Submittee: Hao (Nelson) Chen
Date Submitted: 2017-03-26 09:32
Title: Spring 2017 UBC/SFU Joint Statistical Seminar
Event Type: Conference-Workshop

Location:
Room 7000 in the Harbour Centre, Vancouver, BC

Dates:
March 25th, 2017

Topic:
Statistical research and case-study

Methodology:
Lectures

Objectives Achieved:
Provide an opportunity for students from both UBC and SFU to share their experience and expertise in Statistics.

Scientific Highlights:
N.A.

Organizers:
Chen, Hao, Statistics, UBC.

Speakers:
(1) Chen, Hao, Statistics, UBC. Sequential Computer Experimental Design for Extreme Quantile Estimation.

Abstract: Computer models are often used to study the reliability of engineering processes. The research interest of this talk is to estimate an extreme quantile of the output of a computer model. We first build a statistical surrogate for the input-output relationship of the numerical model with a modest number of evaluations to obtain initial estimates. We then experiment sequentially, guided by a new criterion, to improve the estimate of the quantile. The newly proposed sequential method is compared with the popular expected improvement (EI) method in a real wood computer model that quantifies the relationship between Modulus of Elasticity (MOE) of joists and the corresponding

deflection under a static load on a floor system. We show that the new sequential criterion leads to a faster convergence, which is the major contribution of the research. In addition, the uncertainty of the deflection distribution propagated from estimating input distributions is well quantified using different methods. It is shown that the effect of properly modelling the input distribution is much larger than we would expect.

(2) Dustin Johnson, Statistics, UBC. Online, Large-scale Modeling of Consumer Brand Perception using Latent Dirichlet Allocation.

Abstract: The presentation will consist of a brief overview of a consulting project recently undertaken, where the main purpose is to identify a level of consumer sentiment of a company's brand image through the use of social media data and product reviews. The subject of sentiment analysis will be discussed on a broad level, then attention will be focussed primarily on the methodology of latent Dirichlet allocation (LDA), including its ability to generate and model documents through approximating the underlying data distributions. Toy examples demonstrating the application of LDA for natural language processing (NLP) will be covered along with complementary R and Python code. The presentation will conclude with discussions on extensions of LDA for more complex, online settings as well as how to handle storage and optimization constraints of a large amount of unstructured social media data.

(3) Harry Joe, Statistics, UBC. Statistical computing, software comparisons, data science and numerical methods.

Abstract: This presentation will discuss statistical computing topics that students might not get exposed to in courses. Statistical research and many jobs in statistics / data science require proficiency in statistical software and computing. Through examples, some topics of computing will be discussed, such as (a) comparisons of speed of programming languages and software, (b) writing code for reproducibility, (c) pseudo-code, (d) profiling code for bottlenecks, and (e) available resources for numerical optimization, differentiation and integration.

(4) Christina Nieuwoudt, Statistics, SFU. SimRVPedigree: An R Package to Simulate Pedigrees Ascertained for Multiple Relatives Affected by a Rare Disease

Abstract: Family-based studies are receiving renewed attention because of their ability to identify genetic susceptibility factors associated with rare diseases. These studies have more power to detect rare variants, require smaller sample sizes, and can more accurately detect sequencing errors than case-control studies. However, garnering enough families for analysis of a rare disease could require years of effort, making these studies difficult to replicate. To address this shortcoming we have created an R package, SimRVPedigree, to randomly simulate families ascertained to contain multiple relatives affected by a rare disease. The package aims to mimic the process of family development, while allowing users to incorporate multiple facets of family ascertainment. We illustrate how approximate Bayesian computation with SimRVPedigree may be used to estimate the relative risk of disease for genetic cases in a sample of ascertained families.

(5) Jacob Mortensen, Statistics, SFU. Urban Heat Risk Mapping Using Multiple Point Patterns in Houston, Texas

Abstract: Extreme heat, or persistently high temperatures in the form of heat waves, adversely impacts human health. To study such effects, risk maps are a common epidemiological tool used to identify regions and populations that are more susceptible to these negative outcomes; however, the negative health effects of high temperatures are manifested differently among different segments of the population. In this paper, we propose a novel, hierarchical marked point process model that merges multiple health outcomes into an overall heat risk map. Specifically, we consider health outcomes of heat stress-related 911 calls and mortalities across the city of Houston, Texas. We show that combining multiple health outcomes leads to a broader understanding of the spatial distribution of heat risk than a single health outcome.

Links:
https://www.stat.ubc.ca/~hao.chen/2017/2017joint.html

Comments / Miscellaneous:
The seminar began at 8:30AM and ended at 2:20PM. It went very well and participants provide very positive feedback.
