

Basic Differential Privacy Algorithms and Statistics

Daniel Kifer

NBER Summer Institute

July 17, 2020

Outline

- 1 Basic Design of Differential Privacy Mechanisms
- 2 The Role of Strategy
- 3 Chi-Squared Testing
- 4 Takeaway Messages
- 5 Common Pitfalls
- 6 Additional Mechanisms

What we learned

Definition (Differential Privacy [DMNS06])

Given a privacy loss budget $\epsilon > 0$, an randomized algorithm M satisfies ϵ -differential privacy if for all $E \subset \text{range}(M)$ and all pairs of databases D_1, D_2 that are neighbors of each other,

$$P(M(D_1) \in E) \leq e^\epsilon P(M(D_2) \in E)$$

- Mechanisms: algorithms for Differential Privacy.
- Protects confidentiality of our responses.

What we learned

Definition (Differential Privacy [DMNS06])

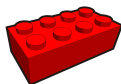
Given a privacy loss budget $\epsilon > 0$, an randomized algorithm M satisfies ϵ -differential privacy if for all $E \subset \text{range}(M)$ and all pairs of databases D_1, D_2 that are neighbors of each other,

$$P(M(D_1) \in E) \leq e^\epsilon P(M(D_2) \in E)$$

- Mechanisms: algorithms for Differential Privacy.
- Protects confidentiality of our responses.
- But how do we design mechanisms M ?
- The conditions have to hold for:
 - All pairs of databases that are neighbors of each other.
 - All sets E .
 - Nearly infinitely many equations to check!

Differential Privacy and Modularity

- Complex Mechanisms built from simpler ones



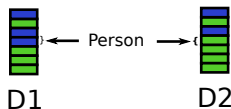
- Basic tools:
 - Sensitivity & Laplace Mechanism.
 - Postprocessing.
 - Composition.

Sensitivity

- Neighbors in Differential Privacy: $D_1 \sim D_2$

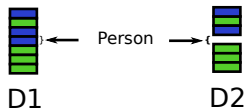
- Bounded neighbors:

- differ on value of one record.
 - use this to ensure response is protected.



- Unbounded neighbors:

- differ on presence/absence of one record.
 - use this to protect participation and response.



- Differential privacy: hide differences between neighbors.

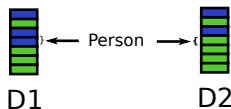
- How to compute $f(D) = \left[\begin{array}{l} \text{average age of voters} \\ \text{average age of non-voters} \end{array} \right]$ with privacy?

Sensitivity

- Neighbors in Differential Privacy: $D_1 \sim D_2$

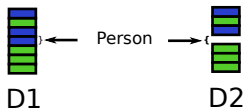
- Bounded neighbors:

- differ on value of one record.
 - use this to ensure response is protected.



- Unbounded neighbors:

- differ on presence/absence of one record.
 - use this to protect participation and response.



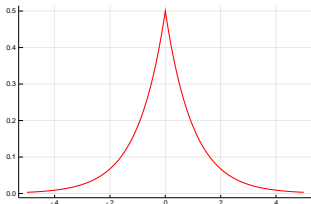
- Differential privacy: hide differences between neighbors.

- How to compute $f(D) = \left[\begin{array}{l} \text{average age of voters} \\ \text{average age of non-voters} \end{array} \right]$ with privacy?

- Inject enough noise to hide any person's response.
 - For any $D_1 \sim D_2$, noise should mask difference between $f(D_1)$ and $f(D_2)$.

Sensitivity

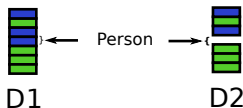
- If we want to add noise, sensitivity tells us how much.



- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.
 - Add Laplace noise with scale Δ_f/ϵ (std= $\sqrt{2}\Delta_f/\epsilon$)

Sensitivity

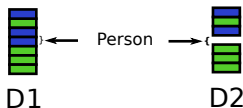
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D) = \text{sum of ages of people in } D$
 - Assume ages are a priori capped at 115

Sensitivity

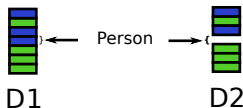
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D) = \text{sum of ages of people in } D$
 - Assume ages are a priori capped at 115
 - Adding or removing 1 person to any database can change sum by at most ± 115
 - $\Delta_f = 115$

Sensitivity

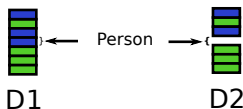
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D)$ = number of people 18 years or older.
 - Adding or removing 1 person can change count by at most ± 1
 - $\Delta_f = 1$

Sensitivity

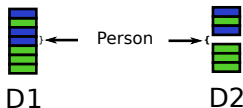
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D) = [\text{number in GQ}, \text{number of Asians}]$.
 - Largest change caused by adding/removing 1 Asian individual in a GQ.
 - $\Delta_f = 2$

Sensitivity

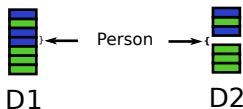
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D) = [\# \text{ of 1-year-olds}, \# \text{ of 2-year-olds}, \dots, \# \text{ of 100-year-olds}]$
 - Any record addition/removal changes exactly one component by ± 1 .
 - $\Delta_f = 1$.

Sensitivity

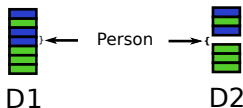
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are a priori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$

Sensitivity

- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.



- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are a priori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
 - Consider $D_1 = \emptyset$, $D_2 = \{115\}$.
 - $f(D_1) = [0, 0]$
 - $f(D_2) = [0, 115]$
 - $\Delta_f = 115$

Sensitivity

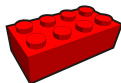
- If we want to add noise, sensitivity tells us how much.
- For Laplace noise:
 - L_1 Sensitivity Δ_f : largest possible impact of one person on f .
 - $\Delta_f = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$.
 - Supremum over all pairs of neighbors.
- Laplace mechanism $M(D)$: add independent $\text{Laplace}(\Delta_f/\epsilon)$ noise to each component of f .

$$f(D) = [\# \text{ of 1-year-olds}, \# \text{ of 2-year-olds}, \dots, \# \text{ of 100-year-olds}]$$

$$M(D) = \begin{bmatrix} \# \text{ of 1-year-olds} & +\text{Laplace}(\Delta_f/\epsilon) \\ \# \text{ of 2-year-olds} & +\text{Laplace}(\Delta_f/\epsilon) \\ \# \text{ of 3-year-olds} & +\text{Laplace}(\Delta_f/\epsilon) \\ \vdots & \vdots \\ \# \text{ of 100-year-olds} & +\text{Laplace}(\Delta_f/\epsilon) \end{bmatrix}$$

Differential Privacy and Modularity

- Complex Mechanisms built from simpler ones
- Basic tools:
 - Sensitivity & Laplace Mechanism.
 - **Postprocessing.**
 - Composition.



Postprocessing

- Suppose M satisfies ϵ -differential privacy.

$$\text{e.g., } M(D) = \begin{bmatrix} \# \text{ of 1-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \# \text{ of 2-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \vdots & \vdots \\ \# \text{ of 100-year-olds} & +Laplace(\Delta_f/\epsilon) \end{bmatrix}$$

Postprocessing

- Suppose M satisfies ϵ -differential privacy.

$$\text{e.g., } M(D) = \begin{bmatrix} \# \text{ of 1-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \# \text{ of 2-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \vdots & \vdots \\ \# \text{ of 100-year-olds} & +Laplace(\Delta_f/\epsilon) \end{bmatrix}$$

- Let g be code that performs chi-squared test.
 - $g \circ M$: run $M(D)$ then run g on the result.
 - Then $g \circ M$ satisfies ϵ -differential privacy (same ϵ parameter)

Postprocessing

- Suppose M satisfies ϵ -differential privacy.

$$\text{e.g., } M(D) = \begin{bmatrix} \# \text{ of 1-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \# \text{ of 2-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \vdots & \vdots \\ \# \text{ of 100-year-olds} & +Laplace(\Delta_f/\epsilon) \end{bmatrix}$$

- Let g be code that performs chi-squared test.
 - $g \circ M$: run $M(D)$ then run g on the result.
 - Then $g \circ M$ satisfies ϵ -differential privacy (same ϵ parameter)
- Let h be code that links to external data.
 - Then $h \circ M$ satisfies ϵ -differential privacy.

Postprocessing

- Suppose M satisfies ϵ -differential privacy.

$$\text{e.g., } M(D) = \begin{bmatrix} \# \text{ of 1-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \# \text{ of 2-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \vdots & \vdots \\ \# \text{ of 100-year-olds} & +Laplace(\Delta_f/\epsilon) \end{bmatrix}$$

- Let g be code that performs chi-squared test.
 - $g \circ M$: run $M(D)$ then run g on the result.
 - Then $g \circ M$ satisfies ϵ -differential privacy (same ϵ parameter)
- Let h be code that links to external data.
 - Then $h \circ M$ satisfies ϵ -differential privacy.
- Let ϕ be any function that does not look directly at the collected data D .
 - Then $\phi \circ M$ satisfies ϵ -differential privacy.

Postprocessing

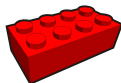
- Suppose M satisfies ϵ -differential privacy.

$$\text{e.g., } M(D) = \begin{bmatrix} \# \text{ of 1-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \# \text{ of 2-year-olds} & +Laplace(\Delta_f/\epsilon) \\ \vdots & \vdots \\ \# \text{ of 100-year-olds} & +Laplace(\Delta_f/\epsilon) \end{bmatrix}$$

- Let g be code that performs chi-squared test.
 - $g \circ M$: run $M(D)$ then run g on the result.
 - Then $g \circ M$ satisfies ϵ -differential privacy (same ϵ parameter)
- Let h be code that links to external data.
 - Then $h \circ M$ satisfies ϵ -differential privacy.
- Let ϕ be any function that does not look directly at the collected data D .
 - Then $\phi \circ M$ satisfies ϵ -differential privacy.
- Differential privacy is closed under post-processing.
- Very few other disclosure avoidance techniques have this property.

Differential Privacy and Modularity

- Complex Mechanisms built from simpler ones



- Basic tools:
 - Sensitivity & Laplace Mechanism.
 - Postprocessing.
 - **Composition.**

Composition

- Week 1: we conduct Senate Poll using ϵ_1 -differential privacy.
 - Release number of “yes” responses + Laplace($1/\epsilon_1$) noise.
- Week 2: we release:
 - Number of “yes” responses from Democrats + Laplace($1/\epsilon_2$) noise.
 - Number of “yes” responses from Republicans + Laplace($1/\epsilon_2$) noise.
 - Sensitivity is 1, so week 2 release satisfies ϵ_2 -differential privacy.
- Surely there is some combined privacy leakage?

Composition

- Week 1: we conduct Senate Poll using ϵ_1 -differential privacy.
 - Release number of “yes” responses + Laplace($1/\epsilon_1$) noise.
- Week 2: we release:
 - Number of “yes” responses from Democrats + Laplace($1/\epsilon_2$) noise.
 - Number of “yes” responses from Republicans + Laplace($1/\epsilon_2$) noise.
 - Sensitivity is 1, so week 2 release satisfies ϵ_2 -differential privacy.
- Surely there is some combined privacy leakage?
 - This is called composition.
 - By itself, Week 1 satisfies ϵ_1 -differential privacy (privacy loss = ϵ_1).
 - By itself, Week 2 satisfies ϵ_2 -differential privacy (privacy loss = ϵ_2).
 - The combined release (Week 1 and Week 2) satisfies $(\epsilon_1 + \epsilon_2)$ -differential privacy (privacy loss = $\epsilon_1 + \epsilon_2$).
 - Hence ϵ is the privacy loss budget.

