

Bayesian and imputation-based analyses for association studies

Matthew Stephens

Presenters

Matthew Stephens, Professor of Statistics and Human Genetics, University of Chicago, USA, mstephens@uchicago.edu

Department of Human Genetics, CLSC 4th Floor, 920 E. 58th Street, CLSC 5th floor Chicago, IL 60637 phone 773.834.0525 fax 773.834.0505.

Biographical data

Matthew Stephens received his PhD in Statistics from the University of Oxford, UK, in 1997. Dr Stephens performed postdoctoral research with Professor Peter Donnelly in Population Genetics, and has since held faculty posts at the University of Washington, and University of Chicago. Dr Stephens research focuses on the development and application of statistical methodology for applications in Biology and Population Genetics. His contributions include methods for inferring population structure, recombination rates, haplotypes and missing genotypes, and identifying polymorphisms from sequence trace data. Current interests include the development of Bayesian and imputation-based approaches to analysis of genome-wide association studies.

Motivation and aims of the tutorial

Several recent papers, both methodological (eg Wakefield, Servin and Stephens, Marchini et al) and applied (WTCCC), have emphasised the potential benefits of using Bayesian methods in the analysis of genetic association studies. However, although Bayesian methods are widely used in statistical applications in biology, and in statistics more generally, many of today's scientists lack formal education in these methods. This creates multiple barriers to the use of Bayesian analyses in practice, and to scientific communication. Those unfamiliar with Bayesian methods will naturally struggle to interpret published Bayesian analyses, and be unlikely to use them themselves in practice. Furthermore, even those familiar and competent with Bayesian methods may elect not to apply them, out of concern that many readers, and referees, will find them difficult to understand.

Our aim in this tutorial will be to help rectify this problem, by providing an introduction to the use of Bayesian methods for the analysis of genome-wide association studies. The tutorial will cover use and interpretation of these methods. We will cover both simple single-SNP analyses, more complex multi-SNP analyses, and some of the most important practical issues that arise in this context (eg multiple comparisons, replication). We will also cover the use of imputation-based methods to test untyped SNPs, and to combine information across studies performed on multiple platforms.

Targeted audience

The review will be aimed at primarily at individuals who are familiar with association studies and standard approaches to statistical hypothesis testing, but less familiar with Bayesian approaches. Familiarity with some basic statistical ideas (e.g. normal distributions, likelihood, p-values) will be assumed, but no exposure to Bayesian statistics will be assumed. We hope the material will be of interest both to those applying methods and developing new methods.

Tutorial outline

- Introduction to Genome-wide association studies (5 minutes)
- Single-SNP tests: the Bayesian approach to Hypothesis testing. Bayes Factors: computation and interpretation. (25 minutes)
- Contrast between Bayesian and Frequentist testing. Limitations of p values. (15 mins)
- Practical issues: Multiple Comparisons, Replication (10 mins)
- Multi-SNP analyses: candidate regions. Bayesian additive models. Comparisons with non-Bayesian approaches. (20 mins)
- Multi-SNP analyses: whole genome. Comparison with LASSO and penalized regression. (15 mins)
- Imputation-based association analysis. Motivation for imputation. Methods for imputating genotypes. Testing imputed variants. (30 mins)