Algorithmic Privacy

What does algo/data privacy mean?

What does algo/data privacy mean?

- Control of access
- Control of use
- Knowledge of access/use
- Freedom from surveillance
- Ownership of data
- Opt in/out
- Anonymity

Privacy vs. Security

- Security: control access to "raw" data
  - Locks, keys, crypto
- Privacy: allow use of data but control inferences/exfiltration
  - Anonymization, differential privacy

Privacy vs. Security

Name	Age	SSN	Major	GPA	…
Joe	19	547…	CIS	2.7…	
Mary	18	233…	Math	3.5…	
Bill	20	713…	English	3.2…	
Eve	17	492…	History	3.4…	
⋮	⋮	⋮	⋮	⋮	

security: keep locked, control access

↓

$\boxed{\text{Algo}}$

↓

GPA prediction model

privacy: model should not reveal Joe's GPA, Eve's SSN

A Bit of (Old School) Crypto

- Suppose I want to send you a message $a \in \{0,1\}$
- We first meet and choose a random $b \in \{0,1\}$
- Later I send you:

$$c = a \oplus b = \begin{cases} 0 & \text{if } a \neq b \\ 1 & \text{if } a = b \end{cases}$$

- You decrypt:

$$\begin{aligned} c \oplus b &= (a \oplus b) \oplus b \\ &= a \oplus (b \oplus b) = a \end{aligned}$$

A Bit of (Old School) Crypto

- Eavesdropper sees only c which is random
- "One-time pad"
- Why? $(a \oplus b) \oplus (a' \oplus b)$
  $= a \oplus a'$
- Longer pads for longer messages/files

Public-Key Crypto

- OTP security is absolute
- But:
  - Have to meet privately
  - Keys are long
  - Every exchange needs new/different keys
- Public-key crypto:
  - separates en/de-cryption
  - encryption keys public
  - security based on computational hardness
  - underlies modern Internet (e.g. https)

Goldilocks & the 3 Privacys

Anonymization: Not private enough (too useful)

Differential Privacy: Strong privacy & utility

"No Harm Whatsoever": Too private (not useful)

"Anonymization"

- Basic idea: start with some sensitive dataset D, transform to anonymized version D'
- $D \to D'$ by two operations:
  - Redaction: remove entire fields/columns
  - Coarsening: reduce resolution of field

"ANONYMIZED DATA ISN'T"

Name	Age	Gender	Zip Code	Smoker	Diagnosis
*	60-70	Male	191**	Y	Heart disease
*	60-70	Female	191**	N	Arthritis
*	60-70	Male	191**	Y	Lung cancer
*	60-70	Female	191**	N	Crohn's Disease
*	60-70	Male	191**	Y	Lung Cancer
*	50-60	Female	191**	N	HIV
*	50-60	Male	191**	Y	Lyme Disease
*	50-60	Male	191**	Y	Seasonal Allergies
*	50-60	Female	191**	N	Ulcerative Colitis

Name	Age	Gender	Zip Code	Diagnosis
*	50-60	Female	191**	HIV
*	50-60	Female	191**	Lupus
*	50-60	Female	191**	Hip Fracture
*	60-70	Male	191**	Pancreatic Cancer
*	60-70	Male	191**	Ulcerative Colitis
*	60-70	Male	191**	Flu Like Symptoms

A Precise Definition

- Say D' is K-anonymous (w.r.t. certain columns) if for any tuple of values in D', there are at least K copies.
- Privacy by confusion?
- What guarantees does this give you?

Anonymization...

- Is brittle
- Pretends D is the only dataset in the world
- Has no meaningful semantics
- Is demonstrably vulnerable to reidentification attacks

BROKEN PROMISES OF PRIVACY: RESPONDING TO THE SURPRISING FAILURE OF ANONYMIZATION

Paul Ohm*

Computer scientists have recently undermined our faith in the privacy-protecting power of anonymization, the name for techniques that protect the privacy of individuals in large databases by deleting information like names and social security numbers. These scientists have demonstrated that they can often "reidentify" or "deanonymize" individuals hidden in anonymized data with astonishing ease. By understanding this research, we realize we have made a mistake, labored beneath a fundamental misunderstanding, which has assured us much less privacy than we have assumed. This mistake pervades nearly every information privacy law, regulation, and debate, yet regulators and legal scholars have paid it scant attention. We must respond to the surprising failure of anonymization, and this Article provides the tools to do so.

