CSCE 658: Randomized Algorithms

Lecture 21

Samson Zhou

Relevant Supplementary Material

- Chapter 3-4 of "The Algorithmic Foundations of Differential Privacy", by Cynthia Dwork and Aaron Roth (https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf)

Last Time: Differential Privacy

- [DMNS06] Given $\varepsilon > 0$ and $\delta \in (0,1)$, a randomized algorithm $A: U^* \to Y$ is (ε, δ)-differentially private if, for every neighboring pair D and D' of datasets, and for all $E \subseteq Y$,
$$\Pr[A(D) \in E] \leq e^{\varepsilon} \cdot \Pr[A(D') \in E] + \delta$$

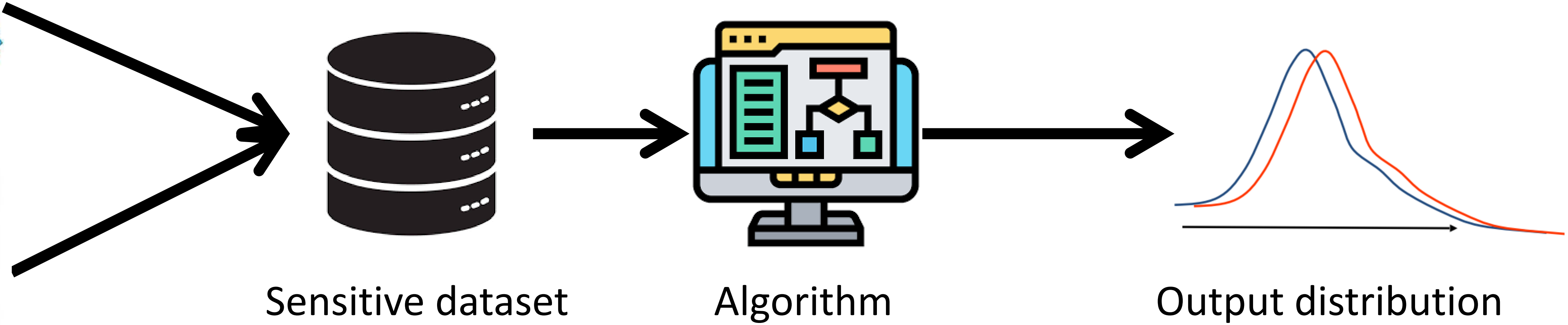
$A($  $) \approx A($  $)$

Last Time: Differential Privacy

- [DMNS06] Given $\varepsilon > 0$ and $\delta \in (0,1)$, a randomized algorithm $A: U^* \to Y$ is (ε, δ)-differentially private if, for every neighboring pair D and D' of datasets, and for all $E \subseteq Y$,

$$\Pr[A(D) \in E] \leq e^{\varepsilon} \cdot \Pr[A(D') \in E] + \delta$$





Last Time: Differential Privacy

- [DMNS06] Given $\varepsilon > 0$ and $\delta \in (0,1)$, a randomized algorithm $A: U^* \to Y$ is (ε, δ)-differentially private if, for every neighboring pair D and D' of datasets, and for all $E \subseteq Y$,
$$\Pr[A(D) \in E] \leq e^{\varepsilon} \cdot \Pr[A(D') \in E] + \delta$$

- Implication: Deterministic algorithms cannot be differentially private unless they are a constant function

Last Time: Local Differential Privacy (LDP)

- [KLNRS08] Given $\varepsilon > 0$ and $\delta \in (0,1)$, a randomized algorithm $A: U^* \to Y$ is (ε, δ)-differentially private if, for every pairs of users' possible data x and x' and for all $E \subseteq Y$,

$$\Pr[A(x) \in E] \le e^{\varepsilon} \cdot \Pr[A(x') \in E] + \delta$$

- Algorithm takes a single user's data
- Compared to previous definition of DP, where algorithm takes all users' data

Last Time: Randomized Response, Revisited

- How many people in this class have a pet?

