Statistical Methodology 27 (2015) 1–11

Contents lists available at ScienceDirect

Statistical Methodology

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

Dynamic Bayesian analysis of generalized odds ratios assuming multivariate skew-normal distribution for the error terms in the system equation

S.K. Ghoreishi [a,*], M.R. Meshkani [b]

[a] *Department of Statistics, Faculty of Sciences, Qom University, Iran*

[b] *Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Iran*

A R T I C L E   I N F O

Article history:
Received 5 February 2014
Received in revised form
6 May 2015
Accepted 7 May 2015
Available online 19 May 2015

Keywords:
Association models
Dynamic models
Generalized odds ratio
Risk ratio
Multivariate skew-normal

A B S T R A C T

In this paper, we develop a methodology for the dynamic Bayesian analysis of generalized odds ratios in contingency tables. It is a standard practice to assume a normal distribution for the random effects in the dynamic system equations. Nevertheless, the normality assumption may be unrealistic in some applications and hence the validity of inferences can be dubious. Therefore, we assume a multivariate skew-normal distribution for the error terms in the system equation at each step. Moreover, we introduce a moving average approach to elicit the hyperparameters. Both simulated data and real data are analyzed to illustrate the application of this methodology.

1. Introduction

In the analysis of contingency tables, the analyst is always interested in testing some hypotheses about the cell probabilities, or more generally, about some generalized odds ratios (GORs). These GORs are very important measures in ascertaining the effects of variables of interest. They can be considered as a measure of genetic risk effects, an association measure, a measure of visual modeling, a measure

* Corresponding author. Tel.: +98 2532902998; fax: +98 2532854973.
E-mail addresses: atty_ghoreishi@yahoo.com (S.K. Ghoreishi), mrmeshkani@gmail.com (M.R. Meshkani).

http://dx.doi.org/10.1016/j.stamet.2015.05.001
1572-3127/

of sociodemographic determinants of abortion, and even a general hypothesis in relational models. For more details, see [11,12,8].

The analysis of GORs (risk ratios and odds ratios) for the 2×2 contingency tables with or without structural zero is popular in various disciplines such as epidemiological, biological, medical and public health studies. Fortunately, several authors have focused on the subject. Recently, testing equality of risk ratios in multiple 2×2 tables with structural zero was investigated by Tang and Jiang [14] where some confounding factors such as age, gender, severity of disease, and region are accounted for. Moreover, some models have been proposed for time series analysis of odds ratios in 2×2 tables by Ghoreishi and Alijani [7]. Eskandari [5] prepared his Ph.D. dissertation regarding empirical Bayes analysis of series of contingency tables. Eskandari and Meshkani [6] dealt with this problem from a view point of finite Markov chains.

To the best of our knowledge, there is no record of works which explicitly investigate dynamic (Bayesian) analysis of GORs for $I \times J$ contingency tables. So this work can be considered as a pioneer in this area.

Here, we focus on developing a methodology to investigate the dynamic Bayesian analysis of GORs for categorical data. We make extensive use of deterministic approximations to avoid MCMC techniques and its undesirable non-sequential nature. Although the state space of our assumed model is described by partially specified distributions, it allows us to take advantage of all features of the dynamic models such as sequential monitoring, intervention, missing data, etc.

Assuming a normal distribution for error terms in the system equations is routine practice, but it may be unrealistic, particularly, whenever the data exhibit some degree of skewness. Indeed, when the data are not symmetrically distributed, failure to include asymmetric effects would certainly lead to some biased estimations. So, we relax the symmetry assumption by letting the error terms follow a skew-normal distribution which is designed to introduce asymmetry into the model. This provides some degrees of flexibility which can capture a broad range of non-normal behaviors. It is crucial to note that assuming a skew-normal distribution for the error terms is not an unrealistic assumption in the real world. Since this model by means of a weighted sum allows a half normal (such as a random effect, a clustering effect, etc.) variable in the model which disturbs the normal structure of the errors at each time point. Equivalently, this phenomenon occurs when one additionally assumes a linear form of a random factor with a half normal distribution in the structure of the system equation. Moreover, dynamic linear models are extended and generalized to various non-normal problems, giving rise to a class of dynamic generalized linear models. For more details see, [17,13,10]. Thus, in order to incorporate asymmetric properties for characterizing features of real data sets, we assume the error terms in the system equation follow a multivariate skew-normal distribution. In this model, setting the non-centrality parameter equal to zero, would reduce it to a multivariate normal for error terms.

The assumption of (univariate/multivariate) skew-normal distribution for error terms and random effects has become common in applied statistical models and there is a relatively smaller but increasing literature on this subject, see [3,9] among others.

It would be useful to note that the dynamic linear models, dynamic generalized linear models and dynamic Bayesian models have been extended to various normal and non-normal problems, West and Harrison, [16] and da-Silva et al. [4]. In these models, using a one parameter exponential family of distributions, the time series $\{y_t\}$ is often modeled with specifications similar to the well-known class of generalized linear models. However, this work proposes to address the dynamic Bayesian analysis of generalized odds ratios (DBGORs) based on the following assumptions:

(i) *Observations*: the observed frequencies have a multinomial distribution at each time point t.
(ii) *Prior*: at time t, the cell probabilities have a corresponding conjugate Dirichlet prior.
(iii) *System equation*: over time, the error terms in the system equation have a multivariate skew-normal distribution.

Since the class of multivariate skew-normal distributions includes multivariate normal distributions as a special case, (see [2]), our approach is an extension of the Bayes linear approach [17] which substitutes conditional multivariate normal means and variances for corresponding parameters to give a Kalman filter type structure. However, our approach to DBGOR uses appropriate conditional multivariate skew-normal means and variances instead.