Program

Seventh Annual Winter Workshop: Longitudinal Data Analysis

Department of Statistics
University of Florida
January 6-8, 2005

Contents

Sponsors

This year's symposium is funded by the National Science Foundation and Info Tech, Inc., along with the Graduate School, the College of Liberal Arts and Sciences and the Department of Statistics of the University of Florida.

Organizing Committee

Alan Agresti
George Casella
Michael Daniels
Cynthia Garvan
André Khuri
Alex Trindade
Linda Young

Invited Speakers

Raymond Carroll, Texas A&M University
Marie Davidian, North Carolina State University
Stephen Fienberg, Carnegie Mellon
Garrett Fitzmaurice, Harvard University
Patrick J. Heagerty, University of Washington
Michael G. Kenward, London School of Hygiene & Tropical Medicine
Bruce G. Lindsay, Pennsylvania State University
Roderick J. Little, University of Michigan
Geert Molenberghs, Limburgs Universitair Centrum, Belgium
Louise Ryan, Harvard University
Oliver Schabenberger, SAS Institute
Edward Vonesh, Baxter Healthcare Corporation

Acknowledgements

The organizers thank the Department of Statistics staff, and especially Carol Rozear, Marilyn Saddler, Robyn Crawford and Tina Greenly for their tremendous efforts in helping to set up this meeting and make it run smoothly.

Conference Schedule

Thursday, January 8, 2004

7:00-10:00 pm Welcome reception at the Keene Faculty Center, Dauer Hall

Friday, January 7, 2005

8:00-8:25am Breakfast (Reitz Union Grand Ballroom)

All Sessions Meet in the Reitz Union Grand Ballroom

8:25-8:30am Welcome by Alan Agresti

8:30-10:30am **Sessioni: SemiparametricMethods**

Chair: Linda Young

Speakers:
Raymond Carroll: *Longitudinal and clustered data and non/semiparametric regression*

Marie Davidian: *Semiparametric estimation of treatment effect in a pretest-posttest study*

10:30-11:00am **Break and Conference photo at JWRU South side**

11:00-l :00pm **Session 2: Methodology for Missing Data**

Chair: Mike Daniels

Speakers:
Garrett Fitzmaurice: *Methods for handling dropouts in longitudinal studies*

Roderick Little: *Robust likelihood-based analysis of multivariate data with missing values.*

1:00-2:30pm **Lunch** (Gator Comer Dining Center)

2:30-4:30pm **Session 3: Inference Functions and Clustering**

Chair: Alex Trindade

Speakers:
Bruce Lindsay: *From estimating functions to inference functions*

Stephen Fienberg: *Applying Bayesian mixed membership models for soft clustering and classification to longitudinal data*

4:30-6:30pm **Poster Session**

Saturday, January 8, 2005

8:00-8:30 Breakfast (Reitz Union Grand Ballroom)

All Sessions Meet in Reitz Union Grand Ballroom

8:30-10:30am **Session 4: Smoothing and Hypothesis Testing in Mixed Models**

Chair: Yongsung Joo

Speakers:
Geert Molenberghs: *The use of score tests for inference on variance components*

Oliver Schabenberger: *Applications of low-rank smoothing in mixed models for longitudinal and spatial data*

10:30-11:00am **Break**

11:00-l :00pm **Session 5: Imputation and Joint Modeling**

Chair: Trevor Park

Speakers:
Michael Kenward: *Multiple imputation for longitudinal/hierarchical data*

Edward Vonesh: *Shared parameter models for the joint analysis of longitudinal data and event times*

1:00-2:3 0pm **Lunch** - free time

2:30-4:30pm **Session 6: ROC Curves and Goodness of Fit**

Chair: Babette Brumback

Speakers:
Patrick Heagerty: *Dynamic criteria and longitudinal accuracy*

Louise Ryan: *Assessing goodness offit for linear models with correlated data*

5:00-8:30pm. **Barbecue** - Alan Agresti's home, 1632 NW 24th Street (map appended)

Invited Talks

Longitudinal and Clustered Data andNon/Semiparametric Regression

Raymond J. Carroll, Texas A&M University

I will review the problem of nonparametric and semiparametric regression for longitudinal/clustered data. I will describe efficient estimation methods for such problems using both splines and kernels. I will describe how methods designed for the Gaussian model can be generalized to any likelihood or criterion function-based method.

