On the Uni- and Bimodality of a Two-Component Gaussian Mixture

N. N. Aprausheva and S. V. Sorokin

Dorodnitcyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia
e-mail: plat@ccas.ru, www2008@ccas.ru

Abstract—Several sufficient conditions are formulated for the uni- and bimodality of a mixture of two Gaussian distributions with equal variances σ^2 and different expectation values μ_i, $i = 1, 2$. An equation governing all the degenerate critical inflection points for the probability density $f(x)$ of the mixture is derived by a statistical method. This equation describes the boundary of the uni- and bimodality domains of $f(x)$.

DOI: 10.1134/S1054661808040056

INTRODUCTION

Gaussian mixtures have found wide applications in various fields of science and practice [1, 7, 8]. This motivated the task of estimating the number of modes in a mixture depending on the distribution parameters. We solve this problem for a mixture of two Gaussian distributions with equal variances σ^2 and different expectation values μ_i, $i = 1, 2$.

1. STATEMENT OF THE PROBLEM AND SOLUTION METHODS

The probability density of such a mixture is given by the formula

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

where $x \in R$ and π_i is an a priori probability of the ith component such that $\pi_i \in (0, 1)$ and $\pi_1 + \pi_2 = 1$.

The mode of the smooth function $f(x)$ is its critical point; it is the root of the equation

$$f'_x(x) = 0,$$
$$f'_x(x) = (\sqrt{2\pi}\sigma^3)^{-1}\sum_{i=1}^{2}\pi_i(\mu_i - x)f_i(x). \tag{2}$$

Equation (2) can be reduced to an equivalent one

$$x = \varphi(x), \tag{3}$$

$$\varphi(x) = \left(\sum_{i=1}^{k}\pi_i\mu_i f_i(x)\right)\left(\sum_{i=1}^{k}\pi_i f_i(x)\right)^{-1}. \tag{4}$$

As was proved in [2], if φ in (4) is a contraction operator, then Eq. (2) has a single root and the function $f(x)$ is unimodal. Below are the results from [2] we need in our presentation.

Theorem 1. For $k = 2$, the function $f(x)$ is unimodal if $\rho^2 \le 4$, where $\rho^2 = (\mu_2 - \mu_1)^2\sigma^{-2}$ is the Mahalanobis distance.

Theorem 2. If $k = 2$, $\rho^2 > 4$, and $\pi_1 \ne \pi_2$, then $f(x)$ is unimodal if

$$\left|\ln(\pi_1\pi_2^{-1})\right| \ge 2^{-1}\rho^2 + 2\ln(2^{-1}(\rho + \sqrt{\rho^2 - 4})). \tag{5}$$

Theorem 3. If $k = 2$, $\rho^2 > 4$, and $\pi_1 = \pi_2$, then $f(x)$ is bimodal.

If an equality holds in (5), then we have an equation for the boundary of the contraction domain of φ for $\rho^2 > 4$ and $\pi_1 \ne \pi_2$. Denoting the right-hand side of (5) by $\psi_3(\rho)$, i.e.,

$$|\psi_3(\rho)| = 2^{-1}\rho^2 + 2\ln(2^{-1}(\rho + \sqrt{\rho^2 - 4})), \tag{6}$$

we write this boundary equation as

$$\left|\ln(\pi_1\pi_2^{-1})\right| = \psi_3(\rho).$$

Numerous experiments have shown that, if (5) in the assumptions of Theorem 2 is replaced by the reverse inequality, i.e., by

$$\left|\ln(\pi_1\pi_2^{-1})\right| < 2^{-1}\rho + 2\ln(2^{-1}(\rho + \sqrt{\rho^2 - 4})),$$

then the number s of critical points of $f(x)$ satisfies

$$1 \le s \le 3.$$

The function $f(x)$ is unimodal for $1 \le s \le 2$ and bimodal for $s = 3$. If $s = 2$, then one of the of critical

