Max-Information, Differential Privacy, and Post-Selection Hypothesis Testing

Ryan Rogers, Aaron Roth, Adam Smith, and Om Thakkar

Non-Adaptive Data Analysis

$X \sim P^n$

Run Experiment

Outcome

??

Test t

Non-Adaptive Data Analysis

$X \sim P^n$

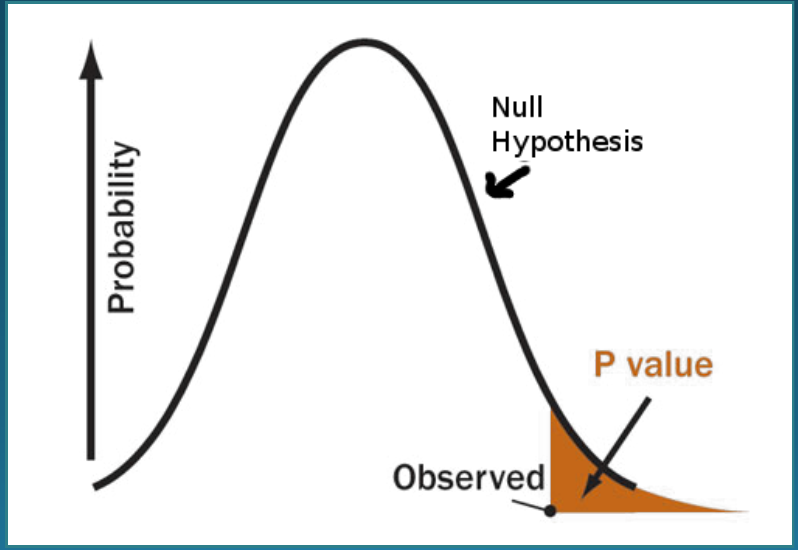
Run Experiment

Outcome

Test t

Application: Hypothesis Testing

- Hypothesis test is defined by a test statistic $t: D^n \to \mathbb{R}$ and a null hypothesis $H_0 \subseteq \Delta(D)$.
- The *p*-value associated with a value a and a distribution $P \in H_0$ is given as $p(a) = \Pr_{x \sim P^n}[t(x) > a]$
  - Denotes the probability of observing a value of the test statistic that is at least as extreme as a.



Application: Hypothesis Testing

- The goal is to reject H_0 if the data is not likely to have been generated from that model.
- Note that $p\big(t(X)\big) \sim U[0,1]$ if $X \sim P^n$ where $P \in H_0$.
- If we reject the model when $p\big(t(x)\big) < \alpha$ then False Discovery < α.

Non-Adaptive Data Analysis

$X \sim P^n$

Run Experiment

Outcome

??

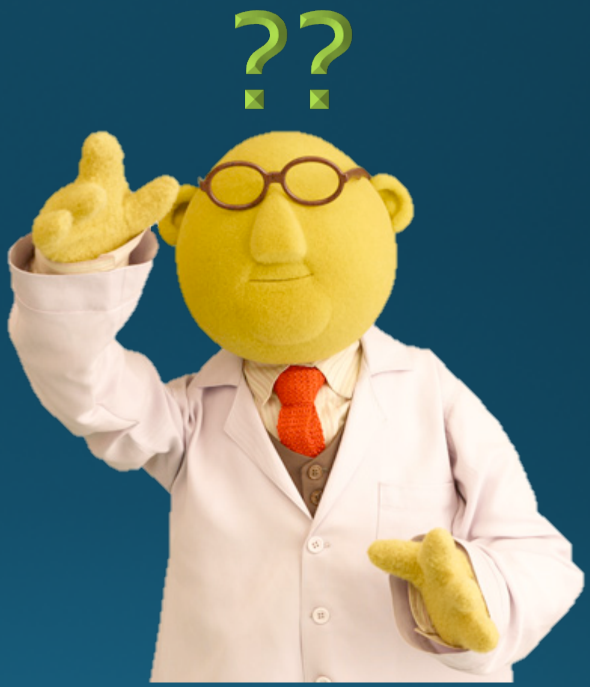
Test t

Nothing Significant

$+t'$

Non-Adaptive Data Analysis

$X \sim P^n$



$+t'$

Test t

Non-Adaptive Data Analysis

Run Experiment

Outcome

$X' \sim P^n$

??

