

# Approximate likelihood inference in generalized linear latent variable models based on the dimension-wise quadrature

Silvia Bianconcini\* and Silvia Cagnone

*Department of Statistical Sciences*

*University of Bologna, Italy*

*e-mail:* [silvia.bianconcini@unibo.it](mailto:silvia.bianconcini@unibo.it); [silvia.cagnone@unibo.it](mailto:silvia.cagnone@unibo.it)

Dimitris Rizopoulos

*Department of Biostatistics*

*Erasmus University Rotterdam, the Netherlands*

*e-mail:* [d.rizopoulos@erasmusmc.nl](mailto:d.rizopoulos@erasmusmc.nl)

**Abstract:** We propose a new method to perform approximate likelihood inference in latent variable models. Our approach provides an approximation of the integrals involved in the likelihood function through a reduction of their dimension that makes the computation feasible in situations in which classical and adaptive quadrature based methods are not applicable. We derive new theoretical results on the accuracy of the obtained estimators. We show that the proposed approximation outperforms several existing methods in simulations, and it can be successfully applied in presence of multidimensional longitudinal data when standard techniques are not applicable or feasible.

**MSC 2010 subject classifications:** 62H25, 62F12, 62P10, 62P25.

**Keywords and phrases:** Binary variables, Laplace approximation, longitudinal data,  $M$ -estimators, numerical integration, random effects.

Received September 2017.

## Contents

1	Introduction . . . . .	4405
2	Inference based on the dimension-wise quadrature method . . . . .	4406
	2.1 Setting . . . . .	4406
	2.2 The proposed method for multidimensional integration . . . . .	4407
	2.3 Statistical properties of the estimators . . . . .	4409
	2.4 Comparison with the adaptive Gauss Hermite quadrature . . . . .	4410
3	Finite sample results . . . . .	4411
	3.1 Monte Carlo simulation study . . . . .	4411
	3.2 Investigation of the approximation error . . . . .	4414
4	Data analysis . . . . .	4415
5	Discussion . . . . .	4417

---

\*Corresponding author.

A	Proof of Proposition 2.1 . . . . .	4418
B	Asymptotic behavior of the dimension-wise quadrature approximation	4419
C	Consistency of the dimension-wise quadrature based estimators . . .	4420
D	Computational details of the dimension-wise quadrature in presence of binary data . . . . .	4421
	Acknowledgements . . . . .	4421
	References . . . . .	4421

## 1. Introduction

Latent variable models are used in many research fields where it is of interest to study constructs that are unobservable but can be indirectly measured by indicators related to them. Examples can be found in sociology where questionnaires are utilized to measure attitudes and opinions, in educational testing where students ability is evaluated using exams, and in medicine where quality of life is assessed by function scores (e.g., been able to walk independently).

A general framework that includes a large variety of latent variable models is represented by the generalized linear latent variable models [18, 28]. Their purpose is to describe the relationship between a set of  $p$  responses or items, and a set of  $q$  latent variables or/and random effects (with  $q < p$ ). The latent variables are supposed to account for the association between the response variables, while random effects are usually introduced in the model to account for some unobserved heterogeneity.

The generalized linear latent variable models can be estimated either using the maximum likelihood method [18, 8], or under the bayesian paradigm [5, 6]. Here, we focus on the former. The maximization procedure requires the latent variables and/or random effects to be integrated out from the likelihood function, and in general analytical solutions do not exist. Numerical quadrature based methods represent a widespread solution to this problem and, among them, the adaptive Gauss Hermite quadrature is considered the gold standard [21, 25]. However, it is computationally unfeasible with a large number of latent variables.

Alternatively, the Laplace approximation does not suffer from the curse of dimensionality since it does not require to solve any integral [4]. It has been applied to latent variable models by [8]. In the bayesian framework, [24] combined it with numerical integration to provide a fast and accurate method, the so called integrated nested Laplace approximation, to approximate the predictive density of the latent variables/random effects. The simplicity of the standard Laplace approximation has a cost related to the fact that the order of the approximation error is  $O(p^{-1})$ . Moreover, it becomes less adequate as the degree of discreteness increases [10]. Several authors developed higher order versions of the method to improve the approximation error [30, 27, 23, 22, 7, 2]. However, their implementation requires to compute many partial derivatives of the integrand.

In this paper, we propose a new approach to approximate the integrals involved in the likelihood of latent variable models, that we refer to as dimension-wise quadrature method. It consists of representing the integrand as a sum of not overlapping components that are the terms of the Taylor series expansion of the function. Truncating the expansion reduces the dimension of the integrals, and makes the computation feasible also when the number of latent variables in the model is large. Moreover, in common circumstances in which existing methods may not converge or have very high bias, the proposed technique still performs robustly and efficiently. The corresponding estimators are asymptotically as accurate as the adaptive Gauss Hermite estimators. Differently from higher order Laplace approximations, the proposed approach does not require any derivative computation, hence it is very simple to implement.

The paper is organized as follows. In Section 2, we illustrate the integration problem when a maximum likelihood estimation of the model is used, and we discuss the proposed dimension-wise quadrature method for multidimensional integration. The asymptotic properties of the corresponding estimators are also derived. The finite sample performance of the proposed estimators are analyzed in a simulation study presented in Section 3, together with an investigation of the approximation error. The application to empirical data is illustrated in Section 4, and Section 5 concludes the paper.

## 2. Inference based on the dimension-wise quadrature method

### 2.1. Setting

Maximum likelihood estimates in the generalized linear latent variable framework are typically obtained by using either the expectation maximization algorithm or the direct maximization through Newton-type algorithms. For a random sample of size  $n$ , the observed data log-likelihood function is

$$\ell(\theta) = \sum_{l=1}^n \log f(y_l; \theta) = \sum_{l=1}^n \log \int_{\mathbb{R}^q} g(y_l | b_l) h(b_l) db_l, \quad (2.1)$$

where  $\theta$  denotes the vector of model parameters, and  $f(y_l; \theta)$  is the probability associated to the individual response pattern  $y_l$ . Typically, the assumption of conditional or local independence is made, which states that the  $p$  observed variables  $y_l$  are independent given the  $q$  latent variables  $b_l$ . The conditional distribution  $g(y_l | b_l)$  is referred to as the measurement part of the model and is taken from the exponential family. The prior density  $h(b_l)$  represents the structural part and can be any continuous density function that ensures unimodality of the integrand in (2.1).

In general, the multidimensional integrals involved in (2.1) cannot be solved analytically. An approximation is needed, on which the bias and variance of resulting estimators depend. For the derivations illustrated here, we follow the notation by [27] based on summation convention, and, for simplicity, we omit the individual subscript  $l$ .

