ESTIMATION ISSUES IN MULTIVARIATE PANEL DATA

Silvia Bianconcini[1] and Silvia Cagnone[1]

[1] Department of Statistical Sciences, University of Bologna
(e-mail: silvia.bianconcini@unibo.it, silvia.cagnone@unibo.it)

ABSTRACT: Latent variable models are a powerful tool in various research fields when the constructs of interest are not directly observable. However, the likelihood-based model estimation can be problematic when dealing with many latent variables and/or random effects since the integrals involved in the likelihood function do not have analytical solutions. In the literature, several approaches have been proposed to overcome this issue. Among them, the pairwise likelihood method and the dimension-wise quadrature have emerged as effective solutions that produce estimators with desirable properties. In this study, we compare a weighted version of the pairwise likelihood method with the dimension-wise quadrature for a latent variable model for binary longitudinal data by means of a simulation study.

KEYWORDS: latent variables, binary data, weighted pairwise likelihood, dimension-wise quadrature

1 Latent variable models for longitudinal binary data

Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p$ be vectors of p binary observed variables each of them observed at T different occasions, $z_1, z_2, \ldots, z_T$ latent variables that account for the associations among the p items at each time point. Let $u_1, u_2, \ldots, u_p$ be p random effects that account for the associations of the same item at different time points. The joint density of the observed variables can be defined as

$$f(\mathbf{y}) = \int_{R^q} g(\mathbf{y} \mid \mathbf{z}, \mathbf{u}) h(\mathbf{z}, \mathbf{u}) d\mathbf{z} d\mathbf{u}$$

where $g(\mathbf{y} \mid \mathbf{z}, \mathbf{u})$ is referred to as measurement part of the model and $h(\mathbf{z}, \mathbf{u})$ as structural part of the model. The dimension of the integral is $q = p + T$.
The measurement part of the model is defined as a generalized linear model with the random component given by

$$g(\mathbf{y}|\mathbf{z}, \mathbf{u}) = \prod_{t=1}^{T} \prod_{j=1}^{p} g(y_{tj}|z_t, u_j) = \prod_{t=1}^{T} \prod_{j=1}^{p} \pi_{tj}(z_t, u_j)^{y_{tj}} (1 - \pi_{tj}(z_t, u_j))^{(1-y_{tj})},$$

where the first equality comes from the conditional independence assumption between items and over time. Each $g(y_{tj}|z_t,u_j)$ follows a Bernoulli distribution of parameter $\pi_{tj}(z_t,u_j)$, that is the probability of success of item j at time t. The systematic component defines the linear predictor $\eta_{tj} = \alpha_{0tj} + \alpha_{tj}z_t + u_j$ where α_{0tj}'s are item and time-dependent intercepts and α_{tj}'s are item and time-dependent factor loadings. We consider the logit as the link function between the systematic component and the conditional means of the random component.

As for the structural part of the model, we assume that the latent variables follow an autoregressive process of the first order (Cagnone *et al.* , 2009) as follows

$$z_t = \phi z_{t-1} + \delta_t \tag{1}$$

where ϕ is the autoregressive coefficient, $\delta_t \sim N(0,1)$ and $z_1 \sim N(0,\sigma_1^2)$. Moreover, the joint density $h(\mathbf{z},\mathbf{u})$ is a multivariate normal with zero mean vector and block diagonal covariance matrix Ψ that contains the matrices $\Omega = diag_{j=1,\dots,p}\{\sigma^2_{uj}\}$ and the autocovariance matrix Γ of the latent variables.

