Journal of the Korean Statistical Society 45 (2016) 451–465

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Bayesian semiparametric analysis for latent variable models with mixed continuous and ordinal outcomes

Yemao Xia *, Jianwei Gou

Department of Applied Mathematics, Nanjing Forestry University, Jiangsu Province, 210037, PR China

ARTICLE INFO

Article history:
Received 5 May 2014
Accepted 18 January 2016
Available online 11 February 2016

AMS 2000 Subject Classifications:
62H25
62G07

Keywords:
Latent variable model
Finite dimensional truncated stick-breaking prior
Blocked Gibbs sampler
Model comparison

ABSTRACT

Latent variable models with continuous and ordinal responses are a useful tool for interpreting the causal interrelationships among the latent variables and building relations between the latent variables and manifest variables. These models have been successfully applied to many different fields, including behavioral, educational, and social and psychological sciences. However, most developments are constrained within parametric families, of which particular distributions are specified for the parameters of interest. This leads to difficulty in dealing with outliers and/or distribution deviations. In this paper, we propose a Bayesian semiparametric modeling for latent variable model with continuous and ordinal variables. A finite dimensional truncated stick-breaking prior is used to model the distributions of the intercepts and/or covariance structural parameters. Within the Bayesian framework, blocked Gibbs sampler is implemented to deal with the posterior analysis. Moreover, the logarithm of pseudo-marginal likelihood is used to compare the competing models. Empirical results are presented to illustrate the methodology.

© 2016 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

Contents

1. Introduction 452
2. Model description 453
   2.1. Parametric LVMs with mixed continuous and ordinal outcomes 453
   2.2. Bayesian semiparametric hierarchical modeling 454
3. Full conditionals and model selection 455
   3.1. Prior specification 455
   3.2. Model selection 456
4. Empirical illustrations 457
   4.1. A simulation study 457
   4.2. An artificial example 459
5. Concluding remarks 461
   Acknowledgments 462
   Appendix. 462
   References 464

* Correspondence to: Department of applied mathematics, Nanjing Forestry University, Nanjing, PR China.
E-mail address: ym_xia71@163.com (Y.M. Xia).

http://dx.doi.org/10.1016/j.jkss.2016.01.005
1226-3192/© 2016 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

1. Introduction

Latent variable models (LVMs) are a widely appreciated multivariate method for exploring the latent variables that are related to the observed data and assessing the causal relationships among them. Such models, in conjunction with popular computer softwares such as LISREL VIII (Jöreskog & Sörbom, 1996) and EQS 6 (Bentler & Wu, 2002), have been used in a wide range of applications in many disciplines including behaviors, educations, social science, psychology and economics.

Owing to its nature, mixed data structure is very common in behavioral and social science researches. Analyzing such data structure within LVM framework becomes increasing interests and has received a lot of attention, see Dunson (2000, 2003), Muthén (1987), Shi and Lee (1998, 2000) and Song and Lee (2002, 2004). However, most of these contributions are mainly developed within particular parametric distribution families such as the exponential family or normal scale mixture family, which have a limited role in dealing with the distributional deviations, particularly when data take on heterogeneity or multimodality. Though some robust methods are developed to downweight the influence of the outliers (see Kano, 1994; Kano, Berkane, & Bentler, 1993; Lee & Xia, 2006), most of them are still confined to dealing with unimodality and are less effective for solving the asymmetric and/or multimodal problems.

A more flexible approach to address this distribution uncertainty is to build a semiparametric Bayesian analysis procedure. Unlike the common parametric methods which heavily depend on specifications of distribution, semiparametric Bayesian approach assumes that the fitted distribution is random and selected from the space of all possible distribution functions according to some priors. This relaxation gives relatively weak assumptions on sampling model, hence providing a more flexible fitting for data. There is rich literature on this topic (see Müller & Quintana, 2004 for a review). Among various nonparametric priors, Dirichlet process prior may be a most widely used nonparametric model for random distribution in Bayesian statistics, see for example, Ferguson (1973), Antoniak (1974), MacEachern (1994), Bush and MacEachern (1996), Escobar (1994), Escobar and West (1995), Lo (1984), MacEachern and Müller (1998) and among others. Because DP puts probability one on the infinite dimensional discrete distribution, it does not allow for direct inferences on its posterior distribution for more general functionals. Naturally, some finite dimensional approximations are proposed to avoid such difficulties, see for example, Doss (1994), Gelfand and Kottas (2002), Muliere and Tardella (1998), and Sethurman and Tiwari (1982). In particular, Ishwaran and Zarepour (2000) suggested truncating DP at a finite level as a surrogate of DP, which results in finite-dimensional stick breaking prior. This offers practical and computational advantages for modeling data. To carry out Bayesian analysis, Ishwaran and Zarepour (2000) developed a so-called blocked Gibbs sampler instead of the usual Pólya urn sampling scheme in DP mixture modeling. By directly sampling values from the posterior of the random measure, blocked Gibbs sampler is more effective and easier to implement when compared to Pólya urn sampling scheme.

Recently, some authors focused on the semiparametric Bayesian modeling for LVMs. For example, for the multivariate categorical data analysis, Kottas, Müller, and Quintana (2005) extended the traditional multivariate probit model (Albert & Chib, 1993; Chen & Dey, 2000; Cowles, Carlin, & Connett, 1996) to a flexible underlying prior probability model. The usual single multivariate normal model is replaced by a mixture of normal priors for the latent variables; Gill and Casella (2009) suggested using DP as nonparametric prior for random intercepts involved in ordinal Bayesian social science models; And for the latent variable model with fixed covariates and continuous responses, Lee, Lu, and Song (2007) established the semiparametric Bayesian hierarchical model by relaxing the common normality of distribution of exogenous factors to follow a finite dimensional Dirichlet process (Ishwaran & James, 2001; Song, Xia, Pan, & Lee, 2009) further considered a Bayesian semiparametric procedure for the multivariate multinominal data. For more recent developments on semiparametric analysis of LVMs, see Xia and Gou (2015) for a reference.

In this paper, we develop a Bayesian semiparametric procedure for analyzing LVMs with mixed continuous and categorical outcomes. The aims are twofold. First, we extended Kottas, Müller and Quintana's model to a more general model which contains measurement model and latent structural equation. The measurement model interprets the relationships between measurement variables and latent variables, and the structural equation is used to explore the interrelation among the latent variables. The second aim is to establish a Bayesian semiparametric procedure for analyzing mixed structure data. A truncated stick-breaking DP prior is introduced as the prior of the distribution of the mean vector and/or variance–covariance parameters. By virtue of the discreteness of truncated DP, we can assess the heterogeneity among the intercepts and/or variance parameters. This will facilitate the interpretation of heterogeneity of the mean and/or covariance structure across the subjects. Moreover, we treat the threshold parameters as unknown and estimate them simultaneously with other model parameters. Hence, it provides a more general and flexible approach to model mixed continuous and ordinal data.

This paper is organized as follows. In Section 2, we present the Bayesian semiparametric model for LVMs with continuous and ordinal outcomes, and discuss the Markov chain Monte Carlo sampling for posterior inferences. In Section 3, we describe prior specification, full conditionals of random quantities. Model selection based on the logarithm of pseudo marginal likelihood is also presented in this section. Section 4 is devoted to the simulation studies and an artificial example to illustrate the performance of the proposal procedure, and Section 5 contains a few final remarks. Some technical details are given in the Appendix.

Download English Version:

https://daneshyari.com/en/article/1144514

Download Persian Version:

https://daneshyari.com/article/1144514

Daneshyari.com