Gaussian Scale Mixtures

Miguel Martins Felgueiras

CEAUL e ESTG do Instituto Politécnico de Leiria

Abstract

We investigate gaussian mixtures with independent components, very effective in modeling real data. Our main purpose is to identify useful approximations to F_X, namely of the Pearson family. A shifted and scaled t-Student approximation was obtained, and as a side result we develop a test on the equality of mean values.

keywords: gaussian mixtures, mean equality, Pearson system.

AMS: 60E05

1 Introduction

Since Pearson (1894) early works that gaussian mixtures occupy an important place on mixtures study, mainly because of worldwide use of gaussian distribution.

Nowadays, finite gaussian mixtures have a several number of applications, from Biology to Economy, passing through Computer and Astronomy (Frühwirth-Schnatter, 2006) and are a very flexible way to model data.

Even though Teicher (1961, 1963) showed that gaussian mixtures are identifiable, mixture parameters estimation is still a difficult problem. Usually we have $3N-1$ unknown parameters, where N is the number of subpopulations, so even for a small subpopulations number there are many unknown parameters to estimate. Besides, usual methods of estimation, like the *expectation-maximization algorithm* (EM) (Hasselblad, 1966, Dempster *et al*, 1977), may lead to poor estimates even when the initial estimates are equal to the true parameters value (Frühwirth-Schnatter, 2006, Verbeek *et al*, 2003). This highlights that the EM algorithm only guarantees that a local maximum is reached and quickly becomes inadequate for a large number of subpopulations.

When it is possible, simplifications like subpopulations equal means lead to a more parsimonious solution (with $2N$ unknown parameters), useful when the number of subpopulations is large.

As gaussian mixtures sharing the same subpopulations mean are always unimodal, we investigate in this paper their approximation to an appropriate member of the Pearson system, which contains at most four unknown parameters. A shifted and scaled t-Student approximation was obtained, diminishing the number of parameters we need to estimate. As a side result we develop a test on the equality of mean values.

2 Definition, Moments and Cumulants

A random variable X is a convex gaussian mixture when

$$f_X(x) = \sum_{j=1}^{N} w_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right\}, \quad \sigma_j > 0, w_j > 0, \sum_{j=1}^{N} w_j = 1. \tag{1}$$

Its characteristic function,

$$\varphi_X(t) = \sum_{j=1}^{N} w_j \varphi_{X_j}(t) = \sum_{j=1}^{N} w_j \exp\left\{it\mu_j - \frac{t^2\sigma_j^2}{2}\right\}, \tag{2}$$

can be used to calculate distribution cumulants and moments, since cumulants generation function is

$$\begin{aligned}\ln[\varphi_X(-it)] &= \mu t + \left(\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)-\mu^2\right)\frac{t^2}{2!}+ \\ &+\left(\sum_{j=1}^{N} w_j\left(\mu_j^3+3\mu_j\sigma_j^2\right)+2\mu^3-3\mu\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)\right)\frac{t^3}{3!}+ \\ &+\left(\begin{array}{c}\sum_{j=1}^{N} w_j\left(\mu_j^4+6\mu_j^2\sigma_j^2+3\sigma_j^4\right)-4\mu\sum_{j=1}^{N} w_j\left(\mu_j^3+3\mu_j\sigma_j^2\right) \\ +12\mu^2\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)-3\left(\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)\right)^2-6\mu^4\end{array}\right)\frac{t^4}{4!}+ \\ &+O\left(t^5\right),\end{aligned}$$

and therefore its cumulants are

$$\begin{aligned}\kappa_1 &= \sum_{j=1}^{N} w_j\mu_j = \mu \qquad (3)\\ \kappa_2 &= \sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)-\mu^2 \\ \kappa_3 &= \sum_{j=1}^{N} w_j\left(\mu_j^3+3\mu_j\sigma_j^2\right)+2\mu^3-3\mu\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right) \\ \kappa_4 &= \sum_{j=1}^{N} w_j\left(\mu_j^4+6\mu_j^2\sigma_j^2+3\sigma_j^4\right)-4\mu\sum_{j=1}^{N} w_j\left(\mu_j^3+3\mu_j\sigma_j^2\right)+ \\ &\quad +12\mu^2\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)-3\left(\sum_{j=1}^{N} w_j\left(\mu_j^2+\sigma_j^2\right)\right)^2-6\mu^4.\end{aligned}$$