Semiparametric Estimation of Treatment Effectin a Pretest-Posttest Study

Marie Davidian, North Carolina State University

The pretest-posttest study is probably the simplest longitudinal data analysis situation there is, and such studies are ubiquitous in research in medicine, public health, and numerous other fields. Subjects are randomized to two treatments ("treatment" and "control," say), response is measured at baseline ("pretest") and at a pre-specified follow-up time ("posttest"), and interest focuses on the effect usually stated as "difference in change of (mean) response from baseline to follow-up between treatment and control." Missing posttest response for some subjects is a routine complication, and disregarding missing cases (often the default approach in practice) can lead to erroneous inference. Despite the popularity of this problem and plentiful literature on the subject, there is stillno consensus on an appropriate strategy for analysis when no data are missing, let alone for taking account of missing follow-up. This is in part due to widespread misconceptions over the need for certain assumptions, such as normality of the pre- and posttest response, to validate certain popular methods.

We take a semiparametric perspective on the pretest-posttest problem, in which limited distributional assumptions on the pre- and posttest response are made, and show how the theory in the landmark paper of Robins, Rotnitzky, and Zhao (1994, JASA) may be used to derive a unified framework for inference both when full data are available and when follow-up response is missing at random. The development provides a demonstration in a familiar context of the application of this theory leading to techniques suitable for practical use. We illustrate the resulting methods and their performance relative to popular approaches by simulation and by application to data from an HIV clinical trial.

This is joint work with Selene Leon and Anastasios Tsiatis.

Applying Bayesian Mixed Membership Models for Soft Clustering

and Classification to Longitudinal Data

Stephen E. Fienberg, Carnegie Mellon University

We present fully Bayesian approach to soft clustering and classification using mixed membership models. Our model structure has assumptions on four levels: population, subject, latent variable, and sampling scheme. Population level assumptions describe the general structure of the population that is common to all subjects. We illustrate the flexibility and utility of the general model through two applications using data from: (i) the National Long Term Care Survey where we explore types of disability; (ii) abstracts and bibliographies from articles published in The Proceedings of the National Academy of Sciences. We suggest adaptations of the basic mixed membership model for these examples that capture aspects of the longitudinal nature of the data.

Methods for Handling Dropouts in Longitudinal Studies

Garrett M. Fitzmaurice, Harvard University

We review the statistical literature on approaches for handling dropouts in longitudinal studies. A variety of ad hoc procedures for handling dropouts are widely used. The rationale for many of these procedures is not well-founded and they can result in biased estimates of effects.
A fundamentally difficult problem arises when the probability of dropout is thought to be related to the specific value that in principle should have been obtained; this is often referred to as informative or non-ignorable dropout. Joint models for the longitudinal outcomes and the dropout times have been proposed to make corrections for non-ignorable dropouts. Two broad classes of joint models are reviewed: selection models and pattern-mixture models. Selection and pattern-mixture modeling approaches each have their own distinct advantages and disadvantages. We describe an alternative approach, based on a marginally-specified pattern-mixture model, that capitalizes on some of the advantageous features of both modeling approaches.

Dynamic Criteria and Longitudinal Accuracy

Patrick J. Heagerty, University of Washington

Longitudinal marker measurements are often used to guide medical decisions such as the timing and choice of intervention. A basic scientific goal is to use the available information to accurately predict those subjects who are likely to experience a key clinical event such as cancer onset or death. With censored survival endpoints prognostic accuracy may be summarized using extensions of the proportion of variation explained by the covariates as commonly used for continuous response models, or using extensions of sensitivity and specificity which are commonly used for binary response models. In this talk we propose new time-dependent accuracy summaries based on time-specific versions of sensitivity and specificity calculated over risk sets. The accuracy summaries are connected to a previously proposed global concordance summary which is a variant of Kendall's tau, yet allow simple extension to longitudinal covariate or marker processes. Standard Cox regression output can be used to obtain estimates of time-dependent sensitivity and specificity, and time-dependent receiver operating characteristic (ROC) curves. Semi-parametri c estimation methods appropriate for both proportional hazards and non-proportional hazards data are introduced and evaluated in simulations. We show that time-specific ROC curves can be naturally averaged to create a single summary ROC curve that displays the prognostic accuracy of a dynamic screening criteria used to signal incident disease events based on available longitudinal marker measurements.

