Rényi Differential Privacy

for Sorbonne Université

Eyal Cohen[1] Eline Pot[1]
[1] LPSM

Supervisor(s):

Christophe Denis[1], Rafael Pinot[1]

Table of Contents

Definition ((ϵ, δ)-DP)

A randomized mechanism $f : \mathcal{D} \mapsto \mathcal{R}$ has (ϵ, δ)- Differential Privacy if for any adjacent $D, D' \in \mathcal{D}$ and $S \subset \mathcal{R}$ we have :

$$\mathbb{P}(f(D) \in S) \leqslant e^{\epsilon}\mathbb{P}(f(D') \in S) + \delta$$

We also define ϵ-DP with $\delta = 0$

Definition (Gaussian mechanism)

$$\mathbf{G}_{\sigma}f(x) = f(x) + N(0, \sigma^2)$$

REMARKS :

- The definition of (ϵ, δ)-DP was initially proposed to capture privacy guarantees of the Gaussian mechanism.
  $\longrightarrow$ However, the *DP* is not always guaranteed.
- It is also interesting and popular for applications of advanced composition theorems. This property allows the control of the cumulative privacy loss over multiple runs of an analysis on the same dataset.
  $\longrightarrow$ But whereas a single mechanism satisfies a continuum of incomparable (ϵ, δ)-DP guarantees, the generalization of this result to the composition of a heterogeneous mechanism is a $\#P$-complete problem.

Table of Contents

Definition (Rényie Divergence)

The Rényi divergence of order $\alpha > 1$ is defined for two probabilities distributions P and Q as :

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_{X\sim Q} \left(\frac{P(X)}{Q(X)} \right)^\alpha$$

Link with KL divergence : We can note that this definition is not valid for $\alpha = 1$ and $\alpha = \infty$. We define them by continuity. In particular, we recognize the Kullback-Leibler divergence for $\alpha = 1$:

$$D_1(P||Q) = \log \mathbb{E}_{X\sim P} \left(\frac{P(X)}{Q(X)} \right) = KL(P||Q) \quad \text{and} \quad D_\infty(P||Q) = \sup_{X \in suppQ} \log \frac{P(X)}{Q(X)}$$

Definition ((α, ϵ)-RDP)

A randomized mechanism $f : \mathcal{D} \mapsto \mathcal{R}$ has ϵ- Rényi Differential Privacy of order α if for any adjacent $D, D' \in \mathcal{D}$, we have:

$$D_\alpha(f(D)||f(D')) \leqslant \epsilon$$

Property (”Bad Outcomes” Guarantee)

The DP guarantee states that the probability of observing a ”bad outcome” will not change by more than a factor of e^{ϵ} whether someone's data is included in the dataset.
This guarantee is relaxed for RDP:
If f is (α, ϵ)-RDP then $\forall D, D' \in \mathcal{D}$ adjacent, we have:

$$e^{-\epsilon}\mathbb{P}(f(D') \in S)^{\alpha/(\alpha-1)} \leqslant \mathbb{P}(f(D) \in S) \leqslant [e^{\epsilon}\mathbb{P}(f(D') \in S)]^{\alpha/(\alpha-1)}$$

PROOF IDEAS: *It is a consequence of the (α, ϵ)-RDP and the Probability Preservation : If we have $\alpha > 1$, P, Q two distributions defined over $\mathcal{R}$ with the same support, $A \subset \mathcal{R}$ an event, then :*

$$\mathbb{P}(A) \leqslant (\exp(D_\alpha(P||Q)) \cdot Q(A))^{(\alpha-1)/\alpha}$$

