Modelling of Magnetic Resonance Spectra Using Mixtures for Binned and Truncated Data

Juan M. Garcia-Gomez[1], Montserrat Robles[1], Sabine Van Huffel[2], and Alfons Juan-Císcar[3]

[1] BET-IBM, Polytechnical University of Valencia
[2] Katholieke Universiteit Leuven, Dept. of Electrical Engineering, ESAT-SCD(SISTA)
[3] DSIC, Polytechnical University of Valencia

Abstract. Magnetic Resonance Spectroscopy (MRS) provides the biochemical composition of a tissue under study. This information is useful for the in-vivo diagnosis of brain tumours. Prior knowledge of the relative position of the organic compound contributions in the MRS suggests the development of a probabilistic mixture model and its EM-based Maximum Likelihood Estimation for binned and truncated data. Experiments for characterizing and classifying Short Time Echo (STE) spectra from brain tumours are reported.

Keywords: Expectation-Maximization, Binned data, ^{1}H magnetic resonance spectroscopy (MRS), automatic classification, brain tumour.

1 Introduction

Magnetic Resonance Spectroscopy (MRS) exploits the magnetic properties of ^{1}H nuclei to provide information about the concentration of the compounds of materials. This makes MRS useful as non-invasive technique for brain tumour diagnosis. The MRS signals are typically interpreted in the frequency domain by visual or automatic procedures to characterize the contribution of the biological compound in the tissue. The amplitude of a compound is proportional to its concentration. This motivates the fitting of MRS spectra by mixture density models.

MRS spectra are typically analyzed by two different approaches. The first approach estimates the underlying model composed by mixtures of components to quantify the concentration of the metabolites. Frequency-domain [1] or time-domain [2] fitting methods based on signal processing are applied to the signals. The second approach extracts features from the spectra using univariate-, multivariate-statistics or pattern recognition methods [3] based on their usefulness on discrimination or regression.

This work proposes the definition and estimation of a probabilistic model based on binned and truncated data to fit ^{1}H magnetic resonance spectra using prior knowledge about the relative position of the components of the organic compounds observed in the tumoral masses of the brain. The estimated parameters for each

J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, pp. 266–273, 2007.
© Springer-Verlag Berlin Heidelberg 2007

spectrum summarize the information from the biological compounds and they are used as features in classification problems of brain tumour diagnosis.

Mixture modelling has been applied in some applications where data are available only in bins and may not be provided along the whole the range [4,5,6]. In [4], red blood cells were collected as volume distributions from a Coulter counter to study the disease status of animals exposed to *A*naplasma marginale. The problem in MRS is similar to the previous problems in the sense that contributions of a mixture of biological compounds are assumed to be observed as counts of bins in a range of the ppm-axis. We present an adaptation of the EM for fitting MR spectra, qualitative results in the characterization of spectra and quantitative results in the classification of brain tumours by the use of the estimated parameters.

The rest of the paper is organized as follows. In sections 2 and 3, the probabilistic model and its EM-based Maximum Likelihood Estimation are presented. Then, results using MRS spectra of brain tumours are reported in section 4.

2 Probabilistic Model

Let $\mathcal{X}$ be a sample space partitioned into B bins, $\mathcal{X}_1, \ldots, \mathcal{X}_B$, of which only the counts on the first B' bins can be recorded, while the counts on the last $B - B'$ can not. For instance, in the univariate case, the first B' bins may be delimited by $B' + 1$ points, $p_0, p_1, \ldots, p_{B'}$, such that $p_0 < p_1 < \cdots < p_{B'}$ and $X_b = (p_{b-1}, p_b]$, $b = 1, \ldots, B'$. N independent samples (draws) from $\mathcal{X}$ are made, but our measuring instrument reports only the number of samples falling in each of these first, *observable* B' bins, but fails to report similar counts for samples out of them, in the $B - B'$ *truncated* regions (e.g. $(-\infty, p_0]$ and $(p_{B'}, \infty)$).

