

Multiple testing procedures for big data analysis

Arnold Janssen Heinrich-Heine-University Duesseldorf
(partially joint work with Marc Ditzhaus, TU Dortmund)

Big data analysis like genome wide association studies require new statistical inference tools. It is a big issue to judge a huge amount of tests for one single data set simultaneously. A well known approach is based on the false discovery rate control (FDR) beginning with Benjamini and Hochberg (1995). Multiple tests select a portion of null hypotheses which are likely to be rejected. The FDR principle controls the ratio of the number of false rejections divided by the amount of all rejections in the mean.

The talk discusses new multiple tests which lead to level alpha control of the FDR. Special attention is devoted to sparsity models when only a few false null are hidden within a big data set. Throughout, exact finite sample control results are presented as well the asymptotics about the stability of the procedures.