Composition

- Week 1: we conduct Senate Poll using ϵ_1 -differential privacy.
 - Release number of “yes” responses + Laplace($1/\epsilon_1$) noise.
- Week 2: we release:
 - Number of “yes” responses from Democrats + Laplace($1/\epsilon_2$) noise.
 - Number of “yes” responses from Republicans + Laplace($1/\epsilon_2$) noise.
 - Sensitivity is 1, so week 2 release satisfies ϵ_2 -differential privacy.
- Surely there is some combined privacy leakage?
 - This is called composition.
 - By itself, Week 1 satisfies ϵ_1 -differential privacy (privacy loss = ϵ_1).
 - By itself, Week 2 satisfies ϵ_2 -differential privacy (privacy loss = ϵ_2).
 - The combined release (Week 1 and Week 2) satisfies $(\epsilon_1 + \epsilon_2)$ -differential privacy (privacy loss = $\epsilon_1 + \epsilon_2$).
 - Hence ϵ is the privacy loss budget.
- In general:
 - If M_1, M_2, \dots, M_k satisfies differential privacy with parameters $\epsilon_1, \dots, \epsilon_k$, respectively
 - Mechanism M : $M(D)$ releases $M_1(D), M_2(D), \dots, M_k(D)$ satisfies $(\sum_{i=1}^k \epsilon_i)$ -differential privacy.

Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are apriori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$

Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are a priori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
- Attempt #1: Laplace Mechanism
 - Sensitivity $\Delta_f = 115$
 - So output $M(D) = \begin{bmatrix} \# \text{ average age of voters} + \text{Laplace}(115/\epsilon) \\ \# \text{ average age of non-voters} + \text{Laplace}(115/\epsilon) \end{bmatrix}$

Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are apriori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
- Attempt #1: Laplace Mechanism
 - Sensitivity $\Delta_f = 115$
 - So output $M(D) = \begin{bmatrix} \# \text{ average age of voters} + \text{Laplace}(115/\epsilon) \\ \# \text{ average age of non-voters} + \text{Laplace}(115/\epsilon) \end{bmatrix}$
 - $\text{std} \approx 163/\epsilon$.



Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are apriori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
- Attempt #2:
 - 1 Use half privacy budget for $f_1(D) = \begin{bmatrix} \text{sum ages of voters} \\ \text{sum ages of non-voters} \end{bmatrix}$
 - 2 Use half privacy budget for $f_2(D) = \begin{bmatrix} \# \text{ of voters} \\ \# \text{ of non-voters} \end{bmatrix}$
 - 3 Then divide.

Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are apriori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
- Attempt #2:
 - ① Use half privacy budget for $f_1(D) = \begin{bmatrix} \text{sum ages of voters} \\ \text{sum ages of non-voters} \end{bmatrix}$
 - $\epsilon_1 = \epsilon/2$
 - Sensitivity $\Delta_{f_1} = 115$
 - $M_1(D) = \begin{bmatrix} \text{sum ages of voters} + \text{Laplace}(115/\epsilon_1) \\ \text{sum ages of non-voters} + \text{Laplace}(115/\epsilon_1) \end{bmatrix}$

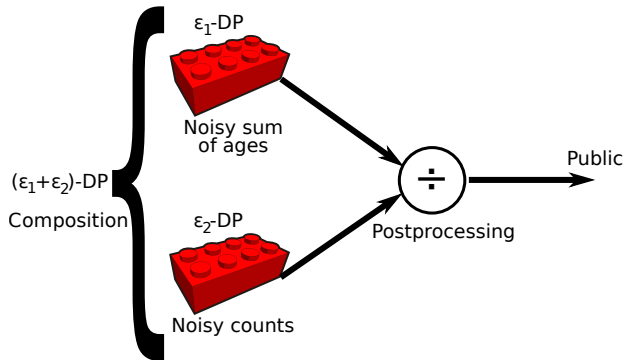
Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are a priori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
- Attempt #2:
 - ① Use half privacy budget for $f_1(D) = \begin{bmatrix} \text{sum ages of voters} \\ \text{sum ages of non-voters} \end{bmatrix}$
 - $\epsilon_1 = \epsilon/2$
 - Sensitivity $\Delta_{f_1} = 115$
 - $M_1(D) = \begin{bmatrix} \text{sum ages of voters} + \text{Laplace}(115/\epsilon_1) \\ \text{sum ages of non-voters} + \text{Laplace}(115/\epsilon_1) \end{bmatrix}$
 - ② Use half privacy budget for $f_2(D) = \begin{bmatrix} \# \text{ of voters} \\ \# \text{ of non-voters} \end{bmatrix}$
 - $\epsilon_2 = \epsilon/2$
 - Sensitivity $\Delta_{f_2} = 1$
 - $M_2(D) = \begin{bmatrix} \# \text{ of voters} + \text{Laplace}(1/\epsilon_2) \\ \# \text{ of non-voters} + \text{Laplace}(1/\epsilon_2) \end{bmatrix}$

