}}{}^{**}$	$l_{\text{Auto}} - l_{\text{MLE}}{}^{*}$	ratio*/**
(A1)	0.020662	0.020616	0.000674	0.0327
(A2)	0.019335	0.019427	0.001006	0.0518
(A3)	0.020639	0.020699	0.000840	0.0406
(A4)	0.021316	0.021341	0.000443	0.0208
(B1)	0.030079	0.030079	0.000081	0.0027
(B2)	0.029524	0.029555	0.000188	0.0064
(B3)	0.032870	0.032869	0.000014	0.0004
(C1)	0.024523	0.024501	0.000318	0.0130
(C2)	0.024957	0.024958	0.000067	0.0268

Note that the difference $l_{\text{Auto}} - l_{\text{MLE}}$ is sufficiently small comparing to l_{MLE}. So considered clusterization procedures are sufficiently efficient.

REFERENCES

Aivazyan, S.A., V.M. Buchstaber, I.S. Yenyukov and L.D. Meshalkin (1989). *Applied Statistics. Classification and Reduction of Dimensionality.* Finansy i Statistika, Moscow (in Russian).

Behboodian, J. (1970). On a mixture of normal distributions. *Biometrika,* **57**, 215–217.

Everitt, B.S., and D.J. Hand (1981). *Finite Mixture Distributions.* Chapman and Hall, London.

Gill, Ph.E., W. Murray and M.H. Wright (1981). *Practical Optimization.* Academic Press, New York (in Russian: Moscow, Mir, 1985).

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics,* **8**, 431–444.

Jakimauskas, G., and J. Sušinskas (1996). *Computational Aspects of Statistical Analysis of Gaussian Mixture Combining EM Algorithm With Non-parametric Estimation (One-dimensional Case).* Preprint No. 96-6, Institute of Mathematics and Informatics, Vilnius.

McLacklan, G.J. and K.E. Basford (1988). *Mixture Models. Inference and Applications to Clustering.* Marcel Dekker, New York.

Rudzkis, R., and M. Radavičius (1995). Statistical estimation of a mixture of Gaussian distributions. *Acta Applicandae Mathematicae,* **38**, 37–54.

Schlesinger, M.I (1965). On spontaneous discrimination of images. In *Reading Automata,* Naukova Dumka, Kiev. pp. 38–45 (in Russian).

Titterington, D.M., A.F.M. Smith and U.E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

Wu, C.F.J (1983). On the convergence properties of the EM algorithm. *Ann. Statist.,* **11**, 95–103.

Received February 1977

G. Jakimauskas was born in 1956. He graduated the Faculty of Mathematics in the Vilnius University in 1979. He is a researcher at the Institute of Mathematics and Informatics.

VIENOS VIENMAČIO GAUSO MIŠINIO VERTINIMO IR KLASTERIZAVIMO PROCEDŪROS EFEKTYVUMO TYRIMAS

Gintautas JAKIMAUSKAS

Nagrinėjamas vienos vienmačio Gauso mišinio vertinimo ir klasterizavimo automatinės procedūros, derinančios EM algoritmą su neparametriniu vertinimu, efektyvumas. Straipsnis paremtas Gauso mišinio statistinio vertinimo metodais, pateiktais R. Rudzkio ir M. Radavičiaus straipsnyje žurnale *Acta Applicandae Mathematicae* **38** (1995). Pagrindinis matematinių metodų pritaikymo rezultatas yra visiškai automatinė procedūra, kuri gali pradėti darbą be jokios informacijos apie nežinomus parametrus ir baigti darbą su galutiniu mišinio modeliu (pratestuotu modelio adekvatumui).