- Generate a random integer:
  - If it is even, answer truthfully
  - Otherwise, proceed below
- Generate another random integer:
  - If it is even, answer YES
  - Otherwise if it is odd, answer NO

Last Time: Randomized Response, Revisited

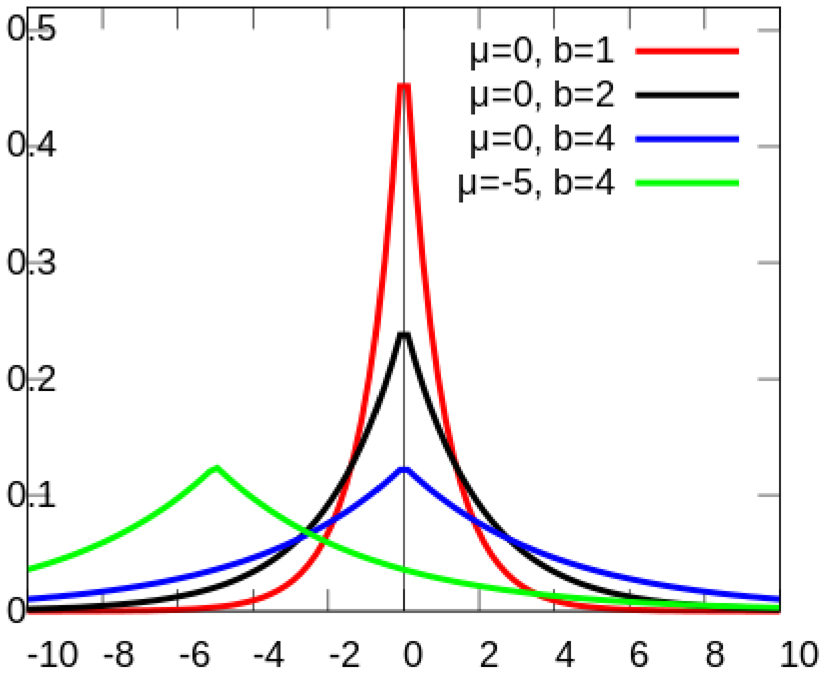
- Answer is correct in expectation, but what is its variance?
- Each answer is incorrect with probability $\frac{1}{4}$
- The variance is $O(n)$

- By Chebyshev, we have additive error $O(\sqrt{n})$ with probability 0.99
- By anti-concentration, error is $\Omega(\sqrt{n})$

Last Time: Laplace Mechanism

- Goal: Algorithm computes $f(x)$ and releases $f(x) + Z$, where $Z \sim \mathrm{Lap}\left(\frac{\sigma_f}{\varepsilon}\right)$
- Laplacian distribution: Probability density function for $\mathrm{Lap}(b)$ is

$$p(x) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right) = \frac{1}{2b}e^{\left(-\frac{|x|}{b}\right)}$$



Laplace Mechanism

- Theorem: Laplace mechanism is ε-differentially private (pure DP)

- What is the algorithm for the private counting problem by the Laplace mechanism?
- What is the error for the private counting problem by the Laplace mechanism?

Beyond Laplace Mechanism

- How do we answer non-numeric queries, e.g., selection?

- Example: What is the most common eye color in the room?

Beyond Laplace Mechanism

- Example: Suppose a study is conducted that finds the current location of individuals, in the two-dimensional plane
- Who is the closest individual to a query location?

Beyond Laplace Mechanism

- What if we want to output the “best” answer, but noise can significantly destroy the answer?

- Example: Suppose we have a large number of apples, and A, B, C each bid \$1.00 and D bids \$4.01. What is the optimal price?
- At \$4.01 the revenue, the revenue is \$4.01, at \$4.00 and at \$1.00 the revenue is \$4.00, but at \$3.02 the revenue is zero!

Exponential Mechanism

- Choose a score function $S: (X^n, Y) \to \mathbb{R}$ and global sensitivity σ
- Sample $y \in Y$ with probability proportional to $\exp\left(\frac{\varepsilon}{2\sigma} S(x, y)\right)$

Beyond Laplace Mechanism

- How do we answer non-numeric queries, e.g., selection?

- Example: What is the most common eye color in the room?

Beyond Laplace Mechanism

- Example: Suppose a study is conducted that finds the current location of individuals, in the two-dimensional plane

- Who is the closest individual to a query location?

