# *Duff*: A Dataset-Based Utility Function Family for the Exponential Mechanism

Google Research

Andrés Muñoz Medina
Jennifer Gillenwater

## Private data analysis

- Unprecedented access to data
- Data driven scientific progress
- Release of data puts user privacy at risk
- Must handle data with care

## Differential Privacy

- Framework for protecting user information
- Outcome of a private mechanism does not depend on an individual's record
- Applications releasing simple statistics:mean, median, ... and learning ML models
- Variety of algorithms
- Hard to select the optimal

**Definition:** We say two datasets $D, D'$ are neighbors if they differ on a single user.

**Definition:** A mechanism $M$ is said to be $(\epsilon, \delta)$-differentially private if for all neighboring datasets

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

## Laplace Mechanism

- Let $f \colon D \mapsto f(D) \in \mathbb{R}$ be a statistic
- A private version of $f$ is given by

$$f(D) + \frac{GS}{\epsilon} Z$$

where $Z \sim \mathrm{Lap}(1)$ and $GS = \max_{D,D'} |f(D) - f(D')|$ is the **global sensitivity** of $f$

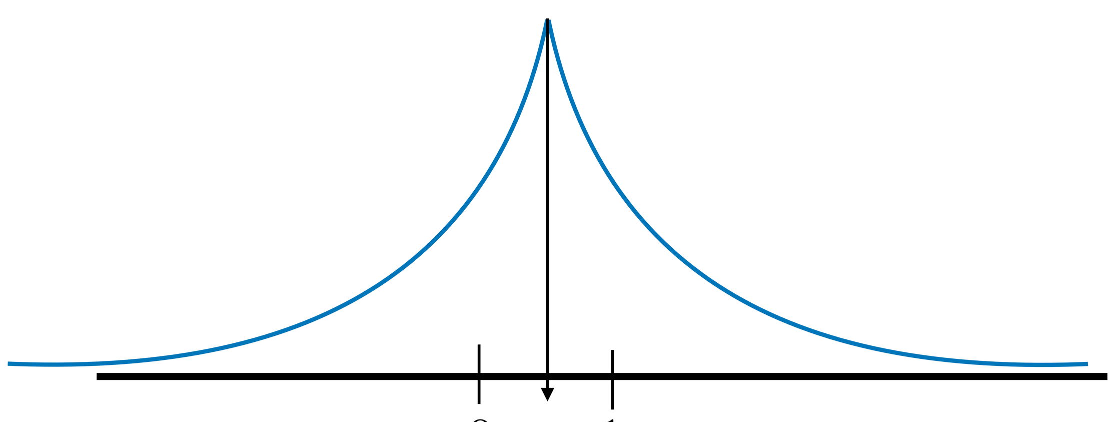
- General mechanism for differential privacy
- Pessimistic: can add too much noise

## Private Medians

- Calculate the median of n numbers in [0,1]

| | | |
|---|---|---|
| $D = \{0,0,0,1,1\}$<br>$Median(D) = 0$ | $GS = 1$ | $D' = \{0,0,1,1,1\}$<br>$Median(D') = 1$ |

- What if $D = \{0.5, \ldots, 0, 5\}$ ?




## Smooth sensitivity

**Definition:** The local sensitivity of a function is given by

$$LS_f(D) = \max_{D'} |f(D) - f(D')|$$

**Definition:** The smooth sensitivity of a function is given by

$$SS_f(\beta, D) = \max_{k} \max_{D' \colon d(D', D) = k} e^{-k\beta} LS_f(D')$$

The mechanism

$$f(D) + \frac{SS_f(\beta, D)}{\alpha} Z$$

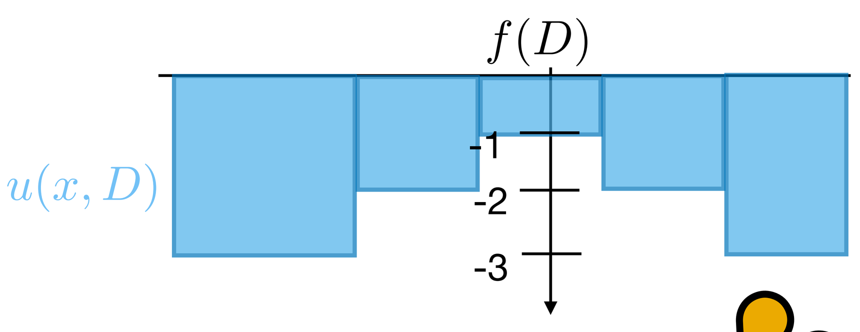
where $Z \sim Lap(1)$ is $(\epsilon, \delta)$-differentially private for some appropriate $\alpha, \beta$.

## Exponential mechanism

- Given a utility function $u \colon \mathbb{R} \times \mathcal{D} \to \mathbb{R}$
- Let $\Delta_u = \max_{x,D,D'} |u(x,D) - u(x,D')|$ denote its sensitivity
- The exponential mechanism samples $x$ with probability proportional to $e^{\frac{\epsilon u(x,D)}{\Delta_u}}$
- Always $(\epsilon, 0)$-differentially private
- Quality depends highly on the utility function

## *Duff*

- Define utility based on distance between datasets $u(x, D) = -\min_{D' \colon f(D') = x} d(D, D')$
- Sensitivity is always 1

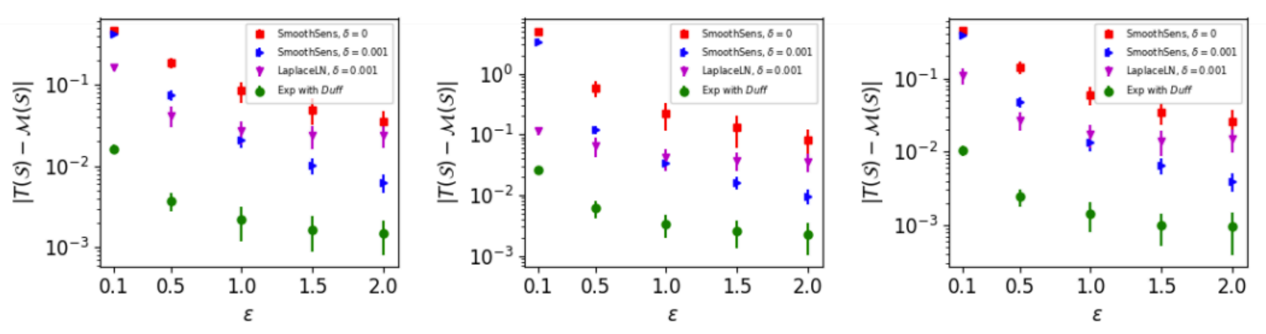



**Theorem:** The output $x$ of the exponential mechanism with Duff satisfies w.p $1 - \delta'$

$$|x - f(D)| \leq \frac{SS_f(O(\epsilon))}{\epsilon} \log(1/\delta')$$

- First connection between the exponential mechanism and smooth sensitivity
- Shows that one can achieve exponential decay with noise scaled by smooth sensitivity and $(\epsilon, 0)$-differential privacy. Before, only $(\epsilon, \delta)$-differential privacy was achievable.

### Experiments



**Up to 10x improvement over state-of-the-art**