2 Model estimation

Model estimation is usually performed by using a full maximum likelihood method. Given a sample of size n, the log-likelihood is given by

$$L(\theta) = \sum_{i=1}^{n} \log f(\mathbf{y}_i,\theta) = \sum_{i=1}^{n} \log \int_{R^q} g(\mathbf{y}_i \mid \mathbf{z}_i,\mathbf{u}_i) h(\mathbf{z}_i,\mathbf{u}_l) d\mathbf{z}_i d\mathbf{u}_i \tag{2}$$

where θ is the vector of parameters to be estimated. A problem related to the maximization of the log-likelihood is that, in general, the multidimensional integral in (2) is not solvable analytically. Recent solutions proposed in the literature to solve this problem include the pairwise likelihood (PL) approach (Lindsay, 1988) and the dimension-wise (DW) quadrature method (Bianconcini *et al.* , 2017). In this work, we compare DW with a weighted version of PL (Varin & Czado, 2010).

The PL estimator is obtained by maximizing bivariate likelihood products that contain the greatest quantity of model parameter information. In the latent variable model for longitudinal binary data described in Section 2, the bivariate density for a pair of responses is

$$f(y_{ijt}, y_{ij't'};\theta) = \int g(y_{ijt}|z_{it},u_{ij}) g(y_{ij't'}|z_{it'},u_{ij'}) h(z_t,z_{t'},u_j,u_{j'}) dz_t dz_{t'} du_j du_{j'}.$$

The dimension of the integrals involved in the expression of $f(y_{ijt}, y_{ij't'}; \theta)$ is four and if $j = j'$ or $t = t'$ it reduces to three. Thus, they can be easily approximated using the Gauss Hermite (GH) quadrature method. As close pairs are more informative, we use a PL likelihood constructed from marginal probabilities of observed pairs less distant than $d \geq 0$ time points. This produces a weighted log PL likelihood of order d defined as

$$pl^{(d)}(\theta;\mathbf{y}) = \sum_{i} \sum_{j,j',t,t'} \log f(y_{ijt}, y_{ij't'}; \theta)\, I_{[0,d]}(t'-t). \tag{3}$$

$I_{[0,d]}$ is the indicator function, equal to 1 if $(t'-t) \in [0,d]$ and 0 otherwise.
The DW method is based on the following representation of the marginal density function

$$\begin{aligned} f(\mathbf{y};\theta) &= |\mathbf{C}_{mo}| \int_{R^q} \frac{\prod_{j=1}^{p} g(y_j|\mathbf{C}_{mo}\mathbf{b}^* + \mathbf{b}_{mo}) h(\mathbf{C}_{mo}\mathbf{b}^* + \mathbf{b}_{mo})}{\phi(\mathbf{b}^*;\mathbf{0},\mathbf{I})} \phi(\mathbf{b}^*;\mathbf{0},\mathbf{I}) d\mathbf{b}^* = \\ &= |\mathbf{C}_{mo}| \int_{R^q} m(\mathbf{b}^*)\phi(\mathbf{b}^*;\mathbf{0},\mathbf{I}) d\mathbf{b}^* = |\mathbf{C}_{mo}| E_{\phi}[m(\mathbf{b}^*)] \end{aligned} \tag{4}$$

where $\mathbf{b} = (\mathbf{z},\mathbf{u})$, $\Sigma_{mo} = \mathbf{C}_{mo}\mathbf{C}'_{mo}$ and $\phi(\cdot)$ is the normal density function.
DW consists in approximating the function $m(\mathbf{b}^*)$ as follows (Bianconcini *et al.* , 2017)

$$\hat{m}(\mathbf{b}^*) = \sum_{l=0}^{s} (-1)^l \left(\begin{array}{c} q-s+l-1 \\ l \end{array} \right) m_{s-l}(\mathbf{b}^*) = \sum_{l=0}^{s} A_l m_{s-l}(\mathbf{b}^*) \tag{5}$$

where $m_{s-l}(\mathbf{b}^*) = m(0,\cdots,b^*_{k_1},0\cdots,b^*_{k_{s-l}},\cdots,0)$ and $A_l = (-1)^l \left(\begin{array}{c} q-s+l-1 \\ l \end{array} \right)$.
Replacing (5) in (4) we obtain the approximate density function