Received May 12, 2008

Values of ψ_3, ψ_2, and ψ for various ρ

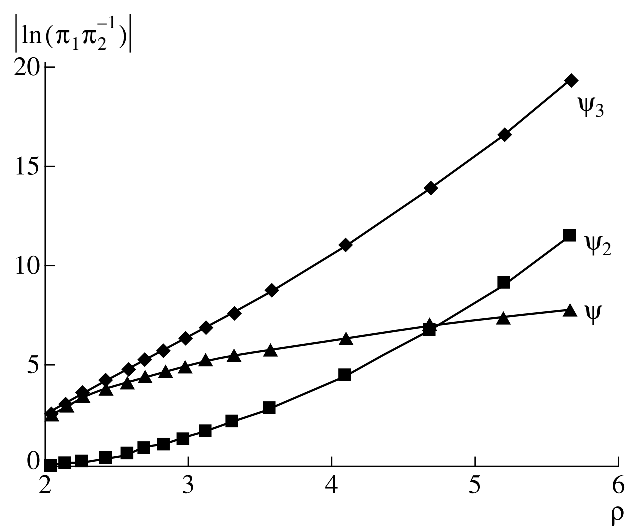
No.	ρ	ψ_3	ψ_2	ψ
1	5.6467	19.3380	11.5129	7.8259
2	5.1894	16.6794	9.2102	7.4691
3	4.6787	13.9328	6.9068	7.0261
4	4.0857	11.0292	4.5951	6.4341
5	3.5807	8.7836	2.9444	5.8391
6	3.3146	7.6764	2.1972	5.4792
7	3.1314	6.9409	1.7346	5.2062
8	2.9807	6.3498	1.3863	4.9635
9	2.8448	5.8255	1.0986	4.7269
10	2.7146	5.3285	0.8473	4.4812
11	2.5631	4.8286	0.6190	4.2095
12	2.4427	4.2907	0.4055	3.8852
13	2.2791	3.6418	0.2007	3.4412
14	2.1522	3.0914	0.0800	3.0113
15	2.0606	2.6141	0.0200	2.5941

points of $f(x)$ is an inflection point; i.e., it satisfies the equation

$$f''_{xx}(x) = 0,$$

$$f''_{xx}(x) = (\sqrt{2\pi}\sigma^3)^{-1}\sum_{i=1}^{2}\pi_i[(x-\mu_i)^2-\sigma^{-2}]f_i(x). \quad (7)$$

Any critical inflection point of the function is a degenerate critical point [3]. The expressions for $f'_x(x)$ and $f''_{xx}(x)$ and Eqs. (2) and (7) yield the equation for all the degenerate critical point of $f(x)$:



Plots of $\psi_3(\rho)$, $\psi_2(\rho)$, and $\psi(\rho)$.

$$x^2 = \left(\sum_{i=1}^{2}\pi_i\mu_i^2 f_i(x)\right)\left(\sum_{i=1}^{2}\pi_i f_i(x)\right)^{-1} - \sigma^2.$$

These results imply that the Gaussian mixture under study is not a Morse function [3].

The equation for the set of critical inflection points of $f(x)$ is one describing the boundary of its uni- and bimodality domains. An equation of the form $|\ln(\pi_1\pi_2^{-1})| = \psi_2(\rho)$ for this boundary is determined by a statistical method.

Fifteen critical inflection points were experimentally found on the number line. The values of ρ and $\psi_2 = |\ln(\pi_1\pi_2^{-1})|$ at each point are presented in the table. Additionally, the table lists the values of $|\ln(\pi_1\pi_2^{-1})| = \psi_3(\rho)$ and $\psi = \psi_3 - \psi_2$, where $\psi_3(\rho)$ is given by (6). The functions $\psi_3(\rho)$, $\psi_2(\rho)$, and $\psi(\rho)$ are also plotted in the figure.

RESULTS

The function $\psi(\rho)$ is approximated using quadratic regression [4, 6]. Specifically, for each point $A_i(\rho, \psi)$, we use the equation

$$\psi_i = a\rho_i^2 + b\rho_i + c + \varepsilon_i, \quad i = 1, 2, \ldots, 15, \quad (8)$$

where ε_i is the deviation of A_i from the parabola

$$\psi = a\rho^2 + b\rho + c. \quad (9)$$

Applying the least squares method [4, 6] to system (8) yields the system of normal equations

$$\begin{cases} \left(\sum_{i=1}^{15}\rho_i^4\right)a + \left(\sum_{i=1}^{15}\rho_i^3\right)b + \left(\sum_{i=1}^{15}\rho_i^2\right)c = \sum_{i=1}^{15}\psi_i\rho_i^2 \\ \left(\sum_{i=1}^{15}\rho_i^3\right)a + \left(\sum_{i=1}^{15}\rho_i^2\right)b + \left(\sum_{i=1}^{15}\rho_i\right)c = \sum_{i=1}^{15}\psi_i\rho_i \\ \left(\sum_{i=1}^{15}\rho_i^2\right)a + \left(\sum_{i=1}^{15}\rho_i\right)b + 15c = \sum_{i=1}^{15}\psi_i. \end{cases}$$

The calculated coefficients of a, b, and c are substituted into these equations to obtain

$$\begin{cases} 3294.188a + 766.418b + 196.320c = 3294.188 \\ 766.418a + 196.320b + 56.220c = 298.181 \\ 196.320a + 56.221b + 15.000c = 85.872. \end{cases} \quad (10)$$

Solving system (10) by Gaussian elimination [5] produces the parameter estimates $\tilde{a} = -0.327$, $\tilde{b} = 3.867$, and $\tilde{c} = -3.738$. Therefore, $\psi(\rho)$ in (9) is approximated by

$$\tilde{\psi} = -0.327\rho^2 + 3.867\rho - 3.738. \qquad (11)$$

For this approximation, the mean absolute error is $\Delta = 0.101$, the mean relative error is $\delta = 0.024$, and the rms deviation is $\sigma_1 = 0.124$ with $\sigma_1^2 = 0.015$.