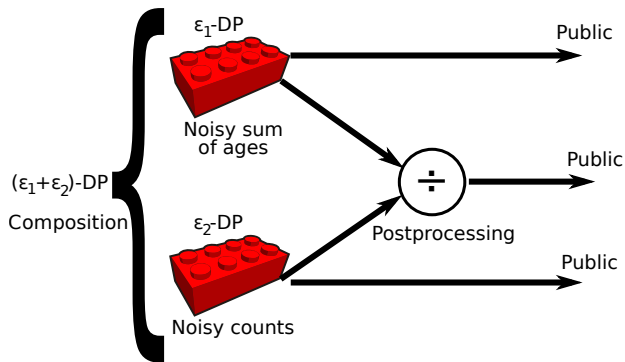
Example 1: Average

- $f(D) = \begin{bmatrix} \# \text{ average age of voters} \\ \# \text{ average age of non-voters} \end{bmatrix}$
 - Assume ages are a priori capped at 115
 - Assume $\text{avg}(\emptyset) = 0$
- Attempt #2:
 - 1 Use half privacy budget for $f_1(D) = \begin{bmatrix} \text{sum ages of voters} \\ \text{sum ages of non-voters} \end{bmatrix}$
 - $\epsilon_1 = \epsilon/2$
 - Sensitivity $\Delta_{f_1} = 115$
 - $M_1(D) = \begin{bmatrix} \text{sum ages of voters} + \text{Laplace}(115/\epsilon_1) \\ \text{sum ages of non-voters} + \text{Laplace}(115/\epsilon_1) \end{bmatrix}$
 - 2 Use half privacy budget for $f_2(D) = \begin{bmatrix} \# \text{ of voters} \\ \# \text{ of non-voters} \end{bmatrix}$
 - $\epsilon_2 = \epsilon/2$
 - Sensitivity $\Delta_{f_2} = 1$
 - $M_2(D) = \begin{bmatrix} \# \text{ of voters} + \text{Laplace}(1/\epsilon_2) \\ \# \text{ of non-voters} + \text{Laplace}(1/\epsilon_2) \end{bmatrix}$
 - 3 Then divide.
 - $\frac{\text{noisy sum of ages of voters}}{\text{noisy count of voters}}, \text{std} \approx \frac{325}{\# \text{ of voters}}$
 - $\frac{\text{noisy sum of ages of non-voters}}{\text{noisy count of non-voters}}, \text{std} \approx \frac{325}{\# \text{ of non-voters}}$

Example 1 Recap



Example 1 Recap



- Noisy measurements:
 - Noisy sum of ages (unbiased)
 - Noisy counts (unbiased)
 - Safe to release as well.
 - Should release them.

What We Learned

- Spend your privacy loss budget wisely!
 - It is easy to waste.
 - Another reason it is called a “budget”
- Carefully choose:
 - What to inject noise into.
 - How to inject the noise.
- Additional improvements possible:
 - e.g., Compute quantiles instead [Smi11].
 - e.g., Compute histograms support age ranges instead [QYL13].

Example 2: Linear Regression

- Linear regression model.
 - Data: $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)\}$
 - Each $\|\vec{x}_i\|_1 \leq C_1$.
 - Each $|y_i| \leq C_2$.
 - Model: $\vec{y} = \mathbf{X}\vec{\beta} + \xi$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

- Classical solution: $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y}$

Example 2: Linear Regression

- Linear regression model.
 - Data: $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)\}$
 - Each $\|\vec{x}_i\|_1 \leq C_1$.
 - Each $|y_i| \leq C_2$.
 - Model: $\vec{y} = \mathbf{X}\vec{\beta} + \xi$
 - Classical solution: $\vec{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y}$
- A differentially private approach:
 - 1 Set $\epsilon_1 = \epsilon_2 = \epsilon/2$.
 - 2 Compute noisy $(\mathbf{X}^T \mathbf{X})^{-1}$ using ϵ_1 budget.
 - 3 Compute noisy $\mathbf{X}^T \vec{y}$ using ϵ_2 budget.
 - 4 Model coefficients: multiply noisy $(\mathbf{X}^T \mathbf{X})^{-1}$ and noisy $\mathbf{X}^T \vec{y}$.
 - 5 Also release the noisy measurements

Example 2: Linear Regression

- Linear regression model.
 - Data: $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)\}$
 - Each $\|\vec{x}_i\|_1 \leq C_1$.
 - Each $|y_i| \leq C_2$.
 - Model: $\vec{y} = \mathbf{X}\vec{\beta} + \xi$
 - Classical solution: $\vec{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y}$
- A differentially private approach:
 - 1 Set $\epsilon_1 = \epsilon_2 = \epsilon/2$.
 - 2 Compute noisy $(\mathbf{X}^T \mathbf{X})^{-1}$ using ϵ_1 budget.
 - Sensitivity of $\mathbf{X}^T \mathbf{X}$ is C_1^2 .
 - Add independent Laplace(C_1^2/ϵ_1) noise to each element of $\mathbf{X}^T \mathbf{X}$.
 - Compute inverse.

Example 2: Linear Regression

- Linear regression model.
 - Data: $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)\}$
 - Each $\|\vec{x}_i\|_1 \leq C_1$.
 - Each $|y_i| \leq C_2$.
 - Model: $\vec{y} = \mathbf{X}\vec{\beta} + \xi$
 - Classical solution: $\vec{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y}$
- A differentially private approach:
 - 1 Set $\epsilon_1 = \epsilon_2 = \epsilon/2$.
 - 2 Compute noisy $(\mathbf{X}^T \mathbf{X})^{-1}$ using ϵ_1 budget.
 - Sensitivity of $\mathbf{X}^T \mathbf{X}$ is C_1^2 .
 - Add independent Laplace(C_1^2/ϵ_1) noise to each element of $\mathbf{X}^T \mathbf{X}$.
 - Compute inverse.
 - 3 Compute noisy $\mathbf{X}^T \vec{y}$ using ϵ_2 budget.
 - Sensitivity of $\mathbf{X}^T \vec{y}$ is $C_1 C_2$
 - Add independent Laplace($C_1 C_2/\epsilon_2$) noise to each element of $\mathbf{X}^T \vec{y}$.

