Privacy: Differential Privacy

Eric Nguyen, Shiyu Feng, Daniel Slyepichev, Sabrina Lopez, Uttam Rao

University of Virginia

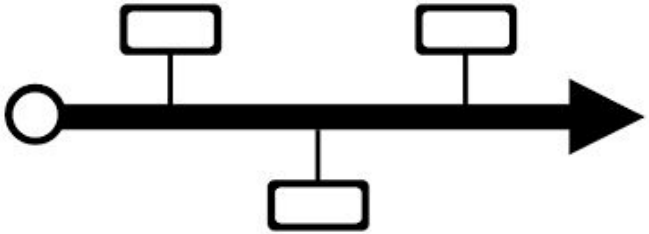
An Introduction

Guess Who?

- Choose an individual to play
- We'll ask Yes/No questions
- What is this metaphor?

Timeline

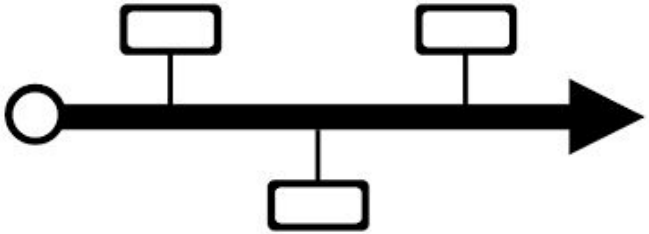
- What is privacy
- Ways privacy can / have been disturbed
- How to protect privacy
- How to implement the privacy protection
- Discussion and Q&A
- How these all come together



University of Virginia

Timeline

- What is privacy
- Ways privacy can / have been disturbed
- How to protect privacy
- How to implement the privacy protection
- Discussion and Q&A



University of Virginia

Introduction to Privacy

What is Privacy?

- Data is collected nearly everywhere
- Result: much of the data contains sensitive personal information ⇒ privacy concerns
- Efforts for privacy include GDPR, Title XIII, Biden's Executive Order on AI
- **Privacy** – the right to be free from being observed or disturbed by other people
- **Information privacy** – the right to have some control over how your personal information is collected and used

A Strong Privacy Definition

- **Compositionality**
  - Privacy protections should degrade in a controlled manner when applied multiple times to prevent cumulative privacy loss
- **Post-processing immunity**
  - Once data is privatized, further analysis should not weaken its privacy guarantees, ensuring future-proof protection
- **Group privacy**
  - Privacy mechanisms should account for groups, ensuring controlled and quantifiable degradation as group size increases
- **Quantifiable privacy-accuracy tradeoffs**
  - Privacy definitions should provide measurable trade-offs between privacy protection and data accuracy to enable informed decision-making

Motivation

- Key Question: What are some motivations for protecting users' privacy?
- Some examples
  - Legal and regulatory compliance
  - Reduce cybersecurity risks
  - Ethical research and data sharing
  - Identity theft and fraud
  - Reducing unwanted marketing or surveillance
  - etc.

University of Virginia

Anonymization

Case Study: Introduction

- **The Netflix Prize Competition (2006 to 2009)**
- Aim: Challenge researchers to improve their recommendation system
- Competition material: training dataset of user data
  - User ID (anonymized)
  - Movie ID
  - Rating
  - Date
- Anonymization – removal or masking of identifying data ⇒ prevent recovery of personal identities

University of Virginia

Case Study: Outcome

- Researchers **cross-referenced the dataset with public information** from online movie database IMDb
  - → matched users between the datasets based on similar movie ratings at similar times
  - IMDb data not de-identified → **re-identification** ⚠️
- Outcome of discovery: class action lawsuit vs Netflix

University of Virginia

Anonymization: Data Anonymization

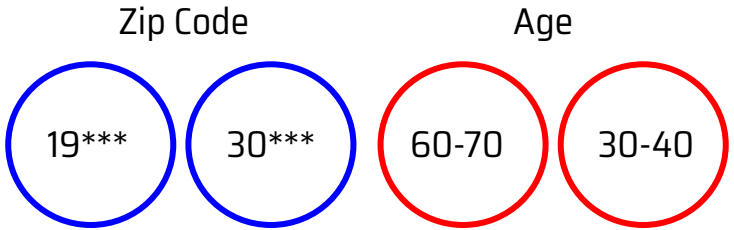
- **Type 1: Data anonymization**
  - E.g., Netflix case study
  - Why does it fail?
    - Lack of formal privacy guarantees
      - Post-processing immunity

Anonymization: K-Anonymization

- **Type 2: K-anonymization**
  - For each record in the dataset, there are least **k-1** other records with identical values based on certain quasi-identifiers
    - ⇒ Generalizing or suppressing identifiable attributes
  - How it works
    - Groups individuals who share the same quasi-identifiers into “anonymized” **k** clusters

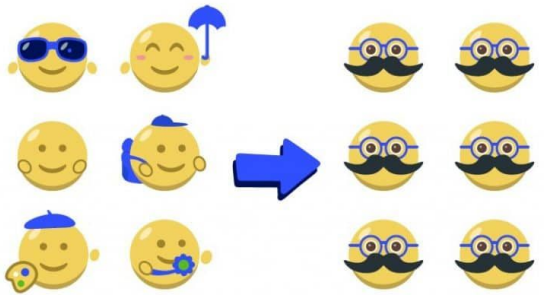
4-anonymized Data

ID	Zip Code	Age
1	19***	60-70
2	30***	30-40
3	19***	60-70
4	30***	30-40
5	19***	60-70
6	30***	30-40
7	30***	30-40
8	19***	60-70



Anonymization: K-Anonymization

- **Type 2: K-anonymization**
  - Why does it fail?
    - <u>Lack of group privacy</u>
      - Protects individuals, not groups
      - E.g., group of same characteristics (e.g., age, zip code)
    - <u>Lack of composition</u>
      - Combination of datasets can ⇒ re-identification
      - E.g., dataset with employee's ages and dataset of credit scores, both with same age and zip code attributes



Database Reconstruction Attacks

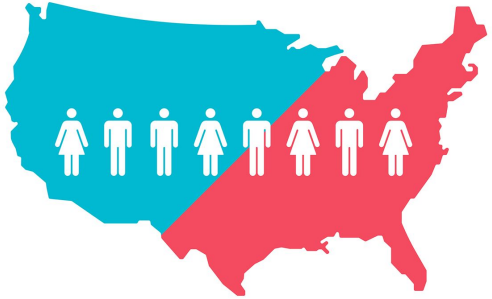
Database Reconstruction Attack (DRA)

- Concept
  - Attackers use published summary statistics to **reconstruct individual records**.
  - Even **without direct access to raw data**, mathematical constraints allow attackers to **infer missing information**.
- Theoretical Fundamental
  - Every statistic leaks some private information (Dinur & Nissim, 2003)

How DRA Works

- **Step 1:** Extract constraints from published data (population, median age, racial distribution, etc.)
- **Step 2:** Translate constraints into mathematical equations (e.g., summation rules)
- **Step 3:** Use mathematical solvers (SAT solvers, optimization models) to infer possible individual record
- **Step 4:** Reconstruct original records partially or fully

An Example of DRA



Summary Statistics

Median age, mean age, racial breakdown, material status distribution

Can attackers figure out each person's age, gender and race from these "harmless" statistics?

