Musings on False Discovery Rate

OMER REINGOLD
MSR

DISCUSSIONS WITH:
CYNTHIA DWORK (MSR), REBECCA POTTENGER (BERKELEY), GREGORY VALIANT (STANFORD), GUY ROTHBLUM (MSR), MORITZ HARDT (IBM), AARON ROTH (UPENN)

***ORIGINAL MOTIVATION**: SYNTACTIC SIMILARITY BETWEEN MULTIPLE HYPOTHESIS TESTING & ERROR ACCUMULATION IN DIFFERENTIAL PRIVACY (MORE LATER)*

Not an Expert, So What's the Plan?

Dear Diary, yesterday we read Benjamini–Hochberg ♥♥♥. I think I know how to re-prove it ☺. Awesome !!!!!1

- More and more, Statistics and CS are overlapping
- So, going through the misunderstandings and raw thoughts of CS theoreticians may be worth it
  - Some may be naïve or flat out wrong but that's good too …
  - I'll omit almost all references (as not to get it wrong).
- We started our investigation 20 years behind, but some of our ideas are now only a couple of years late!
- Plan: (1) missing parameters, (2) separating mixtures of distributions, (3) error of the procedure

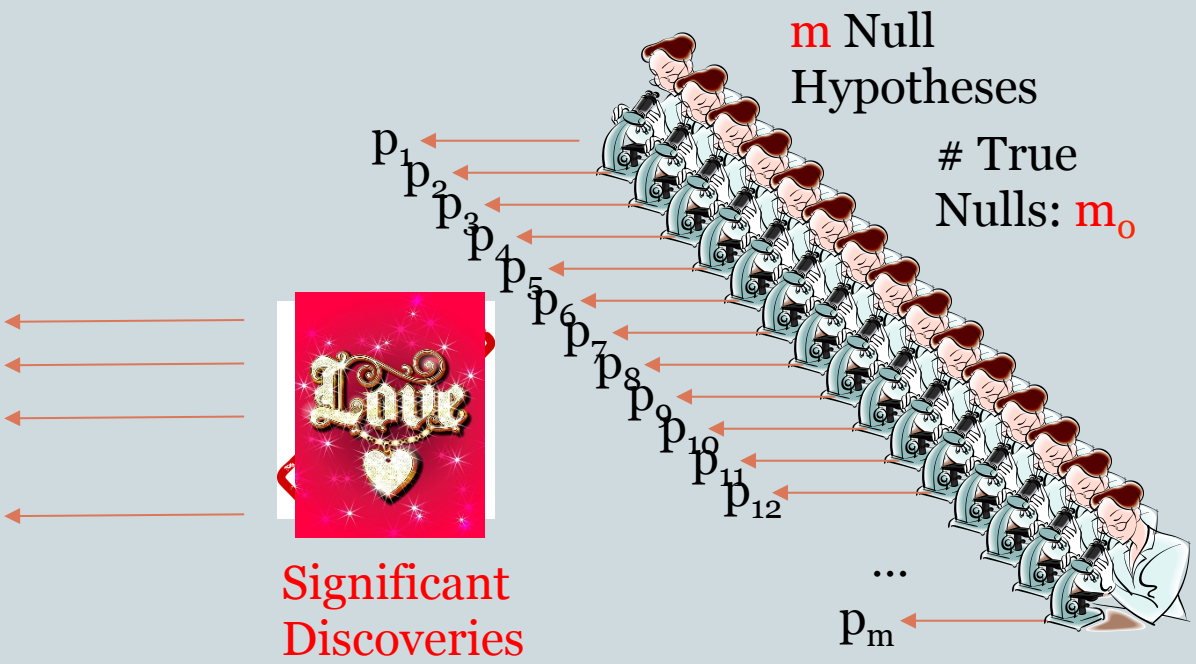
Quick Recap - Multiple Hypothesis Testing

Rejected: R

Rejected True Nulls (type I error): V

FWER (family-wise error): Pr[V>0]

FDR (false discovery rate): E[V/max{R,1}]



Missing Parameters? Is fewer always better?

- Raised on Valiant's Probably Approximately Correct (PAC) learning. Learn an approximation (hypothesis) h to f s.t.

$$\Pr_{\text{learner randomness}} [\Pr_{x} [h(x) \neq f(x)] > \varepsilon] < \delta$$

- Separating these parameters is responsible for some of the most important wo
- p-values seem to comb ds
  - Statistics literature aware o m
- Similarly, natural first R: i
  rate of false discovery good enough, or do we want:

$$\Pr_{\text{tests' randomness}} [\text{false discovery rate} > \varepsilon] < \delta$$

Material Significance

Statistical Significance

A Posteriori Guarantees ?

- A criticism of FDR=E[V/max{R,1}]:
  w. prob. ½, R=0 and E[V/R | R>0]=2α ⇒ FDR= α
- FDR: a priori prob. that a random rejected is truly null. How about a posteriori guarantee (say, given R>0).
- [Storey 01] positive FDR: pFDR=E[V/R | R>0]
- Possible criticism:
  w. prob. ½, R=1, E[V/R | R=1]=0, w. prob. ½, R=100, and E[V/R | R=100]=2α ⇒ FDR= α
- In a Bayesian setting (each null hypothesis is false with a fixed i.i.d. probability), pFDR has an interpretation as a posterior probability.

A Posteriori Guarantees for Frequentists?

- Can we bound $E[V/R \mid p_1, \ldots, p_m]$?
  - Not in the setting of [Benjamini–Hochberg]:
    S - set of true nulls, p_i is i.i.d and uniform $\forall\, i \in S$, no assumption on other p_i's
- What is the right definition then?

