Privacy Loss Distributions

Differential Privacy Team
Google

October 3, 2020

This document is a supplementary material for the implementation of the algorithms for building and manipulating privacy loss distributions in the privacy accounting library.

1 Notation and Preliminaries

For two distributions $\mathcal{D}$ and $\mathcal{D}'$, we use $\mathcal{D} \otimes \mathcal{D}'$ to denote the product distribution of $\mathcal{D}$ and $\mathcal{D}'$. Furthermore, we denote by $\mathcal{D} + \mathcal{D}'$ the distribution of $X + X'$ where X and X' are independently sampled from $\mathcal{D}$ and $\mathcal{D}'$ respectively; $\mathcal{D} - \mathcal{D}'$ is defined similarly. For a real number $k \in \mathbb{R}$, we denote by $k + \mathcal{D}$ the distribution of $k + X$ when $X \sim \mathcal{D}$.

Discrete Distributions. For a discrete distribution $\mathcal{D}$, when there is no ambiguity, we abbreviate $\Pr_{X\sim\mathcal{D}}[X = x]$ as $\mathcal{D}(x)$. For two discrete distributions μ and μ', we use $\mathfrak{D}_{e^\varepsilon}(\mu||\mu')$ to denote their *ε-hockey stick divergence*, i.e.,

$$\mathfrak{D}_{e^\varepsilon}(\mu||\mu') := \sum_{y\in\text{supp}(\mu)} [\mu(y) - e^\varepsilon \cdot \mu'(y)]_+,$$

where $[x]_+$ denotes $\max\{x, 0\}$.

Continuous Distributions. For a continuous distribution $\mathcal{D}$, we use $f_{\mathcal{D}}(\cdot)$ to denote its probability density function. For two continuous distributions μ and μ', their ε-hockey stick divergence is defined as

$$\mathfrak{D}_{e^\varepsilon}(\mu||\mu') := \int [f_\mu(y) - e^\varepsilon \cdot f_{\mu'}(y)]_+ dy.$$

Differential Privacy. For a mechanism $\mathcal{M}$ and an input dataset $\mathbf{x}$, we use $\mathcal{M}(\mathbf{x})$ to denote the distribution of the output. The standard definition of differential privacy [DMNS06, DKM+06] may be rephrased as follows.

Observation 1. *A mechanism $\mathcal{M}$ is (ε, δ)-differentially private (or (ε, δ)-DP for short) if and only if, for any neighboring input datasets $\mathbf{x}, \mathbf{x}'$, it holds that $\mathfrak{D}_{e^\varepsilon}(\mathcal{M}(\mathbf{x})||\mathcal{M}(\mathbf{x}')) \le \delta$.*

2 Privacy Loss Distribution

A notion that will be useful to us is the so-called *Privacy Loss Distribution (PLD)* defined in [DR16]. Here we will mostly follow the notations from [MM18, SMM19, KJH19, KJPH20], from which most of the results we use follow.

Definition 1. *For two discrete distributions μ_{up} and μ_{lo}, their* privacy loss *at $o \in \text{supp}(\mu_{up})$ is defined as*

$$\mathcal{L}_{\mu_{up}/\mu_{lo}}(o) := \ln\left(\frac{\mu_{up}(o)}{\mu_{lo}(o)}\right).$$

The privacy loss distribution (PLD) *of μ_{up} and μ_{lo}, denoted by $PLD_{\mu_{up}/\mu_{lo}}$, is a distribution on $\mathbb{R} \cup \{\infty\}$ where $y \sim PLD_{\mu_{up}/\mu_{lo}}$ is generated as follows: sample $o \sim \mu_{up}$ and let $y = \mathcal{L}_{\mu_{up}/\mu_{lo}}(o)$.*

For two continuous distributions μ_{up} and μ_{lo}, their privacy loss at o is defined as

$$\mathcal{L}_{\mu_{up}/\mu_{lo}}(o) := \ln\left(\frac{f_{\mu_{up}}(o)}{f_{\mu_{lo}}(o)}\right).$$

The PLD of μ_{up} and μ_{lo}, denoted by $PLD_{\mu_{up}/\mu_{lo}}$, is again a distribution on $\mathbb{R} \cup \{\infty\}$ where $y \sim PLD_{\mu_{up}/\mu_{lo}}$ is generated by picking $o \sim \mu_{up}$ and then letting $y = PLD_{\mu_{up}/\mu_{lo}}(o)$.