Example 2: Linear Regression

- Linear regression model.
 - Data: $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)\}$
 - Each $\|\vec{x}_i\|_1 \leq C_1$.
 - Each $|y_i| \leq C_2$.
 - Model: $\vec{y} = \mathbf{X}\vec{\beta} + \xi$
 - Classical solution: $\vec{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y}$
- A differentially private approach:
 - 1 Set $\epsilon_1 = \epsilon_2 = \epsilon/2$.
 - 2 Compute noisy $(\mathbf{X}^T \mathbf{X})^{-1}$ using ϵ_1 budget.
 - Sensitivity of $\mathbf{X}^T \mathbf{X}$ is C_1^2 .
 - Add independent Laplace(C_1^2/ϵ_1) noise to each element of $\mathbf{X}^T \mathbf{X}$.
 - Compute inverse.
 - 3 Compute noisy $\mathbf{X}^T \vec{y}$ using ϵ_2 budget.
 - Sensitivity of $\mathbf{X}^T \vec{y}$ is $C_1 C_2$
 - Add independent Laplace($C_1 C_2/\epsilon_2$) noise to each element of $\mathbf{X}^T \vec{y}$.
 - 4 Model coefficients: multiply noisy $(\mathbf{X}^T \mathbf{X})^{-1}$ and noisy $\mathbf{X}^T \vec{y}$.
 - 5 Also release the noisy measurements
 - noisy $(\mathbf{X}^T \mathbf{X})$
 - noisy $\mathbf{X}^T \vec{y}$

Outline

- 1 Basic Design of Differential Privacy Mechanisms
- 2 The Role of Strategy
- 3 Chi-Squared Testing
- 4 Takeaway Messages
- 5 Common Pitfalls
- 6 Additional Mechanisms

Flexibility

- Differential privacy can be used to:
 - Obtain noisy sub-population totals.
 - Build generalized linear models [CMS11] with confidence intervals [WKL19].
 - Train deep learning models [ACG⁺16].
 - Create synthetic data [LHR⁺10, HLM12].
- Common properties: clever noise strategies.

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 1: add noise to X and Y

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 1: add noise to X and Y
- Sensitivity:
 - For any database, adding/removing one person can
 - Change X by ± 1 .
 - Change Y by ± 1 .
 - Total change at most 2
 - Sensitivity Δ : 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 1: add noise to X and Y
- Sensitivity:
 - For any database, adding/removing one person can
 - Change X by ± 1 .
 - Change Y by ± 1 .
 - Total change at most 2
 - Sensitivity Δ : 2
- Noisy Counts (Measure):
 - $\tilde{X} = X + \text{Laplace}(2/\epsilon)$
 - $\tilde{Y} = Y + \text{Laplace}(2/\epsilon)$

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 1: add noise to X and Y
- Sensitivity:
 - For any database, adding/removing one person can
 - Change X by ± 1 .
 - Change Y by ± 1 .
 - Total change at most 2
 - Sensitivity Δ : 2
- Noisy Counts (Measure):
 - $\tilde{X} = X + \text{Laplace}(2/\epsilon)$
 - $\tilde{Y} = Y + \text{Laplace}(2/\epsilon)$
- Accuracy:
 - $\text{Var}(\tilde{X}) = 8/\epsilon^2$
 - $\text{Var}(\tilde{Y}) = 8/\epsilon^2$

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? In any database, adding/removing individual who is:
 - Neither Hispanic nor VotingAge: S and D unchanged.
 - Hispanic but not VotingAge: S changes by ± 1 , D changes by ± 1
 - Not Hispanic, is VotingAge: S changes by ± 1 , D changes by ± 1
 - Both Hispanic and VotingAge: S changes by ± 2 , D is unchanged.
 - Maximum change: 2
 - Sensitivity: 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? In any database, adding/removing individual who is:
 - **Neither Hispanic nor VotingAge: S and D unchanged.**
 - Hispanic but not VotingAge: S changes by ± 1 , D changes by ± 1
 - Not Hispanic, is VotingAge: S changes by ± 1 , D changes by ± 1
 - Both Hispanic and VotingAge: S changes by ± 2 , D is unchanged.
 - Maximum change: 2
 - Sensitivity: 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? In any database, adding/removing individual who is:
 - Neither Hispanic nor VotingAge: S and D unchanged.
 - **Hispanic but not VotingAge: S changes by ± 1 , D changes by ± 1**
 - Not Hispanic, is VotingAge: S changes by ± 1 , D changes by ± 1
 - Both Hispanic and VotingAge: S changes by ± 2 , D is unchanged.
 - Maximum change: 2
 - Sensitivity: 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? In any database, adding/removing individual who is:
 - Neither Hispanic nor VotingAge: S and D unchanged.
 - Hispanic but not VotingAge: S changes by ± 1 , D changes by ± 1
 - **Not Hispanic, is VotingAge: S changes by ± 1 , D changes by ± 1**
 - Both Hispanic and VotingAge: S changes by ± 2 , D is unchanged.
 - Maximum change: 2
 - Sensitivity: 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? In any database, adding/removing individual who is:
 - Neither Hispanic nor VotingAge: S and D unchanged.
 - Hispanic but not VotingAge: S changes by ± 1 , D changes by ± 1
 - Not Hispanic, is VotingAge: S changes by ± 1 , D changes by ± 1
 - **Both Hispanic and VotingAge: S changes by ± 2 , D is unchanged.**
 - Maximum change: 2
 - Sensitivity: 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? In any database, adding/removing individual who is:
 - Neither Hispanic nor VotingAge: S and D unchanged.
 - Hispanic but not VotingAge: S changes by ± 1 , D changes by ± 1
 - Not Hispanic, is VotingAge: S changes by ± 1 , D changes by ± 1
 - Both Hispanic and VotingAge: S changes by ± 2 , D is unchanged.
 - Maximum change: 2
 - Sensitivity: 2