Yes!

University of Virginia

Public Data

TABLE 1: **FICTIONAL STATISTICAL DATA FOR A FICTIONAL BLOCK**

		AGE		
STATISTIC	**GROUP**	**COUNT**	**MEDIAN**	**MEAN**
1A	total population	7	30	38
2A	female	4	30	33.5
2B	male	3	30	44
2C	black or African American	4	51	48.5
2D	white	3	24	24
3A	single adults	(D)	(D)	(D)
3B	married adults	4	51	54
4A	black or African American female	3	36	36.7
4B	black or African American male	(D)	(D)	(D)
4C	white male	(D)	(D)	(D)
4D	white female	(D)	(D)	(D)
5A	persons under 5 years	(D)	(D)	(D)
5B	persons under 18 years	(D)	(D)	(D)
5C	persons 64 years or over	(D)	(D)	(D)

Note: Married persons must be 15 or over

Translate constraints into Math equations

TABLE 2: **POSSIBLE AGES FOR A MEDIAN OF 30 AND MEAN OF 44**

A	B	C	A	B	C	A	B	C
1	30	101	11	30	91	21	30	81
2	30	100	12	30	90	22	30	80
3	30	99	13	30	89	23	30	79
4	30	98	14	30	88	24	30	78
5	30	97	15	30	87	25	30	77
6	30	96	16	30	86	26	30	76
7	30	95	17	30	85	27	30	75
8	30	94	18	30	84	28	30	74
9	30	93	19	30	83	29	30	73
10	30	92	20	30	82	30	30	72

Each statistic acts as a **constraint**, narrowing down the possible data combination.

- A constraint is a rule that limits possible values.
- Possible combination reduced to 30.
- Combining gender and racial constraints, possible individual records are inferred.

SAT (Satisfiability) Solver

TABLE 3: **VARIABLES ASSOCIATED WITH THE RECONSTRUCTION ATTACK.**

PERSON	AGE	SEX	RACE	MARITAL STATUS
1	A1	S1	R1	M1
2	A2	S2	R2	M2
3	A3	S3	R3	M3
4	A4	S4	R4	M4
5	A5	S5	R5	M5
6	A6	S6	R6	M6
7	A7	S7	R7	M7
KEY				
female		0		
male		1		
black or African American			0	
white			1	
single				0
married				1

```
; First define the integer variables,
; with the range 0..125
(int A1 0 125)
(int A2 0 125)
(int A3 0 125)
(int A4 0 125)
(int A5 0 125)
(int A6 0 125)
(int A7 0 125)

; Statistic 1A: Mean age is 38
(= (+ A1 A2 A3 A4 A5 A6 A7)
    (* 7 38)
)
```

UNIVERSITY of VIRGINIA

Reconstruct the original data

TABLE 4: **A SINGLE SATISFYING ASSIGNMENT**

AGE	SEX	RACE	MARITAL STATUS
8	F	B	S
18	M	W	S
24	F	W	S
30	M	W	M
36	F	B	M
66	F	B	M
84	M	B	M

Even a small dataset with a few public statistics can be nearly fully reconstructed.

Why DRA is dangerous

“Sober Warnings” from Dick et al.

- An attacker **need not reconstruct the entire underlying database** in its internal schema for data subject confidentiality to be at risk.
- Modern optimization techniques increase DRA effectiveness.
- Large-scale, nonconvex optimization techniques can exfiltrate entire rows of sensitive data with confidence.

UNIVERSITY of VIRGINIA

Statistical Disclosure Limitation (SDL)

- Definition: the process of treating confidential data to protect the identity and responses of data subjects' information in the published data.
- Common SDL techniques
  - **Cell suppression**: Hide small counts to prevent direct identification
  - **Top-coding**: Limit the maximum value for certain variable to reduce uniqueness
  - **Input Noise-injection**: Add random variations to prevent reconstruction
  - **Swapping**: Exchange attributes between similar records

Limitations of traditional SDL methods

- **Problem 1:** SDL relies on obfuscation, not mathematical privacy guarantees.
- **Problem 2:** Modern DRA attacks bypass these methods easily.
- **Problem 3:** SDL reduces data utility without ensuring privacy.
- Aggregate statistics with high precision are essentially unprotected microdata.

Case study: 2020 U.S. Census

- Scale: 330 million people across 8.5 million blocks, extensive statistical release
- Law: individual responses must remain confidential (Title 13, U.S. Code).
- Issue: DRA could reveal private information with summary statistics.
- Challenges:
  - Census Bureau previously relied on SDL methods (e.g., cell suppression)
  - Modern optimization + high-precision statistics → SDL protection ineffective

Discussion

- What are ways that we can protect (informational) privacy?
  - What are the benefits and or downsides of your method(s) to protect privacy?

- Is it possible to define a stronger definition of privacy beyond the four properties?
  - Properties: compositionality, post-processing immunity, group privacy, quantifiable privacy-accuracy tradeoffs

Differential Privacy (DP) as a solution

- Adding controlled noise ensure that individual data points do not significantly impact published statistics.
- Key feature:
  - Privacy Budget: controls the trade-off between privacy and accuracy
  - Global noise injection: prevents pattern recognition
- Impact:
  - Stronger privacy protection against DRA
  - Some researchers argued that data accuracy suffered

Differential Privacy: Foundation

Beyond the Census

UNIVERSITY of VIRGINIA

What is Data Privacy, loosely?

Data privacy techniques have the goal of allowing analysts to learn about ***trends*** in sensitive data, without revealing information specific to ***individuals***.

