Statistical Sciences
UNIVERSITY OF TORONTO

SEMINAR

October 2, 2014 at 3:30pm

Sidney Smith Hall, Room 1083

Refreshments will be served at 3:15pm

Model Selection in High-Dimensional Misspecified Models

Yang Feng, Columbia University

Model selection is vital to high-dimensional modeling in selecting the best set of covariates among a sequence of candidate models. Most existing work assumes implicitly that the model under study is correctly specified or of fixed dimensions. Both model misspecification and high dimensionality are, however, common in real applications. In this paper, we investigate two classical Bayesian and Kullback-Leibler divergence principles of model selection in the setting of high-dimensional misspecified models. Asymptotic expansions of these model selection principles in high dimensions reveal that the effect of model misspecification is crucial and should be taken into account, leading to the generalized BIC and generalized AIC. With a natural choice of prior probabilities, we suggest the generalized BIC with prior probability ($\mbox{GBIC}_p$) which involves a logarithmic factor of the dimensionality in penalizing model complexity. We further establish the consistency of the covariance contrast matrix estimator in a general setting. Our results and new method are also supported by numerical studies.

Statistical Sciences
UNIVERSITY OF TORONTO