



Estimation of generalized linear latent variable models via fully exponential Laplace approximation

Silvia Bianconcini, Silvia Cagnone*

Department of Statistics, University of Bologna, Italy

ARTICLE INFO

Article history:

Received 29 August 2011
Available online 1 July 2012

AMS subject classifications:

62H12
62F12

Keywords:

Laplace approximation
Adaptive Gauss–Hermite
EM algorithm
Ordinal data

ABSTRACT

Latent variable models represent a useful tool in different fields of research in which the constructs of interest are not directly observable. In such models, problems related to the integration of the likelihood function can arise since analytical solutions do not exist. Numerical approximations, like the widely used Gauss–Hermite (GH) quadrature, are generally applied to solve these problems. However, GH becomes unfeasible as the number of latent variables increases. Thus, alternative solutions have to be found. In this paper, we propose an extended version of the Laplace method for approximating the integrals, known as fully exponential Laplace approximation. It is computationally feasible also in presence of many latent variables, and it is more accurate than the classical Laplace approximation. The method is developed within the Generalized Linear Latent Variable Models (GLLVM) framework.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Latent variable models represent a useful tool in the social sciences where the analyzed constructs cannot be directly observed and, hence, they are not measurable. However, a set of indicators related to each unobserved variable can be measured. These models can be defined within the Generalized Linear Latent Variable Model (GLLVM) framework [2,16], according to which the entire set of the responses given by an individual to a certain number of items, called the response pattern, is expressed as a function of one or more latent variables through a monotone differentiable link function. The estimation of the model parameters can be obtained by means of a full information maximum likelihood method via the EM algorithm, that guarantees quite accurate estimates [14,15].

The presence of the latent variables causes problems related to the integration of the likelihood function when the observed variables are discrete, since analytical solutions do not exist. In order to overcome this drawback, numerical approximations are usually applied. One of the most often used techniques is the classical Gauss–Hermite (GH) quadrature [3], that provides quite good parameter estimates when many quadrature points are considered per each latent variable. However, it becomes computationally unfeasible as the number of latent variables increases. This represents a serious limitation for a large number of applications where several observed and latent variables are required [5].

As alternative solution to GH, the Adaptive Gauss–Hermite (AGH) quadrature has been discussed for different models with random effects and/or latent variables [17,19,23]. In all these studies, AGH is shown to perform better than GH, also when few quadrature points are used. Indeed, it consists of adjusting the GH nodes with the first and second moments of the posterior density of the latent factors given the manifest variables. This allows a better approximation of the function to be integrated. Nevertheless, the AGH is very computationally intensive, particularly when the observed variables have many categories [4].

* Corresponding author.

E-mail address: silvia.cagnone@unibo.it (S. Cagnone).

An approximation technique that is not affected by the presence of high dimensional integrals is the Laplace method [7,1], that can be viewed as a particular case of the AGH when just one abscissa is used [11]. Given its reduced dimensionality, the Laplace method is one of the fastest techniques, since the computational burden depends only on the calculation of the mode of the integrand [25,20,8,18]. It has been used to estimate GLLVM by [8]. These authors showed that when this approximation is applied to maximize the likelihood function of a general GLLVM using a Newton–Raphson scheme, the asymptotic error is of order $O(p^{-1})$, with p number of items, and hence it is not directly controllable. Joe [9] has investigated the performance of the Laplace technique for a variety of discrete response mixed models, and he has found that it becomes less adequate as the degree of discreteness increases. A way to overcome these limitations is to consider higher order Laplace approximations as proposed by [20]. However, this extension is quite unfeasible in practice, in particular for GLLVM, since it involves combinations of higher (than two) order derivatives of the likelihood function. Recently, [22] developed the Integrated Nested Laplace Approximation (INLA) to perform Bayesian inference of latent Gaussian models with non-Gaussian observations. This procedure combines Laplace approximations with numerical integration to provide a fast and accurate method for approximating the predictive density of the latent variables/random effects. The approach finds a direct application in generalized linear mixed models, and its main computational advantages are in effect when the inverse covariance matrix of the random effects is sparse and the number of parameters is small.

When either the EM algorithm or a direct maximization of the observed data log-likelihood is used for model estimation, an extended version of the Laplace method, called Fully exponential Laplace Approximation (FLA), can be applied. It has been introduced and developed by [26] in the Bayesian context for approximating posterior distributions. Recently, it has been extended by [21] to a variety of models for longitudinal continuous measurements and time-to-event data estimated via the EM algorithm. The main idea proposed by these authors is to apply the FLA to the expected score function of the model parameters with respect to the posterior distribution of the latent variables. With the FLA, a better approximation of the multidimensional integrals is achieved, being the approximation error of order $O(p^{-2})$. Moreover, the computational complexity of this approach is similar to the classical Laplace method since it depends only on the numerical optimization required to compute the mode of the integrand. In this paper, we extend the FLA to the general class of the GLLVM. In detail, in Section 2 the GLLVM is introduced, in Section 3 the estimation problem is discussed considering the FLA. In Section 4, a simulation study for the particular case of ordinal data is performed in order to compare the finite sample and asymptotic properties of the AGH, classical Laplace and FLA under different conditions. Finally, Section 5 gives the conclusions.

2. Model specification

Let $\mathbf{y} = (y_1, \dots, y_p)$ be a vector of p observed variables of any type and $\mathbf{z} = (z_1, \dots, z_q)$ be a vector of q latent variables. The distribution $f(\mathbf{y})$ of the manifest variables can be expressed as

$$f(\mathbf{y}) = \int_{R^q} g(\mathbf{y}|\mathbf{z})h(\mathbf{z})d\mathbf{z}, \quad (1)$$

where $h(\mathbf{z})$ is the density function of \mathbf{z} and it is assumed to be normal with null mean and correlation matrix equal to Ψ , and $g(\mathbf{y}|\mathbf{z})$ is the conditional density function of \mathbf{y} given \mathbf{z} . Under the assumption of conditional independence of \mathbf{y} given \mathbf{z} , the association among the observed variables is assumed to be wholly explained by the latent variables, so that

$$g(\mathbf{y}|\mathbf{z}) = \prod_{i=1}^p g_i(y_i|\mathbf{z}). \quad (2)$$

According to the GLLVM framework, each conditional distribution $g_i(y_i|\mathbf{z})$ belongs to the exponential family and constitutes the random component of the generalized model. More specifically,

$$g_i(y_i|\mathbf{z}) = \exp \left\{ \frac{y_i \theta_i(\mathbf{z}) - b_i(\theta_i(\mathbf{z}))}{\phi_i} + c_i(y_i, \phi_i) \right\}, \quad (3)$$

where $\theta_i(\mathbf{z})$ is the canonical parameter; $b_i(\theta_i(\mathbf{z}))$ and $c_i(y_i, \phi_i)$ are specific functions that assume different forms according to the different nature of y_i ; ϕ_i is a scale parameter. In particular,

$$b'_i(\theta_i(\mathbf{z})) = E(y_i|\mathbf{z}) = \mu_i(\mathbf{z}).$$

The systematic component of the model is given by

$$\eta_i = \tau_i + \sum_{j=1}^q \alpha_{ij} z_j, \quad (4)$$

where τ_i is the intercept and α_{ij} can be interpreted as factor loadings of the model. The link between the systematic component and the conditional mean of the random component is $\eta_i = \nu_i[\mu_i(\mathbf{z})]$, where ν_i can be any monotonic, differentiable link function.