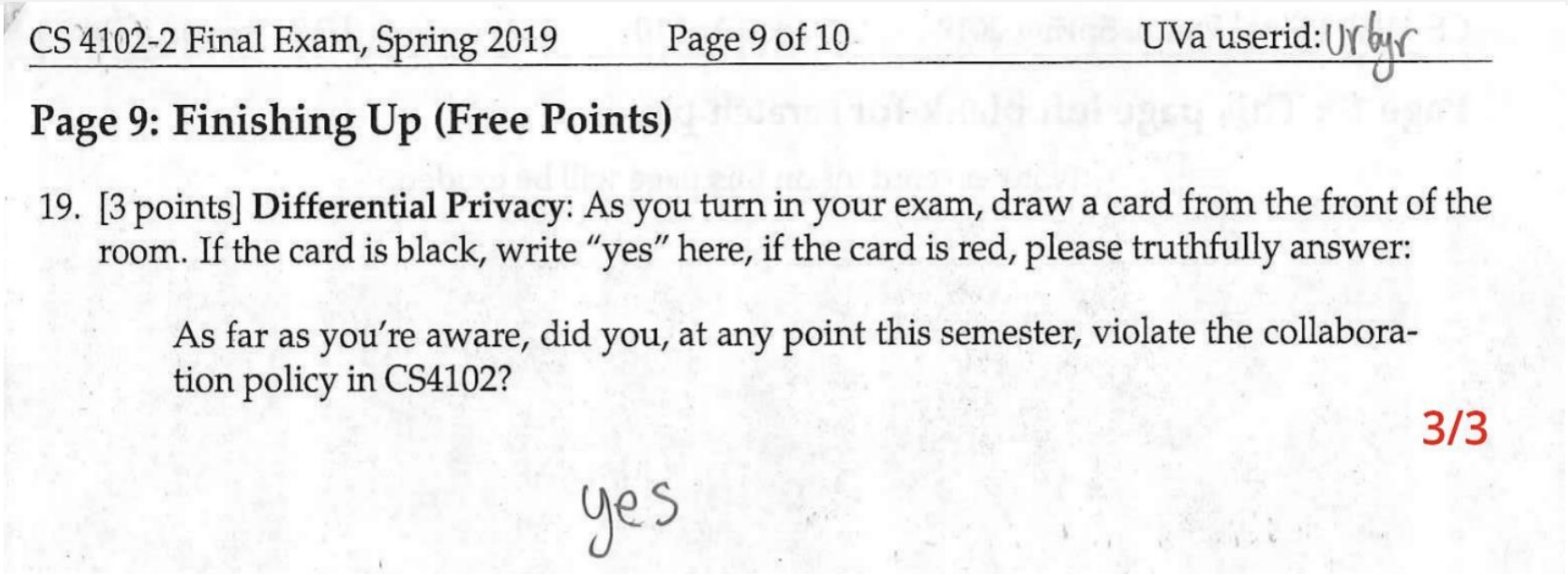
Randomized Response

Did you cheat in this course?

1. Flip a coin
2. If ***tails*** - respond truthfully
3. If ***heads*** - flip a coin again
    a. If ***heads*** - respond “Yes”
    b. If ***tails*** - respond “No”

Plausible deniability:
A “Yes” answer is not incriminating since it occurs with a probability of ¼ regardless of whether or not the student cheated.

Real Example



CS 4102-2 Final Exam, Spring 2019 Page 9 of 10 UVa userid: urbyr

Page 9: Finishing Up (Free Points)

19. [3 points] **Differential Privacy**: As you turn in your exam, draw a card from the front of the room. If the card is black, write "yes" here, if the card is red, please truthfully answer:

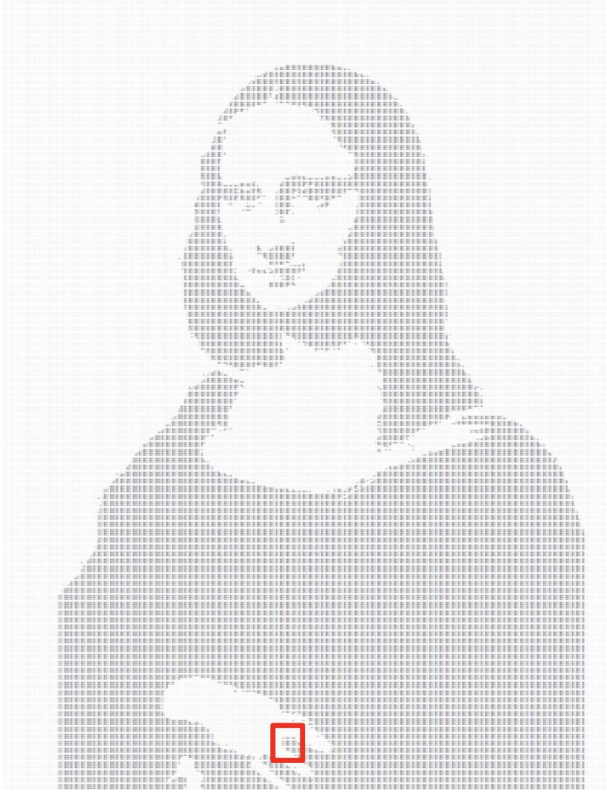
As far as you're aware, did you, at any point this semester, violate the collaboration policy in CS4102?

3/3

yes

How is this different from the scenario on the last slide with coin tosses? Is it better, worse, or the same at creating plausible deniability?

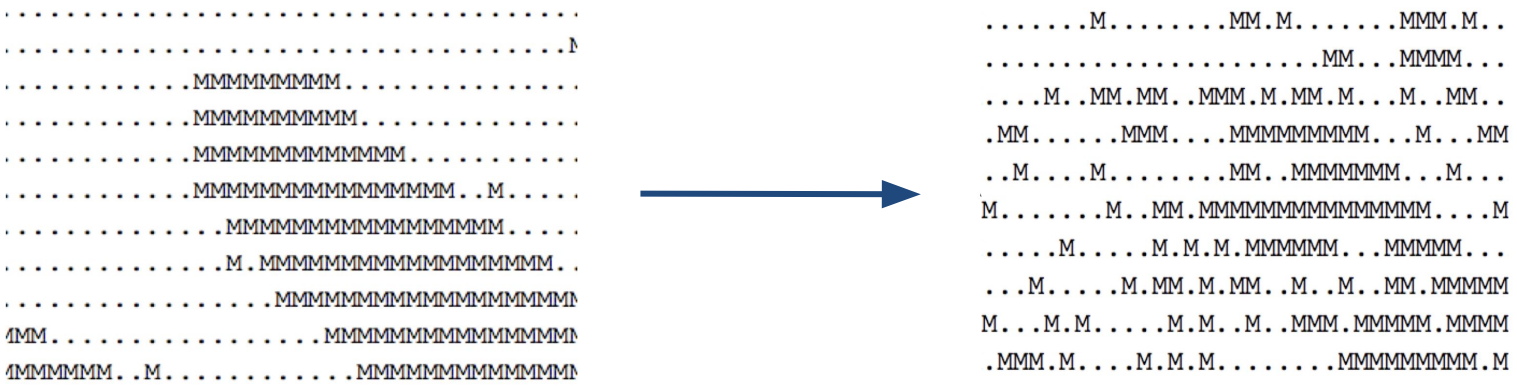
Illustrative Example



```
.....
.....M
.....MMMMMMMMM.....
.....MMMMMMMMMM.....
.....MMMMMMMMMMMMM.....
.....MMMMMMMMMMMMMMMM..M.....
.....MMMMMMMMMMMMMMMMM.....
.....M.MMMMMMMMMMMMMMMMM..
.....MMMMMMMMMMMMMMMMMMM
MM.....MMMMMMMMMMMMMMMM
MMMMMM..M.....MMMMMMMMMMMMM
```

University of Virginia

Illustrative Example



Flip each bit with probability of 25%

Illustrative Example

Overall structure is preserved

Fundamental Property of Privacy

No technique is perfect. No perfectly accurate and deterministic privacy technique can satisfy our requirements, and that ***randomization is essential*** for privacy.

A lack of randomness leads to failure in *composition*.

Datasets and Queries

Dataset D

City	Age	Gender
New York	18	M
New York	21	F
Los Angeles	33	other
Madrid	41	M
San Diego	29	F
St. Louis	38	M

query

$f(D)$ = "**SELECT COUNT**(*) **FROM** D
WHERE Gender = M **AND** Age > 18"

A numeric query:

- $f : D \rightarrow R \subseteq R^r$ that maps a dataset in some real vector space

Adjacency

Two datasets D and D′

Add/remove adjacency

$$D \sim D' \text{ if } |D \Delta D'| = 1$$

where Δ is the symmetric difference of two sets

Exchange adjacency

$$D \sim\!\!\leftrightarrow D'$$

If D' is obtained from D by successively removing one record and then adding a (possibly different) record, then:

There exist elements $d \in D$ and $d' \in U$ such that:

$$D' = (D \setminus \{d\}) \cup \{d'\}$$

This implies that $|D| = |D'|$ and $|D \Delta D'| = 2$.

