

PHD DISSERTATION DEFENSE ANNOUNCEMENT

**MULTIPLE TESTING PROCEDURES FOR COMPLEX STRUCTURED HYPOTHESES AND
DIRECTIONAL DECISIONS**

Anjana Grandhi

Ph.D. candidate in Probability and Applied Statistics
Department of Mathematical Sciences, NJIT, Newark, NJ

Co-Advisors:

Dr. Wenge Guo, Associate Professor, Department of Mathematical Sciences, NJIT
Dr. Ji Meng Loh, Associate Professor, Department of Mathematical Sciences, NJIT

Committee Members:

Dr. Sunil Dhar, Professor, Department of Mathematical Sciences, NJIT
Dr. Sundar Subramanian, Associate Professor, Department of Mathematical Sciences, NJIT
Dr. Zhi Wei, Associate Professor, Department of Computer Science, NJIT

Date: Tuesday, April 28, 2015 Time: 3:00 PM – 5:00 PM Venue: Cullimore Hall, Room 611

ABSTRACT

Several multiple testing procedures are developed based on the inherent structure of the tested hypotheses and specific needs of data analysis. Incorporating the inherent structure of the hypotheses results in development of more powerful and situation specific multiple testing procedures than the existing ones. The focus of this dissertation is on developing multiple testing procedures that utilize the information on this structure of the hypotheses and aim at answering research questions while controlling appropriate error rates.

In the first part of the thesis, a mixed directional false discovery rate (mdFDR) controlling procedure is developed in the context of Uterine Fibroid gene expression data (Davis et al. (2013)). The main question of interest that arises in this research is to find out which genes are associated with onset, growth and size of tumor and find the trends in mean gene expressions across several categories of size of tumor, location etc. To answer these questions, a three-step testing strategy is introduced and a general procedure is proposed that can be used with any mixed directional familywise error rate (mdFWER) controlling procedure for each gene, while controlling the mdFDR as the overall error rate. A specific methodology, based on the Dunnett procedure, is developed and applied to the uterine fibroid gene expression data analysis. Several important genes and pathways are identified that play important role in fibroid formation and growth.

In the second part, the problem of simultaneously testing many two-sided hypotheses is considered when rejections of null hypotheses are accompanied by claims on the direction of the alternative. The fundamental goal is to construct methods that control the mdFWER, which is the probability of making a Type I or Type III (directional) error. In particular, attention is focused on cases where the hypotheses are ordered so that a hypothesis is tested only if all the preceding hypotheses have been rejected. This research proves that the conventional Fixed Sequence procedure, when augmented with directional decisions, can control mdFWER under independence and positive regression dependence of the test statistics. Another more conservative directional procedure is also developed that strongly controls mdFWER under arbitrary dependence of test statistics.

Finally, in the third part, multiple testing procedures are developed for making real time decisions while testing a sequence of a-priori ordered hypotheses. In large scale multiple testing problems in applications such as stream data, statistical process control, etc., the underlying process is regularly monitored and it is desired to control False Discovery Rate (FDR) while making real time decisions about the process being out of control or not. The existing stepwise FDR controlling procedures, such as the Benjamini-Hochberg procedure, are not applicable here because of the implicit assumption that all the p-values are available for applying the testing procedure. In this part of the thesis, powerful Fallback-type procedures are developed under various dependencies for controlling FDR that award the critical constants on rejection of a hypothesis. These procedures overcome the drawback of the conventional FDR controlling procedures by making real time decisions based on partial information available when a hypothesis is tested and allowing testing of each a-priori ordered hypothesis.