The skewness and the kurtosis are given by

$$\beta_1 = \frac{\kappa_3}{\kappa_2^{\frac{3}{2}}} \qquad (4)$$

and

$$\beta_2 = \frac{\kappa_4}{\kappa_2^2}+3. \qquad (5)$$

β_1 and β_2 have therefore complex expressions, and henceforth unimodal mixtures will have good approximations from members of different Pearson type distributions, depending on relationships among the parameters involved.

3 Gaussian Mixtures With Subpopulations Equal Mean (Scale Mixture)

When all the subpopulations have the same mean, it may be difficult, if the phenomenon in study isn't very well known, to see immediately that the data was generated by a mixture, because the data should be unimodal. Note that in this situation

$$f_X(x) = \sum_{j=1}^{N} w_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_j}\right)^2\right\} \tag{6}$$

and

$$f'_X(x) = -(x-\mu)\sum_{j=1}^{N} w_j \frac{1}{\sqrt{2\pi}\sigma_j^2} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_j}\right)^2\right\},$$

leading to a single mode in $x = \mu$.

Under this circunstancies, multimodal data should be a sufficient condition to reject means equality. The main problem arises when data is unimodal, because now the mixture subpopulations means may be or may not be equal. Attending to equations (3) and (4),

$$\begin{aligned} \mu'_1 &= \mu \\ \mu_2 &= \sum_{j=1}^{N} w_j \sigma_j^2 \\ \mu_3 &= 0 \\ \mu_4 &= 3\sum_{j=1}^{N} w_j \sigma_j^4 \end{aligned}$$

and

$$
\begin{aligned}
\beta_1 &= 0 \\
\beta_2 &= \frac{3\sum_{j=1}^{N} w_j \sigma_j^4}{\left(\sum_{j=1}^{N} w_j \sigma_j^2\right)^2}.
\end{aligned}
$$

Using Pearson system, the scaled and shifted mixture can always be approached to a t-Student distribution, as the following theorem states.

Theorem 1.

Let X be a N Gaussian mixture with equal mean μ. Then

$$\alpha\,(X-\mu) \overset{\circ}{\sim} t_{(\nu)}, \tag{7}$$

where

$$\alpha = \sqrt{\frac{1-b_2}{b_0}} \tag{8}$$

and

$$\nu = \frac{1-b_2}{b_2}. \tag{9}$$

In the above expressions, b_0 and b_2 are the Pearson system constants,

$$b_0 = \frac{\mu_2\,(4\beta_2 - 3\beta_1^2)}{10\beta_2 - 18 - 12\beta_1^2} \tag{10}$$

and

$$b_2 = \frac{2\beta_2 - 3\beta_1^2 - 6}{10\beta_2 - 18 - 12\beta_1^2}. \tag{11}$$

Proof.

Since the mixture is symmetric $(\beta_1 = 0)$, it can be approached by a Pearson type VII distribution if $\beta_2 > 3$, that is

$$\frac{3\sum_{j=1}^{N} w_j \sigma_j^4}{\left(\sum_{j=1}^{N} w_j \sigma_j^2\right)^2} > 3 \Longleftrightarrow \sum_{j=1}^{N} w_j \sigma_j^4 > \left(\sum_{j=1}^{N} w_j \sigma_j^2\right)^2.$$

