PROCEEDINGS OF THE 7th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION AND IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES (PRIA-7-2004)
Mathematical Methods in Pattern Recognition

Bounds for the Number of Modes of the Simplest Gaussian Mixture

N. N. Aprausheva[a], N. Mollaverdi[b], and S. V. Sorokin[a]

[a] *Dorodnicyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia*
e-mail: plat@ccas.ru; www2005@ccas.ru

[b] *Department of Computational Mathematics and Cybernetics, Moscow State University, Vorob'evy Gory, ul. Lebedeva 2, Moscow, 119992 Russia*
e-mail: nasermus@yahoo.com

Abstract—Several theorems on sufficient unimodality conditions are formulated for a sum of k normal distributions with the same variance and with different mean values μ_i, $i = 1, \ldots, k$, $2 \le k < \infty$, taken with their a priori probabilities π_i. On the basis of these theorems, estimates for the lower and upper bounds for the mode numbers m are obtained for $k \ge 3$ in the case when the mixture contains k^* components, $2 \le k^* < k$, satisfying the unimodality conditions.

DOI: 10.1134/S1054661806040158

INTRODUCTION

The need for preliminary estimation of the number of modes in Gaussian mixtures is dictated by their wide application in various fields of science and practice, including simulation, pattern recognition, spectroscopy, meteorology, and sea fishery [1–4]. In this study, several theorems on the sufficient unimodality conditions of a mixture of k normal distributions with the same variance σ^2 and with different mean values μ_i, $i = 1, \ldots, k$, $2 \le k < \infty$, taken with their a priori probabilities π_i, are formulated on the basis of the principle of contracting mappings. The unimodality theorems are used as the basis for estimating the upper bound for the mode number m in a mixture with $k \ge 3$ in the case when k^* components of this mixture $(2 \le k^* < k)$ satisfy the unimodality conditions.

1. BASIC THEORETICAL RESULTS

The probability density function of such a mixture has the form

$$f(x) = (\sqrt{2\pi}\sigma)^{-1}\sum_{i=1}^{k}\pi_i\exp[-(x-\mu_i)^2(2\sigma^2)^{-1}],$$
$$x \in R, \quad R = (-\infty, +\infty), \quad \pi_i > 0, \quad \sum_{i=1}^{k}\pi_i = 1. \tag{1.1}$$

The distribution parameters $\pi_1, \ldots, \pi_k, \mu_1, \ldots, \mu_k$, and σ^2 in formula (1.1) are known. The mode of function $f(x)$ (the point of its local maximum) is a root of the equation $f'_x(x) = 0$, which is equivalent to the equation

$$x = \varphi(x), \tag{1.2a}$$

$$\varphi(x) = \left\{\sum_{i=1}^{k}\pi_i\mu_i\exp[-(x-\mu_i)^2(2\sigma^2)^{-1}]\right\} \times \left\{\sum_{i=1}^{k}\pi_i\exp[-(x-\mu_i)^2(2\sigma^2)^{-1}]\right\}^{-1}. \tag{1.2b}$$

Obviously, the number of modes in the mixture under investigation satisfies the inequality [3]

$$1 \le m \le k.$$

For definiteness, we set

$$\mu_1 < \mu_2 < \ldots < \mu_k.$$

In [5], the following proposition is proved.

Theorem 1. All fixed points (FPs) of operator φ lie in the interval (μ_1, μ_k). If operator φ has only one FP, it is a mode of function $f(x)$.

It is known [6] that operator φ has only one fixed point on segment $[\mu_1, \mu_k]$, if it is a contracting operator on this segment; operator φ is a contracting operator on segment $[\mu_1, \mu_k]$ if

$$\varphi'_x(x) < 1. \tag{1.3}$$

Derivative $\varphi'_x(x)$ has the form

$$\varphi'_x(x) = \left(\sum_{s>i}p_{si}^2\psi_s\psi_i\right)\left(\sum_{j=1}^{k}\psi_j\right)^{-2}, \tag{1.4a}$$
$$i = 1, \ldots, k-1, \quad s = 2, \ldots, k,$$

$$\psi_j = \pi_j\exp(-(x-\mu_j)^2(2\sigma^2)^{-1}), \; j = 1, \ldots, k, \tag{1.4b}$$

