Approximate DP

Florian Tramèr

Recap: answering k counting queries with ε-DP can be done with $\tilde{O}(k/\varepsilon n)$ error

This is tight!

To do better: we must relax the definition

ε-DP asks that for all events the probability changes by at most e^{ε} for neighboring databases

Do we really need this for an event w. probability 2^{-128}

ex:

	D	D'
event S	2^{-128}	0

This is not DP!

<u>Def</u>: Approximate DP

an algo $\mathcal{M}: X^n \to Y$ is (ε, δ)-DP if for all $D \sim D'$ and all events $S \subseteq Y$

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

<u>Observations</u>:
- ε-DP $\Leftrightarrow$ $(\varepsilon, 0)$-DP
- δ can be interpreted as a failure probability of DP
- ideally δ would be "cryptographically small" e.g. 2^{-128}
- For pure DP, we could look at $P[\mathcal{M}(D)=y] \leq e^{\varepsilon} P[\mathcal{M}(D')=y]$

  for approx. DP this is no longer true

Properties

- post processing
- group privacy $(k\varepsilon,$ "something weird"$)$-DP

Composition (basic):

Let $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_k)$ be a sequence of algos where $\mathcal{M}_i$ is $(\varepsilon_i, \delta_i)$-DP then $\mathcal{M}$ is $(\sum \varepsilon_i, \sum \delta_i)$-DP.

Def: Privacy loss random variable (PLRV)

let D, D' be two DBs and $\mathcal{M}$ a mechanism.

The PLRV $\mathcal{L}_{M(D) \| M(D')}$ is distributed by drawing $y \leftarrow M(D)$ and outputting

$$\mathcal{L}_{M(D) \| M(D')}(y) = \ln\left(\frac{\Pr[M(D) = y]}{\Pr[M(D') = y]}\right)$$

Why care?

$\mathcal{M}$ is (ε, δ)-DP if $\Pr\left[\left| \mathcal{L}_{M(D) \| M(D')} \right| > \varepsilon \right] \le \delta$

Gaussian Mechanism

Def: ℓ_2-sensitivity of a function $f: X^n \to \mathbb{R}^k$

is $\Delta_2 = \max_{D \sim D'} \underbrace{\| f(D) - f(D') \|_2}_{= \sqrt{\sum (f(D)_i - f(D')_i)}}$

Fact: $\Delta_2 \le \Delta_1 \le \sqrt{k} \cdot \Delta_2$

e.g. $v = [1, 1, \ldots, 1]$ $\quad \|v\|_1 = k$

$\|v\|_2 = \sqrt{k}$

Def: Gaussian distribution

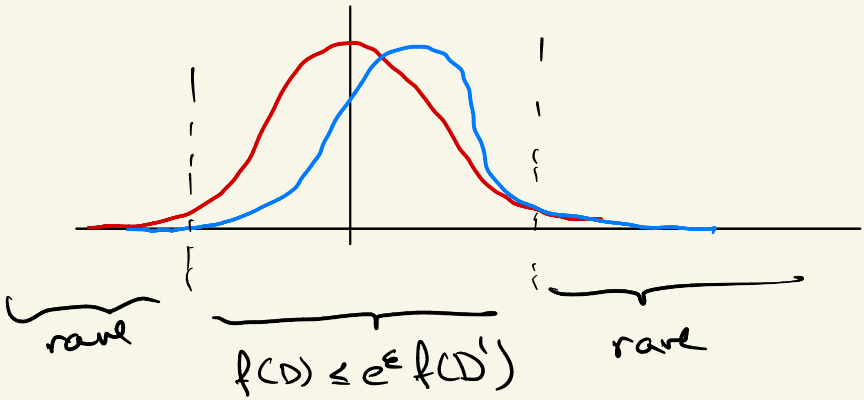
$N(\mu, \sigma^2)$ has density $f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Def: Gaussian mechanism for a function f with sensitivity Δ_2

$M(D) = f(D) + (Y_1, \ldots, Y_k)$

where Y_i are iid samples from $N\left(0, \frac{2 \log\left(\frac{1.25}{\delta}\right) \Delta_2^2}{\varepsilon^2}\right)$

