Joint modeling of Time to Event Data and Repeated Ordered Longitudinal Scores Subject to Linear Mixed Model

Arindom Chakraborty (speaker) and Kalyan Das

Visva-Bharati University, India and University of Calcutta, India

Abstract. Longitudinal studies often receive joint information on time to some event and serial outcome measures. This demands a joint modeling of outcomes as well as cause specific survival time using the observed covariates. In many occasions, data not only involve repeated measures but are also collected on ordered categorical responses. Further in view of the longitudinal variation on the ordinal outcome measure, it is desirable to account for the dependence between ordered categorical responses and survival time for different causes. Depending upon the circumstantial necessity, a new joint model is developed where the survival model is extended to encompass the ordered categorical outcomes. A flexible Monte Carlo EM based method is proposed that can simultaneously handle the longitudinal ordinal data and also the censored survival data. The model is applied to muscle scores and survival data from trials of a treatment for children suffering from Duchenne Muscular dystrophy. Finally a simulation study is conducted to examine the finite sample properties of the proposed estimators in the joint model.

Does Nonparametric Maximum Likelihood Make a Difference in Estimation?

Mido CHANG

Virginia Tech, USA

Abstract. The goal of the present study is to explore the longitudinal effects of all-day kindergarten on the academic achievement of children after adjusting for two covariates: childrens SES and their age upon entry to schooling. The study used the first four measures of the Early Childhood Longitudinal Study (ECLS) done by the National Center for Educational Statistics (NCES). Longitudinal Hierarchical Linear Modeling is the technique best suited to the present study, since it allows us to look

closely at the educational contexts of children who are nested within two types of kindergarten program (half-day and all-day), and to take into account the dependencies of observations that are measured repeatedly. An important methodological issue has, however, emerged during the analysis of the data. When a conventional Hierarchical Linear Modeling (HLM) is carried out where there is a violated distributional assumption, the standard errors can be biased and the conclusions of the study affected. Nonparametric maximum likelihood (NPML) is a technique that does not require the normality assumption required by HLM. Thus, in this study, both HLM and NPML are employed, so as to determine the effects of non-normality, and the results of the two approaches are compared with respect to sensitivity of the normality assumption. This study did not find significant difference between parameter estimates obtained via HLM and NPML. There was, however, a small difference in the magnitude of the random components. Specifically, NPML gave higher variances of the intercept and the slope at Level 2, and a lower variance at Level 1.

Generalized Quasi-likelihood versus Hierarchical Likelihood Inferences in Generalized Linear Mixed Models for Count Data

M. R. I. Chowdhury (speaker) and Brajendra Sutradhar

Memorial University, Canada

Abstract. Inferences for the regression parameters and the variance of the random effects in the generalized linear mixed models (GLMMs) set up, is an extremely important statistical issue. It is however known that the most widely used penalized quasi-likelihood (PQL) approach may not produce consistent estimates for the parameters, especially when the true variance of the random effects is large. In the context of Poisson mixed models, in this paper, we examine the consistency performances of two other competitive estimation approaches, namely, the hierarchical likelihood (HL) and the generalized quasi-likelihood (GQL) approaches. An extensive simulation study shows that the HL approach, similar to the PQL approach, appears to produce highly biased and hence inconsistent estimates for the regression parameters, especially when the variance of the random effects is large. The biases of the

HL estimates also appears to vary depending on the cluster sizes. As an alternative, the GQL approach appears to produce consistent estimates for all parameters of this models irrespective of the size of the cluster and the magnitude of the variance of the random effects. The GQL and HL estimates are also compared in a real life data analysis.

Key words: GLMMs; Method of moments approach; Penalized quasi-likelihood approach; Hierarchical likelihood approach; Generalized quasi-likelihood approach; Relative Bias; Zero-inflated Poisson mixed model.

Investigating the feasibility of binning for analyzing mixed longitudinal data measured at distinct time points: a simulation study

Joel A. Dubin (speaker) and Xiaoqin Xiong

University of Waterloo, Canada

Abstract. For longitudinal data, typically the response and any time-dependent covariates are measured at the same time points within each individual. This generally makes modeling straightforward, with various approaches such as mixed effects modeling, marginal modeling, among others, to answer questions of interest. However, in the case when the response and time-dependent covariates are measured at distinct time points within individuals, a smoothing step is required in order to model the data. One such approach is through a binning technique as described in Xiong and Dubin (submitted, 2008) and demonstrated on a hemodialysis study data set in which an important lagged association between infection events and an important time-varying protein was uncovered through a zero-inflated extension of a generalized linear mixed effect model on binned data. Through binning, the original unaligned response and time-dependent covariates become aligned in equally spaced bins in time. Here, we will conduct an extensive simulation study in order to determine under which data scenarios will binning be (and not be) successful. In short, factors such as the within-subject variability of the response and time-dependent covariates as well as the distance in time within the bins at which the original response and

time-dependent covariates are collected will influence the success of the method. As with the hemodialysis example, we will focus on data scenario where the original longitudinal response is binary (e.g., presence/absence of a health event) and the time-varying predictor (e.g., a serum protein) is continuous.