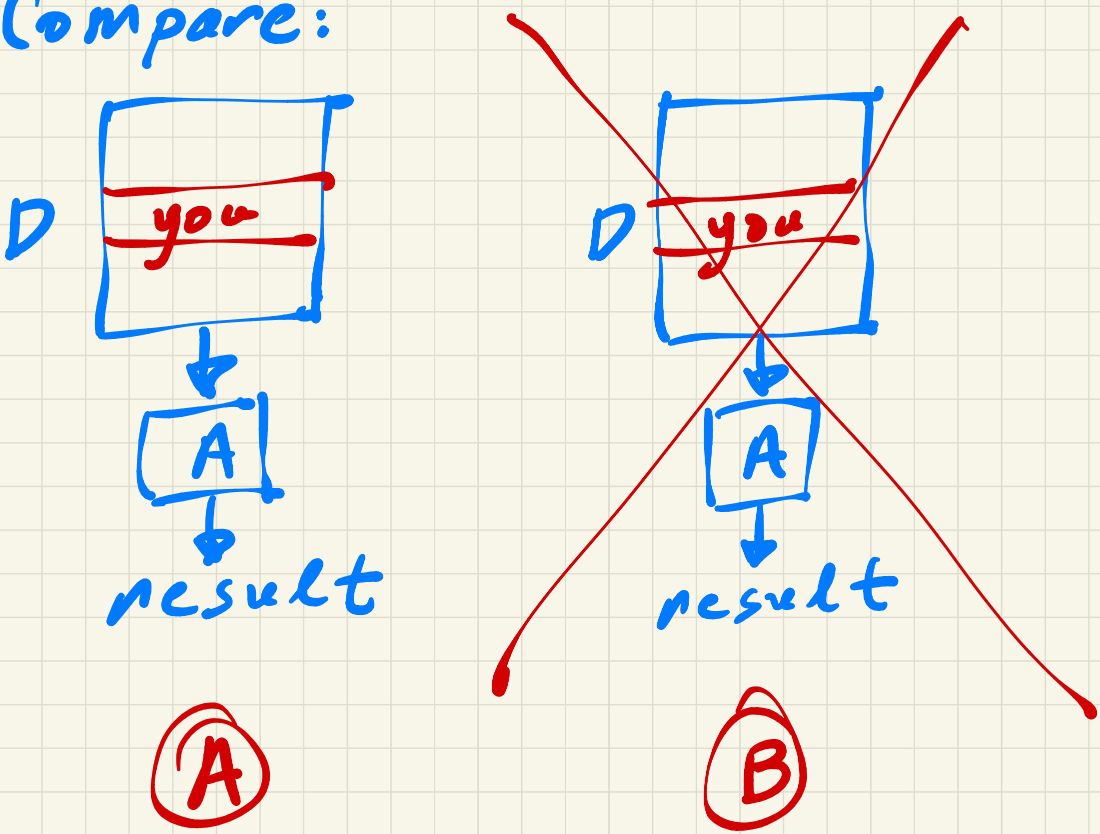
---

* Associate Professor, University of Colorado Law School. This Article was presented at the Privacy Law Scholars Conference and at conferences and faculty workshops at Harvard's Center for Research and Computer Science and Berkman Center, Princeton's Center for Information Technology Policy, Fordham University Center for Law and Information Policy, University of Washington School of Law, University of Washington's Computer Science & Engineering Department, NYU Information Law Institute, DePaul Center for IP Law and Information Technology, International Association of Privacy Professionals Global Privacy Summit, and the University of Colorado Law School. I thank all participants for their comments.

Thanks in particular to Caspar Bowden, Ramon Caceres, Ryan Calo, Deborah Cantrell, Danielle Citron, Nestor Davidson, Pierre de Vries, Vasant Dhar, Cynthia Dwork, Jed Ela, Ed Felten, Victor Fleischer, Susan Freiwald, Brett Frischmann, Michael Froomkin, Simson Garfinkel, Lauren Gelman, Eric Goldman, James Grimmelmann, Mike Hintze, Chris Hoofnagle, Clare Huntington, Jeff Jonas, Jerry Kang, Nancy Kim, Jon Kleinberg, Sarah Krakoff, Tim Lee, William McGeveran, Deven McGraw, Viva Moffat, Tyler Moore, Arvind Narayanan, Helen Nissenbaum, Scott Peppett, Jules Polonetsky, Foster Provost, Joel Reidenberg, Ira Rubinstein, Andrew Schwartz, Ari Schwartz, Vitaly Shmatikov, Chris Soghoian, Dan Solove, Latanya Sweeney, Peter Swire, Salil Vadhan, Michael Waggoner, Phil Weiser, Rebecca Wright, Felix Wu, and Michael Zimmer for their comments. This research was supported by a pre-tenure research leave grant by the University of Colorado Law School, and for this I thank Dean David Getches and Associate Dean Dayna Matthew. Finally, I thank my research assistant, Jerry Green.

Another Try...

- Anonymization: no actual privacy guarantee
- "No harm whatsoever" privacy
- Compare:

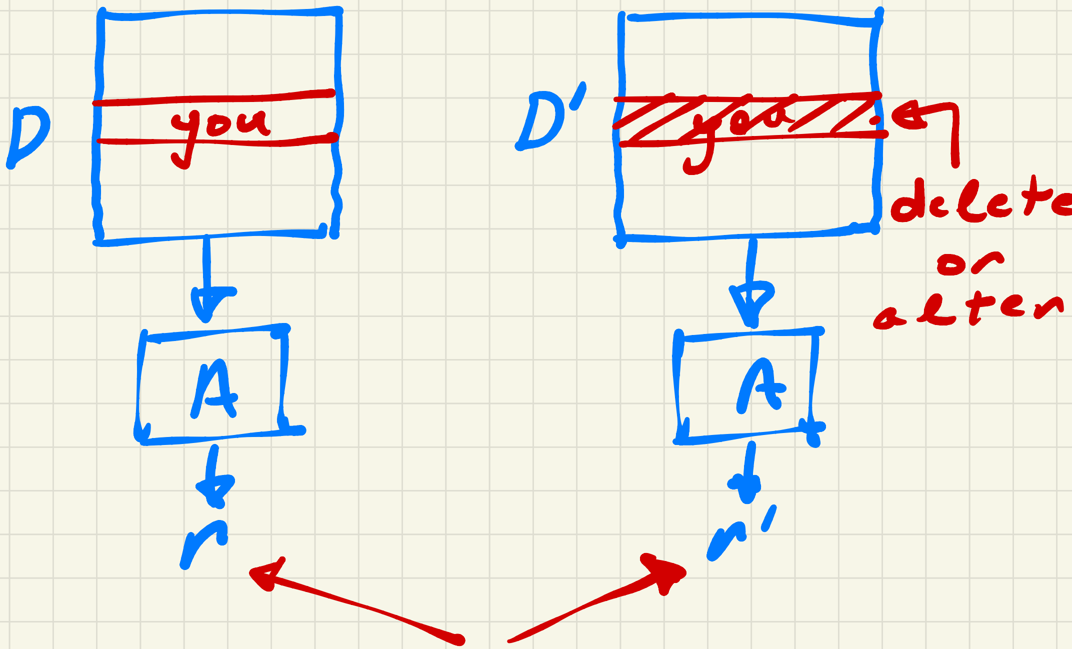


- Chance of (any) harm to you in A & B should be <u>identical</u>
- Good definition?
- Let's consider smoking & lung cancer...