This result is an application of Hölder's inequality : $\forall f, g$ real-valued functions and $\forall p, q > 0$ with $\frac{1}{p} + \frac{1}{q} = 1$, then : $||fg||_1 \leqslant ||f||_p||g||_q$. We apply this inequality with $p = \alpha$, $q = \alpha/(\alpha-1)$, $f(x) = \frac{P(x)}{Q(x)^{1/q}}$ and $g(x) = Q(x)^{1/q}$ to get :

$$\begin{aligned}\int_A P(x)dx &\leqslant \left(\int_A \left(\frac{P(x)}{Q(x)^{\alpha/(\alpha-1)}}\right)^{\alpha} dx\right)^{\frac{1}{\alpha}} \left(\int_A \left(Q(x)^{(\alpha-1)/\alpha}\right)^{\alpha/(1-\alpha)} dx\right)^{\frac{\alpha-1}{\alpha}} \\ &= \left(\int_A P(x)^{\alpha}Q(x)^{1-\alpha}dx\right)^{\frac{1}{\alpha}} \left(\int_A Q(x)dx\right)^{\frac{\alpha-1}{\alpha}} \\ &\leqslant \exp D_\alpha(P||Q)^{(\alpha-1)/\alpha} Q(A)^{(\alpha-1)/\alpha}\end{aligned}$$

We understand privacy as the ability to limit an adversary's knowledge about a given individual's impact over the query. The original framework constrains the difference in probability for two neighboring datasets to a factor e^{ϵ}, then the relaxed version allows for a compromise, adding the δ acts as a breach of privacy in favour of utility.
An interesting property of the Rényi differential privacy framework is that while it is a relaxation of the second framework, it does not allow for constant as a breach of privacy. Indeed, it keeps the difference to a factor and an exponent.
To be more precise:

Corollary

Let f be $(\alpha, \epsilon) - RDP$ with $\alpha > 1$, then:

$$\mathbb{E}\left[\left(\frac{R_{posterior}(D, D')}{R_{prior}(D, D')}\right)^{\alpha-1}\right] \leqslant \exp\left((\alpha - 1)\epsilon\right)$$

And in particular:

$$\mathbb{P}(R_{posterior}(D, D') > R_{prior}(D, D')) < \frac{\exp(\epsilon)}{\beta^{\frac{1}{\alpha-1}}}$$

Limitations of the guarantee

2 Rényi Differential Privacy

The RDP framework suffers from a weaker bound the rarer an event gets, we can see it in this table of bounds for $\epsilon = 0.1$:

Probability	$\alpha = 1.1$	$\alpha = 10$	$\alpha = 100$
0.5	$[0.00044, 0.94751]$	$[0.419, 0.586]$	$[0.449, 0.556]$
0.001	$[9.05 \times 10^{-34}, 0.5385]$	$[0.00042, 0.00218]$	$[0.00084, 0.00118]$
10^{-6}	$[9.04 \times 10^{-67}, 0.2874]$	$[.195 \times 10^{-6}, 4.36 \times 10^{-6}]$	$[7.87 \times 10^{-7}, 1.27 \times 10^{-6}]$

These being more and more vacuous the closer α gets to 1.
This shows a much weaker guarantee for rare events than the pure DP framework with its factor only dependant on ϵ.
Compared to the $(\epsilon, \delta) - DP$ framework, the RDP offers a simpler analysis and the comparison in guarantees will come down to the δ and the α factors, but should offer stronger bounds for events happening with probability smaller than delta.

