

Submittee: Joslin Goh

Date Submitted: 2009-12-09 10:45

Title: Joint SFU/UBC Graduate Student Workshop in Statistics

Event Type: Conference-Workshop

Location:

Simon Fraser University

Dates:

November 21, 2009

Topic:

Statistics and Actuarial Science

Organizers:

Goh, Joslin, Department of Statistics and Actuarial Science, Simon Fraser University -- De Souza, Camila, Department of Statistics, University of British Columbia

Speakers:

Dr. Steven Thompson Statistics and Actuarial Science, Simon Fraser University Title: Sampling in research and statistics Abstract: Sampling design refers to how a sample is selected from a population or distribution. Experimental design refers to how units are assigned to treatments, but in fact an experiment also includes a sampling layer. When a study has not been designed we think of the natural design as the inherent procedure by which the data you see have been selected from all the possibilities. It turns out that in general, one can not make valid inferences from data without taking the design into account. Perhaps of more interest still, improving sampling design usually makes more difference to inference effectiveness than does using one inference approach vs another. Further, the optimal sampling strategy is in general an adaptive one. In this talk I will describe some of the types of designs I have worked on and discuss some of my research experiences along the way. -- Matt Pratola Statistics and Actuarial Science, Simon Fraser University Title: An Overview of Computer Model Calibration Experiments with Application to a Space-Weather Model Abstract: Computer models enable scientists to investigate real-world phenomena in a virtual laboratory using computer experiments. Recently, statistical calibration enabled scientists to incorporate field data. However, the practical application is hardly straightforward. For instance, large and non-stationary computer model output is not well addressed, and model identifiability can also be problematic. Putting aside these difficult issues, this talk will serve as an introduction to computer model calibration experiments, and show some preliminary results of the calibration of a space-weather model of the upper atmosphere. -- Lei Hua Statistics, University of British Columbia, Title: Study tail behavior of multivariate copulas via tail orders Abstract: In this talk, I will introduce the background of the research and present some new results we have obtained. For statistical modeling with copulas, properties such as strengths of upper/lower tail dependence and reflection symmetry or direction of reflection asymmetry are important in deciding on appropriate copulas. For example, for the tail asymmetry phenomena of financial markets, copula families with a variety of tail behavior are useful for statistical modeling. Although

the multivariate Gaussian and t copula families have a wide range of dependence, they are not appropriate when there is reflection or tail asymmetry. But copulas can be constructed from other methods to get different joint tail behavior. Then for use of copulas for inference for joint tail probabilities, sensitivity analysis over different families can be performed. Therefore, an important task is to study and construct new copula families that have different tail behavior and asymmetry than multivariate Gaussian and t copulas. Strong tail dependence has been important in applications where copulas are used for inference on tail probabilities, but the tail order we proposed can also cover intermediate tail dependence. For multivariate Archimedean copulas and other copula constructions based on Laplace transforms, the study of tail order is related to the asymptotic behavior of Laplace transforms at 0 and infinity, and to the tails of the density of the mixing random variable. Some further properties of tail orders that are obtained include the tail relationship between a copula function and its density, and a copula and its margins. -- Saman Muthukumarana Statistics and Actuarial Science, Simon Fraser University, Title: A Bayesian Social Relations Model using Dirichlet Process Priors Abstract: This talk investigates the suitability of Dirichlet process priors in the Bayesian analysis of network data. Dirichlet process priors allow the researcher to weaken prior assumptions by going from a parametric to a semiparametric framework. This is important in the analysis of network data where complex nodal relationships rarely allow a researcher the confidence in assigning parametric priors. The Dirichlet process has a secondary benefit due to the fact that its support is restricted to discrete distributions. This provides a clustering mechanism which is often suitable for network data where groups of individuals in a network can be thought of as arising from the same cohort. The model is applied to real data and the fitness of the model is illustrated. -- Dr. Rollin Brant Statistics, University of British Columbia, Title: How to win collaborators and influence scientists Abstract: My talk will focus on establishing and maintaining scientific collaborations. Being an applied statistician entails working in interdisciplinary teams with investigators from widely varying backgrounds. This type of collaboration is often challenging, but it also offers the greatest potential for making significant contributions to science and the public good. I'll describe some of my most rewarding experiences to provide examples of how success as an applied statistician depends on finding "high potential" collaborators and establishing collegial and mutually rewarding relationships. -- Mike Danilov Title: Robustness under independent contamination model Abstract: Consider the problem of estimating mean vector and covariance matrix of multivariate data. Traditional robustness theory, started by Tukey in 1960, works under the assumptions of mixture contamination model where each observation~(i.e. data case) is either all good or all bad. It is crucial that at least half of the cases in this mixture are good ones because otherwise we cannot tell the difference between the good and the bad anymore. With multivariate data arising in practical applications it is common that different variables originate from different sources, are measured by different instruments, entered by different people etc. Thus variables within one data case can be contaminated independently from each other. Fully discarding observations that are only contaminated in, say one variable out of twenty, as a traditional robust method would, can result in unnecessary data loss or even elimination of all available data. In recent years, a new independent contamination model has been capturing attention of the robust statistical community. In this talk I will start with basic concepts of traditional robustness, then introduce the new contamination model and talk about difficulties that come along with it. -- Kelly Burkett Statistics and Actuarial Science, Simon Fraser University, Title: Sampling genetic ancestries conditional on observed genotype data Abstract: For understanding association of genetic variability with disease outcomes, it can be useful to model the latent genetic ancestry giving rise to the sample's genetic variability. Incorporating ancestry into genetic association statistics requires Monte Carlo methods to sample from the ancestry space as this multidimensional space is too large to enumerate. This presentation will first give a brief background to genetics and the problems addressed by statistical genetics. I then describe our implementation of a Markov Chain Monte Carlo sampler for the topology, node times and data at internal nodes of an ancestral tree conditional on the observed sequence data sampled at present. Finally, the sampler will be applied to data from a case-control study to examine whether the sampled ancestries give any insight into this association. -- Corinne Riddell Statistics, University of British Columbia, Title: A Discussion of Latent Class

Models in a Biostatistical Setting Abstract: Latent Class Models can be used to determine the value of an unknown quantity, in which different measures of the quantity exist, but no measure is completely accurate. In other words, it is used when no gold standard measure exists. In this talk we present one such application, in which a latent class model was used to study the p53 gene status of a cohort of women. Here, four methods were used to measure the women's status, but no method is a gold standard. The primary goal of this talk is to introduce students interested in biostatistics to the subject area through the presentation of an interesting example. We will discuss the use of latent class models to investigate the specific research question followed by a discussion of the advantages and potential limitations of the model in general.

Links:

<http://www.sfu.ca/~jtg3/sfuubc2009.html>
