CSE660
Differential Privacy

November 6, 2017

Marco Gaboardi
Room: 338-B
gaboardi@buffalo.edu
http://www.buffalo.edu/~gaboardi

Differential privacy

Definition

Given $\varepsilon, \delta \geq 0$, a probabilistic query $Q: X^n \rightarrow R$ is (ε,δ)-differentially private iff

for all adjacent database b_1, b_2 and for every $S \subseteq R$:

$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon)\Pr[Q(b_2) \in S] + \delta$$

Composition

Theorem 1.18 (Standard composition for ϵ-differential privacy)**.** Let $\mathcal{M}_i : \mathcal{X}^n \to R_i$ be ϵ_i-differentially private algorithms (for $1 \leq i \leq k$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $\sum_{i=1}^{k} \epsilon_i$-differentially private.

Composition for (ε,δ)-DP

Theorem 1.22 (Standard composition for (ϵ, δ)-differential privacy)**.** Let $\mathcal{M}_i : \mathcal{X}^n \to R_i$ be (ϵ_i, δ_i)-differentially private algorithms (for $1 \leq i \leq k$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $(\sum_{i=1}^k \epsilon_i, \sum_{i=1}^k \delta_i)$-differentially private.

Advanced Composition

Theorem 1.23 (Advanced composition). Let $\mathcal{M}_i : \mathcal{X}^n \to R_i$ be (ϵ, δ)-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')})\epsilon, k\delta + \delta')$-differentially private for every $\delta' > 0$.

Answering multiple queries

We have seen several methods to answer a single query:

- Randomized Response
- Laplace Mechanism
- Exponential Mechanism

And methods to answer multiple queries with small error:

- Standard composition - we can answer $\sqrt{n}$ queries.
- Advanced composition - we can answer n queries.

Gaussian Mechanism

Algorithm 14 Pseudo-code for the Gaussian Mechanism

1: **function** GAUSSMECH(D, q, ϵ)
2: $\quad Y \xleftarrow{\$} \mathsf{Gauss}(0, \frac{2\ln(\frac{1.25}{\delta})(\Delta_2 q)^2}{\epsilon^2})$
3: $\quad$ **return** $q(D) + Y$
4: **end function**

Theorem (Privacy of the Gaussian Mechanism)
The Gaussian mechanism is (ε,δ)-differentially private.

The roles of δ

We have seen three roles that δ plays in DP

1. to account for the probability of failure in a DP computation
2. in the advanced composition theorem to have a better bound on the growth of ε when composing n queries,
3. to allow an analysis of the Gaussian Mechanism.

The point 3 (and 2) were the original motivations for introducing (ε,δ)-differential privacy while the point 1 is somehow undesirable.

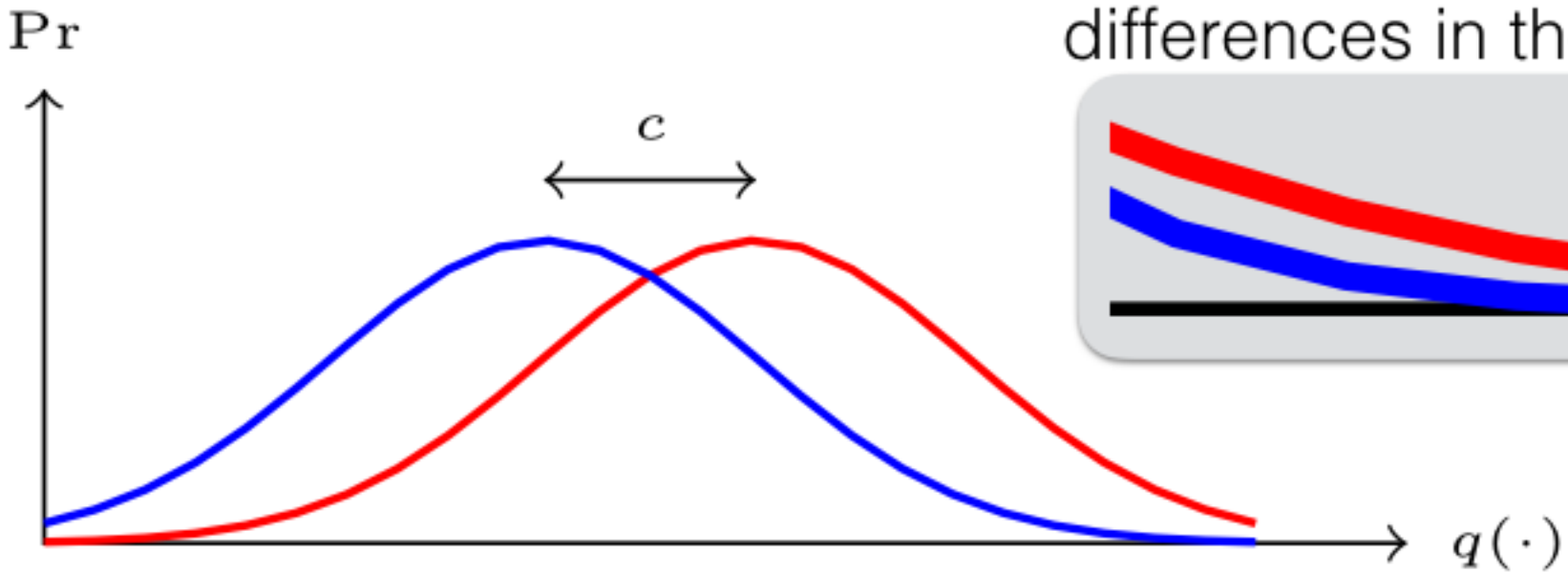
Can we give other privacy definitions that behave well with respect to 3 and 2 and do not require 1?

