Computational Statistics and Data Analysis 68 (2013) 361–374

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Multivariate zero-inflated modeling with latent predictors: Modeling feedback behavior

Jean-Paul Fox *

University of Twente, Faculty of Behavioral Sciences, P.O. Box 217, 7500 AE Enschede, The Netherlands

HIGHLIGHTS

- We propose a multivariate zero-inflated Poisson–Gamma model for counts and times.
- We studied feedback behavior in a computer-based assessment.
- High ability students were more likely to consult feedback.
- The number of feedback pages consulted was negatively related to achievement.
- Fast working students were not likely to consult feedback.

ARTICLE INFO

Article history:
Received 28 November 2012
Received in revised form 1 July 2013
Accepted 1 July 2013
Available online 11 July 2013

Keywords:
Feedback behavior
Latent predictors
Multivariate zero-inflated count data
Item response times
Poisson-log normal model
Poisson–Gamma model

ABSTRACT

In educational studies, the use of computer-based assessments leads to the collection of multiple outcomes to assess student performance. The student-specific outcomes are correlated and often measured in different scales, such as continuous and count outcomes. A multivariate zero-inflated model with random effects is proposed and adapted for the challenging situation where the multiple outcomes are zero-inflated and possibly right truncated. The joint model consists of a Bernoulli component to deal with the problem of extra zeros, and a multivariate truncated component to model correlated mixed response outcomes from the same subject. In a Bayesian modeling approach, MCMC methods are used for parameter estimation. Using a simulation study, it is shown that the within-individual correlation between counts can be accurately estimated together with the other model parameters. The multivariate zero-inflated model is applied to a computer-based feedback study about computer literacy, where first-year bachelor students were given the opportunity to receive additional feedback. The total number of feedback pages visited and the total feedback processing time are modeled using a Poisson and a Gamma distribution, respectively. The joint modeling framework is extended to incorporate explanatory latent variables (student performance and speed of working), to explore individual heterogeneity in feedback behavior in a computer-based assessment.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Computer-based assessment (a.k.a. computer-administered testing), where students are requested to answer items in a computer environment, is receiving increasing attention. Various advantages of computer-based assessment have been exploited such as improved reliability, improved question styles using interactive multimedia technology, and testing on demand, among other things (e.g., van der Linden and Glas, 2010). Recently, the opportunity to provide instant feedback to students has been further developed to improve learning. In different computer-based assessment studies, feedback

* Tel.: +31 53 4893326.
E-mail address: J.P.Fox@utwente.nl.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.07.003

information was integrated to inform students immediately about their deficiencies given the test results. The main object was to investigate the effects of feedback on students' learning outcomes (e.g., Hattie and Timperly, 2007; van der Kleij et al., 2012).

The present study concerns a computer-based formative assessment (CBFA), where first-year bachelor students were given the opportunity to consult feedback after completing an information literacy test. Interest was focused on student-specific propensity to consult feedback and student-specific characteristics of feedback behavior such as the expected number of pages visited and the expected total time to process the feedback information, also referred to as the total reading time. The assessment data provide information about student achievement and working speed.

From a statistical point of view, several important features of the feedback data need to be considered. First, the data are potentially zero-inflated, where a substantial group of students, around 42%, did not consult feedback for different reasons (e.g., time limitations, motivation, not familiar with computer-based assessment). Second, the restricted number of test items leads to right truncated feedback-use counts. Third, the probability of feedback use is most likely correlated with the expected number of consulted feedback pages and the expected time reading feedback. This correlation should be taken into account to avoid biased parameter estimates. Fourth, within-subject correlation needs to be addressed, since the number of pages visited and the total time reading are clustered by subject. Fifth, the total number of feedback pages consulted (count data) and the total feedback reading time (positive continuous data) are measured on different scales. Finally, individual latent predictors, achievement and speed of working, are measured using a response time item response model (RTIRT; Klein Entink et al., 2009; van der Linden, 2007), which requires a proper treatment of the statistical error in the measurements when using them as explanatory variables.

Here, a multivariate zero-inflated model with subject-specific random effects is presented that addresses the specific features of the data. The real-data study about feedback use will be used throughout the paper to illustrate the various modeling issues. The model consists of two components, where one component addresses the excess zeros and the other multivariate component describes the variability of the counts (total visited pages) and the continuous times (total feedback reading time). A truncated Poisson distribution is used to model the counts and a Gamma distribution is used to model the times, where subject-specific random effects are used to model the unobserved between-subject variation. The log of the mean parameters, representing the expected total feedback use and expected total processing time, are considered to be random effects and assumed to be multivariate normally distributed to capture the within-subject correlation. This joint model is referred to as a multivariate zero-inflated Poisson–Gamma model, where the random effect mean parameters are referred to as subject-specific rates.

Several motivations can be given for the joint modeling framework. Inferences can be made for the original mixed outcomes, the joint model implies separate univariate models, marginal inferences can be directly made, and the random effect structure avoids heavy computational problems since it reduces the issue of dimensionality.

In the field of health research, where it is more common to measure mixed outcomes, the joint modeling of mixed outcomes received considerable attention. Many applications consider the modeling of multivariate longitudinal outcomes and a time-to-event outcome (e.g., McCulloch, 2008; Rizopoulos et al., 2010; Yang et al., 2007). Furthermore, in practice, observations might be zero-inflated or censored such that values below or above a limit are not observed (Mihaylova et al., 2011; Rose et al., 2006). For example, in HIV/AIDS studies viral loads below a certain limit cannot be measured and are subject to left censoring (Hu et al., 2011). In the present modeling framework, of typical interest would be, in the case of informative dropouts, the joint modeling of the censored viral load measurements and time-to-event data. When considering healthcare cost data (as the aggregated total cost of different sources) and the number of interventions (i.e. number of times a medical intervention was required during a time period) a joint modeling approach will account for within-subject correlated observations. A zero-inflated multivariate modeling approach was required when interventions were sparse for subjects in the population. In other research fields, zero-inflated multivariate (count) data also arise naturally in different applications. For example, in economics, purchases over time of different households and products are likely to be correlated within households. And purchase observations of products that are rarely used are zero-inflated.

The present approach builds on work on zero-inflated (count) models. This includes the zero-inflated Poisson model, Lambert (1992) and Loeys et al. (2012) for a more general introduction, where recently Bayesian alternatives have been proposed (see Congdon, 2005; Gelman et al., 2004). Wang (2010) also proposed a model to handle zero-inflated count data. Zero-inflated Poisson models with random effects were also considered by Min and Agresti (2005) and Rabe-Hesketh and Skrondal (2007). The excess zeros will be explicitly modeled through a Bernoulli model, where the success probability is defined as feedback use. This will make it possible to investigate the heterogeneity in the feedback use and student characteristics can be used to explain individual differences in the propensity to use feedback.

The multivariate zero-inflated model is further extended by introducing latent student predictors (achievement and speed of working) as covariates. Latent student characteristics are measured using a response time item response model (RTIRT; Klein Entink et al., 2009; van der Linden, 2007) and used as explanatory variables (e.g., Fox and Glas, 2003; Skrondal and Rabe-Hesketh, 2004). A Bayesian estimation approach is adopted, which supports joint estimation of all model parameters using MCMC simulation techniques.

The remainder of this paper is organized as follows. In the next section, the proposed multivariate zero-inflated mixture model is described using the real-data study as an example throughout the paper. Then, the measurement model for the latent predictors is described. An MCMC algorithm is proposed, which represents the sampling steps related to the different

Download English Version:

https://daneshyari.com/en/article/6870789

Download Persian Version:

https://daneshyari.com/article/6870789

Daneshyari.com