In the following, we give the specification of GLLVM for responses characterized by different degrees of discreteness, for which the integration problem arises. That is,

- binary data:

$g_i(y_i|\mathbf{z})$ follows a Bernoulli distribution

$$g_i(y_i|\mathbf{z}) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \tag{5}$$

that can be also written in the exponential form as

$$g_i(y_i|\mathbf{z}) = \exp \left[y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + \log(1 - \pi_i) \right] \tag{6}$$

with

$$\pi_i = \frac{\exp \left(\tau_i + \sum_{j=1}^q \alpha_{ij} z_j \right)}{1 + \exp \left(\tau_i + \sum_{j=1}^q \alpha_{ij} z_j \right)},$$

$$\theta_i(\mathbf{z}) = \text{logit}(\pi_i), \quad b_i(\theta_i(\mathbf{z})) = \log(1 + \exp(\theta_i(\mathbf{z}))),$$

$$\phi_i = 1, \quad c_i(y_i, \phi_i) = 0;$$

- count data:

$g_i(y_i|\mathbf{z})$ follows a Poisson distribution

$$g_i(y_i|\mathbf{z}) = \frac{\mu_i(\mathbf{z})^{y_i}}{y_i!} \exp(-\mu_i(\mathbf{z})) \tag{7}$$

that can be also written as

$$g_i(y_i|\mathbf{z}) = \exp [y_i \log \mu_i(\mathbf{z}) - \mu_i(\mathbf{z}) - \ln(y_i!)] \tag{8}$$

with

$$\theta_i(\mathbf{z}) = \ln \mu_i(\mathbf{z}) = \tau_i + \sum_{j=1}^q \alpha_{ij} z_j \quad b_i(\theta_i(\mathbf{z})) = \exp(\theta_i(\mathbf{z})) = \mu_i(\mathbf{z}),$$

$$\phi_i = 1, \quad c_i(y_i, \phi_i) = -\ln(y_i!);$$

- ordinal data:

$g_i(y_i|\mathbf{z})$ follows a multinomial distribution

$$g_i(y_i|\mathbf{z}) = \prod_{s=1}^{c_i} (\gamma_{i,s}(\mathbf{z}) - \gamma_{i,s-1}(\mathbf{z}))^{y_{i,s}}, \tag{9}$$

with c_i number of categories for the item y_i , $y_{i,s} = 1$ if the response is in category s and 0 otherwise,

$\gamma_{i,s}(\mathbf{z}) = P(y_i \leq s|\mathbf{z}) \quad s = 1, \dots, c_i$ defined as

$$\gamma_{i,s}(\mathbf{z}) = \frac{\exp \left(\tau_i - \sum_{j=1}^q \alpha_{ij} z_j \right)}{1 + \exp \left(\tau_i - \sum_{j=1}^q \alpha_{ij} z_j \right)}.$$

In this case, $\tau_{i,s}$ are category-specific intercepts such that $\tau_{i,1} \leq \tau_{i,2} \leq \dots \leq \tau_{i,s} = +\infty$. The negative sign in front of the loadings indicates that the probability to fall in high categories of y_i increases as \mathbf{z} increases.

To complete the specification of the model, denoting for simplicity $\gamma_{i,s} = \gamma_{i,s}(\mathbf{z})$, we can rewrite $g_i(y_i|\mathbf{z})$ in the exponential form as follows

$$g_i(y_i|\mathbf{z}) = \exp \left[\sum_{s=1}^{c_i-1} \left(y_{i,s}^* \log \frac{\gamma_{i,s}}{\gamma_{i,s+1} - \gamma_{i,s}} - y_{i,s+1}^* \log \frac{\gamma_{i,s+1}}{\gamma_{i,s+1} - \gamma_{i,s}} \right) \right], \tag{10}$$

where $y_{i,s}^* = 1$ if the response is in category s or lower of i th variable and 0 otherwise,

$$\theta_{i,s}(\mathbf{z}) = \log \frac{\gamma_{i,s}}{\gamma_{i,s+1} - \gamma_{i,s}}, \quad b_i(\theta_{i,s}(\mathbf{z})) = \log(1 + \exp(\theta_{i,s}(\mathbf{z}))),$$

$$c_i(y_i, \phi_i) = 0, \quad \phi_i = 1.$$

3. Model estimation

Model estimation is achieved by using the maximum likelihood through the EM algorithm, since the latent variables are unknown. At this regard, we apply a full information maximum likelihood method by which all the parameters of the model are estimated simultaneously. For the sake of simplicity, in the following we consider $\Psi = \mathbf{I}$.

For a random sample of size n , from Eq. (1), the observed data log-likelihood is defined as

$$L = \sum_{l=1}^n \log f(\mathbf{y}_l) = \sum_{l=1}^n \log \int_{R^q} g(\mathbf{y}_l | \mathbf{z}_l) h(\mathbf{z}_l) d\mathbf{z}_l. \quad (11)$$

The EM algorithm consists of an Expectation step (E-step), in which the expected score function $E(S(\mathbf{a}_i))$ of the model parameters $\mathbf{a}'_i = (\tau_i, \alpha_{i1}, \dots, \alpha_{iq})$, $i = 1, \dots, p$, is computed. The expectation is with respect to the posterior distribution $h(\mathbf{z}_l | \mathbf{y}_l)$ of \mathbf{z} given the observations for each individual. In the Maximization step (M-step), updated parameter estimates are obtained by equating to 0 the expected score functions.

Louis [12] proved that the observed data score vector $\partial L / \partial \mathbf{a}_i$ is equivalent to the expected score function $E(S(\mathbf{a}_i))$ with respect to $h(\mathbf{z}_l | \mathbf{y}_l)$, so that

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}_i} &= E(S(\mathbf{a}_i)) = \sum_{l=1}^n \frac{\int S(\mathbf{a}_i) g(\mathbf{y}_l | \mathbf{z}_l) h(\mathbf{z}_l) d\mathbf{z}_l}{\int g(\mathbf{y}_l | \mathbf{z}_l) h(\mathbf{z}_l) d\mathbf{z}_l} \\ &= \sum_{l=1}^n \int \frac{\partial \log g(\mathbf{y}_l | \mathbf{z}_l)}{\partial \mathbf{a}_i} h(\mathbf{z}_l | \mathbf{y}_l) d\mathbf{z}_l \quad i = 1, \dots, p \end{aligned} \quad (12)$$

From Eq. (12), it can be noticed that the computation of the expected score functions involves a multidimensional integral that cannot be solved analytically, hence numerical approximations are required. In particular, in the following, we propose the use of the extended version of the classical Laplace approximation, that is the fully exponential Laplace method.

3.1. Fully exponential Laplace approximation method

The FLA method has been proposed for the first time by [26] in order to approximate posterior distributions in the Bayesian context. It represents an extension of the classical Laplace approximation that, as is known, is based on the second order Taylor expansion of the logarithm of the integrand, with the latent variables evaluated at the mode (see, among the others, [25]).

The Laplace method has the advantage of dealing with integrals of any dimensionality without introducing computational problems but, for the general class of latent variable models discussed in this paper, it produces an approximation error of order $O(p^{-1})$, that can be reduced only increasing the number of observed variables. The FLA leads to an improvement of the approximation error maintaining the same computational complexity as the classical Laplace method. The extension of FLA to joint models for continuous longitudinal measurements and time-to-event data has been proposed by [21]. It requires the computation of the following quantities

$$E(A(\mathbf{z}_l)) = \int A(\mathbf{z}_l) h(\mathbf{z}_l | \mathbf{y}_l) d\mathbf{z}_l = \frac{\int A(\mathbf{z}_l) g(\mathbf{y}_l | \mathbf{z}_l) h(\mathbf{z}_l) d\mathbf{z}_l}{\int g(\mathbf{y}_l | \mathbf{z}_l) h(\mathbf{z}_l) d\mathbf{z}_l} \quad (13)$$

that differ from (12) since $A(\cdot)$ are the components of the score functions $S(\mathbf{a}_i)$ that depend on the latent variables.

