Gaussian Location Mixtures

Miguel Martins Felgueiras

CEAUL e ESTG do Instituto Politécnico de Leiria

Abstract

In this paper we are interested in gaussian mixtures with independent components and common variance. A mixture decomposition is presented and useful approximations to F_X are identified. For an unimodal mixture with two components, a variance equality test is developed.

keywords: gaussian mixtures, variance equality, mixture decomposition, Pearson system.

AMS: 60E05

1 Introduction

Mixtures are very effective in modeling real data since they can accommodate multimodality and many densities shapes. Due to the importance of gaussian distribution, finite gaussian mixtures are the most studied in the mixtures context and relevant applications are found in several knowledge fields (Everitt, Hand, 1981, Frühwirth-Schnatter, 2006).

With the increase of computacional power throughout the last decades, mixtures study gained relevance. Nowadays, demanding algorithms like *expectation-maximization algorithm* (EM) (Hasselblad, 1966, Dempster *et al*, 1977) are already implemented in some software.

However, good estimates are not always obtained, even when the initial estimates are equal to the true parameters value. Mixtures tend to have a large number of unknown parameters (typically $3N-1$, where N is the number of subpopulations), and the estimation problem is still a complex problem.

Variance homogeneity is a rather important question in Statistics, usually assumed in ANOVA. In a finite gaussian mixture with subpopulations common variance (location mixture), the number of unknown parameters reduces to $2N$, leading to a more parsimonious solution.

In this paper we study gaussian location mixtures. A mixture decomposition is presented, allowing the mixture to be seen as a sum of independent random variables. For unimodal mixtures, conditions for the mixture approximation to a beta distribution are presented, leading to the decrease of the number of unknown parameters. Besides, we get beta distribution characterizations. Finally, a variance equality test is presented for an unimodal mixture with two subpopulations.

2 Definition, Moments and Cumulants for Gaussian Location Mixtures

A random variable X is a convex gaussian location mixture when

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

Its characteristic function,

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

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

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

and therefore

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

3 The Location Mixture as an Independent Variable Sum

The mixture characteristic function, $\varphi_X(t)$, shows that X can be obtained as the addition of two independent random variable, a zero mean gaussian (white noise) plus a discrete variable taking values on subpopulations means.

Theorem 1.

Let X be a gaussian location mixture with $\mu_1 < \mu_2 < ... < \mu_N$. Then

$$X \stackrel{d}{=} W + Y \tag{3}$$

where

$$W \sim N(0, \sigma)$$

and

$$Y = \begin{cases} \mu_j \\ w_j \end{cases}, \quad j = 1, ..., N$$

are independent random variables.

Proof.

$W+Y$ characteristic function, with W and Y independent and defined above will be

$$\begin{aligned} \varphi_{W+Y}(t) &= \varphi_W(t)\,\varphi_Y(t) = \exp\left(-\frac{t^2\sigma^2}{2}\right)\left[\sum_{i=1}^{N} w_j \exp(it\mu_j)\right] = \\ &= \sum_{i=1}^{N} w_j \exp\left\{it\mu_j - \frac{t^2\sigma^2}{2}\right\} = \varphi_X(t), \end{aligned}$$

and by characteristic function unicity

$$X \stackrel{d}{=} W + Y,$$

as stated. □

The above sum appears in many known applications. Works under amphibian nervous system conducted by Shapovalov, Shiriaev (1980), Grantyn *et al* (1984) concluded that the junction between primary afferent fibre and motoneurone provides joint electrical and chemical transmission. The mixed synapse can be modeled by the addition of white noise to a binomial or Poisson random variable. In image or signal processing, convolutions between

Poisson probability function and zero mean gaussian density are also used. For example, Murtagh *et al* (1995) refers that astronomical images have additive uncorrelated noise. Poisson noise, due to photon arrival events, and gaussian white noise, due to commonly used digitalized photographic plates.

Mixture cumulants can now be decomposed into a sum. Using cumulants properties,

$$\begin{aligned}\kappa_{X,1} &= \kappa_{Y,1}\\ \kappa_{X,2} &= \kappa_{Y,2}+\sigma^2\\ \kappa_{X,3} &= \kappa_{Y,3}\\ \kappa_{X,4} &= \kappa_{Y,4}\\ &\cdots\end{aligned} \tag{4}$$

The equality $\kappa_{X,2} = \kappa_{Y,2} + \sigma^2$ is important, because the knowledge of $V(Y)$ or σ^2 is enough to estimate all the other parameters by the moments method, and even to fit (in some situations) an appropriate distribution to Y.

