Available online at www.elixirpublishers.com (Elixir International Journal)

Applied Mathematics

Elixir Appl. Math.117 (2018) 50471-50480

ISSN: 2229-712X

The EM Algorithm and Some Expressive Properties for Logistic Mixtures

Xiao Liu and Dirk Greindl

High-Tech, Sogeti Deutschland, Capgemini, Olof-Palme-Straße. 14, 81829 München, Germany.

ARTICLE INFO

Article history:

Received: 14 February 2018;

Received in revised form: 02 April 2018;

Accepted: 13 April 2018;

Keywords

The Expectation-Maximization Algorithm,

Logistic Mixtures.

ABSTRACT

The Expectation-Maximization (EM) algorithm is a parameter estimation method which performs to the general data frame of maximum-likelihood estimation, including some applications to data analysis and statistics. There is the work by Dempster, Laird and Rubin (1977) on the key features of defining the EM algorithm. Logistic mixtures, unlike normal mixtures, have not been studied for the EM algorithm. In this paper, we express the theorem about a relationship between the gradient of the log likelihood and the step in parameter space taken by the EM algorithm in their multivariate extension for the mixture logistic distribution (see Malik and Abraham (1973)). The literature on determination of the number of modes in logistic mixture models has focused primarily on univariate mixtures. In fact, there is a simple description of modality when one is mixing two univariate components. In particular, an analogue of the techniques such as the paper written by Robertson and Fryer (1969) al-lows to attain the results in the case of the convex combination of two item response functions (IRFs) for the unidimensional dichotomous 1-parameter to 4-parameter logistic (1PL to 4PL) item response theory (IRT) models.

1. Introduction

The EM algorithm has become a very popular computational method in statistics. Dempster, Laird and Rubin (1977) introduced a very general framework for defining the EM iterative algorithm which find local maxima of the log-likelihood in problems with analysing incomplete data.

Furthermore, Wu (1983) considerably tightened the theory of EM algorithm by giving strict conditions for convergence to a stationary point of the log-likelihood. Compared with the general experience in numerical optimization, the executation of the expectation step (E-step) and maximization step (M-step) is easy for many statistical problems. Solutions of the M-step often exist in closed form. The M-step can be implemented with a standard statistical package in many cases, thereby saving programming time. Another reason for statisticians preferring to EM is that it does not require large storage space.

Xu and Jordan (1996) provided the EM algorithm for computing maximum likelihood estimates of Gaussian Mixtures. For mixtures of logistic distributions such information is lacking, although a special instance of the EM algorithm prevails therein. In this paper, we propose analogs of the multivariate normal mixture results for the multivariate logistic distribution.

2. Logistic Distribution is An Exponential Family

A K-component mixture of D-dimensional logistic distributions can be represented by the probability density function

$$\varnothing((X|\Theta) = \sum_{i=1}^{K} \pi_i \varnothing(X;\mu_i,s_i), \qquad X \in \mathbb{R}^D \tag{1}$$

where $\pi_i$is the mixing proportion of component i, $\pi_i \in [0;\, 1]$, $\sum_{i=1}^{K} \pi_i = 1$ and $\varnothing((X|\Theta)$ is the density of a multivariate logistic distribution with mean μ and scale s. We will sometimes use $\varnothing_i(X)$ as shorthand notation for $\varnothing\,(X;\ \mu_i, s_i)$, and call $\varnothing_i$ the ith component density, where (see, Malik and Abraham, 1973)

$$\phi_i(\mathbf{x}) = \frac{D!}{\prod_{j=1}^{D} s_{ij}} \exp\left[-\sum_{j=1}^{D}(x_j-\mu_{ij})/s_{ij}\right]\left\{1+\sum_{j=1}^{D}\exp\left[-(x_j-\mu_{ij})/s_{ij}\right]\right\}^{-(D+1)}. \tag{2}$$

Assume that K and N are independent, identically distributed samples $\{X(t)\}_1^N$, we obtain the following log-likelihood:

$$l(\Theta) = \log\prod_{t=1}^{N}\phi\left(\mathbf{x}(t)|\Theta\right) = \sum_{t=1}^{N}\log\phi\left(\mathbf{x}(t)|\Theta\right). \tag{3}$$

The logistic distribution can be written as an exponential family after the transform as below.

$$\begin{aligned}\phi_i(\mathbf{x}) &= \frac{D!}{\prod_{j=1}^{D} s_{ij}} \exp\left[-\sum_{j=1}^{D}(x_j-\mu_{ij})/s_{ij}\right]\left\{1+\sum_{j=1}^{D}\exp\left[-(x_j-\mu_{ij})/s_{ij}\right]\right\}^{-(D+1)} \\ &= \frac{\exp\left[-\sum_{j=1}^{D}(x_j-\mu_{ij})/s_{ij}\right]}{\left\{1+\sum_{j=1}^{D}\exp\left[-(x_j-\mu_{ij})/s_{ij}\right]\right\}}\left\{1+\sum_{j=1}^{D}\exp\left[-(x_j-\mu_{ij})/s_{ij}\right]\right\}^{-D}\cdot\frac{D!}{\prod_{j=1}^{D} s_{ij}}\end{aligned}$$

Tele:

E-mail address: xiao.b.liu@sogeti.co

$$= \frac{\exp\left[-\sum_{j=1}^{D}(x_j-\mu_{ij})/s_{ij}\right]}{\left\{1+\sum_{j=1}^{D}\exp\left[-(x_j-\mu_{ij})/s_{ij}\right]\right\}} \cdot \exp\left\{-D\ln\left\{1+\sum_{j=1}^{D}\exp\left[-(x_j-\mu_{ij})/s_{ij}\right]\right\}+\ln\left(\frac{D!}{\prod_{j=1}^{D}s_{ij}}\right)\right\}. \quad (4)$$

So according to Dempster et al.'s paper in 1977, we can get the following interative algorithm.

$$\pi_i^{(k+1)} = \frac{\sum_{t=1}^{N}\alpha_i^{(k)}(t)}{N}$$

$$\boldsymbol{\mu}_i^{(k+1)} = \frac{(D+1)\boldsymbol{\mu}_i^{(k)}\sum_{t=1}^{N}\alpha_i^{(k)}(t)\left[\sum_{j'=1}^{D}E_{ij'}^{(k)}(t)\right]^{-1}}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)}$$

$$\mathbf{s}_i^{(k+1)} = \frac{\mathbf{s}_i^{(k)}}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)}\sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}y_{j'}^{(k)}(t)\left[1+\frac{D+1}{E_{ij'}^{(k)}(t)}\right] \quad (5)$$