Thm: for $\varepsilon \le 1$, and $\delta > 0$ the Gaussian mechanism is (ε, δ)-DP



Proof (informal)

suppose $f(D) = 0$
$f(D') = \Delta_2$

$$Y \leftarrow M(D), \quad Y \sim N(0, \sigma^2)$$

$$L_{M(D)\|M(D')}(Y) = \ln\left(\frac{\Pr[M(D) = Y]}{\Pr[M(D') = Y]}\right)$$

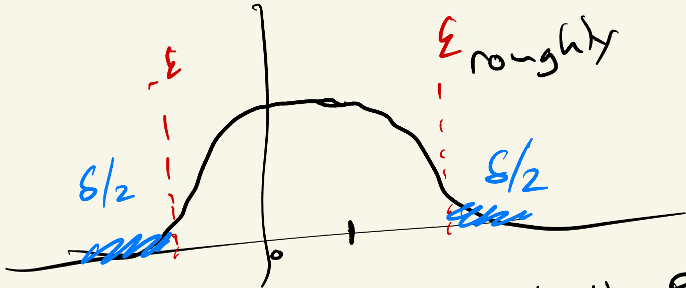
should be PDF

$$= \ln\left(\frac{\exp(-Y^2/2\sigma^2)}{\exp\left(-\frac{(Y-\Delta_2)^2}{2\sigma^2}\right)}\right) = \frac{1}{2\sigma^2}\left((Y-\Delta_2)^2 - Y^2\right)$$

$$= \frac{1}{2\sigma^2}\left(-2Y\Delta_2 + \Delta_2^2\right)$$

The PLRV is a Gaussian!

$$N\left(\frac{\Delta_2^2}{2\sigma^2}, \frac{\Delta_2^2}{\sigma^2}\right)$$



we want to show that the PLRV is larger than ε w.p less than δ

How? Tail bounds for the Gaussian

If $Z \sim N(0, \sigma^2)$ then for all $t > 0$

$$\Pr\left[|Z| > t\sigma\right] \leq 2\exp\left(\frac{-t^2}{2}\right)$$

we show is a weaker result:

$$\Pr\left[\left|N\left(0, \frac{\Delta^2}{\sigma^2}\right)\right| > \varepsilon\right]$$

$2e^{-t^2/2}$

$$= \Pr_{Y \sim N\left(0, \underbrace{\frac{\Delta^2}{\sigma^2}}\right)}\left[|Y| > \underbrace{\frac{\varepsilon\sigma}{\Delta}}_{t} \cdot \underbrace{\frac{\Delta}{\sigma}}_{\text{std dev}}\right] \leq 2e^{-\frac{\varepsilon^2\sigma^2}{2\Delta_2^2}}$$

$$2e^{\frac{-\varepsilon^2\sigma^2}{2\Delta^2}} = \delta$$

$$\sigma = \sqrt{\frac{-2\ln(\delta/2)\Delta_2^2}{\varepsilon^2}} = \sqrt{2\ln(2/\delta)}\frac{\Delta}{\varepsilon}$$

Note: if we add noise $N(0, \sigma^2)$ to a function f with ℓ_2-sensitivity Δ_2 then the output is $(\varepsilon, \delta(\varepsilon))$-DP for any $\varepsilon > 0$

Counting queries with the Gaussian mechanism:

Consider k normalized counting queries $f = (f_1, \dots, f_k)$ applied to $D \in \{0,1\}^n$.

- ℓ_1-sensitivity: k/n
- ℓ_2-sensitivity: $\sqrt{k}/n$

$$\Big\| [\underbrace{\tfrac{1}{n}, \tfrac{1}{n}, \dots, \tfrac{1}{n}}_{k}] \Big\|_2 = \frac{\sqrt{k}}{n}$$

To get (ε, δ)-DP we need to add Gaussian noise of std deviation $\tilde{O}\left(\frac{\Delta_2}{\varepsilon}\right) = \tilde{O}\left(\frac{\sqrt{k}}{\varepsilon n}\right)$

$\Rightarrow$ Maximum error across all k queries is $\tilde{O}\left(\frac{\sqrt{k}}{\varepsilon n}\right)$ with high probability

Laplace: $\tilde{O}\left(\frac{k}{\varepsilon n}\right)$ error for ε-DP

Aside: Bounding the error of DP algorithms

log(1/δ) terms ignored

std deviation of noise is $\tilde{O}\left(\frac{\sqrt{k}}{\varepsilon n}\right)$

$$\text{output} = \left[f_1(D) + N(0,\sigma^2), \ \ldots, \ f_k(D) + N(0,\sigma^2)\right]$$

the maximum error across all k coordinates is bounded by ???

