

# THE 33rd ANNUAL MEETING OF ALBERTA STATISTICIANS

**Saturday, October 22, 2011**  
**University of Calgary**  
**Mathematical Sciences Building**  
**Room MS 431**

*The 33rd Annual Meeting of Alberta Statisticians is sponsored by the Department of Mathematics and Statistics and the Faculty of Science at the University of Calgary, and also by PIMS (Pacific Institute for the Mathematical Sciences).*

## PROGRAM:

12:30 to 13:30: Reception, Registration, Coffee in MS 461.

13:30 to 14:00: **Optimal Designs in the N of 1 Trials.** Yin Li. University of Alberta.

14:00 to 14:30: **On Estimation and Testing for Jumps in the Hazard Density from Right Censored Prevalent Cohort Survival Data.** Yassir Rabhi. University of Alberta.

14:30 to 15:00: **Large Controlling IER, EER and FDR In Replicated Regular Two-Level Experiments.** Joseph Akinlawon. University of Alberta.

15:00 to 15:40: Coffee break in MS 461

15:40 to 16:10: **On Minimax Bias Estimation of a Location Parameter.** John Collins. University of Calgary.

16:10 to 16:40: **Constrained Estimation Methods with Measurement Error and Missing in Covariate.** Hyang M. Kim. University of Calgary.

16:40 to 17:00: Small break.

17:00 to 17:30: **Molecular classification of acute leukemia: an application of minimum Hellinger distance estimation.** Patrick Chen. University of Calgary.

17:30 to 18:00: **Sample Size Estimation Methods for Nuclear Magnetic Resonance (NMR) Metabolomics Data.** Karen A. Kopciuk. Population Health Research, Alberta Health Services – Cancer Care, Calgary.

19:20: Dinner at Blue House Cafe (details to follow).

## **REGISTRATION:**

The registration fee is likely to be **around** \$20 per person. The registration fee will be waived for graduate students.

## **ABSTRACTS:**

1. Yin Li.

### **Optimal Designs in the N of 1 Trials.**

Abstract:

The N of 1 trial is a randomised experiment in a single patient which is designed to study the response of an individual to the treatments and assist to make personal best clinical decision. To compare two treatments, A and B, a popular design in N of 1 trials is a single sequence which consists of AB pairs and BA pairs. In this study, we consider models with carryover effects. The efficiencies of designs are compared and the optimal N of 1 trials are presented under the models. Different covariance structures of the error terms in the models are also considered.

2. Yassir Rabhi.

### **On Estimation and Testing for Jumps in the Hazard Density from Right Censored Prevalent Cohort Survival Data.**

Abstract:

Incident studies are the gold standard in survival analysis. However, it often happens that logistics or other constraints preclude the possibility of recruiting incident cases. One may then recruit prevalent cases, i.e. those who have already experienced the initiation (onset) of their disease before being recruited. A feasible design in such cases is the so-called cross-sectional with follow-up design. It is well known that prevalent cases recruited through such design tend to have a longer survivor-ship. As such observations are subject to biased sampling. This bias is termed length-bias when the disease under study is stable, i.e. the onset of the disease is generated by a stationary Poisson process. Naive analysis of length-bias samples is misleading. Estimation of the hazard and survival function based on prevalent cohort data has been the subject of an intensive research over the past few years. A question, often of prime interest, when estimating the hazard is if there is any abrupt change in the hazard. Such abrupt change can identify rapid acceleration or deceleration of the risk of failure after some age. To the best of our knowledge, there is no result in the literature for estimating the location and size of a change (jump) in the hazard function based on right-censored data collected on prevalent cases. We fill this gap and provide methodology for estimating both the location and the size of a possible jump. Both large and small sample behaviour of the estimators are studied, respectively, analytically and by the means of simulations. The methodology is then applied to analyze a set of survival data collected as part of the Canadian Study of Health and Aging (CSHA) on patients with dementia.

