Model selection approach for genome wide association studies in admixed populations

Piotr Szulc[1]

[1]*Department of Mathematics and Computer Science*
Wroclaw University of Technology, Poland

Abstract

The main purpose of Genome Wide Association Studies (GWAS) is the identification of genes responsible for quantitative traits (Quantitative Trait Loci, QTL) or disease causing genes in human populations. Localization of genes in such outbred populations is relatively difficult. The problem comes from the fact that due to cross-overs (exchange of genetic material among chromatids), which occur during production of reproductive cells, the statistical relation between a QTL genotype and a genotype of the neighboring marker might be very small. Therefore the scientists need to use a huge number of densely spaced markers, which necessitates the application of stringent multiple testing corrections and results in a relatively low power of gene detection.
Modifications of Bayesian Information Criterion, mBIC and mBIC2, were successfully used in GWAS and results can be found *e.g.* in [1]. However, it turns out that we can find much more influential genes if we perform GWAS in *admixed population*, obtained as a result of interbreeding between previously separated ancestral populations. In that case, apart from genotypes, we have information about the origin of genome's fragments.
I will present the problem of localizing genes and show how to use modifications of model selection criteria in admixed population. Finally, I will present results of simulations.

Keywords: linear regression, model selection criteria, GWAS, admixed population
AMS subject classifications: 62J05, 92D20

Bibliography

[1] F. Frommlet, F. Ruhaltinger, P. Twarog and M. Bogdan (2011). A model selection approach to genome wide association studies. *Computational Statistics and Data Analysis* 56, 1038–1051.