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? Equals 2
- Noisy Measurements:
 - $\tilde{S} = S + \text{Laplace}(2/\epsilon)$
 - $\tilde{D} = D + \text{Laplace}(2/\epsilon)$

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? Equals 2
- Noisy Measurements:
 - $\tilde{S} = S + \text{Laplace}(2/\epsilon)$
 - $\tilde{D} = D + \text{Laplace}(2/\epsilon)$
- Reconstruction (postprocessing):
 - $\tilde{X} = (\tilde{S} + \tilde{D})/2$
 - $\tilde{Y} = (\tilde{S} - \tilde{D})/2$

Strategy Example

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 2:
 - Add noise to $S = X + Y$ (Hispanic + VotingAge)
 - Add noise to $D = X - Y$ (Hispanic - VotingAge)
 - Note: not very intuitive quantities.
- Sensitivity? Equals 2
- Noisy Measurements:
 - $\tilde{S} = S + \text{Laplace}(2/\epsilon)$
 - $\tilde{D} = D + \text{Laplace}(2/\epsilon)$
- Accuracy:
 - $\text{Var}(\tilde{S}) = 8/\epsilon^2$
 - $\text{Var}(\tilde{D}) = 8/\epsilon^2$
 - $\text{Var}(\tilde{X}) = 4/\epsilon^2$
 - $\text{Var}(\tilde{Y}) = 4/\epsilon^2$
- Reconstruction (postprocessing):
 - $\tilde{X} = (\tilde{S} + \tilde{D})/2$
 - $\tilde{Y} = (\tilde{S} - \tilde{D})/2$

Summary

- In a given region, suppose we are interested in:
 - $X = \#$ of Hispanic individuals
 - $Y = \#$ of VotingAge individuals
- What do we add noise to?
- Attempt 1:
 - Add noise to X
 - Add noise to Y
 - Variance: $8/\epsilon^2$
- Attempt 2:
 - Add noise to $X + Y$
 - Add noise to $X - Y$
 - Reconstruct
 - Variance: $4/\epsilon^2$
- Select-Measure-Reconstruct Paradigm [LHR⁺10].
- What you want is not always what you should add noise to.

Outline

- 1 Basic Design of Differential Privacy Mechanisms
- 2 The Role of Strategy
- 3 Chi-Squared Testing
- 4 Takeaway Messages
- 5 Common Pitfalls
- 6 Additional Mechanisms

Differential Privacy and Gaussian Noise

- There are versions of differential privacy compatible with Gaussian noise.
 - Approximate differential privacy [DKM⁺06]
 - zCDP [BS16]
 - Renyi Differential Privacy [Mir17]
- Privacy semantics are harder to understand.
- Noise (Gaussian) is easier to understand.
 - Noise scale depends on L_2 sensitivity $\Delta_f^{(2)}$.
 - $\Delta_f^{(2)} = \sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_2$

Classical Chi-Squared Test

- Chi-Squared Tests

- Goodness of fit.
- Test of sample proportions.
- Test of independence.

- Test statistic: $T = \sum_{i=1}^k \frac{(X_i - E_i)^2}{E_i}$

- X_i : number of people of type i

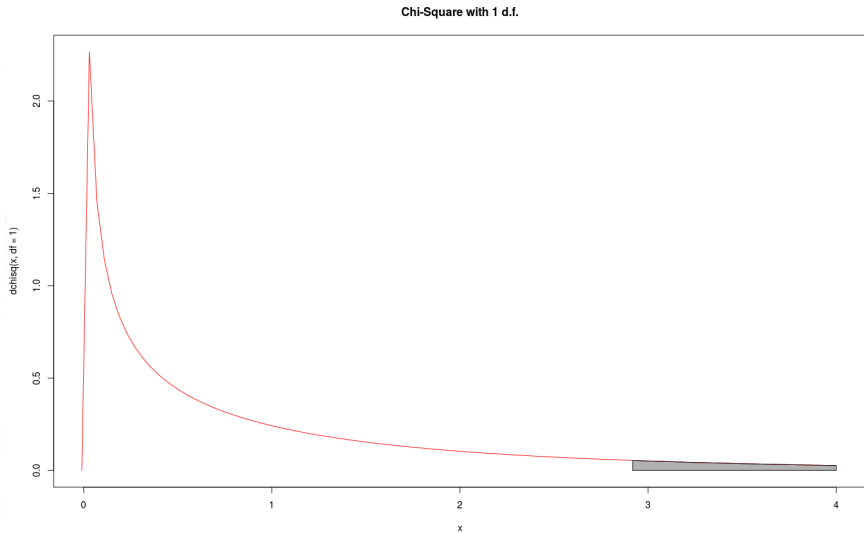
X_1	X_2	\cdots	X_{k-1}	X_k
-------	-------	----------	-----------	-------

- E_i : expected number of people of type i under null hypothesis.

- Asymptotically:

- T has χ_τ^2 distribution.
- τ is degrees of freedom (depends on how E_i is estimated)

Chi-Squared Tails



Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$

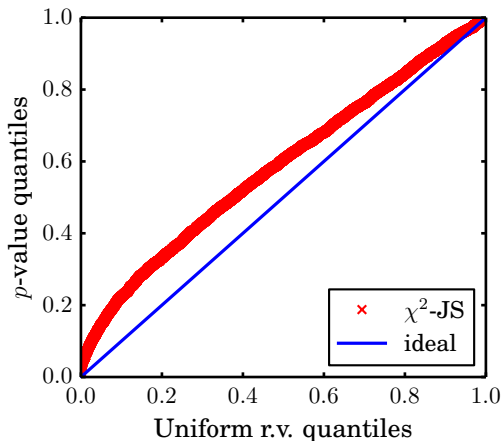
Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 1: pretend \tilde{X}_i are the real data. $T = \sum_{i=1}^k \frac{(\tilde{X}_i - \tilde{E}_i)^2}{\tilde{E}_i}$
 - Run standard chi-squared test on $\tilde{X}_1, \dots, \tilde{X}_k$
 - Reject if p -value below α .