**2.2. The proposed method for multidimensional integration**

Let  $L(b)$  denote the logarithm of the integrand in (2.1). Our proposed approximation relies on the asymptotic expansion of  $L(b)$  around its mode  $b_{mo} = \arg \max_{b \in \mathbb{R}^q} L(b)$ , that is

$$L(b) = L(b_{mo}) + \frac{1}{2} L_{i_1, i_2}(b_{mo})(b - b_{mo})^{i_1, i_2} + \nu(b),$$

where  $L_{i_1, i_2}(b_{mo})$  are the second order partial derivatives of  $L$  evaluated at the mode, and  $(b - b_{mo})^{i_1, i_2}$  refers to specific components of the vector  $b - b_{mo}$ .  $\nu(b)$  includes all the terms of order higher than two in the expansion. Grouping these terms with respect to the number of variables involved in the derivation, we obtain

$$\nu(b) = \sum_{w=1}^q \sum_{k=3}^{\infty} \frac{1}{k!} L_{1_{j_1} i_1 \dots 1_{j_w} i_w}(b_{mo})(b - b_{mo})^{1_{j_1} i_1 \dots 1_{j_w} i_w},$$

where  $\{1_{j_1} i_1 \dots 1_{j_w} i_w\}$  is a set of indices in which the index  $i_1$  is repeated  $j_1$  times, the index  $i_2$  is repeated  $j_2$  times, and so on, with  $\sum_{u=1}^w j_u = k$  and  $i_1 < i_2 < \dots < i_w$ .  $L_{1_{j_1} i_1 \dots 1_{j_w} i_w}(b_{mo})$  are the  $k$ th order partial derivatives of  $L$  evaluated at the mode and taken with respect to  $w$  variables. The probability associated to the individual response pattern is

$$f(y; \theta) = (2\pi)^{q/2} |\Sigma_{mo}|^{1/2} \exp\{L(b_{mo})\} E_{\phi}[\exp\{\nu(b)\}], \tag{2.2}$$

where the expected value is taken with respect to a multivariate normal density function  $\phi(b; b_{mo}, \Sigma_{mo})$  whose mean vector is given by the mode  $b_{mo}$  and its covariance matrix is minus the inverse of the Hessian matrix of  $L(b)$  evaluated at its mode, that is  $\Sigma_{mo} = -L^{i_1, i_2}(b_{mo})$ .

**Corollary 2.1.** *The continuous, differentiable, and real valued function  $\exp\{\nu(b)\}$  in (2.2) has the following convergent Taylor series expansion*

$$\exp\{\nu(b)\} = \sum_{w=1}^q t_w = \sum_{w=1}^q \sum_{k=3}^{\infty} \frac{1}{k!} \sum_P L_{p_1}(b_{mo}) \dots L_{p_t}(b_{mo})(b - b_{mo})^{1_{j_1} i_1 \dots 1_{j_w} i_w},$$

being the third sum over all partitions  $P = p_1 | \dots | p_t$  of the  $k$  indices  $\{1_{j_1} i_1 \dots 1_{j_w} i_w\}$  into  $t$  blocks, each of size greater than or equal to three.

Each term  $t_w, 1 \leq w \leq q$ , is the summation of derivatives involving  $w$  variables. For example,  $t_1 = \sum_{k=3}^{\infty} (1/k!) \sum_P L_{p_1}(b_{mo}) \dots L_{p_t}(b_{mo})(b - b_{mo})^{1_{k} i_1}$  involves all the partitions in  $t$  blocks, each of size greater than three, of the set in which the index  $i_1$  is repeated  $k$  times. That is, it encompasses the derivatives of any order taken with respect to just one variable. A different representation of the terms  $t_w, 1 \leq w \leq q$ , is provided in the following Proposition.

**Proposition 2.1.** *Let the function  $\exp\{\nu_r(b)\} = \exp\{\nu(b_{k_1, \dots, k_r})\}, 1 \leq k_1 < \dots < k_r \leq q$ , define a summation of terms that contain at most  $r$  variables, being the other  $q - r$  variables fixed to the corresponding modes. It follows that*

$$t_w = \sum_{r=0}^w (-1)^{w-r} \binom{q-r}{w-r} \exp\{\nu_r(b)\}.$$

Proposition 2.1 is proved in the Appendix A.

Hence, the function  $\exp\{\nu(b)\}$  admits the exact representation

$$\exp\{\nu(b)\} = \sum_{w=1}^q t_w = \sum_{r=0}^q \sum_{w=0}^{q-r} (-1)^w \binom{q-r}{w} \exp\{\nu_r(b)\}, \quad (2.3)$$

known as cut high dimensional model representation [12, 31].

**Corollary 2.2.** *An approximation of the function  $\exp\{\nu(b)\}$  can be obtained by truncating the exact expansion (2.3) as*

$$\exp\{\nu(b)\} \approx \sum_{w=1}^s t_w = \sum_{r=0}^s \sum_{w=0}^{s-r} (-1)^w \binom{q-r}{w} \exp\{\nu_r(b)\},$$

being  $s \ll q$  or, equivalently,

$$\exp\{\nu(b)\} \approx \sum_{r=0}^s (-1)^r \binom{q-s+r-1}{r} \exp\{\nu_{s-r}(b)\}. \quad (2.4)$$

We refer to (2.4) as dimension-wise approximation. The corresponding approximated marginal density function results

$$\begin{aligned} f_a(y; \theta) &= f_L \left[ \sum_{r=0}^s (-1)^r \binom{q-s+r-1}{r} \sum_{t_1, \dots, t_{s-r}} \exp\{\nu_{s-r}(b_{t_1, \dots, t_{s-r}})\} w_{t_1}^* \dots w_{t_{s-r}}^* \right], \end{aligned} \quad (2.5)$$

where  $f_L = (2\pi)^{q/2} |\Sigma_{mo}|^{1/2} \exp\{L(b_{mo})\}$ , and  $\sum_{t_1, \dots, t_{s-r}} = \sum_{t_1=1}^{n_q} \dots \sum_{t_{s-r}=1}^{n_q}$ , and  $n_q$  is the number of quadrature points selected for each latent variable.  $b_{t_1, \dots, t_{s-r}} = (2)^{1/2} C_{mo}(0, b_{t_1}, 0, \dots, 0, b_{t_j}, 0, \dots, 0, b_{t_{s-r}}, 0)^T + b_{mo}$ , with  $C_{mo}$  obtained from the Cholesky decomposition of the matrix  $\Sigma_{mo}$ .  $w_{t_j}^* = w_{t_j} \pi^{-1/2}$ , where  $b_{t_j}$  and  $w_{t_j}$ ,  $j = 1, \dots, s-r$ , are the classical Gauss Hermite nodes and weights, respectively [29]. (2.5) is obtained by applying what we call dimension-wise quadrature.

When  $s = 0$ , the quantity among squared brackets in (2.5) reduces to one, such that  $f_a(y; \theta)$  reduces to the Laplace approximation of the integral. In other words, the dimension-wise quadrature provides an approximation of  $f(y; \theta)$  that is more accurate than the classical Laplace approximation, due to the inclusion of higher (than two) order terms in the Taylor series expansion of  $L(b)$ . On the other hand, when no integral dimension reduction is performed,

$$f_a(y; \theta) = 2^{q/2} |\Sigma_{mo}|^{1/2} \sum_{t_1, \dots, t_q} g(y | b_{t_1, \dots, t_q}^*) h(b_{t_1, \dots, t_q}^*) w_{t_1}^* \dots w_{t_q}^*,$$

<p><b>Step 1</b> Choose <math>s</math> and <math>nq</math></p> <p><b>Step 2</b> For each <math>l</math> (<math>1 \leq l \leq n</math>)</p> <ul style="list-style-type: none"> <li>• Compute the mode <math>b_{m_o,l}</math> of the integrand in (2.1), and its Hessian evaluated at <math>b_{m_o,l}</math></li> <li>• Compute the <math>f_L</math> term of expression (2.5) where <math>\exp\{L(b_{m_o,l})\} = g(y_l b_{m_o,l})h(b_{m_o,l})</math></li> <li>• For each <math>r</math> (<math>0 \leq r \leq s</math>) <ul style="list-style-type: none"> <li>- Compute the term <math>(-1)^r \binom{q-s+r-1}{r}</math></li> <li>- Compute the higher than two order terms  <math>\nu_{s-r}(b_{t_1,\dots,t_{s-r},l}) = L(b_{t_1,\dots,t_{s-r},l}) - L(b_{m_o,l})</math>  <math>-0.5L_{i_1,i_2}(b_{m_o,l})(b_{t_1,\dots,t_{s-r},l} - b_{m_o,l})^{i_1,i_2}</math></li> <li>- Compute the sum <math>\sum_{t_1,\dots,t_{s-r}} \exp\{\nu_{s-r}(b_{t_1,\dots,t_{s-r},l})\} w_{t_1}^* \dots w_{t_{s-r}}^*</math></li> </ul> </li> <li>• Obtain the approximate marginal density function <math>f_a(y_l; \theta)</math> of expression (2.5)</li> </ul> <p><b>Step 3</b> Get the approximation of the log-likelihood function <math>l_a(\theta)</math> as the sum of <math>\log(f_a(y_l; \theta))</math> over <math>l</math></p>
---

TABLE 1

Pseudo-code description of the dimension-wise quadrature method

where  $b_{t_1,\dots,t_q}^* = (b_{t_1}^*, \dots, b_{t_q}^*)^T = (2)^{1/2}C_{m_o}(b_{t_1}, \dots, b_{t_q})^T + b_{m_o}$  and  $w_{t_j}^* = w_{t_j} \exp[b_{t_j}^2]$  are the adaptive Gauss Hermite nodes and weights, respectively.

