Aggregation of dependent statistics in genome-wide association studies

Mathieu Emily*[1], Florian Hébert[2], and David Causeur[3]

[1]AGROCAMPUS OUEST – Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes – Institut Supérieur des Sciences Agronomiques, Agroalimentaires, Horticoles et du Paysage - 65, rue de St Brieuc - CS 84215 - 35042 Rennes cedex, France

[2]AGROCAMPUS OUEST, Institut de Recherche Mathématique de Rennes – CNRS-Université de Rennes1 : UMR6625 – France

[3]Institut de Recherche Mathématique de Rennes (IRMAR) – Agrocampus Ouest – Campus de Beaulieu, bâtiments 22 et 23, 263 avenue du Général Leclerc, CS 7420535042 RENNES Cédex, France

Résumé

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 , . . . , , 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
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..

*Intervenant