Multiple Imputation for Longitudinal/Hierarchical Data

Michael G. Kenward, London School of Hygiene & Tropical Medicine

Multiple imputation has become a commonly used tool for dealing with missing data in a wide range of settings. Its implementation in highly structured problems such as with unbalanced longitudinal data raises some interesting issues concerning the handling of multivariate multilevel imputation models. These issues have been addressed in a variety of ways. In this talk these approaches are discussed and, in particular, a recently developed implementation using the package MLwiN is described.

From Estimating Functions to Inference Functions

Bruce G. Lindsay, Pennsylvania State University

In the world of statistics there has been some evolution away from likelihood based procedures into methods based on systems of estimating functions designed to achieve various optimality criteria, including semi-parametri c ones. This talk will be about a simple methodology that creates an inference function, something behaving much like a likelihood, for use in estimating function settings. It is hardly new, dating back to Ferguson (1958), but it is quite powerful, as it automatically gives optimal linear combinations of estimating functions as well as ANOVA-like decompositions of variance. Long popular in econometrics, we think that statisticians should consider putting it in their toolbox.

Robust Likelihood-based Analysis of Multivariate Data with Missing Values

Roderick Little and Hyonggin An, University of Michigan

The model-based approach to inference from multivariate data with missing values is reviewed. Regression prediction is most useful when the covariates are predictive of the missing values and the probability of being missing, and in these circumstances predictions are particularly sensitive to model misspecification. The use of penalized splines of the propensity score is proposed to yield robust model-based inference under the missing at random (MAR) assumption, assuming monotone missing data. Simulation comparisons with other methods suggest that the method works well in a wide range of populations, with little loss of efficiency relative to parametric models when the latter are correct. Extensions to more general patterns are outlined.

KEYWORDS: double robustness, incomplete data, penalized splines, regression imputation, weighting

The Use of Score Tests for Inference on VarianceComponents

Geert Molenberghs* and Geert Verbeke†, Limburgs Universitair Centrum, Belgium

Whenever inferences for variance components are required, the choice between one-sided and twosided tests is crucial. This choice is usually driven by whether or not negative variance components are permitted. For two-sided tests, classical inferential procedures can be followed, based on likelihood ratios, score statistics, or Wald statistics. For one-sided tests, however, one-sided test statistics need to be developed, and their null distribution derived. While this has received considerable attention in the context of the likelihood ratio test, there appears to be much confusion about the related problem for the score test. The aim of this paper is to illustrate that classical (two-sided) score test statistics, frequently advocated in practice, cannot be used in this context, but that well-chosen one-sided counterparts could be used instead. A tight relationship with likelihood ratio tests will be established, and all results are illustrated in an analysis of continuous longitudinal data using linear mixed models.

Some References

Cox, D.R. (1983) Some remarks on overdispersion. *Biometrika* 70, 269-274.

Dean, C. (1992) Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association,* 87, 451-457.

Dean, C.B., Ugarte, M.D., and Militino, A.F. (2001) Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics,* 57, 192-202.

Gueorguieva, R. (2001) A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling,* 1, 177-193.

Gray, R.J. (1995) Tests for variation over groups in survival data. *Journal of the American Statistical Association,* 90, 198-203.

Hines, R.J.O. (1997) A comparison of tests for overdispersion in generalized linear models. *Journal of Statistical Computation and Simulation,* 58, 323-342.