The approximation of the log-likelihood function (2.1) is derived by (2.5) following the main steps detailed in Table 1, where the implementation of the proposed dimension-wise quadrature is described. In this study, its maximization is performed using a quasi-Newton algorithm in which the gradient and the Hessian matrix are obtained using numerical derivatives.

### 2.3. Statistical properties of the estimators

To investigate the properties of the dimension-wise quadrature based estimators, denoted by  $\theta_{ML}$ , the asymptotic behavior of the approximation (2.5) is analyzed, and a detailed derivation of the corresponding error rate is given in Appendix B. We establish the asymptotic properties of  $\theta_{ML}$  under the following regularity conditions.

**Condition 2.1.**  $\ell(\theta)$  has a unique maximum at  $\theta_0 \in \Theta$ .

**Condition 2.2.** The parameter space  $\Theta$  is compact. Under concavity of the objective function, the approximated log-likelihood  $l_a(\theta)$ , compactness of  $\Theta$  can be replaced by the assumption that the true parameter vector  $\theta_0$  lies in the interior of the parameter space, and the estimator  $\theta_{ML}$  lies in the interior of a neighborhood containing  $\theta_0$  [19].

**Condition 2.3.**  $\ell(\theta)$  is continuous.

**Condition 2.4.** The empirical approximated log-likelihood function  $l_a(\theta)$  converges uniformly in probability to  $\ell(\theta)$ .

A formal proof of these conditions in presence of ordinal and binary data is given by [9] and [1], respectively.

**Proposition 2.2.** *Under Conditions 2.1–2.4,*

$$(\theta_{ML} - \theta_0) = O_p \left\{ \max \left( n^{-1/2}, p^{-\lceil \frac{n_q}{3} + 1 \rceil} \right) \right\}. \quad (2.6)$$

Thus,  $\theta_{ML}$  is consistent as long as both  $n$  and  $p$  grow to  $\infty$ . A formal proof of Proposition 2.2 is given in Appendix C. The  $n^{-1/2}$  term comes from the standard asymptotic theory, whereas the  $p^{-\lceil n_q/3 + 1 \rceil}$  term derives from the dimension-wise quadrature approximation. For  $n_q \geq 3$ , these estimators are more accurate than the classical Laplace one, that are of order  $O(p^{-1})$ . In particular, the proposed estimators share the same accuracy of the adaptive Gauss Hermite one, but the dimension-wise quadrature method avoids the main computational limitations of the latter.

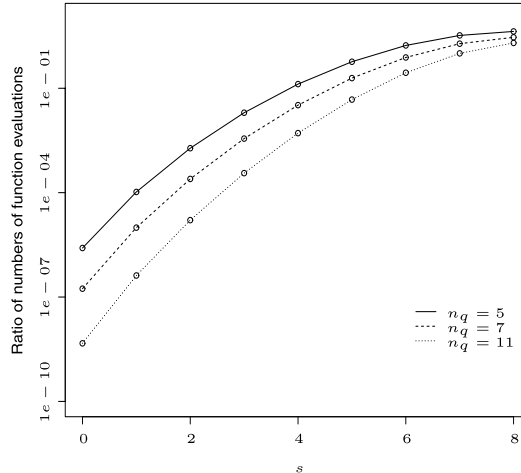
As discussed by [9] and [1] for the Laplace and adaptive Gauss Hermite based estimators, respectively, when numerical techniques are used, the corresponding estimators belong to the class of  $M$ -estimators. Hence,

**Proposition 2.3.** *Under Conditions 2.1–2.4 and if  $H(\theta_0) = -E\{\partial^2 \ell_a(\theta_0) / \partial \theta \partial \theta^T\}$  is nonsingular,  $(n)^{1/2}(\theta_{ML} - \theta_0)$  converges in distribution as  $n \rightarrow \infty$  to a zero mean normal random vector whose covariance matrix is  $H(\theta_0)^{-1} U(\theta_0) \{H(\theta_0)^{-1}\}^T$ , being  $U(\theta_0) = E[\{\partial \ell_a(\theta_0) / \partial \theta\} \{\partial \ell_a(\theta_0) / \partial \theta\}^T]$ .*

#### 2.4. Comparison with the adaptive Gauss Hermite quadrature

The dimension-wise quadrature method is more advantageous than classical and adaptive quadrature techniques from a computational point of view. When evaluating the individual marginal density  $f(y; \theta)$  using the adaptive rule with  $n_q$  quadrature points selected for each dimension, the total number of function or response evaluations is  $n_q^q$ . In contrast,  $\sum_{i=0}^s q! / \{(s-i)!(q-s+i)!\} n_q^{s-i}$  function evaluations are required using the dimension-wise quadrature. Figure 1 shows how the ratio of these two function evaluation numbers varies with respect to  $s$ , the number of terms involved in the dimension-wise quadrature, in presence of eight latent variables and for  $n_q$  equal to 5, 7 and 11. A reduction of the computational effort is achieved when the ratio  $\sum_{i=0}^s q! / \{(s-i)!(q-s+i)!\} n_q^{s-i} / n_q^q < 1$ . The amount of reduction depends on both  $s$  and  $n_q$ . For the univariate dimension-wise quadrature, the ratio has a magnitude of order  $10^{-4}$ ,  $10^{-5}$ , and  $10^{-7}$  when  $n_q = 5, 7$  and  $11$ , respectively. In the bivariate case ( $s = 2$ ), the magnitude of the ratio is of order  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-11}$  when  $n_q = 5, 7$  and  $11$ , respectively. Furthermore, even if not shown here, it is evident that the reduction is dramatically enhanced as the number  $q$  of latent variables in the model increases.

It is worth noting that the dimension-wise quadrature method is feasible in cases in which the classical quadrature techniques are not. At the moment, the adaptive rule is doable in presence of up to ten latent variables in the

FIG 1. Computational load,  $q = 8$ ,  $nq = 5, 7, 11$ .

model, and if a maximum of five quadrature points are used per each dimension (using a Intel(R)Core (TM) i7-2600, 3.40 GHz CPU with 32 Gb RAM). This is particularly restrictive when dealing with longitudinal data, where the number of latent variables in the model increases proportionally with the number of items. On the other hand, the dimension-wise quadrature is feasible for any number of latent variables since it depends on the number of terms ( $s$ ) involved in the approximation.

The choice of  $s$  is fundamental in the dimension-wise quadrature technique. Standard criteria like AIC and BIC are not applicable since they would lead to choose the highest value of  $s$ . Indeed, as  $s$  increases, the value of the log-likelihood typically increases due to the additional information included in it (for further details see [17]). This behaviour of the log-likelihood is shown in the simulation study described in the next Section. For this reason, we follow the same ‘rule of thumb’ applied by [17] for the choice of the number of quadrature points in the traditional quadrature techniques, that is we increase  $s$  until estimates stabilize. This rule is applied in the real data example in Section 4. We underline that, once the value  $s$  is selected, the accuracy of the estimates obtained by applying the dimension-wise quadrature can be improved by increasing the number of quadrature points selected for each dimension.