When it is not possible to find out $V(Y)$ or σ^2, we can still get three equations using the sample characteristics,

$$\begin{aligned}\mu'_{Y,1} &= \mu'_{X,1}\\ \mu_{Y,3} &= \mu_{X,3}\\ \kappa_{Y,4} &= \left(\frac{\mu_{X,3}}{\beta_{X,1}}\right)^{\frac{4}{3}}(\beta_{X,2}-3),\end{aligned} \tag{5}$$

and use them to estimate up to three unknown parameters. This may be sufficient to fit an appropriate Y, namely from the Katz family.

4 Pearson System Approximation

When the mixture have an unimodal density function, it can be approximated by a density from the Pearson system. The goodness of the approximation depends heavily on the values of the skewness β_1 and kurtosis β_2.

For a gaussian location mixture, Pearson type I (four parameters beta) approach arises in a broad range of situations, as the following theorem states.

Theorem 2.
Let $X \stackrel{d}{=} W + Y$ be an unimodal gaussian location mixture according with theorem 1. If

$$V(Y)\sqrt{\frac{1.5\beta_1^2(Y) - \beta_2(Y) + 3}{3}} - V(Y) < \sigma^2 < \frac{1.5\beta_1^2(Y)\,V(Y)}{\beta_2(Y) - 3} - V(Y), \tag{6}$$

the mixture can be approximated by a beta distribution.

Proof.
The mixture can be approximated by a beta distribution if

$$\begin{aligned} 1.5\beta_1^2 &< \beta_2 < 1.5\beta_1^2 + 3 \Longleftrightarrow \\ &\Longleftrightarrow \frac{1.5\kappa_3^2(Y)}{(V(Y) + \sigma^2)^3} < \frac{\kappa_4(Y)}{(V(Y) + \sigma^2)^2} + 3 < \frac{1.5\kappa_3^2(Y)}{(V(Y) + \sigma^2)^3} + 3. \end{aligned}$$

Solving the second inequality,

$$\begin{aligned}
\frac{\kappa_4(Y)}{(V(Y)+\sigma^2)^2}+3 &< \frac{1.5\kappa_3^2(Y)}{(V(Y)+\sigma^2)^3}+3 \Longleftrightarrow \\
&\Longleftrightarrow \frac{\kappa_4(Y)}{(V(Y)+\sigma^2)^2} < \frac{1.5\kappa_3^2(Y)}{(V(Y)+\sigma^2)^3} \Longleftrightarrow \\
&\Longleftrightarrow \left(V(Y)+\sigma^2\right)\kappa_4(Y) < 1.5\kappa_3^2(Y) \Longleftrightarrow \\
&\Longleftrightarrow \sigma^2 < \frac{1.5\kappa_3^2(Y)}{\kappa_4(Y)} - V(Y) \Longleftrightarrow \\
&\Longleftrightarrow \sigma^2 < \frac{\frac{1.5\kappa_3^2(Y)V(Y)}{V^3(Y)}}{\frac{\kappa_4(Y)}{V^2(Y)}} - V(Y) \Longleftrightarrow \\
&\Longleftrightarrow \sigma^2 < \frac{1.5\beta_1^2(Y)V(Y)}{\beta_2(Y)-3} - V(Y).
\end{aligned}$$

For the first inequality,

$$\begin{aligned}
\frac{1.5\kappa_3^2(Y)}{(V(Y)+\sigma^2)^3} &< \frac{\kappa_4(Y)}{(V(Y)+\sigma^2)^2}+3 \Longleftrightarrow \\
&\Longleftrightarrow \frac{1.5\kappa_3^2(Y)}{V(Y)+\sigma^2} < \kappa_4(Y)+3\left(V(Y)+\sigma^2\right)^2
\end{aligned}$$

leading to the sufficient condition

$$\begin{aligned}
\frac{1.5\kappa_3^2(Y)}{V(Y)} &< \kappa_4(Y)+3\left(V(Y)+\sigma^2\right)^2 \Longleftrightarrow \\
&\Longleftrightarrow \frac{1.5\kappa_3^2(Y)-\kappa_4(Y)V(Y)}{3V(Y)} < \left(V(Y)+\sigma^2\right)^2 \Longleftrightarrow \\
&\Longleftrightarrow \sigma^2 > \sqrt{\frac{1.5\kappa_3^2(Y)-\kappa_4(Y)V(Y)}{3V(Y)}} - V(Y) \Longleftrightarrow \\
&\Longleftrightarrow \sigma^2 > \sqrt{\frac{\frac{1.5\kappa_3^2(Y)-\kappa_4(Y)V(Y)}{V^3(Y)}}{\frac{3V(Y)}{V^3(Y)}}} - V(Y) \Longleftrightarrow \\
&\Longleftrightarrow \sigma^2 > \sqrt{\frac{V^2(Y)\left[1.5\beta_1^2(Y)-\beta_2(Y)+3\right]}{3}} - V(Y).
\end{aligned}$$