Adjacency

For two datasets D and D′ ,

Add/remove adjacency

- D and D’ are neighbors if one is a 1-row add/remove difference away from each other
  - According to unbounded differential privacy
  - ⇒ sizes of D and D’ are different by one row

Exchange adjacency

- D and D’ are neighbors if one is a 1-row exchange difference away from each other
  - According to bounded differential privacy
  - ⇒ sizes of D and D’ are equal

Units of Privacy

- Unit of privacy = formal definition of “neighboring” in DP guarantee
- E.g.,
  - **“One person”**
    - Privacy guarantee for whole person
  - Apple’s DP with **“person-day”**
    - Privacy guarantee for one person on a single day

- Unit of privacy can have privacy failures
  - E.g., Apple’s DP does NOT protect trends in data occurring across multiple days, even for individuals
  - “One-person” – good default, usually avoids surprises

- Key assumption for defining neighboring datasets: **each individual’s data is contained in one row of data**

Units of Privacy

- Other units of privacy either to
  - Make data easier to analyze
  - Add difficulty to tie data values to individuals

- Simple assumption to easily formalize the definition of neighboring datasets: **each individual's data is contained in <u>one</u> row of data**
  - Ture ⇒ can define neighboring datasets formally and retain "one person" unit of privacy
  - False ⇒ best to avoid using a different unit of privacy whenever possible

Global Sensitivity

How does a single individual's data impact overall analysis?

The global sensitivity of a function $f : D \to \mathbb{R}$ is defined as the maximum difference in the output of f over all pairs of adjacent datasets $D \sim D' \in D$, measured with respect to the ℓ_p norm:

$$\Delta_p f = \max_{D \sim D'} \|f(D) - f(D')\|_p.$$

Global Sensitivity

Dataset D

City	Age	Gender
New York	18	M
New York	21	F
Los Angeles	33	other
Madrid	41	M
San Diego	29	F
St. Louis	38	M

query

$f(D)$ = "**SELECT COUNT**(*) **FROM** D **WHERE** Gender = M **AND** Age > 18"

Global sensitivity for this query is 1

- Adding or removing a single individual in the dataset can affect the final count by at most 1

Global Sensitivity

Dataset D

City	Age	Gender
New York	18	M
New York	21	F
Los Angeles	33	other
Madrid	41	M
San Diego	29	F
St. Louis	38	M

Task:

compute the average age of all individuals in the dataset

Age range A = [0, 100]

$$\Delta_p f = \frac{\max(A) - \min(A)}{|D|} = \frac{100}{|D|}$$

Differential Privacy, Formally

A randomized mechanism $M : D \to R$ with domain D and range R is (ϵ, δ)-differentially private if, for any event $S \subseteq R$ and any pair $D, D' \in D$ of adjacent datasets:

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

where the probability is calculated over the randomness of M.

ε is called the privacy loss
- controls the level of privacy
- closer to 0 implies stronger privacy

Differential Privacy, Formally

A randomized mechanism $M : D \to R$ with domain D and range R is (ϵ, δ)-differentially private if, for any event $S \subseteq R$ and any pair $D, D' \in D$ of adjacent datasets:

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

where the probability is calculated over the randomness of M.

ε is called the privacy loss
- controls the level of privacy
- closer to 0 implies stronger privacy

δ is the failure threshold
- allows DP not to hold with probability up to δ

Differential Privacy, Formally

Often δ is set to 0

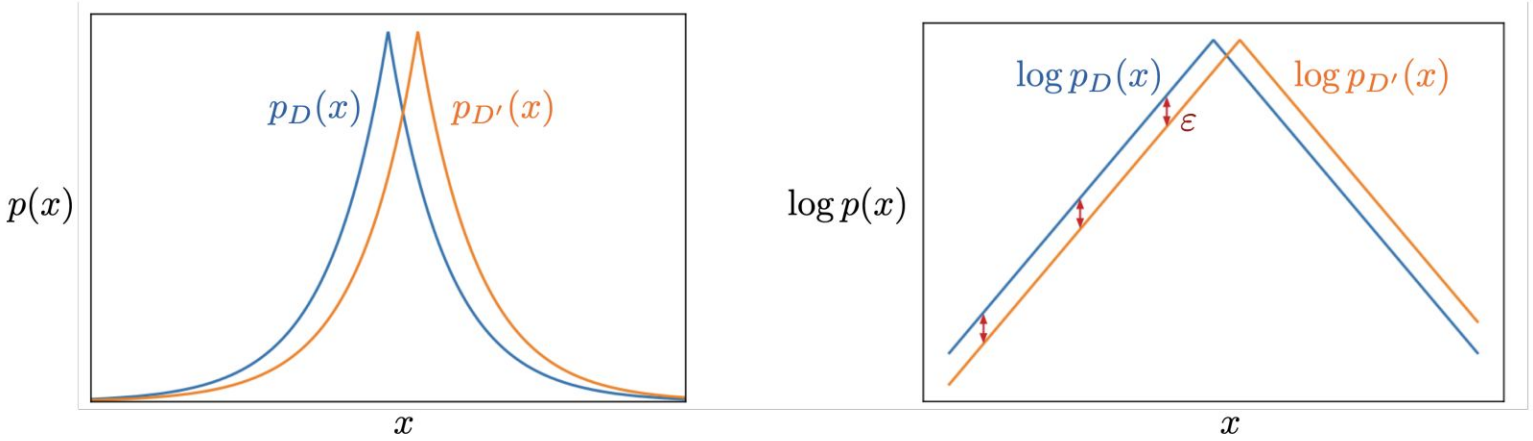
- Called (ε,0)-differentially private or ε-differentially private

$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^{\epsilon}$$

ε is called the privacy loss

- controls the level of privacy
- closer to 0 implies stronger privacy

Differential Privacy, Visually



DP of Randomized Response

Did you cheat in this course?

1. Flip a coin
2. If ***tails*** - respond truthfully
3. If ***heads*** - flip a coin again
   a. If ***heads*** - respond “Yes”
   b. If ***tails*** - respond “No”