Property	Differential Privacy	Rényi Differential Privacy
Change in probability of outcome S	$\Pr[f(D) \in S] \leq e^{\epsilon} \Pr[f(D') \in S]$ $\Pr[f(D) \in S] \geq e^{-\epsilon} \Pr[f(D') \in S]$	$\Pr[f(D) \in S] \leq (e^{\epsilon} \Pr[f(D') \in S])^{(\alpha-1)/\alpha}$ $\Pr[f(D) \in S] \geq e^{-\epsilon} \Pr[f(D') \in S]^{\alpha/(\alpha-1)}$
Change in the Bayes factor	$\frac{R_{\text{post}}(D, D')}{R_{\text{prior}}(D, D')} \leq e^{\epsilon}$ always	$\mathrm{E}\left[\left\{\frac{R_{\text{post}}(D, D')}{R_{\text{prior}}(D, D')}\right\}^{\alpha}\right] \leq \exp[(\alpha-1)\epsilon]$
Change in log of the Bayes factor	$\lvert\Delta \log R(D, D')\rvert \leq \epsilon$ always	$\mathrm{E}[\Delta \log R(D, D')] \leq \epsilon$
Post-processing	f is ϵ-DP (or (α, ϵ)-RDP) $\Rightarrow g \circ f$ is ϵ-DP (or (α, ϵ)-RDP, resp.)	
Adaptive sequential composition (basic)	f, g are ϵ-DP (or (α, ϵ)-RDP) $\Rightarrow (f, g)$ is 2ϵ-DP (resp., $(\alpha, 2\epsilon)$-RDP)	
Group privacy, pre-processing	f is ϵ-DP (or (α, ϵ)-RDP), g is 2^c-stable $\Rightarrow f \circ g$ is $2^c\epsilon$-DP (resp., $(\alpha/2^c, 3^c\epsilon)$-RDP)	

Figure: Summary of the properties shared by *DP* and *RDP* (from the paper)

Table of Contents

We note that :

- The definition of (ϵ, δ)-RDP is equivalent to ϵ-differential privacy.
- As the Rényi divergence is monotonous in α, we have that (∞, δ)-RDP implies (α, δ)-RDP for any $\alpha < \infty$.

Theorem (From RDP to (ϵ, δ)-DP)

If f is an(α, δ)-RDP mechanism, then it is also $\left(\epsilon + \frac{\log(1/\delta)}{\alpha - 1}, \delta\right)$ *- DP for any* $0 < \delta < 1$.

<u>PROOF IDEAS:</u> *We re-use the argument of Probability Preservation which states that : If we have* $\alpha > 1, P, Q$ *two distributions defined over* $\mathcal{R}$ *with the same support,* $A \subset \mathcal{R}$ *an event, then we have:*
$\mathbb{P}(A) \leqslant (\exp(D_\alpha(P||Q) \cdot Q(A)))^{(\alpha-1)/\alpha}$.
And we prove that $\mathbb{P}(f(D) \in S) \leqslant \max\left(\exp\left(\epsilon + \frac{\log(1/\delta)}{\alpha - 1}\right)\mathbb{P}(f(D') \in S); \delta\right)$

Table of Contents

Lemma

If P, Q verify $D_\infty(P||Q) \leqslant \epsilon$ and $D_\infty(Q||P) \leqslant \epsilon$, then $\forall \alpha \geqslant 1 : D_\alpha(P||Q) \leqslant 2\alpha\epsilon^2$

Theorem (Advanced Composition Theorem)

If $f : \mathcal{D} \mapsto \mathcal{R}$ is an adaptive composition of n mechanisms all satisfying ϵ-DP, and D, D' are two adjcents inputs, then $\forall S \subset \mathcal{R}$, we have :

$$\mathbb{P}(f(D) \in S) \leqslant \exp\left(2\epsilon\sqrt{n\log(1/\mathbb{P}(f(D') \in S))}\right) \cdot \mathbb{P}(f(D') \in S)$$

PROOF IDEAS: *Let us note $Q = \mathbb{P}(f(D') \in S)$. We apply the lemma to the RDP curve of the underlying mechanisms and we use the property of adaptive sequential composition. Then, we have $\forall \alpha \geqslant 1, D_\alpha(f(D)||f(D')) \leqslant 2\alpha n\epsilon^2$.*
If $\log(1/Q) \geqslant \epsilon^2 n$, then, by choosing $\alpha = \sqrt{\log(1/Q)}/(\epsilon\sqrt{n})$, and applying the property of probability preservation, we have : $\mathbb{P}(f(D) \in S) \leqslant \exp(2\epsilon\sqrt{n\log(1/Q)}) \cdot Q$
Else, it is trivial since we have $\exp(2\epsilon\sqrt{n\log(1/Q)}) \cdot Q \geqslant \exp(2\log(1/Q)) \cdot Q = 1/Q > 1$