The main idea of FLA is to approximate both the numerator and the denominator in Eq. (13) with the classical Laplace method. Tierney and Kadane [25] proved that the error terms of order $O(p^{-1})$ in the numerator and the denominator cancel out, leading to a smaller error term of order $O(p^{-2})$. The development of FLA is strictly related to the form of the score function, which strongly depends on the type of data, that is on the associated member of the exponential family. For the specific cases discussed above we obtain

- Binary data

$$\begin{aligned} A_{1,i}(\mathbf{z}_l) &= -\frac{\partial b_i(\theta_i(\mathbf{z}_l))}{\partial \tau_i} = -\pi_i(\mathbf{z}_l) \\ A_{2,i}(\mathbf{z}_l) &= \frac{\partial \log g_i(y_{i,l} | \mathbf{z}_l)}{\partial \alpha_i} = \mathbf{z}_l(y_i - \pi_i(\mathbf{z}_l)). \end{aligned}$$

- Count data

$$\begin{aligned} A_{1,i}(\mathbf{z}_l) &= -\frac{\partial b_i(\theta_i(\mathbf{z}_l))}{\partial \tau_i} = -\mu_i(\mathbf{z}_l) \\ A_{2,i}(\mathbf{z}_l) &= \frac{\partial \log g_i(y_{i,l} | \mathbf{z}_l)}{\partial \alpha_i} = \mathbf{z}_l(y_i - \mu_i(\mathbf{z}_l)). \end{aligned}$$

• Ordinal data

$$\begin{aligned}
 A_{1,i,s}(\mathbf{z}_i) &= -\frac{\partial \theta_{i,s-1,l}(\mathbf{z}_i)}{\partial \tau_{i,s}} = \begin{cases} (1 - \gamma_{i,s,l}), & \text{if } s = 1; \\ (1 - \gamma_{i,s,l}) \frac{\gamma_{i,s,l}}{(\gamma_{i,s,l} - \gamma_{i,s-1,l})}, & \text{if } s = 2, \dots, c_i - 1. \end{cases} \\
 A_{2,i,s}(\mathbf{z}_i) &= \frac{\partial b_i(\theta_{i,s,l}(\mathbf{z}_i))}{\partial \tau_{i,s}} = (1 - \gamma_{i,s,l}) \frac{\gamma_{i,s,l}}{(\gamma_{i,s+1,l} - \gamma_{i,s,l})} \quad s = 1, \dots, c_i - 1. \\
 A_{3,i,s}(\mathbf{z}_i) &= -\frac{\partial \log g_i(y_{i,s,l} | \mathbf{z}_i)}{\partial \alpha_i} = \begin{cases} (1 - \gamma_{i,s,l}) \mathbf{z}_i, & \text{if } s = 1; \\ (1 - \gamma_{i,s,l} - \gamma_{i,s-1,l}) \mathbf{z}_i, & \text{if } s = 2, \dots, c_i. \end{cases}
 \end{aligned}$$

The FLA approximation can be applied only to strictly positive functions $A(\cdot)$. In our case, this condition is not necessarily guaranteed since the $A(\cdot)$ are components of the score functions, not constrained to be positive. To overcome this problem, the method of the moment generating function can be used. According to this approach, since the quantity $\exp\{t'A(\mathbf{z}_i)\}$ is always positive, the FLA approximation can be applied to the moment generating function $M(t) = E[\exp\{t'A(\mathbf{z}_i)\}]$, with latent variables \mathbf{z}_i evaluated at the mode $\hat{\mathbf{z}}_i = \arg \max_{\mathbf{z}_i} [\log g(\mathbf{y}_i | \mathbf{z}_i) + \log h(\mathbf{z}_i) + t'A(\mathbf{z}_i)]$. In doing so, we get the approximated moment generating function $\hat{M}(t)$. Hence, from the corresponding cumulant-generating function $\log \hat{M}(t)$, we obtain the approximated expected values $\hat{E}(A(\mathbf{z}_i))$. These latter values are the quantities of interest, and they are given by

$$\hat{E}(A(\mathbf{z}_i)) = \left. \frac{\partial}{\partial t} \log \hat{M}(t) \right|_{t=0} = \left. \frac{\partial}{\partial t} \log \hat{E}[\exp\{t'A(\mathbf{z}_i)\}] \right|_{t=0}. \tag{14}$$

Tierney et al. [26] proved (Theorem 2, p. 712) that Eq. (14) is equivalent to the following expression

$$\begin{aligned}
 \hat{E}(A(\mathbf{z}_i)) &= A(\hat{\mathbf{z}}_i) + \left. \frac{\partial \log \det(\Sigma_i^{(t)})^{-1/2}}{\partial t} \right|_{(\mathbf{z}_i=\hat{\mathbf{z}}_i, t=0)} + O(p^{-2}) \\
 &= A(\hat{\mathbf{z}}_i) - \frac{1}{2} \text{tr}(\Omega) \Big|_{(\mathbf{z}_i=\hat{\mathbf{z}}_i, t=0)} + O(p^{-2}),
 \end{aligned} \tag{15}$$

where

$$\Sigma_i^{(t)} = -\frac{\partial^2 \{\log g(\mathbf{y}_i | \mathbf{z}_i) + \log h(\mathbf{z}_i) + t'A(\mathbf{z}_i)\}}{\partial \mathbf{z}_i' \partial \mathbf{z}_i} \tag{16}$$

and

$$\Omega = \Sigma_i^{-1} \{ \partial \Sigma_i^{(t)} / \partial t \}.$$

The expressions of the first derivatives of Σ_i with respect to t are reported in the [Appendix](#).

3.2. EM algorithm

The steps of the EM algorithm are defined as follows:

1. Choose initial values for the parameters of the model and for the mode.
2. Compute the mode $\hat{\mathbf{z}}_i, i = 1, \dots, n$, by using a Newton–Raphson iterative scheme. In more detail, for the (m) -th iteration

$$\hat{\mathbf{z}}_i^m = \hat{\mathbf{z}}_i^{(m-1)} - [(\Sigma_i^{(m-1)})^{-1} S(\mathbf{z}_i^{(m-1)})] \Big|_{(\mathbf{z}_i=\hat{\mathbf{z}}_i^{(m-1)}, t=0)},$$

where Σ_i is the Hessian matrix defined in expression (16) and $S(\mathbf{z}_i)$ is defined as follows

$$S(\mathbf{z}_i) = -\frac{\partial \{\log g(\mathbf{y}_i | \mathbf{z}_i) + \log h(\mathbf{z}_i) + t'A(\mathbf{z}_i)\}}{\partial \mathbf{z}_i}. \tag{17}$$

3. E-step. Compute the FLA expected values $\hat{E}(A(\mathbf{z}_i))$, and the approximated expected score function $\hat{E}(S(\mathbf{a}_i))$, where $i = 1, \dots, p$.
4. M-step. Obtain improved estimates for the model parameters. For all of them, a Newton–Raphson iterative scheme is used in order to solve the corresponding nonlinear maximum likelihood equations.
5. Repeat steps 2–4 until convergence is attained.