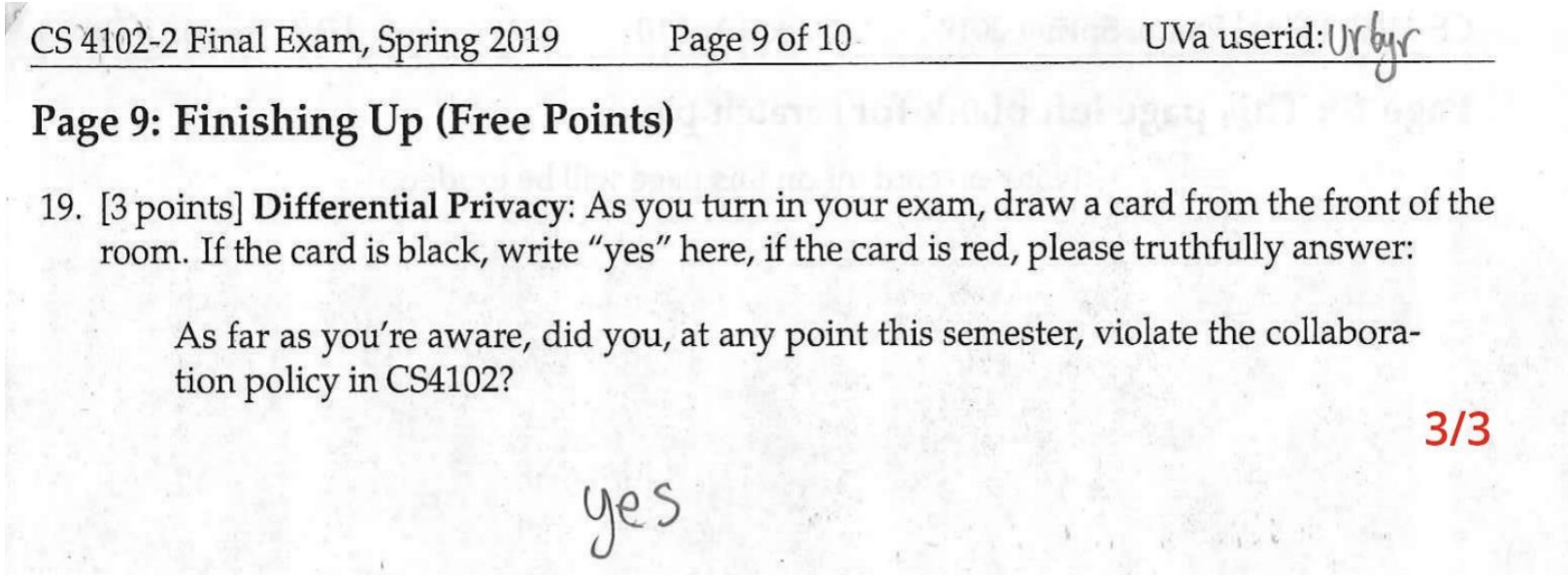
$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^{\epsilon}$$

$$\frac{\Pr[\text{Response} = \text{Yes} | \text{Truth} = \text{Yes}]}{\Pr[\text{Response} = \text{Yes} | \text{Truth} = \text{No}]}$$

$$= \frac{3/4}{1/4} = \frac{\Pr[\text{Response} = \text{No} | \text{Truth} = \text{No}]}{\Pr[\text{Response} = \text{No} | \text{Truth} = \text{Yes}]} = 3$$

ln3-differential privacy

DP of Real Example



CS 4102-2 Final Exam, Spring 2019 Page 9 of 10 UVa userid: Urbyr

Page 9: Finishing Up (Free Points)

19. [3 points] **Differential Privacy**: As you turn in your exam, draw a card from the front of the room. If the card is black, write "yes" here, if the card is red, please truthfully answer:

As far as you're aware, did you, at any point this semester, violate the collaboration policy in CS4102?

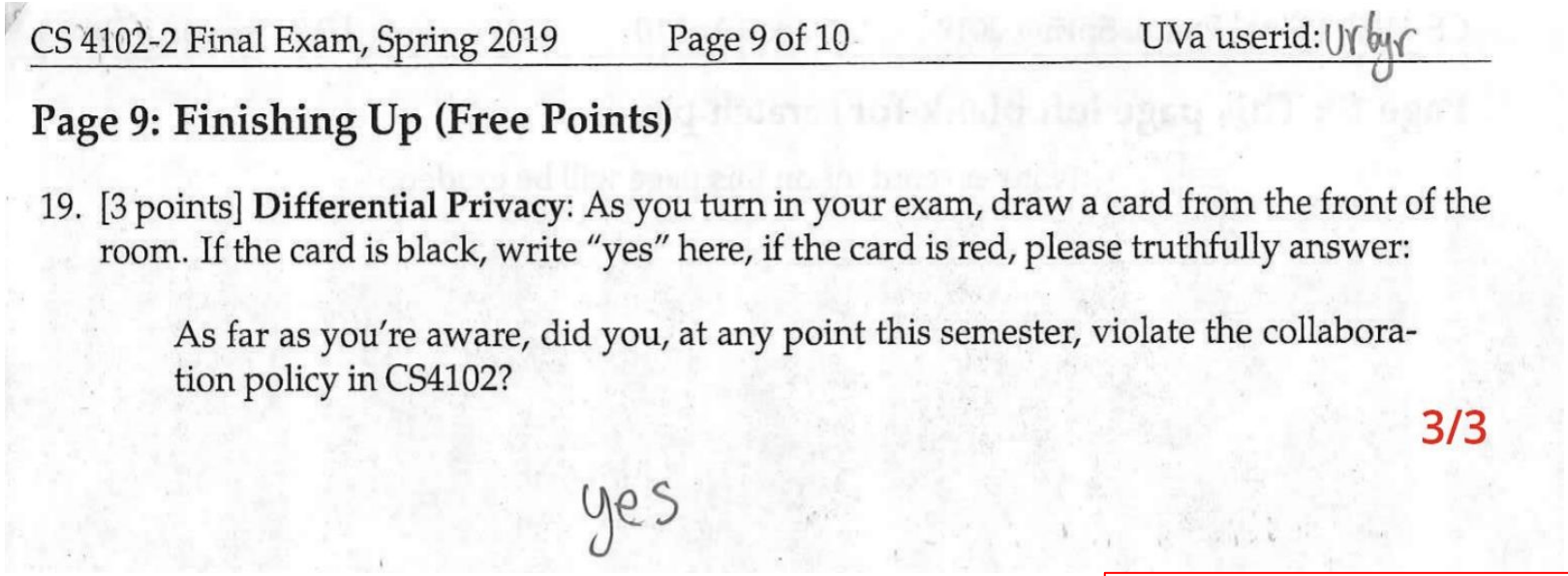
3/3

yes

$$\frac{\Pr[\text{Response} = \text{Yes} | \text{Truth} = \text{Yes}]}{\Pr[\text{Response} = \text{Yes} | \text{Truth} = \text{No}]} = \frac{1}{0.5}$$

ln2-differential privacy

DP of Real Example



CS 4102-2 Final Exam, Spring 2019 Page 9 of 10 UVa userid: Urbyr

Page 9: Finishing Up (Free Points)

19. [3 points] **Differential Privacy**: As you turn in your exam, draw a card from the front of the room. If the card is black, write "yes" here, if the card is red, please truthfully answer:

As far as you're aware, did you, at any point this semester, violate the collaboration policy in CS4102?

