DOI: 10.1561/9781638284772.ch3

Chapter 3

Composition of Differential Privacy & Privacy Amplification by Subsampling

By Thomas Steinke

3.1 Introduction

Our data is subject to many different uses. Many entities will have access to our data, including government agencies, healthcare providers, employers, technology companies, and financial institutions. Those entities will perform many different analyses that involve our data and those analyses will be updated repeatedly over our lifetimes. The greatest risk to privacy is that an attacker will combine multiple pieces of information from the same or different sources and that the combination of these will reveal sensitive details about us. Thus we cannot study privacy leakage in a vacuum; it is important that we can reason about the accumulated privacy leakage over multiple independent analyses.

As a concrete example to keep in mind, consider the following simple differencing attack: Suppose your employer provides healthcare benefits. The employer pays for these benefits and thus may have access to summary statistics like how many employees are currently receiving pre-natal care or currently are being treated for cancer. Your pregnancy or cancer status is highly sensitive information, but intuitively the aggregated count is not sensitive as it is not specific to you. However, this count may be updated on a regular basis and your employer may notice that the count increased on the day you were hired or on the day you took off for a medical appointment. This example shows how multiple pieces of information – the date

of your hire or medical appointment, the count before that date, and the count afterwards – can be combined to reveal sensitive information about you, despite each piece of information seeming innocuous on its own. Attacks could combine many different statistics from multiple sources and hence we need to be careful to guard against such attacks, which leads us to differential privacy.

Differential privacy has strong composition properties – if multiple independent analyses are run on our data and each analysis is differentially private on its own, then the combination of these analyses is also differentially private. This property is key to the success of differential privacy. Composition enables building complex differentially private systems out of simple differentially private subroutines. Composition allows the re-use data over time without fear of a catastrophic privacy failure. And, when multiple entities use the data of the same individuals, they do not need to coordinate to prevent an attacker from learning private details of individuals by combining the information released by those entities. To prevent the above differencing attack, we could independently perturb each count to make it differentially private; then taking the difference of two counts would be sufficiently noisy to obscure your pregnancy or cancer status.

Composition is quantitative. The differential privacy guarantee of the overall system will depend on the number of analyses and the privacy parameters that they each satisfy. The exact relationship between these quantities can be complex. There are various composition theorems that give bounds on the overall parameters in terms of the parameters of the parts of the system. In this chapter, we will study several composition theorems (including the relevant proofs) and we will also look at some examples that demonstrate how to apply the composition theorems and why we need them.

Composition theorems provide privacy bounds for a given system. A system designer must use composition theorems to design systems that simultaneously give good privacy and good utility (i.e., good statistical accuracy). This process often called "privacy budgeting" or "privacy accounting." Intuitively, the system designer has some privacy constraint (i.e., the overall system must satisfy some final privacy guarantee) which can be viewed as analogous to a monetary budget that must be divided amongst the various parts of the system. Composition theorems provide the accounting rules for this budget. Allocating more of the budget to some part of the system makes that part more accurate, but then less budget is available for other parts of the system. Thus the system designer must also make a value judgement about which parts of the system to prioritize.

Overview of the Chapter

This chapter provides an in-depth discussion of composition theorems and privacy amplification techniques in Differential Privacy. It begins by introducing the basic

composition theorem in Section 3.2, and examining whether basic composition strategies achieve optimal privacy guarantees. Next, in Section 3.3, it reviews the concept of privacy loss distributions and offers a statistical hypothesis testing perspective to understand approximate Differential Privacy. The chapter then discusses advanced composition via the privacy loss distribution, in Section 3.4, revisiting basic composition and exploring composition through Gaussian approximation. It reviews the notion of Concentrated Differential Privacy, adaptive composition, and post-processing, and examines the composition of approximate Differential Privacy. Then, the chapter focuses on privacy amplification by subsampling, in Section 3.6, covering subsampling techniques for pure and approximate Differential Privacy, the differences between addition/removal and replacement for neighboring datasets, and how subsampling interacts with composition. Here, the concept of Rényi Differential Privacy is introduced, along with analytic bounds for privacy amplification and practical guidance on the use of privacy amplification by subsampling in real-world applications. Finally, the chapter concludes, in Section 3.7, with a reflection on the historical development of the discussed concepts and provides further reading for a deeper understanding of these concepts.

3.2 Basic Composition

The simplest composition theorem is what is known as basic composition. This applies to pure ε-DP (although it can be extended to approximate (ε, δ)-DP). Basic composition says that, if we run k independent ε-DP algorithms, then the composition of these is $k\varepsilon$-DP. More generally, we have the following result.

Theorem 3.1 (Basic Composition)**.** *Let $M_1, M_2, \cdots, M_k : \mathcal{X}^n \to \mathcal{Y}$ be randomized algorithms. Suppose M_j is ε_j-DP for each $j \in [k]$. Define $M : \mathcal{X}^n \to \mathcal{Y}^k$ by $M(x) = (M_1(x), M_2(x), \cdots, M_k(x))$, where each algorithm is run independently. Then M is ε-DP for $\varepsilon = \sum_{j=1}^{k} \varepsilon_j$.*

Proof. Fix an arbitrary pair of neighboring datasets $x, x' \in \mathcal{X}^n$ and output $y \in \mathcal{Y}^k$. To establish that M is ε-DP, we must show that $e^{-\varepsilon} \leq \frac{\mathbb{P}[M(x)=y]}{\mathbb{P}[M(x')=y]} \leq e^{\varepsilon}$. By independence, we have

$$\frac{\mathbb{P}\left[M(x) = y\right]}{\mathbb{P}\left[M(x') = y\right]} = \frac{\prod_{j=1}^{k} \mathbb{P}\left[M_j(x) = y_j\right]}{\prod_{j=1}^{k} \mathbb{P}\left[M_j(x') = y_j\right]}$$

$$= \prod_{j=1}^{k} \frac{\mathbb{P}\left[M_j(x) = y_j\right]}{\mathbb{P}\left[M_j(x') = y_j\right]} \leq \prod_{j=1}^{k} e^{\varepsilon_j} = e^{\sum_{j=1}^{k} \varepsilon_j} = e^{\varepsilon},$$

where the inequality follows from the fact that each M_j is ε_j-DP and, hence, $e^{-\varepsilon_j} \leq \frac{\mathbb{P}[M_j(x)=y_j]}{\mathbb{P}[M_j(x')=y_j]} \leq e^{\varepsilon_j}$. Similarly, $\prod_{j=1}^{k} \frac{\mathbb{P}[M_j(x)=y_j]}{\mathbb{P}[M_j(x')=y_j]} \geq \prod_{j=1}^{k} e^{-\varepsilon_j}$, which completes the proof. □

Basic composition is already a powerful result, despite its simple proof; it establishes the versatility of differential privacy and allows us to begin reasoning about complex systems in terms of their building blocks. For example, suppose we have k functions $f_1, \cdots, f_k : \mathcal{X}^n \to \mathbb{R}$ each of sensitivity 1. For each $j \in [k]$, we know that adding Laplace$(1/\varepsilon)$ noise to the value of $f_j(x)$ satisfies ε-DP. Thus, if we add independent Laplace$(1/\varepsilon)$ noise to each value $f_j(x)$ for all $j \in [k]$, then basic composition tells us that releasing this vector of k noisy values satisfies $k\varepsilon$-DP. If we want the overall system to be ε-DP, then we should add independent Laplace(k/ε) noise to each value $f_j(x)$.

3.2.1 Is Basic Composition Optimal?

If we want to release k values each of sensitivity 1 (as above) and have the overall release be ε-DP, then, using basic composition, we can add Laplace(k/ε) noise to each value. The variance of the noise for each value is $2k^2/\varepsilon^2$, so the standard deviation is $\sqrt{2}k/\varepsilon$. In other words, the scale of the noise must grow linearly with the number of values k if the overall privacy and each value's sensitivity is fixed. It is natural to wonder whether the scale of the Laplace noise can be reduced by improving the basic composition result. We now show that this is not possible.

For each $j \in [k]$, let $M_j : \mathcal{X}^n \to \mathbb{R}$ be the algorithm that releases $f_j(x)$ with Laplace(k/ε) noise added. Let $M : \mathcal{X}^n \to \mathbb{R}^k$ be the composition of these k algorithms. Then M_j is ε/k-DP for each $j \in [k]$ and basic composition tells us that M is ε-DP. The question is whether M satisfies a better DP guarantee than this – i.e., does M satisfy ε_*-DP for some $\varepsilon_* < \varepsilon$? Suppose we have neighboring datasets $x, x' \in \mathcal{X}^n$ such that $f_j(x) = f_j(x') + 1$ for each $j \in [k]$. Let $y = (a, a, \cdots, a) \in \mathbb{R}^k$ for some $a \geq \max_{j=1}^{k} f_j(x)$. Then

$$\begin{aligned}
\frac{\mathbb{P}\left[M(x) = y\right]}{\mathbb{P}\left[M(x') = y\right]} &= \frac{\prod_{j=1}^{k} \mathbb{P}\left[f_j(x) + \text{Laplace}(k/\varepsilon) = y_j\right]}{\prod_{j=1}^{k} \mathbb{P}\left[f_j(x') + \text{Laplace}(k/\varepsilon) = y_j\right]} \\
&= \prod_{j=1}^{k} \frac{\mathbb{P}\left[\text{Laplace}(k/\varepsilon) = y_j - f_j(x)\right]}{\mathbb{P}\left[\text{Laplace}(k/\varepsilon) = y_j - f_j(x')\right]} \\
&= \prod_{j=1}^{k} \frac{\frac{\varepsilon}{2k} \exp\left(-\frac{\varepsilon}{k}|y_j - f_j(x)|\right)}{\frac{\varepsilon}{2k} \exp\left(-\frac{\varepsilon}{k}|y_j - f_j(x')|\right)}
\end{aligned}$$

$$= \prod_{j=1}^{k} \frac{\exp\left(-\frac{\varepsilon}{k}(y_j - f_j(x))\right)}{\exp\left(-\frac{\varepsilon}{k}(y_j - f_j(x'))\right)} \qquad (y_j \geq f_j(x) \text{ and } y_j \geq f_j(x'))$$

$$= \prod_{j=1}^{k} \exp\left(\frac{\varepsilon}{k}\left(f_j(x) - f_j(x')\right)\right)$$

$$= \exp\left(\frac{\varepsilon}{k}\sum_{j=1}^{k}\left(f_j(x) - f_j(x')\right)\right) = e^{\varepsilon}.$$

This shows that basic composition is optimal. For this example, we cannot prove a better guarantee than what is given by basic composition.

Is there some other way to improve upon basic composition that circumvents this example? Note that we assumed that there are neighboring datasets $x, x' \in \mathcal{X}^n$ such that $f_j(x) = f_j(x') + 1$ for each $j \in [k]$. In some settings, no such worst case datasets exist. In that case, instead of scaling the noise linearly with k, we can scale the Laplace noise according to the ℓ_1 sensitivity $\Delta_1 := \sup_{\substack{x,x' \in \mathcal{X}^n \\ \text{neighboring}}} \sum_{j=1}^{k} |f_j(x) - f_j(x')|$.

Instead of adding assumptions to the problem, we will look more closely at the example above. We showed that there exists some output $y \in \mathbb{R}^d$ such that $\frac{\mathbb{P}[M(x)=y]}{\mathbb{P}[M(x')=y]} = e^{\varepsilon}$. However, such outputs y are very rare, as we require $y_j \geq \max\{f_j(x), f_j(x')\}$ for each $j \in [k]$ where $y_j = f_j(x) + \text{Laplace}(k/\varepsilon)$. Thus, in order to observe an output y such that the likelihood ratio is maximal, all of the k Laplace noise samples must be positive, which happens with probability 2^{-k}. The fact that outputs y with maximal likelihood ratio are exceedingly rare turns out to be a general phenomenon and not specific to the example above.