3. Joseph Akinlawon (joint work with Peingfei LI and Rohana J. Karunamuni).

### **Large Controlling IER, EER and FDR In Replicated Regular Two-Level Experiments.**

Abstract:

Experimental Designs involve laying out of detailed experimental plans in advance of doing experiments for better output or response, adequate data analyzes and reasonable conclusions based on the analyzes. The most important class of these designs is replicated regular two-level factorial designs, where the experiment is replicated more than one time at each setting of input factors. In literature, two important concepts-IER and EER, are taken into consideration in identifying the factors that have significant effects on the location and dispersion modelling of the response. The existing methods give results that are either too liberal or conservative in controlling IER and EER for both location and dispersion modelling of the response. In this work, we propose new methods for identifying significant location and dispersion factors. Monte Carlo studies show that our proposed methods perform extremely well in terms of controlling the IER and EER. We also extend our proposed methods to control the FDR introduced by Benjamini and Hochberg in 1995. The results from the simulation studies also show that the proposed methods control FDR tightly. Two real data sets were used as case study to illustrate the performance of the proposed methods.

4. John Collins.

### **On Minimax Bias Estimation of a Location Parameter.**

Abstract:

We consider the problem of finding the asymptotically minimax bias estimator, among all location-invariant and scale-equivariant estimators, of the centre of symmetry of a symmetric strongly unimodal distribution under arbitrary  $\varepsilon$ -contamination. More specifically, the model is that random samples are observed from a distribution

$$G(x) = (1 - \varepsilon) F((x - \theta) / \sigma) + \varepsilon H(x),$$

where  $F$  is a fixed known distribution function with strongly unimodal density function symmetric about 0;  $\theta$  and  $\sigma$  are unknown location and scale parameters;  $\varepsilon$  is a fixed proportion of contamination; and  $H$  is a completely unknown contaminating distribution. The objective is find an asymptotically (as the sample size  $\rightarrow \infty$ ) minimax estimator of  $\theta$  in the presence of both the unknown scale parameter  $\sigma$  and the unknown contaminating distribution  $H$ . We define two different reasonable location-invariant and scale-equivariant bias functionals  $B_1$  and  $B_2$  for this estimation problem, and then compute lower bounds on minimax bias by explicitly taking into account the structure of the unidentifiability of the location parameter which is induced by the  $\varepsilon$ -contamination in the model. The minimax bias estimators corresponding to  $B_1$  and  $B_2$  are shown to be the median and the midpoint of the shortest  $100\gamma\%$  of the distribution (where  $\gamma$  depends on  $\varepsilon$ ), respectively. These results are extensions of Huber's classical 1964 result that the median has minimax bias among all location-invariant estimators of  $\theta$  in the case that the parameter  $\sigma$  in the above model is taken to be known rather than unknown.

5. Hyang Kim.

**Constrained Estimation Methods with Measurement Error and Missing in Covariate.**

Abstract:

In many studies for exposure-disease association, group-based exposure assessment arises when exposure to each member of a group is assigned on the basis of group mean exposure estimated from a sample of measurements among members of the same group. Very often only a small number of observations are available to estimate the group mean exposure and hence the observed exposure-disease association is expected to be biased. In many cases groups are naturally ordered in a certain way and this ordering information could be incorporated in estimation procedure. We introduce constrained estimation methods for regression models having measurement error and missing in covariate, and then illustrate the methods by analyzing decline in lung function due to exposures to carbon black.

6. Patrick Chen.

**Molecular classification of acute leukemia: an application of minimum Hellinger distance estimation.**

Abstract:

Forthcoming.

7. Dr. Karen A. Kopciuk.

**Sample Size Estimation Methods for Nuclear Magnetic Resonance (NMR) Metabolomics Data.**

Abstract:

Typical projection-based methods used to analyze NMR metabolomics data, such as Partial Least Squares Regression, do not assume an underlying stochastic model for the data. Thus, this popular algorithmic modelling approach does not have statistical inference methods or the corresponding sample size methods. To overcome this limitation, we investigated whether sample size estimation approaches for high dimensional deoxyribonucleic acid (DNA) microarray data were suitable for use with NMR metabolomics data, as the data types share many key properties. Simulation studies compared three popular DNA microarray sample size estimation methods across a number of NMR metabolomics data features. Comparisons of the Variable Importance in Projections (VIP) with the  $p$ -values from test statistics confirmed the same significant metabolites were selected in both modelling approaches. Simulation study results and an application to a pancreatic cancer data set suggested two of the three methods were appropriate to use for sample size estimation of NMR metabolomics data. Future work will investigate these and other sample size estimation methods for use with metabolomics data measured on other high throughput platform.