$\iff$ $\max\{N(0,\sigma^2), \ \ldots, \ N(0,\sigma^2)\}$ is bounded

1) bound the error of <u>one</u> Gaussian

If $Z \sim N(0,\sigma^2)$ then $\Pr[|Z| > t\cdot\sigma] \leq 2e^{-\frac{t^2}{2}}$

$$\Rightarrow \Pr[N(0,\sigma^2) > \log(n)\cdot\sigma] \approx 1/n^2 = o(1)$$

with high probability

$\Rightarrow$ the error of <u>one</u> query is

$$\tilde{O}(\sigma) = \tilde{O}\left(\frac{\sqrt{k}}{\varepsilon n}\right) \text{ w.h.p}$$

2) bound the noise across all k queries

$$\Pr\left[\max_{1 \le i \le k} |Z_i| \ge t \cdot \sigma\right] \le k \cdot 2 \cdot e^{-t^2/2}$$

set $t = \log(n \cdot k)$ $\qquad o(1)$

The error is less than $\log(nk) \cdot \sigma = \tilde{O}\left(\frac{\sqrt{k}}{\varepsilon n}\right)$

$\underbrace{\text{with high probability}}_{1 - o_n(1)}$ (e.g. $1 - 1/n$ or $1 - 1/n^2$ or $1 - e^{-n}$ ...)

What about arbitrary (ε,δ)-DP algos?

We just saw that the composition of k Gaussian mechanisms that are (ε,δ)-DP is $(\tilde{O}(\sqrt{k}\,\varepsilon), \delta)$-DP!

This applies to <u>all</u> (ε,δ)-DP mechanisms

Theorem: (Advanced composition)

For all $\varepsilon, \delta, \delta' > 0$, let $M = (M_1, \dots, M_k)$ be a sequence of (ε,δ)-DP algorithms.

Then M is $(\hat{\varepsilon}, \hat{\delta})$-DP where

$$\tilde{\varepsilon} = \varepsilon\sqrt{2k\log(1/\delta')} + \underbrace{k \cdot \varepsilon \frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}}_{\text{"if } \varepsilon \text{ is small" this is roughly } \varepsilon^2/2}$$

$$\tilde{\delta} = k\delta + \delta'$$

$$\hat{\varepsilon} \approx \varepsilon\sqrt{k} + k \cdot \varepsilon^2$$

Cleaning this up:

let $\varepsilon < \frac{1}{\sqrt{k}}$, $\delta' = \delta$

Then $\tilde{\varepsilon} = \Theta\left(\varepsilon\sqrt{k\log(1/\delta)}\right)$

$\tilde{\delta} = \Theta(k\delta)$

How many counting queries can we answer privately?

Guarantee	Error	Mechanism	lower bound
local ε-DP	$\tilde{O}\left(\frac{1}{\varepsilon\sqrt{n}}\right)$	Randomized R.	$\Omega\left(\frac{1}{\varepsilon\sqrt{n}}\right)$
ε-DP	$\tilde{O}\left(k/\varepsilon n\right)$	Laplace	$\Omega\left(k/\varepsilon n\right)$
(ε,δ)-DP	$\tilde{O}\left(\sqrt{k}/\varepsilon n\right)$	Gaussian	$\Omega\left(\sqrt{k}/\varepsilon n\right)$
(ε,δ)-DP	$\tilde{O}\left(\frac{\log k}{\varepsilon n}\sqrt{\log\lvert X\rvert}\right)^{1/2}$	Private Mult. Weights	$\tilde{O}\left(\frac{\log k}{\varepsilon n}\sqrt{\log\lvert X\rvert}\right)^{1/2}$

X is the domain of the data

for k counting queries on bits, $X = \{0,1\}^k$

then you win nothing...