Where the coefficients α_i^K are defined as follows:

$$\alpha_i^{(k)}(t) = \frac{\pi_i^{(k)}\phi\left(\mathbf{x}(t);\boldsymbol{\mu}_i^{(k)},\mathbf{s}_i^{(k)}\right)}{\sum_{i'=1}^{K}\pi_{i'}^{(k)}\phi\left(\mathbf{x}(t);\boldsymbol{\mu}_{i'}^{(k)},\mathbf{s}_{i'}^{(k)}\right)},$$

$$E_{ij'}^{(k)}(t) = 1+e^{\frac{x_{j'}(t)-\mu_{ij'}^{(k)}}{s_{ij'}^{(k)}}}+e^{\frac{x_{j'}(t)-\mu_{ij'}^{(k)}}{s_{ij'}^{(k)}}}\sum_{j\neq j'}^{D}e^{-\frac{x_j(t)-\mu_{ij}^{(k)}}{s_{ij}^{(k)}}},$$

$$\mathbf{s}_i^{(k)} = \left[\sum_{j'=1}^{D}s_{ij'}^{(k)-1}\right]^{-1},$$

and

$$y_{j'}^{(k)}(t) = \frac{x_{j'}(t)-\mu_{ij'}^{(k)}}{s_{ij'}^{(k)}}.$$

3. The EM Algorithm for Logistic Mixtures and Theorem of Connection between EM and Gradient Ascent

For the most central part we confine our attention to the following theorem about a relationship between the gradient of the log likelihood and the step in parameter space taken by the EM algorithm.

Theorem 1. At each iteration of the EM algorithm in equation (5), we can get

$$A^{(k+1)}-A^{(k)} = \phi_A^{(k)}\left.\frac{\partial l}{\partial A}\right|_{A=A^{(k)}} \quad (6)$$

$$\boldsymbol{\mu}_i^{(k+1)}-\boldsymbol{\mu}_i^{(k)} = \phi_{\boldsymbol{\mu}_i}^{(k)}\left.\frac{\partial l}{\partial \boldsymbol{\mu}_i}\right|_{\boldsymbol{\mu}_i=\boldsymbol{\mu}_i^{(k)}} \quad (7)$$

$$\mathbf{s}_i^{(k+1)}-\mathbf{s}_i^{(k)} = \phi_{\mathbf{s}_i}^{(k)}\left.\frac{\partial l}{\partial \mathbf{s}_i}\right|_{\mathbf{s}_i=\mathbf{s}_i^{(k)}} \quad (8)$$

Where

$$\phi_A^{(k)} = \frac{1}{N}\left\{\mathrm{diag}[\pi_1^{(k)},\cdots,\pi_K^{(k)}]-A^{(k)}(A^{(k)})^T\right\} \quad (9)$$

$$\phi_{\boldsymbol{\mu}_i}^{(k)} = -\frac{\mathbf{s}_i^{(k)}\boldsymbol{\mu}_i^{(k)}}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)} \quad (10)$$

$$\phi_{\mathbf{s}_i}^{(k)} = \frac{\mathbf{s}_i^{(k)}\mathbf{s}_i^{(k)}}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)} \quad (11)$$

where A denotes the vector of mixing proportions $[\pi_1,\ldots,\pi_K]^T$, i indexes the mixture components (i = 1; ; K), k denotes the iteration number. Moreover, given the constraints $\sum_{i=1}^{K}\pi_i^{(K)}=1$ and $\pi_i^{(K)}\geq 0$, $\emptyset_A^{(K)}$ is a positive definite matrix, and the matrix $\emptyset_{\mu_i}^{(K)}$ is negative definite, and $\emptyset_{s_i}^{(K)}$ matrix with probability one for N sufficiently large.

Proof.

(i) First we consider the EM Algorithm update for the mixing proportions pi. It follows from equations (1) and (3) that

$$\left.\frac{\partial l}{\partial A}\right|_{A=A^{(k)}} = \sum_{t=1}^{N}\frac{\left[\phi(\mathbf{x}(t),\theta_1^{(k)}),\cdots,\phi(\mathbf{x}(t),\theta_K^{(k)})\right]^T}{\sum_{i=1}^{K}\pi_i\phi\left(\mathbf{x}(t),\theta_i^{(k)}\right)}. \quad (12)$$

Multiplying by $\emptyset_A^{(K)}$, we have

$$\phi_A^{(k)} \left.\frac{\partial l}{\partial A}\right|_{A=A^{(k)}} \tag{13}$$

$$= \frac{1}{N}\sum_{t=1}^{N} \frac{\left[\pi_1^{(k)}\phi(\mathbf{x}(t),\theta_1^{(k)}),\cdots,\pi_K^{(k)}\phi(\mathbf{x}(t),\theta_K^{(k)})\right]^T - A^{(k)}\sum_{i=1}^{K}\pi_i\phi\left(\mathbf{x}(t),\theta_i^{(k)}\right)}{\sum_{i=1}^{K}\pi_i\phi\left(\mathbf{x}(t),\theta_i^{(k)}\right)}$$

$$= \frac{1}{N}\sum_{t=1}^{N}\left[\alpha_1^{(k)}(t),\cdots,\alpha_K^{(k)}(t)\right]^T - A^{(k)}.$$

The update formula for A in equation (5) can be rewritten as

$$A^{(k+1)} = A^{(k)} + \frac{1}{N}\sum_{t=1}^{N}\left[\alpha_1^{(k)}(t),\cdots,\alpha_K^{(k)}(t)\right]^T - A^{(k)}.$$

Combining the last two equations establishes the update rule for A (equation (9)). Furthermore, for an arbitrary vector u, we can get

$$Nu^T\phi_A^{(k)}u = u^T \operatorname{diag}\left[\pi_1^{(k)},\cdots,\pi_K^{(k)}\right]u - \left(u^TA^{(k)}\right)^2.$$

By Jensen's inequality we obtain

$$u^T \operatorname{diag}\left[\pi_1^{(k)},\cdots,\pi_K^{(k)}\right]u = \sum_{i=1}^{K}\pi_i^{(k)}u_i^2 \quad > \left(\sum_{i=1}^{K}\pi_i^{(k)}u_i\right)^2 \quad = \left(u^TA^{(k)}\right)^2. \tag{14}$$

Thus, $u^T\emptyset_A^{(K)}u > 0$ and $\emptyset_A^{(K)}$ is positive definite given the constraints $\sum_{i=1}^{K}\pi_i^{(K)} = 1$ and $\pi_i^{(K)} \geq 0$ for $\forall i$.

