Data Simulations

We simulated a genomic region with LD structure and MAF mimicking observed genetic variants in the *Kallikrein* (*KLK)* region [1], which spans about 266 Kb on chromosome 19q13. The original data set included 978 SNPs with MAF>2%. We then pruned SNPs based on pairwise Pearson correlation >95%, which resulted in a total of 872 SNPs. The LD blocks were determined with the software *Haploview* [2] with the method "Solid Spine of LD". This internally developed method searches for a "spine" of strong LD running from one marker to another along the legs of the triangle in the LD chart (this would mean that the first and last markers in a block are in strong LD with all intermediate markers but that the intermediate markers are not necessarily in LD with each other. We found a total of 85 LD blocks with this method. The SNP numbers, SNP names and LD blocks are given in the file "*block.info.long.csv*".

A total of 10 simulated files, each containing 2000 cases and 2000 controls, is included as example. The SNP genotypes were sampled from observed genotype data in the *KLK* region and for each individual, the case–control status was generated from a Bernoulli trial with probability p of being a case given by:

$$logit(p) = \alpha + \sum_{p=1,..,7} \beta_p X_p,$$

where we chose α=-2.0 and $\beta_1=\beta_2=\ldots=\beta_7$=1.0. We generated a population of 35,000 individuals according to this model and then extracted 2000 cases and 2000 controls from it. The files for cases are "*gent_case*X.*dat*" and for controls "*gent_cont*X.*dat*" with X=*1,..,10*. The list of true SNPs for the 10 simulation replicates is given in the file "*list.SNPsTrue10rep.csv*".

The Birth-Death MCMC algorithm (BDMCMC)

The BDMCMC method is our implementation of the Birth-Death MCMC algorithm [3] to regression models. A technical report is provided in the document "*Bayesian_regression_BDMCMC.pdf*". An *R* code running the BDMCMC algorithm is given in "*BDmcmcbin.R*". An example on how to run the code on the 10 simulation datasets is given in "*Run_BDMCMC_Github.R*" and how to read the final results and compute the FDR and TDR statistics in "*read_BDMCMC_Github.R*".

References

1. Briollais L, et al. Germline Mutations in the Kallikrein 6 Region and Predisposition for Aggressive Prostate Cancer. J Natl Cancer Inst. 2017 Apr 1;109(4).
2. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21(2):263-265.
3. Dobra, A. and Mohammadi, A. (2018). Loglinear model selection and human mobility. Annals of Applied Statistics 12(2): 815-845.