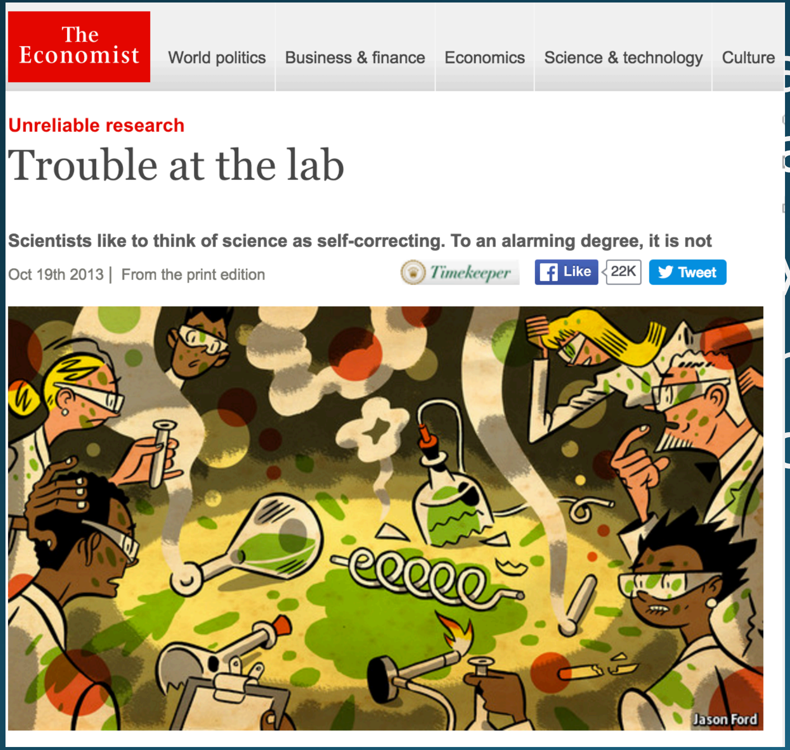
$+t'$

Test t

Need the choice of test to be independent of the data

False Discovery

- Rejecting a true null hypothesis should occur in at most an α fraction of the tests.


The Economist
World politics | Business & finance | Economics | Science & technology | Culture