For a mechanism $\mathcal{M}$ and two input vectors $\mathbf{x}$ and $\mathbf{x}'$, the privacy loss distribution (PLD) between $\mathbf{x}$ and $\mathbf{x}'$ is defined as $PLD_{\mathcal{M}(\mathbf{x})/\mathcal{M}(\mathbf{x}')}$.

The main observation that makes PLD useful is that it allows one to calculate the ε-hockey stick divergence between the two distributions, or equivalently to check whether a mechanism is (ε, δ)-DP.

Observation 2 ([SMM19, KJH19])**.** *For any two distributions μ_{up} and μ_{lo} where both are discrete or both are continuous, it holds that*

$$\mathfrak{D}_{\varepsilon}(\mu_{up}||\mu_{lo}) = \mathbb{E}_{y \sim PLD_{\mu_{up}/\mu_{lo}}}[1 - e^{\varepsilon - y}]_+.$$

Due to Observation 1, a mechanism $\mathcal{M}$ is (ε, δ)-DP if and only if the following holds for all neighboring input datasets $\mathbf{x}$ and $\mathbf{x}'$:

$$\delta \geq \mathbb{E}_{y \sim PLD_{\mathcal{M}(\mathbf{x})/\mathcal{M}(\mathbf{x}')}}[1 - e^{\varepsilon - y}]_+.$$

Another observation is that PLD is very compatible with composition of mechanisms. When the composition is non-adaptive, i.e., when mechanisms $\mathcal{M}_1$ and $\mathcal{M}_2$ are run independently, the output distribution on input vector $\mathbf{x}$ is simply the product distribution $\mathcal{M}_1(\mathbf{x}) \otimes \mathcal{M}_2(\mathbf{x})$. The observation here is that the PLD of the product distribution is simply the *convolution* of the two PLDs. We state that formally and also recall the definition of convolution below.

Definition 2. *Let μ and μ' be any distributions on real numbers. Their convolution, denoted by $\mu * \mu'$, is a distribution on real numbers where a sample $t \sim \mu * \mu'$ is drawn by first independently sampling $a \sim \mu$, $a' \sim \mu'$ and then letting $t = a + a'$.*

Observation 3 ([SMM19])**.** *Let $\mu_{up}, \mu'_{up}, \mu_{lo}$ and μ'_{lo} be any distributions such that all of them are discrete or all are continuous. Then, we have*

$$PLD_{(\mu_{up} \otimes \mu'_{up})/(\mu_{lo} \otimes \mu'_{lo})} = PLD_{\mu_{up}/\mu_{lo}} * PLD_{\mu'_{up}/\mu'_{lo}}.$$

2.1 Composition via Privacy Loss Buckets

Observations 2 and 3 provide a way to compute the privacy parameters for compositions of multiple mechanisms: first we calculate the PLD of each mechanism, find their convolutions, and finally compute the ε-hockey stick divergence of their convolution. An issue here is that the trivial implementation of this algorithm is not efficient; for instance, PLD itself can be a continuous distribution which cannot be represented finitely. Another consideration is that the convolution of multiple PLDs may blow up the support size. (That is, if we compose k mechanisms each with PLD support size n, then the resulting PLD may have support as large as n^k.)