In the following section, we perform a simulation study for the model with ordinal variables reported in expression (9). The choice of this particular model has been mainly motivated by three reasons. The first one is that ordinal data are very diffused in analysis carried out in the social sciences. The second one is that latent variable models with ordinal data are the most affected by computational burden due to integration problems. Finally, the FLA method for the model with ordinal variables is the most complex to derive among the cases discussed in Section 2. All the technical details for the application of FLA in presence of ordinal data are reported in the [Appendix](#).

Table 1Mean, bias and MSE of the parameter estimates for AGH_{me}, and AGH_{mo}, Lap, and FLA in the generated data.

% Valid samples	AGH						Lap			FLA			
	Mean			Mode			66	Bias	MSE	70	Mean	Bias	MSE
	82			72									
True	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	
$\alpha_{11} = 1.03$	1.37	0.34	0.63	1.47	0.44	0.61	1.17	0.14	0.21	1.02	-0.01	0.03	
$\alpha_{21} = 1.44$	1.41	-0.02	0.41	1.26	-0.18	0.34	0.82	-0.61	0.62	1.62	0.18	0.17	
$\alpha_{31} = 2.11$	2.22	0.10	0.68	1.96	-0.15	0.37	1.21	-0.90	1.04	2.31	0.20	0.33	
$\alpha_{41} = 1.80$	1.91	0.11	0.61	1.75	-0.05	0.43	1.21	-0.59	0.56	1.52	-0.28	0.15	
$\alpha_{51} = 1.53$	1.58	0.05	0.40	1.42	-0.10	0.25	1.23	-0.30	0.30	1.80	0.27	0.26	
$\alpha_{12} = 0.00$	-	-	-	-	-	-	-	-	-	-	-	-	
$\alpha_{22} = 2.42$	1.99	-0.42	0.74	2.09	0.33	0.46	1.85	-0.57	0.57	2.01	-0.41	0.36	
$\alpha_{32} = 1.52$	1.70	0.18	0.51	1.80	0.27	0.44	1.38	-0.14	0.24	1.98	0.46	0.44	
$\alpha_{42} = 0.75$	0.76	0.01	0.49	0.93	0.18	0.35	1.37	0.62	0.56	1.01	0.26	0.12	
$\alpha_{52} = 1.34$	1.50	0.16	0.44	1.43	0.26	0.41	1.96	0.62	0.39	1.70	0.36	0.31	

4. Simulation study

The properties of the FLA method in the GLLVM for ordinal data can be evaluated by performing a simulation study in which several conditions are taken into account. The results will be compared with those obtained with the classical Laplace (Lap) as applied by [8,9], and with the AGH quadrature. In recent years, the latter has been widely applied in latent variable models, since it allows to obtain estimates that are as accurate as those derived by the GH technique, but using a smaller number of quadrature points. It essentially consists of scaling and translating the classical Gaussian quadrature locations to place them under the peak of the integrand, and two different procedures have been adopted in the literature. According to the first one, the mode of the integrand and the inverse of the information matrix of the integrand evaluated at the mode are computed [11,17,23]. The advantage of this approach lies in the fact that the quadrature points are not involved in these computations. However, this method is computationally demanding since it requires numerical optimization routines and the computation of second derivatives. Moreover, when parameter estimates are obtained by using iterative algorithms, like in our case, the first and second order moments have to be computed at each step, hence the algorithm becomes very slow.

An alternative procedure consists of computing the posterior means and covariance matrices at each step of the algorithm [19]. Although this method requires the use of quadrature points themselves, the posterior moments should better describe the integrand in those cases in which its tails are heavier than the normal density. In the following, we show how both these techniques work in latent variable models for ordinal data, and we compare their performances with FLA.

The software used for the analyses are Fortran 95 and R. The codes are available from the authors upon request.

4.1. Finite sample properties of the estimators

To investigate empirically the finite sample performance of the FLA, Lap and AGH, based on both the posterior mean (AGH_{me}) and mode (AGH_{mo}), we generated data from a population that consists of five variables and satisfies a two factor model. The number of categories is the same for each observed variable, and equal to 4. 500 random samples were considered with $n = 200$ subjects. We chose 5 quadrature points per each latent variable for both the adaptive approximations. We also considered 7 quadrature points, but there was a little difference with 5 nodes, suggesting that the latter provides sufficient accuracy for this example.

The population parameters were chosen in such a way that the thresholds range from -3 to 3 . The factor loadings are the following: $\alpha_1 = (1.03, 1.44, 2.11, 1.8, 1.53)$ and $\alpha_2 = (0, 2.42, 1.52, 0.75, 1.34)$ with not null values generated from a log-normal distribution, and one loading fixed to 0 to get a unique solution.

Table 1 reports the mean, bias, and Mean Square Error (MSE) of the parameter estimates obtained by applying all the techniques. The results show that the percentage of valid samples is quite high for all the procedures apart from Lap, ranging from 66% to 82%. The FLA presents much better MSE values than those achieved by Lap, AGH_{me} and AGH_{mo}. The superior performance of FLA compared with Lap is mainly due to the larger bias that characterizes the values obtained with the classical method. This is in agreement with the findings given by [9] in the case of ordinal data. Furthermore, FLA performs better than the quadrature techniques mainly because of a smaller variability of the estimates. Comparing the adaptive techniques, AGH_{me} estimates are less biased than those determined by AGH_{mo}, and present an opposite sign of the bias for several values of α_1 . On the other hand, the latter behaves better in terms of MSE values. The different performance of the two adaptive techniques can be due to the fact that the individual posterior densities to be approximated are not always symmetric. In latent variable models for ordinal data, [6] proved that the posterior densities asymptotically follow a multivariate normal distribution. However, for a small number of observed variables, the integrand could be skewed, and the numerical procedures could provide quite different results.

To analyze the shape of the individual posterior densities in the generated population, we computed measures of multivariate skewness $\beta_{1,q}$ and kurtosis $\beta_{2,q}$ proposed by Mardia (1970). In the case of two latent variables, they are given

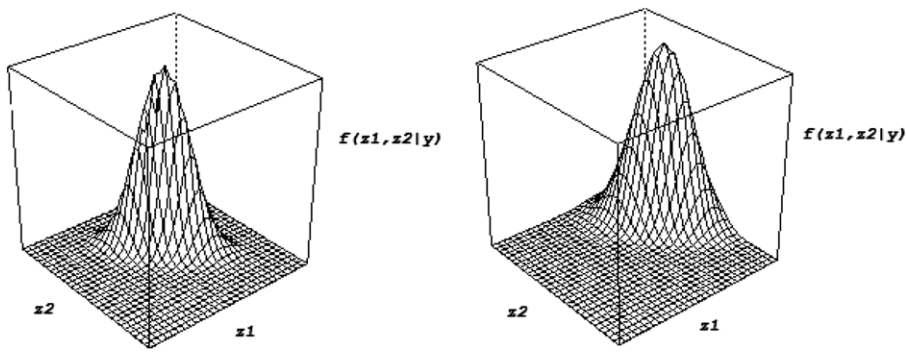


Fig. 1. Individual posterior densities with different shapes (on the left side $\beta_{1,2} = 0.000$, and on the right side $\beta_{1,2} = 0.168$) in the generated population of 200 subjects.

by

$$\beta_{1,2}(l) = \mu_{30}^2 + \mu_{03}^2 + 3\mu_{12}^2 + 3\mu_{21}^2 \quad l = 1, \dots, n$$

and

$$\beta_{2,2}(l) = \mu_{04} + \mu_{40} + 2\mu_{22} \quad l = 1, \dots, n$$

with $\mu_{ij} = E(z_{1i}^i z_{2j}^j)$, whereas z_{1i} and z_{2j} are the latent factors standardized with respect to the posterior densities. Mardia (1970) also derived the asymptotic distributions of both $\beta_{1,q}$ and $\beta_{2,q}$, and the corresponding statistical tests to evaluate the null hypotheses $H_0 : \beta_{1,q} = 0$ and $H_0 : \beta_{2,q} = q(q + 2)$, being $q(q + 2)$ the kurtosis in q -variate normal densities.