So, a sufficient condition to approximate the mixture by a beta distribution will be

$$V(Y)\sqrt{\frac{1.5\beta_1^2(Y)-\beta_2(Y)+3}{3}} - V(Y) < \sigma^2 < \frac{1.5\beta_1^2(Y)V(Y)}{\beta_2(Y)-3} - V(Y).$$

□

For instance, if $Y \sim Bi(n,p)$ then the approach is valide when

$$p \in [0; 0.21] \cup [0.79; 1] \wedge n > \frac{1}{6p(1-p)} \wedge \sigma^2 < \frac{np(1-p)}{2(1-6p+6p^2)}. \qquad (7)$$

For $Y \sim P(\lambda)$, theorem 2 leads to the sufficient condition

$$\lambda > \frac{2}{27} \wedge \sqrt{\frac{\lambda}{6}} - \lambda < \sigma^2 < \frac{\lambda}{2},$$

which contains

$$\lambda > \frac{1}{6} \wedge \sigma^2 < \frac{\lambda}{2}. \qquad (8)$$

Other examples are possible for known Y distributions. An important question here is how to guarantee unimodalitie. Unfortunately, useful sufficient conditions are not yet available.

5 Two Populations with equal Variance

When only two subpopulations are considered, the problem is much more easy to deal with, mainly because we now have a maximum of four unknown parameters $(w, \mu_1, \mu_2, \sigma)$.

Mixture cumulants expressions are

$$\begin{aligned}
\kappa_1 &= w\mu_1 + (1-w)\mu_2 \\
\kappa_2 &= \sigma^2 + (1-w)w(\mu_1-\mu_2)^2 \\
\kappa_3 &= (1-w)w(2w-1)(\mu_1-\mu_2)^3 \\
\kappa_4 &= (1-w)w(1-6(1-w)w)(\mu_1-\mu_2)^4,
\end{aligned}$$

and therefore

$$\beta_1^2 = \frac{(\kappa_3)^2}{(\kappa_2)^3} = \frac{\left[(1-w)w(2w-1)(\mu_1-\mu_2)^3\right]^2}{\left[\sigma^2+(1-w)w(\mu_1-\mu_2)^2\right]^3}$$

and

$$\beta_2 = \frac{\kappa_4}{(\kappa_2)^2} + 3 = \frac{(1-w)\,w\,(1-6\,(1-w)\,w)\,(\mu_1-\mu_2)^4}{\left[\sigma^2 + (1-w)\,w\,(\mu_1-\mu_2)^2\right]^2} + 3.$$

For a broad w interval, the mixture can be approximated by a beta distribution.

Theorem 3.
Let X be an unimodal gaussian location mixture with two components. If

$$w \in \left[\frac{1}{2} \pm \frac{\sqrt{3}}{6}\right],$$

then the mixture can be approximated by a beta distribution.

Proof.
The beta distribution approximation is possible when

$$1.5\beta_1^2 < \beta_2 < 3 + 1.5\beta_1^2.$$

For the second inequality, $\beta_2 < 3 + 1.5\beta_1^2$,

$$\frac{(1-w)\,w\,(1-6\,(1-w)\,w)\,(\mu_1-\mu_2)^4}{\left[\sigma^2 + (1-w)\,w\,(\mu_1-\mu_2)^2\right]^2} < \frac{1.5\left[(1-w)\,w\,(2w-1)\,(\mu_1-\mu_2)^3\right]^2}{\left[\sigma^2 + (1-w)\,w\,(\mu_1-\mu_2)^2\right]^3} \iff$$

$$\iff \frac{1-6\,(1-w)\,w}{(1-w)\,w\,(2w-1)^2} < \frac{1.5\,(\mu_1-\mu_2)^2}{\sigma^2 + (1-w)\,w\,(\mu_1-\mu_2)^2},$$

whose exact solution is

$$0.5 - 0.5\sqrt{\frac{4\sigma^2 + (\mu_1-\mu_2)^2}{12\sigma^2 + (\mu_1-\mu_2)^2}} \le w \le 0.5 + 0.5\sqrt{\frac{4\sigma^2 + (\mu_1-\mu_2)^2}{12\sigma^2 + (\mu_1-\mu_2)^2}},$$

which contains the solution

$$w \in \left[\frac{1}{2} \pm \frac{\sqrt{3}}{6}\right] \approx [0.2113; 0.7887]\,.$$

