Information Flow Experiments
Michael Tschantz*, Amit Datta, Anupam Datta, Jeanette Wing^ 
*University of California Berkeley, Carnegie Mellon University, ^Microsoft Research

Motivation

Ads are omnipresent on the web.

Statsistics

- Non-parametric statistical tests make no assumptions about the ad distributions.
- Null hypothesis: Both ad sets come from the same distribution.

Goals

- Design experiments to detect information flows.
- Develop a statistical theory that supports our experimental design and allows us to draw useful conclusions.

Experiments

- Permutation Test
  
  ![Diagram of experiments with user, later ads, Google, and advertisers]

  Cosine Similarity  |  Keyword Analysis  |  $\chi^2$
  |  |  |
  control-experiment  |  18/20  |  18/20  |  20/20
  control-control  |  1/20  |  1/20  |  12/20
  experiment-experiment  |  0/20  |  1/20  |  5/20

Related Work

- Guha et al, 2010: Developed techniques for measuring online advertising systems: cosine similarity, display URL.
- Balebako et al, 2012: Analyzed effects of privacy tools in limiting behavioral advertisements: Ghostery, DoNotTrack, OptOut, etc.
- Wills and Tatar, 2012: Miscellaneous studies on online advertisements: multiple signals, keywords.
- Sweeney, 2013: Discrimination in online ad delivery: correlation in racially stereotypical names and the term ‘arrest’.

- We observed a cross-browser effect among the different instances running together.
- Permutation tests for observing effect does not require lack of cross-unit effects.
- Statistical tests allow us to show interference exists. Coupled with the theorem:
  Interference $\iff$ Causal Effect
  we conclude ‘the signals have a causal effect on the ads’.