Let $\boldsymbol{N}' = (N_1, \ldots, N_{B'})$ be the vector of observed counts and let $N' = \sum_{b=1}^{B'} N_b$. Clearly, the probability of $\boldsymbol{N}'$ can be computed by marginalisation of the joint probability of both, observed and truncated counts,

$$p(\boldsymbol{N}') = \sum_{N_{B'+1}, \ldots, N_B} p(\boldsymbol{N}) \tag{1}$$

where $\boldsymbol{N} = (N_1, \ldots, N_{B'}, N_{B'+1}, \ldots, N_B)$ is the complete vector of counts. We do not know the truncated counts, nor even the total number of samples N, but we know that $\boldsymbol{N}$ has a multinomial distribution defined by N samples from B categories,

$$p(\boldsymbol{N}) = \frac{N!}{\prod_{b=1}^{B} N_b!} \prod_{b=1}^{B} p(b)^{N_b} \tag{2}$$

where $p(b)$ is the probability for a sample to fall in bin $\mathcal{X}_b$, $b = 1, \ldots, B$.

We assume that (2) can also be computed by marginalisation of the joint density for counts and (missing) samples,

$$p(\boldsymbol{N}) = \int d\boldsymbol{X}\, p(\boldsymbol{N}, \boldsymbol{X}) \tag{3}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_B)$ is the whole collection of N independent samples, $\boldsymbol{X}_b = (\boldsymbol{x}_{b1}, \ldots, \boldsymbol{x}_{bN_b})$ is the collection of those N_b from bin $\mathcal{X}_b$ ($b = 1, \ldots, B$), and

$$p(\boldsymbol{N}, \boldsymbol{X}) = \frac{N!}{\prod_{b=1}^{B} N_b!} \prod_{b=1}^{B} \prod_{n=1}^{N_b} p(\boldsymbol{x}_{bn}) \tag{4}$$

where $p(\boldsymbol{x}_{bn})$ is the (unknown) probability density for a sample from bin $\mathcal{X}_b$.

At this point, we assume that samples come from a common probability density function, irrespective of their originating bins. This density function is a parametric, C-component mixture,

$$p_{\boldsymbol{\Theta}}(\boldsymbol{x}) = \sum_{c=1}^{C} \pi_c \, p_{\boldsymbol{\Theta}'}(\boldsymbol{x} \mid c) \tag{5}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\pi}, \boldsymbol{\Theta}')$ is the parameter vector of the mixture; $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_C)$ is the vector of mixture coefficients, subject to $\sum_c \pi_c = 1$, and $\boldsymbol{\Theta}'$ includes the parameters required to define each mixture component $p_{\boldsymbol{\Theta}'}(\boldsymbol{x} \mid c)$, $c = 1, \ldots, C$. As usual with finite mixtures, we may think of $\boldsymbol{x}$ as an incomplete component-labelled sample which may be completed by addition of an indicator variable (component label) $\boldsymbol{z} \in \{0, 1\}^C$ with 1 in the position of the indicated component and zeros elsewhere. Therefore, we can rewrite (5) as

$$p_{\boldsymbol{\Theta}}(\boldsymbol{x}) = \sum_{\boldsymbol{z}} p_{\boldsymbol{\Theta}}(\boldsymbol{x}, \boldsymbol{z}) \tag{6}$$

with

$$p_{\boldsymbol{\Theta}}(\boldsymbol{x}, \boldsymbol{z}) = \prod_{c=1}^{C} (\pi_c \, p_{\boldsymbol{\Theta}'}(\boldsymbol{x} \mid c))^{z_c} \tag{7}$$

By substitution of (7) in (6), (6) in (4) and some straightforward manipulations, we can rewrite (4) as

$$p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}) = \sum_{\boldsymbol{Z}} p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z}) \tag{8}$$

where $\boldsymbol{Z}$ is the collection of component labels for $\boldsymbol{X}$, that is, $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_B)$, with $\boldsymbol{Z}_b = (\boldsymbol{z}_{b1}, \ldots, \boldsymbol{z}_{bN_b})$ and $\boldsymbol{z}_{bn} \in \{0, 1\}^C$ ($b = 1, \ldots, B$; $n = 1, \ldots, N_b$); and

$$p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z}) = \frac{N!}{\prod_{b=1}^{B} N_b!} \prod_{b=1}^{B} \prod_{n=1}^{N_b} \prod_{c=1}^{C} (\pi_c \, p_{\boldsymbol{\Theta}'}(\boldsymbol{x}_{bn} \mid c))^{z_{bnc}} \tag{9}$$

Note that we have added the parameter vector $\boldsymbol{\Theta}$ as a subscript to the joint densities $p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X})$ and $p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z})$ to emphasize their dependence on the parameters governing the hidden mixture (5).