Can we improve on basic composition if we only ask for a high probability bound? That is, instead of demanding $\frac{\mathbb{P}[M(x)=y]}{\mathbb{P}[M(x')=y]} \leq e^{\varepsilon_*}$ for all $y \in \mathcal{Y}$, we demand $\mathbb{P}_{Y \leftarrow M(x)}\left[\frac{\mathbb{P}[M(x)=Y]}{\mathbb{P}[M(x')=Y]} \leq e^{\varepsilon_*}\right] \geq 1 - \delta$ for some $0 < \delta \ll 1$. Can we prove a better bound $\varepsilon_* < \varepsilon$ in this relaxed setting? The answer turns out to be yes.

The limitation of pure ε-DP is that events with tiny probability – which are negligible in real-world applications – can dominate the privacy analysis. This motivates us to move to relaxed notions of differential privacy, such as approximate (ε, δ)-DP and concentrated DP, which are less sensitive to low probability events. In particular, these relaxed notions of differential privacy allow us to prove quantitatively better composition theorems. The rest of this chapter develops this direction further.

3.3 Privacy Loss Distributions

Qualitatively, an algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ is differentially private if, for all neighboring datasets $x, x' \in \mathcal{X}^n$, the output distributions $M(x)$ and $M(x')$ are "indistinguishable" or "close." The key question is how do we quantify the closeness or indistinguishability of a pair of distributions?

Pure DP (a.k.a. pointwise DP) [DMNS06] uniformly bounds the likelihood ratio $-\frac{\mathbb{P}[M(x)=y]}{\mathbb{P}[M(x')=y]} \leq e^{\varepsilon}$ for all $y \in \mathcal{Y}$. As discussed at the end of the section on basic composition (Section 3.2), this can be too strong as the outputs y that maximize this likelihood ratio may be very rare.

We could also consider the total variation distance (a.k.a. statistical distance):

$$\mathrm{d}_{\mathrm{TV}}\left(M(x), M(x')\right) := \sup_{S \subset \mathcal{Y}} \left(\mathbb{P}\left[M(x) \in S\right] - \mathbb{P}\left[M(x') \in S\right]\right).$$

Another option would be the KL divergence (a.k.a. relative entropy). Both TV distance and KL divergence turn out to give poor privacy-utility tradeoffs; that is, to rule out bad algorithms M, we must set these parameters very small, but that also rules out all the good algorithms. Intuitively, both TV and KL are not sensitive enough to low-probability bad events (whereas pure DP is too sensitive). We need to introduce a parameter (δ) to determine what level of low probability events we can ignore.

Approximate (ε, δ)-DP [Dwo+06] is a combination of pure ε-DP and δ TV distance. Specifically, M is (ε, δ)-DP if, for all neighboring datasets $x, x' \in \mathcal{X}^n$ and all measurable $S \subset \mathcal{Y}$, $\mathbb{P}\left[M(x) \in S\right] \leq e^{\varepsilon} \cdot \mathbb{P}\left[M(x') \in S\right] + \delta$. Intuitively, (ε, δ)-DP is like ε-DP except we can ignore events with probability $\leq \delta$. That is, δ represents a failure probability, so it should be small (e.g., $\delta \leq 10^{-6}$), while ε can be larger (e.g., $\varepsilon \approx 1$); having two parameters with very different values allows us to circumvent the limitations of either pure DP or TV distance as a similarity measure.

All of these options for quantifying indistinguishability can be viewed from the perspective of the privacy loss distribution. The privacy loss distribution also turns out to be essential to the analysis of composition. Approximate (ε, δ)-DP bounds are usually proved via the privacy loss distribution.

We now formally define the privacy loss distribution and relate it to the various quantities we have considered. Then, in Section 3.3.1, we will calculate the privacy loss distribution corresponding to the Gaussian mechanism, which is a particularly nice example. In the next Section 3.3.2, we explain how the privacy loss distribution arises naturally via statistical hypothesis testing. To conclude, in Section 3.3.3, we precisely relate the privacy loss back to approximate (ε, δ)-DP. In the next section

(Section 3.4), we will use the privacy loss distribution as a tool to analyze composition.

Definition 3.2 (Privacy Loss Distribution)**.** *Let P and Q be two probability distributions on $\mathcal{Y}$. Define $f_{P\|Q} : \mathcal{Y} \to \mathbb{R}$ by $f_{P\|Q}(y) = \log(P(y)/Q(y))$.*[i] *The privacy loss random variable is given by $Z = f_{P\|Q}(Y)$ for $Y \leftarrow P$. The distribution of Z is denoted* PrivLoss$(P\|Q)$.

In the context of differential privacy, the distributions $P = M(x)$ and $Q = M(x')$ correspond to the outputs of the algorithm M on neighboring inputs x, x'. Successfully distinguishing these distributions corresponds to learning some fact about an individual person's data. The randomness of the privacy loss random variable Z comes from the randomness of the algorithm M (e.g., added noise). Intuitively, the privacy loss tells us which input (x or x') is more likely given the observed output ($Y \leftarrow M(\cdot)$). If $Z > 0$, then the hypothesis $Y \leftarrow P = M(x)$ explains the observed output better than the hypothesis $Y \leftarrow Q = M(x')$ and vice versa. The magnitude of the privacy loss Z indicates how strong the evidence for this conclusion is. If $Z = 0$, both hypotheses explain the output equally well, but, if $Z \to \infty$, then we can be nearly certain that the output came from P, rather than Q. A very negative privacy loss $Z \ll 0$ means that the observed output $Y \leftarrow P$ strongly supports the wrong hypothesis (i.e., $Y \leftarrow Q$).

As long as the privacy loss distribution is well-defined,[ii] we can easily express almost all the quantities of interest in terms of it:

- Pure ε-DP of M is equivalent to demanding that $\mathbb{P}_{Z \leftarrow \text{PrivLoss}(M(x)\|M(x'))}[Z \leq \varepsilon] = 1$ for all neighboring x, x'.[iii]

i. The function $f_{P\|Q}$ is called the log likelihood ratio of P with respect to Q. Formally, $f_{P\|Q}$ is the natural logarithm of the Radon-Nikodym derivative of P with respect to Q. This function is defined by the property that $P(S) = \mathbb{E}_{Y \leftarrow P}[\mathbb{I}[Y \in S]] = \mathbb{E}_{Y \leftarrow Q}\left[e^{f_{P\|Q}(Y)} \cdot \mathbb{I}[Y \in S]\right]$ for all measurable $S \subset \mathcal{Y}$. For this to exist, we must assume that P and Q have the same sigma-algebra and that P is absolutely continuous with respect to Q and vice versa – i.e., $\forall S \subset \mathcal{Y} \quad Q(S) = 0 \iff P(S) = 0$.

ii. The privacy loss distribution is not well-defined if absolute continuity fails to hold. Intuitively, this corresponds to the privacy loss being infinite. We can extend most of these definitions to allow for an infinite privacy loss. For simplicity, we do not delve into these issues.

iii. Note that, by the symmetry of the neighboring relation (i.e., if x, x' are neighboring datasets then x', x are also neighbors), we also have $\mathbb{P}_{Z \leftarrow \text{PrivLoss}(M(x)\|M(x'))}[Z \geq -\varepsilon] = 1$ as a consequence of $\mathbb{P}_{Z' \leftarrow \text{PrivLoss}(M(x')\|M(x))}\left[Z' \leq \varepsilon\right] = 1$.

- The KL divergence is the expectation of the privacy loss: $\mathrm{D}_1(P\|Q) := \mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(P\|Q)}[Z]$.[iv]
- The TV distance is given by

$$\begin{aligned}\mathrm{d}_{\mathrm{TV}}(P,Q) &= \mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(P\|Q)}\left[\max\{0,1-\exp(-Z)\}\right]\\ &= \frac{1}{2}\mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(P\|Q)}\left[\left|1-\exp(-Z)\right|\right].\end{aligned}$$

- Approximate (ε,δ)-DP of M is implied by $\mathop{\mathbb{P}}_{Z\leftarrow\mathrm{PrivLoss}(M(x)\|M(x'))}[Z\le\varepsilon]\ge 1-\delta$ for all neighboring x,x'. So we should think of approximate DP as a tail bound on the privacy loss. To be precise, (ε,δ)-DP of M is equivalent to

$$\mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(M(x)\|M(x'))}\left[\max\{0,1-\exp(\varepsilon-Z)\}\right]\le\delta,$$

for all neighboring x,x'. (See Proposition 3.7.)

3.3.1 Privacy Loss of Gaussian Noise Addition

As an example, we will work out the privacy loss distribution corresponding to the addition of Gaussian noise to a bounded-sensitivity query. This example is particularly clean, as the privacy loss distribution is also a Gaussian, and it will turn out to be central to the story of composition.

Proposition 3.3 (Privacy Loss Distribution of Gaussian)**.** *Let* $P=\mathcal{N}(\mu,\sigma^2)$ *and* $Q=\mathcal{N}(\mu',\sigma^2)$. *Then* $\mathrm{PrivLoss}(P\|Q)=\mathcal{N}(\rho,2\rho)$ *for* $\rho=\frac{(\mu-\mu')^2}{2\sigma^2}$.

Proof. We have $P(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ and $Q(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu')^2}{2\sigma^2}\right)$. Thus the log likelihood ratio is

$$\begin{aligned}f_{P\|Q}(y) &= \log\left(\frac{P(y)}{Q(y)}\right)\\ &= \log\left(\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu')^2}{2\sigma^2}\right)}\right)\\ &= -\frac{(y-\mu)^2}{2\sigma^2}+\frac{(y-\mu')^2}{2\sigma^2}\end{aligned}$$

iv. The expectation of the privacy loss is always non-negative. Intuitively, this is because we take the expectation of the log likelihood ratio $f_{P\|Q}(Y)$ with respect to $Y\leftarrow P$ – i.e., the true answer is P, so on average the log likelihood ratio should point towards the correct answer.

$$= \frac{(y^2 - 2\mu' y + \mu'^2) - (y^2 - 2\mu y + \mu^2)}{2\sigma^2}$$
$$= \frac{2(\mu - \mu')y - \mu^2 + \mu'^2}{2\sigma^2}$$
$$= \frac{(\mu - \mu')(2y - \mu - \mu')}{2\sigma^2}.$$

The log likelihood ratio $f_{P\|Q}$ is an affine linear function. Thus the privacy loss random variable $Z = f_{P\|Q}(Y)$ for $Y \leftarrow P = \mathcal{N}(\mu, \sigma^2)$ will also follow a Gaussian distribution. Specifically, $\mathbb{E}[Y] = \mu$, so

$$\mathbb{E}[Z] = \frac{(\mu - \mu')(2\mathbb{E}[Y] - \mu - \mu')}{2\sigma^2} = \frac{(\mu - \mu')^2}{2\sigma^2} = \rho$$

and, similarly, $\mathbb{V}[Y] = \sigma^2$, so

$$\mathbb{V}[Z] = \frac{((2(\mu - \mu'))^2}{(2\sigma^2)^2} \cdot \mathbb{V}[Y] = \frac{(\mu - \mu')^2}{\sigma^2} = 2\rho.$$

□

To relate Proposition 3.3 to the standard Gaussian mechanism $M : \mathcal{X}^n \to \mathbb{R}$, recall that $M(x) = \mathcal{N}(q(x), \sigma^2)$, where q is a sensitivity-Δ query – i.e., $|q(x) - q(x')| \leq \Delta$ for all neighboring datasets $x, x' \in \mathcal{X}^n$. Thus, for neighboring datasets x, x', we have $\text{PrivLoss}\left(M(x)\middle\|M(x')\right) = \mathcal{N}(\rho, 2\rho)$ for some $\rho \leq \frac{\Delta^2}{2\sigma^2}$.