Using the Cauchy-Schwarz inequality,

$$\left(\sum_{j=1}^{N} x_j^2\right)\left(\sum_{j=1}^{N} y_j^2\right) \geq \left(\sum_{j=1}^{N} x_j y_j\right)^2,$$

with $x_j = \sqrt{w_j}\sigma_j^2$ and $y_j = \sqrt{w_j}$, we get

$$\begin{aligned}\left(\sum_{j=1}^{N} w_j\sigma_j^4\right)\left(\sum_{j=1}^{N} w_j\right) &\geq \left(\sum_{j=1}^{N} w_j\sigma_j^2\right)^2 \Longleftrightarrow \\ &\Longleftrightarrow \sum_{j=1}^{N} w_j\sigma_j^4 \geq \left(\sum_{j=1}^{N} w_j\sigma_j^2\right)^2.\end{aligned}$$

This establishes that the mixture can be approximated by a Pearson type VII distribution. Subtracting μ, the approximating density will be

$$\begin{aligned}f_{X-\mu}(x) &= K_1\left(b_0 + b_2x^2\right)^{-\frac{1}{2b_2}} = K_2\left(1 + \frac{b_2}{b_0}x^2\right)^{-\frac{1}{2b_2}} \underset{b_2=\frac{1}{\nu+1}}{=} \\ &= K_2\left(1 + \frac{x^2}{b_0(\nu+1)}\right)^{-\frac{\nu+1}{2}},\end{aligned}$$

where K_1, K_2 are constants such that $\int\limits_{-\infty}^{+\infty} f_{X-\mu}(x)\,dx = 1$. Considering

$$Y = \sqrt{\frac{\nu}{b_0(\nu+1)}}(X-\mu),$$

then the density of Y is

$$f_Y(y) = K_3\left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

t-Student density function for

$$K_3 = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\upsilon\pi}\,\Gamma\left(\frac{\nu}{2}\right)},$$

leading to the conclusion that

$$\sqrt{\frac{1-b_2}{b_0}}(X-\mu) \overset{\circ}{\sim} t_{\left(\frac{1-b_2}{b_2}\right)}.$$

□

4 Testing $\mu_1 = \mu_2 = ... = \mu_N = \mu$

Theorem 1 allows us to test the equality of mean values

$$H_0 : \mu_1 = \mu_2 = ... = \mu_N, \tag{12}$$

because under H_0

$$\sqrt{\frac{1-b_2}{b_0}} (X - \mu) \overset{\circ}{\sim} t_{\left(\frac{1-b_2}{b_2}\right)}.$$

For $N = 2$, gaussian mixture symmetry implies (Everitt, Hand, 1981) $\mu_1 = \mu_2$ or $w = 0.5$ and $\sigma_1^2 = \sigma_2^2$. For the second condition, the fourth cumulant on equation (3) becomes

$$\kappa_4 = -0.125 (\mu_1 - \mu_2)^4,$$

leading to $\beta_2 < 3$, and so t-Student approach is only valide when $\mu_1 = \mu_2$.

Unfortunately, for $N \geq 3$ it is possible (although very unlikely) that $\beta_1 = 0$ and $\beta_2 > 3$ without $\mu_1 = \mu_2 = ... = \mu_N$. So, not rejecting H_0 may not mean $\mu_1 = \mu_2 = ... = \mu_N$, even theoretically.

On the other hand, rejecting H_0 implies that at least one of the mean values is different from the others.

As usual, unknown parameters (in this case b_0, b_2 and μ) need to be estimated. As $\beta_1 = 0$, estimation can be done by the moments method.

5 Simulation Results

To evaluate the quality of theorem 1 approach, a small simulation work was carried out considering Gaussian mixtures with two, three and four components. The hypothesis of equal mean values was tested using Kolmogorov-Smirnov (K-S) statistics, at the 5% significance level. For each parameter vector were simulated 1000 samples with 1000 observations each. $P\,(Rej.H_0)$ represents the ratio between the number of runs where mean equality was rejected over total number of runs. The results are presented in the following tables, together with the representation of some of the theoretical densities. Note that the approximation seems to work pretty well, under the null hypothesis, and it is very sensitive to skewness and multimodality, contributing to a high test power in this situations. For three and four subpopulations only unimodal mixtures were considered, because the results for two subpopulations clearly shown that multimodality always leads to H_0 rejection.