*Biostatistics, Center for Statistics Limburgs Universitair Centrum, Universitaire Campus, Building D, B-3590 Diepenbeek, Belgium, tel: +32/11/26 8238, fax: +32/11/26 8299, email: geert.molenberghsfiîluc˛ac.be

■fBiostatistical Centre, Katholieke Universiteit Leuven, U.Z. Sint-Rafa'el,˙Kapucijnenvoer 35, B-3000 Leuven, Belgium, tel: +32/16/33 6891, fax: +32/16/33 6900, email: geert.verbeke@med.kuleuven.ac.be

Assessing Goodness of Fit for Linear Models with Correlated Data

Louise Ryan, Harvard School of Public Health & Dana-Farber Cancer Institute

We present graphical and test-based methods for assessing goodness of fit for linear models with correlated data. The methods derive from the use of a Cholesky decomposition to transform model residuals to approximate independence. The methods generalize results from Lange and Ryan (1989) and can be tailored to be sensitive to specific types of model departures (e.g. misspecification of random effects distributions). The methods are illustrated with several Datasets.

This work is join with Andres Houseman and Brent Coull.

Applications of Low-Rank Smoothing in Mixed Models for Longitudinal and Spatial Data

Oliver Schabenberger, SAS Institute Inc.

Because of a simple mathematical connection between spline models and mixed models, mixed model software can be used to fit splines. One appealing side-effect of this approach—when the spline coefficients are random—is that the degree of smoothing is directly related to a variance component; making likelihood-based estimates available. **A** natural step is to go beyond scatter-plot smoothing and to consider smooth components in the mixed model framework in much the same way as random intercepts and slopes are used to explain subject-specific (conditional) trends. For example, in a longitudinal study smooth trends over time can be combined with random intercepts among patient cohorts and fixed treatment effects.

This talk presents applications using low-rank radial smoothers and their comparisons to marginal models. The problem of knot selection and placement is approached with efficient space partitioning algorithms. Other practical aspects and implementation are discussed.

Shared Parameter Models for the Joint Analysis of Longitudinal Data and Event Times

Edward F. Vonesh, Baxter Healthcare Corporation

Longitudinal studies often gather joint information on time to some event (survival analysis, time to dropout) and serial outcome measures (repeated measures, growth curves). Depending on the purpose of the study, one may wish to estimate and compare serial trends over time while accounting for possibly non-ignorable missing data or one may wish to investigate any associations that may exist between the event time of interest and various longitudinal trends. In this paper, we consider a class of random-effects models known as shared parameter models that are particularly useful for jointly analyzing such data; namely repeated measurements and event time data. Specific attention will be given to the longitudinal setting where the primary goal is to estimate and compare serial trends over time while adjusting for possible informative censoring due to patient dropout. Parametric and semi-parametric survival models for event times together with generalized linear or nonlinear mixed-effects models for repeated measurements are proposed for jointly modeling serial outcome measures and event times. Methods of estimation are based on a generalized nonlinear mixed-effects model that may be easily implemented using existing software. This approach allows for flexible modeling of both the distribution of event times and of the relationship of the longitudinal response variable to the event time of interest. The model and methods are illustrated with several examples including data from a multi-center study of the effects of diet and blood pressure control on progression of renal disease, the Modification of Diet in Renal Disease Study.

Poster Abstracts

Marginal Models for Bivariate Markov Chains and Granger Non causality

Roberto Colombi & Sabrina Giordano, University of Calabri

In this work we model categorical time series as bivariate Markov chains with transition probabilities that rely on some time-dependent covariates. We are interested in adopting the Marginal approach, introduced in litterature to model joint probabilities in contingency tables, to determine the joint transition matrix of bivariate Markov chains. We stress the use of Marginal models in this dynamic context. Moreover, we deal with Granger noncausality between the two processes jointly composing a bivariate Markov chain. Roughly speaking, a process is said to not Granger cause another process if the past of the former can not be used to better predict present or future values of the latter. The main issue in testing for noncausality is that the definition of Granger involves an infinite number of random variables. Modelling bivariate Markov chains, we propose a simple condition on a finite set of variables that implies the general definition of Granger noncausality on an infinite temporal domain. Besides, we show that for Marginal models only, to check the proposed condition simply reduces to test if some coefficients are not different from zero. Therefore the original problem of testing for causality becomes a simple problem of hypothesis testing on linear constraints on parameters. Furthermore, maximum likelihood method for dependent observations is presented referring to martingale theory in order to obtain some asymptotic results. Finally, Marginal models are applied to sales demand sequences of five soft-drinks to test the proposed methodology in a company production planning and inventory control. Both MLE and hypothesis testing procedures has been developed with the software GAUSS.