(ii) Next, we prove the second part of the theorem. It follows from equations (1) and (3) that

$$\left.\frac{\partial l}{\partial \boldsymbol{\mu}_i}\right|_{\boldsymbol{\mu}_i=\boldsymbol{\mu}_i^{(k)}} = \sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}s_{ij'}^{(k)^{-1}}\left(1-\frac{D+1}{E_{ij'}^{(k)}(t)}\right). \tag{15}$$

Premultiplying by $\emptyset_{\mu i}^{(K)}$, considering the convergence of summation in series, In we obtain

$$\phi_{\boldsymbol{\mu}_i}^{(k)}\left.\frac{\partial l}{\partial \boldsymbol{\mu}_i}\right|_{\boldsymbol{\mu}_i=\boldsymbol{\mu}_i^{(k)}}$$

$$= -\frac{\mathbf{s}_i^{(k)}\boldsymbol{\mu}_i^{(k)}}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)}\sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}s_{ij'}^{(k)^{-1}} + \frac{(D+1)\boldsymbol{\mu}_i^{(k)}\sum_{t=1}^{N}\alpha_i^{(k)}(t)}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}E_{ij'}^{(k)}(t)}$$

$$= \boldsymbol{\mu}_i^{(k+1)} - \boldsymbol{\mu}_i^{(k)}. \tag{16}$$

Following from equation (5), we have $\sum_{t=1}^{N}\alpha_i^{(K)}(t) > 0$ moreover, $s_i^{(k)}$ and $\mu_i^{(k)}$ are both positive definite. Thus, according to equation (10), we can get that $\emptyset_{\mu i}^{(k)}$ is negative definite.

(iii) Finally we end by considering the EM update for the scale parameters s_i: From equations (1) and (3), we obtain

$$\left.\frac{\partial l}{\partial \mathbf{s}_i}\right|_{\mathbf{s}_i=\mathbf{s}_i^{(k)}} = \sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}s_{ij'}^{(k)^{-1}}\left(-1+y_{j'}^{(k)}(t)+y_{j'}^{(k)}(t)\frac{D+1}{E_{ij'}^{(k)}(t)}\right). \tag{17}$$

Permultiplying by $\emptyset_{s i}^{(k)}$, considering the associative property of scalar product of vectors yields

$$\phi_{\mathbf{s}_i}^{(k)}\left.\frac{\partial l}{\partial \mathbf{s}_i}\right|_{\mathbf{s}_i=\mathbf{s}_i^{(k)}}$$

$$= \phi_{\mathbf{s}_i}^{(k)}\sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}s_{ij'}^{(k)^{-1}}\left(-1+y_{j'}^{(k)}(t)+y_{j'}^{(k)}(t)\frac{D+1}{E_{ij'}^{(k)}(t)}\right)$$

$$= -\mathbf{s}_i^{(k)} + \frac{\mathbf{s}_i^{(k)}}{\sum_{t=1}^{N}\alpha_i^{(k)}(t)}\sum_{t=1}^{N}\alpha_i^{(k)}(t)\sum_{j'=1}^{D}y_{j'}^{(k)}(t)\left(1+\frac{D+1}{E_{ij'}^{(k)}(t)}\right)$$

$$= \mathbf{s}_i^{(k+1)} - \mathbf{s}_i^{(k)}. \tag{18}$$

Otherwise, according to the strong law of large numbers, $s_i^{(k)}$ is positive definite with probability one when N is large enough; then from equation (11), we can obviously get that $\emptyset_{s_i}^{(k)}$ is positive definite.
On account of parts (i) (ii) and (iii), we complete the proof.

4. The Convex Combination of Two 3-, and 4-Parameter Logistic Mixtures in 1-Dimentional Case

4.1. Convex Combination of IRFs as a Mixture

The convex combination of two IRFs for the unidimensional dichotomous 1PL to 4PL IRT models can be viewed as a two-components mixture problem using the logistic distribution.

We consider the 2PL model first. The discussion for the 1PL or Rasch model can be obtained as a special case. The cases of lower and upper asymptotes, the 3PL and 4PL models, are discussed thereafter. For the 3PL and 4PL models, the mixing proportion in the results for modality and tangentiality obtained for the 2PL model are expressed in terms of the guessing and careless error (slipping) parameters.

Let $F_{a1,b1^{-1}}$ and $F_{a2,b2^{-1}}$ be IRFs of the 2PL model: for i=1,2, with $a_i \in \mathbb{R}$ and reals $b_i > 0,$

$$F_{a_i,b_i^{-1}}(\theta) = \frac{\exp((\theta - a_i)/b_i^{-1})}{1+\exp((\theta - a_i)/b_i^{-1})},$$

As function of $\theta \in \mathbb{R}$ taking values in (0,1). Let $0 < \lambda < 1$ be the convexity scalar. The convex combination

$$G_{a_1,a_2,b_1^{-1},b_2^{-1},\lambda} = \lambda F_{a_1,b_1^{-1}} + (1-\lambda) F_{a_2,b_2^{-1}},$$

as a function of θ, defines a mapping from R to (0, 1), which we call the 'mixture IRF.'