QQ Plot for Attempt 1

- Red: sampling distribution under null hypothesis.
- Blue: ideal behavior for valid p-values.



Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 1: pretend \tilde{X}_i are the real data. $T = \sum_{i=1}^k \frac{(\tilde{X}_i - \tilde{E}_i)^2}{\tilde{E}_i}$
 - Run standard chi-squared test on $\tilde{X}_1, \dots, \tilde{X}_k$
 - Reject if p -value below α .
- In this procedure, added noise:
 - does not change underlying phenomena (fit, independence, etc. of original data)
 - tends to make test statistic larger.
 - “p-values” appear smaller
 - leads to increased false discovery

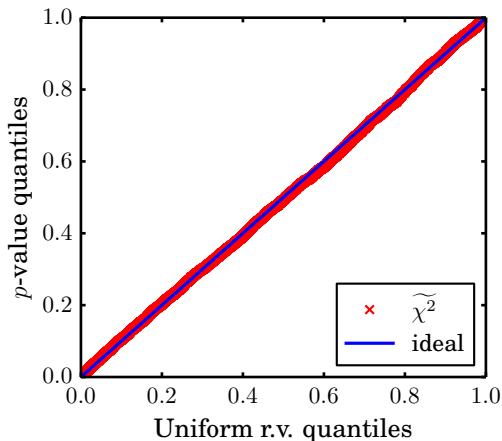
Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 2:
 - Re-use noisy data in test statistic. $T = \sum_{i=1}^k \frac{(\tilde{X}_i - \tilde{E}_i)^2}{\tilde{E}_i}$.
 - Estimate sampling distribution of T more accurately.

QQ Plot for Attempt 2

- Red: sampling distribution under null hypothesis.
- Blue: ideal behavior for valid p-values.



Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 2:
 - Re-use noisy data in test statistic. $T = \sum_{i=1}^k \frac{(\tilde{X}_i - \tilde{E}_i)^2}{\tilde{E}_i}$.
 - Estimate sampling distribution of T more accurately.
- p-values are valid.
- Are we done?

Testing with Differential Privacy

- Data:

X_1	X_2	\cdots	X_{k-1}	X_k
-------	-------	----------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 3:
 - Sampling distribution (under null) from prior attempts is not approximately chi-squared.
 - Is there a test statistic over the \tilde{X}_i that is?

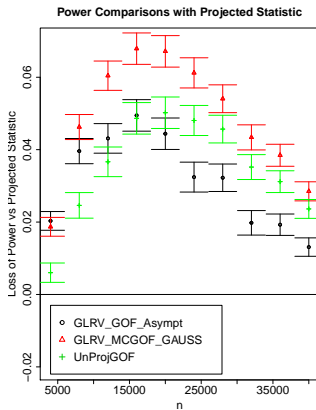
Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 3:
 - Sampling distribution (under null) from prior attempts is not approximately chi-squared.
 - Is there a test statistic over the \tilde{X}_i that is?
 - Yes! [RK17] (projected statistic)
 - Appears to be more powerful.

Projected Statistic

- Loss of power of other statistics compared to projected statistic [RK17].



Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 3:
 - Sampling distribution (under null) from prior attempts is not approximately chi-squared.
 - Is there a test statistic over the \tilde{X}_i that is?
 - Yes! [RK17] (projected statistic)
 - Appears to be more powerful.
 - **Now are we done?**

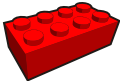
Testing with Differential Privacy

- Data:

X_1	X_2	\dots	X_{k-1}	X_k
-------	-------	---------	-----------	-------
- Suppose we are given noisy measurements.
 - Added Gaussian Noise.
 - Scale depends on privacy parameters.
 - $\tilde{X}_1 = X_1 + N(0, \sigma^2)$
 - $\tilde{X}_2 = X_2 + N(0, \sigma^2)$
 - \vdots
 - $\tilde{X}_k = X_k + N(0, \sigma^2)$
- Attempt 3:
 - Sampling distribution (under null) from prior attempts is not approximately chi-squared.
 - Is there a test statistic over the \tilde{X}_i that is?
 - Yes! [RK17] (projected statistic)
 - Appears to be more powerful.
 - Now are we done?
 - **We could pick a better noise distribution! [AS20]**

Outline

- 1 Basic Design of Differential Privacy Mechanisms
- 2 The Role of Strategy
- 3 Chi-Squared Testing
- 4 Takeaway Messages
- 5 Common Pitfalls
- 6 Additional Mechanisms

- Differential Privacy is like  .
- Also like spending money.
 - Easy to waste privacy loss budget without “financial” planning.
 - Where do you add the noise?
 - What do you do after the noise?
 - Accurate tracking of total privacy cost [Mir17, BW18].

Takeaway Message II

- Differentially private algorithms produce many data products:
- e.g.,
 - Intermediate noisy measurements (safe to release)
 - Synthesized microdata (safe to release)
 - Source code (safe to release)
- Demystified:
 - Noisy measurements are often just counts + noise
 - Noise is often unbiased
 - Variance and distribution are known

Thank You



Outline

- 1 Basic Design of Differential Privacy Mechanisms
- 2 The Role of Strategy
- 3 Chi-Squared Testing
- 4 Takeaway Messages
- 5 Common Pitfalls
- 6 Additional Mechanisms

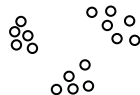
Normalizing Data

- Data normalization and feature selection prior to model fitting.
- In a dataset of Age, Weight, Height, Income:
 - Subtract off the mean age, divide by std of ages.
 - Subtract off mean weight, divide by std of weights.
 - Subtract off mean height, divide by std of heights.
 - Subtract off mean income, divide by std of income.
- This affects sensitivity: adding/removing 1 record can affect entire dataset.
 - Adding 1 billionaire can throw off mean and standard deviation.
 - Most of the normalized incomes are near 0.
 - Causes sensitivity to equal n , number of records.
- Better: use some privacy budget for:
 - normalization.
 - feature selection
- Suggestion: use robust statistical models.