### 3. Finite sample results

#### 3.1. Monte Carlo simulation study

The performance of the proposed method is evaluated in finite samples by means of a Monte Carlo simulation study in presence of binary data. In this specific case,



$$g(y | b) = \exp [y^T \eta - 1^T \log\{1 + \exp(\eta)\}], \quad (3.1)$$

where  $\eta$  represents the systematic component of the model defined as  $\eta = \alpha_0 + \alpha^T b$ .  $\alpha_0$  is a  $p$ -dimensional vector of item specific intercepts, and  $\alpha$  is a structure matrix that contains unknown factor loadings.  $b = (b_1, \dots, b_q)^T$  is a vector of  $q$  factors that account for the associations among the observed variables, assumed to be normally distributed with zero mean vector and correlation matrix  $\Psi$ . The relationship between the systematic and the random component is expressed through a logit link function. The technical details for the application of the dimension-wise quadrature in this particular case are reported in Appendix D.

We consider a four-factor model with eight indicators. Moreover, we use five quadrature points per dimension to compare the performance of the proposed method with the Laplace approximation and the adaptive Gauss Hermite quadrature method, that occurs when the approximated integral is four-dimensional. The true values of the model parameters have been selected so that at least  $q^2$  constraints are imposed to guarantee invariance to factor rotations [11]. Four constraints are obtained from the correlation matrix  $\Psi$ , being the main diagonal elements equal to one. This also ensures the identification of the scale of the latent variables. Further constraints are obtained by assuming a simple structure for  $\alpha$ , that is by partitioning the manifest variables into four non-overlapping groups, each indicative of a different latent variable.

The true values of the free loadings are fixed to 1.5 and the correlation parameters to 0.6. This results in low correlations between the variables underlying the binary items. In this case, it is known that, in general, the Laplace approximation performs well.

We set  $n = 100, n = 500, n = 1000$  and used 500 replicates for each sample size. Figure 2 reports the box plots of the Monte Carlo estimators for  $\alpha_{11}$  and for  $\psi_{12}$  based on Laplace ( $s = 0$ ), the dimension-wise quadrature method ( $s = 1, 2, 3$ ), and the adaptive Gauss Hermite quadrature for the different sample sizes. The results for all the other parameters are very similar, and they are not reported here for simplicity.

The Laplace estimators are highly biased for both the loading and the correlation parameter. The bias does not decrease as  $n$  increases since the error rate of the approximation, that is  $O(p^{-1})$ , always predominates as  $n \rightarrow \infty$ . On the contrary, both the dimension-wise quadrature with  $s = 1, 2, 3$  and the adaptive Gauss Hermite approximation produce estimators with small bias and the bias tends to zero as  $n$  increases. The Laplace approximation strongly underestimates the variability of the loading estimators compared with the gold standard adaptive Gauss Hermite quadrature. On the other hand, the variability of both the loading and the correlation estimators based on the dimension-wise quadrature is increasingly similar to the corresponding adaptive based estimators as  $s$  increases.

The computational performance of the algorithm based on the different approximation methods is analyzed in Table 2. The average log-likelihood (*Av-loglik*) increases for increasing values of  $s$ . As for the computational time (*Av-time*), the algorithm is slower when  $s = 3$  and with the adaptive quadrature. For

FIG 2. Box-plots of  $\alpha_{11}$  (a) and  $\psi_{12}$  (b) estimates for the four-factor model under Laplace ( $s = 0$ ), dimension-wise quadrature ( $s = 1, 2, 3$ ) and adaptive Gauss Hermite quadrature approximation.

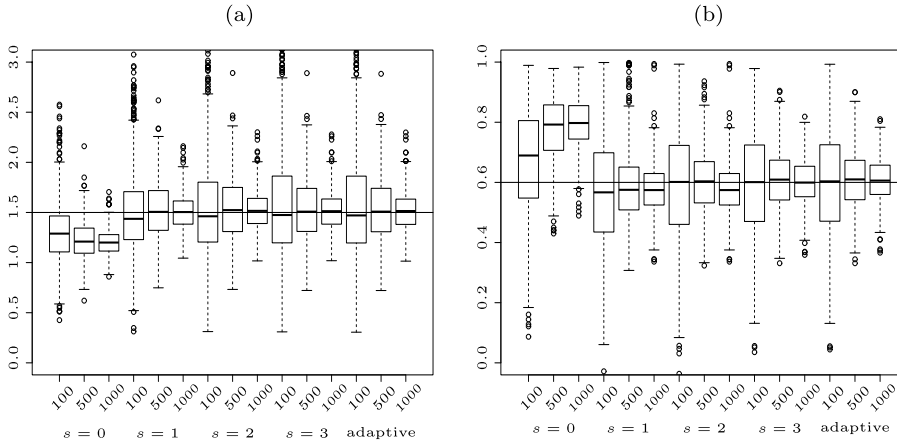


TABLE 2

Computational performance of the algorithms based on the Laplace approximation ( $s = 0$ ), the dimension-wise quadrature ( $s = 1, 2, 3$ ) and the adaptive Gauss Hermite quadrature for the four factor model,  $n = 100, 500, 1000$ .

$n = 100$					
	$s=0$	$s=1$	$s=2$	$s=3$	adaptive
<i>Av-loglik</i>	-504.12	-501.34	-499.21	-498.67	-498.59
<i>Nrfev</i>	33.50	52.35	63.33	70.68	71.41
<i>Avtime (s)</i>	20.42	32.41	73.96	459.18	470.28
$n = 500$					
	$s=0$	$s=1$	$s=2$	$s=3$	adaptive
<i>Av-loglik</i>	-2550.53	-2538.69	-2531.31	-2530.26	-2530.22
<i>Nrfev</i>	55.16	62.19	53.32	55.20	55.50
<i>Avtime (s)</i>	42.85	109.65	479.39	1488.51	1522.72.
$n = 1000$					
	$s=0$	$s=1$	$s=2$	$s=3$	adaptive
<i>Av-loglik</i>	-5113.20	-5090.11	-5075.77	-5073.73	-5073.65
<i>Nrfev</i>	53.83	49.93	45.76	46.96	46.92
<i>Avtime (s)</i>	63.38	152.78	805.01	2245.13	2611.01

example, when  $n = 500$  and  $n = 1000$ , in these cases on average it requires about three times to converge compared to  $s = 2$ . Although, as expected, the Laplace approximation appears to be the fastest method, in certain circumstances it may not converge or has very high bias, as shown above. On the other hand, the proposed method still performs robustly and efficiently. In terms of number of function evaluations (*Nrfev*) the methods perform quite similarly apart from the case of  $n = 100$  where there are noticeable differences.

The behaviour of the approximation methods in presence of high correlations between the underlying variables have been also investigated. The results are similar to those shown here, provided good parameter starting values [17].

### 3.2. Investigation of the approximation error

We analyzed the theoretical error induced by each approximation technique in two simple cases: a three and four-factor model. Table 3 illustrates the error produced by the dimension-wise quadrature and the Laplace method in approximating the marginal density  $f(y, \theta)$  overall possible response patterns. For the former, we used five quadrature points for dimension coherently with the previous simulation study. The benchmark is represented by the adaptive Gauss Hermite quadrature with 15 quadrature points per dimension that provides a very good approximation of the function up to the ninth digit, as shown in the third column of Table 3. The differences are evaluated in terms of Maximum Absolute Error (Max Abs Error) and Average Absolute Error (Avg Abs Error) observed over all possible response patterns computed for each rule with respect to the benchmark approximation.