This convex combination of two 2PL-IRFs can be viewed as a mixture of two logistic distributions. Each component function $F_{ai,bi}(i = 1,2)$ is the cumulative distribution function of the logistic distribution with the location parameter (mean of the logistic distribution) $\mu_i : a_i$(item difficulty in the 2PL model) and scale parameter (proportional to the standard deviation of the logistic distribution) $\sigma_i := 1/b_i$(reciprocal of item discrimination in the 2PL model). With $P := \lambda$as the mixing proportion, $G_{a1,b_1^{-1},b_2^{-1},\lambda}$ is the cumulative distribution function of the two-components logistic mixture. To recap, in terms of the probability density functions $g_{a1,b_1^{-1},b_2^{-1},\lambda} f_{a1,b_1^{-1}},$ and $f_{a2,b_2^{-1}},$ corresponding to $G_{a1,a2,b_1^{-1},b_2^{-1},\lambda} F_{a1,b_1^{-1}},$ respectively, the mixture can be represented as

$$g_{a_1,a_2,b_1^{-1},b_2^{-1},\lambda} = \lambda f_{a_1,b_1^{-1}} + (1-\lambda) f_{a_2,b_2^{-1}}, \tag{19}$$

Where, for i=1,2,

$$f_{a_i,b_i^{-1}}(\theta) = \frac{\exp(-(\theta - a_i)/b_i^{-1})}{b_i^{-1}(1+\exp(-(\theta - a_i)/b_i^{-1}))^2}.$$

Note at this point, in order to view the convex combination of two 3PL- or 4PL-IRFs with lower or upper asymptotes as a mixture of logistic distributions (up to an additive constant, which is the λ convex combination of the two guessing parameters of the mixed IRFs, and which does not affect modality and tangentiality considerations), the mixing proportion must be redefined in terms of the convexity scalar and guessing or slipping parameters. We will discuss that in detail later in the paper.

4.2 Equivalent Formulation of the Modality Problem Using Fewer Parameters

The mixture in Equation (19) with five parameters $\lambda, a_1, a_2, b_1^{-1}$ and b_2^{-2} can be reduced to a mixture with three parameters $\lambda, a := (a_1 - a_2)/b_1^{-1}$ and $b^{-1} := b_2^{-1}/b_1^{-1}$without affecting the modality of that mixture. Then it is easier to explore this property.

This amounts to applying the formula of change of variables for probability density functions. Consider the linear transformation $\psi := (\theta - a_1)/b_1^{-1}$Under this transformation, the mixture in Equation (19) transforms to the mixture

$$g_{0,a,1,b^{-1},\lambda} = \lambda f_{0,1} + (1-\lambda) f_{a,b^{-1}} \tag{20}$$

of the unit logistic distribution (μ=0 and σ=1) with density

$$f_{0,1}(\psi) = \frac{\exp(-\psi)}{(1+\exp(-\psi))^2},$$

and the logistic distribution with parameters $\mu := a(= (a_1 - a_2)/b_1^{-1}$ and $\sigma := b^{-1}(= b_2^{-1})/b_1^{-1}$, with density

$$f_{a,b^{-1}}(\psi) = \frac{\exp(-(\psi - a)/b^{-1})}{b^{-1}(1+\exp(-(\psi - a)/b^{-1}))^2},$$

each as a function of $\Psi \epsilon \mathbb{R}$.

Now we have: If $\theta_0 \epsilon \mathbb{R}$ is a mode of the mixture in Equation (19), then $\psi_0 := (\theta_0 - a_1)/b_1^{-1}$ is a mode of the mixture in Equation (20), and vice versa, if $\Psi \epsilon \mathbb{R}$ is a mode of the mixture in Equation (20), then $\theta_0 := \psi_0 b_1^{-1} + a_1$ is a mode of the mixture in Equation (19). Therefore, in the sequel, the modes of the mixture probability density function $g_{0,a,1,b_1^{-1}b^{-1},\lambda}$ are studied, and as a consequence, the modes $g_{a1,a2,b_1^{-1},b_2^{-1},\lambda}$ of are obtained by applying the transformation.

4.3 Modality Properties of the Logistic Mixture

Without loss of generality, let $a \geq 0$. the theorem of modality properties of the logistic mixture is as follows.

Theorem 2. Let the prerequisites be as mentioned above. Let $g := g_0, a, 1, b^{-1}, \lambda$. It holds:

1.The 'mixture IRF' g is unimodal for all $0 < \lambda < 1, if 0 \leq a \leq a_0(b^{-1})$

$$a_0(b^{-1}) := \log \frac{y^{b^{-1}}[2b^{-1}(y^2-1) + \sqrt{3b^{-2}(y^2-1)^2 + (y^2-4y+1)^2}]}{(y^2-4y+1) + b^{-1}(y^2-1)}.$$

2. For $a > a_0(b^{-1})$,the 'mixture IRF' g is bimodal if and only if λ lies in the open interval (λ_1, λ_2) where

$$\lambda_i^{-1} = 1 + \frac{e^a b^{-1} y^{b^{-1}} (y+1)(e^a - y^{b^{-1}})}{(y-1)(e^a + y^{b^{-1}})^3}$$

and y_i (for i = 1; 2) are the roots of the equation

$$(1-b^{-1})y^{2b^{-1}+2} - 4y^{2b^{-1}+1} + y^{2b^{-1}}(1+b^{-1}) + 4b^{-1}e^a y^{b^{-1}+2}$$
$$-4b^{-1}e^a y^{b^{-1}} - (1+b^{-1})e^{2a}y^2 + 4e^{2a}y + (b^{-1}-1)e^{2a} = 0$$

With $(2-\sqrt{3})e^{ab} < y_1 < y_2 < e^{ab}$. otherwise g is unimodal.

Proof. 1. If a = 0 (i.e., $a_2 = a_1$), the function g is unimodal for all 0 < λ < 1, regardless of the values b^{-1} may take. This is obvious geometrically. Analytically, one would require the derivative g' (with respect to ψ) to be zero. We have, for any parameter values $a \epsilon \mathbb{R}$ and b > 0,

$$f'_{a,b^{-1}}(\psi) = \frac{\exp(-(\psi - a)/b^{-1})\left(\exp(-(\psi - a)/b^{-1}) - 1\right)}{b^{-2}(1+\exp(-(\psi - a)/b^{-1}))^3}$$
$$= f_{a,b^{-1}}(\psi)\frac{\exp(-(\psi - a)/b^{-1}) - 1}{b^{-1}(1+\exp(-(\psi - a)/b^{-1}))},$$

So, if a=0

$$g'(\psi) = \lambda f_{0,1}(\psi)\frac{\exp(-\psi)-1}{1+\exp(-\psi)} + (1-\lambda) f_{0,b^{-1}}(\psi)\frac{\exp(-\psi/b^{-1})-1}{b^{-1}(1+\exp(-\psi/b^{-1}))}.$$