The privacy loss of the Gaussian mechanism is unbounded; thus it does not satisfy pure ε-DP. However, the Gaussian distribution is highly concentrated, so we can say that with high probability the privacy loss is not too large. This is the basis of the privacy guarantee of the Gaussian mechanism.

3.3.2 Statistical Hypothesis Testing Perspective

To formally quantify differential privacy, we must measure the closeness or indistinguishability of the distributions $P = M(x)$ and $Q = M(x')$ corresponding to the outputs of the algorithm M on neighboring inputs x, x'. Distinguishing a pair of distributions is precisely the problem of (simple) hypothesis testing in the field of statistical inference. Thus it is natural to look at hypothesis testing tools to quantify the (in)distinguishability of a pair of distributions.

In the language of hypothesis testing, the two distributions P and Q would be the null hypothesis and the alternate hypothesis, which correspond to a positive or negative example. We are given a sample Y drawn from one of the two distributions and our task is to determine which. Needless to say, there is, in general, no

hypothesis test that perfectly distinguishes the two distributions and, when choosing a hypothesis test, we face a non-trivial tradeoff between false positives and false negatives. There are many different ways to measure how good a given hypothesis test is.

For example, we could measure the accuracy of the hypothesis test evenly averaged over the two distributions. In this case, given the sample Y, an optimal test chooses P if $P(Y) \geq Q(Y)$ and otherwise chooses Q; the accuracy of this test is

$$\frac{1}{2}\mathop{\mathbb{P}}_{Y \leftarrow P}[P(Y) \geq Q(Y)] + \frac{1}{2}\mathop{\mathbb{P}}_{Y \leftarrow Q}[P(Y) < Q(Y)] = \frac{1}{2} + \frac{1}{2}\mathrm{d}_{\mathrm{TV}}(P, Q).$$

This measure of accuracy thus corresponds to TV distance. The greater the TV distance between the distributions, the more accurate this test is. However, as we mentioned earlier, TV distance does not yield good privacy-utility tradeoffs. Intuitively, the problem is that this hypothesis test doesn't care about how confident we are. That is, the test only asks whether $P(Y) \geq Q(Y)$, but not how big the difference or ratio is. Hence we want a more refined measure of accuracy that does not count false positives and false negatives equally.

Regardless of how we measure how good the hypothesis test is, there is an optimal test statistic, namely the log likelihood ratio. This test statistic gives a real number and thresholding that value yields a binary hypothesis test; *any* binary hypothesis test is dominated by some value of the threshold. In other words, the tradeoff between false positives and false negatives reduces to picking a threshold. This remarkable – yet simple – fact is established by the Neyman-Pearson lemma:

Lemma 3.4 (Neyman-Pearson Lemma [NP33])**.** *Fix distributions P and Q on $\mathcal{Y}$ and define the log-likelihood ratio test statistic $f_{P\|Q} : \mathcal{Y} \to \mathbb{R}$ by $f_{P\|Q}(y) = \log\left(\frac{P(y)}{Q(y)}\right)$. Let $T : \mathcal{Y} \to \{P, Q\}$ be any (possibly randomized) test. Then there exists some $t \in \mathbb{R}$ such that*

$$\mathop{\mathbb{P}}_{Y \leftarrow P}[T(Y) = P] \leq \mathop{\mathbb{P}}_{Y \leftarrow P}\left[f_{P\|Q}(Y) \geq t\right] \quad \textit{and}$$

$$\mathop{\mathbb{P}}_{Y \leftarrow Q}[T(Y) = Q] \leq \mathop{\mathbb{P}}_{Y \leftarrow Q}\left[f_{P\|Q}(Y) \leq t\right].$$

How is this related to the privacy loss distribution? The test statistic $Z = f_{P\|Q}(Y)$ under the hypothesis $Y \leftarrow P$ is precisely the privacy loss random variable $Z \leftarrow \mathrm{PrivLoss}(P\|Q)$. Thus the Neyman-Pearson lemma tells us that the privacy loss distribution $\mathrm{PrivLoss}(P\|Q)$ captures everything we need to know about distinguishing P from Q.

Note that the Neyman-Pearson lemma also references the test statistic $f_{P\|Q}(Y)$ under the hypothesis $Y \leftarrow Q$. This is fundamentally not that different from the

privacy loss. There are two ways we can relate this quantity back to the usual privacy loss: First, we can relate it to PrivLoss $(Q\|P)$ and this distribution is something we should be able to handle due to the symmetry of differential privacy guarantees.

Remark 3.5. *Fix distributions P and Q on $\mathcal{Y}$ such that the log likelihood ratio $f_{P\|Q}(y) = \log\left(\frac{P(y)}{Q(y)}\right)$ is well-defined for all $y \in \mathcal{Y}$. Since $f_{P\|Q}(y) = -f_{Q\|P}(y)$ for all $y \in \mathcal{Y}$, if $Z \leftarrow \text{PrivLoss}(Q\|P)$, then $-Z$ follows the distribution of $f_{P\|Q}(Y)$ under the hypothesis $Y \leftarrow Q$.*

Second, if we need to compute an expectation of some function g of $f_{P\|Q}(Y)$ under the hypothesis $Y \leftarrow Q$, then we can still express this in terms of the privacy loss PrivLoss $(P\|Q)$:

Lemma 3.6 (Change of Distribution for Privacy Loss)**.** *Fix distributions P and Q on $\mathcal{Y}$ such that the log likelihood ratio $f_{P\|Q}(y) = \log\left(\frac{P(y)}{Q(y)}\right)$ is well-defined for all $y \in \mathcal{Y}$. Let $g : \mathbb{R} \to \mathbb{R}$ be measurable. Then*

$$\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[g(f_{P\|Q}(Y))\right] = \mathop{\mathbb{E}}_{Z\leftarrow \text{PrivLoss}(P\|Q)}\left[g(Z)\cdot e^{-Z}\right].$$

Proof. By the definition of the log likelihood ratio (see Definition 3.2), we have $\mathop{\mathbb{E}}_{Y\leftarrow P}[h(Y)] = \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[h(Y)\cdot e^{f_{P\|Q}(Y)}\right]$ for all measurable functions h. Setting $h(y) = g(f_{P\|Q}(y))\cdot e^{-f_{P\|Q}(y)}$ yields $\mathop{\mathbb{E}}_{Z\leftarrow \text{PrivLoss}(P\|Q)}\left[g(Z)\cdot e^{-Z}\right] = \mathop{\mathbb{E}}_{Y\leftarrow P}[h(Y)] = \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[h(Y)\cdot e^{f_{P\|Q}(Y)}\right] = \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[g(f_{P\|Q}(Y))\right]$, as required. We can also write these expressions out as an integral to obtain a more intuitive proof:

$$\begin{aligned}
\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[g(f_{P\|Q}(Y))\right] &= \int_{\mathcal{Y}} g(f_{P\|Q}(y))\cdot Q(y)\mathrm{d}y \\
&= \int_{\mathcal{Y}} g(f_{P\|Q}(y))\cdot \frac{Q(y)}{P(y)}\cdot P(y)\mathrm{d}y \\
&= \int_{\mathcal{Y}} g(f_{P\|Q}(y))\cdot e^{-\log(P(y)/Q(y))}\cdot P(y)\mathrm{d}y \\
&= \int_{\mathcal{Y}} g(f_{P\|Q}(y))\cdot e^{-f_{P\|Q}(y)}\cdot P(y)\mathrm{d}y \\
&= \mathop{\mathbb{E}}_{Y\leftarrow P}\left[g(f_{P\|Q}(Y))\cdot e^{-f_{P\|Q}(Y)}\right] \\
&= \mathop{\mathbb{E}}_{Z\leftarrow \text{PrivLoss}(P\|Q)}\left[g(Z)\cdot e^{-Z}\right].
\end{aligned}$$

□

3.3.3 Approximate DP & the Privacy Loss Distribution

So far, in this section, we have defined the privacy loss distribution, given an example, and illustrated that it is a natural quantity to consider that captures essentially everything we need to know about the (in)distinguishability of two distributions. To wrap up this section, we will relate the privacy loss distribution back to the definition of approximate (ε, δ)-DP:

Proposition 3.7 (Conversion from Privacy Loss Distribution to Approximate Differential Privacy)**.** *Let P and Q be two probability distributions on $\mathcal{Y}$ such that the privacy loss distribution* $\mathrm{PrivLoss}(P\|Q)$ *is well-defined. Fix $\varepsilon \geq 0$ and define*

$$\delta := \sup_{S \subset \mathcal{Y}} P(S) - e^{\varepsilon} \cdot Q(S).$$

Then

$$\begin{aligned}
\delta &= \mathop{\mathbb{P}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)}[Z > \varepsilon] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \mathrm{PrivLoss}(Q\|P)}\left[-Z' > \varepsilon\right] \\
&= \mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)}\left[\max\{0, 1 - \exp(\varepsilon - Z)\}\right] \\
&= \int_{\varepsilon}^{\infty} e^{\varepsilon - Z} \cdot \mathop{\mathbb{P}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)}[Z > z] \mathrm{d}z \\
&\leq \mathop{\mathbb{P}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)}[Z > \varepsilon].
\end{aligned}$$

Proof. For any measurable $S \subset \mathcal{Y}$, we have

$$P(S) - e^{\varepsilon} \cdot Q(S) = \int_{\mathcal{Y}} \mathbb{I}[y \in S] \cdot \left(P(y) - e^{\varepsilon} \cdot Q(y)\right) \mathrm{d}y,$$

where $\mathbb{I}$ denotes the indicator function – it takes the value 1 if the condition is true and 0 otherwise. To maximize this expression, we want $y \in S$ whenever $P(y) - e^{\varepsilon} \cdot Q(y) > 0$ and we want $y \notin S$ when this is negative. Thus $\delta = P(S_*) - e^{\varepsilon} \cdot Q(S_*)$ for

$$S_* := \left\{y \in \mathcal{Y} : P(y) - e^{\varepsilon} \cdot Q(y) > 0\right\} = \left\{y \in \mathcal{Y} : f_{P\|Q}(y) > \varepsilon\right\}.$$

Now

$$P(S_*) = \mathop{\mathbb{P}}_{Y \leftarrow P}\left[f_{P\|Q}(Y) > \varepsilon\right] = \mathop{\mathbb{P}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)}[Z > \varepsilon],$$

and, by Remark 3.5,

$$Q(S_*) = \mathop{\mathbb{P}}_{Y \leftarrow Q}\left[f_{P\|Q}(Y) > \varepsilon\right] = \mathop{\mathbb{P}}_{Z' \leftarrow \mathrm{PrivLoss}(Q\|P)}\left[-Z' > \varepsilon\right].$$

This gives the first expression in the result:

$$\delta = P(S_*) - e^{\varepsilon} \cdot Q(S_*) = \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > \varepsilon] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \text{PrivLoss}(Q\|P)}\left[-Z' > \varepsilon\right].$$

Alternatively, $P(S_*) = \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[\mathbb{I}[Z > \varepsilon]]$ and, by Lemma 3.6,

$$Q(S_*) = \mathop{\mathbb{E}}_{Y \leftarrow Q}\left[\mathbb{I}[f_{P\|Q}(Y) > \varepsilon]\right] = \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}\left[\mathbb{I}[Z > \varepsilon] \cdot e^{-Z}\right],$$

which yields

$$\delta = P(S_*) - e^{\varepsilon} \cdot Q(S_*) = \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}\left[(1 - e^{\varepsilon} \cdot e^{-Z}) \cdot \mathbb{I}[Z > \varepsilon]\right].$$

Note that $(1 - e^{\varepsilon} \cdot e^{-z}) \cdot \mathbb{I}[z > \varepsilon] = \max\{0, 1 - e^{\varepsilon - z}\}$ for all $z \in \mathbb{R}$. This produces the second expression in our result.