Keywords: Binning; Generalized linear mixed effects models; Longitudinal modeling; Simulation; Time-varying covariates.

On Conditional AIC for Random Coefficient Models

Yasunori Fujikoshi (speaker), Tetsuro Sakurai and Yuka Kobayashi

Chuo University, Bunkyo-ku, Tokyo, Japan

Abstract. This paper is concerned with the problem of selecting candidate models in random coefficient models when the research focus is on cluster or individual. Vaida and Blanchahard (2005) proposed the use of conditional AIC (cAIC) in mixed-effects models, not the conventional AIC based on the consideration of marginal likelihood, However, the cAIC was derived under the assumption that the variance of error terms and the scaled covariance matrix of random effects are known. In this paper we derive a cAIC in random coefficient models, without these strong assumptions, but using an unbiased estimator for the variance parameter. Key Words and Phrases: conditional AIC, mixed-effects model, random coefficient model, selection of candidate models.

Efficient and Robust Estimation for Longitudinal Mixed Models for Binary Data

René Holst

University of Southern Denmark, Denmark

Abstract. This paper proposes a longitudinal mixed model for binary data. The model extends the classical Poisson trick, in which a binomial regression is fitted by switching to a Poisson framework. A recent estimating equations method for generalized linear longitudinal mixed models, called GEEP, is used as a vehicle for

fitting the conditional Poisson regressions, given a latent process of serial correlated Tweedie variables. The regression parameters are estimated using a quasi-score method, whereas the dispersion and correlation parameters are estimated by use of bias-corrected Pearson-type estimating equations, using second moments only. Random effects are predicted by BLUPs. The method provides a computationally efficient and robust approach to the estimation of longitudinal clustered binary data and accommodates linear and non-linear models. A simulation study is used for validation and finally the method is applied to some fishing gear selectivity data.

Analysis of Longitudinal Binary Data with Misclassification

Yunqi Ji (speaker) and Zhaozhi Fan

Memorial University, Canada

Abstract. In medical studies, it is almost impossible to obtain completely exact binary response due to imperfect diagnostic procedures. Ignoring misclassification on binary response results in attenuation on estimations of interested parameters. In this paper, we present a new misclassification model linking the true responses and observed responses by an explicit expression. Methods based on generalized quasi-likelihood (GQL), Optimal GQL (OGQL) and maximin likelihood (ML) approaches are developed to obtain consistent estimates of regression coefficients and the dynamic dependence parameter for longitudinal binary response. We also investigate the robustness of the three approaches about the estimated sensitivity and specificity. Asthma data from Harvard Six Cities Study (H6CS) will be analyzed to illustrate the methods.

Keywords: Misclassification; Children Asthma; GQL; EM; dynamic binary response.

Maximum Quasi-likelihood (MQL) versus Generalized Absolute-deviation based Quasi-likelihood (GADQL) Estimation for Median Regression

Vandna Jowaheer (Speaker), Gary Sneddon (Co-speaker) and Brajendra C. Sutradhar

Mauritius University, Mauritius and Memorial University, Canada

Abstract. There exists an indicator function based maximum quasi-likelihood (MQL) approach [Jung (1996, JASA, 251-257)] for the regression analysis in a linear longitudinal set up for asym- metric data. In a recent unpublished work we have, however, shown that this MQL approach, as compared to a generalized absolute-deviation based quasi-likelihood (GADQL) approach [see Morgenthaller (1992, Biometrika, 747-754) for ADQL approach], produces less efficient regression estimates even when the errors are symmetric but negatively correlated. Since the relative performances of these MQL and GADQL approaches are not known for the correlated asymmetric errors, we have further developed the GADQL approach for an exponential model with correlated errors. This we have done following Hasan, Sutradhar and Sneddon (2007, Sankhya, 548-580). In this paper, we examine the relative performance of this GADQL approach for the correlated exponential data as compared to the MQL approach, both for asymptotic and finite small sample cases.

Second-order least squares estimation of mixed effects models

Daniel Li (speaker) and Liqun Wang

University of Manitoba, Canada

Abstract. The main approach to estimation of mixed effects models focuses on the likelihood methods which rely heavily on the normality assumption for random effects and regression random errors. In addition, they often have computational difficulties. We apply the second-order least squares method to estimate GLMM where the distributions of the regression errors are nonparametric while those of random effects are parametric but not necessarily normal. These estimators are based on the first two marginal moments of the response variable. We present simulation studies of finite sample properties of the second-order least squares estimators and compare them with the ML estimators.

Analysis of non linear multinomial time series

JC Loredo-Osti (Speaker) and Brajendra C. Sutradhar

Memorial University, Canada

Abstract. Applications consisting of a multinomial time series with time dependent covariates are very common but modeling this type of multi-dimensional non-stationary data, especially the modeling of their auto-correlation structure is not easy. In this paper, as one of the basic properties of the multinomial time series, we provide an algebra friendly non-stationary correlation structure. As far as the inference is concerned, we develop a maximum likelihood approach for the estimation of the regression effects, dynamic dependence parameter, as well as category effects. These maximum likelihood estimators are known to be asymptotically efficient. However, to understand the small sample performances of these estimators, we carry out a simulation study.