For all values of λ and b^{-1}, the only root of the equation $(\psi) = 0 \, is \, \psi = 0 (= a)$.(For any $\psi \neq 0$ the terms 'exp(ψ) -1' and $'\exp(\psi/b^{-1}) - 1$ are not zero, and they have the same sign.) For the second derivative we have, for any parameter values $a \epsilon \mathbb{R}$ and b > 0,

$$f''_{a,b^{-1}}(\psi) = f_{a,b^{-1}}(\psi)\frac{(\exp(-(\psi - a)/b^{-1}))^2 - 4\exp(-(\psi - a)/b^{-1}) + 1}{b^{-2}(1+\exp(-(\psi - a)/b^{-1}))^2}.$$

Hence, if $a = 0, g''(\psi = 0) = -\lambda/8 - (1-\lambda)/8b^{-3} < 0$ for all values of λ and b^{-1}

Now, let a > 0. Let us assume g be bimodal. Then there is a (local) minimum point ψ_{min}> 0 such that $g'\psi_{min} = 0 \, and g''\psi_{min}$ > 0. In other words, at this point ψ_{min},

$$\lambda f_{0,1}(\psi_{\min})\frac{\exp(-\psi_{\min})-1}{1+\exp(-\psi_{\min})}$$
$$+(1-\lambda) f_{a,b^{-1}}(\psi_{\min})\frac{\exp(-(\psi_{\min}-a)/b^{-1})-1}{b^{-1}(1+\exp(-(\psi_{\min}-a)/b^{-1}))} = 0 \quad (21)$$

and

$$\lambda f_{0,1}(\psi_{\min})\frac{(\exp(-\psi_{\min}))^2-4\exp(-\psi_{\min})+1}{(1+\exp(-\psi_{\min}))^2}$$
$$+(1-\lambda)f_{a,b^{-1}}(\psi_{\min})\frac{(\exp(-(\psi_{\min}-a)/b^{-1}))^2-4\exp(-(\psi_{\min}-a)/b^{-1})+1}{b^{-2}(1+\exp(-(\psi_{\min}-a)/b^{-1}))^2}>0. \tag{22}$$

Solving for $\lambda f_{0,1}(\psi_{min})$ in Equation (21), substituting that in Equation (22), and canceling out $(1-\lambda f_{a,b^{-1}}(\psi_{min})$ gives (note , $\psi_{min}\neq 0$)

$$\frac{(\exp(-(\psi_{\min}-a)/b^{-1}))^2-4\exp(-(\psi_{\min}-a)/b^{-1})+1}{b^{-2}(1+\exp(-(\psi_{\min}-a)/b^{-1}))^2}$$
$$-\frac{(\exp(-(\psi_{\min}-a)/b^{-1})-1)\left((\exp(-\psi_{\min}))^2-4\exp(-\psi_{\min})+1\right)}{b^{-1}(1+\exp(-(\psi_{\min}-a)/b^{-1}))\left((\exp(-\psi_{\min}))^2-1\right)}>0.$$

Let $y=\exp(-(\psi_{\min}-a)/b^{-1})$, then we get $\exp(-\psi_{\min})=\frac{y^{b^{-1}}}{\exp(a)}$. According to formulae (21) and (22), we get

$$\frac{y^2-4y+1}{b^{-1}(1+y)^2}-\frac{(y-1)[(\exp(-\psi_{\min}))^2-4\exp(-\psi_{\min})+1]}{(1+y)((\exp(-\psi_{\min}))^2-1)}>0.$$

Corresponding that $y>1,(\exp(-\psi_{\min}))^2<1,$ we get

$$(y^2-4y+1)((\exp(-\psi_{\min}))^2-1)\;-\;b^{-1}(y^2-1)[(\exp(-\psi_{\min}))^2$$
$$-\;4\exp(-\psi_{\min})+1]<0.$$

Bringing $\exp(-\psi_{\min})=\frac{y^{b^{-1}}}{\exp(a)}$ after arrangement, We get

$$h(y)\equiv(1-b^{-1})y^{2b^{-1}+2}-4y^{2b^{-1}+1}+y^{2b^{-1}}(1+b^{-1})$$
$$+4b^{-1}\exp(a)y^{b^{-1}+2}-4b^{-1}\exp(a)y^{b^{-1}}-(1+b^{-1})\exp(2a)y^2$$
$$+4\exp(2a)y+(b^{-1}-1)\exp(2a)<0 \tag{23}$$

At this moment, we treat exp(a) as a variable, then we can get

$$h(\exp(a))=[b^{-1}(1-y^2)-(y^2-4y+1)]\exp(2a)+4b^{-1}(y^{b^{-1}+2}-y^{b^{-1}})\exp(a)$$
$$+(1-b^{-1})y^{2b^{-1}+2}-4y^{2b^{-1}+1}+(1+b^{-1})y^{2b^{-1}}<0 \tag{24}$$

h(exp(a)) is a quadratic equation about exp(a), the discriminant

$$\Delta_h=16b^{-2}(y^{b^{-1}+2}-y^{b^{-1}})^2$$
$$-4[b^{-1}(1-y^2)-(y^2-4y+1)][(1-b^{-1})y^{2b^{-1}+2}-4y^{2b^{-1}+1}+(1+b^{-1})y^{2b^{-1}}]$$
$$=4\{4b^{-2}(y^2-1)^2+[y^2-4y+1+b^{-1}(y^2-1)][y^2-4y+1-b^{-1}(y^2-1)]\}$$
$$=4[4b^{-2}(y^2-1)^2+(y^2-4y+1)^2-b^{-2}(y^2-1)^2]$$
$$=4[3b^{-2}(y^2-1)^2+(y^2-4y+1)^2] \tag{25}$$

(i) Formula (25) is permanently larger than zero, so the left hand side of (24) can not be permanently less than zero other than

$$b^{-1}(1-y^2)-(y^2-4y+1)=0. \tag{26}$$

Solving equation (26), considering y > 1; b^{-1}> 0, we get

$$y=\frac{2+\sqrt{3+b^{-2}}}{1+b^{-1}}.$$

Following formula (24), we get

$$4b^{-1}(y^{b^{-1}+2}-y^{b^{-1}})\exp(a)+(1-b^{-1})y^{2b^{-1}+2}-4y^{2b^{-1}+1}+(1+b^{-1})y^{2b^{-1}}<0$$

$$\Rightarrow \exp(a)<\frac{(1-b^{-1})y^{2b^{-1}+2}-4y^{2b^{-1}+1}+(1+b^{-1})y^{2b^{-}}}{4b^{-1}(y^{b^{-1}}-y^{b^{-1}+2})}$$

$$=\frac{y^{b^{-1}}[b^{-1}(y^2-1)-(y^2-4y+1)]}{4b^{-1}(y^2-1)}$$

$$=\frac{y^{b^{-1}}}{2}.$$

The above formula is permanently untenable, i.e., in this case, the extreme point is only the maximal point.