Received December 24, 2004

$$\rho_{si} = |\mu_s - \mu_i|\sigma^{-1}, \quad i < s, \tag{1.4c}$$

where ρ_{si} is the Mahalanobis distance. In this case, the unimodality condition (1.3) for the mixture assumes the form

$$\sum_{s>i}(\rho_{si}^2 - 2)\psi_s\psi_i - \sum_{j=1}^{k}\psi_j^2 < 0. \tag{1.5}$$

This inequality was used for deriving the theorems of unimodality for the given mixture, which will be formulated below. The proofs of Theorems 3 and 4 are given in [1].

Theorem 2. For $k \geq 2$, a Gaussian mixture is unimodal if

$$\rho_{k1}^2 \leq 2. \tag{1.6}$$

Proof. The sufficient conditions for operator φ to be contracted on segment $[\mu_1, \mu_k]$ are as follows: inequality (1.5) holds if all distances ρ_{si} satisfy the inequality

$$\rho_{si}^2 \leq 2, \; i < s, \; i = 1, 2, \ldots, k-1, \; s = 2, 3, \ldots, k. \tag{1.7}$$

On the basis of the third axiom of metric [6], inequalities (1.7) are equivalent to inequality (1.6). If inequality (1.6) holds, inequality (1.5) is valid, which is the sufficient unimodality condition for the mixture.

Corollary 2.1 For $k \geq 3$, if there exists a value of $\rho_{s,\,s+t}^2$ such that

$$\rho_{s,\,s+t}^2 \leq 2, \quad s \geq 1, \quad t \geq 1, \quad s+t \leq k,$$

the number m of modes in the mixture satisfies the inequality

$$m \leq k - t.$$

Theorem 3. For $k = 2$, a mixture is unimodal if

$$\rho_{21}^2 \leq 4.$$

Corollary 3.1. For $k \geq 3$, if there exists at least one value

$$\rho_{s,\,s+1}^2 \leq 4, \quad s \in \{1, 2, \ldots, k-1\},$$

the number m of modes satisfies the inequality

$$m \leq k - 1.$$

Corollary 3.2. For $k \geq 4$, if set $p = \{\rho_{i,\,i+1}\}$, $i = 1, 2, \ldots, k-1$, contains at least one nonempty subset P_1,

$$P_1 = \{\rho_{i,\,i+1}, \rho_{j,\,j+1}, \ldots, \rho_{s,\,s+1}, \rho_{t,\,t+1}\}, \tag{1.8a}$$

consisting of n_1 elements satisfying the conditions

$$\begin{gathered}\rho_{i,\,i+1} \leq 2, \; \rho_{j,\,j+1} \leq 2, \ldots, \rho_{s,\,s+1} \leq 2, \; \rho_{t,\,t+1} \leq 2, \\ j \leq i+2, \ldots, s \geq j+2, \quad t \geq s+2, \\ i \geq 1, \quad t+1 \leq k,\end{gathered} \tag{1.8b}$$

and

$$\begin{gathered}\rho_{i+1,\,i+2} \geq \rho_0, \ldots, \rho_{j-1,\,j} \geq \rho_0, \\ \rho_{j+1,\,j+2} \geq \rho_0, \ldots, \rho_{t-1,\,t} \geq \rho_0, \quad \rho_{t+1,\,t+2} \geq \rho_0,\end{gathered} \tag{1.8c}$$

where ρ_0 is a sufficiently large positive number, $\rho_0 > 2$, and the number m of modes in the mixture satisfies the inequality

$$m \leq k - n_1. \tag{1.9}$$

It was found experimentally that $\rho_0 = 6$. The maximum number of elements in set P_1 is defined as

$$n_{\max} = E[2^{-1}k + \varepsilon], \tag{1.10}$$

where ε is a small positive number, $0 < \varepsilon \leq 0.01$, and $E[y]$ is the integer part of y. The number of elements of set P_1 satisfies the inequalities

$$1 \leq n_1 \leq n_{\max}.$$

Theorem 4. For $k = 2$, $\rho_{12}^2 > 4$, the mixture is unimodal if

$$\begin{gathered}\left|\ln(\pi_1\pi_2^{-1})\right| \geq \rho_{12}^2 2^{-1} + 2\ln[(\rho_{12} + \sqrt{\rho_{12}^2 - 4})2^{-1}], \\ \pi_1 \neq \pi_2.\end{gathered} \tag{1.11}$$