Estimation in Generalized Linear Models with Heterogeneous Random Effects

Woncheol Jang, Duke University

The penalized quasi-likelihood (PQL) approach is the most common estimation procedure for the generalized linear mixed model (GLMM). However, it has been noticed that the PQL tends to underestimate variance components as well as regression coefficients in the previous literature. In this paper, we numerically show that the biases of the variance components are systematically related to the biases of the regression coefficient estimates, and also show that the biases of the variance component estimates of the PQL increase as random effects become more heterogeneous.

This is joint work with Johan Lim

Tests for Multivariate Binary Data

Bernhard Klingenberg, Williams College

This poster presents global tests of equality between two vectors of binomial probabilities, based on data from two or several independent or dependent multivariate binary samples. Such data commonly arise in safety studies for newly developed drugs as part of longitudinal clinical trials. In such studies, the occurrence of several, possibly correlated, side effects are compared under a placebo and a drug treatment. Equality between the treatments is defined as either simultaneous homogeneity in the marginal distributions or as identity of the joint distributions. Following a multivariate approach to inference rather than combining results from univariate tests, we construct binary data analogs of the well known Hotelling T-squared test for multivariate
normal data and extensions of McNemar's test. Likelihood ratio tests based on fitting marginal

models for binary data can be computationally intensive for even a moderate number of side effects. Instead, we focus on simpler quadratic form statistics that reduce to well known Wald and score type tests in the univariate case. For either type of statistic, asymptotic inference is often inappropriate due to sparseness, and we also present exact permutation tests. The methods are illustrated with safety data from Phase II clinical trials.

Rank-Based Techniques for Two-Sample Repeated Measures

John D. Kloke and Joseph W. McKean, Western Michigan University

We explore some rank-based techniques for analyzing longitudinal data. In this presentation we focus on a two-sample repeated measures problem. In this design the subjects are randomized to one of two treatments, then measurements on the response are taken over time. Though main effects are the primary interest, in the presence of interaction the difference between treatments can be difficult to interpret. Hence, we utilize a two-stage analysis. Emphasis will be placed on a rank-based Wilcoxon analysis. Examples and simulation results will be provided.

Classifying Disease Progression byLongitudinal Biomarkers in the Absence of aGold Standard

Mareike Kohlmann and Leonhard Held, Ludwig-Maximilians-Universität München Munich
Veit P. Grunert, Roche Diagnostics GmbH, Penzberg, Germany

Correct, early and fast available classification of disease progression can be challenging when we are faced with complex diseases like rheumatoid arthritis. Longitudinal data analysis can be applied to find a prognostic biomarker panel for classifying disease progression in the absence of a gold standard. To overcome the difficulty of a missing gold standard, reference disease progression groups have to be extracted from longitudinal radiographic scores by heterogeneity mixed models. The heterogeneity in the progression groups is captured by a mixture normal distribution for the random effects. Subsequent application of longitudinal discriminant analysis reveals the potentials of the biomarker panel to classify disease progression. This is achieved by
modelling the trajectories of the markers with covariance pattern models to account for the time dependence. Afterwards, we use the resulting mean and covariance estimates as input into a classical discriminant analysis to select the best subset of biomarkers. Results are evaluated by ROC curves varying with the number of biomarkers and the number of time points needed for reliable classification. The usefulness of the models are demonstrated on real life data.

