Submittee: Audrey Beliveau
Date Submitted: 2012-10-04 11:26
Title: Fall 2012 SFU-UBC Joint Graduate Students Seminar in Statistics
Event Type: Conference-Workshop

Location:

Simon Fraser University, Harbour Center Campus, Room 7000, 515 West Hastings Street, Vancouver

---

Dates:

Saturday, September 29th 2012

Topic:

Various topics in statistics.

Methodology:

Lectures given by three SFU graduate students, three UBC graduate students, one SFU faculty and one UBC faculty.

Objectives Achieved:

Students were able to show their work and practice their presentation skills, as well as to see other students' research. Students also learned about some of our faculty members research. A lot of interaction and networking by students from both SFU and UBC was achieved.

Organizers:

Principal Organizer: Beliveau, Audrey, Dept. of Statistics and Actuarial Science, Simon Fraser University // Secondary Organizers: Karim, Ehsan, Dept. of Statistics, University of British Columbia; Luo, Hao, Dept. of Statistics, University of British Columbia; Zhang, Hongyang, Dept. of Statistics, University of British Columbia; Casquilho, Camila, Dept. of Statistics, University of British Columbia

---

Speakers:

Presenter: Joslin Goh; Affiliation: Dept. of Statistics and Actuarial Science, SFU; Title: Prediction and Calibration Using Outputs From Multiple; Simulators Abstract: Deterministic simulators are widely used to describe physical processes in lieu of physical observations. In some cases, more than one computer simulator can be used to explore the physical system. Through the combination of field observations and simulated outputs, predictive models are developed for the real physical system. The resulting model can be used to perform sensitivity analysis for the system, solve inverse problems and make predictions. The proposed approach is Bayesian and will be illustrated through applications in predictive science at the Centre for Radiative Shock Hydrodynamics at the

University of Michigan. // Presenter: Seong-Hwan Jun; Affiliation: Dept. of Statistics, UBC; Title: Entangled Monte Carlo; Abstract: We propose a novel method for scalable parallelization of SMC algorithms, Entangled Monte Carlo simulation (EMC). EMC avoids the transmission of particles between nodes, and instead reconstructs them from the particle genealogy. In particular, we show that we can reduce the communication to the particle weights for each machine while efficiently maintaining implicit global coherence of the parallel simulation. We explain methods to efficiently maintain a genealogy of particles from which any particle can be reconstructed. We demonstrate using examples from Bayesian phylogenetic that the computational gain from parallelization using EMC significantly outweighs the cost of particle reconstruction. The timing experiments show that reconstruction of particles is indeed much more efficient as compared to transmission of particles. // Presenter: Zheng Sun; Affiliation: Dept. of Statistics and Actuarial Science, SFU; Title: EDF Tests for Ordered Categorical Data; Abstract: In this talk, we consider a general class of EDF (Empirical Distribution Function) tests for ordered categorical data (ordered contingency tables), that is when the cells have a natural ordering, for example, letter grades on exams. Asymptotic distributions are found under the null hypothesis H_0: each row follows the same distribution. Asymptotic distributions under some contiguous alternatives are also found and asymptotic power of these tests can be calculated. A theorem is proved connecting the cases when parameters are known with those when parameters must be estimated. Components of these test statistics are examined and the first 4 components can be interpreted as tests that are aimed at specific alternatives: location, scale, skewness and kurtosis. We compare powers of the EDF tests with many competing tests including tests derived from the Neyman Pearson Lemma. EDF tests compare favourably. A example data set is analyzed. // Presenter : Dr. Ruben Zamar; Affiliation: Dept. of Statistics, UBC; Title: Robustness and Other Things; Abstract: Data quality is typically affected by the presence of outliers and other forms of data contamination. It may also be affected by missing data, data duplication, etc. From a broad perspective I am interested in the study of the detrimental effect of poor data quality on statistical inference, and in developing appropriate alternative methods to address these problems. The purpose of this talk is to give students a broad picture of my research interests and some current research projects. "Other things" in the title refers to other related topics I am interested in, such as cluster analysis, model selection, bootstrap and data mining. // Presenter: Dr. Joan Hu; Affiliation: Dept. of Statistics and Actuarial Science, SFU; Title: Statistical Analysis for Forest Fire Control; Abstract: This talk discusses statistical issues arising from forest fire control. We start with brief background information to motivate the statistical problems. Models and inference procedures are then proposed. A set of Canadian forest fire data is used throughout the talk for illustration. This is an on-going project jointly with W. John Braun. // Presenter: Jabed Hossain Tomal; Dept. of Statistics, UBC; Title: Ensembling Descriptor Sets using Phalanxes of Variables to Rank Activity of Compounds in QSAR Studies; Abstract: In QSAR studies, molecular descriptors are used to model biological activity of compounds. The statistical model aims to rank rare actives early in a list of compounds. The classifier "random forest" has been found highly accurate in QSAR studies. To enhance its performance in terms of predictive ranking, we propose an ensemble method by grouping variables together. The variables in a group (we call phalanx) are good to put together, whereas the variables in different groups (phalanxes) are good to ensemble. Finally, our method aggregates the phalanxes. There exist several molecular descriptor sets in QSAR studies, and a particular set might do well in ranking activity of compounds for some assays, and fail to do well for other assays. We have considered four assays and five descriptor sets for each. We apply the ensemble of phalanxes to each descriptor set and further ensemble across the five descriptor sets we generated. The performance of our ensemble is compared with random forest. Specifically, random forest was applied to each of the five descriptor sets and to the pool of descriptor sets. We found our method superior to any of the random forests using two rigorous evaluation procedures. // Presenter: Shirin Golchi; Affiliation: Dept. of Statistics and Actuarial Science, SFU; Title: Monotone Interpolation: Sampling from a Constrained Gaussian Posterior; Abstract: Gaussian process (GP) models are popular tools for non-parametric modelling and function estimation. They are commonly used in the area of computer experiments where a finite number of function evaluations are available from a simulator and the underlying functin is to be

estimated using a statistical model while interpolating the given points. However, in the case that extra information such as monotonicity of the underlying function is available, it is not straight-forward to incorporate the constraints in a GP model. I will talk about the constrained posterior distribution together with a recipe to sample from it.   // Presenter: Vincenzo Coia; Affiliation: Dept. of Statistics, UBC; Title: A New Sieve Model for Extreme Values; Abstract: Although rare, extreme events leave a lasting impact on our lives and the world in general. It is therefore important to determine the potential magnitude and frequency of such events, especially when these extremes are dangerous. We focus on the case when these extreme values are heavy tailed. Extreme Value Theory provides a theoretical basis for extrapolating and making inference into these heavy tails; however, there is room for improvement in the extrapolation methods. One modification to the heavy tail is to add an upper truncation; we propose a modification which "progressively truncates" the tail with permeable filters like a sieve. The techniques are then applied to the largest Atlantic hurricanes and the largest black sea bass in Buzzard's Bay. We find that, in most cases, the sieve model provides the best fit, followed by the truncated model.

Links:
http://www.sfu.ca/~abelivea/Workshop/Home.html