Corollary 4.1. For $k \geq 3$, if there exists at least one value

$$\rho_{i,\,i+1}^2 > 4, \quad i \in \{1, 2, \ldots, k-1\},$$

and if the triple $\rho_{i,\,i+1}$, π_i, π_{i+1} satisfies conditions (1.11), the number of modes in the mixture satisfies the inequality

$$m \leq k - 1.$$

Corollary 4.2. For $k \geq 4$, if set $P = \{\rho_{i,\,i+1}\}$, $i = 1, 2, \ldots, k-1$, has at least one nonempty subset P_2,

$$P_2 = \{\rho_{i,\,i+1}, \rho_{j,\,j+1}, \ldots, \rho_{s,\,s+1}, \rho_{t,\,t+1}\}, \tag{1.12a}$$

consisting of n_2 elements satisfying the conditions

$$\begin{gathered}\rho_{i,\,i+1} > 2, \rho_{j,\,j+1} > 2, \ldots, \\ \rho_{s,\,s+1} > 2, \rho_{t,\,t+1} > 2, \; i \geq 1, \; t+1 \leq k, \\ j \geq i+2, \ldots, s \geq j+2, t \geq s+2,\end{gathered} \tag{1.12b}$$

and if for each triple of quantities

$$\begin{gathered}\rho_{i,\,i+1}, \pi_i, \pi_{i+1}; \\ \rho_{j,\,j+1}, \pi_j, \pi_{j+1}; \ldots, \rho_{t,\,t+1}, \pi_t, \pi_{t+1}\end{gathered}$$

inequalities (1.11) hold and

$$\begin{gathered}\rho_{i+1,\,i+2} \geq \rho_0, \\ \rho_{j+1,\,j+2} \geq \rho_0, \ldots, \rho_{t+1,\,t+2} \geq \rho_0,\end{gathered} \tag{1.12c}$$

where ρ_0 is a sufficiently large positive number, $\rho_0 > 2$, we have the following inequality for the number m of modes in the mixture:

$$m \le k - n_2. \quad (1.13)$$

It was found experimentally that $\rho_0 = 6$.

The maximum number $n_{\max}$ of elements in set P_2 is defined by formula (1.10); the number n_2 of elements in set P_2 satisfies the inequalities

$$1 \le n_2 \le n_{\max}.$$

Theorem 5. For $k \ge 3$ and $\rho_{k1}^2 > 2$, a mixture is unimodal if

$$\sum_{\rho_{si} \in P_3} (\rho_{si}^2 - 2)\psi_s(x)\psi_i(x) < \sum_{\rho_{si} \in P_4} (2 - \rho_{si}^2)\psi_s(x)\psi_i(x) + \sum_{i=1}^{k} \psi_i^2(x), \quad (1.14a)$$

$$P_3 = \{\rho_{si};\, \rho_{si}^2 > 2\}, \quad P_4 = \{\rho_{si};\, \rho_{si}^2 \le 2\}, \quad s > i, \quad s = 2, \ldots, k, \quad i = 1, \ldots, k-1. \quad (1.14b)$$

Proof. Inequality (1.14a) follows from inequality (1.5), which represents sufficient conditions for the compactness of operator φ and of the uniqueness of its FPs, which is a mode of function $f(x)$ in accordance with Theorem 1.

Inequality (1.14a) is inconvenient for defining the unimodality of function $f(x)$ since both sides of this inequality are functions of x. Consequently, it is expedient to find the relations between the upper bounds of both of its sides.