Now, by substitution of (8) in (3), and (3) in (1), we can write our probabilistic model as

$$p_{\boldsymbol{\Theta}}(\boldsymbol{N}') = \sum_{N_{B'+1}, \ldots, N_B} \int d\boldsymbol{X} \sum_{\boldsymbol{Z}} p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z}) \tag{10}$$

Note that $p_{\boldsymbol{\Theta}}(\boldsymbol{N}')$ can be seen as an *incomplete* model which results from marginalisation (many-to-one mapping) of the *complete* model $p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z})$.

Obviously, model (10) still needs adoption of a particular parametric form for the mixture components. Taking into account the specific application considered in this work, we will assume that samples are drawn from a C-component mixture of univariate normal densities, of means known up to a *global shift* μ_0, and independent variances $\sigma_1^2, \ldots, \sigma_C^2$; that is, for all $c = 1, \ldots, C$,

$$p_{\boldsymbol{\Theta}'}(\boldsymbol{x} \mid c) \sim N(\mu_0 + \delta_c, \sigma_c^2) \tag{11}$$

where δ_c is the known displacement from μ_0 of the cth component mean. Thus, the vector of parameters governing the mixture components is $\boldsymbol{\Theta}'=(\mu_0, \sigma_1^2, \ldots, \sigma_C^2)$.

3 EM-Based Maximum Likelihood Estimation

Maximum likelihood estimation of $\boldsymbol{\Theta}$ using the EM algorithm has been previously considered in [4] and [5] for the univariate and multivariate normal cases, respectively. Our case is similar to, but slightly different from the general, parameter-independent univariate case. More precisely, the general univariate model assumes that component means are independent, while in our model all of them are known up to a global shift. This makes our estimation problem simpler, but the EM algorithm is almost identical. In what follows, we briefly review the EM algorithm for the general model and then we provide the necessary modifications for our modelling variation. The reader is referred to [4] for more details.

The log-likelihood function of $\boldsymbol{\Theta}$ w.r.t. a given $\boldsymbol{N}'$ is

$$L(\boldsymbol{\Theta}; \boldsymbol{N}') = \log \sum_{N_{B'+1}, \ldots, N_B} \int d\boldsymbol{X} \sum_{\boldsymbol{Z}} p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z}) \tag{12}$$

which is exactly the logarithm of $p_{\boldsymbol{\Theta}}(\boldsymbol{N}')$ as defined in (10), but interpreted as a function of $\boldsymbol{\Theta}$ only, and assuming that mixture components are univariate normals. The EM algorithm maximises (12) iteratively, through the application of two basic steps in each iteration: the E(xpectation) step and the M(aximisation) step. On the one hand, the E step computes a lower bound of (12) for all $\boldsymbol{\Theta}$; the so-called Q function,

$$Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(k)}) = E[\log p_{\boldsymbol{\Theta}}(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{Z}) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}] \tag{13}$$

that is, the expectation of the logarithm of the complete model, conditional to the incomplete data, $\boldsymbol{N}'$, and a current estimation of the model parameters, $\boldsymbol{\Theta}^{(k)}$. On the other hand, the M step obtains a new estimate for $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}^{(k+1)}$, by maximisation of the Q function,

$$\boldsymbol{\Theta}^{(k+1)} = \arg\max_{\boldsymbol{\Theta}} \; Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(k)}) \qquad \text{s.t.} \sum_c \pi_c = 1 \tag{14}$$

Given an initial value of the parameters, $\boldsymbol{\Theta}^{(0)}$, these two steps are repeated until convergence to a local maximum of the likelihood function.

