Computational Statistics and Data Analysis 57 (2013) 392–403

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Limited information estimation in binary factor analysis: A review and extension

Jianmin Wu [a], Peter M. Bentler [b,*]

[a] Beijing Institute of Technology, China

[b] University of California, Los Angeles, United States

ARTICLE INFO

Article history:
Received 8 October 2011
Received in revised form 27 June 2012
Accepted 27 June 2012
Available online 2 July 2012

Keywords:
Limited information
Laplace approximation
Binary response
Marginal likelihood
Factor scores

ABSTRACT

Based on the Bayes modal estimate of factor scores in binary latent variable models, this paper proposes two new limited information estimators for the factor analysis model with a logistic link function for binary data based on Bernoulli distributions up to the second and the third order with maximum likelihood estimation and Laplace approximations to required integrals. These estimators and two existing limited information weighted least squares estimators are studied empirically. The limited information estimators compare favorably to full information estimators based on marginal maximum likelihood, MCMC, and multinomial distribution with a Laplace approximation methodology. Among the various estimators, Maydeu-Olivares and Joe's (2005) weighted least squares limited information estimators implemented with Laplace approximations for probabilities are shown in a simulation to have the best root mean square errors.

1. Introduction

Factor analysis is a very popular and well-developed model for the analysis of continuous data (see e.g., Yanai and Ichikawa, 2007). While the factor analysis model is equally applicable to binary response data, the development of a reliable statistical machinery for it has resisted consensus. An ideal class of methods involves the use of full information maximum likelihood. However, with the inevitable sparseness of data for models with many variables and dimensions and the difficult integrations involved, limited information methods recently have been revived as a serious contender for routine use.

The development of marginal maximum likelihood (MML) estimation of factor loadings provided an important boost to full information methods (Bock and Aitkin, 1981; Meng and Schilling, 1996). Maximum marginal likelihood estimation is, however, often complicated because the marginal likelihood includes an intractable integral in the estimation of categorical data models in factor analysis. Several approaches have been developed to treat this troublesome integral. Numerical integration is one option, for example, Naylor and Smith (1982) used Gaussian quadrature which leads to the efficient calculation of posterior densities. In numerical analysis, Gauss–Hermite (G–H) quadrature is an extension of the Gaussian quadrature method for approximating the value of integrals. A main problem with G–H quadrature is the inaccurate approximation of integrals. While adaptive G–H quadrature (Pinheiro and Chao, 2006; Schilling and Bock, 2005) can improve accuracy, it also is not optimal when the dimensionality of the integral is high because the number of quadrature points can grow exponentially with the number of latent variables. A second approach involves the use of EM and MCEM algorithms to fit various latent variable models (Meng and van Dyk, 1998, 1997). They have advantages of numerical stability and implementation simplicity, but also have two disadvantages. They are often criticized for slow convergence when the

* Corresponding author. Tel.: +1 310 825 2893; fax: +1 310 206 4315.
E-mail address: bentler@ucla.edu (P.M. Bentler).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.06.022

fraction of missing information is large. Also, the computational efficiencies of MCEM algorithms (An and Bentler, 2012; Chan and Kuk, 1997; Levine and Casella, 2001) rely heavily on the convergence rate of corresponding MCMC samplers. Still another approach involves Laplace approximation, which uses a large sample normal approximation to the posterior distribution and is generally very accurate. Applications of this approach can be found in Breslow and Clayton (1993), Lee and Nelder (1996, 2006), Lin and Breslow (1996) and Huber et al. (2004). Laplace approximation maximum likelihood (LAML) estimation has been widely used for integrals in Bayesian inference (Tierney and Kadane, 1986). The Laplace approximation is also useful for approximating the likelihood in various nonlinear random effects models, when the integrals in the likelihood do not have closed form solutions. Liu and Pierce (1993) showed that Laplace approximation to the marginal likelihood is very accurate. Yun and Lee (2004) implemented the h-likelihood based Laplace approximation for fixed effects, and compared it to G–H for binary outcomes. Raudenbush et al. (2000) applied a Laplace approximation to the integral and maximized the approximate integrated likelihood via Fisher scoring. On the other hand, Joe (2008) reported that the Laplace approximation is biased in generalized linear mixed models with binary response variables. Thus, it is not known whether the Laplace method can be used in binary factor analysis models. This paper reviews LAML in this context (Wu and Bentler, 2012) as well as an alternative method based on the multinomial distribution, and also utilizes MML and MCMC methods to compare to the main methods of interest, limited information methods based on the Bernoulli and multinomial distributions.