Table 1: Some Gaussian mixtures for two subpopulations

	$P\left(Rej.H_0\right)$		$P\left(Rej.H_0\right)$
(1) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.1;0;0;0.2;1)$	0.001	**(16)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.1;0;0;5;1)$	0
(2) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.1;1;0;0.2;1)$	0.840	**(17)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.1;1;0;5;1)$	0.059
(3) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.1;2;0;0.2;1)$	0.998	**(18)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.1;2;0;5;1)$	0.754
(4) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.3;0;0;0.2;1)$	0.002	**(19)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.3;0;0;5;1)$	0.011
(5) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.3;1;0;0.2;1)$	0.999	**(20)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.3;1;0;5;1)$	0.905
(6) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.3;2;0;0.2;1)$	1	**(21)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.3;2;0;5;1)$	1
(7) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.5;0;0;0.2;1)$	0.020	**(22)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.5;0;0;5;1)$	0.028
(8) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.5;1;0;0.2;1)$	1	**(23)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.5;1;0;5;1)$	0.958
(9) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.5;2;0;0.2;1)$	1	**(24)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.5;2;0;5;1)$	1
(10) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.7;0;0;0.2;1)$	0.003	**(25)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.7;0;0;5;1)$	0.005
(11) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.7;1;0;0.2;1)$	1	**(26)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.7;1;0;5;1)$	0.572
(12) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.7;2;0;0.2;1)$	1	**(27)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.7;2;0;5;1)$	0.994
(13) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.9;0;0;0.2;1)$	0	**(28)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.9;0;0;5;1)$	0.108
(14) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.9;1;0;0.2;1)$	0.914	**(29)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.9;1;0;5;1)$	0.259
(15) $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.9;2;0;0.2;1)$	1	**(30)** $\left(w,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=(0.9;2;0;5;1)$	0.744

