

The Fourth University of Manitoba Statistics Research Workshop

Abstracts of all Talks

Day 1 (Wednesday August 4, 2004):

Location: Theatre E, Pathology Building, 770 Bannatyne Avenue, University of Manitoba, Bannatyne Campus.

9:40 – 10:20 **A mixed model approach for fitting semiparametric regression models with longitudinal data**

C.-L. Huynh,
University of Manitoba,
Department of Psychology.

Smoothing methodology offers a means by which hidden, seasonal and cyclic trends, or non-linear relationships, can be handled without the restrictions of parametric models (e.g., normality, linearity, and homoscedasticity.). With the development of computer software, smoothing has become a widely used tool for data analysis and inference, and its integration into complex models as well as its applications in applied research are growing. Smoothing techniques currently available in SAS are the procedures (PROC) for generalized additive models (GAM), kernel density estimation (KDE), kriging for binary variables (KRIGE2D), local regression (LOESS), and thin plate smoothing splines (TPSLINE). Although useful in many circumstances, these nonparametric methods are subject to some restrictions that may yield unwanted or invalid, results. For example, they do not account for random effects in varying coefficient models, or correlated errors with longitudinal data. Moreover, it may be hard to implement model diagnostics and fit sensitivity with the nonparametric methods. Mixed model representations of penalized splines has been developed by Ruppert, Wand, & Carroll (2003) to deal with these shortcomings.

In this talk, the basic steps to form and implement some simple smoothing mixed models (SMM) are explained. Examples are used to demonstrate the application of SMM for some common seminonparametric regression settings such as partial parametric (PPM), partial linear (PLM), additively separable (ASM) and varying coefficient (VCM) models. The major advantage of this approach is that, the fitting of these seminonparametric models can be achieved using PROC MIXED in SAS (with some minor modifications). Mixed model software also provides automatic smoothing parameter choice via (restricted) maximum likelihood estimation and other minimization methods of variance components (ML, REML, MIVQUE and MINQUE). Mixed model representations of smoothers also allow for straightforward combination of smoothing with other modelling tools such as random effects for longitudinal data. Diagnostics and sensitivity of SMM fit are also discussed.

10:20 – 11:00 **Longitudinal, multi-level and spatial analyses of non-normal health outcomes**

R. Ma,
Canadian Research Institute for Social Policy and
University of New Brunswick,
Department of Mathematics and Statistics.

In large-scale health studies, we often encounter non-normally distributed outcomes which are not only temporally, but also spatially correlated. One such an example is the National Longitudinal Survey of Children and Youth (NLSCY) in Canada where numerous children across Canada were followed every two years since 1994/5. Traditionally, various longitudinal and/or multi-level data analysis techniques are adopted to handle such data. In this talk, we will present a new random effects modeling approach to the analyses of such spatially and temporally correlated health outcomes.

11:20 – 12:00 **Spatial statistics for epidemiological research**

C. Green,
Manitoba Health.

We will describe a number of data preparation and statistical issues which arise when working with spatial data in a Public Health context. Specifically, the presentation will briefly cover the history of the use of spatial analysis in Public Health, provide some critical perspectives on the use of space and place in Public Health surveillance and research, and describe the three stages of spatial- data visualization, data exploration and data modeling. The remainder of the presentation will then focus on issues relating to rate instability resulting from small numbers and inaccuracies in geocoding, and the impact of spatial dependency on regression analysis. Time permitting, a number of open domain software applications (GEODA, Satscan) which can help address these issues will be demonstrated using several public health examples.

2:00 – 2:40

Estimation of nonlinear regression models with Berkson measurement errors

L. Wang,
University of Manitoba,
Department of Statistics.

The problem of measurement errors (or errors-in-variables) arises frequently in regression analysis of real life data. For example, health scientists may be interested in how coronary heart disease is related to systolic blood pressure, or how lung cancer incidence is associated with long-term exposure to certain air pollutants. In such studies, the explanatory variables (long-term systolic blood pressure and exposure) are either too costly or impossible to measure exactly. It is well known that standard statistical inference procedures will lead to biased estimation and inconsistent conclusion, if measurement errors are present but are ignored.

My talk will focus on the Berkson-type measurement errors in the context of nonlinear regression models. For this set-up, most widely applied statistical inference methods, such as regression calibration and simulation-extrapolation (SIMEX), provide only approximate solutions. In my talk, I will propose a minimum distance estimator based on the first two conditional moments of the response variable given the observed predictors. I will also discuss computational issues and propose a simulation-based estimator. Both these estimators are exactly consistent under fairly general conditions. Some numerical examples will be presented.

2:40 – 3:20

On some aspects of Berkson's bioassay problem

B.K. Sinha,
University of Maryland Baltimore County,
Department of Mathematics and statistics.