Mixture of Distributions (A Framework)

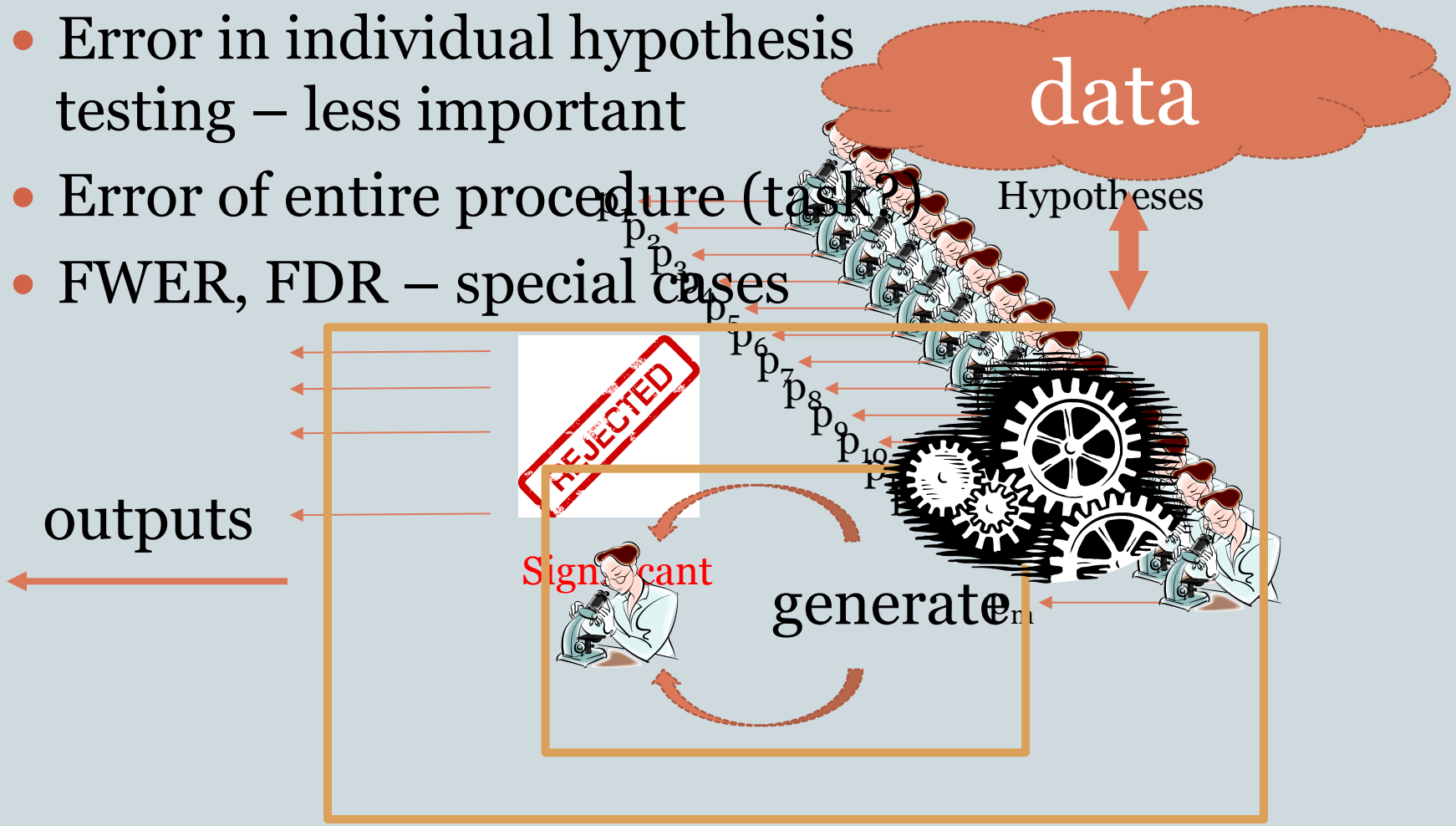
- Concentrate on estimation m_0 (# true nulls)
- m_0 values p_i uniform in [0,1] (denote U([0,1])).
- For the specific observations $\{p_j \mid j \text{ truly not null}\}$, define X to be the uniform distribution over these values.
- The distribution p_i where i is uniform in [m] is a mixture of X and U([0,1]), with weights $(m- m_0)/m$ and m_0/m.
- Approach: find a (provable) estimator m_0 s.t

1. $(m \geq) E[m_0] \geq m_0$
2. As X gets far from uniform, $E[m_0]$ gets closer to m_0

- "Far"? Several options: earth mover distance, moments…
- Each choice suggests a different estimator

Picture is Changing – Error of Computations

- Error in individual hypothesis testing – less important
- Error of entire procedure (task?)
- FWER, FDR – special cases



Adaptive Choice of Hypotheses

- Can we search an hypothesis space smartly to avoid full price in errors?
- Many existing instantiations: filtering, hierarchical testing, pair-wise comparisons, model selection …
- Controlling the procedure error as a framework may give more and help connect to machine learning.
- Connection to Differential Privacy: when querying a database privacy loss may accumulate.
  - But, does not accumulate linearly [DRV10].
  - Similar phenomenon if testing related hypotheses?
- Adaptive choice of hypotheses has dangers too (especially if reusing data). CS may help here too.

Conclusions

- Many connection points between Multiple Hypothesis Testing and Theory of Computing (Machine Learning, Differential Privacy, Property Testing, ..)
- Perspective may be a bit different, but bridging the gap could be fruitful
- Controlling the procedure error?