So... What's a better idea?

A More Refined Comparison

- Compare:



r & r' should be

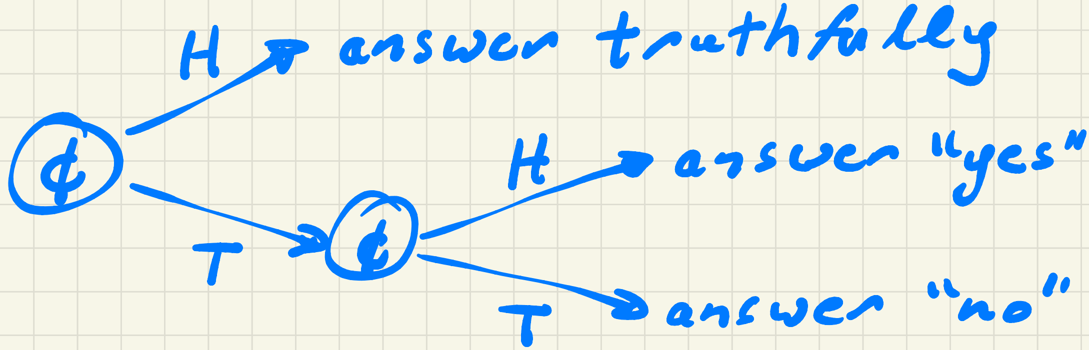
"indistinguishable"

Indistinguishability

- Intuition: Observer seeing only output "can't tell" if input was D or D'
- But if $r = 1.0$ & $r' = 0.999$, can tell!
- Example: average salary
- Going to need randomized algos & probabilistic indistinguishability

Randomized Response (1965)

Have you deliberately violated social distancing?



- Privacy: Plausible deniability
- Utility:

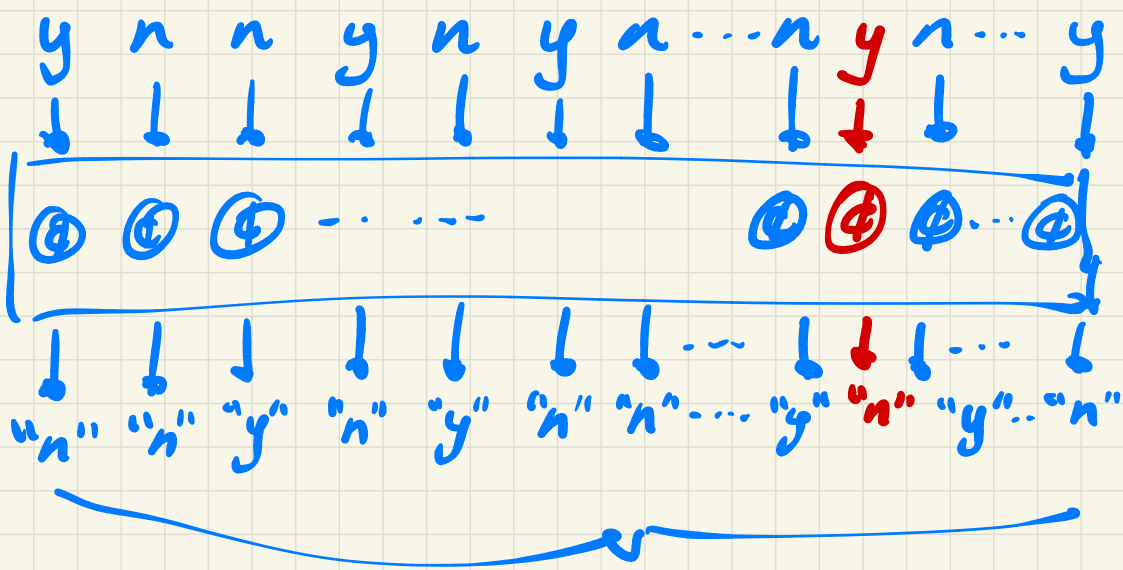
$$\Pr[\text{"yes"} \mid \text{yes}] = 1/2 + 1/2 \cdot 1/2 = 3/4$$

$$\Pr[\text{"yes"} \mid \text{no}] = 0 + 1/2 \cdot 1/2 = 1/4$$

$$\Pr[\text{"yes"}] = p(3/4) + (1-p)(1/4)$$

true fraction of violators

RR as an "algorithm"