For a logistic regression dose-response model with the probability of survival $P(S|d, \theta, \beta) = [1 + \exp(\theta + d\beta)]^{-1}$, we address the problem of estimation of the slope parameter β when θ is known. Berkson's minimum chisquare estimate is contrasted with Fisher's maximum likelihood estimate in both small and large samples. The problem brings out an amazing phenomenon, leading to some novel concepts of asymptotic efficiency.

3:40 – 4:20

Note: this talk will take place at Theatre A, Medical Services Building.

Modeling the longitudinal changes in prevalence of respiratory symptoms among Canadian grain elevator workers

P. Pahwa,

University of Saskatchewan,

Department of Epidemiology and Community Health.

It is well known that human respiratory health is affected due to grain dust exposure. Longitudinal changes in the estimates of prevalence of respiratory symptoms in Canadian grain workers were obtained by analyzing data collected from five regions of Canada over 15 years. Data on respiratory symptoms, smoking status, and pulmonary function were obtained approximately every 3 years during an observational interval called “cycle”. The periods of cycles (number of grain workers) were 1978-81 (Cycle I)(n=5702), 1981-84 (Cycle II)(n=5491), 1984-87 (Cycle III)(n=3713), 1987-1990 (Cycle IV)(n=2847), and 1990-1993 (Cycle V)(n=3078) respectively. A logistic regression model for longitudinal correlated data based on generalized estimating equations approach was utilized to investigate the longitudinal changes in the prevalence of respiratory symptoms.

Day 2 (Thursday August 5, 2004):

Location: Room 260, Helen Glass Centre for Nursing, University of Manitoba, Fort Garry Campus.

9:40 – 10:20 **Using genetic admixture to control for population stratification and residual confounding**

T.M. Beasley,
University of Alabama at Birmingham,
Department of Biostatistics, Section on Statistical Genetics.

Recent reviews have noted contradictory findings in the genetics of many health-related outcomes, including obesity (e.g., Rankinen et al., 2002). This contradictory evidence may be due the inclusion of different ethnic groups or genetically admixed samples, which has lead to population stratification. Population stratification can lead to spurious associations at markers that are unlinked to quantitative-trait or disease loci.

Historically, proxy variables such as skin color, race, and ethnic label have been used to make inferences about population structure, even in the absence of corroborative genetic data (Cooper, 1994; Laveist, 1997; Williams, 1997). Many researchers assume that by using self-reported ethnicity as a categorical covariate in a statistical model to predict outcome variables of interest, the model statistically controls for population stratification, and thus, the results for the marker locus under study are valid. Self-identified ethnicity as a proxy variable for confounding genetic factors (population stratification) can yield imprecise estimates of genetic effects and result in residual confounding. Using a measure of admixture can better represent ethnic differences that are due to genetic variability.

Several approaches, which use multilocus data to enable valid case-control tests of association, even in the presence of population stratification, have been proposed. Genomic Control (GC) methods use the independent marker loci to adjust the distribution of a standard test statistic (Devlin & Roeder, 1999), but it may have a conservative bias in that test statistics are shrunk toward the null value. Structured Association (SA) methods infer the details of population structure before conducting tests for association and do not necessarily suffer from a conservative bias (Hoggart et al., 2003).

All of these approaches assume that population structure can be accurately inferred, and thus do not directly account for the unreliability of estimating population structure or admixture. If the subpopulation information is not estimated with a high degree of reliability subsequent tests may be invalid due to residual confounding (Becher, 1992). Issues in developing analytic approaches to estimate admixture in order to control for population stratification include how to: (a) determine a parsimonious set of markers; (b) assess the reliability of the admixture estimates; and (c) account for the unreliability of these estimates in order to control for residual confounding.

10:20 – 11:00

Uncertainty and sensitivity analysis of a model for SARS transmission

J.F. Brewster¹, R.G. McLeod, A.B. Gumel² and D.A. Slonowsky¹
University of Manitoba,
Institute of Industrial Mathematical Sciences and
¹Department of Statistics, ²Department of Mathematics.

In this talk we conduct a statistical study of a deterministic model for the transmission dynamics and control of severe acute respiratory syndrome (SARS). The effect of the model parameters on the dynamics of the disease is analyzed using uncertainty and sensitivity analyses. The response of interest is the control reproduction number, which is an epidemiological threshold governing the persistence or eradication of SARS. The model includes components describing SARS control measures such as the quarantine and isolation of asymptomatic and symptomatic individuals. One feature of our modelling strategy is the incorporation of functions describing the progressive refinement of these SARS control measures over time. Consequently, the model contains both time-varying inputs and outputs. In this setting, established uncertainty and sensitivity analytical techniques are adapted in order to ascertain key parameters driving the epidemiological model. It is envisioned that this approach could assist in the formulation and assessment of appropriate health strategies for SARS and other infectious diseases.