To obtain the third expression in the result, we apply integration by parts to the second expression: Let $F(z) := \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > z]$ be the complement of the cumulative distribution function of the privacy loss distribution. Then the probability density function of Z evaluated at z is given by the negative derivative, $-F'(z)$.[v] Then

$$\begin{aligned}
\delta &= \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}\left[\max\{0, 1 - e^{\varepsilon - Z}\}\right] \\
&= \int_{\mathbb{R}} \max\{0, 1 - e^{\varepsilon - z}\} \cdot (-F'(z))\mathrm{d}z \\
&= \int_{\varepsilon}^{\infty} (1 - e^{\varepsilon - z}) \cdot (-F'(z))\mathrm{d}z \\
&= \int_{\varepsilon}^{\infty} \left(\frac{\mathrm{d}}{\mathrm{d}z}(1 - e^{\varepsilon - z}) \cdot (-F(z))\right) - (0 - e^{\varepsilon - z} \cdot (-1)) \cdot (-F(z))\mathrm{d}z \\
&\qquad\qquad \text{(product rule)} \\
&= \lim_{z \to \infty} (1 - e^{\varepsilon - z}) \cdot (-F(z)) - (1 - e^{\varepsilon - \varepsilon}) \cdot (-F(\varepsilon)) - \int_{\varepsilon}^{\infty} e^{\varepsilon - z} \cdot (-F(z))\mathrm{d}z \\
&\qquad\qquad \text{(fundamental theorem of calculus)} \\
&= -\lim_{z \to \infty} \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > z] + \int_{\varepsilon}^{\infty} e^{\varepsilon - z} \cdot \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > z]\mathrm{d}z.
\end{aligned}$$

If the privacy loss is well-defined, then $\lim_{z \to \infty} \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > z] = 0$.

v. In general, the privacy loss may not be continuous – i.e., F may not be differentiable. Nevertheless, the final result still holds in this case.

The final expression (an upper bound, rather than a tight characterization) is easily obtained from any of the other three expressions. In particular, dropping the second term $-e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \text{PrivLoss}(Q\|P)}\left[-Z' > \varepsilon\right] \leq 0$ from the first expression yields the upper bound. □

The expression $\delta = \sup_{S \subset \mathcal{Y}} P(S) - e^{\varepsilon} \cdot Q(S)$ in Proposition 3.7 is known as the "hockey stick divergence" and it determines the smallest δ for a given ε such that $P(S) \leq e^{\varepsilon} Q(S) + \delta$ for all $S \subset \mathcal{Y}$. If $P = M(x)$ and $Q = M(x')$ for arbitrary neighboring datasets x, x', then this expression gives the best approximate (ε, δ)-DP guarantee.

Proposition 3.7 gives us three equivalent ways to calculate δ, each of which will be useful in different circumstances. To illustrate how to use Proposition 3.7, we combine it with Proposition 3.3 to prove a tight approximate differential privacy guarantee for Gaussian noise addition:

Corollary 3.8 (Tight Approximate Differential Privacy for Univariate Gaussian)**.** *Let $q : \mathcal{X}^n \to \mathbb{R}$ be a deterministic function and let $\Delta := \sup_{\substack{x,x' \in \mathcal{X}^n \\ \text{neighboring}}} |q(x) - q(x')|$ be its sensitivity. Define a randomized algorithm $M : \mathcal{X}^n \to \mathbb{R}$ by $M(x) = \mathcal{N}(q(x), \sigma^2)$ for some $\sigma^2 > 0$. Then, for any $\varepsilon \geq 0$, M satisfies (ε, δ)-DP with*

$$\delta = \overline{\Phi}\left(\frac{\varepsilon - \rho_*}{\sqrt{2\rho_*}}\right) - e^{\varepsilon} \cdot \overline{\Phi}\left(\frac{\varepsilon + \rho_*}{\sqrt{2\rho_*}}\right),$$

where $\rho_* := \Delta^2/2\sigma^2$ and $\overline{\Phi}(z) := \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}[G > z] = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} \exp(-t^2/2) \mathrm{d}t$.

Furthermore, this guarantee is optimal – for every $\varepsilon \geq 0$, there is no $\delta' < \delta$ such that M is (ε, δ')-DP for general q.

Proof. Fix arbitrary neighboring datasets $x, x' \in \mathcal{X}^n$ and $S \subset \mathcal{Y}$. Let $\mu = q(x)$ and $\mu' = q(x')$. Let $P = M(x) = \mathcal{N}(\mu, \sigma^2)$ and $Q = M(x') = \mathcal{N}(\mu', \sigma^2)$. We must show $P(S) \leq e^{\varepsilon} \cdot Q(S) + \delta$ for arbitrary $\varepsilon \geq 0$ and the value δ given in the result.

By Proposition 3.3, $\text{PrivLoss}(P\|Q) = \text{PrivLoss}(Q\|P) = \mathcal{N}(\rho, 2\rho)$, where $\rho = \frac{(\mu - \mu')^2}{2\sigma^2} \leq \rho_* = \frac{\Delta^2}{2\sigma^2}$.

By Proposition 3.7, we have $P(S) \leq e^{\varepsilon} \cdot Q(S) + \delta$, where

$$\begin{aligned}
\delta &= \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > \varepsilon] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \text{PrivLoss}(Q\|P)}\left[-Z' > \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{Z \leftarrow \mathcal{N}(\rho, 2\rho)}[Z > \varepsilon] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \mathcal{N}(\rho, 2\rho)}\left[-Z' > \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[\rho + \sqrt{2\rho} \cdot G > \varepsilon\right] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[-\rho + \sqrt{2\rho} \cdot G > \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[G > \frac{\varepsilon - \rho}{\sqrt{2\rho}}\right] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[G > \frac{\varepsilon + \rho}{\sqrt{2\rho}}\right]
\end{aligned}$$

$$= \overline{\Phi}\left(\frac{\varepsilon - \rho}{\sqrt{2\rho}}\right) - e^{\varepsilon} \cdot \overline{\Phi}\left(\frac{\varepsilon + \rho}{\sqrt{2\rho}}\right).$$

Since $\rho \leq \rho_*$ and the above expression is increasing in ρ, we can substitute in ρ_* as an upper bound.

Optimality follows from the fact that both Propositions 3.3 and 3.7 give exact characterizations. Note that we must assume that there exist neighboring x, x' such that $\rho = \rho_*$. □

The guarantee of Corollary 3.8 is exact, but it is somewhat hard to interpret. We can easily obtain a more interpretable upper bound:

$$\begin{aligned}
\delta &= \overline{\Phi}\left(\frac{\varepsilon - \rho_*}{\sqrt{2\rho_*}}\right) - e^{\varepsilon} \cdot \overline{\Phi}\left(\frac{\varepsilon + \rho_*}{\sqrt{2\rho_*}}\right) \\
&\leq \overline{\Phi}\left(\frac{\varepsilon - \rho_*}{\sqrt{2\rho_*}}\right) = \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[G > \frac{\varepsilon - \rho_*}{\sqrt{2\rho_*}}\right] \\
&\leq \frac{\exp\left(-\frac{(\varepsilon-\rho_*)^2}{4\rho_*}\right)}{\max\left\{2, \sqrt{\frac{\pi}{\rho_*}} \cdot (\varepsilon - \rho_*)\right\}}. \qquad \text{(assuming } \varepsilon \geq \rho_*\text{)}
\end{aligned}$$

3.4 Composition via the Privacy Loss Distribution

The privacy loss distribution captures essentially everything about the (in)distinguishability of a pair of distributions. It is also the key to understanding composition. Suppose we run multiple differentially private algorithms on the same dataset and each has a well-defined privacy loss distribution. The composition of these algorithms corresponds to the convolution of the privacy loss distributions. That is, the privacy loss random variable corresponding to running all of the algorithms independently is equal to the sum of the independent privacy loss random variables of each of the algorithms:

Theorem 3.9 (Composition is Convolution of Privacy Loss Distributions)**.** *For each $j \in [k]$, let P_j and Q_j be distributions on $\mathcal{Y}_j$ and assume* $\text{PrivLoss}\left(P_j \middle\| Q_j\right)$ *is well defined. Let $P = P_1 \times P_2 \times \cdots \times P_k$ denote the product distribution on $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_k$ obtained by sampling independently from each P_j. Similarly, let $Q = Q_1 \times Q_2 \times \cdots \times Q_k$ denote the product distribution on $\mathcal{Y}$ obtained by sampling independently from each Q_j. Then* $\text{PrivLoss}(P \| Q)$ *is the convolution of the distributions* $\text{PrivLoss}\left(P_j \middle\| Q_j\right)$ *for all $j \in [k]$. That is, sampling $Z \leftarrow \text{PrivLoss}(P \| Q)$ is equivalent to $Z = \sum_{j=1}^{k} Z_j$ when $Z_j \leftarrow \text{PrivLoss}\left(P_j \middle\| Q_j\right)$ independently for each $j \in [k]$.*

Proof. For all $y \in \mathcal{Y}$, the log likelihood ratio (Definition 3.2) satisfies

$$\begin{aligned} f_{P\|Q}(y) &= \log\left(\frac{P(y)}{Q(y)}\right) \\ &= \log\left(\frac{P_1(y_1)\cdot P_2(y_2)\cdot\cdots\cdot P_k(y_k)}{Q_1(y_1)\cdot Q_2(y_2)\cdot\cdots\cdot Q_k(y_k)}\right) \\ &= \log\left(\frac{P_1(y_1)}{Q_1(y_1)}\right) + \log\left(\frac{P_2(y_2)}{Q_2(y_2)}\right) + \cdots + \log\left(\frac{P_k(y_k)}{Q_k(y_k)}\right) \\ &= f_{P_1\|Q_1}(y_1) + f_{P_2\|Q_2}(y_2) + \cdots + f_{P_k\|Q_k}(y_k). \end{aligned}$$

Since P is a product distribution, sampling $Y \leftarrow P$ is equivalent to sampling $Y_1 \leftarrow P_1, Y_2 \leftarrow P_2, \cdots, Y_k \leftarrow P_k$ independently.

A sample from the privacy loss distribution $Z \leftarrow \text{PrivLoss}(P\|Q)$ is given by $Z = f_{P\|Q}(Y)$ for $Y \leftarrow P$. By the above two facts, this is equivalent to $Z = f_{P_1\|Q_1}(Y_1) + f_{P_2\|Q_2}(Y_2) + \cdots + f_{P_k\|Q_k}(Y_k)$ for $Y_1 \leftarrow P_1, Y_2 \leftarrow P_2, \cdots, Y_k \leftarrow P_k$ independently. For each $j \in [k]$, sampling $Z_j \leftarrow \text{PrivLoss}\left(P_j \middle\| Q_j\right)$ is given by $Z_j = f_{P_j\|Q_j}(Y_j)$ for $Y_j \leftarrow P_j$. Thus sampling $Z \leftarrow \text{PrivLoss}(P\|Q)$ is equivalent to $Z = Z_1 + Z_2 + \cdots + Z_k$ where $Z_1 \leftarrow \text{PrivLoss}(P_1\|Q_1), Z_2 \leftarrow \text{PrivLoss}(P_2\|Q_2), \cdots, Z_k \leftarrow \text{PrivLoss}(P_k\|Q_k)$ are independent. □

Theorem 3.9 is the key to understanding composition of differential privacy. More concretely, we should think of a pair of neighboring inputs x, x' and k algorithms $M_1, \cdots, M_k$. Suppose M is the composition of $M_1, \cdots, M_k$. Then the the differential privacy of M can be expressed in terms of the privacy loss distribution $\text{PrivLoss}\left(M(x)\middle\|M(x')\right)$. Theorem 3.9 allows us to decompose this privacy loss as the sum/convolution of the privacy losses of the constituent algorithms $\text{PrivLoss}\left(M_j(x)\middle\|M_j(x')\right)$ for $j \in [k]$. Thus if we have differential privacy guarantees for each M_j, this allows us to prove differential privacy guarantees for M.

