Bayesian variable selection with a focus on the analysis of genomic data

E. Lesaffre [1,*], and V. Rockova[2]

1 Department of Biostatistics, Erasmus MC, Rotterdam, the Netherlands and
L-Biostat, K.U.Leuven, Leuven Belgium
2 Department of Biostatistics, Erasmus MC, Rotterdam, the Netherlands
Corresponding author: e.lesaffre@erasmusmc.nl

Keywords: Bayesian, Variable selection, Spike and slab, high dimensionality, genomic data

Abstract:

We first give a brief overview of Bayesian variable selection in the context of linear regression, focusing on the main stream ideas and approaches. We would classify them as: (1) search strategies for selecting the model (MC^3), (2) search strategies for selecting the model and estimating the regression coefficients (Spike and slab) and (3) estimating the regression coefficients and thereby inducing a variable selection procedure (Lasso).

In the second part of the talk we focus on the scenario when p > n (p = number of regressors and n = sample size) and when there is a known structure among the covariates. We give an overview of elaborate stochastic search approaches for high-dimensional variable selection. Towards the end of the talk, we discuss recent methodological contributions in Bayesian variable selection that implement deterministic inferential procedures for posterior (model) exploration. We illustrate the usefulness of these approaches on a real genomic dataset.