References

Verbeke, G. and Lesaffre, E. (1996): A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association, 91, 217-221.*

Marshall, G. and Barón, A. E. (2000): Linear discriminant models for unbalanced longitudinal data. *Statistics in Medicine, 19, 1969 —1981.*

Zolg, J. W. and Langen, H. (2004): How industry is approaching the search for new diagnostic markers and biomarkers. *Molecular and Cellular Proteomics, 3, 345 -354.*

Keywords: Heterogeneity Mixed Model, Longitudinal Discriminant Analysis, Biomarker Selection, Roc Curve

Semiparamatric Analysis of Longitudinal Data with Informative Observation Times

Do-Hwan Park, University of Missouri-Columbia

Statistical analysis of longitudinal data has been discussed by many authors and a number of methods have been proposed. Most of the researches focus on the situations where observation times are independent of or carry no information about the response variable, therefore, rely on conditional inference procedures given the observation times. This poster session considers a different situation where the independence assumption may not hold. It means the observation times may carry information about the response variable. We proposed a conditional model to integrate observation times and response variable for such cases. Andersen and Gill's multiplicative intensity model is used for modelling the observation process and a semiparametric model relates the longitudinal processes with the time. For inference, estimating equation approaches are proposed and both large and finite sample properties of the proposed method are established. We perform extensive simulation studies that the proposed inference procedures are appropriate for the practical use. The methodology has been applied to a bladder cancer study that motivated this investigation.

Analysis of Survival and Longitudinal Genearray Data

Natasa Rajicic, Dianne Finkelstein, David Schoenfeld,
Harvard School of Public Health

We extend the approach of semiparametric parameter estimation from a joint longitudinal failure time model as developed by Tsiatis & Davi dian [2], to construct a score test based on estimating equations. We relate the longitudinally collected covariate data to the event process by following the specifications given in [2]. No assumption about the distribution of the random effects is made. Rather than being interested in parameter estimation as in [2], we use the derived estimating equations to develop a score test for the association between the event process and the longitudinal covariate. Furthermore, we consider the situation where many different, correlated longitudinal covariates exist in the data (e.g. gene expression arrays collected over time). Following the development of the score test for each individual covariate separately, we address the issue of multiple testing by using a multiple testing procedure based on permutations. We also address the missing data problem that occurs as a result of using permutations on possibly censored, longitudinal data. Our proposed method is illustrated on a genomic dataset from a multi-center research study of Inflammation and the Host Response to Injury that aims to uncover the biological reasons why patients can have dramatically different outcomes after suffering a traumatic injury [1].

References:

1. Inflammation and Host Response to Injury: a multi-disciplinary research project. National Institute of General Medical Sciences (NIGMS) http://www.gluegrant.org, last accessed on November 1, 2004.
2. Tsiatis A, Davi dian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika, 88(2):447-458, 2001.

Evaluation of cognition Effects on Human-Body Movements in Mobile Computing using Multi-Level Mixed Model

Bin Shi, Georgia Institute of Technology

Quantitative evaluations of cognition-context effects on the measurement of human body movements (in terms of accelerations) are studied using the multi-level mixed-effects models. Orthogonal polynomial model is chosen among other mixed-effects model. The evaluations are fundamental for building a reliable algorithm to recognize the cognitive activities during the use of mobile devices, especially in wireless environments. Different summaries derived from the 3-D acceleration time series data, including mean, standard deviation, entropy, autocorrelations, Q-statistics, Engle statistics, are compared in this study. The results show that subtle changes in the difficulty of reading comprehension tasks can induce statistically significant changes of acceleration data in several summaries among these. The findings will contribute to laying a foundation for utilizing the accelerometer as a channel to observe cognitive activities in mobile computing.

Random Effects Logistic Models for Analyzing Efficacy of a Longitudinal Randomized Treatment with Non-adherence

Dylan Small, Wharton School, University of Pennsylvania

We present a random effects logistic approach for estimating the efficacy of treatment for compilers in a randomized trial with treatment non-adherence and longitudinal binary outcomes. We use our approach to analyze a primary care depression intervention trial. Our use of a random effects model to estimate efficacy supplements intent-to-treat longitudinal analyses based on random effects logistic models that are commonly used in primary care depression research. Our estimation approach is an extension of Nagelkerke (et al.) 2000, *Statistics in Medicine's* instrumental variables approximation for cross-sectional binary outcomes. Our approach is easily implementable with standard random effects logistic regression software. We show through a simulation study that our approach provides reasonably accurate inferences for the setting of the depression trial under model assumptions. We also evaluate the sensitivity of our approach to model assumptions for the depression trial. This is joint work with Tom Ten Have, Marshall Joffe and Jing Cheng.