Basic Composition, Revisited

We can revisit basic composition (Theorem 3.1 in Section 3.2) with the perspective of privacy loss distributions. Suppose $M_1, M_2, \cdots, M_k : \mathcal{X}^n \to \mathcal{Y}$ are each ε-DP. Fix neighboring datasets $x, x' \in \mathcal{X}^n$. This means that $\mathbb{P}_{Z_j \leftarrow \text{PrivLoss}\left(M_j(x)\middle\|M_j(x')\right)}\left[Z_j \le \varepsilon\right] = 1$ for each $j \in [k]$. Now let $M : \mathcal{X}^n \to \mathcal{Y}^k$ be the composition of these algorithms. We can express the privacy loss $Z \leftarrow \text{PrivLoss}\left(M(x)\middle\|M(x')\right)$ as $Z = Z_1 + Z_2 + \cdots + Z_k$ where $Z_j \leftarrow \text{PrivLoss}\left(M_j(x)\middle\|M_j(x')\right)$ for each $j \in [k]$. Basic composition simply adds up the upper bounds:

$$Z = Z_1 + Z_2 + \cdots + Z_k \le \varepsilon + \varepsilon + \cdots + \varepsilon = k\varepsilon.$$

This bound is tight if each Z_j is a point mass (i.e., $\mathbb{P}\left[Z_j = \varepsilon\right] = 1$). However, this is not the case. (It is possible to prove, in general, that $\mathbb{P}\left[Z_j = \varepsilon\right] \leq \frac{1}{1+e^{-\varepsilon}}$.) The way we will prove better composition bounds is by applying concentration of measure bounds to this sum of independent random variables. That way we can prove that the privacy loss is small with high probability, which yields a better differential privacy guarantee.

Intuitively, we will apply the central limit theorem. The privacy loss random variable of the composed algorithm M can be expressed as the sum of independent bounded random variables. That means the privacy loss distribution $\text{PrivLoss}\left(M(x)\|M(x')\right)$ is well-approximated by a Gaussian, which is the information we need to prove a composition theorem. What is left to do is to obtain bounds on the mean and variance of the summands and make this Gaussian approximation precise.

Gaussian Composition

It is instructive to look at composition when each constituent algorithm M_j is the Gaussian noise addition mechanism. In this case the privacy loss distribution is exactly Gaussian and convolutions of Gaussians are also Gaussian. This is the ideal case and our general composition theorem will be an approximation to this ideal.

Specifically, we can prove a multivariate analog of Corollary 3.8:

Corollary 3.10 (Tight Approximate Differential Privacy for Multivariate Gaussian)**.** *Let $q : \mathcal{X}^n \to \mathbb{R}^d$ be a deterministic function and let $\Delta := \sup_{\substack{x,x' \in \mathcal{X}^n \\ neighboring}} \|q(x) - q(x')\|_2$ be its sensitivity in the 2-norm. Define a randomized algorithm $M : \mathcal{X}^n \to \mathbb{R}^d$ by $M(x) = \mathcal{N}(q(x), \sigma^2 I)$ for some $\sigma^2 > 0$, where I is the identity matrix. Then, for any $\varepsilon \geq 0$, M satisfies (ε, δ)-DP with*

$$\delta = \overline{\Phi}\left(\frac{\varepsilon - \rho_*}{\sqrt{2\rho_*}}\right) - e^{\varepsilon} \cdot \overline{\Phi}\left(\frac{\varepsilon + \rho_*}{\sqrt{2\rho_*}}\right),$$

where $\rho_* := \Delta^2/2\sigma^2$ and $\overline{\Phi}(z) := \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}[G > z] = \frac{1}{\sqrt{2\pi}} \int_z^\infty \exp(-t^2/2)\mathrm{d}t$.

Furthermore, this guarantee is optimal – for every $\varepsilon \geq 0$, there is no $\delta' < \delta$ such that M is (ε, δ')-DP for general q.

Proof. Fix arbitrary neighboring datasets $x, x' \in \mathcal{X}^n$ and $S \subset \mathcal{Y}$. Let $\mu = q(x), \mu' = q(x') \in \mathbb{R}^d$. Let $P = M(x) = \mathcal{N}(\mu, \sigma^2 I)$ and $Q = M(x') = \mathcal{N}(\mu', \sigma^2 I)$. We must show $P(S) \leq e^{\varepsilon} \cdot Q(S) + \delta$ for arbitrary $\varepsilon \geq 0$ and the value δ given in the result.

Now both P and Q are product distributions: For $j \in [d]$, let $P_j = \mathcal{N}(\mu_j, \sigma^2)$ and $Q_j = \mathcal{N}(\mu'_j, \sigma^2)$. Then $P = P_1 \times P_2 \times \cdots P_d$ and $Q = Q_1 \times Q_2 \times \cdots \times Q_d$.

By Theorem 3.9, $\text{PrivLoss}(P\|Q) = \sum_{j=1}^{d} \text{PrivLoss}\left(P_j \middle\| Q_j\right)$ and $\text{PrivLoss}(Q\|P) = \sum_{j=1}^{d} \text{PrivLoss}\left(Q_j \middle\| P_j\right)$.

By Proposition 3.3, $\text{PrivLoss}\left(P_j \middle\| Q_j\right) = \text{PrivLoss}\left(Q_j \middle\| P_j\right) = \mathcal{N}(\rho_j, 2\rho_j)$, where $\rho_j = \frac{(\mu_j - \mu'_j)^2}{2\sigma^2}$ for all $j \in [d]$.

Thus $\text{PrivLoss}(P\|Q) = \text{PrivLoss}(Q\|P) = \sum_{j=1}^{d} \mathcal{N}(\rho_j, 2\rho_j) = \mathcal{N}(\rho, 2\rho)$, where $\rho = \sum_{j=1}^{d} \rho_j = \frac{\|\mu - \mu'\|_2^2}{2\sigma^2} \leq \rho_* = \frac{\Delta^2}{2\sigma^2}$.

By Proposition 3.7, we have $P(S) \leq e^{\varepsilon} \cdot Q(S) + \delta$, where

$$
\begin{aligned}
\delta &= \mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z > \varepsilon] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \text{PrivLoss}(Q\|P)}\left[-Z' > \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{Z \leftarrow \mathcal{N}(\rho, 2\rho)}[Z > \varepsilon] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{Z' \leftarrow \mathcal{N}(\rho, 2\rho)}\left[-Z' > \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[\rho + \sqrt{2\rho} \cdot G > \varepsilon\right] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[-\rho + \sqrt{2\rho} \cdot G > \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[G > \frac{\varepsilon - \rho}{\sqrt{2\rho}}\right] - e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{G \leftarrow \mathcal{N}(0,1)}\left[G > \frac{\varepsilon + \rho}{\sqrt{2\rho}}\right] \\
&= \overline{\Phi}\left(\frac{\varepsilon - \rho}{\sqrt{2\rho}}\right) - e^{\varepsilon} \cdot \overline{\Phi}\left(\frac{\varepsilon + \rho}{\sqrt{2\rho}}\right).
\end{aligned}
$$

Since $\rho \leq \rho_*$ and the above expression is increasing in ρ, we can substitute in ρ_* as an upper bound.

Optimality follows from the fact that Propositions 3.3 and 3.7 and Theorem 3.9 give exact characterizations. Note that we must assume that there exist neighboring x, x' such that $\rho = \rho_*$. □

The key to the analysis of Gaussian composition in the proof of Corollary 3.10 is that sums of Gaussians are Gaussian. In general, the privacy loss of each component is not Gaussian, but the sum still behaves much like a Gaussian and this observation is the basis for improving the composition analysis.

Composition Via Gaussian Approximation

After analyzing Gaussian composition, our next step is to analyze the composition of k independent ε-DP algorithms. We will use the same tools as we did for Gaussian composition and we will develop a new tool, which is called concentrated differential privacy.

Let $M_1, \cdots, M_k : \mathcal{X}^n \to \mathcal{Y}$ each be ε-DP and let $M : \mathcal{X}^n \to \mathcal{Y}^k$ be the composition of these algorithms. Let $x, x' \in \mathcal{X}^n$ be neighboring datasets. For notational convenience, let $P_j = M_j(x)$ and $Q_j = M_j(x')$ for all $j \in [k]$ and let $P = M(x) = P_1 \times P_2 \times \cdots \times P_k$ and $Q = M(x') = Q_1 \times Q_2 \times \cdots \times Q_k$.

For each $j \in [k]$, the algorithm M_j satisfies ε-DP, which ensures that the privacy loss random variable $Z_j \leftarrow \text{PrivLoss}\left(P_j \middle\| Q_j\right) = \text{PrivLoss}\left(M_j(x) \middle\| M_j(x')\right)$ is supported on the interval $[-\varepsilon, \varepsilon]$. The privacy loss being bounded immediately implies a bound on the variance: $\mathbb{V}\left[Z_j\right] \leq \mathbb{E}\left[Z_j^2\right] \leq \varepsilon^2$. We also can prove a bound on the expectation: $\mathbb{E}\left[Z_j\right] \leq \frac{1}{2}\varepsilon^2$. We will prove this bound formally later (in Proposition 3.16). For now, we give some intuition: Clearly $\mathbb{E}\left[Z_j\right] \leq \varepsilon$ and the only way this can be tight is if $Z_j = \varepsilon$ with probability 1. But $Z_j = \log(P_j(Y_j)/Q_j(Y_j))$ for $Y_j \leftarrow P_j$. Thus $\mathbb{E}\left[Z_j\right] = \varepsilon$ implies $P_j(Y_j) = e^{\varepsilon} \cdot Q_j(Y_j)$ with probability 1. This yields a contradiction: $1 = \sum_y P_j(y) = \sum_y e^{\varepsilon} \cdot Q_j(y) = e^{\varepsilon} \cdot 1$. Thus we conclude $\mathbb{E}\left[Z_j\right] < \varepsilon$ and, with a bit more work, we can obtain the bound $\mathbb{E}\left[Z_j\right] \leq \frac{1}{2}\varepsilon^2$ from the fact that $|Z_j| \leq \varepsilon$ and $\sum_y P_j(y) = \sum_y Q_j(y) = 1$.

Our goal is to understand the privacy loss $Z \leftarrow \text{PrivLoss}\left(P \| Q\right) = \text{PrivLoss}\left(M(x) \middle\| M(x')\right)$ of the composed algorithm. Theorem 3.9 tells us that this is the convolution of the constituent privacy losses. That is, we can write $Z = \sum_{j=1}^{k} Z_j$ where $Z_j \leftarrow \text{PrivLoss}\left(P_j \middle\| Q_j\right) = \text{PrivLoss}\left(M_j(x) \middle\| M_j(x')\right)$ independently for each $j \in [k]$.

By independence, we have

$$\mathbb{E}\left[Z\right] = \sum_{j=1}^{k} \mathbb{E}\left[Z_j\right] \leq \frac{1}{2}\varepsilon^2 \cdot k \quad \text{and} \quad \mathbb{V}\left[Z\right] = \sum_{j=1}^{k} \mathbb{V}\left[Z_j\right] \leq \varepsilon^2 \cdot k.$$

Since Z can be written as the sum of independent bounded random variables, the central limit theorem tells us that it is well approximated by a Gaussian – i.e.,

$$\text{PrivLoss}\left(P \| Q\right) = \text{PrivLoss}\left(M(x) \middle\| M(x')\right) \approx \mathcal{N}(\mathbb{E}\left[Z\right], \mathbb{V}\left[Z\right]).$$

Are we done? Can we substitute this approximation into Proposition 3.7 to complete the proof of a better composition theorem? We must make this approximation precise. Unfortunately, the approximation guarantee of the quantitative central limit theorem (a.k.a., the Berry-Esseen Theorem) is not quite strong enough. To be precise, converting the guarantee to approximate (ε, δ)-DP would incur an error of $\delta \geq \Omega(1/\sqrt{k})$, which is larger than we want.