Non-Numerical Operations

- Operations that don't return numbers still affect sensitivity.

- Consider reporting the results of a clustering.

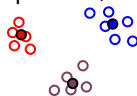


- Cluster the data



- Within each cluster compute the sum of points + Laplace noise
- Within each cluster compute the number of points + Laplace noise

- Divide, to get approximate cluster centers



- Publish these cluster centers.

- Common mistake: forgetting to use differential privacy in the initial clustering.
 - Adding 1 record can result in a completely different clustering.
 - Instead, use a differentially private clustering algorithm (e.g., [McS09]).

Neighbors I

- Bounded differential privacy:
 - Neighbors D_1, D_2 differ on value of one record.
 - n (# of respondents) comes for free.
 - n is the same for all records.
- Unbounded differential privacy:
 - Neighbors D_1, D_2 differ on presence/absence of one record.
 - D_1 and D_2 have different number of respondents.
 - n is not free. If you need it, use privacy budget to get an approximate value.

Neighbors 2

- Must consider all possible D_1 and D_2 that are neighbors of each other.
- Common mistake: only considering neighbors of current database.
- Example database of ages capped at 115:
 $D^* = \{1, 2, 35, 36, 36, 99, 115\}$
 - What is sensitivity of the median?
 - Adding/removing 1 record for this dataset changes median by at most 1.
 - 1 is not the sensitivity.
 - Consider $D_1 = \{0, 0, 0, 115, 115\}$, $D_2 = \{0, 0, 0, 115, 115, 115\}$
 - So sensitivity is $115/2$.
 - More advanced techniques add less noise when median is stable (like in D^*)
 - Smooth sensitivity [NRS07].
 - Private quantiles and Exponential Mechanism [Smi11, MT07].

Outline

- 1 Basic Design of Differential Privacy Mechanisms
- 2 The Role of Strategy
- 3 Chi-Squared Testing
- 4 Takeaway Messages
- 5 Common Pitfalls
- 6 Additional Mechanisms

Additional Mechanisms

- Exponential mechanism [DR14, MT07]
- Noisy Max [DR14, DWZK19]
- Sparse Vector [DR14, DWZK19]
- Smooth Sensitivity [NRS07]
- Example usage: [HLM12]

Basic Mechanism Comparisons

- Pure differential privacy (no δ).
 - L_1 Sensitivity $\Delta_f^{(1)}$: $\sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1$
 - Laplace mechanism. Noise scale: β .
 - Privacy is a function of $\Delta_f^{(1)}/\beta$ (this equals ϵ).
- Approximate differential privacy
 - L_2 Sensitivity $\Delta_f^{(2)}$: $\sup_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_2$
 - Laplace mechanism. Noise scale: β .
 - Privacy (ϵ, δ curve) is a function of $\Delta_f^{(2)}/\sigma$ [BW18]

References I



Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang.
Deep learning with differential privacy.
In CCS, 2016.



Jordan Alexander Awan and Aleksandra Slavkovic.
Differentially private inference for binomial data.
Journal of Privacy and Confidentiality, 10(1), Jan. 2020.



Mark Bun and Thomas Steinke.
Concentrated differential privacy: Simplifications, extensions, and lower bounds.
In Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985, 2016.

References II



Borja Balle and Yu-Xiang Wang.

Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising.

In [Proceedings of the 35th International Conference on Machine Learning, 2018](#).



Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate.

Differentially private empirical risk minimization.

[J. Mach. Learn. Res.](#), 12:1069–1109, 2011.






Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor.

Our data, ourselves: Privacy via distributed noise generation.

In [EUROCRYPT](#), pages 486–503, 2006.

References III

-  Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In [TCC](#), 2006.
-  Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. [Foundations and Trends in Theoretical Computer Science](#), 9(3–4):211–407, 2014.
-  Zeyu Ding, Yuxin Wang, Danfeng Zhang, and Daniel Kifer. Free gap information from the differentially private sparse vector and noisy max mechanisms. [Proc. VLDB Endow.](#), 13(3):293–306, November 2019.

References IV



Moritz Hardt, Katrina Ligett, and Frank McSherry.

A simple and practical algorithm for differentially private data release.

In [NIPS](#), 2012.



Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor.

Optimizing linear counting queries under differential privacy.

In [PODS](#), 2010.



Frank D. McSherry.

Privacy integrated queries: An extensible platform for privacy-preserving data analysis.

In [SIGMOD](#), pages 19–30, 2009.

References V



Ilya Mironov.

Rényi differential privacy.

In 30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21-25, 2017, pages 263–275, 2017.



Frank McSherry and Kunal Talwar.

Mechanism design via differential privacy.

In Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, pages 94–103, 2007.



Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith.

Smooth sensitivity and sampling in private data analysis.

In STOC, pages 75–84, 2007.

References VI



Wahbeh Qardaji, Weining Yang, and Ninghui Li.

Understanding hierarchical methods for differentially private histograms.

[Proc. VLDB Endow.](#), 6(14), 2013.



Ryan Rogers and Daniel Kifer.

A new class of private chi-square hypothesis tests.

In [Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, \(AISTATS\)](#), 2017.



Adam Smith.

Privacy-preserving statistical estimation with optimal convergence rates.

In [STOC](#), 2011.

References VII



YUE WANG, DANIEL KIFER, and JAEWOO LEE.

Differentially private confidence intervals for empirical risk minimization.

[Journal of Privacy and Confidentiality](#), 2019.