TABLE 3  
Comparison of integration rules based on Avg Abs Error and Max Abs Error over all possible response patterns for each rule with respect to the benchmark adaptive Gaussian-Hermite quadrature with 15 nodes in each dimension.

Three-factor model					
Rule	$s$	$n_q$	$\sum f(y; \theta)$	Avg Abs Error	Max Abs Error
Adaptive		15	1.0000000	-	-
Gauss Hermite		5	0.9997155	4.54e-06	1.08e-04
Dimension-wise quadrature	1	5	0.9875631	1.94e-04	2.72e-03
	2	5	0.9986880	2.05e-05	4.87e-04
Laplace	0	-	0.9620533	5.93e-04	1.47e-03
Four-factor model					
Rule	$s$	$n_q$	$\sum f(y; \theta)$	Avg Abs Error	Max Abs Error
Adaptive		15	1.0000000	-	-
Gauss Hermite		5	0.99972328	1.10e-06	6.92e-05
Dimension-wise quadrature	1	5	0.98238138	6.88e-05	2.85e-03
	2	5	0.99715424	1.11e-05	8.05e-04
	3	5	0.99965834	1.34e-06	5.84e-05
Laplace	0	-	0.95340396	1.82e-04	8.87e-04

As expected, the biggest differences are observed for the Laplace approximation. In both models, it provides the worst approximation, with a percentage of uncovered area around 5%. On the other hand, the dimension-wise method performs well, and better as  $s$  increases. Specifically, with  $s = 2$  for the three-factor model and  $s = 3$  in the four-factor model, the area under the curve is almost completely covered. The approximation provided by the dimension-wise quadrature method with  $s = 3$  is, on average, equal to that obtained with the adaptive quadrature up to the sixth digit, and provides very similar results to the adaptive quadrature with the same number of quadrature points ( $n_q = 5$ ).

#### 4. Data analysis

We analyze data from the Aging, Demographics and Memory Study, that is the first population based study of dementia in the United States (the data are available on the website <http://www.rand.org>). The participants, aged 70 or older individuals, received a thorough in-home clinical and neuropsychological assessment that allowed researchers to estimate the prevalence, predictors, and outcomes of dementia in the United States elderly population. They were asked to answer to the items of a questionnaire devoted to the cognitive assessment. A detailed description of the measures of the cognitive functioning can be found in [20]. For our analysis, we selected a subset of these measures corresponding to binary items concerning the following subtractions: (1) 100 minus 7 equals what? [SER71], (2) And 7 from that? [SER72], (3) And 7 from that? [SER73], (4) And 7 from that? [SER74], (5) And 7 from that? [SER75]. For all of them, 1 means correct answer and 0 means wrong answer.

Previous exploratory and confirmatory analysis on the Aging, Demographics and Memory Study data [14, 13] highlighted that these items are measures of a latent factor called serial 7's test correspondent to a specific sub-scale related to the mental status of the interviewed people. Three waves of data (2001–2003, 2002–2005, 2006–2008) were collected on a sub-sample of  $n = 71$  individuals. The aim is to investigate how the working memory and mental processing task in which respondents counted backward from 100 by 7s changes over time.

The analysis of multidimensional longitudinal data can be challenging since a large number of random effects/latent variables can be required to take into account for different sources of variability in the data.

We evaluate the performance of the proposed approach by applying the generalized linear latent variable models for multidimensional longitudinal data proposed by [6] and [3] to  $p$  binary variables observed at  $T$  different time points, such that the response vector  $y = (y_{11}, \dots, y_{pT})^T$  is  $pT$ -dimensional. Beyond the association between several items at the same time point, their dependence over time as well as the variability of the same item at different time points have to be considered. The latter is accounted by  $p$  item-specific random effects  $u = (u_1, \dots, u_p)^T$ , whereas the former is explained by means of  $T$  time-dependent latent factors  $z = (z_1, \dots, z_T)^T$ , such that the vector of latent variables  $b = (z, u)^T$  is  $(p + T)$ -dimensional. The correlation of the latent variables  $z$  over time is modeled through a first order non-stationary autoregressive process of parameter  $\phi$  and  $Var(z_1) = \sigma_1^2$ . Hence, the covariance matrix of  $b$  is a block matrix  $\Psi = \text{diag}(\Gamma, \Sigma_u)$  with  $\Sigma_u = \text{diag}_{i=1, \dots, p} \{\sigma_{u_i}^2\}$ . The elements of  $\Gamma$  are the variances and covariances of the latent variables  $z_t$ ,  $t = 1, \dots, T$ , that is  $\gamma_{t,t} = Var(z_t) = \phi^{2(t-1)}\sigma_1^2 + I(t \geq 2) \sum_{k=1}^{t-1} \phi^{2(k-1)}$  and  $\gamma_{t,t'} = Cov(z_t, z_{t'}) = \phi^{t+t'-2}\sigma_1^2 + I(t \geq 2) \sum_{k=0}^{t-2} \phi^{t'-t+2k}$ , where  $I(\cdot)$  is the indicator function and  $t < t'$ . The measurement part of the model is specified as in (3.1). The linear predictor is characterized by a  $pT$ -dimensional vector  $\alpha_0$  of item- and time-specific intercepts, and a matrix  $\alpha$  of dimension  $(pT \times (p + T))$  whose generic sub-matrix of the  $j$ -th item is given by  $[\alpha_j I_T \quad O[j, 1_T]]$ .  $I_T$  is the identity matrix, and  $O[j, 1_T]$  is a null matrix of

dimension  $T \times (T + 1)$  whose  $j$ -th column is substituted by a  $T$ -dimensional vector of ones.

In the longitudinal setting, the number of latent variables increases linearly with the number of observations. In this particular example, the number of latent variables and random effects is equal to eight ( $q = 8$ ), being  $p = 5$  and  $T = 3$ . Although for these data the adaptive Gauss Hermite approximation is feasible if five quadrature points are selected per dimension, its computational burden is too heavy. More than two hours are required for just one function evaluation. Thus, we only apply the dimension-wise quadrature method, whose computational advantages have been discussed before. We consider a fully constrained model (equal thresholds and equal loadings over time) in order to obtain the same metric of the latent variable at the three time points. The resulting measurement invariance allows comparisons of the factor across time [15, 16].

As for the choice of  $s$ , we start fitting the model with  $s$  equal to 0 (Laplace approximation), and we increase its value until the mean of the absolute relative differences in parameter estimates ( $\Delta$ ) is sufficiently small (order of  $10^{-3}$ ).

TABLE 4  
Estimated model parameters for the non-stationary model and absolute relative differences between consecutive values of  $s$ , Aging, Demographics and Memory Study dataset.