Limited information methods were developed to avoid the problems of high dimensional normal integrals, as well as to address the problem that, due to the consequent sparseness of contingency tables, estimation of parameters in a multinomial distribution framework becomes increasingly difficult as the number of multinomial categories increases (e.g., Agresti, 2002, Section 9.8). Although these methods have a long history (e.g., Christoffersson, 1975), in recent years Maydeu-Olivares (Forero and Maydeu-Olivares, 2009; Lee et al., 1995; Maydeu-Olivares, 2001; Maydeu-Olivares and Joe, 2005; Maydeu-Olivares, 2006) has been the main proponent for the use of limited information methods to estimate latent trait models based on univariate and bivariate Bernoulli sample moments with weighted least squares. See also (Bartholomew and Knott, 1999; Christoffersson, 1975). More recently, in a longitudinal modeling framework Fu et al. (2011) have shown that a pairwise likelihood approach with an EM algorithm and quadrature also can perform comparably to full information methods, but they do not compare their method to other limited information methods. As far as we know, outside of their developers, there has been no systematic evaluation of the proposed limited information weighted least squares estimator. We study these estimators in factor analysis models where they have not yet been applied with Laplace approximations to joint probabilities. Furthermore, we develop an alternative limited information estimator based on the first and second order marginal moment maximum likelihood, using marginal moments obtained by mapping between marginal moment and joint probability as approximated by the Laplace method and an extended likelihood or Bayes model approach to factor scores as estimated by Newton–Raphson. We compare this new approach to limited information estimation with that of weighted least squares as well as with full information methods.

In the context of latent trait and factor analysis models, an important issue is testing the adequacy of any proposed model. Thus we also report goodness of model fit using limited information by using a quadratic form statistic that is asymptotically chi-square distributed (Christoffersson, 1975; Maydeu-Olivares and Joe, 2005; Reiser, 1996).

The paper is organized as follows: in Section 2, we introduce a model for factor analysis with binary response data. In Section 3 we discuss factor score and factor loading estimation. In Section 4, we review limited information test statistics of model fit. In Section 5 we provide a comparison of estimates of several methods, and report on simulations that study their performance. Section 6 gives the conclusions.

2. A model for factor analysis

In this section we introduce notation for a factor analysis model with a logistic link function for binary response variables. These observed or manifest variables are denoted by $Y_i = (y_{i1}, y_{i2}, \ldots, y_{ic})^T,\ i = 1, 2, \ldots, N, j = 1, 2, \ldots, c$. y_{ij} is ith independent response from the jth item, and $y_{ij} = \begin{cases} 1 & \text{success} \\ 0 & \text{others} \end{cases}$ (or: yes vs. no, agree vs. disagree, etc.). A latent structure model supposes the binary variables are related to a set of k unobservable variables denoted by $\eta_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{ik})^T$, where η_i denotes a $k \times 1$ vector of factor scores from ith response. For simplicity of argument, as is usually assumed in exploratory factor analysis, the latent variables are assumed to follow a multivariate normal distribution with a null mean vector $E(\eta_i) = 0$ and identity covariance matrix $\mathrm{cov}(\eta_i) = I_{k\times k}$. Letting $\Lambda_j = (\lambda_{1j}, \lambda_{2j}, \ldots, \lambda_{kj})^T$ be a $k \times 1$ vector and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_c)$ be a $1 \times c$ intercept vector, conditioning on the parameters α, Λ and η, the conditional probability of success is given by

$$\pi_{ij} = p(y_{ij}|\alpha_j, \eta_i, \Lambda_j) = \frac{\exp(\alpha_j + \eta_i^T \Lambda_j)}{1 + \exp(\alpha_j + \eta_i^T \Lambda_j)}. \tag{1}$$

3. Estimation of factor scores and factor loadings

In standard applications of MML in item response theory (IRT), latent trait scores do not impact estimates of the item parameters. They are computed once, conditional on estimates of the parameters. In our approach to the factor analysis