Ignoring an additive term not involving $\boldsymbol{\Theta}$, the Q function can be written as

$$Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(k)}) = \sum_{c=1}^{C} \sum_{b=1}^{B} N_b^{(k)} \, E_b[z_c^{(k)}(x_b)(\log \pi_c + \log p_{\boldsymbol{\Theta}'}(x_b \mid c)) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}] \quad (15)$$

where $N_b^{(k)}$ is the expected number of samples drawn from bin $\mathcal{X}_b$,

$$N_b^{(k)} = \begin{cases} N_b & \text{if } b \leq B' \\ N' \dfrac{p(b)^{(k)}}{\sum_{b'=1}^{B'} p(b')^{(k)}} & \text{otherwise} \end{cases} \quad (16)$$

with $p(b)^{(k)}$ being the probability for a sample to fall in bin $\mathcal{X}_b$,

$$p(b)^{(k)} = \int_{\mathcal{X}_b} dx \; p_{\boldsymbol{\Theta}^{(k)}}(x) \quad (17)$$

The expectation in (15) is with respect to a sample x_b from bin $\mathcal{X}_b$; i.e., with respect to the truncated density of the bin $\mathcal{X}_b$

$$p^{\text{trunc}}_{\boldsymbol{\Theta}^{(k)}}(x_b) = \frac{p_{\boldsymbol{\Theta}^{(k)}}(x_b)}{p(b)^{(k)}} \quad (18)$$

and involves the posterior probability for x_b to belong to component c of the mixture, given a current parameter estimate $\boldsymbol{\Theta}^{(k)}$,

$$z_c^{(k)}(x_b) = \frac{\pi_c^{(k)} \, p_{\boldsymbol{\Theta}'^{(k)}}(x_b \mid c)}{p_{\boldsymbol{\Theta}^{(k)}}(x_b)} \quad (19)$$

Maximisation of (15), as indicated in (14), leads to the following re-estimates for each component c ($c = 1, \ldots, C$)

$$\pi_c^{(k+1)} = \frac{\sum_{b=1}^{B} N_b^{(k)} \, E_b[z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]}{\sum_{b=1}^{B} N_b^{(k)}} \quad (20)$$

$$\mu_c^{(k+1)} = \frac{\sum_{b=1}^{B} N_b^{(k)} \, E_b[x_b \, z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]}{\sum_{b=1}^{B} N_b^{(k)} \, E[z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]} \quad (21)$$

$$\sigma_c^{2(k+1)} = \frac{\sum_{b=1}^{B} N_b^{(k)} \, E_b[(x_b - \mu_c^{(k+1)})^2 \, z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]}{\sum_{b=1}^{B} N_b^{(k)} \, E[z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]} \quad (22)$$

where, as in (15), all expectations are with respect to the truncated density (18). Their derivations were shown by McLachlan and Jones in [4,7].

Equations (20), (21) and (22) are the basic equations of an EM iteration in the general, parameter-independent univariate case (EMBTD). In our case

(EM4BTDr), with means known up to shift μ_0, the basic equations are (20), (22) and $\mu_c^{(k+1)} = \mu_0^{(k+1)} + \delta_c, c = 1, \ldots, C$, where

$$\mu_0^{(k+1)} = \frac{\sum_{c=1}^{C} \sum_{b=1}^{B} N_b^{(k)} \, E_b[(x_b - \delta_c) \, z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]}{\sum_{c=1}^{C} \sum_{b=1}^{B} N_b^{(k)} \, E_b[z_c^{(k)}(x_b) \mid \boldsymbol{N}', \boldsymbol{\Theta}^{(k)}]}, \tag{23}$$

4 Experimental Results

The mixture models presented in the previous sections were applied to 147 multicenter signals acquired during the Interpret project [8]. The brain tumour diagnosis and quality of signals were validated by two committees of experts [3]. The number of cases per diagnosis are 77 Glioblastoma Multiforme (GM), 50 Meningioma (MM) and 20 Astrocytoma grade II (A2). STE, (TE $\in$ [20,30]ms) Single Voxel (SV) spectra were used in the study [3]. The reference point for each spectra was set and validated visually by a spectroscopist, then little or null shifting is expected in the components.

