Writing Project Proposal

Connor Demorest

10/22/2021

1) Topic of interest: Write a few sentences trying to describe what you want to work on and what you might do to address it. Include a reference to one source you have or plan to read or note that you are looking for pertinent literature (make sure the reference is included in a proper citation style) - course notes could also be a reference at this point but would just be a starting point.

Genetics and metabolomics data are data that commonly have few observations and many response variables. A common method to determine if a gene or metabolite is different between a control and treated unit is using a hypothesis test, but there are often thousands or more genes being tested in a metabolomics study. Multiple hypothesis testing (MHT) methods are needed to correct multiplicity errors arising from the many hypothesis tests being performed simultaneously to control for the error rate. One of the most common MHT methods was established in Benjamini and Hochberg (1995), which introduced the concept of False Discovery Rates (FDR). FDR controlling methods control the expected proportion of incorrect rejections of the null hypotheses to a reasonable level, while making correct rejections with high power. A Bayesian perspective of MHT was established in Efron et al. (2001) and advanced in Sun and Cai (2007) with the Lfdr, the "local" false discovery rate, as a Bayesian analogue of the p-value. The focus of this project is to explore the Bayesian approach to MHT, compare Lfdr-based techniques to the frequentist procedures, and apply an Lfdr-based procedure to correct multiplicity errors and analyze a metabolomics dataset.

2) Planned faculty advisor(s)

Dr. Nandi is planning to be my faculty advisor for this project. Dr. Greenwood and Dr. Hoegh are the other two committee members. Their expertise in multivariate statistics and Bayesian statistics will be helpful.

3) Current status of work

We have found a potential data set to use, or a contact that could provide real experimental data if we use a different data set. A brief literature review has been done. Simulation studies have been done to show the necessity of MHT methods. Additional simulation studies and data analysis using MHT methods remain to be done.

References:

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300.

Efron, Bradley, Robert Tibshirani, John D Storey, and Virginia Tusher. 2001. "Empirical Bayes Analysis of a Microarray Experiment." *Journal of the American Statistical Association* 96 (456): 1151–60. https://doi.org/10.1198/016214501753382129.

Sun, Wenguang, and T Tony Cai. 2007. "Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control." *Journal of the American Statistical Association* 102 (479): 901–12.