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

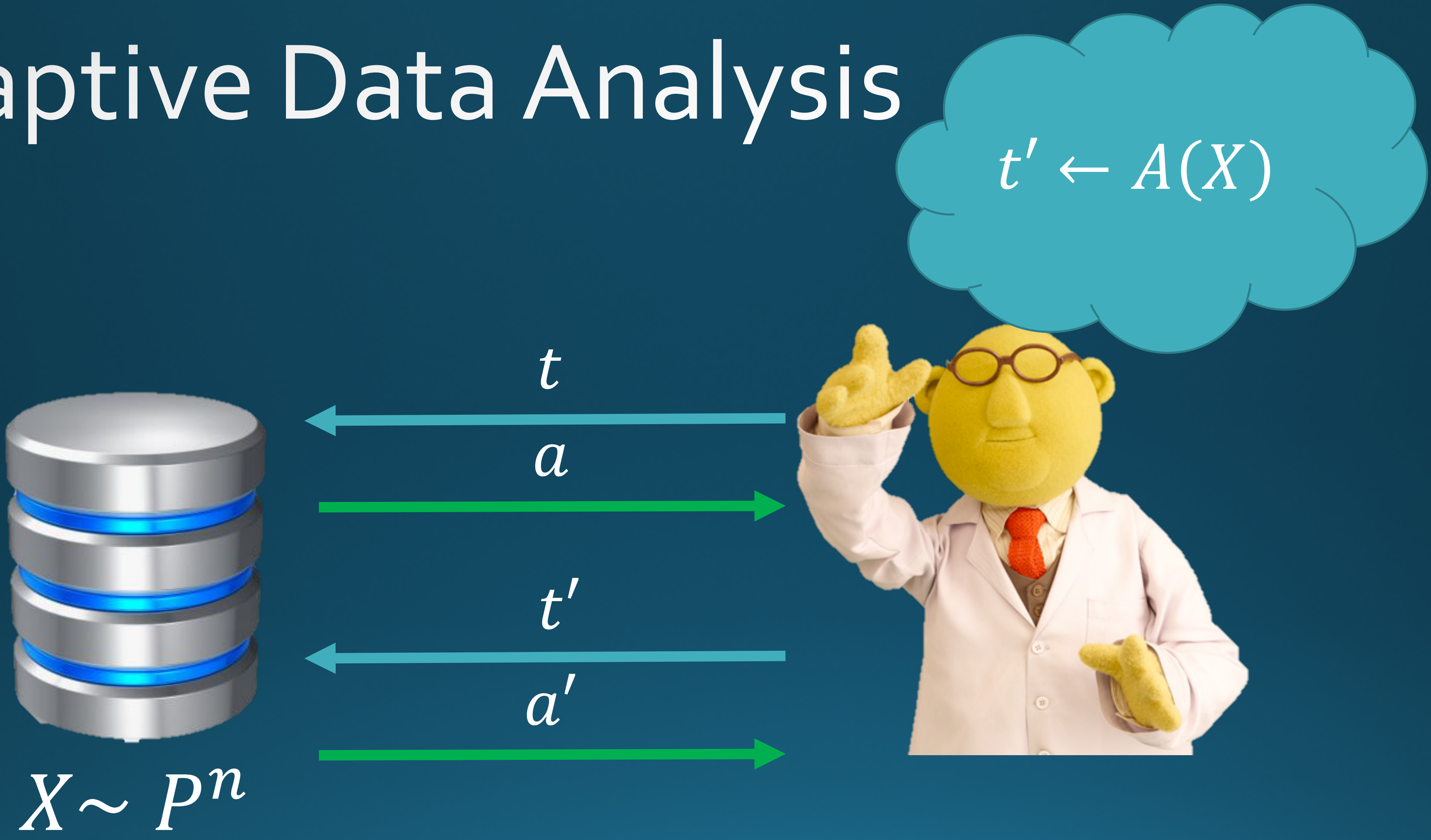
Timekeeper | Like 22K | Tweet

ally relies
ata.

ysis is inh

n the scie
ch highe

Adaptive Data Analysis



Problem: $p(t'(X))$ is no longer uniform.

Valid p-Value Correction

- Even when we use the data to determine a test, we still want to be able to control the false discovery rate.
- A function $\gamma: [0,1] \rightarrow [0,1]$ is a valid *p*-value correction function for a selection procedure $A: D^n \rightarrow T$ if for every α the procedure:
  1. Select test $t \leftarrow A(X)$
  2. Reject H_0 if $p\big(t(X)\big) < \gamma(\alpha)$

  has probability at most α of false discovery.
- We will assume that the selection procedure *A* satisfies some property

Max-Information [DFHPRR15]

- An algorithm $A: D^n \to T$ with bounded max-info allows the analyst to treat $A(X)$ as if it is independent of data X up to a correction factor determined be the max-info bound.
- The β-approximate max-info between two random variables Y and Z is

$$I_\infty^\beta(Y;Z) = \log\left(\sup_{O:\Pr[(Y,Z)\in O]>\beta} \frac{\Pr[(Y,Z)\in O]-\beta}{\Pr[Y\otimes Z\in O]}\right)$$

- If $\Pr_{(y,z)\sim(Y,Z)}\left[\frac{\Pr[(Y,Z)=(y,z)]}{\Pr[Y=y]\Pr[Z=z]} \geq 2^k\right] \leq \beta$ then $I_\infty^\beta(Y;Z) \leq k$.