$$f_a(\mathbf{y};\theta) = f_L + |\mathbf{C}_{mo}| \left[\sum_{l=0}^{s-1} A_l \int_{R^{s-l}} \sum_{k_1<\ldots<k_{s-l}} m_{s-l}(\mathbf{b}^*)\phi(b^*_{k_1})\cdots\phi(b^*_{k_{s-i}}) db^*_{k_1}..db^*_{k_{s-l}} \right].(6)$$

where f_L denotes the classical Laplace approximation of the integral when $s = 0$. The dimension of the integrals in expression (6) depends on the choice of s. With low values of s, the integrals can be easily approximated using the GH quadrature. In the extreme cases of $s = 0$ and $s = q$, we obtain the classical Laplace and the adaptive GH quadrature method respectively.

3 Simulation study: preliminary results

We perform a simulation study with $p = 3$, $T = 6$, $n = 200$. We consider the UnWeighted (UW) PL function where all the pairs are involved and the PL of order $d = 1,2,3$. As for DW, we set $s = 0,1,2$. For both methods, the number of quadrature points of GH is fixed at 8. 500 replications are generated for each condition of the study. From the results (Table 1) it is evident that DW with $s = 2$ shows the best performance for almost all the parameter estimates. As for PL, in this design, we don't observe relevant differences for different d and UW. We will further explore the effect of T on the PL method by increasing it.

Table 1. *Estimated bias and rmse (in brackets), $p = 3$ and $T = 6$, $n = 200$.*

True	PL				DW		
	UW	$d = 1$	$d = 2$	$d = 3$	$s = 0$	$s = 1$	$s = 2$
$\alpha_1 = 1.00$							
$\alpha_2 = 0.96$	−0.11(0.55)	0.08(0.39)	0.16(0.58)	0.13(0.49)	−0.21(0.24)	−0.12(0.22)	−0.02(0.18)
$\alpha_3 = 1.07$	−0.02(0.33)	0.05(0.40)	0.14(0.55)	0.09(0.44)	−0.27(0.30)	−0.24(0.29)	−0.09(0.21)
$\phi = 0.50$	0.01(0.11)	−0.02(0.11)	−0.02(0.11)	−0.02(0.10)	0.07(0.10)	0.01(0.09)	−0.02(0.09)
$\sigma_1^2 = 2$	−0.19(1.22)	0.26(1.24)	0.21(1.13)	0.17(1.10)	0.23(0.93)	0.46(1.11)	0.29(0.93)
$\sigma_{u1}^2 = 1$	−0.02(0.29)	0.02(0.30)	0.03(0.32)	0.02(0.31)	−0.30(0.42)	−0.26(0.41)	−0.08(0.31)
$\sigma_{u2}^2 = 1$	−0.07(0.38)	0.07(0.39)	0.08(0.40)	0.07(0.42)	−0.20(0.34)	−0.14(0.34)	−0.01(0.33)
$\sigma_{u3}^2 = 2$	0.01(0.63)	0.09(0.74)	0.12(0.78)	0.09(0.71)	−0.40(0.57)	−0.30(0.51)	−0.09(0.47)

References

BARTHOLOMEW, D, KNOTT, M, & MOUSTAKI, I. 2011. *Latent Variable Models and Factor Analysis: A Unified Approach.* Wiley series in Probability and Statistics.

BIANCONCINI, S, CAGNONE, S, & RIZOPOULOS, D. 2017. Approximate likelihood inference in generalized linear latent variable models based on the dimension-wise quadrature. *Electronic Journal of Statistics.*, **11**, 4404–4423.

CAGNONE, S, MOUSTAKI, I, & VASDEKIS, V. 2009. Latent variable models for multivariate longitudinal ordinal responses. *British Journal of Mathematical and Statistical Psychology.*, **62**, 401–415.

LINDSAY, B. 1988. *Statistical inference from stochastic processes.* Providence: Am. Math. Soc. Chap. Composite likelihood methods, pages 221–239.

VARIN, C, & CZADO, C. 2010. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics.*, **11**, 127–138.