Semi-nonparametric Mixed Models for Longitudinal Data

Guei-Feng (Cindy) Tsai, Oregon State University

We develop nonparametric marginal and random effects models for the analysis of longitudinal data. In particular, we consider a time-varying coefficient model because it provides a flexible statistical model and takes both time and covariates effects into account for longitudinal data. In order to incorporate correlation of longitudinal measurements, the quadratic inference function method is applied. We perform a goodness-of-flt test for testing whether the coefficients are time varying in marginal models method. Furthermore, we take subject-specific effects into consideration and develop random effects models in our setting. In addition, we construct additional estimating equations for estimating variance components. We also propose a goodness-of-fit test for testing whether the variance components of random effects are significant. A real data example on cell cycle microarray data and simulations are illustrated using our methods.

*Polynomial Spline Confidence Bands for Regression Curves**

Jing Wang and Lij ian Yang, Michigan State University

Asymptotically exact and conservative confidence bands are obtained for nonparametric regression function, based on piecewise constant and piecewise linear polynomial spline estimation, respectively. Compared to the pointwise nonparametric confidence interval of Huang (2003), the confidence bands are inflated only by a factor of $(\log(n)\}^{1/2}$, similar to the Nadaraya-Watson confidence bands of Härdle (1989), and the local polynomial bands of Xia (1998) and Claeskens and Van Keilegom (2003). Simulation experiments have provided strong evidence that corroborates with the asymptotic theory.

* *AMS 2000 subject classification'.* Primary 62G15; secondary 62G08.
* *Key words and phrases*: Brownian bridge, B spline, knots, nonparametric regression, quantile transformation.

Semiparametric Joint Modeling of Longitudinal Measurements and Time-to-event Data

Wen Ye, Xihong Lin, Jeremy M.G. Taylor, University of Michigan

Longitudinal studies in medical research often generate both censored time-to-event data and repeated measurements on biomarkers. Recently, joint models using both types of data have been developed. Commonly, the longitudinal covariate is modeled by a linear mixed model. However, in some cases, the biomarker's time trajectory is not linear, such as the prostate specific antigen PSA profile after radio-therapy in prostate cancer study. We propose a two-stage regression calibration approach which models the longitudinal biomarker using a semiparametric mixed model, where covariate effects are modeled parametrically and the individual time trajectories are modeled nonparametrically using a population smoothing spline and subject-specific random stochastic processes. Estimates of the biomarker level and change rate at each time-event are then used as time dependent covariates in the second stage survival model. The performance of the approach is illustrated by application to a prostate cancer study.

Statistical Inference for the AFT Model

Yichuan Zhao, Georgia State University

The accelerated failure time model is an alternative regression tool to study the association between

survival time and covariates. Semiparametric inference procedures have been proposed in extensive literatures. In this paper, the rank-based inference procedures are studied with empirical likelihood ratio method. An empirical likelihood confidence region is constructed for the regression parameters. The proposed method is further compared with wald type based method through extensive simulation studies.

This is a joint work with Professor Eugene Huang, Emory University.

Generalized Score Test of Homogeneity For Mixed Effects Models

Hongtu Zhu and Heping Zhang,
Columbia University and Yale University

Many important problems in psychology and biomedical studies require testing for overdis-persion, correlation and heterogeneity in mixed effects and latent variable models, and score tests are particularly useful for this purpose. But, the existing testing procedures depend on restrictive assumptions. In this paper, we propose a class of test statistics based on a general mixed effect model to test the homogeneity hypothesis that all of the variance components are zero. Under some mild conditions, not only do we derive asymptotic distributions of the test statistics, but also propose a resampling procedure for approximating their asymptotic distributions conditional on the observed data. To overcome the technical challenge, we establish an invariance principle for random quadratic forms indexed by a parameter. Simulation studies are conducted to investigate the empirical performance of the test statistics. A real dataset is analyzed to illustrate the application of our theoretical results.

Key words and phrases. Functional central limit theorem, latent variable, random quadratic form, score test, variance component.

Directions to Dr. Agresti's Home

Travel west on NW 16th Avenue

Turn right onto NW 24th Street

Dr. Agresti's house is the second one on the left (two-story split level).