Our approach is to look at the moment generating function – i.e., the expectation of an exponential function – of the privacy loss distribution. To be precise, we will show that, for all $t \geq 0$,

$$\mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P \| Q)}\left[\exp(tZ)\right] = \prod_{j=1}^{k} \mathop{\mathbb{E}}_{Z_j \leftarrow \text{PrivLoss}\left(P_j \| Q_j\right)}\left[\exp(tZ_j)\right]$$

$$\leq \exp\left(\frac{1}{2}\varepsilon^2 t(t+1)\cdot k\right)$$
$$= \mathop{\mathbb{E}}_{\tilde{Z}\leftarrow\mathcal{N}(\frac{1}{2}\varepsilon^2 k,\varepsilon^2 k)}\left[\exp(t\tilde{Z})\right].$$

In other words, rather than attempting to prove a Gaussian approximation, we prove a one-sided bound. Informally, this says that $\text{PrivLoss}(P\|Q) \leq \mathcal{N}(\frac{1}{2}\varepsilon^2 k, \varepsilon^2 k)$. The expectation of an exponential function turns out to be a nice way to formalize this inequality, because, if X and Y are independent, then $\mathbb{E}\left[\exp(X+Y)\right] = \mathbb{E}\left[\exp(X)\right]\cdot\mathbb{E}\left[\exp(Y)\right]$.

To formalize this approach, we next introduce concentrated differential privacy.

3.4.1 Concentrated Differential Privacy

Concentrated differential privacy [DR16; BS16] is a variant of differential privacy (like pure DP and approximate DP). The main advantage of concentrated DP is that it composes well. Thus we will use it as a tool to prove better composition results.

Definition 3.11 (Concentrated Differential Privacy)**.** *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized algorithm. We say that M satisfies ρ-concentrated differential privacy (ρ-zCDP) if, for all neighboring inputs $x, x' \in \mathcal{X}^n$, the privacy loss distribution $\text{PrivLoss}\left(M(x)\middle\|M(x')\right)$ is well-defined (see Definition 3.2) and*

$$\forall t \geq 0 \qquad \mathop{\mathbb{E}}_{Z\leftarrow\text{PrivLoss}(M(x)\|M(x'))}\left[\exp(tZ)\right] \leq \exp(t(t+1)\cdot\rho).$$

To contextualize this definition, we begin by showing that the Gaussian mechanism satisfies it.

Lemma 3.12 (Gaussian Mechanism is Concentrated DP)**.** *Let $q : \mathcal{X}^n \to \mathbb{R}^d$ have sensitivity Δ – that is, $\|q(x) - q(x')\|_2 \leq \Delta$ for all neighboring $x, x' \in \mathcal{X}^n$. Let $\sigma > 0$. Define a randomized algorithm $M : \mathcal{X}^n \to \mathbb{R}^d$ by $M(x) = \mathcal{N}(q(x), \sigma^2 I_d)$. Then M is ρ-zCDP for $\rho = \frac{\Delta^2}{2\sigma^2}$.*

Proof. Fix neighboring inputs $x, x' \in \mathcal{X}^n$ and $t \geq 0$. By Proposition 3.3, for each $j \in [d]$,
$\text{PrivLoss}\left(M(x)_j\middle\|M(x')_j\right) = \mathcal{N}(\hat{\rho}_j, 2\hat{\rho}_j)$ for $\hat{\rho}_j = \frac{(q(x)_j - q(x')_j)^2}{2\sigma^2}$. By Theorem 3.9, $\text{PrivLoss}\left(M(x)\middle\|M(x')\right) = \sum_{j=1}^d \mathcal{N}(\hat{\rho}_j, 2\hat{\rho}_j) = \mathcal{N}(\hat{\rho}, 2\hat{\rho})$ for $\hat{\rho} = \sum_{j=1}^d \hat{\rho}_j = \frac{\|q(x)-q(x')\|_2^2}{2\sigma^2} \leq \rho$. Thus $\mathop{\mathbb{E}}_{Z\leftarrow\text{PrivLoss}(M(x)\|M(x'))}\left[\exp(tZ)\right] = \exp(t(t+1)\hat{\rho}) \leq \exp(t(t+1)\rho)$, as required. □

To analyze the composition of k independent ε-DP algorithms, we will prove three results: (i) Pure ε-DP implies $\frac{1}{2}\varepsilon^2$-zCDP. (ii) The composition of k independent $\frac{1}{2}\varepsilon^2$-zCDP algorithms satisfies $\frac{1}{2}\varepsilon^2 k$-zCDP. (iii) $\frac{1}{2}\varepsilon^2 k$-zCDP implies approximate (ε', δ)-DP with $\delta \in (0,1)$ arbitrary and $\varepsilon' = \varepsilon \cdot \sqrt{2k\log(1/\delta)} + \frac{1}{2}\varepsilon^2 k$. We begin with composition, as this is the raison d'être for concentrated DP:

Theorem 3.13 (Composition for Concentrated Differential Privacy)**.** *Let $M_1, M_2, \cdots, M_k : \mathcal{X}^n \to \mathcal{Y}$ be randomized algorithms. Suppose M_j is ρ_j-zCDP for each $j \in [k]$. Define $M : \mathcal{X}^n \to \mathcal{Y}^k$ by $M(x) = (M_1(x), M_2(x), \cdots, M_k(x))$, where each algorithm is run independently. Then M is ρ-zCDP for $\rho = \sum_{j=1}^k \rho_j$.*

Proof. Fix neighboring inputs $x, x' \in \mathcal{X}^n$. By our assumption that each algorithm M_j is ρ_j-zCDP,

$$\forall t \geq 0 \qquad \mathop{\mathbb{E}}_{Z_j \leftarrow \mathrm{PrivLoss}\left(M_j(x) \| M_j(x')\right)} \left[\exp(tZ_j)\right] \leq \exp(t(t+1)\cdot\rho_j).$$

By Theorem 3.9, $Z \leftarrow \mathrm{PrivLoss}\left(M(x)\middle\|M(x')\right)$ can be written as $Z = \sum_{j=1}^k Z_j$, where $Z_j \leftarrow \mathrm{PrivLoss}\left(M_j(x)\middle\|M_j(x')\right)$ independently for each $j \in [k]$.

Thus, for any $t \geq 0$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(M(x)\|M(x'))} \left[\exp(tZ)\right] &= \mathop{\mathbb{E}}_{\substack{\forall j \in [k] \quad Z_j \leftarrow \mathrm{PrivLoss}\left(M_j(x)\middle\|M_j(x')\right) \\ \text{independent}}} \left[\exp\left(t\sum_{j=1}^k Z_j\right)\right] \\
&= \prod_{j=1}^k \mathop{\mathbb{E}}_{Z_j \leftarrow \mathrm{PrivLoss}\left(M_j(x)\|M_j(x')\right)} \left[\exp(tZ_j)\right] \\
&\leq \prod_{j=1}^k \exp(t(t+1)\cdot\rho_j) \\
&= \exp\left(t(t+1)\cdot\sum_{j=1}^k \rho_j\right) \\
&= \exp(t(t+1)\cdot\rho).
\end{aligned}$$

Since x and x' were arbitrary, this proves that M satisfies ρ-zCDP, as required. □

Next we show how to convert from concentrated DP to approximate DP, which applies the tools we developed earlier.

Proposition 3.14 (Conversion from Concentrated DP to Approximate DP)**.** *For any $M : \mathcal{X}^n \to \mathcal{Y}$ and any $\varepsilon, t \geq 0$, M satisfies (ε, δ)-DP with*

$$\begin{aligned}\delta &= \sup_{\substack{x,x' \in \mathcal{X}^n \\ \text{neighboring}}} \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(M(x)\|M(x'))} \left[\exp(tZ)\right] \cdot \frac{\exp(-\varepsilon t)}{t+1} \cdot \left(1 - \frac{1}{t+1}\right)^t \\ &\leq \sup_{\substack{x,x' \in \mathcal{X}^n \\ \text{neighboring}}} \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(M(x)\|M(x'))} \left[\exp(t(Z-\varepsilon))\right].\end{aligned}$$

In particular, if M satisfies ρ-zCDP, then M satisfies (ε, δ)-DP for any $\varepsilon \geq \rho$ with

$$\begin{aligned}\delta &= \inf_{t>0} \exp(t(t+1)\rho - \varepsilon t) \cdot \frac{1}{t+1} \cdot \left(1 - \frac{1}{t+1}\right)^t \\ &\leq \exp(-(\varepsilon - \rho)^2/4\rho).\end{aligned}$$

Proof. Fix arbitrary neighboring inputs x, x'. Fix $\varepsilon, t \geq 0$. We must show that for all S we have $\mathbb{P}[M(x) \in S] \leq e^\varepsilon \cdot \mathbb{P}\left[M(x') \in S\right] + \delta$ for the value of δ given in the statement above.

Let $Z \leftarrow \text{PrivLoss}\left(M(x)\middle\|M(x')\right)$. By Proposition 3.7, it suffices to show

$$\mathbb{E}\left[\max\{0, 1 - \exp(\varepsilon - Z)\}\right] \leq \delta,$$

for the value of δ given in the statement above.

Let $c > 0$ be a constant such that, with probability 1,

$$\max\{0, 1 - \exp(\varepsilon - Z)\} \leq c \cdot \exp(tZ).$$

Taking expectations of both sides we have $\mathbb{E}\left[\max\{0, 1 - \exp(\varepsilon - Z)\}\right] \leq c \cdot \mathbb{E}\left[\exp(tZ)\right]$, which is the kind of bound we need. It only remains to identify the appropriate value of c to obtain the desired bound.

We trivially have $0 \leq c \cdot \exp(tZ)$ as long as $c > 0$. Thus we only need to ensure $1 - \exp(\varepsilon - Z) \leq c \cdot \exp(tZ)$. That is, for any value of $t > 0$, we can set

$$\begin{aligned}c &= \sup_{z \in \mathbb{R}} \frac{1 - \exp(\varepsilon - z)}{\exp(tz)} \\ &= \sup_{z \in \mathbb{R}} \exp(-tz) - \exp(\varepsilon - (t+1)z) \\ &= \frac{\exp(-\varepsilon t)}{t+1} \cdot \left(1 - \frac{1}{t+1}\right)^t,\end{aligned}$$

where the final equality follows from using calculus to determine that $z = \varepsilon + \log(1 + 1/t)$ is the optimal value of z. Thus $\mathbb{E}\left[\max\{0, 1 - \exp(\varepsilon - Z)\}\right] \leq \mathbb{E}\left[\exp(tZ)\right] \cdot \frac{\exp(-\varepsilon t)}{t+1} \cdot \left(1 - \frac{1}{t+1}\right)^t$, which proves the first part of the statement.

Now assume M is ρ-zCDP. Thus

$$\forall t \geq 0 \quad \mathbb{E}\left[\exp(tZ)\right] \leq \exp(t(t+1)\cdot\rho),$$

which immediately yields the equality in the second part of the statement.

To obtain the inequality in the second part of the statement, we observe that

$$\max\{0, 1-\exp(\varepsilon - Z)\} \leq \mathbb{I}[Z > \varepsilon] \leq \exp(t(Z-\varepsilon)),$$

whence $c \leq \exp(-\varepsilon t)$. Substituting in this upper bound on c and setting $t = (\varepsilon-\rho)/2\rho$ completes the proof □

Remark 3.15. *Proposition 3.14 shows that ρ-zCDP implies $(\varepsilon, \delta = \exp(-(\varepsilon-\rho)^2/4\rho))$-DP for all $\varepsilon \geq \rho$. Equivalently, ρ-zCDP implies $(\varepsilon = \rho + 2\sqrt{\rho\cdot\log(1/\delta)}, \delta)$-DP for all $\delta > 0$. Also, to obtain a given a target (ε,δ)-DP guarantee, it suffices to have ρ-zCDP with*

$$\frac{\varepsilon^2}{4\log(1/\delta)+4\varepsilon} \leq \rho = \left(\sqrt{\log(1/\delta)+\varepsilon} - \sqrt{\log(1/\delta)}\right)^2 \leq \frac{\varepsilon^2}{4\log(1/\delta)}.$$

This gives a sufficient condition; tighter bounds can be obtained from Proposition 3.14.