(ii) Rethinking about formula (24), if

$$b^{-1}(1-y^2)-(y^2-4y+1)\neq 0,$$

then there are two different roots of $h(e^a) = 0$.

Without loss of generality, let $b^{-1}(1-y^2)-(y^2-4y+1) < 0$; then the parabola opens downward. Solving them, we can get

$$e^{a_{1,2}}=\frac{4b^{-1}(y^{b^{-1}}-y^{b^{-1}+2})\mp\sqrt{4y^{2b^{-1}}[3b^{-2}(y^2-1)^2+(y^2-4y+1)^2]}}{2[b^{-1}(1-y^2)-(y^2-4y+1)]}, \quad (27)$$

Where $y>\frac{2+\sqrt{3+b^{-2}}}{1+b^{-1}}$.

Thus

$$a_1=\log\frac{y^{b^{-1}}[2b^{-1}(1-y^2)+\sqrt{3b^{-2}(y^2-1)^2+(y^2-4y+1)^2}]}{b^{-1}(1-y^2)-(y^2-4y+1)},$$

$$a_2=\log\frac{y^{b^{-1}}[2b^{-1}(y^2-1)+\sqrt{3b^{-2}(y^2-1)^2+(y^2-4y+1)^2}]}{(y^2-4y+1)+b^{-1}(y^2-1)}.$$

It is not difficult to check that $e^{a_1}<\frac{y^{b^{-1}}}{2}$.

So if

$$0\le a\le a_0(b^{-1}): \quad (28)$$

$$=a_2=\log\frac{y^{b^{-1}}[2b^{-1}(y^2-1)+\sqrt{3b^{-2}(y^2-1)^2+(y^2-4y+1)^2}]}{(y^2-4y+1)+b^{-1}(y^2-1)},$$

The mixture g is unimodal for all $0 < \lambda < 1$.

2. Following formula (21), after calculation, we can get

$$\frac{1}{\lambda}=1+\frac{\exp(-\psi_{\min})}{(1+\exp(-\psi_{\min}))^2}\cdot\frac{1-\exp(-\psi_{\min})}{1+\exp(-\psi_{\min})}\cdot\frac{b^{-1}(y+1)}{y-1}$$

$$=1+\frac{e^a b^{-1}y^{b^{-1}}(y+1)(e^a-y^{b^{-1}})}{(y-1)(e^a+y^{b^{-1}})^3} \quad (29)$$

Then for $a > a_0(b^{-1})$, the mixture g is bimodal if and only if λ lies in the open interval (λ_1, λ_2),where

$$\lambda_i^{-1}=1+\frac{e^a b^{-1}y_i^{b^{-1}}(y_i+1)(e^a-y_i^{b^{-1}})}{(y_i-1)(e^a+y_i^{b^{-1}})^3} \quad (30)$$

and y_i (for i = 1; 2) are the roots of the equation

$$(1-b^{-1})y^{2b^{-1}+2}-4y^{2b^{-1}+1}+y^{2b^{-1}}(1+b^{-1})+4b^{-1}e^a y^{b^{-1}+2}$$
$$-4b^{-1}e^a y^{b^{-1}}-(1+b^{-1})e^{2a}y^2+4e^{2a}y+(b^{-1}-1)e^{2a}=0 \quad (31)$$

With $(2-\sqrt{3})e^{ab}<y_1<y_2<e^{ab}$. Otherwise, g is unimodal.

Following formulae (21) and (22), do not eliminate l , we can get

$$K(\lambda) \equiv 2\lambda(1-\lambda)b^{-2} f_{0,1} f_{a,b^{-1}} \exp(-\psi_{\min})[1+\exp(-\psi_{\min})]^2$$
$$\cdot[(\exp(-\psi_{\min}))^2 - 4\exp(-\psi_{\min})+1]$$
$$+3b^{-2}\lambda^2(\exp(-\psi_{\min}))^2[1-\exp(-\psi_{\min})]^2 f_{0,1}^2$$
$$-(1-\lambda)^2[1+\exp(-\psi_{\min})]^6 f_{a,b^{-1}}^2 > 0 \quad (32)$$

Next we treat λ as a variable, and calculate the discriminant of K(λ), denoted by Δ_K .

$$\Delta_K = \{2b^{-2} f_{0,1} f_{a,b^{-1}} \exp(-\psi_{\min})[1+\exp(-\psi_{\min})]^2[(\exp(-\psi_{\min}))^2 - 4\exp(-\psi_{\min})+1]$$
$$+2f_{a,b^{-1}}^2[1+\exp(-\psi_{\min})]^6\}^2$$
$$+4\{-2b^{-2} f_{0,1} f_{a,b^{-1}} \exp(-\psi_{\min})[1+\exp(-\psi_{\min})]^2$$
$$\cdot[(\exp(-\psi_{\min}))^2 - 4\exp(-\psi_{\min})+1]$$
$$+3b^{-2} f_{0,1}^2(\exp(-\psi_{\min}))^2[1-\exp(-\psi_{\min})]^2$$
$$-f_{a,b^{-1}}^2[1+\exp(-\psi_{\min})]^6\} f_{a,b^{-1}}^2[1+\exp(-\psi_{\min})]^6. \quad (33)$$

After calculation and reduction, we can get the formula below

$$b^{-2}[(\exp(-\psi_{\min}))^2 - 4\exp(-\psi_{\min})+1]^2 + 3[1-\exp(-\psi_{\min})]^2 > 0 \quad (34)$$