Combining the definitions of ψ_3, ψ_2, and ψ with expressions (6) and (11) gives the following approximate equation for the set of degenerate critical inflection points of $f(x)$:

$$\left|\ln(\pi_1\pi_2^{-1})\right| = 0.827\rho^2 - 3.867\rho + 2\ln[2^{-1}(\rho + \sqrt{\rho^2 - 4})] + 3.738, \qquad (12)$$

which is an approximate equation for the boundary of the uni- and bimodality domains of $f(x)$ for $\rho > 2$ and $\pi_1 \neq \pi_2$ in the parameter space (π_1, π_2, ρ).

Using the Chebyshev inequality [4] for the unknown function ψ, namely,

$$P\{|\psi - \tilde{\psi}| < t\sigma_1\} \geq 1 - t^{-2}$$

with $t = 4$ and $\sigma_1 = 0.124$, we obtain the confidence interval

$$P\{\tilde{\psi} - 0.496 < \psi < \tilde{\psi} + 0.496\} \geq 0.9375, \qquad (13)$$

where $\tilde{\psi}$ is given by (11).

Since $\psi = \psi_3 - \psi_2$ and $\tilde{\psi} = \psi_3 - \tilde{\psi}_2$, combining (12) with (13), we find for $\psi_2(\rho) = |\ln(\pi_1\pi_2^{-1})|$ that

$$P\{d + 3.242 < \psi_2 < d + 4.234\} \geq 0.9375, \qquad (14)$$

where

$$d = 0.827\rho^2 - 3.867\rho + 2\ln[2^{-1}(\rho + \sqrt{\rho^2 - 4})].$$

Expression (14) gives a confidence interval for all the critical inflection points of $f(x)$ with a probability greater than or equal to 0.9375. Our argument implies that $f(x)$ is bimodal with a probability greater than or equal to 0.9375 if

$$\left|\ln(\pi_1\pi_2^{-1})\right| < d + 3.242$$

for $\rho > 2$ and $\pi_1 \neq \pi_2$ and is unimodal with the same probability if

$$\left|\ln(\pi_1\pi_2^{-1})\right| > d + 4.234.$$

CONCLUSIONS

For the probability density of the mixture under study, we found an exact equation for all its degenerate critical points lying on the number line and derived an approximate equation for its degenerate critical inflection points in the space of the parameters π_1, π_2, and ρ. A confidence interval was constructed for these points. The uni- and bimodality domains of the mixture were determined.

REFERENCES

1. N. N. Aprausheva and M. B. Radzhabova, *Classification of Raw Cotton Based on a Statistical Algorithm* (Vychisl. Tsentr Akad. Nauk SSSR, Moscow, 1997), pp. 4–5 [in Russian].
2. N. N. Aprausheva and S. V. Sorokin, "On the Unimodality of a Simple Gaussian Mixture," Comput. Math. Math. Phys. **44**, 785–793 (2004).
3. V. I. Arnold, A. N. Varchenko, and S. M. Gusein-Zade, *Singularities of Differentiable Maps* (Nauka, Moscow, 1982; Birkhäuser, Boston, 1985).
4. H. Cramer, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, 1946; Mir, Moscow, 1975).
5. A. G. Kurosh, *Higher Algebra* (Mir, Moscow, 1972; Mir, Moscow, 1980).
6. C. R. Rao, *Linear Statistical Inference and Its Applications* (Nauka, Moscow, 1968; Wiley, New York, 1973).
7. Ya. A. Fomin and G. R. Tarlovskii, *Statistical Theory of Pattern Recognition* (Radio i Svyaz', Moscow, 1986) [in Russian].
8. M. A. Carreira-Perpiñán and C. Williams, "On the Number of Modes of a Gaussian Mixture," Res. Report EDI-INF-RR-0159 (School of Inf. Univ. of Edinburgh, Edinburgh, 2003).

Nina Nikolaevna Aprausheva was born in 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 75 publications). Member of Association of Pattern Recognition and Association of. "Women in science and education." N.N.Aprausheva's publications have been awarded several prizes at the scientific publications contests in Computing Center of Russian Academy of Sciences.

Stanislav Viktorovich Sorokin was born in 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 (20 publications).