

JPSM/MPSM

Survey Methodology Seminar Series

Comparison of MCMC with Integrated Nested Laplace Approximation in Estimating County by Demographic Diabetes Prevalence

ROBERT BASKIN

Presenter

Senior Mathematical Statistician at Agency for Healthcare Research and Quality

Monday, November 3, 2014

12:00-1:00pm

Speaker will be at the University of Maryland
Video Room 368 - ISR South Basement
Room 1208 LeFrak Hall– Maryland

Abstract

This is a very nontechnical introduction to an R package, R-INLA, which performs Integrated Nested Laplace Approximation (INLA) for fitting Bayesian models. R-INLA has seen major growth in use in the area of spatial statistics but may not be as well known outside that area. The use of MCMC has driven a major growth in the use of Bayesian models across many statistical fields and MCMC can fit quite general models. However, Laplace approximation, introduced by Tierney and Kadane, JASA, 1986 and expanded on in Rue, Martino and Chopin, JRSSB, 2009 can have certain advantages over MCMC in fitting certain kinds of Bayesian models which fall in the class of Gaussian Markov Random Fields (GMRF). For models fitting within the GMRF framework, INLA can achieve the same accuracy as MCMC with considerable speed improvements. This allows the modeler to more easily investigate model fit through techniques such as group cross validation or leave-one-out cross validation. Another performance advantage is that INLA can fit quite large datasets in a reasonable amount of time.

The R-INLA package will be introduced through a simple disease mapping model of London borough suicides found in Congdon's 'Bayesian Statistical Modelling' (2006) (code is available on the R-INLA website <http://www.r-inla.org/>). Next, will be an introduction to the AHRQ Quality Indicators (<http://www.qualityindicators.ahrq.gov/>) as a motivation for Small Area Estimates of diabetes prevalence at the level of U.S. county by demographic groups. The diabetes models will be based on the Behavioral Risk Factor Surveillance Survey datasets (2008 – 2010). Estimation of some very simple models on this dataset will demonstrate the speed effectiveness of INLA versus MCMC. R-INLA will be used for investigation of more diabetes models through leave-one-out cross validation. The talk will end with an open question related to the AHRQ QIs and SAE.