By computing these measures for the individual posterior densities generated in this simulation study, we observed that about 35% of these functions have a significant skewness, and a kurtosis always not significantly different from 8. In particular, $\beta_{1,2}$ is on average equal to 0.044 and it ranges from 0.000 to the significant value 0.168. The presence of individual posterior densities having different shapes could justify the different behavior of AGHs and FLA. In Fig. 1, we show two different functions obtained from our generated data. In order to better analyze the finite sample properties of FLA and AGHs, we also generated data from two hypothetical extreme scenarios: one in which all the posterior densities are symmetric, and another one in which a high percentage (more than 60%) of the densities are skewed. As before, we consider five observed variables, each with 4 categories, satisfying a two factor model. The results for both the populations are shown in Table 2.

In the first scenario, the thresholds for each item are equal to -2 for the first category, 0 for the second, and 2 for the third one, whereas the loadings are all fixed to 0.5 except one set equal to zero. In this population, all the individual posterior densities are symmetric, with $\beta_{1,2}$ on average equal to 0.005 , and $\beta_{2,2}$ always not significantly different from 8 . As in the previous simulation study, we generated 500 random samples with 200 subjects.

For all the samples the algorithm achieves the convergence for FLA and AGH_{me} , and in the 96% of the cases for AGH_{mo} . The FLA improves a lot with respect to the previous case, with a reduction of almost one digit in the MSE values, mainly due to smaller bias values for α_2 . On the other hand, both the AGH techniques provide better results in terms of bias and MSE, even if they still perform worse than FLA. We can also notice that the results provided by the two adaptive procedures are almost the same, with an equal sign of the bias for all the estimates, and slight discrepancies due to the different computational techniques involved. Indeed, as discussed by [19], the two procedures should provide similar results when the posterior densities are symmetric.

In the second scenario, the thresholds for each item are equal to -1 for the first category, 0 for the second, and 1 for the third one, whereas the loadings are fixed equal to $\alpha_1 = (2.5, 2.5, 2.5, 2.5, 2.5)$ and $\alpha_2 = (0, 1, 1, 1, 1)$. In this case, 65% of the posterior densities are skewed. $\beta_{1,2}$ ranges from 0.000 to 0.239 , the latter being significantly different from zero, and it is on average equal to 0.127 . On the other hand, there is no significant kurtosis for all the subjects. The main consequence of this high percentage of skew densities is that, for both FLA and AGH_{mo} , a very small number of samples (26% for the former, 18% for the latter) converge properly. Hence, the results are not reliable. On the other hand, AGH_{me} seems to be unaffected by the different shapes of the posterior densities. Its results are more stable in terms of mean, bias, and MSE of the estimates as well as in terms of percentage of valid samples, that also in this case is 73% .

From these results, we can argue that FLA will be superior than AGH when the majority of the posterior densities is symmetric. In these cases, the former provides better MSE values for the estimates than the latter, mainly due to a reduced variability in the estimates. Moreover, the bias introduced in the estimates using FLA is quite comparable with the one in the AGH estimates. On the other hand, we have also shown that the AGH_{me} provides more stable results, that are not affected by the shape of the integrand. Its use is then suggested in populations characterized by a high percentage of skew distributions. It is interesting to highlight that, even if crucial when assessing simulations in such kind of studies, we cannot draw conclusions on the relation between the degree of dependence between items and the behavior of the FLA. Indeed, further investigations of the shape of the individual posterior densities have highlighted that their skewness

Table 2
Mean, bias and MSE for FLA, AGH_{me}, and AGH_{mo} for different scenarios in finite samples.

% Valid samples	AGH						FLA		
	Mean			Mode			100	Bias	MSE
	90			100					
True	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
$\alpha_{11} = 0.5$	0.90	0.40	0.69	0.84	0.34	0.61	0.59	0.09	0.02
$\alpha_{21} = 0.5$	0.52	0.02	0.37	0.55	0.04	0.46	0.61	0.11	0.02
$\alpha_{31} = 0.5$	0.47	-0.02	0.41	0.47	-0.03	0.47	0.60	0.10	0.02
$\alpha_{41} = 0.5$	0.48	-0.02	0.42	0.55	0.05	0.60	0.60	0.10	0.02
$\alpha_{51} = 0.5$	0.51	0.01	0.42	0.53	0.03	0.48	0.60	0.10	0.02
$\alpha_{12} = 0.0$	-	-	-	-	-	-	-	-	-
$\alpha_{22} = 0.5$	0.73	0.23	0.57	0.72	0.22	0.65	0.60	0.10	0.01
$\alpha_{32} = 0.5$	0.73	0.23	0.68	0.70	0.20	0.71	0.59	0.09	0.01
$\alpha_{42} = 0.5$	0.65	0.15	0.54	0.63	0.13	0.67	0.59	0.09	0.01
$\alpha_{52} = 0.5$	0.72	0.22	0.58	0.68	0.19	0.56	0.59	0.09	0.02
% Valid samples	73			18			26		
True	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
$\alpha_{11} = 2.5$	2.47	-0.03	0.31	2.34	-0.16	0.25	2.97	0.47	0.86
$\alpha_{21} = 2.5$	2.71	0.21	0.44	2.41	-0.08	0.14	2.28	-0.22	0.25
$\alpha_{31} = 2.5$	2.73	0.23	0.48	2.44	-0.06	0.11	2.21	-0.29	0.22
$\alpha_{41} = 2.5$	2.71	0.21	0.36	2.57	0.07	0.19	2.28	-0.22	0.20
$\alpha_{51} = 2.5$	2.77	0.27	0.50	2.52	0.02	0.11	2.27	-0.23	0.29
$\alpha_{12} = 0.0$	-	-	-	-	-	-	-	-	-
$\alpha_{22} = 1.0$	1.12	0.13	0.49	0.93	-0.07	0.38	1.92	0.92	0.93
$\alpha_{32} = 1.0$	1.15	0.15	0.59	0.97	-0.03	0.33	1.84	0.84	0.80
$\alpha_{42} = 1.0$	1.09	0.09	0.52	0.98	-0.02	0.34	1.91	0.91	0.89
$\alpha_{52} = 1.0$	1.24	0.24	0.65	0.89	-0.11	0.45	1.97	0.97	0.98

also depends on the thresholds. As these latter increase, the skewness decreases for given values of the loadings. It means that scenarios characterized by high values of thresholds and quite high values of the loadings imply high and significant polychoric correlations between the observations and a high percentage of symmetric posterior densities. For example, using the loadings selected in the second scenario of Table 2 with thresholds ranging from -3 to 3 instead of ranging from -1 to 1, the percentage of asymmetric posterior densities decreases from more than 60% to almost 15%. In this case, differently from what is obtained for the second scenario, FLA performs very well. However, in both cases the polychoric correlations between items are in general quite high, indicating that there is no evidence of a direct relation between the FLA behavior and the degree of dependence between manifest variables.