The final piece of the puzzle is the conversion from pure DP to concentrated DP.

Proposition 3.16. *Suppose M satisfies ε-DP, then M satisfies $\frac{1}{2}\varepsilon^2$-zCDP.*

Proof. Fix neighboring inputs x, x'. Let $Z \leftarrow \mathrm{PrivLoss}\left(M(x)\middle\|M(x')\right)$. By our ε-DP assumption, Z is supported on the interval $[-\varepsilon, +\varepsilon]$. Our task is to prove that $\mathbb{E}\left[\exp(tZ)\right] \leq \exp(\frac{1}{2}\varepsilon^2 t(t+1))$ for all $t > 0$.

The key additional fact is the following consequence of Lemma 3.6

$$\begin{aligned}\mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(P\|Q)}\left[e^{-Z}\right] &= \mathop{\mathbb{E}}_{Y\leftarrow P}\left[e^{-f_{P\|Q}(Y)}\right] \\ &= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[e^{f_{P\|Q}(Y)}\cdot e^{-f_{P\|Q}(Y)}\right] = \mathop{\mathbb{E}}_{Y\leftarrow Q}[1] = 1.\end{aligned}$$

We can write this out as an integral to make it clear:

$$\begin{aligned}\mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(P\|Q)}\left[\exp(-Z)\right] &= \mathop{\mathbb{E}}_{Y\leftarrow P}\left[\exp(-f_{P\|Q}(Y))\right] \\ &= \mathop{\mathbb{E}}_{Y\leftarrow P}\left[\exp(-\log(P(Y)/Q(Y)))\right] \\ &= \mathop{\mathbb{E}}_{Y\leftarrow P}\left[\frac{Q(Y)}{P(Y)}\right] \\ &= \int_y \frac{Q(y)}{P(y)}P(y)\mathrm{d}y\end{aligned}$$

$$= \int_{\mathcal{Y}} Q(y)\mathrm{d}y$$
$$= 1.$$

The combination of these two facts – $Z \in [-\varepsilon, \varepsilon]$ and $\mathbb{E}\left[\exp(-Z)\right] = 1$ – is all we need to know about Z to prove the result. The technical ingredient is Hoeffding's lemma [Hoe63]:

Lemma 3.17 (Hoeffding's lemma)**.** *Let Z be a random variable supported on the interval $[-\varepsilon, +\varepsilon]$. Then for all $t \in \mathbb{R}$, $\mathbb{E}\left[\exp(tZ)\right] \leq \exp(t\mathbb{E}[Z] + t^2\varepsilon^2/2)$.*

Proof. To simplify things, we can assume without loss of generality that Z is supported on the discrete set $\{-\varepsilon, +\varepsilon\}$. To prove this claim, let $\tilde{Z} \in \{-\varepsilon, +\varepsilon\}$ be a randomized rounding of Z. That is, $\mathop{\mathbb{E}}_{\tilde{Z}}\left[\tilde{Z}\middle|Z = z\right] = z$ for all $z \in [-\varepsilon, +\varepsilon]$. By Jensen's inequality, since $\exp(tz)$ is a convex function of $z \in \mathbb{R}$ for any fixed $t \in \mathbb{R}$, we have

$$\mathop{\mathbb{E}}_{Z}\left[\exp(tZ)\right] = \mathop{\mathbb{E}}_{Z}\left[\exp\left(t\mathop{\mathbb{E}}_{\tilde{Z}}\left[\tilde{Z}\middle|Z\right]\right)\right] \leq \mathop{\mathbb{E}}_{Z}\left[\mathop{\mathbb{E}}_{\tilde{Z}}\left[\exp(t\tilde{Z})\middle|Z\right]\right] = \mathop{\mathbb{E}}_{\tilde{Z}}\left[\exp(t\tilde{Z})\right].$$

Note that $\mathbb{E}\left[\tilde{Z}\right] = \mathbb{E}[Z]$. Thus it suffices to prove $\mathbb{E}\left[\exp(t\tilde{Z})\right] \leq \exp(t\mathbb{E}\left[\tilde{Z}\right] + \frac{1}{2}\varepsilon^2 t^2)$ for all $t \in \mathbb{R}$.

The final step in the proof is some calculus: Let $p := \mathbb{P}\left[\tilde{Z} = \varepsilon\right] = 1 - \mathbb{P}\left[\tilde{Z} = -\varepsilon\right]$. Then $\mathbb{E}[Z] = \mathbb{E}\left[\tilde{Z}\right] = \varepsilon p - \varepsilon(1-p) = \varepsilon(2p-1)$. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(t) := \log \mathbb{E}\left[\exp(t\tilde{Z})\right] = \log(p \cdot e^{t\varepsilon} + (1-p) \cdot e^{-t\varepsilon}) = \log(1 - p + p \cdot e^{2t\varepsilon}) - t\varepsilon.$$

For all $t \in \mathbb{R}$,

$$f'(t) = \frac{2\varepsilon p \cdot e^{2t\varepsilon}}{1 - p + p \cdot e^{2t\varepsilon}} - \varepsilon,$$

and

$$\begin{aligned} f''(t) &= \frac{(2\varepsilon)^2 p \cdot e^{2t\varepsilon} \cdot (1 - p + p \cdot e^{2t\varepsilon}) - (2\varepsilon p \cdot e^{2t\varepsilon})^2}{(1 - p + p \cdot e^{2t\varepsilon})^2} \\ &= (2\varepsilon)^2 \cdot \frac{p \cdot e^{2t\varepsilon}}{1 - p + p \cdot e^{2t\varepsilon}} \cdot \left(1 - \frac{p \cdot e^{2t\varepsilon}}{1 - p + p \cdot e^{2t\varepsilon}}\right) \\ &= (2\varepsilon)^2 \cdot x \cdot (1 - x) \leq (2\varepsilon)^2 \cdot \frac{1}{4} = \varepsilon^2. \end{aligned}$$

The final line sets $x = \frac{p \cdot e^{2t\varepsilon}}{1-p+p \cdot e^{2t\varepsilon}}$ and uses the fact that the function $x \cdot (1-x)$ is maximized at $x = \frac{1}{2}$.

Note that $f(0) = 0$ and $f'(0) = 2\varepsilon p - \varepsilon = \mathbb{E}\left[\tilde{Z}\right] = \mathbb{E}[Z]$. By the fundamental theorem of calculus, for all $t \in \mathbb{R}$,

$$f(t) = f(0) + f'(0) \cdot t + \int_0^t \int_0^s f''(r)\mathrm{d}r\mathrm{d}s \leq 0 + \mathbb{E}[Z] \cdot t + \int_0^t \int_0^s \varepsilon^2 \mathrm{d}r\mathrm{d}s = \mathbb{E}[Z] \cdot t + \frac{1}{2}\varepsilon^2 t^2.$$

This proves the lemma, as $\mathbb{E}\left[\exp(tZ)\right] \leq \mathbb{E}\left[\exp(t\tilde{Z})\right] = \exp(f(t)) \leq \exp(\mathbb{E}[Z] \cdot t + \frac{1}{2}\varepsilon^2 t^2)$. □

If we substitute $t = -1$ into Lemma 3.17, we have

$$1 = \mathbb{E}\left[\exp(-Z)\right] \leq \exp(-\mathbb{E}[Z] + \frac{1}{2}\varepsilon^2),$$

which rearranges to $\mathbb{E}[Z] \leq \frac{1}{2}\varepsilon^2$.

Substituting this bound on the expectation back into Lemma 3.17 yields the result: For all $t > 0$, we have

$$\mathbb{E}\left[\exp(tZ)\right] \leq \exp\left(t \cdot \mathbb{E}[Z] + \frac{1}{2}\varepsilon^2 t^2\right) \leq \exp\left(\frac{1}{2}\varepsilon^2 t(t+1)\right).$$

□

Combining these three results lets us prove what is known as the advanced composition theorem where we start with each individual algorithm satisfying pure DP [DRV10]:

Theorem 3.18 (Advanced Composition Starting with Pure DP)**.** *Let $M_1, M_2, \cdots, M_k : \mathcal{X}^n \to \mathcal{Y}$ be randomized algorithms. Suppose M_j is ε_j-DP for each $j \in [k]$. Define $M : \mathcal{X}^n \to \mathcal{Y}^k$ by $M(x) = (M_1(x), M_2(x), \cdots, M_k(x))$, where each algorithm is run independently. Then M is (ε, δ)-DP for any $\delta > 0$ with*

$$\varepsilon = \frac{1}{2}\sum_{j=1}^{k} \varepsilon_j^2 + \sqrt{2\log(1/\delta)\sum_{j=1}^{k} \varepsilon_j^2}.$$

Proof of Theorem 3.18. By Proposition 3.16, for each $j \in [k]$, M_j satisfies ρ_j-zCDP with $\rho_j = \frac{1}{2}\varepsilon_j^2$. By composition of concentrated DP (Theorem 3.13), M satisfies

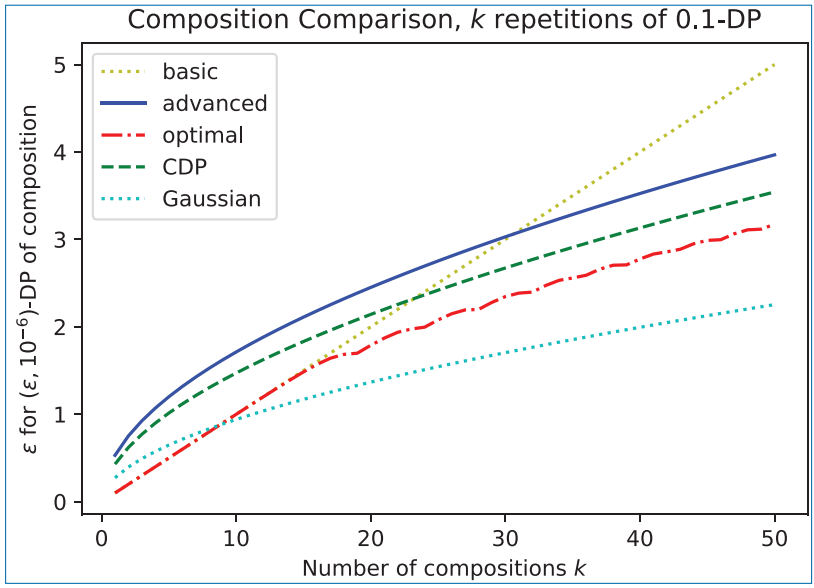

Figure 3.1. Comparison of different composition bounds. We compose k independent 0.1-DP algorithms to obtain a $(\varepsilon, 10^{-6})$-DP guarantee. Theorem 3.1 – *basic* composition – gives $\varepsilon = k \cdot 0.1$. For comparison, we have *advanced* composition (Theorem 3.18), an *optimal* bound [KOV15], and Concentrated DP (*CDP*) with the improved conversion from Proposition 3.14. For comparison, we also consider composing the *Gaussian* mechanism using Corollary 3.10, where the Gaussian noise is scaled to have the same variance as Laplace noise would have to attain 0.1-DP.