Gaussian Mechanism

Theorem (Privacy of the Gaussian Mechanism)
The Gaussian mechanism is (ε,δ)-differentially private.

Proof: Intuitively

We need δ to account for bigger differences in the tail



Advanced Composition

Theorem 1.23 (Advanced composition). Let $\mathcal{M}_i : \mathcal{X}^n \to R_i$ be (ϵ, δ)-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')})\epsilon, k\delta + \delta')$-differentially private for every $\delta' > 0$.

Intuition: some of the outputs have positive privacy loss (i.e. give evidence for dataset D) and some have negative privacy loss (i.e. give evidence for dataset D'). The cancellations gives a smaller overall privacy loss.

Strategy:

1-considering the expected value of the privacy loss,

2-bound the privacy loss of all the variables together

3-compute the probability

Privacy Loss

In general we can think about the following quantity as the privacy loss incurred by observing r as output of $\mathcal{M}$ on the databases D and D'.

$$\mathcal{L}_{\mathcal{M}}^{D \to D'}(r) = \ln\left(\frac{\Pr[\mathcal{M}(D) = r]}{\Pr[\mathcal{M}(D') = r]}\right) = -\mathcal{L}_{\mathcal{M}}^{D' \to D}(r)$$

The $(\epsilon, 0)$-differential privacy requirement corresponds to requiring that for every r and every adjacent D, D' we have:

$$\left|\mathcal{L}_{\mathcal{M}}^{D \to D'}(r)\right| \leq \epsilon$$

(ε,δ)-Differential Privacy

This corresponds to a privacy loss of the form:

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

The (ϵ, δ)-differential privacy requirement corresponds to requiring that for every r and every adjacent D, D' we have:

$$\Pr\left[\left|\mathcal{L}_{\mathcal{M}}^{D\to D'}(r)\right| \leq \epsilon\right] \geq 1-\delta$$

Bounding the moments

A random variable can be described using its moments.

$$\mu_n = \mathbb{E}[X^n]$$

Here we consider central moments. For instance, the first central moment is the mean, the second is the variance, the third is the skewness, etc.

Can we bound the moments of the privacy loss?

Moment generating function

The probability distribution of a random variable X can be described by its moment generating function:

$$\mathtt{m}_X(\alpha) = \mathbb{E}[e^{\alpha X}]$$

This function can be used to compute, or give upper bounds on the moments of the random variable X.

$$m_X(\alpha) = 1 + \alpha\mu_1 + \frac{\alpha^2\mu_2}{2!} + \ldots + \frac{\alpha^n\mu_n}{n!} + \ldots$$

Renyi Divergence

Definition 3 (Rényi divergence)**.** For two probability distributions P and Q defined over $\mathcal{R}$, the Rényi divergence of order $\alpha > 1$ is

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha - 1} \log \mathrm{E}_{x\sim Q}\left(\frac{P(x)}{Q(x)}\right)^\alpha .$$

It provides a way to measures the closeness of two probability distributions.

It is parametrized by $\boldsymbol{\alpha}$. Each $\boldsymbol{\alpha}$ gives different informations about the closeness of the two probability distributions.

