The Bayesian paradigm offers a flexible modeling framework for analyzing data with complex structures, but relative to penalization-based methods, little is known about the consistency of Bayesian model selection methods for high dimensional data settings that include datasets from many biological applications including gene expression data. I will present a new framework for understanding Bayesian model selection consistency, using sample size dependent spike and slab priors that help achieve appropriate shrinkage. More specifically, strong selection consistency is established in the sense that the posterior probability of the true model converges to one even when the number of covariates grows nearly exponentially with the sample size. Furthermore, the posterior on the model space is asymptotically similar to the L0 penalized likelihood. I will also introduce a new Gibbs sampling algorithm for posterior computation, which is much more scalable than the standard Gibbs sampler for large datasets, and yet it retains the strong selection consistency property. The new algorithm and the consistency theory work for a variety of problems including linear and logistic regressions, and a more challenging problem of censored quantile regression where a non-convex loss function is involved.

Consistent and Scalable Bayesian Model Selection for High Dimensional Data