For the first inequality,

$$\frac{(1-w)\,w\,(1-6\,(1-w)\,w)\,(\mu_1-\mu_2)^4}{\left[\sigma^2+(1-w)\,w\,(\mu_1-\mu_2)^2\right]^2}+3>\frac{1.5\left[(1-w)\,w\,(2w-1)\,(\mu_1-\mu_2)^3\right]^2}{\left[\sigma^2+(1-w)\,w\,(\mu_1-\mu_2)^2\right]^3}\Longleftrightarrow$$

$$\Longleftrightarrow 3\sigma^4+(1-w)\,w[6\sigma^2+(1-3\,(1-w)\,w)\,(\mu_1-\mu_2)^2]\,(\mu_1-\mu_2)^2>$$

$$>\frac{1.5\left[(1-w)\,w\,(2w-1)\,(\mu_1-\mu_2)^3\right]^2}{\sigma^2+(1-w)\,w\,(\mu_1-\mu_2)^2},$$

a sufficient condition will be

$$(1-w)\,w[(1-3\,(1-w)\,w)\,(\mu_1-\mu_2)^2]\,(\mu_1-\mu_2)^2>$$

$$>1.5\,(1-w)\,w\,(2w-1)^2\,(\mu_1-\mu_2)^4\Longleftrightarrow$$

$$\Longleftrightarrow 1-3\,(1-w)\,w>1.5\,(2w-1)^2\Longleftrightarrow -3w^2+3w-0.5>0,$$

implying that

$$w\in\left[\frac{1}{2}\pm\frac{\sqrt{3}}{6}\right]\approx[0.2113;0.7887]\,.$$

Both inequalities lead to the same solution, so on the above interval an unimodal mixture can be approximated to a beta distribution. □

If $w\notin\left[\frac{1}{2}\pm\frac{\sqrt{3}}{6}\right]$, the mixture can still be approximated to a beta distribution, depending on μ_1, μ_2 and σ.

Note that the usefulness of this approximation is somehow arguable, since the model have the same number of unknown parameters; the main interest of the previous theorem is to provide a variance equality test.

6 Testing $\sigma_1=\sigma_2=\sigma$

If the mixture is approximated to a beta distribution, as previously stated, then

$$X\overset{\circ}{\sim}beta(a,b,p,q).$$

Under this assumptions,

$$Y = \frac{X-a}{b-a} \overset{\circ}{\sim} beta(p,q)$$

and

$$Z = \frac{Y}{1-Y} \overset{\circ}{\sim} betaprime(p,q),$$

leading to

$$W = \frac{q}{p} Z \overset{\circ}{\sim} F(2p, 2q).$$

Finally, note that

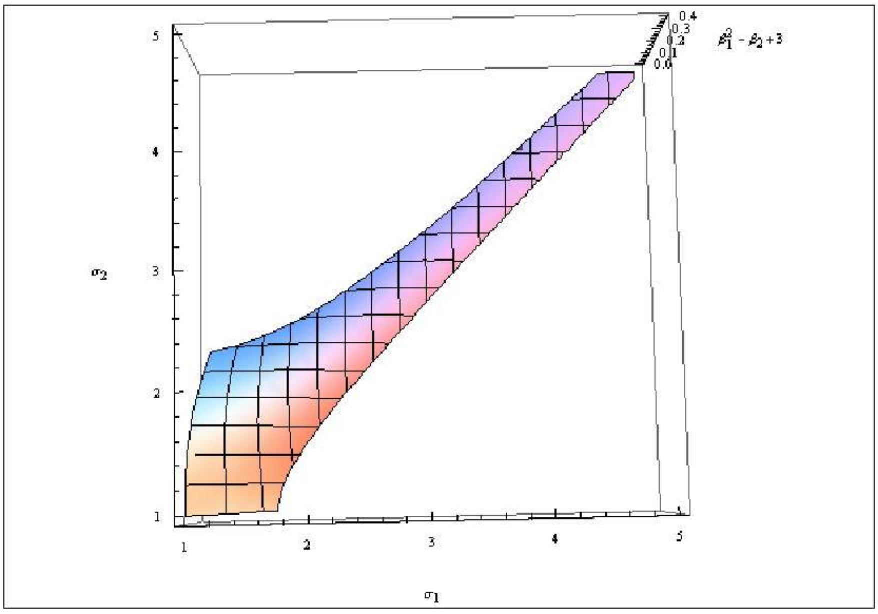
$$W = \frac{q}{p} Z = \frac{q}{p} \frac{\frac{X-a}{b-a}}{1-\frac{X-a}{b-a}} = \frac{q}{p} \frac{X-a}{b-X},$$

