A nonparametric Bayesian model for meta-analysis of genetic association studies

Maria De Iorio[1], Claudio J. Verzilli[2] and John C. Whittaker[2]

[1] Department of Epidemiology and Public Health, Imperial College London
[2] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK

We present a Bayesian nonparametric model for the meta-analysis of candidate gene studies with a binary outcome. Such studies often report results from association tests for different, possibly study- specific and non- overlapping markers (typically SNPs) in the same genetic region. Meta-analyses of the results at each marker in isolation are seldom appropriate as they ignore the correlation that may exist between markers due to linkage disequlibrium (LD) and cannot assess the relative importance of variants at each marker. Also such marker-wise meta-analyses are restricted to only those studies that have typed the marker in question, with a potential loss of power. A better strategy is one which incorporates information about the LD between markers so that any combined estimate of the effect of each variant is corrected for the effect of other variants, as in multiple regression. Here we develop a Bayesian nonparametric model which models the observed genotype group frequencies conditional to the case/control status and uses pairwise LD measurements between markers as prior information to make posterior inference on adjusted effects. The approach allows borrowing strength across studies and across markers. The analysis is based on a mixture of Dirichlet processes model as the underlying nonparametric model. Full posterior inference is performed through Markov chain Monte Carlo algorithms. The approach is demonstrated on simulated and real data.