Max-Information [DFHPRR15]

- We say that an algorithm A has β-approximate max-info at most k, denoted as $I_{\infty}^{\beta}(A, n) \leq k$ if for every distribution S over datasets D^n we have $I_{\infty}^{\beta}(X; A(X)) \leq k$ where $X \sim S$.
- It will be important to distinguish max-info over product distributions, denoted $I_{\infty,\Pi}^{\beta}(A, n)$, which is the same as above except S can only be a product distribution, i.e. $S = P^n$ for some P over D.

Max-Info gives Valid p-Value Corrections

- If we have selection procedure A such that $I^0_{\infty,\Pi}(A,n) \leq k$ then a valid *p*-value correction function is
$$\gamma(\alpha) = \frac{\alpha}{2^k}$$
- Proof: Let $O \subseteq D^n \times T$ be the event that A selects a test statistic where the *p*-value is at most $\gamma(\alpha)$, but the null is true.
$$\begin{aligned} &\Pr\left[p\left(t(X)\right) \leq \gamma(\alpha) \cap t = A(X)\right] \\ &= \Pr\left[\left(X, A(X)\right) \in O\right] \\ &\leq 2^k \underbrace{\Pr[X \otimes A(X') \in O\,]}_{\leq \gamma(\alpha)} \end{aligned}$$

Max-Info gives Valid p-Value Corrections

- If we have selection procedure *A* such that $I^{\beta}_{\infty,\Pi}(A,n) \leq k$ then a valid *p*-value correction function is
$$\gamma(\alpha) = \frac{\alpha - \beta}{2^k}$$
- Proof: Let $O \subseteq D^n \times T$ be the event that *A* selects a test statistic where the *p*-value is at most $\gamma(\alpha)$, but the null is true.
$$\begin{aligned}\Pr\left[p\left(t(X)\right) \leq \gamma(\alpha) \cap t = A(X)\right] \\ = \Pr\left[\left(X, A(X)\right) \in O\right] \\ \leq 2^k \underbrace{\Pr[X \otimes A(X') \in O]}_{\leq \gamma(\alpha)} + \beta\end{aligned}$$

Mutual Info gives Valid p-Value Corrections

- For test selection $A: D^n \to T$ with mutual info $I\left(X; A(X)\right) \le m$ for any P where $X \sim P^n$, we can also obtain a valid p–value correction with the result from [RZ16], which leads to

$$\gamma(\alpha) = \min\left\{\frac{2^{\frac{-\log(e)m}{\alpha^2}}}{2}, \frac{\alpha}{2}\right\}$$

- However, a bound on mutual information of m giv
for any $k > 0$,

$$I_{\infty,\Pi}^{\beta(k)}(A, n) \le k \text{ where } \beta(k) \le$$

- Thus, when we have a bound on the mutual informatio
following valid p-value correction

$$\gamma(\alpha) = \frac{\alpha 2^{\frac{-2}{\alpha}(m+.54)}}{2}$$

We improve by using max-info for $\alpha \le 0.05$ when $m \ge 0.05$

Stability with Low-Sensitivity Queries

- From [BNSSSU'16] we know that other notions of stability lead to ways to estimate the values of adaptively chosen queries on the data:
  - Bound $|q(X) - q(X')|$ where $q \leftarrow A(X)$ w.h.p. over $X \sim P^n$ and A.
- A query $q: D^n \to \mathbb{R}$ is **low sensitive** if for any two datasets x, x' that differ in one entry we have

$$|q(x) - q(x')| \leq \Delta$$

- However, *p*-values are NOT low-sensitive enough:
  - Requires $\Delta > \frac{0.37}{\sqrt{n}}$
  - This sensitivity leads to a trivial error guarantee using results from [BNSSSU'16].