so

$$\frac{q}{p} \frac{X-a}{b-X} \overset{\circ}{\sim} F(2p, 2q). \tag{9}$$

Unfortunately, there are some situations where the variances are different but the approximation still holds, even theoretically (see figure 1).

Figure 1: region where $1.5\beta_1^2 < \beta_2 < 3 + 1.5\beta_1^2$ for $(w, \mu_1, \mu_2, \sigma_1, \sigma_2) = (0.35, 0, 2, \sigma_1, \sigma_2)$



In conclusion, rejecting H_0 implies $\sigma_1 \neq \sigma_2$, but if H_0 is not rejected then variances may or may not be equal; even so, σ_1 and σ_2 are at least close.

7 Simulation Results

To evaluate the quality of theorem 3, a small simulation work was carried out. The hypothesis of equal variances was tested using Kolmogorov-Smirnov (K-S) test at the 5% significance level. For each parameter vector 1000 samples with 1000 observations each have been simulated. $P\left(Rej.H_0\right)$ represents the ratio between the number of runs where variance equality was rejected over total number of runs. The results are presented in the following table.

Note that the approximation seems to work pretty well, under the null hypothesis, but the test power should be higher.

Table 1: testing $\sigma_1 = \sigma_2$ on two subpopulations gaussian mixtures

	$P\left(Rej.H_0\right)$		$P\left(Rej.H_0\right)$
(1) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 0.5; 0.2; 0.2)$	0.006	**(19)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 2; 4; 4)$	0.005
(2) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 0.5; 0.2; 0.8)$	0.160	**(20)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 2; 1.4; 3)$	0.022
(3) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 0.5; 0.2; 1.2)$	0.436	**(21)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 2; 0.4; 2)$	0.300
(4) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 0.5; 0.2; 0.2)$	0.003	**(22)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 2; 4; 4)$	0.001
(5) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 0.5; 0.2; 0.8)$	0.700	**(23)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 2; 4; 3)$	0.015
(6) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 0.5; 0.2; 1.2)$	0.977	**(24)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 2; 4; 2)$	0.307
(7) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 0.5; 0.2; 0.2)$	0.000	**(25)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 2; 4; 4)$	0.000
(8) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 0.5; 0.2; 0.5)$	0.359	**(26)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 2; 4; 3)$	0.005
(9) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 0.5; 0.2; 0.8)$	0.966	**(27)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 2; 4; 2)$	0.124
(10) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 1; 0.5; 0.5)$	0.019	**(28)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 3; 5; 5)$	0.008
(11) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 1; 0.5; 1.2)$	0.014	**(29)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 3; 5; 3.5)$	0.152
(12) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 1; 0.5; 2.0)$	0.122	**(30)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.2; 0; 3; 5; 2.5)$	0.759
(13) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 1; 0.5; 0.5)$	0.004	**(31)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 3; 5; 5)$	0.001
(14) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 1; 0.5; 1.2)$	0.090	**(32)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 3; 5; 3.5)$	0.089
(15) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 1; 0.5; 2)$	0.681	**(33)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.35; 0; 3; 5; 2.5)$	0.591
(16) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 1; 0.5; 0.5)$	0.000	**(34)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 3; 5; 5)$	0.000
(17) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 1; 0.5; 1.2)$	0.330	**(35)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 3; 5; 3.5)$	0.005
(18) $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 1; 0.5; 2)$	0.972	**(36)** $\left(w, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = (0.5; 0; 3; 5; 2.5)$	0.139

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] Grantyn, R., Shapovalov, A., Shiriaev, B. (1984) Relation between structural and release parameters at frog sensory-motor synapse.. *The Journal of Physiology 349,* 459-474.

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

[6] Murtagh, F., Starck, J., Bijaoui, A. (1995). Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics, Supplement Series, 112,* 179-189.

[7] Shapovalov, A., Shiriaev, B. (1980) Dual mode of junctional transmission at synapses between single primary afferent fibres and motoneurones in the amphibian. *The Journal of Physiology 306,* 1-15.

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