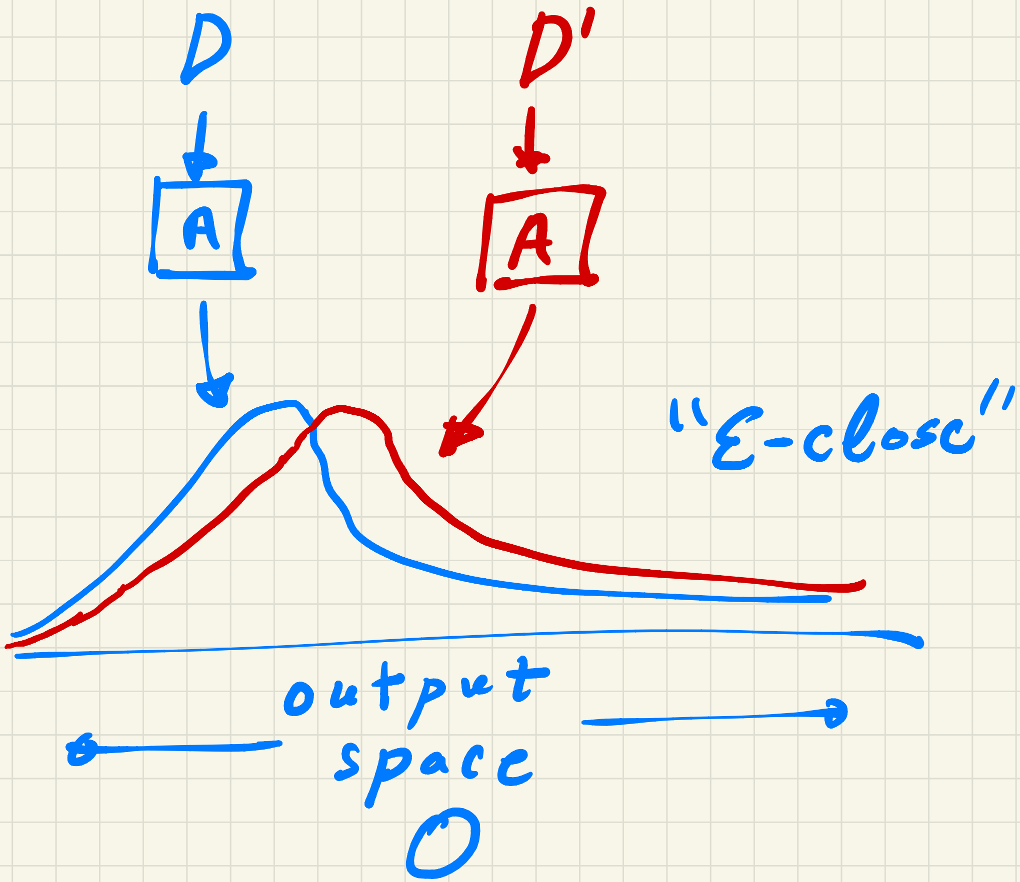


output distribution

<u>Idea:</u> Output distribution should be almost the same if you are y or n.

Differential Privacy

Defn (Randomized) algo A is ε-DP if for any neighboring D & D' if:



Formally: For any subset $S \subseteq O$:

$$\Pr[A(D) \in S] \leq e^{\varepsilon} \Pr[A(D') \in S]$$

$$\Pr[A(D) \in S] \geq \frac{1}{e^{\varepsilon}} \Pr[A(D') \in S]$$

- $\varepsilon \to 0$: perfect privacy
- $\varepsilon \to \infty$: no privacy
- S as your "privacy fear"
- **Important**: Does not promise $A(D), A(D')$ small, just that they are close
- Property of algo, not data!
- Contrast anonymization

DP analysis of RR

- Let $o_i \in \{\text{"yes"}, \text{"no"}\}$ be output of participant i
- **Note:** for any inputs (y/n),

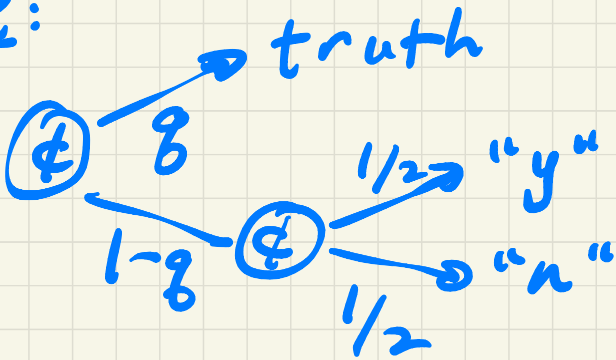
$$\Pr[o_1 o_2 o_3 \cdots o_n] = \Pr[o_1]\Pr[o_2]\cdots\Pr[o_n]$$

(independence)

so we can just analyze you.

- Need to relate $\Pr[\text{"yes"} \mid \text{yes}]$ & $\Pr[\text{"yes"} \mid \text{no}]$
- But $\dfrac{\Pr[\text{"yes"} \mid \text{yes}]}{\Pr[\text{"yes"} \mid \text{no}]} = \dfrac{3/4}{1/4} = 3.$
- $e^{\varepsilon} = 3$, $\varepsilon = \ln(3) \doteq 1.1$

q-RR:



Then:

$$\Pr[\text{"y"} \mid y] = q + (1-q)/2$$
$$= (1+q)/2$$

$$\Pr[\text{"y"} \mid n] = 0 + (1-q)/2$$
$$= (1-q)/2$$

$$\text{Ratio} = \frac{1+q}{1-q} = e^{\varepsilon}$$

$$\boxed{\varepsilon = \ln\left(\frac{1+q}{1-q}\right)}$$

General Tools for Differential Privacy: The Laplace Mechanism

The Laplace Mechanism

- Consider data $x \in [0,1]^n$
- So i's data is number x_i
- Want to (DP) compute some function:

$$f(x) = f(x_1, x_2, \ldots, x_n) \in \mathbb{R}$$

- E.g. average, median, std, max, $x_1 x_3^2 + 10 x_7^3 + \cdots$
- Want a general way of DP-computing $f(x)$ accurately

The Laplace Mechanism

- Compute $f(x)$ exactly then "add noise"
- What kind? How much?
- How much? Define sensitivity of $f(x)$:

$$\Delta f \triangleq \max_{\substack{\text{neighboring} \\ x, x' \in [0,1]^n}} \{ |f(x) - f(x')| \}$$

The Laplace Mechanism

Examples:

