

Bayesian multi-SNP analysis methods for association studies

Yongtao Guan¹ and Matthew Stephens^{1,2}

¹*Department of Human Genetics, University of Chicago, USA*

²*Department of Statistics, University of Chicago, USA*

Ongoing large-scale genetic association studies are collecting data on hundreds of thousands of genetic variants in thousands of phenotyped individuals, in an attempt to identify genetic variants that are associated with phenotypes of interest. Standard analyses of such studies involve testing each genetic variant, one at a time, for association with phenotype, and reporting a p value for each such test. Here we will describe Bayesian approaches to this problem, with a particular emphasis on the simultaneous analysis of multiple genetic variants, perhaps even all observed genetic variants. From a general statistical view, this problem can be considered a variable selection problem with the number of variable vastly exceeding the number of sampled individuals ($p \gg n$), and many of the methods we describe are generally applicable to such problems, not only to association studies. Our talk will focus particularly on choice of suitable priors for this kind of variable-selection problem, advantages of multi-variable analyses over standard single-variable analyses, and comparisons of the Bayesian approach with alternative variable-selection methods such as the LASSO.