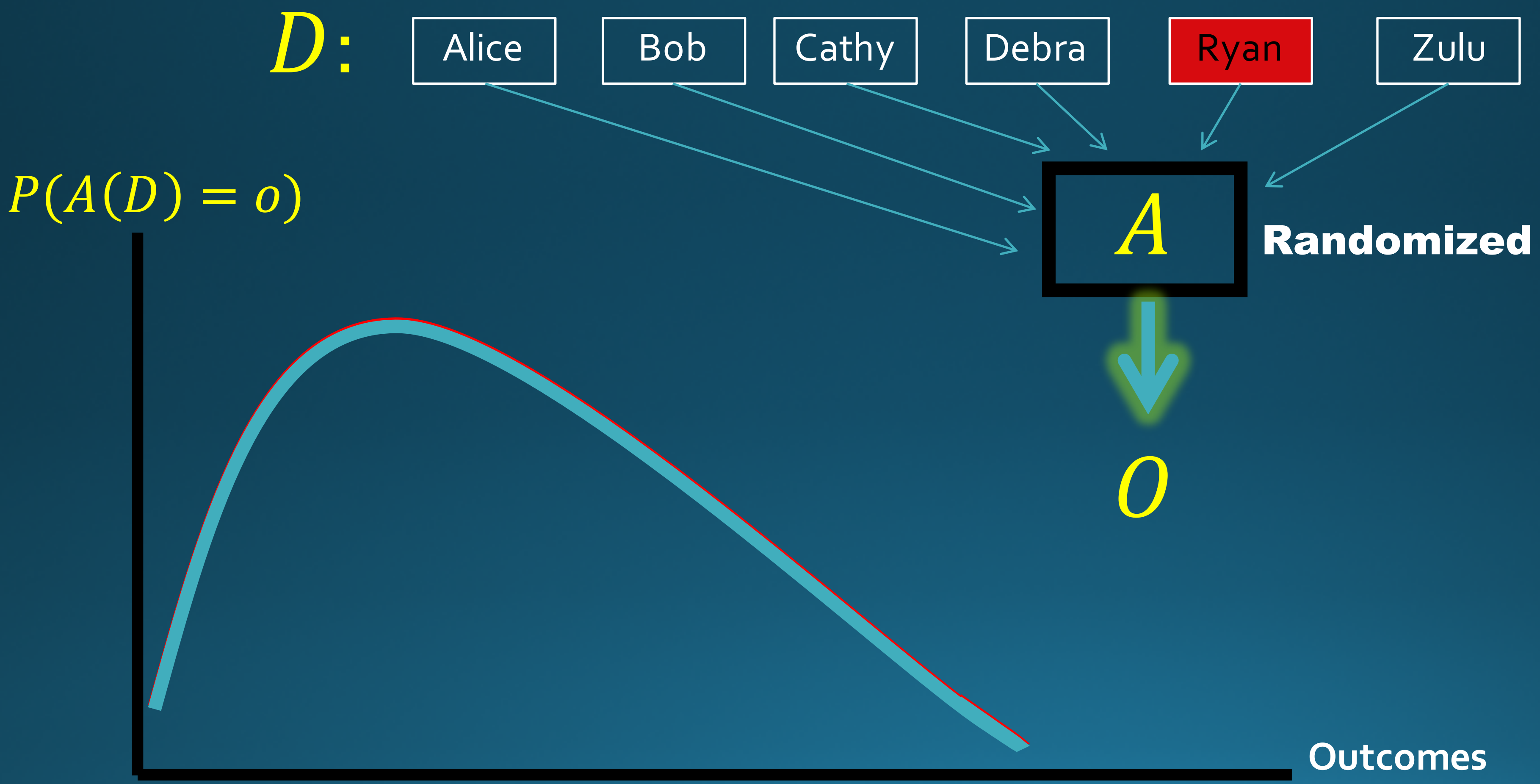
Max-Info Test Selection Procedures

- Question becomes, what test selection procedures $A: D^n \to T$ have bounded max-info?
- Recent work from [DFHPRR15] shows that the following procedures have bounded max-info:
  - (Pure) Differential Privacy – algorithmic stability condition.