Figure 1: Theoretical densities for some mixtures from table 1

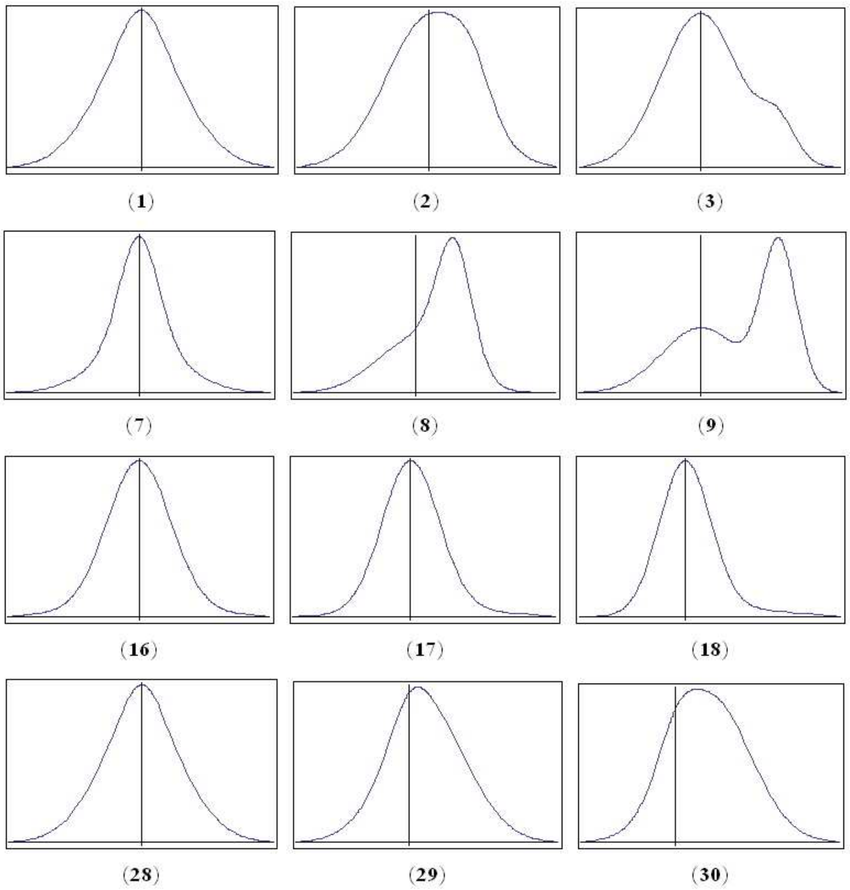
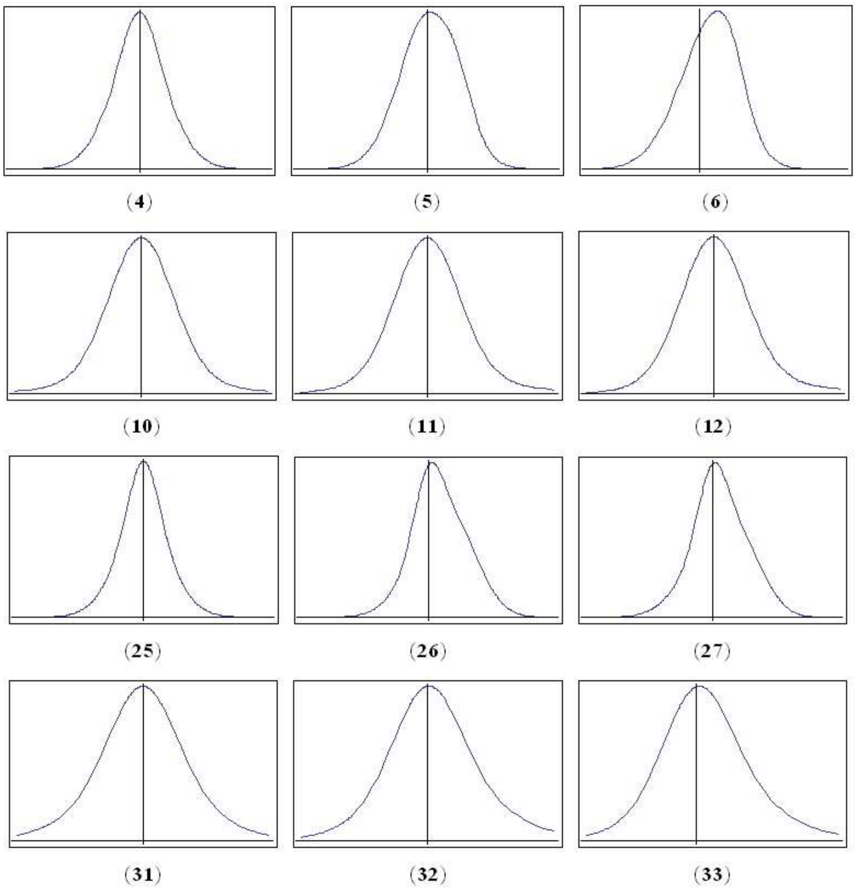


