



THE UNIVERSITY OF
CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

XIANG ZHU

Department of Statistics
The University of Chicago

A Bayesian Large-Scale Multiple Regression Model for
Genome-Wide Association Summary Statistics

FRIDAY, June 30, 2017, at 3:00 PM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Bayesian methods for large-scale multiple regression provide attractive approaches to the analysis of genome-wide association studies (GWAS). For example, they can estimate heritability of complex traits, allowing for both polygenic and sparse models; and by incorporating external genomic information into the priors they can increase statistical power and yield new biological insights. However, these methods require access to individual genotypes and phenotypes, which are often not easily available for large studies.

Here we provide a framework for performing these analyses without individual-level data. Specifically, we introduce a "Regression with Summary Statistics" (RSS) likelihood, which relates the multiple regression coefficients to univariate regression results that are often easily available. The RSS likelihood requires estimates of correlations among covariates (SNPs), which also can be obtained from public databases. We perform Bayesian multiple regression analysis by combining the RSS likelihood with previously-proposed prior distributions, and then using Markov chain Monte Carlo or Variational Bayes algorithms to compute posteriors. In a wide range of simulations RSS performs similarly to analyses using individual-level data, including SNP heritability estimation, genetic association detection and gene set enrichment analysis.

We apply RSS methods to analyze published GWAS summary statistics of 1.1 millions common variants from 31 human phenotypes, 3,913 biological pathways retrieved from nine public databases, and 113 tissue-associated gene sets derived from gene expression profiles of 53 human tissues. We identify many previously-unreported genes, pathways and tissues that show strong evidence for association with complex traits in our large-scale integrated analyses. Software is available at <https://github.com/stephenslab/rss>.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.