3/3

yes

$$\frac{\Pr[\text{Response} = \text{No} | \text{Truth} = \text{No}]}{\Pr[\text{Response} = \text{No} | \text{Truth} = \text{Yes}]} = \frac{0.5}{0}$$

Privacy is broken for "no" responders

What Differential Privacy Promises

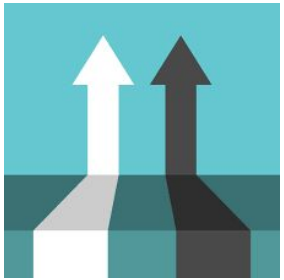
- Compositionality
- Group Privacy
- Post-processing immunity,
- Quantifiable privacy-accuracy tradeoffs

Ensures that individuals are not exposed to any **additional harm** due to their data being included in the private database *x* compared to if their data had not been part of *x*

What Differential Privacy Promises

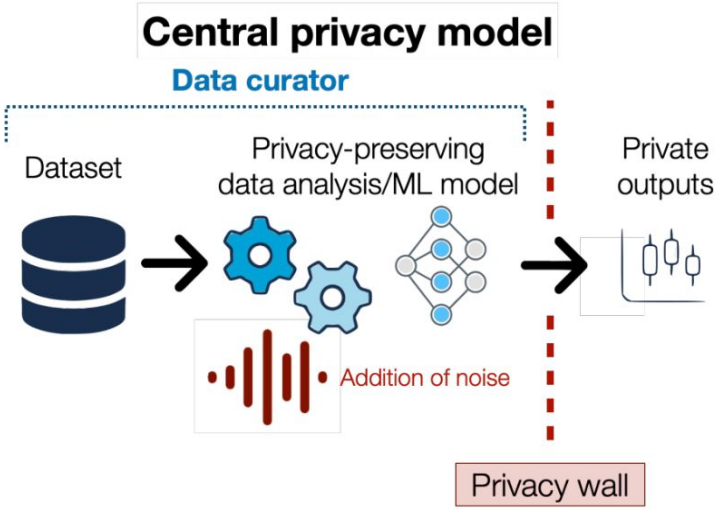
- Compositionality
  - Sequential Composition
    - Property applies when multiple DP mechanisms are applied to the same dataset
    - Cumulative privacy loss is bounded by the sum of the individual privacy parameters

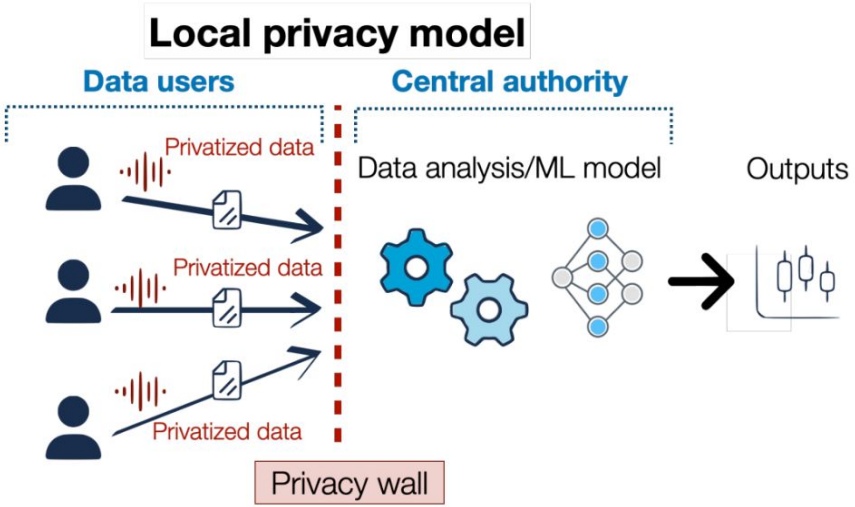
  - Parallel Composition
    - Property applies when dataset is split into disjoint chunks and mechanism is ran on each chunk
    - Privacy cost remains at the level of a single mechanism's privacy parameter ε
      - Mechanism applied k times on k disjoint subsets ⇒ privacy cost = ε, not kε



University of Virginia

Where to Guarantee DP





Which is better?

University of Virginia

Where to Guarantee DP

Centralized Model:

- Data is collected, stored, and processed at a central location managed by a trusted data curator
- The curator has direct access to raw data and ensures privacy mechanisms are properly applied.
- This model assumes the central entity will responsibly handle and protect data

Distributed (Local) Model:

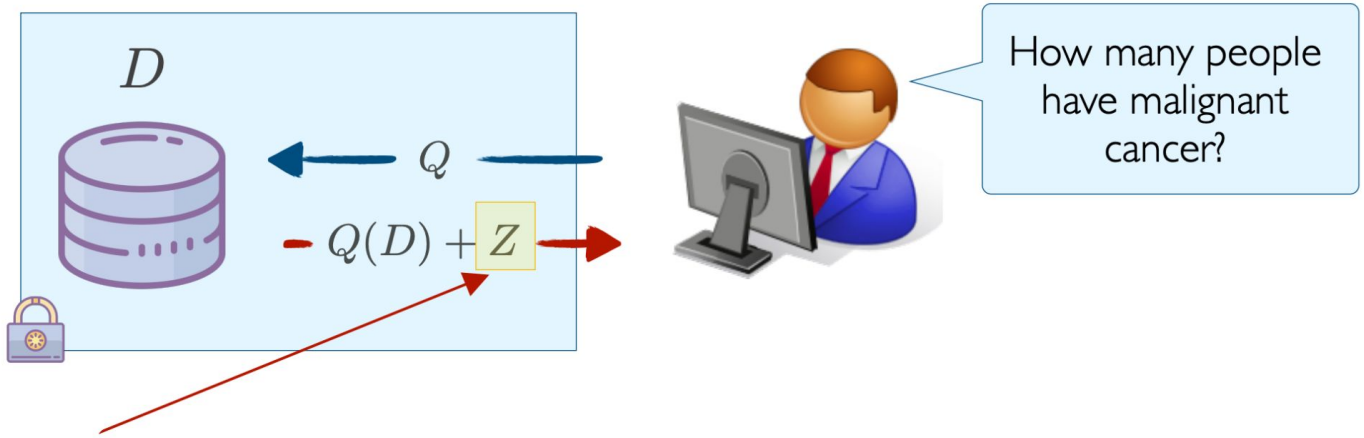
- Data remains decentralized, residing on personal devices or local databases
- Privacy-preserving algorithms are applied locally before sending processed data to a central authority

Which is better?

Statistical Queries vs Private Selection

Differential Privacy: Design & Implementation

How Do We Design a DP Algorithm



Add noise to the real answer while

1. Not leaking too much information about the dataset
2. Keeping noisy answers closer to the real answer

How Do We Add Noise

Laplace distribution

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

μ is a location parameter and *b* is a scale parameter

DP definition

$$\frac{\Pr[F(x) = S]}{\Pr[F(x') = S]} \leq e^{\epsilon}$$

Laplace mechanism

$$F(x) = f(x) + \text{Lap}\left(\frac{s}{\epsilon}\right)$$

s is global sensitivity,

Lap(S) is the Laplace distribution sampling,

ε is the privacy parameter controlling the level of privacy

How Do We Add Noise

Laplace distribution

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

μ is a location parameter and *b* is a scale parameter

DP definition

$$\frac{\Pr[F(x) = S]}{\Pr[F(x') = S]} \leq e^{\epsilon}$$

Laplace mechanism

$$F(x) = f(x) + \text{Lap}\left(\frac{s}{\epsilon}\right)$$

s is global sensitivity,

Lap(S) is the Laplace distribution sampling,

ε is the privacy parameter controlling the level of privacy

** Proof of Laplace mechanism is **ε**-DP: https://arxiv.org/abs/2411.04710

University of Virginia

Laplace: ε-Differential Privacy

Proof. Let $D \sim D'$ be any two neighboring datasets in $\mathcal{D}$, and let p_D and $p_{D'}$ be the probability density functions of $\mathcal{M}_{\text{Lap}}(D; f, \varepsilon)$ and $\mathcal{M}_{\text{Lap}}(D'; f, \varepsilon)$, respectively. Then for any $r \in \mathcal{R}$,

$$
\begin{aligned}
\frac{p_D(r)}{p_{D'}(r)} &= \prod_{i=1}^{d} \left(\frac{\exp\left(-\frac{\varepsilon |f(D)_i - r_i|}{\Delta f}\right)}{\exp\left(-\frac{\varepsilon |f(D')_i - r_i|}{\Delta f}\right)} \right) \\
&= \prod_{i=1}^{d} \exp\left(\frac{\varepsilon(|f(D')_i - r_i| - |f(D)_i - r_i|)}{\Delta f} \right) \\
&\leq \prod_{i=1}^{d} \exp\left(\frac{\varepsilon(|f(D')_i - f(D)_i|)}{\Delta f} \right) && \text{(By the triangle inequality.)} \\
&= \prod_{i=1}^{d} \exp\left(\frac{\varepsilon \cdot (\|f(D) - f(D')\|_1)}{\Delta f} \right) && \text{(By the definition of } \Delta f.\text{)} \\
&\leq \exp(\varepsilon).
\end{aligned}
$$

UNIVERSITY of VIRGINIA

How Much Noise is Enough?

