High-dimensional Bayesian model selection: ideas, theory and examples

Wednesday, January 16, 2019
3:30pm - 4:30pm, CHS 33-105
Refreshments served at 3:00 PM in room 51-254 CHS

ABSTRACT:

Bayesian model selection is a general framework for two fundamental tasks in Statistics: choosing the best model among a set under consideration and quantifying the selection uncertainty. Applications include hypothesis testing, variable selection, dimension reduction or mixture models, to name a few examples. We discuss the state of the art, with an emphasis on high-dimensional settings where the number of models may far exceed the sample size. A natural question is whether the solution provided by BMS is good in a frequentist sense, e.g. what are the guarantees that one selects the optimal model and that posterior probabilities adequately portray uncertainty. We overview current results and discuss limitations, particularly that an important consequences of unduly enforcing sparsity or model misspecification is a drop in power to detect signal that can be quite problematic in practice. We discuss results that offer guidance as to how much sparsity does one really need to enforce, and how to address misspecification by deciding how flexible does a model need to be (e.g. include non-normal errors or non-linear terms) on a data-based manner. We illustrate practical implications via a variety of examples, including high-dimensional regression, mixture and factor models.