

# Report for 04-A Multiple Test Correction for Streams and Cascades of Statistical Hypothesis Tests

Xinyu Chen

February 16, 2017

## 1 Summary

This paper presented the Subfamilywise Multiple Testing. This is a multiple-testing correction method to control the familywise error rate(FWER) without the knowledge of specific hypotheses or the number of hypotheses beforehand.

The paper used a *Monte Carlo* simulation on learning graphic models to demonstrate this Subfamilywise Multiple Testing is important in current machine learning problems. The authors used both synthetic and real-world datasets to show the Subfamilywise Multiple Testing can produce better results in model selection learning tasks.

## 2 Key Takeaway

Multiple Testing Correction or Multiple Comparison problem is a important issue in many researches. I am more aware of this problem after I've read this paper and looked up for relevant content. The authors presented insights into this problem with regards to machine learning algorithms. They used *Monte Carlo* simulation and this is a great method to gain better understanding of problems from the perspective of probabilities.

### 3 Discussions

- *hypothesis streams.* The authors emphasized the number of hypotheses is undetermined and they mentioned hypothesis streams as subsequent hypotheses are unknown until some hypotheses are rejected. However, this violated the assumption for FWER that hypotheses are independent. Under the situation of hypothesis streams, figure 1 does not hold. We should treat FWER as conditional probability in such cases.
- *graphical model selection.* The number of edges at step  $s$  is  $M = n \cdot (n - 1)/2$  is correct, but the number of models  $M \cdot (M + 1)/2$  is not quite right. Maybe the authors want to show there are that many *configurations*.
- *Monte Carlo simulation.* The p-value are *uniform* distributed from  $[0, 1]$  for true hypotheses and  $[0, \text{maxFalsePValue}]$  for false hypotheses, which is not quite reasonable. The p-values are probabilities, they should confirm with the model distribution. In the Monte Carlo simulation, p-values are uniformly distributed, the number of *rejected true* hypotheses increases faster when the size of a *subfamily* increases. Figure 2 is not a good visualization example.
- *Subfamily size.* The authors claim the probability of a subfamily containing false as well as true hypotheses increases as the size of subfamilies increases. This is not true, unless they designed to change this probability. It should remain the same regardless the size. The rejections of *rfalse* hypotheses has no effect on FWER. Maybe the simulation can omit false ones?
- *Model Selection.* The algorithm update the *budget*  $\alpha$  iteratively. Each time it subtracts  $p \cdot |E|$ . Can this *budget*  $\alpha$  becomes negative with some big *p-values*?