- $f = \text{average}$, $\Delta f = 1/n$

  $(111\cdots1 \to 111\cdots10)$

- $f = \text{std} = \sqrt{\frac{1}{n}\sum_i (x_i - \mu)^2}$

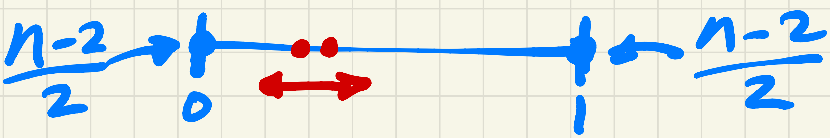
  $\Delta f \approx 1/\sqrt{n}$

- $f(x) = \max(x_1, \dots x_n)$, $\Delta f = 1$

  $(00\cdots0 \to 00\cdots01)$

- $f(x) = x_1 \cdot x_2 \cdots x_n$, $\Delta f = 1$

- $f(x) = \text{median}(x_1, \dots x_n) = 1$



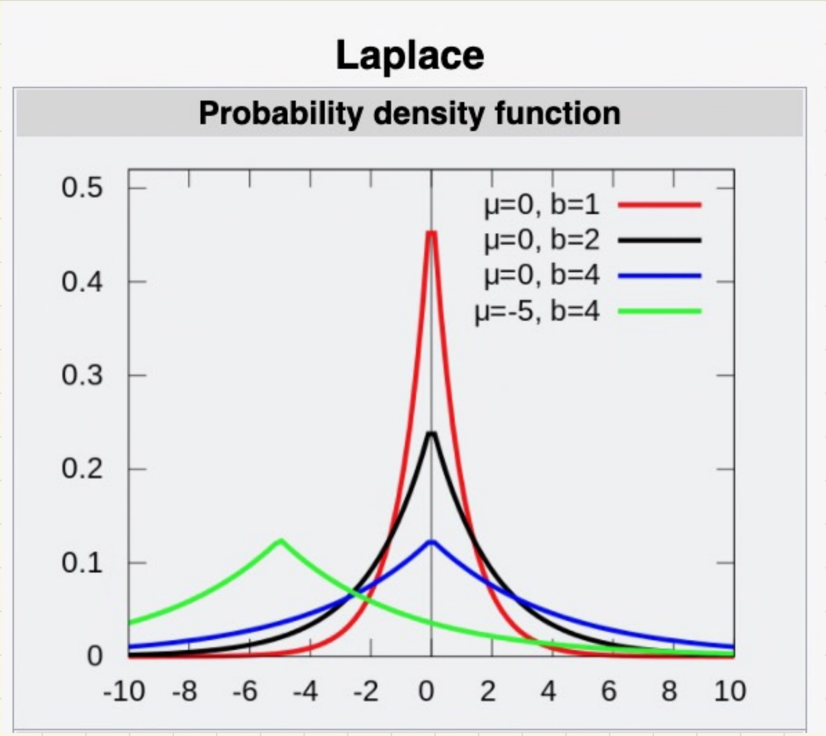
The Laplace Mechanism

- Laplace distribution:
  - randomly choose value v with probability:

$$\frac{1}{2b} e^{-|v|/b}$$

  - b is a parameter
  - larger values of $|v|$ are exponentially less probable

The Laplace Mechanism



- mean = 0, variance = 2b
- larger b: more noise
  smaller b: less noise

The Laplace Mechanism

- **Finally:**
  - compute $f(x)$
  - output $f(x)+v$
    where v is Laplace
    with $b = \Delta f / \varepsilon$

sensitivity of f

for ε-DP

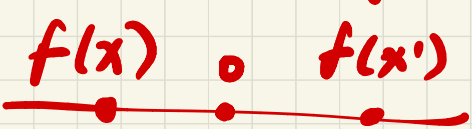
Theorem: Laplace mech. satisfies ε-DP.

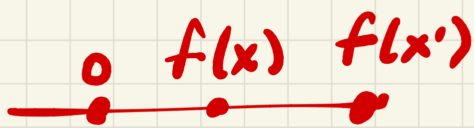
Proof: Let $x, x' \in [0,1]^n$ be neighbors with LM distributions $p_x, p_{x'}$.

For any output value o:

$$\frac{p_x(o)}{p_{x'}(o)} = \frac{\left(\frac{1}{2b} e^{-|f(x)-o|/b}\right)}{\left(\frac{1}{2b} e^{-|f(x')-o|/b}\right)}$$

$$= e^{\left(\underbrace{|f(x')-o| - |f(x)-o|}_{\le |f(x')-f(x)|}\right)/b}$$





$$\leq e^{|f(x') - f(x)|/b}$$

$$\leq e^{\Delta f/b}$$

$$= e^{\Delta f/(\Delta f/\varepsilon)} = e^{\varepsilon}. \checkmark$$

If $S = \{o_1, o_2, \ldots o_\ell\}$ is a set:

$$\frac{p_x(S)}{p_{x'}(S)} = \frac{\sum_{i=1}^{\ell} p_x(o_i)}{\sum_{i=1}^{\ell} p_{x'}(o_i)} \leq \frac{\sum_{i=1}^{\ell} e^{\varepsilon} p_{x'}(o_i)}{\sum_{i=1}^{\ell} p_{x'}(o_i)}$$

$$\leq e^{\varepsilon}.$$

LM is also useful

- E.g. if $\Delta f = 1/n$ (average)
  then $b = \Delta f/\varepsilon = 1/\varepsilon n$.
  $\to 0$ as $n \to \infty$ for
  fixed ε.
- $\Delta f = 1/\sqrt{n}$, $b = \frac{1}{\varepsilon \sqrt{n}}$
- General: any f s.t.
  $\Delta f \to 0$ as $n \to \infty$

General Tools for DP II: The Exponential Mechanism

Complex Inputs/Outputs

- Laplace: numbers → number
- What about:

  dataset → decision tree, neural net, ...

  social network → clustering

  votes → chosen/ideal outcome

  auction bids → winners & prices

Can't just "add noise" to output!

The Setting

- Input space I
- Output space O
- For $x \in I$, $o \in O$, notion of quality or utility of o:

$$u(x,o) \in \mathbb{R}$$

- Examples:
  - x = dataset, o = neural net, u = error of o on x
  - x = votes, o = outcome, u = agreement of o on x
  - x = bids, o = winners, u = social welfare

The Setting

- Absent privacy goal is

$$o^* = \underset{o \in \mathcal{O}}{\operatorname{argmax}} \{ u(x, o) \}$$

i.e. pick best output.

- But this won't be DP!