  - Bounded Description Length – bound in terms of $|T|$.
- Nice composition rules for procedures with bounded max-info: if $I_\infty^{\beta_1}(A_1, n) \le k_1$ and $I_\infty^{\beta_2}(A_2, n) \le k_2$ then

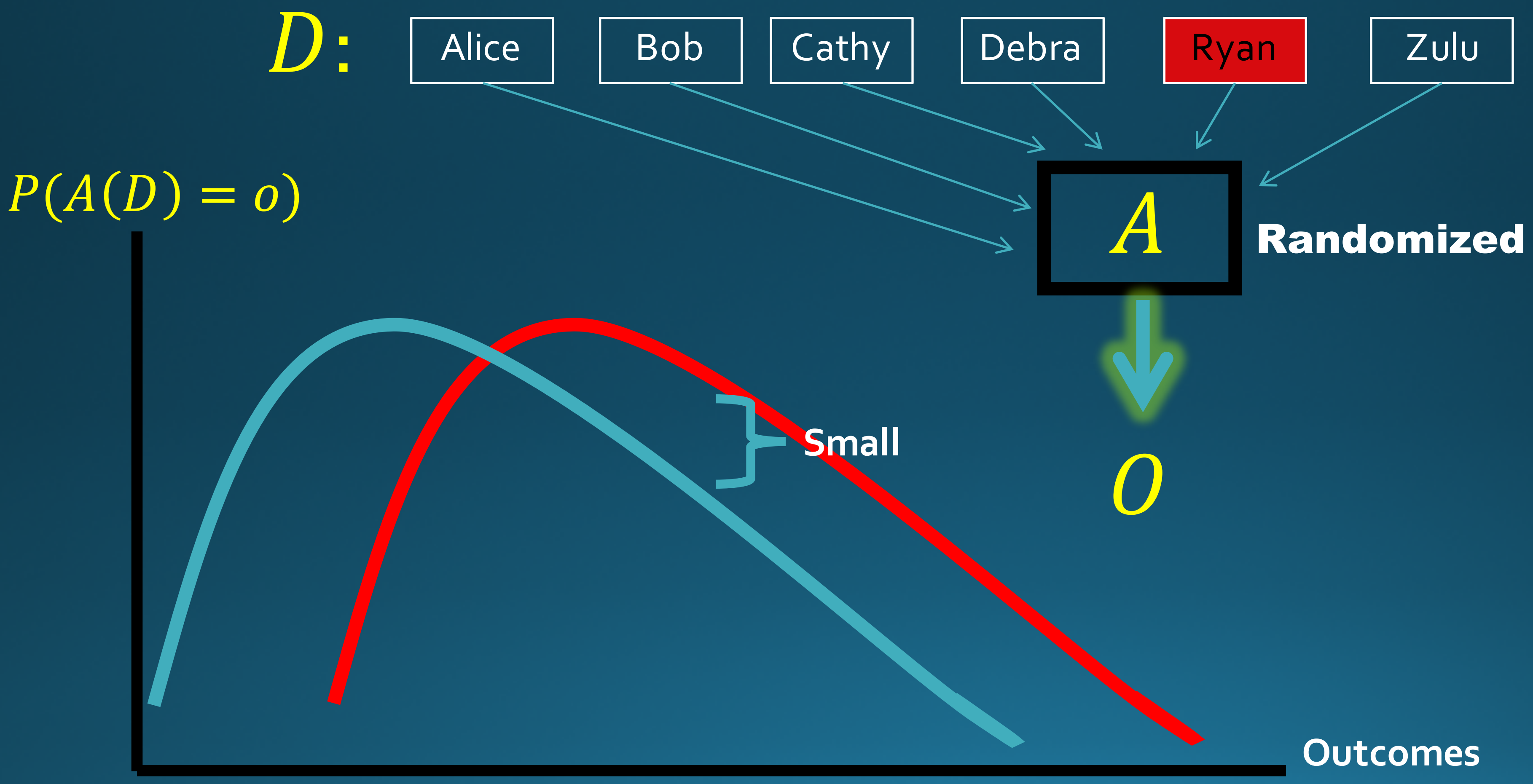
$$I_\infty^{\beta_1+\beta_2}(A_1 \circ A_2, n) \le k_1 + k_2$$

