Н.Н. Апраушева[1], *С.В. Сорокин*[2] (Учреждение Российской академии наук
Вычислительный центр им. А.А. Дородницына РАН, Москва, Россия
E-mail: [1]plat@ccas.ru, [2]www2009@ccas.ru)

Об унимодальности одномерной гауссовой смеси

Широкое использование гауссовых смесей в различных областях науки и практики вызывает необходимость решения таких задач, как определение их мод и предварительное оценивание их числа [1]–[3]. В этой работе сформулировано несколько достаточных условий унимодальность одномерной гауссовой смеси, компоненты которой имеют равные дисперсии σ^2 и различные математические ожидания μ_i, $i = 1, 2, \ldots, k$, $2 \le k < \infty$.

Плотность вероятности такой смеси выражается формулой

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

$x \in R$, π_i — априорная вероятность i-й компоненты, $\pi_i \in (0,1)$, $\pi_1 + \ldots + \pi_k = 1$.

Без умаления общности полагаем $\mu_1 < \mu_2 < \ldots < \mu_k$. Мода гладкой функции $f(x)$ является ее критической точкой (КТ), т.е. является корнем уравнения

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

Уравнение (2) приводится к равносильному уравнению

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

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

Доказано [4], что если оператор φ в (4) является сжимающим, то уравнение в (2) имеет один корень и функция $f(x)$ унимодальна. Из условия сжатости оператора φ имеем следующие теоремы.

Теорема 1. *При $k = 2$ функция $f(x)$ унимодальна, если $\rho_{21}^2 \le 4$. $\rho_{21}^2 = (\mu_2 - \mu_1)^2\sigma^{-2}$, ρ_{21}^2 — расстояние Махаланобиса.*

Теорема 2. *При $k = 2$ и $\rho^2 > 4$, $\pi_1 \ne \pi_2$, функция $f(x)$ унимодальна, если*

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

Теорема 3. *При $k \geq 3$ функция $f(x)$ унимодальна, если*

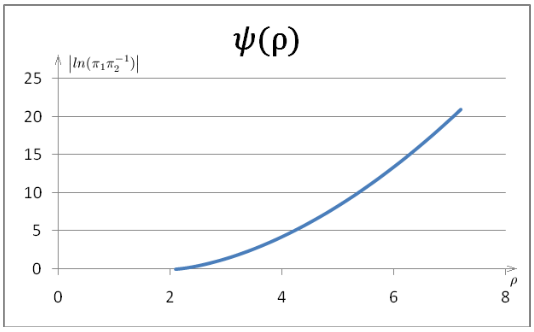
$$\rho_{k1}^2 \leq 4, \quad \sum_{s=2}^{k} \rho_{s,s-1} = \rho_{k1}, \qquad \max_{\rho_{si} \in P} \rho_{si} \leq 2,$$

$$P = \{\rho_{si} : s > i\} \setminus \rho_{k1}, \qquad s = 2, \dots, k, \qquad i = 1, \dots, k-1.$$

Многочисленные эксперименты показали, что при $\rho_{21}^2 > 4$ функция $f(x)$ может быть унимодальной и тогда, когда оператор φ в (4) не является сжимающим. В этой ситуации для частного случая ($k = 2$) нам удалось установить, что границей между семействами унимодальных и бимодальных смесей является множество вырожденных критических точек перегиба (ВКТП) функции $f(x)$ и найти приближенное уравнение этого множества в виде

$$\left|\ln(\pi_1 \pi_2^{-1})\right| = \psi(\rho), \qquad \rho_{21} = \rho. \tag{6}$$

Зафиксировав значение μ_1 и меняя значения μ_2, π_1, нашли 52 ВКТП функции $f(x)$, для каждой из которых вычислялись значения ρ_{21}, $\left|\ln(\pi_1 \pi_2^{-1})\right|$. График функции (6) на отрезке $[2.1, 7.2]$ приведен на рисунке.



Экспериментальную кривую аппроксимировали полиномом 2-й степени,

$$\left|\ln(\pi_1 \pi_2^{-1})\right| = a\rho^2 + b\rho + c.$$

Вычислив неизвестные коэффициенты a, b, c методом наименьших квадратов [5], получим аппроксимирующее выражение

$$\tilde{\psi}(\rho) = 0.579\rho^2 - 1.242\rho + 0.004. \tag{7}$$

При такой аппроксимации функции $\psi(\rho)$ имеем: средняя абсолютная погрешность $\Delta = 0.040$, средняя относительная погрешность $\delta = 0.060$, квадратическое отклонение $\sigma = 0.046$, $\sigma^2 = 0.002$.

Используя неравенство Чебышева

$$P\left\{|\psi(\rho) - \tilde{\psi}(\rho)| < t\sigma\right\} > 1 - t^2,$$

для неизвестной функции $\psi(\rho)$ получим доверительный коридор при $t = 4$, $\sigma = 0.046$

$$P\left\{\tilde{\psi}(\rho) - 0.184 < \psi(\rho) < \tilde{\psi}(\rho) + 0.184\right\} > 0.937$$

с доверительной вероятностью 0.937. В силу (6) последнее неравенство представим в виде

$$P\left\{\tilde{\psi}(\rho) - 0.184 < |\ln(\pi_1\pi_2^{-1})| < \tilde{\psi}(\rho) + 0.184\right\} > 0.937, \qquad (8)$$

в котором $\tilde{\psi}(\rho)$ выражается формулой (7).

Выражение (8) дает доверительный коридор множества вырожденных критических точек перегиба семейства функций $f(x)$ при $k = 2$, $\rho^2 > 4$, $\pi_1 \neq \pi_2$.

Из наших рассуждений следует, что если

$$\left|\ln(\pi_1\pi_2^{-1})\right| > 0.579\rho^2 - 1.242\rho + 0.188,$$

то функция $f(x)$ унимодальна с вероятностью > 0.937, а при выполнении неравенства

$$\left|\ln(\pi_1\pi_2^{-1})\right| < 0.579\rho^2 - 1.242\rho + 0.180$$

она бимодальна с той же вероятностью.

[1] *Апраушева Н.Н., Раджабова М.Б.* Классификация хлобка-сырца по статистическому алготиму. — М.: ВЦ АН СССР, 1990.

[2] *Апраушева Н.Н., Горлач И.А., Желнин А.А., Сорокин С.В.* Об опыте автоматического статистического распознавания облачности // ЖВМиМФ. — 1998. — Т. 38, № 10. — С. 1788–1792.

[3] *Carreira-Perpiñán M.A., Williams C.* On the Number of Modes of a Gaussian Mixture. Inform. Res. Report EDI-INF-RR-0159. School of Inf. Univ. of Edinburg, 2003.

[4] *Апраушева Н.Н., Сорокин С.В.* Об унимодальности простейшей гауссовой смеси // ЖВМиМФ. — 2004. — Т. 44, № 5. — С. 838–846.

[5] *Крамер Г.* Математические методы статистики. — М.: Мир, 1975.