Corollary

If f is the composition of the n ϵ-DP mechanisms, and $0 < \delta < 1$ is such that $\log(1/\delta) \geqslant \epsilon^2 n$, then f is $(4\epsilon\sqrt{2n\log(1/\delta)}, \delta)$-DP

Table of Contents

Definition

Let $f : \mathcal{D} \mapsto \{0, 1\}$. We define the Random Response mechanism as :

$$RR_p f(D) = \begin{cases} f(D) & \text{with probability } p \\ 1 - f(D) & \text{with probability } 1 - p \end{cases}$$

Theorem

$RR_p(f)$ is $\left(\alpha, \frac{1}{\alpha-1} \log(p^\alpha(1-p)^{1-\alpha} + (1-p)^\alpha p^{1-\alpha})\right)$-RDP if $\alpha > 1$ and $\left(\alpha, (2p-1)\log(\frac{p}{1-p})\right)$-RDP if $\alpha = 1$

Definition (Laplace noise)

Let us suppose that $f : \mathcal{D} \mapsto \mathbb{R}$ is such that $\forall$ adjacents $D, D' \in \mathcal{D}$, $|f(D) - f(D')| \leqslant 1$ (*i.e.* f has sensitivity 1).
We define the Laplace mechanism : $\mathbf{L}_\lambda f(D) = f(D) + \Lambda(0, \lambda)$
with $\Lambda(\mu, \lambda)$ is the Laplace distribution of density : $\frac{1}{2\lambda} \exp\left(-\frac{|x-\mu|}{\lambda}\right)$

Proposition (Rényi divergence for Laplace distribution and its offset)

$$\forall \alpha \geqslant 1 \text{ and } \lambda > 0 : \qquad D_\alpha(\Lambda(0, \lambda)||\Lambda(1, \lambda)) = \frac{1}{\alpha - 1} \log\left(\frac{\alpha}{2\alpha - 1} \exp\left(\frac{\alpha - 1}{\lambda}\right) + \frac{\alpha - 1}{2\alpha - 1} \exp\left(-\frac{\alpha}{\lambda}\right)\right)$$

<u>Proof ideas:</u> *We use the definition of Rényie Divergence with integrals as the Laplace distribution admits density and we evaluate the integral separately over the intervals* $(-\infty, 0], [0, 1]$ *and* $[1, +\infty)$

Theorem

If $f : \mathcal{D} \mapsto \mathbb{R}$ *has sensitivity* 1*, then the Laplace mechanism* $\mathbf{L}_\lambda f$ *is*
$\left(\alpha, \frac{1}{\alpha-1} \log\left(\frac{\alpha}{2\alpha-1} \exp((\alpha - 1)/\lambda) + \frac{\alpha-1}{2\alpha-1} \exp(-\alpha/\lambda)\right)\right)$*- RDP.*

<u>Proof ideas:</u> *We use the precedent proposition combined with the fact that the Laplace mechanism is additive and that the Rényi divergence between* $\mathbf{L}_\lambda f(D)$ *and* $\mathbf{L}_\lambda f(D')$ *only depends on* α *and* $|f(D) - f(D')|$

Definition (Gaussian mechanism)

Let us suppose that $f : \mathcal{D} \mapsto \mathbb{R}$.
We define the Gaussian mechanism : $\mathbf{G}_\sigma f(D) = f(D) + N(0, \sigma^2)$

Proposition (Rényi Divergence between a Gaussian and its offset)

$$D_\alpha \left(N(0, \sigma^2) || N(\mu, \sigma^2) \right) = \alpha \mu^2 / (2\sigma^2)$$

Theorem

If $f : \mathcal{D} \mapsto \mathbb{R}$ has sensitivity 1, then the Gaussian mechanism $\mathbf{G}_\sigma f$ is $(\alpha, \alpha/(2\sigma^2))$

<u>Proof ideas:</u> *We use the same approach as before and use the previous result.*

Rényi Differential Privacy

Thank you for listening !
Any Questions ?