- What is enough noise to prevent re-identification?
- Attempt to break into dataset:

```
karries_row = adult[adult['Name'] == 'Karrie Trusslove']
karries_row[karries_row['Target'] == '<=50K'].shape[0]
```

  - Result: 1 == income column for Karrie's row (privacy violation)

- Attempt to break into dataset with Laplace Mechanism:

```
sensitivity = 1
epsilon = 0.1

karries_row = adult[adult['Name'] == 'Karrie Trusslove']
karries_row[karries_row['Target'] == '<=50K'].shape[0] + \
  np.random.laplace(loc=0, scale=sensitivity/epsilon)
```

  - Result: 0.00437885... == too noisy to reliably tell

How Much Noise is Enough?

- DP adds noise to gradient when training
- Noise in DP mechanism CAN ⇒ model accuracy to become worse

Tradeoffs

- Smaller values of ε ⇒ less accurate models
- More iterations ⇒ larger privacy cost for DP mechanism (worse model)
  - Need to use smaller ε for each iteration ⇒
  - Upwards noise scaling

Hard to balance iterations/accuracy and scale of noise

Laplace: Accuracy Guarantee

For any numerical query $f : \mathcal{D} \to \mathcal{R} \subseteq \mathbb{R}^d$, *and any database* $D \in \mathcal{D}$,

$$Pr\left[|f(D) - \mathcal{M}_{Lap}(D; f; \varepsilon)| \geq \ln\left(\frac{d}{\beta}\right) \cdot \left(\frac{\Delta f}{\varepsilon}\right)\right] \leq \beta$$

Laplace: Accuracy Guarantee

Proof. The proof is for $d = 1$ for simplicity, but it generalizes for $d > 1$. The proof follows from characterizations of the tails of the Laplace distribution. For a random variable $Z \sim Lap(b)$ and a real number $\alpha > 0$,

$$Pr\left[|Z| \geq \alpha\right] = \exp\left(-\alpha/b\right).$$

Therefore, given that $f(D) - \mathcal{M}_{Lap}(D; f; \varepsilon)$ is Laplace with parameter $b = \frac{\Delta f}{\varepsilon}$, it follows that

$$Pr\left[|f(x) - \mathcal{M}_{Lap}(D; f; \varepsilon)| \geq \alpha\right] = \exp\left(-\alpha \cdot \frac{\varepsilon}{\Delta f}\right) \triangleq \beta.$$

Solving for α in $\exp\left(-\alpha \cdot \frac{\varepsilon}{\Delta f}\right) = \beta$ leads to

$$\alpha \cdot \frac{\varepsilon}{\Delta f} = \ln\left(\frac{1}{\beta}\right),$$

hence

$$\alpha = \ln\left(\frac{1}{\beta}\right) \cdot \left(\frac{\Delta f}{\varepsilon}\right).$$

Laplace Mechanism: Computing Average

Consider a dataset containing the ages of 10, 000 individuals:

- ages ranging from 0 to 100 years

Task: compute the average age while ensuring differential privacy

Laplace Mechanism: Computing Average

1. *Determine the query function and its sensitivity.* In this task the query function is the average age,

$$f(\text{data}) = \frac{1}{n}\sum_{i=1}^{n} \text{age}_i,$$

$$\Delta f = \frac{\text{max age} - \text{min age}}{n} = \frac{100 - 0}{10{,}000} = 0.01.$$

Laplace Mechanism: Computing Average

2. *Apply the Laplace Mechanism.* The next step is to select the privacy parameter ε and add noise drawn from the Laplace distribution with scale parameter $\frac{\Delta f}{\varepsilon}$. Selecting $\varepsilon = 0.5$ to obtain a strong privacy guarantee adds the following noise:

$$\text{noise} \sim \text{Lap}\left(\frac{\Delta f}{\varepsilon}\right) = \text{Lap}\left(\frac{0.01}{0.5}\right) = \text{Lap}(0.02).$$

The private query thus reports $f(\text{data}) + \text{noise}$.

Laplace Mechanism: Computing Average

3. *Analyze the error bound.* Additionally, by setting a confidence level $\beta = 0.05$ (meaning that one is 95% confident in the error bound), the error bound can be computed as,

$$\text{Error Bound} = \frac{\Delta f}{\varepsilon} \ln\left(\frac{1}{\delta}\right) = \frac{0.01}{0.5} \ln\left(\frac{1}{0.05}\right) \approx 0.06 \text{ years}.$$

How Do We Add Noise

Exponential mechanism

- DP method that selects an object from a set based on a scoring function
  - With dataset D, set of objects H, score function s(D, h), it outputs object h with probability proportional to

$$\exp\left(\frac{\varepsilon s(D,h)}{2\Delta s}\right), \text{ where } \Delta s \triangleq \max_{h\in\mathcal{H}} \max_{D\sim D'} |s(D,h) - s(D',h)|.$$

  - ε = the privacy parameter controlling the level of privacy
  - Δs = global sensitivity of the utility / score function
- ε controls trade-off between privacy and utility
  - Higher ⇒ more weight on quality of selection, less privacy
  - Lower ⇒ more randomness / privacy, less utility

How Else Do We Add Noise

Exponential mechanism

- DP method that selects an object from a set based on a scoring function
  - With dataset D, set of objects H, score function s(D, h), it outputs object h with probability proportional to

$$\exp\left(\frac{\varepsilon s(D,h)}{2\Delta s}\right), \text{ where } \Delta s \triangleq \max_{h\in\mathcal{H}} \max_{D\sim D'} |s(D,h) - s(D',h)|.$$

  - ε = the privacy parameter controlling the level of privacy
  - Δs = global sensitivity of the utility / score function

How Do We Add Noise

Exponential mechanism

- DP method that selects an object from a set based on a scoring function
  - With dataset D, set of objects H, score function s(D, h), it outputs object h with probability proportional to

$$\exp\left(\frac{\varepsilon s(D,h)}{2\Delta s}\right), \text{ where } \Delta s \triangleq \max_{h\in\mathcal{H}} \max_{D\sim D'} |s(D,h) - s(D',h)|.$$

  - ε = the privacy parameter controlling the level of privacy
  - Δs = global sensitivity of the utility / score function

