

# Multiple Testing of Causal Hypotheses

Samantha Kleinberg and Bud Mishra

April 30, 2008

Courant Institute of Mathematical Sciences, New York University

## Abstract

A primary problem in causal inference is the following: From a set of time course data, such as that generated by gene expression microarrays, is it possible to infer all significant causal relationships between the elements described by this data? In prior work [2], we have proposed a framework that combines notions of causality in philosophy, with the algorithmic approaches built on model checking and statistical techniques for multiple hypotheses testing. The causal relationships can be then described in terms of temporal logic formulas, reframing the problem in terms of model checking. The logic used, PCTL, allows description of both the time between cause and effect and the probability of this relationship being observed. Borrowing from philosophy, we define prima facie causes in terms of probability raising, and then determine genuine causality by computing the average difference a prima facie cause makes to the occurrence of its effect, given each of the other prima facie causes of that effect. However, it faces many interesting issues confronted in statistical theories of hypotheses testing, namely, given these causal formulas with their associated probabilities and our average computed differences, instead of choosing an arbitrary threshold, how do we decide which are “significant”? To address this problem rigorously, we use the concepts of multiple hypothesis testing (treating each causal relationship as a hypothesis), and false discovery control. In particular, we apply the empirical Bayesian formulation proposed by Efron in [1]. This method uses an empirical rather than theoretical null, which has been shown to be better equipped for cases where the test statistics are dependent - as may be true in the case of complex causal structures.

## References

- [1] B. Efron. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association*, 99(465):96–105, 2004.
- [2] S. Kleinberg and B. Mishra. Inferring causation in time course data with temporal logic. Under review, 2008.