Differential Privacy [DMNS '06]

Differential Privacy [DMNS '06]



Differential Privacy [DMNS '06]



Differential Privacy [DMNS '06]

- A randomized algorithm $A: X^n \to Y$ is (ε, δ)-differentially private if for any neighboring data sets $D, D' \in X^n$ and for any outcome $S \subseteq Y$ we have

$$P(A(D) \in S) \leq e^{\varepsilon} P(A(D') \in S) + \delta$$

If $\delta = 0$ we say pure DP, and otherwise approximate DP.

Technical Contributions

- Past result [DFHPRR15]: If $A: D^n \rightarrow T$ is $(\epsilon, 0)$-DP then for $\beta > 0$,

$$I_\infty(A, n) \leq O(\epsilon n) \text{ and } I^{\beta}_{\infty,\Pi}(A, n) \leq O\left(\epsilon^2 n + \epsilon\sqrt{n \cdot \log\left(\frac{1}{\beta}\right)}\right)$$

- We achieve a bound on max-info with product distributions for the much larger class of (ϵ, δ)-DP algorithms.
- Important because adaptively composing ℓ many $(\epsilon', 0)$-DP algorithms leads to an overall $(\epsilon'\ell, 0)$-DP algorithm, but also for any $\delta > 0$, we get an $\left(O\left(\epsilon'\sqrt{\ell \log\left(\frac{1}{\delta}\right)}\right), \delta\right)$-DP algorithm.

Technical Contributions

- **Positive Result**: If $A: D^n \to T$ is (ϵ, δ)-DP then

$$I^{\beta}_{\infty,\Pi}(A, n) = O\left(n\,\epsilon^2 + n\,\sqrt{\epsilon\delta}\right) \qquad \text{for} \quad \beta = O\left(n\sqrt{\frac{\delta}{\epsilon}}\right)$$

In the case of low sensitive queries, this bound nearly gives the optimal generalization bound for approx DP algorithms from [BNSSSU16]

Technical Contributions

- **Positive Result**: If $A: D^n \rightarrow T$ is (ϵ, δ)-DP then

$$I^{\beta}_{\infty,\Pi}(A, n) = O\left(n\,\epsilon^2 + n\,\sqrt{\epsilon\delta}\right) \qquad \text{for} \quad \beta = O\left(n\sqrt{\frac{\delta}{\epsilon}}\right)$$

- **Lower bound for non-product distributions**: There is an (ϵ, δ)-DP mechanism A such that for any $\beta \leq \frac{1}{4} - \delta$ we have

$$I^{\beta}_{\infty}(A, n) = n\; - O\left(\log\left(\frac{1}{\delta}\right)\frac{\log(n)}{\epsilon}\right)$$

Consequences of Results

- Max-Info also satisfies strong composition guarantees.
- Pure DP and bounded description length algorithms can be composed in arbitrary ord … eralization.
- Not the case … a product distributi…
- Even if dat … ibution on the data afte… er from a product distribution.
- Ordering matters: important to do DP computations first!

In fact, our lower bound shows that if we do BDL + approx DP, then we can reconstruct the dataset drawn from a product distribution

Thanks!