

[Marina Vannucci](#)

Bayesian Methods for variable Selection with Applications to High-Dimensional Genomic Data

Description: I will cover Bayesian methods for variable selection that use spiked priors and consider applications to high-dimensional data. Various modeling settings will be considered, starting with the widely used linear regression models. Bayesian models for variable selection have been successfully employed in such settings, making problems with hundreds of regressor variables and a few samples quite feasible. These methods use mixing priors on the regression coefficients to do the selection and fast Markov Chain Monte Carlo approaches to sample from posterior distributions. I will focus on applications to high-dimensional data and give a brief demonstration of a Matlab software I have written. Emphasis will be on data from Genomics studies that use high-throughput expression levels of genes and proteins. I will then address extensions of the variable selection methodologies to other linear settings, in particular to probit models for classification settings and to models for survival data. I will use some more examples from genomic studies and show how models and inferential algorithms can be modified to incorporate additional information, such as data substructure and/or biological knowledge on gene functions. If time permits I will also address variable selection for model-based clustering with finite mixture models and an unknown number of components. Variable selection is achieved via latent binary vectors that identify the discriminating variables and are updated via a Metropolis algorithm. Inference on the cluster structure is obtained via reversible jump MCMC. I will illustrate performances of the methodology on simulated data and on DNA microarray data.

WED: Bayesian methods for variable selection in regression models, with applications.

THU: Extensions to classification models and to survival data. Relaxing prior assumptions to include biological knowledge.

FRI: More on prior models. Extensions to model-based clustering.

Assumed background: Foundations of Bayesian Analysis (not strictly necessary).