This brings us to an algorithm of Meiser and Mohammadi [MM18] called *Privacy Buckets*. This simple algorithm allows us to "approximate" PLDs in such a way that the convolution is efficient, and still gives good numerical approximation for privacy parameters. As its name suggest, privacy buckets rounds the value of the PLD into buckets, which are integer multiples of a chosen positive real number (called `value_discretization_interval` in our implementation). The point here is that, once the values are integer multiples of such a number, we may use (inverse) Fast Fourier Transform (FFT) to quickly compute the convolution. (The idea of using FFT has been suggested by [KJH19, KJPH20].) We basically implement this; there are several subtleties in the implementation, which are listed below:

- We separately account for the "infinity mass", i.e., the probability of $o \sim \mu_{up}$ such that $\mu_{lo}(o) = 0$.
- Our code allows one to compute both the "pessimistic" (i.e., safe) and "optimistic" estimates of the hockey stick divergence between μ_{up} and μ_{lo}. In the former case, the PLD values are rounded up. In the latter case, the PLD values are rounded down. The pessimistic estimate results in a larger δ value than the true value, whereas the optimistic estimate results in a smaller δ than the true value.
- To make the implementation efficient, we also make sure that the array is not too long. This is done by truncating any outcome o (resp. a set S of outcomes) such that $\mu_{up}(o)$ (resp. $\mu_{up}(S)$) is smaller than a certain threshold. For the pessimistic case, this mass is accounted in `infinity_mass`. For the optimistic case, this mass is completely thrown away.

3 PLDs of Specific Mechanisms

In this section, we calculate the privacy loss distributions for several well-known mechanisms. Throughout this section, we only consider a scalar-valued function f. Recall that its sensitivity is defined as $\Delta(f) := \max_{\mathbf{x},\mathbf{x}'} |f(\mathbf{x}) - f(\mathbf{x}')|$ where the maximum is over two neighboring datasets $\mathbf{x}$ and $\mathbf{x}'$.

3.1 Laplace Mechanism

The Laplace mechanism [DMNS06] simply outputs $f(x)+\mathrm{Lap}(b)$ where $\mathrm{Lap}(b)$ is the Laplace random variable with parameter b; its probability density function at point x is equal to $\frac{1}{2b} \cdot e^{-|x|/b}$.

In this case, the "worst case"[1] PLD of the mechanism is the same as the PLD between $\mathrm{Lap}(b)$ and $\mathrm{Lap}(b) + \Delta(f)$. Let $\tilde{\Delta} := \Delta(f)/b$. The aforementioned PLD is the same as the PLD between $\mathrm{Lap}(1)$ and $\mathrm{Lap}(1) + \tilde{\Delta}$. That is, the privacy loss variable is generated by first picking $x \sim \mathrm{Lap}(1)$ and letting the privacy loss be

$$\ln\left(\frac{\frac{1}{2}\cdot e^{-|x|}}{\frac{1}{2}\cdot e^{-|x-\tilde{\Delta}|}}\right) = |x-\tilde{\Delta}| - |x| = \begin{cases} \tilde{\Delta} & \text{if } x \le 0, \\ -\tilde{\Delta} & \text{if } x \ge \tilde{\Delta}, \\ \tilde{\Delta} - 2x & \text{if } -\tilde{\Delta} < x < \tilde{\Delta}. \end{cases}$$

3.2 Gaussian Mechanism

The Gaussian mechanism (see [BW18] and the references therein) simply outputs $f(x) + \mathcal{N}(0, \sigma^2)$ where $\mathcal{N}(0, \sigma^2)$ is the centered Gaussian random variable with standard deviation σ; its probability density function at point x is equal to $\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$.

Here, the worst-case PLD of the mechanism is the same as the PLD between $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(\Delta(f), \sigma^2)$. Let $\tilde{\Delta} := \Delta(f)/\sigma$. The aforementioned PLD is the same as the PLD between $\mathcal{N}(0, 1)$ and $\mathcal{N}(\Delta(f)/\sigma, 1)$. That is, the privacy loss variable is generated by first picking $x \sim \mathcal{N}(0, 1)$ and letting the privacy loss be

$$\ln\left(\frac{\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-x^2/2}}{\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-(x-\tilde{\Delta})^2/2}}\right) = \frac{\tilde{\Delta}}{2}\cdot\left(\tilde{\Delta} - 2x\right).$$