Formula (34) is permanently larger than zero, so there are two different roots of the equation K(λ) = 0 about λ . According to formula (29), we get

$$\lambda = \frac{1}{1+f_{0,1}\cdot\frac{1-\exp(-\psi_{\min})}{1+\exp(-\psi_{\min})}\cdot\frac{b^{-1}[\exp(-(\psi_{\min}-a)/b^{-1})+1]}{\exp(-(\psi_{\min}-a)/b^{-1})-1}} \quad (35)$$

Calculating $\frac{d\lambda}{d\psi_{\min}}$, after reduction, we get the formula below

$$-b^{-1}[1-\exp(-\psi_{\min})]^2[(\exp(-(\psi_{\min}-a)/b^{-1}))^2-1]$$
$$+2b^{-1}\exp(-\psi_{\min})[(\exp(-(\psi_{\min}-a)/b^{-1}))^2-1]$$
$$+2[1-(\exp(-\psi_{\min}))^2]\exp(-(\psi_{\min}-a)/b^{-1}) \quad (36)$$

When $\exp(-\psi_{min}))^2 - 4\exp(-\psi_{min})+1 \le 0,$ this means $0 < \psi_{min} \le (2+\sqrt{3}),$ formula (36) is permanently larger than zero.

$\frac{d\lambda}{d\psi_{min}} \neq 0,$ this implies that λ is monotone with ψ_{min}. So λ and ψ_{min} given at formula (29) is one on one. Substitution of $\lambda_1 and\ \lambda_2$ in formula (30) yields two values of y,y_1,y_2.

Then we obtain immediately the necessary and sufficient condition started in part 2 of theorem 2.

Thus for $a > a_0(b^{-1}),$ the mixture g is bimodal if only if λ lies in the open interval (λ_1, λ_2), where

$$\lambda_i^{-1} = 1 + \frac{e^a b^{-1} y_i^{b^{-1}}(y_i+1)(e^a - y_i^{b^{-1}})}{(y_i-1)(e^a+y_i^{b^{-1}})^3}$$

and y_i, (for i=1,2) are the roots of the equation (31) with $(2-\sqrt{3})e^{ab} < y_1 < y_2 < e^{ab}$. Otherwise g is unimodal.

Corollary1. If $a \le b^{-1}\ln\frac{2+\sqrt{3+b^{-2}}}{1+b^{-1}}$,g is unimodal.

Corollary 2. If $a \ge \ln\frac{2b^{-1}(2b^{-1}+\sqrt{3+b^{-2}+1})}{b^{-1}-1}$, then $\lambda_1\ and\ \lambda_2$, such that g is bimodal for $\lambda_1 < \lambda < \lambda_2$.

Proof. Proof of Corollary 1 and 2:

Following the theorem 2, we can get that $e^{a_0} \ge y^{b^{-1}}$

Where $y > \frac{2+\sqrt{3+b^{-2}}}{1+b^{-1}}$

and $e^{a_0} \leq a_2(y=2)$ because of

$$\frac{2+\sqrt{3+b^{-2}}}{1+b^{-1}} \leq 2.$$

Then we obtain

$$\left(\frac{2+\sqrt{3+b^{-2}}}{1+b^{-1}}\right)^{b^{-1}} \leq e^{a_0} \leq a_2(y=2) = \frac{2^{b^{-1}}(2b^{-1}+\sqrt{3b^{-2}+1})}{b^{-1}-1}$$

Thus we get Corollary 1 and 2.

4.4 The Plot of Bimodal Regions

Being similar with the paper of Robertson and Fryer (1969), we get the plot in logistic mixture of bimodal regions (also see Liu and Ünlü (2014)) in figure 1 when $b^{-1} \geq 1$

5. Plots between the constrained and EM-equivalent Hessian

Let k[H] = λ_M [H]= λ_M [H] be the condition number of H, where λ_M [H] and λ_M [H] denote the largest and smallest eigenvalues of $H(\Theta^*)$ respectively, and Θ^* is the maximum likelihood estimate of l, then $|K|^{-1}$denotes the reciprocal of condition number without sign.

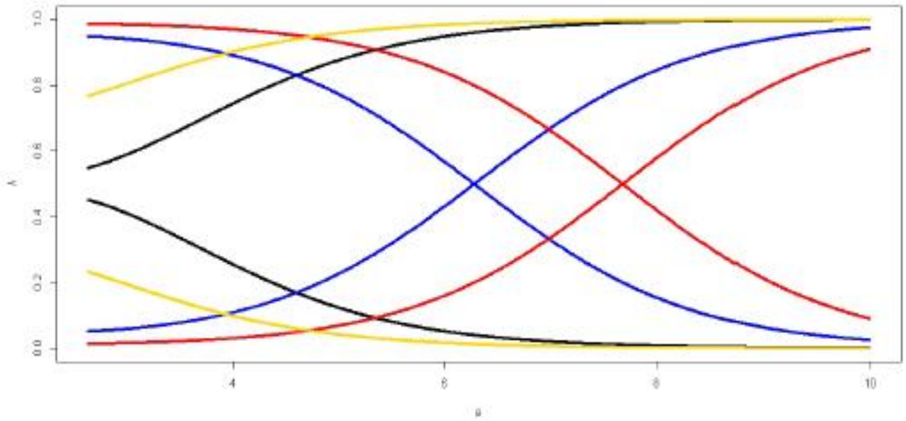

Fig. 1. Biomdal regions for $b^{-1} \geq 1$. $b^{-1} = 1$ Is gold, $b^{-1} = 2$ is black, $b^{-1} = 10$ is blue, $b^{-1} = 20$ is red.

Example 1. Consider the logistic mixture with D = 1 and K = 2, and with the parameters

$$\mu_1 = 2 \quad s_1 = 2 \quad \pi_1 = \frac{1}{2}$$

$$\mu_2 = 1 \quad s_1 = 2 \quad \pi_1 = \frac{1}{2}$$

The next figure (Figure 2) shows the plots between the constrained and EM-equivalent Hessians of Example 1.

6. The Example for PISA

According to formula (2.2) and example 3 in Liu (2015), this section introduces examples applied to PISA (see Adams, Wilson and Wang (1997)) in the context of cognitive diagnosis modeling.

