

Bayesian Post model Selection Inference

Daniel Yekutieli

It is common practice to use the observed data to select the variables that are included in the model that is fitted for the data, and use the data again to derive statistical inferences from the resulting model that are only valid if the model was specified a priori. Berk et al. (2012) propose a novel perspective on the use and interpretation of the selected models. According to their sub-model interpretation of parameters, the subset of selected explanatory variables does not specify a model for the data, but rather determines the sub-model for which the statistical inference is provided. Berk et al. (2012) produce valid “post selection inference” by constructing wider marginal confidence intervals for the coefficients of the selected sub-model that offer simultaneous coverage for all the coefficient in all the sub-models that can be selected.

I will present a Bayesian framework for providing comprehensive post model selection inference (point estimators, credible sets and predictive distributions) that corresponds to the sub-model interpretation of parameters, and apply it to a simplified simulated example and to a small model selection example (cement data set -- Woods et al., 1932). This is joint work with Edward George from the University of Pennsylvania.