4.2. Asymptotic properties of estimators

The asymptotic properties of the Laplace maximum likelihood estimators $\hat{\theta}$ have been derived and discussed by [21]. Under suitable regularity conditions, these authors showed that

$$\hat{\theta} - \theta_0 = O_p \left[\max \left\{ n^{-1/2}, p^{-2} \right\} \right],$$

where θ_0 denotes the true parameter value. $\hat{\theta}$ will be consistent as long as both n and p grow to ∞ . FLA is superior than standard Laplace method, since the latter produces estimators with an approximation error of order $O_p \left[\max \left\{ n^{-1/2}, p^{-1} \right\} \right]$. On the other hand, following [11,26], it can be shown that FLA shares the same approximation error of the AGH with 5 quadrature points.

To assess the asymptotic accuracy of the FLA estimator as well as of the classical Laplace, we generated 500 random samples with 1000 subjects from the population described in the previous section. We also applied both the adaptive techniques, and the results are shown in Table 3.

The percentage of valid samples is quite high for all the techniques, apart from Lap, ranging from 65% to 94%. FLA has a good performance as before with small MSE and bias values. On the other hand, both AGHs have an analogous behavior: the MSE values are drastically reduced with respect to the finite sample situation, and the bias is small for all the parameters.

It is worth noting that again FLA performs much better than the classical Laplace approximation, also in this case for the higher biases of the estimates obtained with the latter method. On the other hand, the bias in the AGH and FLA estimates is quite comparable, with a slightly better performance of the former for the second factor loadings (Table 3).

To better investigate the asymptotic properties of the FLA as function not only of the number of individuals n , but also of the number of items p , we performed a numerical study in which the performance of the FLA is compared with AGH_{me} in the presence of ten observed items, each of them characterized by four categories, assumed to satisfy a four factor model. We compared FLA only with AGH_{me} since the computational burden of this simulation is quite heavy and in the previous study, for $n = 1000$, AGH_{me} and AGH_{mo} gave very similar results, whereas Lap performed worse than all the other methods. Moreover, we did not consider higher dimensions to allow a direct comparison with the AGH.

Table 3
Mean, bias and MSE for AGH_{me}, AGH_{mo}, Lap and FLA for generated data with $n = 1000$.

% Valid samples	AGH						Lap			FLA		
	Mean			Mode			65	90	90	90	90	90
	92	92	92	94	94	94						
True	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
$\alpha_{11} = 1.03$	1.11	0.07	0.07	1.14	0.11	0.09	1.12	0.09	0.24	1.00	0.03	0.01
$\alpha_{21} = 1.44$	1.41	-0.03	0.10	1.36	-0.07	0.10	0.78	-0.66	0.66	1.59	0.15	0.07
$\alpha_{31} = 2.11$	2.08	-0.03	0.08	2.04	-0.07	0.09	1.17	-0.93	1.15	2.25	0.14	0.08
$\alpha_{41} = 1.80$	1.82	0.02	0.11	1.79	0.01	0.13	1.08	-0.71	0.77	1.52	0.28	0.09
$\alpha_{51} = 1.53$	1.52	-0.01	0.05	1.48	-0.05	0.06	1.19	-0.33	0.33	1.72	0.19	0.08
$\alpha_{12} = 0.00$	-	-	-	-	-	-	-	-	-	-	-	-
$\alpha_{22} = 2.42$	2.21	-0.21	0.18	2.24	-0.18	0.19	1.86	-0.56	0.53	2.00	-0.42	0.25
$\alpha_{32} = 1.52$	1.58	0.06	0.11	1.64	0.12	0.12	1.31	-0.21	0.32	1.88	0.36	0.18
$\alpha_{42} = 0.75$	0.74	-0.01	0.11	0.79	0.04	0.09	1.34	0.59	0.58	0.96	0.21	0.06
$\alpha_{52} = 1.34$	1.40	0.06	0.07	1.44	0.10	0.07	1.75	0.41	0.26	1.64	0.30	0.13

We generated 500 random samples with $n = 1000$ subjects. The population parameters were chosen in such a way that the thresholds range from -3 to 3 . The factor loadings are the following:

$$\begin{aligned} \alpha_1 &= (0.38, 1.49, 1.87, 1.12, 0.47, 0.40, 0.37, 0.42, 1.14, 2.18), \\ \alpha_2 &= (0, 0.83, 1.61, 1.69, 2.18, 0.79, 0.79, 2.54, 1.08, 0.68), \\ \alpha_3 &= (0, 0, 1.01, 0.82, 0.19, 2.17, 0.37, 1.27, 0.90, 1.40), \quad \text{and} \\ \alpha_4 &= (0, 0, 0, 1.85, 0.49, 1.06, 0.39, 0.66, 0.83, 0.81) \end{aligned}$$

with not null values generated from a log-normal distribution, and six loadings fixed to 0 to get a unique solution.

Table 4 reports the mean, bias, and Mean Square Error (MSE) of the parameter estimates obtained by applying the two techniques. We can observe that a high percentage of the sample reached convergence properly. As before, the variance are generally higher for the quadrature approximation, whereas the reverse occurs for the bias. However, FLA generally seems to behave better than AGH in terms of MSE.

5. Discussion

This paper is concerned with the adequacy of several approximations of the likelihood function in generalized linear latent variable models with particular reference to ordinal data. In particular, we proposed an extended version of the Laplace method for approximating integrals, known as fully exponential Laplace approximation. Classical Laplace methods are known to work poorly in presence of discrete response variables [9], because of the not negligible bias that characterizes the estimates. This result has been confirmed in our study, whereas we have shown how the FLA is generally appropriate in models for ordinal data in both finite and large samples. The comparison with the adaptive Gauss–Hermite quadrature techniques has highlighted that in finite samples the FLA provides better results in terms of MSE values when the majority of the posterior densities is symmetric. Indeed, for a small number of observed variables, the symmetry of the individual posterior densities is not always guaranteed, and the percentage of skewed distributions tends to vary according to the parameter values. When the majority of the densities are skewed, FLA and AGH_{mo} do not achieve convergence in most cases. On the other hand, AGH_{me} is more stable, and it is not affected by the shape of the functions to be approximated.

The main strength of the FLA approach is that it effectively copes with high dimensional latent structures without increasing substantially the computational burden. This is one of the main drawbacks in the application of AGH techniques in latent variable models. Five quadrature points can provide accurate estimates, but the computational effort increases exponentially as the number of latent factors increases. At this regard, we have studied the performance of the FLA in presence of two- and four-dimensional latent structures. This allowed us to analyze the asymptotic properties of all the methods as functions of the number of both the individuals and items. In large samples, in general the FLA achieves the same approximation of the AGH with five quadrature points, and all the techniques behave similarly, apart from the classical Laplace that performs worse.

The main limitation of the FLA approach is that it is not possible to control the magnitude of the approximation error of the integral, as done in AGH by modifying the number of quadrature points. However, as discussed by [21], a virtue of the fully exponential Laplace approximation is that it is very general, and it can be used in almost all the general linear latent variable models. Overall, for latent variable models with ordinal data, the FLA is very adequate to approximate the likelihood function, and it should be considered as a valid alternative to adaptive Gaussian quadrature techniques.

Further lines of research will be oriented to compare the performance of FLA with the multidimensional splines. The latter represents a useful alternative to approximate the posterior densities [24] and to investigate the main assumptions on the prior distribution of the latent variables that is still an open issue in the GLLVM framework [10,13].

Table 4
Mean, bias and MSE for FLA, and AGH_{me}, for generated data with $n = 1000$ and $p = 10$.

