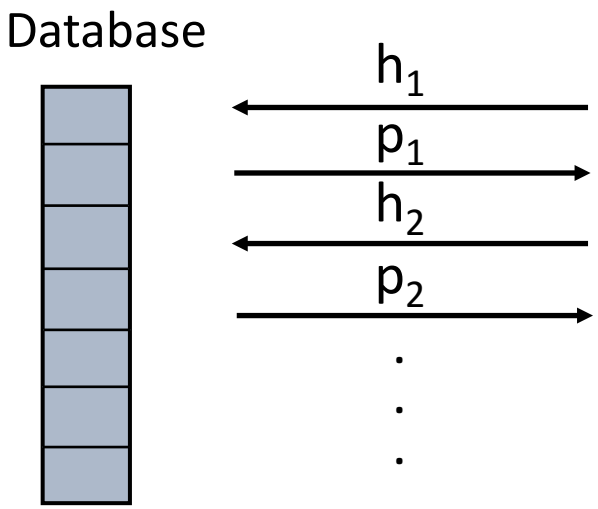
PAPRIKA: Private Online False Discovery Rate Control

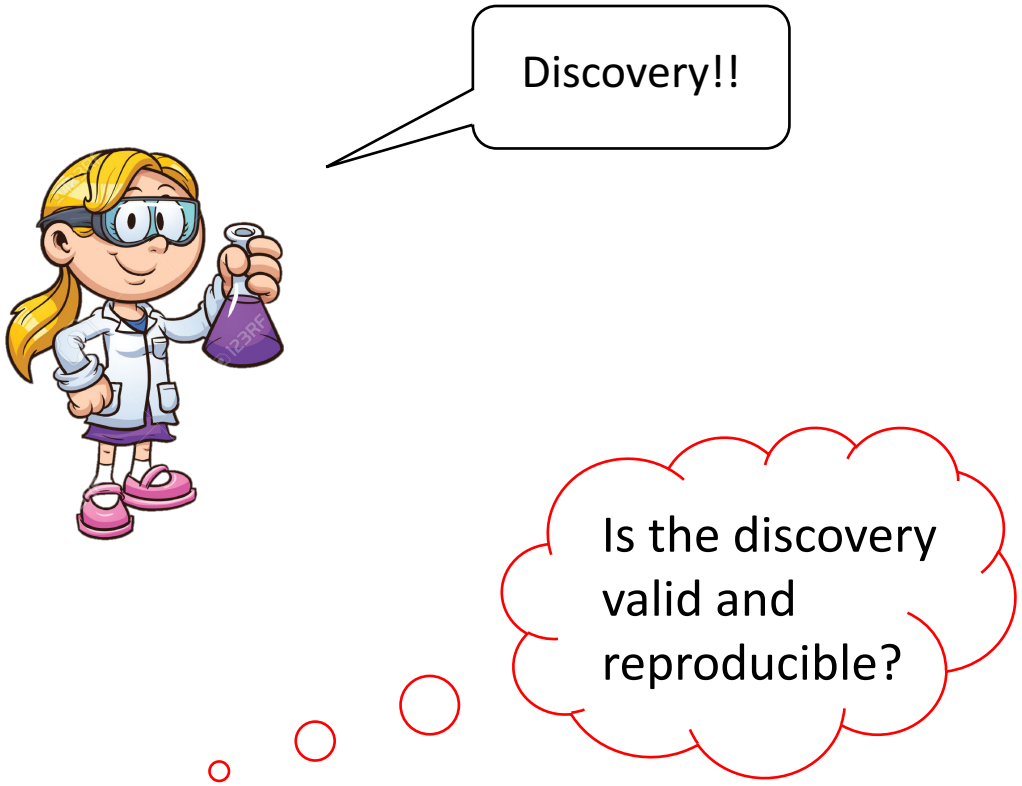
Wanrong Zhang, Georgia Tech
Gautam Kamath, University of Waterloo
Rachel Cummings, Columbia University