Beyond numerical queries

<u>example</u>: All ETH students vote for their favorite class

We use "approval voting" (you can vote for as many classes as you'd like)

For k classes, a student's vote is $x_i \subseteq [k]$

The score of the jth class is

$$q(j;D) = \left|\{ i \in D : j \in x_i \}\right|$$

we want to release $\underset{1 \leq j \leq k}{\operatorname{argmax}}\, q(j;D)$

How to do this privately?

"Naive": add noise to each score $q(j;D)$ and release maximum (post-processing & composition)

for ε-DP, we need noise $\tilde{O}\left(\frac{k}{\varepsilon}\right)$ for each score

Exponential mechanism

Def: Selection Problem

- A set $\mathcal{Y}$ of possible outcomes
- A score function $q: \mathcal{Y} \times X^n \to \mathbb{R}$ that measures the "quality" of our output for a given dataset
- A sensitivity bound Δ such that $q(y; \cdot)$ has sensitivity at most Δ for every $y \in \mathcal{Y}$

$\Rightarrow \forall y \in \mathcal{Y}$, and $D \sim D'$:

$$|q(y; D) - q(y; D')| \leq \Delta$$

Def: Exponential mechanism

Let $(D, \mathcal{Y}, q, \Delta)$ be a selection problem

Then return a sample from the distribution over $\mathcal{Y}$ defined by:

$$\Pr[Y = y] \propto \exp\left(\frac{\varepsilon}{2\Delta} q(y; D)\right)$$

Thm : Exponential mechanism is ε-DP

Proof: we want to show: $\frac{\Pr[Y=y \mid D]}{\Pr[Y=y \mid D']} \le e^{\varepsilon}$

$$\frac{\Pr[Y=y \mid D]}{\Pr[Y=y \mid D']} \overset{(*)}{=} \frac{\exp\left(\frac{\varepsilon}{2\Delta} q(y;D)\right) / \sum_{y} \exp(\dots)}{\exp\left(\frac{\varepsilon}{2\Delta} q(y;D')\right) / \sum_{y'} \exp(\dots)}$$

$$\exp\left(\frac{\varepsilon}{2\Delta} q(y;D)\right) \le \exp\left(\frac{\varepsilon}{2\Delta}\left(q(y;D')+\Delta\right)\right)$$

$$= \exp\left(\frac{\varepsilon}{2\Delta} q(y;D')\right) \cdot \exp\left(\frac{\varepsilon}{2}\right)$$

$$(*) \le \frac{\cancel{\exp\left(\frac{\varepsilon}{2\Delta} q(y;D')\right)} \cdot \exp\left(\frac{\varepsilon}{2}\right)}{\cancel{\exp\left(\frac{\varepsilon}{2\Delta} q(y;D')\right)}} \cdot \underbrace{\frac{\sum_{y'} \exp\left(\frac{\varepsilon}{2\Delta} q(y;D')\right)}{\sum_{y'} \exp\left(\frac{\varepsilon}{2\Delta} q(y;D)\right)}}$$

$$\le \exp\left(\frac{\varepsilon}{2}\right) \cdot \exp(\varepsilon/2)$$

$$= \exp(\varepsilon)$$

Utility of the exponential mechanism

$$\mathrm{OPT}(D) = \max_{y \in \mathcal{Y}} q(y; D)$$

<u>Thm</u> Suppose $\mathcal{Y}$ is finite and has size k, and let $\delta > 0$.

Then the output of the exponential mechanism M satisfies

$$\Pr_{y \leftarrow M(D)} \left[q(y; D) \leq \mathrm{OPT}(D) - \frac{2\Delta(\log k + \log(\frac{1}{\delta}))}{\varepsilon} \right] \leq \delta$$

Compare with "naive" approach:

Laplace: error $\tilde{O}(k/\varepsilon)$

Gaussian: error $\tilde{O}(\sqrt{k}/\varepsilon)$

Exp. Mech: error $\tilde{O}\left(\frac{\log k}{\varepsilon}\right)$

Ex: $k = 1000$ classes, $\varepsilon = 1/2$

Gaussian: noise $\tilde{O}\left(\frac{\sqrt{1000}}{0.5}\right) \approx 63$

E.M. with p 99% the selected class has a score within 46 of OPT