- Example: exfiltrating training data in ML.

- What can we do?

Generalized Sensitivity

- Let's define

$$\Delta u = \max_{\substack{x, x' \in I \\ \text{neighbors}}} \max_{o \in O} \left\{ |u(x,o) - u(x',o)| \right\}$$

- How much can changing a single input change the quality of some output?

- E.g. ML: $\Delta u = 1/n$

- E.g. average of $x_1, \dots x_n$: $\Delta u = 1/n$

- E.g. winning auction price: Δu large

The Exponential Mechanism

- On input x, output each $o \in O$ with probability:

$$p_x(o) = \frac{e^{\varepsilon u(x,o)/2\Delta u}}{Z(x)}$$

where $Z(x) = \sum_{o \in O} e^{\varepsilon \cdot u(x,o)/2\Delta u}$

- Every $o \in O$ might be output, but better o more likely

Exp. Mechanism is ε-DP

Proof: $\forall$ nbrs $x, x' \in \mathcal{I}$, $o \in \mathcal{O}$:

$$\frac{p_x(o)}{p_{x'}(o)} = \frac{e^{\varepsilon \cdot u(x,o)/2\Delta u} / Z(x)}{e^{\varepsilon \cdot u(x',o)/2\Delta u} / Z(x')}$$

$$= e^{\varepsilon(u(x',o) - u(x,o))/2\Delta u} \cdot \left(\frac{Z(x')}{Z(x)} \right)$$

$$\leq e^{\varepsilon|u(x',o) - u(x,o)|/2\Delta u} \cdot \left(\downarrow \right)$$

$$\leq e^{\varepsilon \Delta u / 2\Delta u} \cdot \left(\quad \right)$$

$$= e^{\varepsilon/2} \cdot \left(\quad \right)$$

Now:

$$\frac{Z(x')}{Z(x)} = \frac{\sum_o e^{\varepsilon \cdot u(x',o)/2\Delta u}}{\sum_o e^{\varepsilon \cdot u(x,o)/2\Delta u}}$$

$$\leq \frac{\sum_o e^{\varepsilon(u(x,o)+\Delta u)/2\Delta u}}{\sum_o e^{\varepsilon u(x,o)/2\Delta u}}$$

$$= e^{\varepsilon \Delta u/2\Delta u} \frac{\sum_o e^{\varepsilon u(x,o)/2\Delta u}}{\sum_o e^{\varepsilon u(x,o)/2\Delta u}}$$

$$= e^{\varepsilon/2}, \text{ and}$$

$$e^{\varepsilon/2} \cdot e^{\varepsilon/2} = e^{\varepsilon}.$$

Utility of Exp. Mech.

- Special case: Θ finite
- Fix x, let $EM(x)$ denote Exp. Mech.
- Then with high probability:

$$u(x, \theta^*) - u(x, EM(x))$$

$$\lesssim \frac{2\Delta u}{\varepsilon} \ln |\Theta|.$$

- E.g. ML: $\Delta u = 1/n$, $|\Theta| = 2^d$,

set $\frac{2d}{\varepsilon n} \ll 1 \Rightarrow n \gg \frac{2d}{\varepsilon}$.

So I, O and $u(x,0)$ can be anything and we can be DP!

What's the catch?

DP Immunity to Post-Processing

$$x \to \boxed{A} \xrightarrow{A(x)} \boxed{B} \to y \in \mathcal{O}'$$

ε-DP   arbitrary algo

Then $B(A(x))$ also ε-DP.

Proof: $\forall$ nbrs x, x', $\forall o' \in \mathcal{O}'$,

let $T = \{o \in \mathcal{O} : B(o) = o'\}$

Then:

$$\Pr[B(A(x)) = o'] = \Pr[A(x) \in T]$$
$$\le e^{\varepsilon} \Pr[A(x') \in T]$$
$$= e^{\varepsilon} \Pr[B(A(x')) = o']$$

Special case:

$$\mathcal{O}' = \begin{array}{l}\text{"input was } x\text{"} \\ \text{"input was } x'\text{"}\end{array}$$

Then for small ε, have:

$\Pr[B(A(x)) = \text{"}x\text{"}]$ (right)

$\approx \Pr[B(A(x')) = \text{"}x\text{"}]$ (wrong)

$= 1 - \Pr[B(A(x')) = \text{"}x'\text{"}]$

Or:

$\Pr[B(A(x)) = \text{"}x\text{"}]$

$+ \Pr[B(A(x')) = \text{"}x'\text{"}] \approx 1.$

"Not (much) better than random guessing"

Composition Properties of DP

Parallel DP Algos

- Suppose A_1 is ε_1-DP & A_2 is ε_2-DP
- $A(x) = (A_1(x), A_2(x))$
- Then A is $(\varepsilon_1+\varepsilon_2)$-DP
- Proof: $p_x^A(o_1, o_2)$