% Valid samples	AGH			FLA		
	89			91		
	Mean	Bias	MSE	Mean	Bias	MSE
True						
$\alpha_{011} = 0.38$	0.43	0.05	0.01	0.40	0.02	0.00
$\alpha_{021} = 1.49$	1.30	-0.19	0.13	1.57	0.08	0.02
$\alpha_{031} = 1.87$	1.55	-0.32	0.38	2.32	0.45	0.28
$\alpha_{041} = 1.12$	1.45	0.33	0.24	1.45	0.32	0.28
$\alpha_{051} = 0.47$	0.38	-0.09	0.20	0.61	0.14	0.03
$\alpha_{061} = 0.40$	0.36	-0.04	0.22	0.60	0.20	0.05
$\alpha_{071} = 0.37$	0.32	-0.05	0.03	0.41	0.04	0.01
$\alpha_{081} = 0.42$	0.23	-0.19	0.28	0.78	0.36	0.15
$\alpha_{091} = 1.14$	1.05	-0.09	0.10	1.27	0.13	0.06
$\alpha_{101} = 2.18$	2.39	0.21	0.16	1.96	-0.22	0.19
$\alpha_{012} = 0.00$	-	-	-	-	-	-
$\alpha_{022} = 0.83$	1.20	0.37	0.26	0.98	0.15	0.03
$\alpha_{032} = 1.61$	2.06	0.45	0.56	1.86	0.25	0.12
$\alpha_{042} = 1.69$	2.20	0.51	0.61	1.97	0.28	0.24
$\alpha_{052} = 2.18$	1.85	-0.33	0.37	1.99	-0.19	0.05
$\alpha_{062} = 0.79$	0.81	0.02	0.28	1.00	0.21	0.06
$\alpha_{072} = 0.79$	0.70	-0.09	0.04	0.92	0.13	0.02
$\alpha_{082} = 2.54$	2.24	-0.30	0.29	2.34	-0.20	0.07
$\alpha_{092} = 1.08$	1.14	0.06	0.08	1.33	0.25	0.10
$\alpha_{102} = 0.68$	0.89	0.21	0.33	1.24	0.56	0.43
$\alpha_{013} = 0.00$	-	-	-	-	-	-
$\alpha_{023} = 0.00$	-	-	-	-	-	-
$\alpha_{033} = 1.01$	1.11	0.10	0.25	1.27	0.26	0.11
$\alpha_{043} = 0.82$	1.14	0.32	0.53	1.24	0.42	0.27
$\alpha_{053} = 0.19$	0.35	0.16	0.41	0.54	0.35	0.13
$\alpha_{063} = 2.17$	2.15	-0.02	0.35	1.89	-0.28	0.09
$\alpha_{073} = 0.37$	0.43	0.06	0.05	0.47	0.10	0.01
$\alpha_{083} = 1.27$	1.39	0.12	0.37	1.32	0.05	0.02
$\alpha_{093} = 0.90$	1.00	0.10	0.11	1.18	0.28	0.10
$\alpha_{103} = 1.40$	1.50	0.10	0.35	1.83	0.43	0.34
$\alpha_{014} = 0.00$	-	-	-	-	-	-
$\alpha_{024} = 0.00$	-	-	-	-	-	-
$\alpha_{034} = 0.00$	-	-	-	-	-	-
$\alpha_{044} = 1.85$	1.70	-0.15	0.33	1.42	-0.43	0.26
$\alpha_{054} = 0.49$	0.71	0.22	0.18	0.50	0.01	0.00
$\alpha_{064} = 1.06$	0.81	-0.25	0.53	0.91	-0.15	0.03
$\alpha_{074} = 0.39$	0.47	0.08	0.04	0.44	0.05	0.00
$\alpha_{084} = 0.66$	0.74	0.08	0.29	0.68	0.02	0.01
$\alpha_{094} = 0.83$	0.66	-0.17	0.21	1.10	0.27	0.09
$\alpha_{104} = 0.81$	0.56	-0.25	0.37	0.54	-0.27	0.10

Appendix

In order to apply the fully exponential Laplace approximation to the model for ordinal data, we have to compute the first derivative of Σ_l with respect to t . At this regard, given that

$$\Sigma_l^{(t)} = \sum_{i=1}^p \sum_{s=1}^{c_i-1} \{ \alpha_i \alpha'_i [-y_{i,s,l}^* \gamma_{i,s+1,l} (1 - \gamma_{i,s+1,l}) + y_{i,s+1,l}^* \gamma_{i,s,l} (1 - \gamma_{i,s,l})] \} + \mathbf{I} - \frac{\partial^2 t' A(\mathbf{z}_l)}{\partial \mathbf{z}'_l \partial \mathbf{z}_l}$$

we make use of the following result

$$\frac{\partial \Sigma_l^{(t)}}{\partial t} \Big|_{(\mathbf{z}_l = \hat{\mathbf{z}}_l, t=0)} = \frac{\partial \Sigma_l^{(t)}}{\partial \mathbf{z}} \frac{\partial \hat{\mathbf{z}}_l}{\partial t} \Big|_{(t=0)}$$

according to which

$$\begin{aligned} \frac{\partial \Sigma_l^{(t)}}{\partial t} \Big|_{(\mathbf{z}_l = \hat{\mathbf{z}}_l, t=0)} &= \sum_{i=1}^p \sum_{s=1}^{c_i-1} \alpha_i \alpha'_i [y_{i,s,l}^* \gamma_{i,s+1,l} (1 - 3\gamma_{i,s+1,l} + 2\gamma_{i,s+1,l}^2) + \\ &\quad - y_{i,s+1,l}^* \gamma_{i,s,l} (1 - 3\gamma_{i,s,l} + 2\gamma_{i,s,l}^2)] \times \alpha_i \Sigma_l^{-1} A'(\hat{\mathbf{z}}_l) - A''(\hat{\mathbf{z}}_l), \end{aligned}$$

where $A'(\hat{\mathbf{z}}_l) = \frac{\partial A(\mathbf{z}_l)}{\partial \mathbf{z}_l} \Big|_{\mathbf{z}_l = \hat{\mathbf{z}}_l}$ and $A''(\hat{\mathbf{z}}_l) = \frac{\partial^2 A(\mathbf{z}_l)}{\partial \mathbf{z}'_l \partial \mathbf{z}_l} \Big|_{\mathbf{z}_l = \hat{\mathbf{z}}_l}$.

For the thresholds, the first-order partial derivatives result in

$$A'_{1,i,s}(\mathbf{z}_i) = \alpha_i \gamma_{i,s,l} (1 - \gamma_{i,s,l}), \quad s = 1, \dots, c_i - 1$$

and

$$A'_{2,i,s}(\mathbf{z}_i) = -\alpha_i \gamma_{i,s,l} (1 - \gamma_{i,s,l}) \quad s = 1, \dots, c_i - 1$$

for $A_{1,i,s}(\mathbf{z}_i)$ and $A_{2,i,s}(\mathbf{z}_i)$, respectively. Furthermore, the corresponding second-order partial derivatives are given by

$$A''_{1,i,s}(\mathbf{z}_i) = -\alpha_i \gamma_{i,s,l} (1 - 3\gamma_{i,s,l} + 2\gamma_{i,s,l}^2), \quad s = 1, \dots, c_i - 1$$

and

$$A''_{2,i,s}(\mathbf{z}_i) = \alpha_i \gamma_{i,s,l} (1 - 3\gamma_{i,s,l} + 2\gamma_{i,s,l}^2) \quad s = 1, \dots, c_i - 1.$$