** Proof of Exponential mechanism is ε-DP: https://arxiv.org/abs/2411.04710

UNIVERSITY of VIRGINIA

Exponential Mechanism

Proof. The proof assumes that $\mathcal{H}$ is a finite set. For any two neighbouring datasets $D \sim D'$ and some outcome $h \in \mathcal{H}$,

$$\begin{aligned}
\frac{\Pr[\mathcal{M}_{\exp}(D) = h]}{\Pr[\mathcal{M}_{\exp}(D') = h]} &= \frac{\left(\frac{\exp(\varepsilon s(D,h)/2\Delta s)}{\sum_{h' \in \mathcal{H}} \exp\left(\varepsilon s(D,h')/2\Delta s\right)}\right)}{\left(\frac{\exp\left(\varepsilon s(D',h)/2\Delta s\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\varepsilon s(D',h')/2\Delta s\right)}\right)} \\
&= \exp\left(\frac{\varepsilon(s(D,h) - s(D',h))}{2\Delta s}\right) \frac{\sum_{h' \in \mathcal{H}} \exp\left(\varepsilon s(D',h')/2\Delta s\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\varepsilon s(D,h')/2\Delta s\right)} \\
&\leq \exp\left(\frac{\varepsilon}{2}\right) \exp\left(\frac{\varepsilon}{2}\right) \frac{\sum_{h' \in \mathcal{H}} \exp\left(\varepsilon s(D,h')/2\Delta s\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\varepsilon s(D,h')/2\Delta s\right)} \\
&= \exp(\varepsilon).
\end{aligned}$$

The inequality follows due to the definition of Δs. □

Exponential Mechanism

For the exponential mechanism, accuracy is determined by guaranteeing, with high probability, that the output by the mechanism has a high score, as close as possible to optimality

Theorem 6.2. *Let us fix a database D, and let $\mathcal{H}_{OPT} = \{h^* \in \mathcal{H} \text{ s.t. } s(D,h) = \max_h s(D,h)\}$ be the set of elements in $\mathcal{H}$ that achieve the maximum possible utility score. Then, the exponential mechanism guarantees*

$$Pr\left[s(D, \mathcal{M}_{exp}(D)) \geq OPT - \frac{2\Delta s}{\varepsilon}\left(\ln\left(|\mathcal{H}|/\beta\right)\right)\right] \geq 1 - \beta.$$

where $OPT = \max_h s(D,h)$.

Exponential Mechanism

Proof. Take any $c \in \mathbb{R}$. It follows that

$$\begin{aligned}
\Pr\left[s(D, \mathcal{M}_{\exp}(D) \leq c\right] &= \frac{\sum_{h:\, s(D,h)\leq c} \exp\left(\varepsilon s(D,h)/2\Delta s\right)}{\sum_{r\in\mathcal{H}} \exp(\varepsilon s(D,h)/2\Delta s)} \\
&\leq \frac{\sum_{r:\, s(D,h)\leq c} \exp\left(\varepsilon c/2\Delta s\right)}{\sum_{r\in\mathcal{H}_{\mathrm{OPT}}} \exp(\varepsilon \mathrm{OPT}/2\Delta s)} \\
&\leq \frac{|\mathcal{H}| \exp(\varepsilon c/2\Delta s)}{|\mathcal{H}_{\mathrm{OPT}}| \exp(\varepsilon \mathrm{OPT}/2\Delta s)} \\
&= \frac{|\mathcal{H}|}{|\mathcal{H}_{\mathrm{OPT}}|} \exp\left(\frac{\varepsilon(c-\mathrm{OPT})}{2\Delta s}\right) \\
&\leq |\mathcal{H}| \exp\left(\frac{\varepsilon(c-\mathrm{OPT})}{2\Delta s}\right).
\end{aligned}$$

The result follows by plugging in

$$c \triangleq \mathrm{OPT} - \frac{2\Delta s}{\varepsilon} \ln\left(|\mathcal{H}|/\delta\right).$$

DP Efficiency

- Time efficiency
  - Iterative algorithms and complex query processing can ⇒ high computational costs
    - Especially with high-dimensional queries
- Space efficiency
  - Storing large histograms, high counts, and detailed representations ⇒ significant memory usage

- Optimization techniques: convex optimization, iterative refinement, dimension reduction, query batching

- Efficiency bottlenecks
  - High computational complexity, significant memory overhead
  - Tradeoff between privacy guarantees and performance

DP Synthetic Data

- Data can be generated from DP algorithms that
  - Have added calibrated noise to preserve privacy
  - Capture key statistical properties of original data
- **Goal:** answer queries of original while maintaining privacy

- Simple synthetic representations e.g., counts, histograms to capture key data features
- Marginals – projections of the joint data distribution onto a subset of dimensions

- Tradeoff: high-dimensional synthetic data
  - Exponential growth in combos ⇒ high computational overhead
  - Increased noise requirements; can degrade performance

Discussion

Q&A

- There's a tradeoff between accuracy and privacy. How much privacy do we need? When do we need both accuracy and privacy?
  - Think of some applications where accuracy is more important than privacy and vice versa.

- How could we apply differential privacy to deep learning algorithms? Could it improve an algorithm's fairness or robustness? What mechanisms could we implement it into?

- The census is an important database that affects millions of people, where differential privacy is needed for proper protection to the block level. Are there other databases that need differential privacy implementation?

NEWS

NIST Finalizes Guidelines for Evaluating 'Differential Privacy' Guarantees to De-Identify Data

March 6, 2025

https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-226.pdf

Consumers urged to delete 23andMe data as bankruptcy sparks privacy fears

By **Bhanvi Satija** and **Siddhi Mahatole**

March 25, 2025 6:36 PM EDT · Updated 7 days ago

Putting It All Together

Key takeaways

- Traditional privacy methods are insufficient - DRAs are real threats
- Differential privacy provides a stronger defense, but it's still evolving
- Future challenge: balancing data usability and privacy protection
- To learn more about and practice code for DP, you can go to the following resource: https://programming-dp.com/cover.html

Thank You

CS 6501: Responsible AI

References

- S.A. Keller, & J.M. Abowd, Database reconstruction does compromise confidentiality, *Proc. Natl. Acad. Sci. U.S.A.* 120 (12) e2300976120, https://doi.org/10.1073/pnas.2300976120 (2023).
- Theofanos, Mary. “Differential Privacy: A Q&A with NIST’s Mary Theofanos.” *NIST*, Mary Theofanos, 9 Aug. 2019, www.nist.gov/blogs/taking-measure/differential-privacy-qa-nists-mary-theofanos.
- Fioretto, Ferdinando, Pascal Van Hentenryck, and Juba Ziani. "Differential Privacy Overview and Fundamental Techniques." *arXiv preprint arXiv:*2411.04710 (2024).
- Kamath, Gautam. “Lecture 1 - Some Attempts at Data Privacy.” *CS 860 - Algorithms for Private Data Analysis - Fall 2020*, 2020, www.gautamkamath.com/CS860-fa2020.html.
- Martindale, Christian, et al. “Understanding Database Reconstruction Attacks on Public Data.” *DigitalCollections@ILR*, 1 Jan. 2018, ecommons.cornell.edu/items/046034b9-9365-436b-88aa-e8c3fae94b7c.
- “Programming Differential Privacy.” *Programming Differential Privacy - Programming Differential Privacy*, programming-dp.com/cover.html.