Exponential Mechanism

- Theorem: Exponential mechanism is ε-differentially private (pure DP)
- Suppose the "answer" for dataset D is z and the "answer" for dataset D is z'

- Let y be the output of the exponential mechanism for D and let y' be the output of the exponential mechanism for D'

Exponential Mechanism

- Want to show that for any fixed x, $\frac{\Pr[y=x]}{\Pr[y'=x]} \leq e^{\varepsilon}$

$$\frac{\Pr[y = x]}{\Pr[y' = x]} \leq \exp\left(\frac{\varepsilon S(D, x)}{2\sigma_f}\right) / \sum_x \exp\left(\frac{\varepsilon S(D, x)}{2\sigma_f}\right)$$

$$\cdot \sum_x \exp\left(\frac{\varepsilon S(D', x)}{2\sigma_f}\right) / \exp\left(\frac{\varepsilon S(D', x)}{2\sigma_f}\right)$$

Exponential Mechanism

- Want to show that for any fixed x, $\frac{\Pr[y=x]}{\Pr[y'=x]} \leq e^{\varepsilon}$

$$\frac{\Pr[y = x]}{\Pr[y' = x]} \leq \exp\left(\frac{\varepsilon(S(D,x) - S(D',x))}{2\sigma_f}\right)$$

$$\cdot \sum_x \exp\left(\frac{\varepsilon S(D,x)}{2\sigma_f}\right) / \sum_x \exp\left(\frac{\varepsilon S(D',x)}{2\sigma_f}\right)$$

Exponential Mechanism

- Want to show that for any fixed x, $\frac{\Pr[y=x]}{\Pr[y'=x]} \leq e^{\varepsilon}$

$$\frac{\Pr[y = x]}{\Pr[y' = x]} \leq \exp\left(\frac{\varepsilon}{2}\right)$$

$$\cdot \exp\left(\frac{\varepsilon}{2}\right) \sum_x \exp\left(\frac{\varepsilon S(D,x)}{2\sigma_f}\right) / \sum_x \exp\left(\frac{\varepsilon S(D,x)}{2\sigma_f}\right)$$

$$= \exp(\varepsilon)$$

Exponential Mechanism

- Theorem: Exponential mechanism is ε-differentially private (pure DP)
- Note we can still apply exponential mechanism when Y is the set of the real numbers
- How does it compare to the Laplace mechanism?

Mechanisms: Exponential vs. Laplace

- Consider a query with sensitivity σ_f
- Suppose the "answer" for dataset D is z and the "answer" for dataset D is z'
- Let $y = z + \mathrm{Lap}\left(\frac{\sigma_f}{\varepsilon}\right)$
- $p(x) = \frac{\varepsilon}{2\sigma_f} \exp\left(-\frac{\varepsilon|x|}{\sigma_f}\right) = \frac{\varepsilon}{2\sigma_f} e^{\left(-\frac{\varepsilon|x|}{\sigma_f}\right)}$ for a Laplace distribution with scale parameter $\frac{\sigma_f}{\varepsilon}$

Mechanisms: Exponential vs. Laplace

- Laplace mechanism: Output $y = z + x$ with probability proportional to $\frac{\varepsilon}{2\sigma_f} e^{\left(-\frac{\varepsilon|x|}{\sigma_f}\right)}$

Mechanisms: Exponential vs. Laplace

- Consider a query with sensitivity σ_f
- Suppose the “answer” for dataset D is z and the “answer” for dataset D is z'
- Choose score function $-2|y - x|$
- Exponential mechanism: Output $y = z + x$ with probability proportional to $e^{\left(-\frac{\varepsilon|x|}{\sigma_f}\right)}$

Mechanisms: Exponential vs. Laplace

- Laplace mechanism: Output $y = z + x$ with probability proportional to $\frac{\varepsilon}{2\sigma_f} e^{\left(-\frac{\varepsilon|x|}{\sigma_f}\right)}$

- Exponential mechanism: Output $y = z + x$ with probability proportional to $e^{\left(-\frac{\varepsilon|x|}{\sigma_f}\right)}$
- Recovers the Laplace mechanism!

Exponential Mechanism Drawbacks

- Sampling process may be inefficient

- Error can be large