As for the loadings, the elements of the gradient $A_{3,i,s}(\mathbf{z}_i)$ with respect to the latent variables result in

$$\frac{\partial A_{3,i,s}(z_{jl})}{\partial z_{jl}} = \begin{cases} (1 - \gamma_{i,1,l})(1 + \gamma_{i,1,l}\alpha_{ij}z_{jl}), & \text{if } s = 1; \\ (1 - \gamma_{i,s,l} - \gamma_{i,s-1,l}) + \alpha_{ij}z_{jl}(\gamma_{i,s,l}(1 - \gamma_{i,s,l}) + \gamma_{i,s-1,l}(1 - \gamma_{i,s-1,l})), & \text{if } s = 2, \dots, c_i. \end{cases}$$

$$\frac{\partial A_{3,i,s}(z_{jl})}{\partial z_{kl}} = \begin{cases} \alpha_{ik}z_{jl}\gamma_{i,1,l}(1 - \gamma_{i,1,l}), & \text{if } s = 1; \\ \alpha_{ik}z_{jl}(\gamma_{i,s,l}(1 - \gamma_{i,s,l}) + \gamma_{i,s-1,l}(1 - \gamma_{i,s-1,l})), & \text{if } s = 2, \dots, c_i \end{cases}$$

On the other hand, the elements of the corresponding Hessian matrix are given by

$$\frac{\partial^2 A_{3,i,s}(z_{jl})}{\partial z_{jl}^2} = \begin{cases} \alpha_{ij}\gamma_{i,1,l}(1 - \gamma_{i,1,l})(2 - \alpha_{ij}z_{jl}(1 - 2\gamma_{i,1,l})), & \text{if } s = 1; \\ \alpha_{ij}[\gamma_{i,s,l}(1 - \gamma_{i,s,l})(2 - \alpha_{ij}z_{jl}(1 - 2\gamma_{i,s,l})) \\ + \gamma_{i,s-1,l}(1 - \gamma_{i,s-1,l})(2 - \alpha_{ij}z_{jl}(1 - 2\gamma_{i,s-1,l}))], & \text{if } s = 2, \dots, c_i. \end{cases}$$

$$\frac{\partial^2 A_{3,i,s}(z_{jl})}{\partial z_{jl}\partial z_{kl}} = \begin{cases} \alpha_{ik}\gamma_{i,1,l}(1 - \gamma_{i,1,l})(1 - \alpha_{ij}z_{jl}(1 - 2\gamma_{i,1,l})), & \text{if } s = 1; \\ \alpha_{ik}[\gamma_{i,s,l}(1 - \gamma_{i,s,l})(1 - \alpha_{ij}z_{jl}(1 - 2\gamma_{i,s,l})) \\ + \gamma_{i,s-1,l}(1 - \gamma_{i,s-1,l})(1 - \alpha_{ij}z_{jl}(1 - 2\gamma_{i,s-1,l}))], & \text{if } s = 2, \dots, c_i. \end{cases}$$

$$\frac{\partial^2 A_{3,i,s}(z_{jl})}{\partial z_{kl}^2} = \begin{cases} -\alpha_{ik}^2 z_{jl} \gamma_{i,1,l} (1 - 3\gamma_{i,1,l} + 2\gamma_{i,1,l}^2), & \text{if } s = 1; \\ -\alpha_{ik}^2 z_{jl} (\gamma_{i,s-1,l} - 3\gamma_{i,s-1,l}^2 + 2\gamma_{i,s-1,l}^3 + \gamma_{i,s,l} - 3\gamma_{i,s,l}^2 + 2\gamma_{i,s,l}^3), & \text{if } s = 2, \dots, c_i. \end{cases}$$

References

- [1] O.E. Barndorff-Nielsen, D.R. Cox, *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, New York, 1989.
- [2] D.J. Bartholomew, M. Knott, *Latent Variable Models and Factor Analysis*, second ed., Kendall's Library of statistics, London, 1999.
- [3] R.D. Bock, M. Aitkin, Marginal maximum likelihood estimation of item parameters: application of an em algorithm, *Psychometrika* 46 (1981) 433–459.
- [4] S. Cagnone, P. Monari, Latent variable models for ordinal data by using the adaptive quadrature approximation, *Computational Statistics* (2012) <http://dx.doi.org/10.1007/s00180-012-0319-z>.
- [5] S. Cagnone, I. Moustaki, V. Vasdekis, Latent variable models for multivariate longitudinal ordinal responses, *British Journal of Mathematical and Statistical Psychology* 62 (2009) 401–415.
- [6] H.H. Chang, The asymptotic posterior normality of latent trait for a polytomous irt model, *Psychometrika* 61 (1996) 445–463.
- [7] N. De Bruijn, *Asymptotic Methods in Analysis*, Dover Publications, New York, 1981.
- [8] P. Huber, E. Ronchetti, M.P. Victoria-Feser, Estimation of generalized linear latent variable models, *Journal of the Royal Statistical Society B* 66 (2004) 893–908.
- [9] H. Joe, Accuracy of laplace approximation for discrete response mixed models, *Computational Statistics and Data Analysis* 52 (2008) 5066–5074.
- [10] M. Knott, P. Tzamourani, Bootstrapping the estimated latent distribution of the two-parameter latent trait model, *British Journal of Mathematical and Statistical Psychology* 60 (2007) 175–191.
- [11] Q. Liu, D.A. Pierce, A note on gauss-hermite quadrature, *Biometrika* 81 (1994) 624–629.
- [12] T.A. Louis, Finding the observed information matrix when using the em algorithm, *Journal of the Royal Statistical Society, B* 44 (1982) 226–233.
- [13] Y. Ma, M.G. Genton, Explicit estimating equations for semiparametric generalized linear latent variable models, *Journal of the Royal Statistical Society, Series B* 72 (2010) 475–495.
- [14] I. Moustaki, A latent variable model for ordinal data, *Applied Psychological Measurement* 24 (2000) 211–223.
- [15] I. Moustaki, A general class of latent variable models for ordinal manifest variables with covariates effects on the manifest and latent variables, *British Journal of Mathematical and Statistical Psychology* 56 (2003) 337–357.
- [16] I. Moustaki, M. Knott, Generalized latent trait models, *Psychometrika* 65 (2000) 391–411.
- [17] J.C. Pinhero, D.M. Bates, Approximation to the loglikelihood function in the nonlinear mixed effects model, *Journal of Computational and Graphical Statistics* 4 (1995) 12–35.
- [18] J. Pinhero, E.C. Chao, Efficient laplace and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models, *Journal of Computational and Graphical Statistics* 15 (2006) 58–81.
- [19] S. Rabe-Hesketh, A. Skrondal, A. Pickles, Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects, *Journal of Econometrics* 128 (2005) 301–323.
- [20] S.W. Raudenbush, M.-L. Yang, M. Yosef, Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation, *Journal of Computational and Graphical Statistics* 9 (2000) 141–157.
- [21] D. Rizopoulos, G. Verbeke, E. Lesaffre, Fully exponential laplace approximations for the joint modelling of survival and longitudinal data, *Journal of the Royal Statistical Society B* 71 (2009) 637–654.
- [22] H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent gaussian models by using integrated nested laplace approximations, *Journal of the Royal Statistical Society B* 71 (2009) 319–392.
- [23] S. Schilling, R.D. Bock, High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature, *Psychometrika* 70 (2005) 533–555.
- [24] D. Thissen, C.M. Woods, Item response theory with estimation of the latent population distribution using spline-based densities, *Psychometrika* 71 (2) (2006) 281–301.
- [25] L. Tierney, J. Kadane, Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* 81 (1986) 82–86.
- [26] L. Tierney, R. Kass, J. Kadane, Fully exponential laplace approximations to expectations and variances of nonpositive functions, *Journal of the American Statistical Association* 84 (1989) 710–716.