We will prove auxiliary theorems for two continuous functions $a_1(x)$ and $a_2(x)$, which are positive-definite on segment $[\mu_1, \mu_k]$,

$$a_1(x) > 0, \quad a_2(x) > 0, \quad (1.15)$$

and for their upper bounds b_1 and b_2,

$$b_1 > a_1(x), \quad b_2 > a_2(x). \quad (1.16)$$

Lemma 1. If two continuous positive-definite functions $a_1(x)$ and $a_2(x)$ and their upper bounds b_1 and b_2 satisfy inequalities

$$b_1 < b_2, \quad (1.17)$$

$$b_2 - b_1 \le a_2(x) - a_1(x) \quad (1.18)$$

on segment $[\mu_1, \mu_k]$, then

$$a_1(x) < a_2(x). \quad (1.19)$$

Proof. We write inequality (1.18) in the form

$$a_1(x) \le a_2(x) - b_2 + b_1.$$

The last inequality leads to inequality (1.19) provided that $-b_2 + b_1 < 0$ (i.e., if inequality (1.17) is valid), QED.

Lemma 2. If two continuous positive-definite functions $a_1(x)$ and $a_2(x)$ and their upper bounds b_1 and b_2 satisfy the conditions

$$a_1(x) < a_2(x), \quad (1.20)$$

$$b_2 - b_1 \ge a_2(x) - a_1(x) \quad (1.21)$$

on segment $[\mu_1, \mu_k]$, then the following inequality holds:

$$b_1 < b_2. \quad (1.22)$$

Proof. We write inequality (1.22) in the form

$$b_1 \le b_2 - a_2(x) + a_1(x).$$

From the latter inequality, we obtain inequality (1.22) only if inequality (1.20) holds, i.e.,

$$-a_2(x) + a_1(x) < 0,$$

QED.

To use Lemmas 1 and 2 for formulating the unimodality conditions for a mixture, we introduce the following notation for the functions appearing in inequality (1.14a):

$$a_1(x) = \sum_{\rho_{si} \in P_3} (\rho_{si}^2 - 2)\psi_s\psi_i, \qquad a_2(x) = \sum_{\rho_{si} \in P_4} (2 - \rho_{si}^2)\psi_s\psi_i + \sum_{j=1}^{k} \psi_j^2. \quad (1.23)$$

Using expression (1.4b), we can easily find the upper bounds for functions $a_1(x)$ and $a_2(x)$ in expression (1.23):

$$b_1 = \sum_{\rho_{si} \in P_3} (\rho_{si}^2 - 2)\pi_s\pi_i, \qquad b_2 = \sum_{\rho_{si} \in P_4} (2 - \rho_{si}^2)\pi_s\pi_i + \sum_{j=1}^{k} \pi_j^2. \quad (1.24)$$

Then Theorem 5 and Lemma 1 lead to the following proposition.

Theorem 6. For $k \ge 3$ and $\rho_{k1}^2 > 2$, a mixture is unimodal if the following two inequalities hold:

$$\sum_{\rho_{si} \in P_3} (\rho_{si}^2 - 2)\pi_s\pi_i < \sum_{\rho_{si} \in P_4} (2 - \rho_{si}^2)\pi_s\pi_i + \sum_{j=1}^{k} \pi_j^2, \quad (1.25)$$

$$\sum_{\rho_{si} \in P_4} (2 - \rho_{si}^2)\pi_s\pi_i + \sum_{j=1}^{k} \pi_j^2 - \sum_{\rho_{si} \in P_3} (\rho_{si}^2 - 2)\pi_s\pi_i \le \sum_{\rho_{si} \in P_4} (2 - \rho_{si}^2)\psi_s(x)\psi_i(x) \quad (1.26)$$

Table

No. of mixture	Mixture parameters						Values of DR		
	ρ_{21}	ρ_{32}	ρ_{31}	π_1	π_2	m	b_1	b_2	p
1	1.41	0.59	2	0.33	0.33	1	0.22	0.52	1
2	1.80	0.20	2	0.50	0.25	1	0.40	0.50	1
3	1.00	1.20	2.20	0.10	0.20	1	0.20	0.64	1
4	1.41	1.41	2.82	0.60	0.20	1	0.71	0.44	1
5	1.41	1.41	2.82	0.45	0.10	2	1.20	0.42	2
6	1.50	1.50	3.00	0.33	0.33	1	0.83	0.33	1
7	1.50	1.50	3.00	0.40	0.20	2	1.16	0.36	2
8	2.00	3.00	5.00	0.80	0.10	2	2.07	0.66	2
9	2.50	2.50	5.00	0.40	0.20	3	4.36	0.36	3
10	2.50	2.50	5.00	0.33	0.33	3	3.50	0.33	3

