2016 PIMS-UBC Statistics Constance van Eeden Lecture

19 January, 2016
11:00 am (pre-lecture reception at 10:30 am)
MSL Lecture Theatre 102 (2185 East Mall)
The University of British Columbia

BI-CROSS-VALIDATION FOR FACTOR ANALYSIS

Art B. Owen (Stanford University)

ABSTRACT: Factor analysis is a core technique in applied statistics with implications for biology, education, finance, psychology and engineering. It represents a large matrix of data through a small number k of latent variables or factors. Despite more than 100 years of use, it remains challenging to choose k from the data. Ad hoc and subjective methods are popular, but subject to confirmation bias and they do not scale to automatic uses. There are many recent tools in random matrix theory (RMT) that apply to the factor analysis setting, so long as the noise has constant variance. Real data usually involves heteroscedasticity foiling those techniques. There are also tools in the econometrics literature, but those apply mostly to the strong factor setting unlike RMT which handles weaker factors. The best published method is parallel analysis, but that is only justified by simulations. We propose a bi-cross-validation approach holding out some rows and some columns of the data matrix, predicting the held out data via a factor analysis on the held in data. We also use simulations to justify the method, though our simulations are designed using recent findings from RMT. The new approach outperforms previous methods that we found, as measured by recovery of a true underlying factor matrix.

This is joint work with Jingshu Wang of Stanford University.

BIO: ART OWEN is a professor of statistics at Stanford University. He is best known for developing empirical likelihood and randomized quasi-Monte Carlo. Empirical likelihood is an inferential method that uses a data driven likelihood without requiring the user to specify a parametric family of distributions. It yields very powerful tests and is used in econometrics. Randomized quasi-Monte Carlo sampling, is a quadrature method that can attain nearly $O(n^{-3})$ mean squared errors on smooth enough functions. It is useful in valuation of options and in computer graphics. His present research interests focus on large scale data matrices. Professor Owen's teaching is focused on doctoral applied courses including linear modeling, categorical data, and stochastic simulation (Monte Carlo).