3.2.1 Calculating ε-hockey stick divergence of Gaussian Mechanism

The ϵ-hockey stick divergence between $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(\Delta(f), \sigma^2)$ is equal to the ϵ-hockey stick divergence between $\mathcal{N}(0, 1)$ and $\mathcal{N}(\tilde{\Delta}, 1)$. Let ϕ and Φ denote the PDF and CDF of the standard normal distribution respectively. We can write

$$\mathfrak{D}_{e^\varepsilon}(\mathcal{N}(0,1)\|\mathcal{N}(\tilde{\Delta},1)) = \int_{-\infty}^{\infty} [\phi(x) - e^\varepsilon \cdot \phi(x-\tilde{\Delta})]_+ dx.$$

Now, as stated above, $\frac{\phi(x)}{\phi(x-\tilde{\Delta})} = e^{\frac{\tilde{\Delta}}{2}\cdot(\tilde{\Delta}-2x)}$, which is greater than e^ε if and only if $x < x_{upper} := 0.5\tilde{\Delta} - \varepsilon/\tilde{\Delta}$. As a result, we have

$$\begin{aligned}\mathfrak{D}_{e^\varepsilon}(\mathcal{N}(0,1)\|\mathcal{N}(\tilde{\Delta},1)) &= \int_{-\infty}^{x_{upper}} \left(\phi(x) - e^\varepsilon\cdot\phi(x-\tilde{\Delta})\right) dx. \\ &= \Phi(x_{upper}) - e^\varepsilon\cdot\Phi(x_{upper}-\tilde{\Delta}). \end{aligned} \tag{1}$$

3.3 Discrete Laplace Mechanism

The Discrete Laplace Mechanism (also known as the Symmetric Geometric Mechanism; e.g., see [GRS12]) outputs $f(x) + \mathrm{DLap}(a)$ where $\mathrm{DLap}(a)$ is the Discrete Laplace distribution with parameter a; its probability mass function at $x \in \mathbb{Z}$ is $\frac{e^a-1}{e^a+1}\cdot e^{-a|x|}$. (For simplicity, we assume that the image of f is a subset of the integers.)

[1] See section 3.5 for more discussion on "worst case" PLD.

In this case, the PLD of the mechanism is the same as the PLD between $\mathrm{DLap}(a)$ and $\mathrm{DLap}(a)+\Delta(f)$. That is, the privacy loss variable is generated by first picking $x \sim \mathrm{DLap}(a)$ and letting the privacy loss be

$$\ln\left(\frac{\frac{e^a-1}{e^a+1}\cdot e^{-a|x|}}{\frac{e^a-1}{e^a+1}\cdot e^{-a|x-\Delta(f)|}}\right) = a(|x-\Delta(f)|-|x|) = \begin{cases} a\cdot\Delta(f) & \text{if } x \le 0, \\ -a\cdot\Delta(f) & \text{if } x \ge \Delta(f), \\ a(\Delta(f)-2x) & \text{if } -\Delta(f) < x < \Delta(f). \end{cases}$$

3.4 k-Randomized Response

In the k-Randomized Response [War65], the input is one of k values. The protocol outputs the input with probability $1-p$. With the remaining probability p, the protocol outputs a uniformly random element from the k possible values (including the input itself).

Let $\mathcal{R}_k$ denote the randomized response. In this case, the PLD of the mechanism is equal to the PLD between $\mathcal{R}_k(x)$ and $\mathcal{R}_k(x')$ where x and x' are two distinct inputs. That is, the privacy loss variable is generated by first picking $o \sim \mathcal{R}_k(x)$ and letting it be

$$\ln\left(\frac{\Pr[\mathcal{R}_k(x)=o]}{\Pr[\mathcal{R}_k(x')=o]}\right) = \begin{cases} \ln\left(\frac{k(1-p)+p}{p}\right) & \text{if } o = x, \\ \ln\left(\frac{p}{k(1-p)+p}\right) & \text{if } o = x', \\ 0 & \text{if } o \notin \{x, x'\}. \end{cases}$$

In other words, the privacy loss variable is equal to

$$\begin{cases} \ln\left(\frac{k(1-p)+p}{p}\right) & \text{with probability } 1-p+\frac{p}{k}, \\ \ln\left(\frac{p}{k(1-p)+p}\right) & \text{with probability } \frac{p}{k}, \\ 0 & \text{with probability } \frac{p(k-2)}{k}. \end{cases}$$