$$+\sum_{j=1}^{k}\psi_j^2(x)-\sum_{\rho_{si}\in P_3}(\rho_{si}^2-2)\psi_s(x)\psi_i(x);$$

in these inequalities, sets P_3 and P_4 are defined in (1.14b).

Corollary 6.1. For $k \geq 4$ and $\rho_{k1}^2 > 2$, if the set of k mixture components contains a subset k^* of components with numbers $i, i+1, \ldots, i+k^*-1$, $1 \leq i < k-k^*+1$, $3 \leq k^* < k$, which satisfy conditions (1.25) and (1.26), the number m of modes of the mixture satisfies the inequality

$$m \leq k - k^* + 1. \quad (1.27)$$

Theorem 5 and Lemma 2 lead to the following proposition.

Theorem 7. For $k \geq 3$ and $\rho_{k1}^2 > 2$, inequalities (1.14a) and

$$\sum_{\rho_{si}\in P_4}(2-\rho_{si}^2)\pi_s\pi_i+\sum_{j=1}^{k}\pi_j^2-\sum_{\rho_{si}\in P_3}(\rho_{si}^2-2)\pi_s\pi_i$$

$$\geq \sum_{\rho_{si}\in P_4}(2-\rho_{si}^2)\psi_s(x)\psi_i(x) \quad (1.28)$$

$$+\sum_{j=1}^{k}\psi_j^2(x)-\sum_{\rho_{si}\in P_3}(\rho_{si}^2-2)\psi_s(x)\psi_i(x)$$

lead to inequality (1.25).

Thus, for $k \geq 3$ and $\rho_{k1}^2 > 2$, sufficient conditions of unimodality of a mixture, viz., inequalities (1.25) and (1.26), are obtained from the principle of contracting mappings. It is practically impossible to verify the fulfillment of inequality (1.26) at each point $x \in [\mu_1, \mu_k]$.

2. RESULTS OF EXPERIMENTS

Numerous experiments performed for $k = 3, 4, 10, 20, 40$, and 100 proved that a mixture is unimodal if inequality (1.25) holds; operator φ can be either contracting or noncontracting. If inequality (1.25) is violated, the mixture can be either unimodal or multimodal. In this case, for correct decision making, it is necessary to calculate the values of function $f'_x(x)$ at points

$$\mu_1, \mu_1+h, \mu_1+2h, \ldots, \mu_1+lh, \mu_k \quad (2.1)$$

and calculate the number of sign reversals from plus to minus (this number will be denoted by p). If $p = 1$, the mixture is probably unimodal. For $p \geq 2$, the mixture is multimodal. For the number m of modes in the mixture, the following inequality holds:

$$m \geq p. \quad (2.2)$$

It was established experimentally that the optimal value of h in (2.1) is equal to $4^{-1}\sigma$ since for this value the equality holds in (2.2):

$$m = p. \quad (2.3)$$

The table contains the results obtained for ten three-component mixtures ($k = 3$): their parameters ρ_{21}, ρ_{32}, ρ_{31}, π_1, π_2, and m (the number m of modes was obtained using the computational algorithm for fixed points of function $f(x)$ [7]), as well as the values of decision rules (DRs) b_1, b_2, and p (1.25), (2.1)–(2.3).

In accordance with DR (1.25), mixtures 1–3 are unimodal ($b_1 < b_2$), and the number of modes in mixtures 4–10 is not defined ($b_1 > b_2$). Applying the second decision rule (2.1)–(2.3) to these modes, we obtained the exact number of modes ($m = p$).

Although we have determined the exact number m of modes for each mixture from the table using our criteria, we will still illustrate the estimation of the upper bound for m. On the basis of Corollary 3.1 for mixtures 1–8, we have $m \leq 2$ since $\rho_{21} \leq 2$.