	$s = 0$	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$\Delta_{01}$	$\Delta_{12}$	$\Delta_{23}$	$\Delta_{34}$
$\hat{\alpha}_1$	1.000	1.000	1.000	1.000	1.000	-	-	-	-
$\hat{\alpha}_2$	0.512	0.487	0.555	0.585	0.585	0.048	0.139	0.055	0.000
$\hat{\alpha}_3$	0.438	0.437	0.508	0.529	0.529	0.001	0.163	0.040	0.001
$\hat{\alpha}_4$	0.517	0.520	0.589	0.621	0.620	0.006	0.134	0.054	0.001
$\hat{\alpha}_5$	0.609	0.579	0.686	0.727	0.726	0.051	0.185	0.060	0.001
$\hat{\phi}$	1.015	1.007	0.982	0.992	0.992	0.008	0.025	0.011	0.001
$\hat{\sigma}_1^2$	6.799	7.178	6.102	5.753	5.756	0.056	0.150	0.057	0.001
$\hat{\sigma}_{u_1}^2$	1.988	1.550	0.443	0.579	0.566	0.221	0.714	0.307	0.023
$\hat{\sigma}_{u_2}^2$	1.109	1.022	1.159	1.367	1.356	0.079	0.134	0.179	0.008
$\hat{\sigma}_{u_3}^2$	0.425	0.526	0.588	0.744	0.736	0.238	0.119	0.264	0.010
$\hat{\sigma}_{u_4}^2$	0.491	0.549	0.512	0.705	0.693	0.117	0.067	0.376	0.017
$\hat{\sigma}_{u_5}^2$	0.113	0.107	0.299	0.388	0.383	0.052	1.788	0.296	0.014
$Av(\Delta)$	-	-	-	-	-	0.080	0.329	0.155	0.007
$log-lik$	-520.38	-519.97	-518.89	-517.74	-517.78	-	-	-	-
$time (s)$	259.18	547.28	541.76	2930.92	16251.971	-	-	-	-

In Table 4, we report the estimated model parameters under the proposed method using five quadrature points for  $s = 0, 1, 2, 3, 4$ . Parameter estimates sensibly change from  $s = 1$  to  $s = 2$  (the average  $\Delta_{12} = 0.329$ ) whereas they are all very close when  $s = 3$  and  $s = 4$  (the average  $\Delta_{34} = 0.007$ ). This indicates that, for this particular example, the additional information included in the likelihood when  $s$  is greater than 3 is almost irrelevant. From a computational point of view, there is a substantial difference in the time to convergence as the value of  $s$  increases, that is as a greater number of terms are involved in the approximation. The computational time for  $s$  equal to four is more than five times the computational time for  $s$  equal to three. It is evident that the latter can be considered as the reference solution. Table 5 shows the parameter estimates for this case with associated asymptotic standard errors in brackets.

TABLE 5  
*Estimated model parameters with standard errors in brackets,  $s=3$ , Aging, Demographics and Memory Study dataset.*

	$\hat{\alpha}$	$\hat{\sigma}_{u_i}^2$
<i>SER71</i>	1.00	0.44 (0.22)
<i>SER72</i>	0.56 (0.27)	1.16 (1.54)
<i>SER73</i>	0.51 (0.31)	0.59 (0.79)
<i>SER74</i>	0.59 (0.38)	0.51 (0.62)
<i>SER75</i>	0.69 (0.30)	0.30 (0.24)

The significance of the variances of the random effects can be tested by using the likelihood ratio statistics (LR) for consecutive nested models that are asymptotically mixture of two weighted chi-squared distributions [26]. In Table 6, we report the values of these statistics with associated  $p$ -values. All the variances are not significantly different from zero, indicating that there is no relevant individual heterogeneity on each item over time. On the other hand, the autoregressive parameter is significant ( $\phi = 0.99$  with standard error equal to 0.15), and indicates a highly persistent latent process over time.

TABLE 6  
*Likelihoods and likelihood ratio statistics with associated  $p$ -values for comparing nested random-effects models,  $s=3$ , Aging, Demographics and Memory Study dataset.*

	<i>log-lik</i>	LR	<i>p-value</i>
No RE model	-524.66		
One RE model	-523.85	1.617	0.102
Two RE model	-522.71	2.273	0.226
Three RE model	-522.51	0.414	0.875
Four RE model	-519.98	5.057	0.225
Five RE model	-517.74	4.468	0.415

### 5. Discussion

One of the main problems in the estimation of the generalized linear latent variable models is that the integrals involved in the likelihood do not always have an analytical solution. A gold standard approach is represented by the adaptive Gauss Hermite quadrature, that is known to provide quite accurate estimates, but becomes computational unfeasible in presence of a large number of factors. A typical case is represented by generalized linear latent variable models for longitudinal data where the number of latent variables/random effects increases proportionally with the number of items.

In this paper, we have proposed a new method that overcomes these limitations since it decomposes the  $q$ -dimensional integral into the sum of  $1, \dots, s$ -dimensional integrals, being  $s \ll q$ , that can be easily approximated using classical Gauss Hermite quadrature techniques. We have shown that the dimension-wise quadrature uses less information than the adaptive based one but the estimators obtained with the two methods share the same asymptotic properties. Moreover, in finite samples, the dimension-wise quadrature based estimators

present similar accuracy to the adaptive based estimators, whereas Laplace estimators tend to be less accurate in general.

We implemented the proposed method for generalized linear latent variable models for mixed data in R combined with Fortran, and the code is available from the authors upon request. However, the extension of the dimension-wise method to other models is straightforward.

## Appendix A: Proof of Proposition 2.1

Let  $\exp\{\nu_r(b)\}$  define the summation of terms that contains at most  $r$  variables, that is

$$\begin{aligned} \exp\{\nu_r(b)\} &= \exp\{\nu(b_{k_1, \dots, k_r})\} \\ &= \sum_{1 \leq k_1 < \dots < k_r \leq q} \exp\{\nu(b_{m_o, 1}, \dots, b_{m_o, k_1-1}, b_{k_1}, b_{m_o, k_1+1}, \dots, b_{m_o, k_r-1}, \\ &\quad b_{k_r}, b_{m_o, k_r+1}, \dots, b_{m_o, q})\}, \end{aligned} \quad (\text{A.1})$$

being the other  $q - r$  variables fixed to the corresponding modes. Using Taylor expansion at  $b_{m_o}$ ,

$$\begin{aligned} \exp\{\nu(b_{m_o, 1}, \dots, b_{k_1}, b_{m_o, k_1+1}, \dots, b_{m_o, k_r-1}, b_{k_r}, \dots, b_{m_o, q})\} \\ = \sum_{w=0}^r \sum_{k=3}^{\infty} \frac{1}{k!} \sum_P L_{p_1}(b_{m_o}) \dots L_{p_t}(b_{m_o}) (b - b_{m_o})^{1_{j_1} k_{i_1}, \dots, 1_{j_w} k_{i_w}}, \end{aligned}$$

where the third sum is over all partitions  $P = p_1 \mid \dots \mid p_t$  of  $k$  indices  $\{1_{j_1} k_{i_1}, \dots, 1_{j_w} k_{i_w}\}$ ,  $1 \leq i_1 < \dots < i_w \leq r$ ,  $\sum_{u=1}^w j_u = k$ , into  $t$  blocks of size greater than or equal to three, being  $\{1_{j_1} k_{i_1}, \dots, 1_{j_w} k_{i_w}\}$  a set of indices in which  $i_1$  is repeated  $j_1$  times,  $i_2$  is repeated  $j_2$  times, and so on. Hence, (A.1) becomes

$$\begin{aligned} \exp\{\nu_r(b)\} &= \sum_{1 \leq k_1 < \dots < k_r \leq q} \sum_{w=0}^r \sum_{k=3}^{\infty} \frac{1}{k!} \sum_P L_{p_1}(b_{m_o}) \dots \\ &\quad L_{p_t}(b_{m_o}) (b - b_{m_o})^{1_{j_1} k_{i_1}, \dots, 1_{j_w} k_{i_w}} \\ &= \sum_{w=0}^r \sum_{1 \leq k_1 < \dots < k_r \leq q} t_w = \sum_{w=0}^r \binom{q-w}{r-w} t_w, \end{aligned} \quad (\text{A.2})$$

being  $t_w$  defined as in Corollary 2.1. Making use of the relationships induced by the Pascal's triangle of binomial coefficients, (A.2) can be conversely written as