11:20 – 12:00

Step-down tests for multivariate independent group designs

L. Lix,
University of Manitoba,
Department of Community Health Sciences.

A step-down analysis is one approach that researchers may use to probe a statistically significant multivariate test of group means if some a priori ordering of the outcome variables has been established. Under this approach, tests of group differences are conducted using an analysis of covariance (ANCOVA) in which higher-ranked dependent variables serve as covariates for tests on lower-ranked dependent variables. This research will describe the development of a step-down test that is robust to departures from derivational assumptions which underlie conventional multivariate tests. A numeric example will be used to illustrate its application.

2:00 – 2:40

Application of hierarchical models to administrative health data

D. Chateau and O. Ekuma,
University of Manitoba,
Manitoba Centre for Health Policy.

Two applications of hierarchical modeling will be described in this presentation. The first focuses on clustered data where patients are nested within physicians. Attempts to describe the effects of physician characteristics on outcomes have frequently assumed that observations of patients are independent, or have aggregated data at the physician level. Multilevel modeling is an ideal alternative as it accounts for nested observations and simultaneously estimates patient and physician effects on outcomes. The second application investigates time trends and factors that describe changes in health care costs at the end of life. Thus, we analyze person-specific, longitudinal data, where time points are the first level and individuals are the second level.

2:40 – 3:20

Diallel crosses and incomplete block designs

D.K. Ghosh,
Saurashtra University,
Department of Statistics.

The concept of combining ability is becoming important in plant and animal breeding. It is especially useful in connection with testing procedures in which it is desired to study and compare the performance of lines in hybrid combinations. Several authors have carried out the construction and analysis of Complete Diallel Crosses plans. We discuss several more methods for the construction of CDC plans which also covers the methods carried out by other authors. Analysis of such CDC plans are discussed. Further, the conditions for Universal optimal CDC plans are explained. We also find out the efficiency factor of such optimal CDC plans. A relationship between CDC plans and Balanced Incomplete Block Designs, from which CDC plans are generated, is established. Also the robust CDC plans against the loss of some crosses in one block and/or all the crosses in a block is studied. It is found that some of the CDC plans are robust.

3:40 – 4:20

Title TBA
A.B. Gumel,
University of Manitoba,
Department of Mathematics and
Institute of Industrial Mathematical Sciences.

Abstract not yet available.

Day 3 (Friday August 6, 2004):

Location: Room 260, Helen Glass Centre for Nursing, University of Manitoba, Fort Garry Campus.

9:40 – 10:20 **The use of bandit processes in mathematical finance and economics**

X. Wang,
University of Manitoba,
Department of Statistics.

Bandit processes are statistical decision models with the goal of optimal sequential selections from several populations. Some or all of the populations have unknown distributions. The essential feature of bandit processes is the conflict between information gathering (so as to make more informed decisions later) and immediate payoff (so as to balance some objective of overall optimality). The application of bandit processes in mathematical finance and economics will be examined, and the problem of dynamically pricing a product in the face of unknown demand functions will be discussed in detail.

10:20 – 11:00

On valuation of life insurance contracts with equity-linked maturity guarantees

A. Melnikov and V. Skornyakova,
University of Alberta,
Department of Mathematical and Statistical Sciences.

Insurance companies have, in recent decades, become more innovative in their product designs due to increasing pressure from competing financial institutions. The benefits of such products (segregated funds in Canada, variable annuities in USA, equity-linked life insurance in Europe) depend on the investment performance.

The contracts under our consideration have two types of uncertainty: the evolution of a financial market and a survival status of an insured. We consider the contracts with fixed and flexible maturity guarantees. For a fixed guarantee, the payoff is equal to the maximum between a fixed constant and the value of a risky asset at maturity time provided the insured is alive. For a flexible guarantee, the payoff equals the maximum between two risky assets one of which is supposed to be more risky and more profitable playing the role of a potential gain while the other is more reliable and less profitable serving as a flexible guarantee.

In our setting, the evolution of risky assets is described by the Black-Scholes model and a jump-diffusion model. We exploit the method of quantile hedging, which, together with Black-Scholes (fixed guarantee) and Margrabe (flexible guarantee) formulae, creates effective pricing methodology of such products. The further reduction in the price is possible by pooling homogeneous insureds together and diversifying the mortality risk. Finally, we give numerical examples to illustrate how our results could be applied to actuarial practice.

11:20 – 12:00

Inference for random coefficient GARCH models

A. Thavaneswaran,
University of Manitoba,
Department of Statistics.

Both volatility clustering and conditional nonnormality induce the leptokurtosis typically observed in financial data. An ARMA representation is used to derive the kurtosis of the various class of GARCH-models such as power GARCH, non-Gaussian GARCH, non-stationary and random coefficient GARCH and mixed GARCH. Following Thavaneswaran and Heyde (1999), state space formulation is used to study inference for the class of Random Coefficient GARCH models. Hidden Markov GARCH models are also discussed in some detail.

Last updated: July 30, 2004.