As regards the determination of a large value of ρ_0 ($\rho_0 = 6$) in the wording of Corollaries 3.2 and 4.2, we proceeded from the behavior of each component of function $f(x)$,

$$f_i(x) = (\sqrt{2\pi}\sigma)^{-1}\exp((x-\mu_i)^2(2\sigma^2)^{-1}),$$
$$i = 1, 2, \dots, k,$$

namely,

$$\lim_{x \to \pm\infty} f_i(x) = 0.$$

Each function $f_i(x)$ has inflections at points $x_i = \mu_i \pm \sigma$. In addition, the landmark was the 3σ rule for the Gaussian distribution [8].

Numerical experiments aimed at determining the value of ρ_0 were performed for various mixtures with $k = 4$, 10, 20, 40, and 100. In this case, the values of parameters π_i, $\rho_{i,\,i+1}$, $i = 1, 2, \dots, k - 1$ were defined as follows: $\pi_i = k^{-1}$,

$$\rho_{i,\,i+1} = \begin{cases} 2 & \text{for odd } i, \\ 6 & \text{for even } i. \end{cases}$$

In all experiments, the number of modes calculated in accordance with the algorithm developed in [7] was equal to $2^{-1}k$, which coincides with formula (1.10).

CONCLUSIONS

The main result of this research is the refinement of the upper bound for the number of modes in the simplest Gaussian mixture, which was obtained as corollaries of the theorems on the unimodality of the mixture. The results formulated here can be helpful in solving problems of approximation, pattern recognition, spectroscopy, and other fields of science.

REFERENCES

1. E. A. Patrick, *Fundamentals of Pattern Recognition* (Prentice Hall Englewood Cliffs, NJ, 1972).
2. Yu. I. Zhuravlev and I. V. Isaev, "Construction of Recognition Algorithms, which are Correct for a Preset Control Sample," Zh. Vychisl. Mat. i Mat. Fiz. **19** (3), 726–738 (1979).
3. M. A. Carreira-Perpiñán, "Mode-Finding for Mixture of Gaussian Distributions," IEEE Trans. on Pattern Analys. and Mach. Intell. **22** (11), 1318–1323 (2000).
4. L. G. Malinovskii, *Analysis of Statistical Relations. Construction–Model Approach* (Nauka, Moscow, 2002) [in Russian].
5. N. N. Aprausheva and S. V. Sorokin, "On Unimodality of the Simplest Gaussian Mixture," Zh. Vychisl. Mat. i Mat. Fiz. **44** (5), 838–846 (2004).
6. A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis* (Nauka, Moscow, 1972) [in Russian].
7. N. N. Aprausheva, N. Mollaverdi, S. V. Sorokin, and N. A. Torkhova, *On Modes of the Simplest Gaussian Mixture* (Computer Center of the Russian Academy of Sciences, Moscow, 2003) [in Russian].
8. H. Cramer, *Mathematical Methods of Statistics* (Almquist and Wiskell, Stockholm, 1946).

Aprausheva Nina Nikolaevna (b. 1940). Graduated from the Moscow State University (Department of Mechanics and Mathematics) in 1966. In 1982, defended the Cand. Sci. thesis "On the Application of Mixtures of Normal Distributions in Pattern Recognition."

Senior scientist at Dorodnicyn Computing Center, Russian Academy of Sciences.

Scientific interests: mathematical statistics, computational mathematics, mathematical methods in pattern recognition, and decision theory (about 70 publications).

Member of Association of Pattern Recognition.

Mollaverdi Naser (b. 1967). Graduated from the Isfahan University of Technology (Iran) in 1990. In 1995, graduated from the Educational University (Teheran, Iran).

Third-year scholar at Moscow State University (Department of Computational Mathematics and Cybernetics).

Scientific interests: computational mathematics, pattern recognition, programming, and statistics (12 publications).

Sorokin Stanislav Viktorovich (b. 1947). Graduated from Moscow State University (Department of Computational Mathematics and Cybernetics) in 1971.

Programmer at Dorodnicyn Computing Center, Russian Academy of Sciences.

Scientific interests: computational mathematics, system programming, and pattern recognition (15 publications).