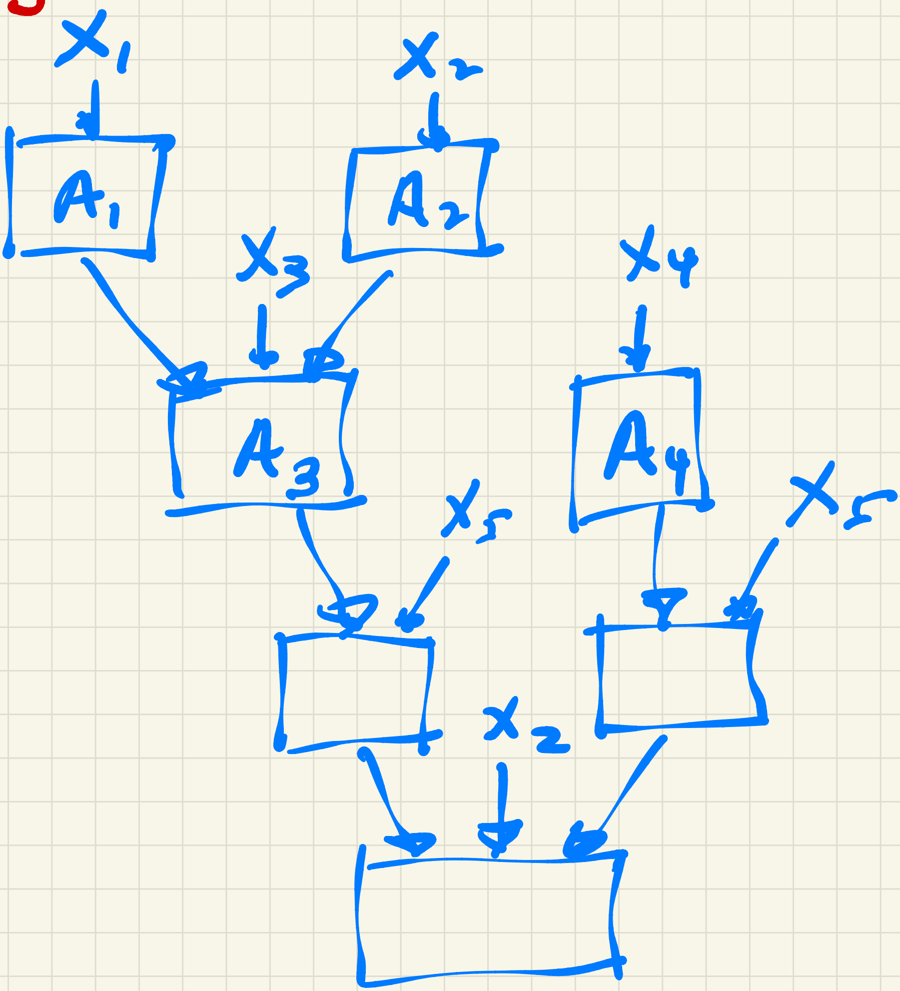
$$= p_x^{A_1}(o_1)\, p_x^{A_2}(o_2) \quad \text{indep.}$$

$$\le e^{\varepsilon_1} p_{x'}^{A_1}(o_1)\, e^{\varepsilon_1} p_{x'}^{A_2}(o_2)$$

$$= e^{\varepsilon_1+\varepsilon_2} p_{x'}^{A}(o_1, o_2). \checkmark$$

General Composition

- Suppose your data is in **many** datasets $X_1, X_2, X_3, \ldots$



etc.

"Theorem": If there are K datasets & algos, each ε-DP, then composition is $K \cdot \varepsilon$-DP.

Has led to development of an algo toolkit for DP.

Applied Case Study:

DP Synthetic Data Generation

The Goal

$$\begin{matrix} D \longrightarrow \\ \varepsilon \longrightarrow \end{matrix} \boxed{\;A\;} \longrightarrow D'$$

$\leftarrow \varepsilon$-DP

- D' is a dataset that is DP but still "looks like" D
- Like anonymization but <u>better</u>

"Looks Like D"

- D′ approx. preserves statistical props of D′
- Conditional queries:
  - e.g. fraction of rows s.t. age $\geq$ 25 & gender = F & job = VP
  - e.g. avg. salary of same
- Generally low sensitivity ($\sim 1/n$)
- $\forall$ query q, want $|q(D) - q(D')|$ small

How?

- Generally hard
- Could use Exp. Mechanism, but...

Instead:

- Add Laplace noise to the $q(D) \to q'(D)$
- Treat entries of D' as variables
- Use gradient descent to reduce $\max_q |q'(D) - q(D')|$

Algorithm 2 Relaxed Adaptive Projection (RAP)

Input: A dataset D, a collection of m statistical queries Q, a "queries per round" parameter $K \leq m$, a "number of iterations" parameter $T \leq m/K$, a synthetic dataset size n', and differential privacy parameters ϵ, δ.

Let ρ be such that:

$$\epsilon = \rho + 2\sqrt{\rho \log(1/\delta)}$$

if $T = 1$ **then**
  for $i = 1$ to m **do**
    Let $\hat{a}_i = G(D, q_i, \rho/m)$.
  end for
  Randomly initialize $D' \in (\mathcal{X}^r)^{n'}$.
  Output $D' = RP(q, \hat{a}, D')$.
else
  Let $Q_S = \emptyset$ and $D'_0 \in (\mathcal{X}^r)^{n'}$ be an arbitrary initialization.
  for $t = 1$ to T **do**
    for $k = 1$ to K **do**
      Define $\hat{q}^{Q \setminus Q_S}(x) = (\hat{q}_i(x) : q_i \in Q \setminus Q_S)$ where $\hat{q}_i$ is an equivalent extended differentiable query for q_i.
      Let $q_i = RNM(D, q^{Q \setminus Q_S}, q^{Q \setminus Q_S}(D'_{t-1}), \frac{\rho}{2T \cdot K})$.
      Let $Q_S = Q_S \cup \{q_i\}$.
      Let $\hat{a}_i = G(D, q_i, \frac{\rho}{2T \cdot K})$.
    end for
    Define $q^{Q_S}(x) = (q_i(x) : q_i \in Q_S)$ and $\hat{a} = \{\hat{a}_i : q_i \in Q_S\}$ where $\hat{q}_i$ is an equivalent extended differentiable query for q_i. Let $D'_t = RP(q^{Q_S}, \hat{a}, D'_{t-1})$.
  end for
  Output D'_T.
end if

```
import jax.numpy as np
def threeway_marginals(D):
    return np.einsum('ij,ik,il->jkl', D, D, D)/D.shape[0]
```

