Conditional false discovery rates in high dimensional data sets

Supervisor: Chris Wallace

One of the challenges in high-dimensional data analysis is the number of parallel hypotheses that are simultaneously tested. Methods to control the number of false discoveries, such as Benjamini-Hochberg, have been widely adopted as a pragmatic solution that avoids the considerable loss of power that can be induced by attempting to control family-wise error rates.

Often, multiple datasets exist which explore the same dimensions in different contexts, such as testing the same genes for association with different diseases. When the diseases are related, information provided by one study can be used to increase power in the other. Conditional false discovery rates, which exploit empirical Bayes ideas, are an intuitive and powerful method to do this, by defining variable thresholds for significance whilst maintaining strong control of false discovery rates (FDR). By using cumulative density functions, they are more robust to sampling variation in lower-powered and lower-dimensional datasets than methods to modulate FDR which model the probability density functions. However, as currently defined, they can incorporate information from only a single quantitative covariate. There is considerable scope for their extension to multiple and/or ordered covariates, or even high-dimensional covariates through dimensionality-reducing mapping functions.

This PhD offers the chance to develop these extensions, and apply them in a range of areas in high-dimensional omics data, for example to the detection of interactions in high-dimensional data, or through developing custom mapping functions to detecting rare variants that cause rare, extreme forms of autoimmunity, building on information available from their common disease counterparts already studied.

Reference:

Informal enquiries to Chris Wallace are welcome: cew54@cam.ac.uk

More about our group: http://chr1swallace.github.io

Start date: Easter Term (April) or Michaelmas Term (October) 2019

All application queries regarding eligibility should be directed to phdstudy@mrc-bsu.cam.ac.uk

How to Apply: Applications should be made on-line via www.graduate.study.cam.ac.uk/applicant-portal selecting course details MDBI22 PhD in Biostatistics

Deadline for applications: 3rd January 2019