Models for copy number alterations data

Arief Gusnanto

Department of Statistics, University of Leeds, Leeds LS2 9PT, United Kingdom

Copy number alterations (CNA) are structural variation in the genome, in which some regions exhibit more or less than the normal two chromosomal copies. This genomic CNA profile provides critical information in tumour progression and is therefore informative for patients' clinical outcome or survival. Two important features of such data are, first, the data are high-dimensional as they cover the whole genome. Second, due to molecular mechanisms, CNA data exhibit high correlations in 'blocks' since CNA appear in segments. These create challenges in statistical modelling and interpretation such as, for example, in prediction of pathological subtypes, prediction of patients' survival, or in performing dimensional reduction. In dealing with these challenges we investigated different statistical models that mainly utilise random effects. Specifically, we consider three approaches with different assumptions of random effects to arrive at sparse solution for simpler interpretation. The term 'sparse' here refers to the case where most of the model parameters are zero estimated while the others are estimated away from zero. This talk will describe how these approaches can deal with the challenges and show that the results have a sensible and simpler interpretation. We will illustrate the methods using real data from lung cancer patients.

Keywords: Copy number alterations; cancer; high-dimensional; correlation; random effects; sparse solution.