False Discovery Rates - A New Deal

Dr. Matthew Stephens, University of Chicago

Abstract

False Discovery Rate (FDR) methodology, first put forward by Benjamini and Hochberg, and further developed by many authors - including Storey, Tibshirani, and Efron - is now one of the most widely used statistical methods in genomics, among other areas of application. A typical genomics workflow consists of i) estimating thousands of effects, and their associated p values; ii) feeding these p values to software (e.g. the widely used q-value package) to estimate the FDR for any given significance threshold. In this talk we take a fresh look at this problem, and highlight two deficiencies of this standard pipeline that we believe could be improved. First, current methods, being based directly on p values (or z scores), fail to fully account for the fact that some measurements are more precise than others. Second, current methods assume that the least significant p values (those near 1) are all null – something that initially appears intuitive, but will not necessarily hold in practice. We suggest simple approaches to address both issues, and demonstrate the potential for these methods to increase the number of discoveries at a given FDR threshold. We also discuss the connection between this problem and shrinkage estimation, and problems involving sparsity more generally.

Note: Taking photos during the seminar is prohibited

For disability access information or listening devices, please contact the Office of Support Services at 410-955-1197 or on the Web at www.jhsph.edu/SupportServices. EO/AA