$$t_w = \sum_{r=0}^w (-1)^{w-r} \binom{q-r}{w-r} \exp\{\nu_r(b)\}.$$

**Appendix B: Asymptotic behavior of the dimension-wise quadrature approximation**

Motivated by the simulation study and real data application, we derive the properties of the dimension-wise quadrature based estimators in generalized linear latent variable models for binary data. We consider a simple structure for the loading matrix, that is the manifest variables are partitioned into  $q$  non-overlapping groups of size  $p_i, i = 1, \dots, q$ , each indicative of a different latent variable. The logarithm of the integrand in (1) is

$$L(b) = \log g(y | b) + \log h(b) = \sum_{i=1}^q \sum_{j=1}^{p_i} [y_j(\alpha_{0j} + \alpha_{ji}b_i) + \log(1 + \exp(\alpha_{0j} + \alpha_{ji}b_i))] - \frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\Psi| - \frac{1}{2} b^T \Psi^{-1} b,$$

where  $\sum_{i=1}^q p_i = p$ ,  $\alpha_{0j}, j = 1, \dots, p_i, i = 1, \dots, q$ , are item specific intercepts, and  $\alpha_{ji}, j = 1, \dots, p_i, i = 1, \dots, q$ , are the factor loadings. The  $p_i$ 's have the same order, that is  $\min(p_i) = O(p)$ .  $\Psi$  is a correlation matrix for identification reasons. To analyze the statistical properties of the approximated integral solution (1), we consider the expansion of  $\exp\{\nu(b)\}$  given in Corollary 2.1, that is

$$\exp\{\nu(b)\} = \sum_{w=1}^q \sum_{k=3}^{\infty} \frac{1}{k!} \sum_P L_{p_1}(b_{m_o}) \dots L_{p_t}(b_{m_o}) (b - b_{m_o})^{1_{j_1} i_1 \dots 1_{j_w} i_w} \quad (\text{B.1})$$

where the third sum is over all partitions  $P$  of the  $k$  indices  $\{1_{j_1} i_1 \dots 1_{j_w} i_w\}$ , in which the index  $i_1$  is repeated  $j_1$  times, the index  $i_2$  is repeated  $j_2$  times, and so on, such that  $\sum_{u=1}^w j_u = k$  and  $i_1 < i_2 < \dots < i_w$ . In particular,  $P = p_1 | \dots | p_t$  is a partition of the  $k$  indices  $\{1_{j_1} i_1 \dots 1_{j_w} i_w\}$  into  $t$  blocks, each of size greater than or equal to three. Based on (B.1), the exact solution of the integral is given by

$$f(y; \theta) = f_L \left\{ \sum_{w=0}^q \sum_{k=2}^{\infty} \sum_{P,Q} L_{p_1}(b_{m_o}) \dots L_{p_t}(b_{m_o}) L^{q_1}(b_{m_o}) \dots L^{q_k}(b_{m_o}) \right\},$$

where  $f_L = (2\pi)^{q/2} |\Sigma_{m_o}|^{1/2} \exp\{L(b_{m_o})\}$ . Since only even order moments of multivariate normal variables do not vanish, both partitions  $P$  and  $Q$  refer to  $2k$  indices  $\{1_{j_1} i_1 \dots 1_{j_w} i_w\}$ , such that  $\sum_{u=1}^w j_u = 2k$ . Specifically,  $Q = q_1 | \dots | q_k$  denotes the partition of these  $2k$  indices into  $k$  blocks, each of size two. Each component  $L^{q_m}(b_{m_o}), m = 1, \dots, k$ , refers to specific elements of the matrix  $\Sigma_{m_o}$ . Then, the approximation of  $f(y; \theta)$  based on the dimension-wise quadrature can be rewritten as

$$f_a(y; \theta) = f_L \left\{ \sum_{w=0}^s \sum_{k=2}^{n_q} \sum_{P,Q} L_{p_1}(b_{m_o}) \dots L_{p_t}(b_{m_o}) L^{q_1}(b_{m_o}) \dots L^{q_k}(b_{m_o}) \right\}. \quad (\text{B.2})$$



$L(b)$  and its derivatives are  $O(p)$  when the differentiation is performed with respect to the same component of  $b$ , otherwise they are  $O(1)$  or  $O(0)$ . Furthermore,

**Corollary B.1.** *If  $q = o(p)$ , then the diagonal elements of  $\Sigma_{mo}$  are  $O(p^{-1})$  and the off-diagonal are  $O(p^{-2})$ .*

The asymptotic order of the proposed approximation is derived by analyzing the order of the bipartition  $(P, Q)$  with  $\sum_{u=1}^w j_u = 2(n_q + 1)$  for  $w = 1, \dots, s$ . For fixed  $q$ , when  $w = 1$  the asymptotic order in (B.2) is  $O(p^{t-m})$ . On the other hand, when  $w \geq 2$ , the asymptotic order of the term corresponding to the bipartition  $(P, Q)$  in (B.2) is  $O(p^{t-k})$  if the blocks in the bipartition  $P$  are of even size, and is of order  $O(p^{-k+1})$  when at least one of the blocks in the bipartition  $P$  is of odd size. Hence, the asymptotic error of the proposed approximation based on  $n_q$  quadrature points is the same associated to the bipartition  $(P, Q)$  of  $2(n_q + 1)$  indices  $\{1_{2(n_q+1)}i_1\}$ , that is  $O(p^{-\lceil n_q/3+1 \rceil})$ , where notation  $\lceil r \rceil$  indicates the largest integer not exceeding  $r$ . It is equivalent to the asymptotic error associated to the adaptive Gauss Hermite quadrature [1].

### Appendix C: Consistency of the dimension-wise quadrature based estimators

This section concerns with the consistency of the dimension-wise quadrature based estimators. All the following proofs proceed along the lines of [1], who derived the rate of convergence of the estimator based on the adaptive Gauss Hermite quadrature for generalized linear latent variable models.

Let  $S_a(\cdot)$  denote the approximated score vector based on the dimension-wise quadrature approximation (2.5); then we obtain

$$\begin{aligned} \sum_{l=1}^n E_{b|y} \{S_l(\theta_{ML}; b)\} &= S(\theta_{ML}) = \sum_{l=1}^n \left\{ S_l(\theta_{ML}, b_{mo}) + \dots + O\left(p^{-\lceil \frac{n_q}{3} + 1 \rceil}\right) \right\} \\ n^{-1} S(\theta_{ML}) &= n^{-1} S_a(\theta_{ML}) + O\left(p^{-\lceil \frac{n_q}{3} + 1 \rceil}\right), \end{aligned} \quad (\text{C.1})$$

since  $\theta_{ML}$  is chosen such that  $\sum_{l=1}^n E_{a,b|y} \{S_l(\theta_{ML}; b)\} = S_a(\theta_{ML}) = 0$ . Under common regularity conditions, we apply the Taylor series expansion in  $S(\theta)$  around the true parameter vector  $\theta_0$ :

$$S(\theta_{ML}) = S(\theta_0) + H(\theta^*)(\theta_{ML} - \theta_0), \quad (\text{C.2})$$

where  $\theta^*$  lies on the segment joining  $\theta_0$  and  $\theta_{ML}$ , and  $H(\theta^*) = \partial S(\theta) / \partial \theta|_{\theta=\theta^*} = \sum_{l=1}^n \partial S_l(\theta, b_{mo}) / \partial \theta|_{\theta=\theta^*} = \sum_{l=1}^n H_l(\theta^*, b_{mo})$ . From equations (C.1) and (C.2) we obtain

$$(\theta_{ML} - \theta_0) = - \left\{ n^{-1} \sum_{l=1}^n H_l(\theta^*, b_{mo}) \right\}^{-1} \left\{ n^{-1} S(\theta_0) + O\left(p^{-\lceil \frac{n_q}{3} + 1 \rceil}\right) \right\}.$$