Conditional Weighted Generalized Quasilikelihood Inferences in Incomplete Longitudinal Binary Models

Brajendra C. Sutradhar and Taslim S. Mallick (speaker)

Memorial University, Canada

Abstract. There exist many studies dealing with complete longitudinal binary data. But in practice one may often encounter the problem of missing-ness in such longitudinal binary set-up. The pattern of this missing-ness or incompleteness in the data can be monotonic or arbitrary and it is mostly expected that the data would be (a) Missing Completely at Random (MCAR) or (b) Missing at Random (MAR) (Rubin, 1976). The inference for the data MAR is however complicated. Our objective is to explore this difficult inference issue for the data MAR and develop procedures for the consistent estimation of the parameters of the model. For a widely used inference procedure in this area, we refer to Robins, Rotnitzky and Zhao (1995) [RRZ (1995)]. See also Fitzmaurice et al (1996) and Paik (1997). To be specific, they have used weighted generalized estimating equation (WGEE) approach for the consistent and efficient estimation of the regression parameters of the model. In this paper, we demonstrate that the WGEE approach in fact produces highly biased and hence inconsistent estimate. We propose a new conditional weighted generalized quasi-likelihood (CWGQL) approach to reduce the bias in order to obtain consistent as well as efficient estimates.

On Quasi-likelihood Estimation with Negative Binomial Offspring and Immigration

Brajendra Sutradhar and Alwell Oyet (speaker)

Memorial University, Canada

Abstract: It is well known that as opposed to the sub-critical case, the so-called conditional least squares and the conditional weighted least squares methods may not yield unbiased and hence consistent estimates for the mean parameters of the negative binomial offspring and immigration distributions in the super-critical case. In this paper, we propose a new conditional quasi-likelihood method that provides mean estimates with smaller mean squared errors in the super-critical case as compared to the conditional least and weighted least squares approaches. Further, we simplify the conditional quasi-likelihood estimating equations both for the mean and the variance parameters under a special model with binary offspring distribution appropriate for a controlled population such as Chinese system. It is also demonstrated empirically that a reasonable estimation for the variance or over-dispersion parameter requires that the data be collected over a large period of time.

Bias corrected quasilikelihood inferences in generalized linear longitudinal models for count data with measurement error

R. Prabhakar Rao (speaker) and Brajendra Sutradhar

Sri Sathya Sai University, India and Memorial University, Canada

Abstract. When the covariates in generalized linear models (GLM) for independent data are subject to measurement errors, there exists several bias correction approaches such as M-estimation based bias correction and correction for score functions to obtain consistent estimator for the regression parameter of the model. In practice, when

the observations are collected over time, the application of the existing bias corrected approaches ignoring the correlations of these repeated responses will yield inefficient regression estimates. Moreover, these approaches, especially the best known correction for score functions, are also not easy to extend to the generalized linear longitudinal models (GLLM) to obtain consistent as well as efficient estimators. In this paper, we propose a bias corrected generalized quasi-likelihood (GQL) method for the consistent and efficient estimation of the regression parameters involved in a GLLM for count data whose covariates are subject to normal measurement errors. The proposed bias corrected GQL (BCGQL) estimators for the dependence case accommodates the existing corrected score functions based estimators for the independent data.

GQL Inferences in Linear Mixed Models with Dynamic Mean Structure

Bingrui Sun (Speaker) and Brajendra C. Sutradhar

Memorial University, Canada

Abstract. In some panel data studies for continuous data, the expectation of the response variable of an individual (or individual firm) at a given point of time may depend on the covariate history up to the present time. Also, the response at a given point of time may be influenced by an individual random effect. This type of data are usually analyzed by fitting a linear mixed model with dynamic mean structure. When the distribution of the random effects and error component of the model are not known, the likelihood inferences can not be used any longer. As a possible remedy, there exists some alternatives such as bias corrected least squares dummy variable (BCLSDV) and instrumental variable based generalized method of moments (IVGMM), which however may produce inefficient estimates. In this paper, we develop a new GMM as well as a generalized quasi-likelihood (GQL) estimating approach and demonstrate that they perform well in estimating all parameters of the model, the GQL being in general more efficient than the GMM approach.

GMM and GQL Inferences in Linear Dynamic Models with Time Varying Correlation Structure

Vickneswary Tagore (speaker), Brajendra Sutradhar and R. Prabhakar Rao

Memorial University, Canada and Sri Sathya Sai University, India

Abstract. There is an increasing interest among econometricians and statisticians to study the financial market volatility through time dependent variance models. In this paper, we consider a linear dynamic model with time varying correlations structure to study such economic data. With regard to the estimation of the parameters of the model, we exploit a generalized quasi-likelihood (GQL) approach and demonstrate that this proposed approach produces estimates with smaller asymptotic variances as compared to an existing popular generalized method of moments (GMM) approach.