The following parameters in Example 2 show the scores from PISA 2012 between Germany and Luxembourg corresponding to Math and Read.

Example 2. The mixture logistic density with D = 2 and K = 4, let the parameters from PISA 2012 between Germany and Luxembourg corresponding to Math be

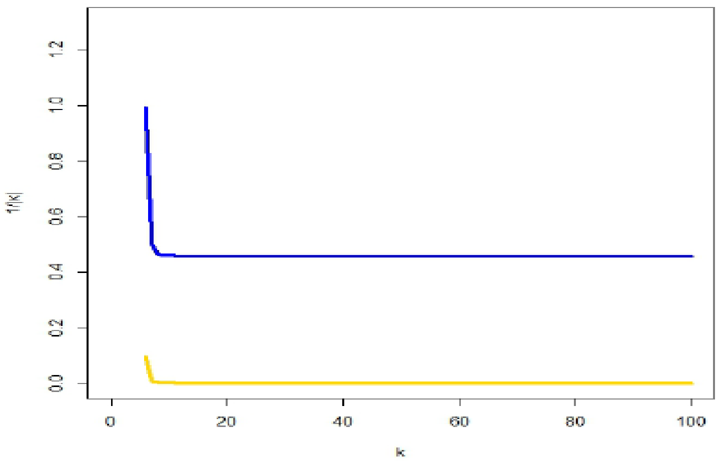

Fig. 2. The constrained Hessian is blue, EM-equivalent Hessian is gold. The terminology "constrained and EM-equivalent Hessians" refers to the matrices $A^T HA$ and $A^T \emptyset HA$ respectively.

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 533 \\ 503 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 4.52 \\ 4.07 \end{pmatrix},$$
$$\boldsymbol{\mu}_2 = \begin{pmatrix} 528 \\ 496 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 4 \\ 2.74 \end{pmatrix},$$
$$\boldsymbol{\mu}_3 = \begin{pmatrix} 511 \\ 482 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 4.92 \\ 2.9 \end{pmatrix},$$
$$\boldsymbol{\mu}_4 = \begin{pmatrix} 496 \\ 478 \end{pmatrix}, \qquad \sigma_4 = \begin{pmatrix} 9.95 \\ 5.59 \end{pmatrix},$$

$$\pi_1 = \pi_2 = \pi_3 = \pi_4 = \frac{1}{4}$$

and let the parameters corresponding to Read be

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 532 \\ 499 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 4.43 \\ 4.74 \end{pmatrix},$$
$$\boldsymbol{\mu}_2 = \begin{pmatrix} 524 \\ 495 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 3.42 \\ 3.32 \end{pmatrix},$$
$$\boldsymbol{\mu}_3 = \begin{pmatrix} 509 \\ 483 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 5.1 \\ 3.09 \end{pmatrix},$$
$$\boldsymbol{\mu}_4 = \begin{pmatrix} 494 \\ 476 \end{pmatrix}, \qquad \sigma_4 = \begin{pmatrix} 8.62 \\ 6.26 \end{pmatrix},$$

$$\pi_1 = \pi_2 = \pi_3 = \pi_4 = \frac{1}{4}$$

Figure 3. shows the contours of the density given in Example 2.

7. Conclusion

Future developments of the work (Liu (2015)) described here consists of improving over the technique for displaying the contour plot when K = 3 fol-lowing the Taylor expansion to take into account of the solution of ridgeline equation. The mathematical expression of the constructive implicit function theorem in logistic case is also very interesting (see Liu (2016)).

Otherwise, an application of this work can be used in PISA (Programme for International Student Assessment) analysis. According to Reckase (2009), this model can also be useful as compensatory extensions in didactics of mathematics.

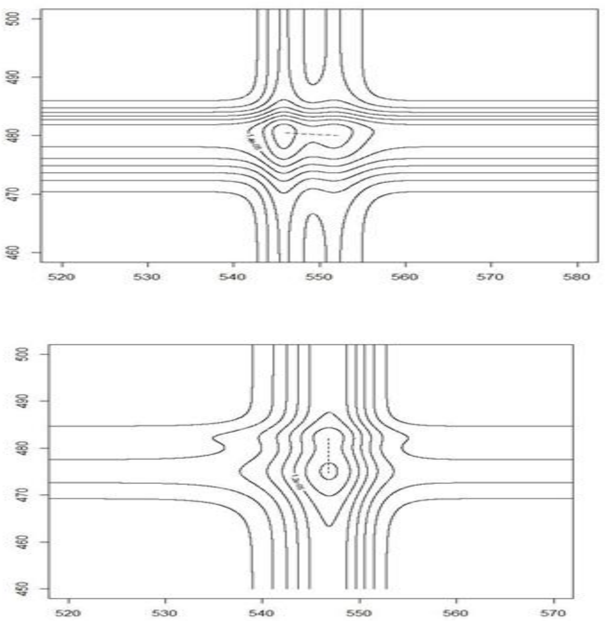

Fig 3. Contour plot and ridgeline curve (- - -) for the mixture density given in Example 2 from PISA 2012 between Germany and Luxembourg corresponding to Math and Read.

References

Adams RJ, Wilson M, Wang W (1997) The multidimensional random coefficients multinomial logit model. Applied Psychological Measurement 21: 1–23

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B39: 1–38

Liu X (2016) The Constructive Implicit Function Theorem and Proof in Logistic Mixtures. Mathematics and Statistics 4(1): 40–45

Liu X (2015) Multivariate Logistic Mixtures. Universal Journal of Applied Mathematics 3(4): 77–87

Liu X, Ünlü A (2014) Multivariate Logistic Mixtures. In: European Conference on Data Analysis (ECDA), Bremen, Germany, pp 105

Malik HJ, Abraham B (1973) Multivariate logistic distributions. The Annals of Statistics 1(3): 588–590

Reckase MD (2009) Multidimensional Item Response Theory. Springer, New York, 2009.

Robertson CA, Fryer JG (1969) Some descriptive properties of normal mixtures. Skand. AktuarTidskr 52: 137–146

Wu CFJ (1983) On the Convergence Properties of the EM Algorithm. The Annals of Statistics 11: 95–103

Xu L, Jordan MI (1996) On Convergence Properties of the EM Algorithm for Gaussian Mixtures. Neural Computation 8: 129–151