The a priori information from biochemical knowledge used in the experiments was the chemical shift (in ppm units) of metabolites L2 (0.92), Glx (2.04), L1 (1.25), Glx2 (2.46), LAC (1.31), Glx3 (3.76), ALA (1.48, 1.46), mI (3.26), NAc (2.02, 2.00), mI2 (3.53), Cr (3.03), mI3 (3.61), Cr2 (3.92), Tau (3.25), Cho (3.19), Tau2 (3.42), Gly (3.55), ALA2 (3.78). The initial models for both algorithms was exactly the same, equal prior probability and variance for each component were established.

Figure 1 summarizes the main behaviour of the EMBTD (EM for Binned and Truncated data in general form) and EMBTDr (EM for Binned and Truncated data with means known up to shift μ_0) estimates. In the first example (top-left), the parameters estimated by both EMBTD and EMBTDr are quite similar, hence the spectrum is fitted in a similar way. In the second example (top-right), the related means restriction incorporated in the EMBTDr model keeps the position of the compounds better than the EMBTD model according to the underlying biological mixture. In the third example (bottom), the EMBTD model fits a lipid contribution at 2.75ppm not specified in the prior knowledge, but the meaning of the initial components based on biological knowledge is lost.

Two studies were carried out to characterize the behaviour of the models on average. In the first study, we measure the mean of the differences between the estimated shifting ($\hat{\mu_c}$) of the components with respect to the typical chemical shift (μ_c). None or small shifting is assumed in the spectra of the database. Therefore, the smaller difference, the closer the estimated component is to the organic compound. Table 1 shows the results obtained by EMBTD and EMBTDr on the MRS database. The differences obtained by both models are small, considering that the range is 3.6ppm and the frequency resolution 0.02ppm. However, the difference obtained on average by EMBTD is 3.6 times the difference obtained by EMBTDr. Hence, EMBTDr keeps better the position of the biological compound in the estimated model.

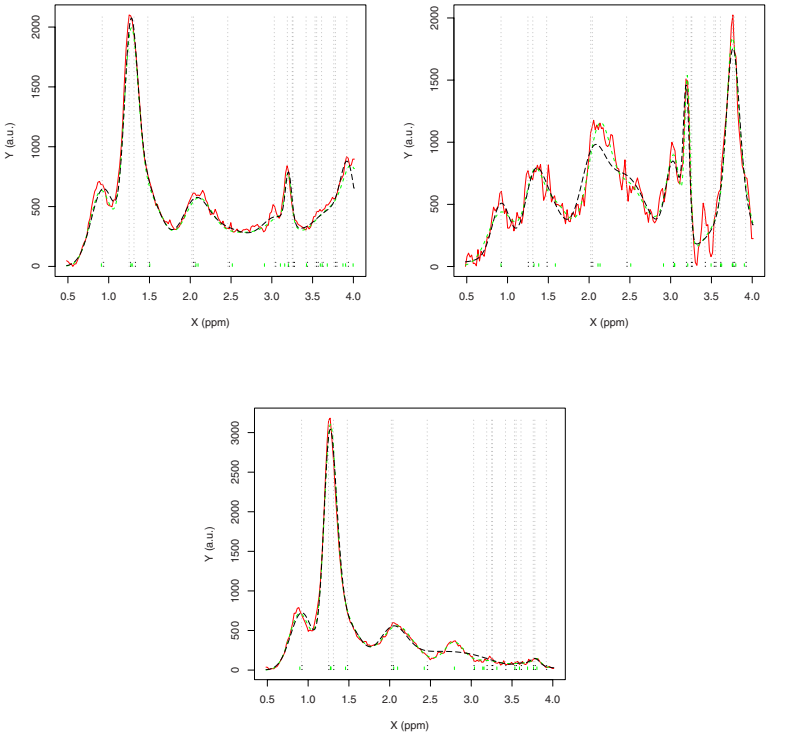


Fig. 1. Three spectra showing the behaviour of the EMBTD and EMBTDr models. Real spectra are drawn in solid lines, EMBTD models in dashed lines and EMBTDr in double-dashed lines. μ_c are marked with vertical lines, $\hat{\mu_c}$ of each model are marked with a small vertical line for EMBTD and two dots for EMBTDr.