In addition, we have that, as  $n \rightarrow \infty$ ,  $n^{-1}H(\theta^*)$  converges in probability to  $E_y\{H(\theta_0)\}$ , where the expectation is taken with respect to  $f(y; \theta)$ , and  $H(\theta^*) = \sum_{l=1}^n H_l(\theta^*, b_{mo})$ . By further assuming that  $E_y\{H(\theta_0)\}$  is non-singular,  $\{n^{-1}H(\theta^*)\}^{-1}$  converges in probability to  $E_y\{H(\theta_0)\}^{-1}$ . It follows that

$$\begin{aligned} n^{1/2}(\theta_{ML} - \theta_0) &= -E[H(\theta_0)]^{-1} \left[ n^{-1/2}S(\theta_0) + O\left(n^{1/2}p^{-[\frac{nq}{3}+1]}\right) \right] \Rightarrow \\ (\theta_{ML} - \theta_0) &= -E_y[H(\theta_0)]^{-1} \left[ n^{-1}S(\theta_0) + O\left(p^{-[\frac{nq}{3}+1]}\right) \right] \\ &= O_p \left[ \max\left(n^{-1/2}, p^{-[\frac{nq}{3}+1]}\right) \right]. \end{aligned}$$

#### Appendix D: Computational details of the dimension-wise quadrature in presence of binary data

For the latent variable model for binary data described in Subsection 3.1, the quantities  $L(b)$  and  $L_{i_1, i_2}(b)$  involved in the expression of  $f_L$ , and  $\nu(b)$  involved in the computation of the approximated marginal density function  $f_a(y, \theta)$  in (2.5), have the following expressions

$$L(b) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\Psi| + y^T \eta - 1^T \log\{1 + \exp(\eta)\} - \frac{1}{2} b^T \Psi^{-1} b,$$

and

$$L_{i_1, i_2}(b) = -\Sigma^{-1} = \frac{\partial L(b)}{\partial b \partial b^T} = \sum_{i=1}^p \alpha_i^T \alpha_i \{\eta_i / (1 + \eta_i)^2\} - \Psi^{-1}.$$

The mode  $b_{mo}$  is obtained by using a Newton-Raphson iterative scheme as follows

$$b_{mo}^{(it)} = b_{mo}^{(it-1)} - [(\Sigma_{mo}^{(it-1)})^{-1} G(b_{mo}^{(it-1)})],$$

where ‘it’ is the iteration counter and  $G(b)$  is the first derivative of  $L(b)$  with respect to  $b$ . For binary data, it has the following expression

$$G(b) = \frac{\partial L(b)}{\partial b} = \alpha^T \frac{\eta}{1 + \eta} - \Psi^{-1} b.$$

#### Acknowledgements

Silvia Bianconcini and Silvia Cagnone acknowledge the financial support from the grant RBFR12SHVV funded by the Italian Government (FIRB project “Mixture and latent variable models for causal inference and analysis of socio-economic data”).

#### References

- [1] BIANCONCINI S. (2014). Asymptotic properties of adaptive maximum likelihood estimators in latent variable models. *Bernoulli*. 20 (3), pp. 1507–1531. [MR3217452](#)

- [2] BIANCONCINI S. AND CAGNONE S. (2012). Estimation of generalized linear latent variable models via fully exponential Laplace approximation. *Journal of Multivariate Analysis*. 112, pp. 183–193. [MR2957295](#)
- [3] CAGNONE S. AND MOUSTAKI I. AND VASDEKIS V. (2009). Latent variable models for multivariate longitudinal ordinal responses. *British Journal of Mathematical and Statistical Psychology*. 62, pp. 401–415. [MR2668750](#)
- [4] DE BRUIJN N. (1981). *Asymptotic Methods in Analysis*. Dover Publications, New York. [MR0671583](#)
- [5] DUNSON D. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society B*. 62, pp. 355–366. [MR1749544](#)
- [6] DUNSON D. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*. 98, pp. 555–563. [MR2011671](#)
- [7] EVANGELOU E. AND ZHU Z. AND SMITH R. L. (2011). Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation. *Journal of Statistical Planning and Inference*. 141, pp. 3564–3577. [MR2817363](#)
- [8] HUBER P. AND RONCHETTI E. AND VICTORIA-FESER M. P. (2004). Estimation of Generalized Linear Latent Variable Models. *Journal of the Royal Statistical Society B*. 66, pp. 893–908. [MR2102471](#)
- [9] HUBER P. AND SCAILLET O. AND VICTORIA-FESER M. P. (2009). Assessing multivariate predictors of financial market movements: a latent factor framework for ordinal data. *Annals of Applied Statistics*. 3, pp. 249–271. [MR2668707](#)
- [10] JOE H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*. 52, pp. 5066–5074. [MR2526575](#)
- [11] JÖRESKOG K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*. 34, pp. 183–202. [MR0221659](#)
- [12] LI G. AND ROSENTHAL C. AND RABITZ H. (2001). High dimensional model representations. *Journal of Physical Chemistry A*. 105, pp. 7765–7777. [MR1921625](#)
- [13] MCARDLE J. J. (2010). Contemporary Challenges of Longitudinal Measurement using HRS data. In Walford, G. and Tucker E. and Viswanathan M. eds. *The SAGE Handbook of Measurement*. Pp. 509–536.
- [14] MCARDLE J. J. AND FISHER G. G. AND KADLEC K. M. (2007). Latent Variable Analyses of Age Trends of Cognition in the Health and Retirement Study. *Psychology and Aging*. 22 (3), pp. 525–545.
- [15] MEREDITH W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*. 58, pp. 525–543. [MR1248361](#)
- [16] MILLSAP R. E. AND YUN-TEINN J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*. 39, pp. 479–515.
- [17] MOLENBERGHS G. AND VERBEKE G. (2005). *Models for discrete longitudinal data*. New York: Springer-Verlag. [MR2171048](#)

- [18] MOUSTAKI I. AND KNOTT M. (2000). Generalized latent trait models. *Psychometrika*. 65, pp. 391–411. [MR1792703](#)
- [19] NEWEY W. K. AND MCFADDEN D. (1994). Large sample estimation and hypothesis testing. In Engle R. F. and McFadden D. (Eds.), *Handbook of Econometrics*. Amsterdam: North-Holland. [MR1315971](#)
- [20] OFSTEDAL M. B. AND FISHER G. G. AND HERZOG A. R. (2005). *Documentation of cognitive functioning measures in the Health and Retirement Study*. Ann Arbor: University of Michigan. (HRS/AHEAD Documentation Report DR-006).
- [21] RABE-HESKETH S. AND SKRONDALA. AND PICKLES A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*. 128, pp. 301–323.
- [22] RAUDENBUSH S. W. AND YANG M. L. AND YOSEF M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*. 9, pp. 141–157.
- [23] RIZOPOULOS D. AND VERBEKE G. AND LESAFFRE E. (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society B*. 71, pp. 637–654.
- [24] RUE H. AND MARTINO S. AND CHOPIN N. (2009). Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*. 71, pp. 319–391.
- [25] SCHILLING S. AND BOCK R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*. 70, pp. 533–555.
- [26] SELF S. G. AND LIANG K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*. 82, pp. 605–610.
- [27] SHUN Z. AND MCCULLAGH P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society B*. 57, pp. 749–760.
- [28] SKRONDAL A. AND RABE-HESKETH S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.
- [29] STROUD A. AND SECREST D. (1966). *Gaussian Quadrature Formulas*. Prentice Hall.
- [30] TIERNEY L. AND KASS R. AND KADANE J. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*. 84, pp. 710–716.
- [31] XU H. AND RAHMAN S. (2004). A generalized dimension-reduction method for multi-dimensional integration in stochastic mechanics. *International Journal for Numerical Methods in Engineering*. 61, pp. 1992–2019.