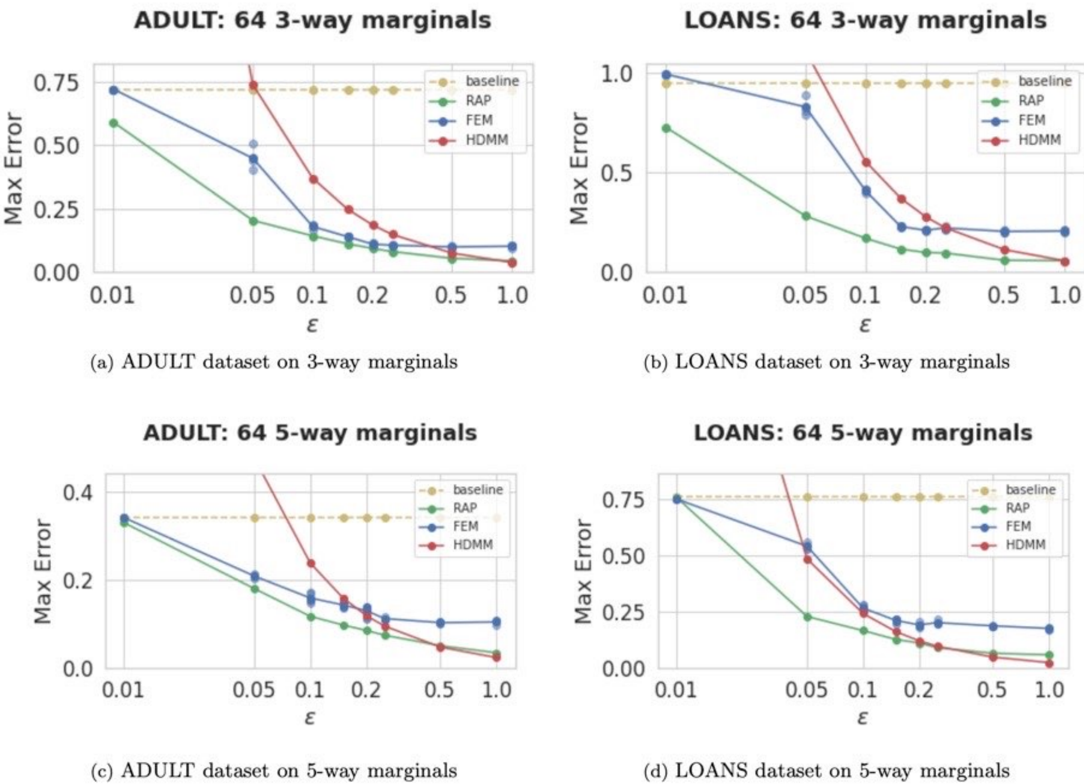
Figure 1: Python function used to compute 3-way product queries

Dataset	Records	Features	Transformed Binary Features
ADULT	48842	15	588
LOANS	42535	48	4427

Table 1: Datasets. Each dataset starts with the given number of original (categorical and real valued) features. After our transformation, it is encoded as a dataset with a larger number of binary features.

Parameter	Description	Values
K	Queries per round	5 10 25 50 100
T	Number of iterations	2 5 10 25 50

Table 2: RAP hyperparameters tested in our experiments



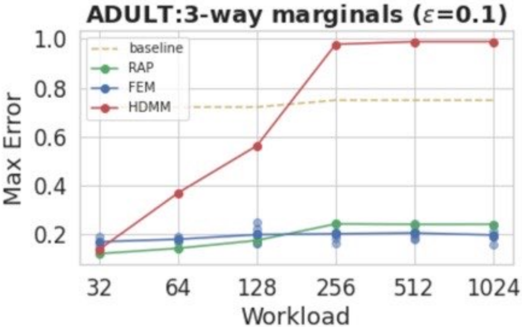
(a) ADULT dataset on 3-way marginals

(b) LOANS dataset on 3-way marginals

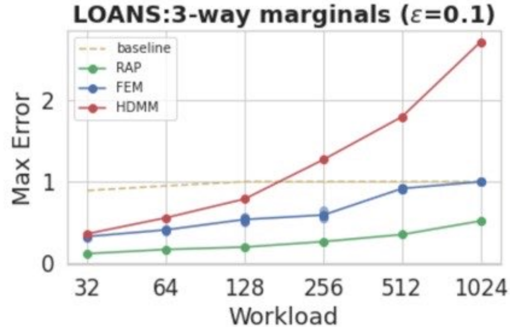
(c) ADULT dataset on 5-way marginals

(d) LOANS dataset on 5-way marginals

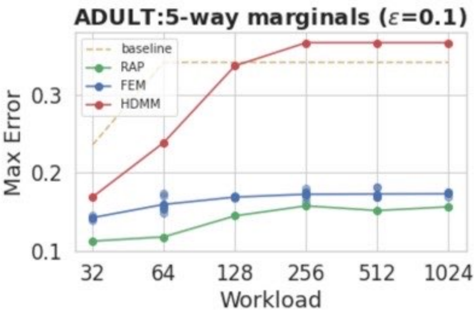
Figure 2: Max-error for 3 and 5-way marginal queries on different privacy levels. The number of marginals is fixed at 64.



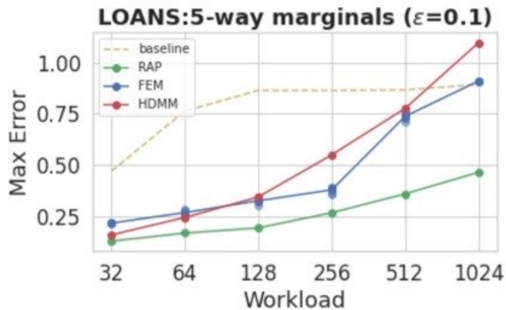
(a) ADULT dataset on 3-way marginals



(b) LOANS dataset on 3-way marginals



(c) ADULT dataset on 5-way marginals



(d) LOANS dataset on 5-way marginals

Figure 3: Max error for increasing number of 3 and 5-way marginal queries with $\epsilon = 0.1$