Table 1. Mean of the difference between the estimated shifting of the components by the EM algorithms with respect to the typical chemical shift

	EMBTD	EMBTDr
$\mu_c - \hat{\mu_c}$	0.0079	0.0022

Table 2. Results in classification. Cells are composed by ϵ [95%-CI], where ϵ is the error estimation and [95%-CI] the 95% credibility intervals [9].

classes	PCA+LDA	EMBTD+LDA	EMBTDr+LDA
MM GM	17.32 [11.50,24.54]	**10.24 [5.86 16.31]**	14.17 [8.93,20.94]
MM A2	**7.14 [2.84 14.70]**	14.29 [7.59,23.70]	**7.14 [2.84,14.70]**
GM A2	8.25 [4.00,14.81]	9.28 [4.71,16.11]	**5.15 [2.05,10.74]**

In the second study, the ($\hat{\boldsymbol{\pi}}$) parameters estimated by the EMBTD and EMBTDr were the input variables of binary classifiers based on Linear Discriminant Analysis (LDA). The performace of these classifiers where compared with similar LDA classifiers based on PCA (Principal Component Analysis). Table 2 shows the estimation of the error, carried out by Cross Validation with 10 stratified partitions. EMBTDr achieves the best performance when classifing A2 from GM and A2 from MM (equal to PCA). In the discrimination of MM from GM, both EMBTDr and EMBTD estimates are considerable better than the PCA model.

5 Conclusions and Further Work

A probabilistic mixture model for binned and truncated data with univariate mixture densities of means known up to a global shift has been proposed for Magnetic Resonance Spectroscopy data characterization. The model can be efficiently estimated by means of the E(xpectation)-M(aximisation) algorithm. The new version of the algorithm keeps the biological information in the model and fits properly STE MR Spectra. The incorporation of the classifier in a Decision Support System (DSS) could be of interest for clinicians to decide the diagnosis of routine or special patients. In further work, more applications of the proposed mixture model will be considered in MRS analysis.

Acknowledgments. This work was partially funded by the EU projects eTumour (FP6-2002-LH 503094) and HealthAgents (IST-2004-27214). We thank INTERPRET partners, C. Arús (UAB), C. Majós (IDI), A. Moreno (CDP), J. Griffiths (SGUL), A. Heerschap (RU), W. Gajewicz (MUL), and J. Calvar (FLENI), for providing data.

References

1. Mierisova, S., Ala-Korpela, M.: MR spectroscopy quantitation: a review of frequency domain methods. NMR Biomed. 14(4), 247–259 (2001)
2. Vanhamme, L., Sundin, T., Hecke, P.V., Huffel, S.V.: MR spectroscopy quantitation: a review of time-domain methods. NMR Biomed. 14(4), 233–246 (2001)
3. Tate, A., et al.: Development of a DSS for diagnosis and grading of brain tumours using in vivo magnetic resonance SV spectra. NMR Biomed. 19(4), 411–434 (2006)
4. McLachlan, G.J., Jones, P.N.: Fitting mixture models to grouped and truncated data via the EM algorithm. Biometrics 44, 571–578 (1988)
5. Cadez IV, et al.: Maximum likelihood estimation of mixture densities for binned and truncated multivariate data. Mach. Learn. 47(1), 7–34 (2002)
6. Same, A., Ambroise, C., Govaert, G.: A classification EM algorithm for binned data. Computational Statistics and Data Analysis (2005)
7. Jones, P.N., McLachlan, G.J.: Statistical algorithms: Algorithm AS 254. Applied Statistics 39(2), 273–282 (1990)
8. Julia-Sape, M., et al.: A multicentre, web-accessible and quality control-checked database of in vivo MRS of brain tumour patients. MAGMA 19(1), 22–33 (2006)
9. Martin, J.K., Hirschberg, D.S.: Small sample statistics for classification error rates II. Technical Report ICS-TR-96-22 (1996)