Renyi Divergence

Having

$$D_\alpha(\mathcal{Q}(b_1)||\mathcal{Q}(b_2)) \leq \rho$$

Corresponds to

$$\mathbb{E}[e^{(\alpha-1)\left(\frac{\mathcal{Q}(b_1)}{\mathcal{Q}(b_2)}\right)}] \leq e^{(\alpha-1)\rho}$$

Renyi Differential Privacy

Definition 1.11 (Rényi Differential Privacy). Given $\alpha > 1$ and $0 \leq \rho \leq 1$, a probabilistic query $\mathcal{Q} : X^n \to R$ is (α, ρ)-Rényi Differentially Private (RDP) iff for all adjacent databases b_1, b_2:

$$D_\alpha(\mathcal{Q}(b_1) || \mathcal{Q}(b_2)) \leq \rho$$

This definition shares some good properties with (ε,δ)-differential privacy: resilience to post-processing, composition and group privacy.

[Mironov'17]

zero-Concentrated Differential Privacy

Definition 1.12 (zero-Concentrated Differential Privacy). Given $0 \leq \rho \leq 1$, a probabilistic query $\mathcal{Q} : X^n \to R$ is ρ-zero Concentrated Differentially Private (zCDP) iff for all adjacent databases b_1, b_2:

$$\forall \alpha > 1.\ D_\alpha(\mathcal{Q}(b_1) || \mathcal{Q}(b_2)) \leq \alpha\rho.$$

This definition shares some good properties with (ε,δ)-differential privacy: resilience to post-processing, composition and group privacy.

[Dwork&Rothblum'15,Bun&Steinke'16]

Gaussian Mechanism revisited

Algorithm 15 Pseudo-code for the Gaussian Mechanism

1: **function** GAUSSMECH(D, q, ϵ)
2: $Y \xleftarrow{\$} \mathsf{Gauss}(0, \sigma^2)$
3: **return** $q(D) + Y$
4: **end function**

The Gaussian mechanism satisfies $\boldsymbol{\alpha}$-Renyi differential privacy for:

$$\rho = \frac{\alpha \Delta_2 q}{2\sigma^2}$$

It also satisfies zero-concentrated differential privacy for:

$$\rho = \frac{\Delta_2 q}{2\sigma^2}$$

Composition for RDP

Theorem 1.26 (Composition for (α, ϵ)-RDP)**.** Let $\mathcal{M}_1 : \mathcal{X}^n \to R_1$ be an (α, ϵ_1)-RDP algorithm and let $\mathcal{M}_2 : \mathcal{X}^n \to R_2$ be an (α, ϵ_2)-RDP algorithm. Then their composition defined to be $\mathcal{M}_{1,2} : \mathcal{X}^n \to R_1 \times R_2$ by the mapping $\mathcal{M}_{1,2}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D))$ is $(\alpha, \epsilon_1 + \epsilon_2)$-RDP.

[Mironov'17]

Composition for zCDP

Theorem 1.27 (Composition for (ξ, ρ)-zCDP)**.** Let $\mathcal{M}_1 : \mathcal{X}^n \to R_1$ be an (ξ_1, ρ_1)-zCDP algorithm and let $\mathcal{M}_2 : \mathcal{X}^n \to R_2$ be an (ξ_2, ρ_2)-zCDP algorithm. Then their composition defined to be $\mathcal{M}_{1,2} : \mathcal{X}^n \to R_1 \times R_2$ by the mapping $\mathcal{M}_{1,2}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D))$ is $(\xi_1 + \xi_2, \rho_1 + \rho_2)$-zCDP.

[Dwork&Rothblum'15,Bun&Steinke'16]

Back to DP from RDP

Proposition 3 (From RDP to (ϵ, δ)-DP)**.** *If f is an (α, ϵ)-RDP mechanism, it also satisfies $(\epsilon + \frac{\log 1/\delta}{\alpha - 1}, \delta)$-differential privacy for any $0 < \delta < 1$.*

[Mironov'17]

Back to DP from zCDP

Proposition 1.3. *If M provides ρ-zCDP, then M is $(\rho + 2\sqrt{\rho \log(1/\delta)}, \delta)$-differentially private for any $\delta > 0$.*

[Dwork&Rothblum'15,Bun&Steinke'16]

Using zCDP and RDP in data analysis

- They behave well with respect to the Gaussian distribution which composes well,
- They have better composition properties,
- They provide more refined analysis,
- Still under study.