ICML, 2021

https://arxiv.org/abs/2002.12321

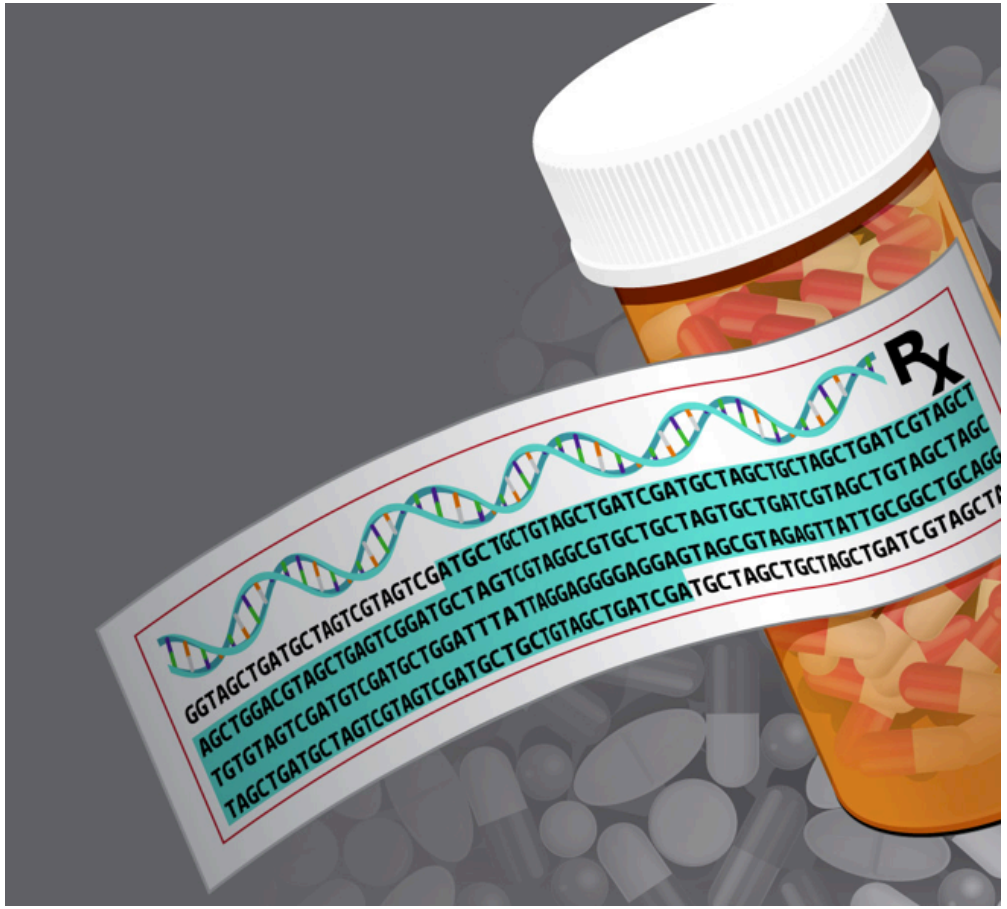
False Discovery





False Discovery Rate (FDR) Control

- Goal: design a procedure that takes in p-values of all hypotheses and decides which to reject.
  - Control FDR to be below than a threshold
  - Maintain high true positive rate (power).



Need formal privacy guarantees for FDR control algorithms

False Discovery Rate (FDR) Control

Let $\mathcal{R}$ be set of all rejected hypotheses, and let $\mathcal{H}^0$ be the set of hypotheses for which the null is true

$$FDR = \mathbb{E}[FDP(\mathcal{R})] = \mathbb{E}\left[\frac{|\mathcal{H}^0 \cap \mathcal{R}|}{|\mathcal{R}|}\right]$$

Private method [Dwork et al., 2018]

Offline vs Online

- Offline: know all p-values in advance and make reject decisions at the end
- Online: p-values arrive one at a time and must make reject decisions immediately

Open problem: no known privately tools

Differential privacy [DMNS '06]

Bound the maximum amount that one person's data can change the distribution of an algorithm's output

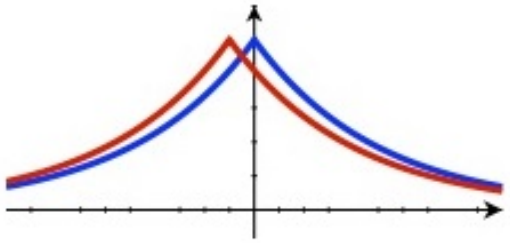
An algorithm $M: T^n \rightarrow R$ is (ϵ, δ)-**differentially private** if $\forall$ neighboring $x, x' \in T^n$ and $\forall\, S \subseteq R$,

$$P[M(x) \in S] \leq e^{\epsilon}\, P[M(x') \in S] + \delta$$

$(t_1, \ldots, t_i, \ldots, t_n)$

$(t_1, \ldots, t_i', \ldots, t_n)$

$\Rightarrow$



PAPRIKA

GAI + SVT

- Generalized Alpha-investing rules: LORD++ [RYWJ'17].
- Dynamic thresholds in SVT [DPNR '10]: to match the alpha-investing rule
- Adding noise that scales with the multiplicative sensitivity [DSZ'18]

  of p-values to reduce the noise for privacy
- Shifting the threshold to accommodate FDR as a novel accuracy metric
- The candidacy indicator step from SAFFRON [RZWJ'18] cannot be done privately and requires new analysis for both privacy and accuracy.

Main Result

Theorem: PAPRIKA is (ϵ, δ)-differentially private and controls FDR to be below $\alpha + \delta t$ for testing t hypotheses if *p*-values are independent .

δ is cryptographically small for privacy, so this term tiny

- Controls mFDR to be below $\alpha + \delta t$ for testing t hypotheses if *p*-values are adaptive under the null.

Empirical Results (In paper)

PAPRIKA: Private Online False Discovery Rate Control

Wanrong Zhang, Georgia Tech
Gautam Kamath, University of Waterloo
Rachel Cummings, Columbia University

ICML, 2021

https://arxiv.org/abs/2002.12321