Table 2: Some Gaussian mixtures for three and four subpopulations

	β_1	β_2	$P\left(Rej.H_0\right)$
(1) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.1; 0; 0; 0; 0.2; 0.6; 1)$	0	3.254	0.064
(2) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.1; 1; 0; 0; 0.2; 0.6; 1)$	-0.138	2.906	0.764
(3) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.1; 1; 0; 1; 0.2; 0.6; 1)$	0.013	3.110	0.266
(4) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.4; 0; 0; 0; 0.2; 0.6; 1)$	0	3.366	0.026
(5) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.4; 1; 0; 0; 0.2; 0.6; 1)$	-0.122	2.978	0.602
(6) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.4; 1; 0.5; 0; 0.2; 0.6; 1)$	-0.295	3.117	0.371
(7) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.3; 0.4; 0; 0; 0; 0.2; 0.6; 1)$	0	3.800	0.005
(8) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.3; 0.4; 0.5; 0.5; 0; 0.2; 0.6; 1)$	-0.361	3.838	0.270
(9) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.3; 0.4; 0; 1; 0; 0.2; 0.6; 1)$	0.062	3.258	0.188
(10) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.1; 0; 0; 0; 6; 3; 1)$	0	5.501	0.001
(11) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.1; 1; 0; 0; 6; 3; 1)$	0.568	5.914	0.071
(12) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.1; 1.5; 0; 0; 6; 3; 1)$	0.830	6.338	0.353
(13) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.4; 0; 0; 0; 6; 3; 1)$	0	4.367	0.000
(14) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.4; 1; 0; 0; 6; 3; 1)$	0.320	4.584	0.028
(15) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.1; 0.4; 2; 0; 0; 6; 3; 1)$	0.644	5.119	0.408
(16) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.3; 0.4; 0; 0; 0; 6; 3; 1)$	0	4.050	0.000
(17) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.3; 0.4; 1; 0; 0; 6; 3; 1)$	0.382	4.081	0.313
(18) $\left(w_1, w_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2\right) = (0.3; 0.4; 2; 0; 0; 6; 3; 1)$	0.657	4.070	0.993
(19) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.1; 0.1; 0; 0; 0; 0; 0.2; 0.5; 0.8; 1)$	0	3.248	0.051
(20) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.1; 0.1; 0.5; 0; 0; 0; 0.2; 0.5; 0.8; 1)$	-0.108	3.179	0.183
(21) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.1; 0.1; 1; 0; 0; 0; 0.2; 0.5; 0.8; 1)$	-0.134	2.933	0.682
(22) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.2; 0.3; 0; 0; 0; 0; 0.2; 0.5; 0.8; 1)$	0	3.241	0.029
(23) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.2; 0.3; 0; 1; 0; 0; 0.2; 0.5; 0.8; 1)$	-0.068	3.028	0.469
(24) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.2; 0.3; 0.5; 1; 0; 0; 0.2; 0.5; 0.8; 1)$	-0.186	3.043	0.443
(25) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (1/4; 1/4; 1/4; 0; 0; 0; 0; 0.2; 0.5; 0.8; 1)$	0	3.418	0.000
(26) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (1/4; 1/4; 1/4; 0; 1; 0; 0.5; 0.2; 0.5; 0.8; 1)$	0.000	3.240	0.201
(27) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (1/4; 1/4; 1/4; 0; 1; 0; 0; 0.2; 0.5; 0.8; 1)$	0.115	3.186	0.105
(28) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.1; 0.1; 0; 0; 0; 0; 6; 4; 2; 1)$	0	5.038	0.001
(29) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.1; 0.1; 1; 0; 0; 0; 6; 4; 2; 1)$	0.464	5.545	0.065
(30) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.1; 0.1; 1; 0.6; 0.3; 0; 6; 4; 2; 1)$	0.587	5.421	0.306
(31) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.2; 0.3; 0; 0; 0; 0; 6; 4; 2; 1)$	0	4.209	0.000
(32) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.2; 0.3; 0; 1; 0; 0; 6; 4; 2; 1)$	0.258	4.385	0.085
(33) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.1; 0.2; 0.3; 1; 0.6; 0.3; 0; 6; 4; 2; 1)$	0.397	4.402	0.185
(34) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.25; 0.25; 0.25; 0; 0; 0; 0; 6; 4; 2; 1)$	0	4.766	0.002
(35) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.25; 0.25; 0.25; 0; 1; 0; 0; 6; 4; 2; 1)$	0.103	3.982	0.029
(36) $\left(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right) = (0.25; 0.25; 0.25; 0; 2; 0; 0; 6; 4; 2; 1)$	0.234	3.809	0.485

Figure 2: Theoretical densities for some mixtures from table 2



References

[1] Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B 39*, 1–37.

[2] Everitt, B. S., Hand D. J. (1981). *Finite Mixture Distributions.* London: Chapman & Hall.

[3] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.

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

[5] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A 185*, 71-110.

[6] Teicher, H. (1961). Identifiability of mixtures. *The Annals of Mathematical Statistics 32*, 244–248.

[7] Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics 34*, 1265–1269.

[8] Verbeek, V., Vlassis, N., Krose, B. (2003). Efficient Greedy Learning of Gaussian Mixture Models. *Neural Computation, 15, 2,* 469-485.