3.5 Pessimistic PLD for (ε, δ)-DP Algorithms

In some scenarios, we may not know the specific algorithm being applied or it may be hard to write down the PLD exactly, but we do know that the algorithm is (ε, δ)-DP. In this case, it is possible to define the "worst case" PLD, which is a pessimistic estimate of the true PLD. Specifically, [KOV15] proves the following:[2]

Theorem 1. *For any (ε, δ)-DP mechanism $\mathcal{M}$ and neighboring input datasets $\mathbf{x}, \mathbf{x}'$, Let $\mathcal{M}^*$ be the following algorithm:*

$$\begin{aligned} &\Pr[\mathcal{M}^*(\mathbf{x})=0] = \delta, && \Pr[\mathcal{M}^*(\mathbf{x})=0] = 0, \\ &\Pr[\mathcal{M}^*(\mathbf{x})=1] = (1-\delta)\cdot\frac{e^\varepsilon}{1+e^\varepsilon}, && \Pr[\mathcal{M}^*(\mathbf{x})=1] = (1-\delta)\cdot\frac{1}{1+e^\varepsilon}, \\ &\Pr[\mathcal{M}^*(\mathbf{x})=2] = (1-\delta)\cdot\frac{1}{1+e^\varepsilon}, && \Pr[\mathcal{M}^*(\mathbf{x})=2] = (1-\delta)\cdot\frac{e^\varepsilon}{1+e^\varepsilon}, \\ &\Pr[\mathcal{M}^*(\mathbf{x})=3] = 0, && \Pr[\mathcal{M}^*(\mathbf{x})=3] = \delta. \end{aligned}$$

Then, there exists an algorithm T such that $T(\mathcal{M}^*(\mathbf{x}))$ and $T(\mathcal{M}^*(\mathbf{x}'))$ are identically distributed as $\mathcal{M}(\mathbf{x})$ and $\mathcal{M}(\mathbf{x}')$ respectively.

[2] See also [MV18] for an alternative proof.

By post-processing property of differential privacy, the above theorem means that $\mathcal{M}$ is more private than $\mathcal{M}^*$. As such, we may use the PLD of $\mathcal{M}^*$ as a pessimistic estimate of the PLD of $\mathcal{M}$. The privacy loss of $\mathcal{M}^*$ is equal to

$$\begin{cases} \infty & \text{with probability } \delta, \\ \varepsilon & \text{with probability } (1-\delta) \cdot \frac{e^\varepsilon}{1+e^\varepsilon}, \\ -\varepsilon & \text{with probability } (1-\delta) \cdot \frac{1}{1+e^\varepsilon}. \end{cases}$$

References

[BW18] Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. *arXiv preprint arXiv:1805.06530*, 2018.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

[DR16] Cynthia Dwork and Guy N. Rothblum. Concentrated differential privacy. *CoRR*, abs/1603.01887, 2016.

[GRS12] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

[KJH19] Antti Koskela, Joonas Jälkö, and Antti Honkela. Computing exact guarantees for differential privacy. *arXiv preprint arXiv:1906.03049*, 2019.

[KJPH20] Antti Koskela, Joonas Jälkö, Lukas Prediger, and Antti Honkela. Tight approximate differential privacy for discrete-valued mechanisms using FFT. *arXiv preprint arXiv:2006.07134*, 2020.

[KOV15] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *ICML*, pages 1376–1385, 2015.

[MM18] Sebastian Meiser and Esfandiar Mohammadi. Tight on budget?: Tight bounds for r-fold approximate differential privacy. In *CCS*, pages 247–264, 2018.

[MV18] Jack Murtagh and Salil P. Vadhan. The complexity of computing the optimal composition of differential privacy. *Theory Comput.*, 14(1):1–35, 2018.

[SMM19] David M. Sommer, Sebastian Meiser, and Esfandiar Mohammadi. Privacy loss classes: The central limit theorem in differential privacy. *PoPETs*, 2019(2):245–269, 2019.

[War65] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.