Aggregation of dependent statistics in genome-wide association studies

Mathieu Emily[1], Florian Hébert[1], and David Causeur[1]

[1]IRMAR, Agrocampus Ouest, Rennes, France

Case-control genome-wide association consists in testing an association between Y, a binary variable (a case-control phenotype), and a set of categorical explanatory variables, $X_1, \ldots, , X_p$, where X_i is the ith Single Nucleotide Polymorphism along the genome. Associations are usually tested in a pointwise approach where each X_i is tested sequentially. Due to the block structure of the genome, pointwise tests are correlated and a proper handling of the dependence is needed.

In this work, we focus on SNPSet tests where a block of variables are jointly tested in an approach similar to the global testing introduced in [1]. In our context, both the dependence pattern and the association signal can be very different between regions of the genome. The presentation will first show that the two extreme choices consisting in ignoring dependence or on the contrary whitening the pointwise test statistics cannot be uniformly powerful over the variety of dependence and association patterns. We therefore introduce a new class of aggregation methods spanning the range between ignorance of dependence and complete decorrelation. We also propose a method minimizing a distance between the null and non-null moment generating functions of the test statistics within the former class to choose the more appropriate handling of dependence.

References

[1] Arias-Castro, E., Candès, E.J. and Plan, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *The Annals of Statistics. 39, 5, 2533-2556.*.