ρ-zCDP with $\rho = \sum_{j=1}^{k} \rho_j$. Finally, Proposition 3.14 can convert this concentrated DP guarantee to approximate DP: M satisfies (ε, δ)-DP for all $\varepsilon \geq \rho$ and $\delta = \exp(-(\varepsilon - \rho)^2/4\rho)$. We can rearrange this so that $\delta > 0$ is arbitrary and $\varepsilon = \rho + \sqrt{4\rho \log(1/\delta)}$. □

Recall that the basic composition theorem (Theorem 3.1) gives $\delta = 0$ and $\varepsilon = \sum_{j=1}^{k} \varepsilon_j$. That is, basic composition scales with the 1-norm of the vector $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_k)$, whereas advanced composition scales with the 2-norm of this vector (and the squared 2-norm). Neither bound strictly dominates the other. However, asymptotically (in a sense we will make precise in the next paragraph) advanced composition dominates basic composition.

Suppose we have a fixed (ε, δ)-DP guarantee for the entire system and we must answer k queries of sensitivity 1. Using basic composition, we can answer each query by adding Laplace(k/ε) noise to each answer. However, using advanced composition, we can answer each query by adding Laplace$(\sqrt{k/2\rho})$ noise to each answer,

where

$$\rho \geq \frac{\varepsilon^2}{4\log(1/\delta) + 4\varepsilon},$$

(per Remark 3.15). If the privacy parameters $\varepsilon, \delta > 0$ are fixed (which implies ρ is fixed) and $k \to \infty$, we can see that asymptotically advanced composition gives noise per query scaling as $\Theta(\sqrt{k})$, while basic composition results in noise scaling as $\Theta(k)$.

3.4.2 Adaptive Composition & Post-processing

Thus far we have only considered non-adaptive composition. That is, we assume that the algorithms $M_1, M_2, \cdots, M_k$ being composed are independent. More generally, adaptive composition considers the possibility that M_j can depend on the outputs of $M_1, \cdots, M_{j-1}$. This kind of dependence arises very often, either in an iterative algorithm, or an interactive system where a human chooses analyses to perform sequentially. Fortunately, adaptive composition is easy to deal with.

Proposition 3.19 (Adaptive Composition of Concentrated DP)**.** *Let $M_1 : \mathcal{X}^n \to \mathcal{Y}_1$ be ρ_1-zCDP. Let $M_2 : \mathcal{X}^n \times \mathcal{Y}_1 \to \mathcal{Y}_2$ be such that, for all $y_1 \in \mathcal{Y}_1$, the algorithm $x \mapsto M(x, y_1)$ is ρ_2-zCDP. That is, M_2 is ρ_2-zCDP in terms of its first argument for any fixed value of the second argument. Define $M : \mathcal{X}^n \to \mathcal{Y}_2$ by $M(x) = M_2(x, M_1(x))$. Then M is $(\rho_1 + \rho_2)$-zCDP.*

Proposition 3.19 only considers the composition of two algorithms, but it can be extended to k algorithms by induction.

Proof. Fix neighboring inputs $x, x' \in \mathcal{X}^n$. Fix $t \geq 0$. Let $Z \leftarrow \text{PrivLoss}\left(M(x) \middle\| M(x')\right)$. We must prove $\mathbb{E}\left[\exp(tZ)\right] \leq \exp(t(t+1)(\rho_1 + \rho_2))$.

For non-adaptive composition, we could write $Z = Z_1 + Z_2$ where $Z_1 \leftarrow \text{PrivLoss}\left(M_1(x) \middle\| M_1(x')\right)$ and $Z_2 \leftarrow \text{PrivLoss}\left(M_2(x) \middle\| M_2(x')\right)$ are independent. However, we cannot do this in the adaptive case – the two privacy losses are not independent. Instead, we use the fact that, conditioned on the value of the first privacy loss Z_1, the privacy loss Z_2 still satisfies the bound on the moment generating function. That is, for all z_1, we have $\mathbb{E}\left[\exp(tZ_2) \mid Z_1 = z_1\right] \leq \exp(t(t+1)\rho_2)$. To make this argument precise, we must expand out the relevant definitions.

For now, we make a simplifying technical assumption (which we will justify later): We assume that, given $y_2 = M(x, y_1)$, we can determine y_1. This means we can decompose $f_{M(x)\|M(x')}(y_2) = f_{M_1(x)\|M_1(x')}(y_1) + f_{M_2(x,y_1)\|M_2(x',y_1)}(y_2)$. Thus

$$\mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(M(x)\|M(x'))}\left[\exp(tZ)\right]$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{Y \leftarrow M_2(x, M_1(x))} \left[\exp\left(t \cdot f_{M(x)\|M(x')}(Y)\right)\right] \\
&= \mathop{\mathbb{E}}_{Y_1 \leftarrow M_1(x)} \left[\mathop{\mathbb{E}}_{Y_2 \leftarrow M_2(x, Y_1)} [\exp(t \cdot (f_{M_1(x)\|M_1(x')}(Y_1) \right. \\
&\quad \left. + f_{M_2(x,Y_1)\|M_2(x', Y_1)}(Y_2)))] \right] \\
&= \mathop{\mathbb{E}}_{Y_1 \leftarrow M_1(x)} \left[\exp\left(t \cdot f_{M_1(x)\|M_1(x')}(Y_1)\right) \right. \\
&\quad \left. \cdot \mathop{\mathbb{E}}_{Y_2 \leftarrow M_2(x, Y_1)} \left[\exp\left(t \cdot f_{M_2(x,Y_1)\|M_2(x', Y_1)}(Y_2)\right)\right] \right] \\
&\leq \mathop{\mathbb{E}}_{Y_1 \leftarrow M_1(x)} \left[\exp\left(t \cdot f_{M_1(x)\|M_1(x')}(Y_1)\right)\right] \\
&\quad \cdot \sup_{y_1} \mathop{\mathbb{E}}_{Y_2 \leftarrow M_2(x, y_1)} \left[\exp\left(t \cdot f_{M_2(x,y_1)\|M_2(x', y_1)}(Y_2)\right)\right] \\
&= \mathop{\mathbb{E}}_{Z_1 \leftarrow \mathrm{PrivLoss}(M_1(x)\|M_1(x'))} \left[\exp\left(t \cdot Z_1\right)\right] \\
&\quad \cdot \sup_{y_1} \mathop{\mathbb{E}}_{Z_2 \leftarrow \mathrm{PrivLoss}\left(M_2(x,y_1)\|M_2(x', y_2)\right)} \left[\exp\left(t \cdot Z_2\right)\right] \\
&\leq \exp(t(t+1)\rho_1) \cdot \exp(t(t+1)\rho_2) \\
&= \exp(t(t+1)(\rho_1 + \rho_2)),
\end{aligned}
$$

as required. All that remains is to justify our simplifying technical assumption. We can perforce ensure this assumption holds by defining $\hat{M} : \mathcal{X}^n \to \mathcal{Y}_1 \times \mathcal{Y}_2$ by $\hat{M}(x) = (y_1, y_2)$ where $y_1 = M_1(x)$ and $y_2 = M_2(x, y_1)$ and proving the theorem for $\hat{M}$ in lieu of M. Since the output of $\hat{M}$ includes both outputs, rather than just the last output, the above decomposition works. The result holds in general because M is a *post-processing* of $\hat{M}$. That is, we can obtain $M(x)$ by running $\hat{M}(x)$ and discarding the first part of the output. Intuitively, discarding part of the output cannot hurt privacy. Formally, this is the post-processing property of concentrated DP, which we prove in Lemma 3.20 and Corollary 3.21. □

Lemma 3.20 (Post-processing for Concentrated DP)**.** *Let $\hat{P}$ and $\hat{Q}$ be distributions on $\hat{\mathcal{Y}}$ and let $g : \hat{\mathcal{Y}} \to \mathcal{Y}$ be an arbitrary function. Define $P = g(\hat{P})$ and $Q = g(\hat{Q})$ to be the distributions on $\mathcal{Y}$ obtained by applying g to a function from $\hat{P}$ and $\hat{Q}$ respectively. Then, for all $t \geq 0$,*

$$
\mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)} \left[\exp(tZ)\right] \leq \mathop{\mathbb{E}}_{\hat{Z} \leftarrow \mathrm{PrivLoss}\left(\hat{P}\middle\|\hat{Q}\right)} \left[\exp(t\hat{Z})\right].
$$

Proof. To generate a sample from $Y \leftarrow Q$, we sample $\hat{Y} \leftarrow \hat{Q}$ and set $Y = g(\hat{Y})$. We consider the reverse process: Given $y \in \mathcal{Y}$, define $\hat{Q}_y$ to be the conditional distribution of $\hat{Y} \leftarrow \hat{Q}$ conditioned on $g(\hat{Y}) = y$. That is, $\hat{Q}_y$ is a distribution such that we can generate a sample $\hat{Y} \leftarrow \hat{Q}$ by first sampling $Y \leftarrow Q$ and then sampling $\hat{Y} \leftarrow \hat{Q}_Y$. Note that if g is an injective function, then $\hat{Q}_y$ is a point mass.

We have the following key identity. Formally, this relates the Radon-Nikodym derivative of the postprocessed distributions (P with respect to Q) to the Radon-Nikodym derivative of the original distributions ($\hat{P}$ with respect to $\hat{Q}$) via the conditional distribution $\hat{Q}_y$.

$$\forall y \in \mathcal{Y} \quad \frac{P(y)}{Q(y)} = \mathop{\mathbb{E}}_{\hat{Y} \leftarrow \hat{Q}_y}\left[\frac{\hat{P}(\hat{Y})}{\hat{Q}(\hat{Y})}\right].$$

To see where this identity comes from, write

$$\begin{aligned}
\mathop{\mathbb{E}}_{\hat{Y} \leftarrow \hat{Q}_y}\left[\frac{\hat{P}(\hat{Y})}{\hat{Q}(\hat{Y})}\right] &= \int_{\{\hat{y}:g(\hat{y})=y\}} \frac{\hat{P}(\hat{y})}{\hat{Q}(\hat{y})} \cdot \hat{Q}_y(\hat{y}) \mathrm{d}\hat{y} \\
&= \int_{\{\hat{y}:g(\hat{y})=y\}} \frac{\hat{P}(\hat{y})}{\hat{Q}(\hat{y})} \cdot \frac{\hat{Q}(\hat{y})}{\int_{\{\tilde{y}:g(\tilde{y})=y\}} \hat{Q}(\tilde{y}) \mathrm{d}\tilde{y}} \mathrm{d}\hat{y} \\
&= \frac{\int_{\{\hat{y}:g(\hat{y})=y\}} \hat{P}(\hat{y}) \mathrm{d}\hat{y}}{\int_{\{\tilde{y}:g(\tilde{y})=y\}} \hat{Q}(\tilde{y}) \mathrm{d}\tilde{y}} \\
&= \frac{P(y)}{Q(y)}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(P\|Q)}\left[\exp(tZ)\right] &= \mathop{\mathbb{E}}_{Y \leftarrow P}\left[\exp(t \cdot f_{P\|Q}(Y))\right] \\
&= \mathop{\mathbb{E}}_{Y \leftarrow Q}\left[\exp((t+1) \cdot f_{P\|Q}(Y))\right] \qquad \text{(Lemma 3.6)} \\
&= \mathop{\mathbb{E}}_{Y \leftarrow Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^{t+1}\right] \\
&= \mathop{\mathbb{E}}_{Y \leftarrow Q}\left[\left(\mathop{\mathbb{E}}_{\hat{Y} \leftarrow \hat{Q}_Y}\left[\frac{\hat{P}(\hat{Y})}{\hat{Q}(\hat{Y})}\right]\right)^{t+1}\right] \\
&\le \mathop{\mathbb{E}}_{Y \leftarrow Q}\left[\mathop{\mathbb{E}}_{\hat{Y} \leftarrow \hat{Q}_Y}\left[\left(\frac{\hat{P}(\hat{Y})}{\hat{Q}(\hat{Y})}\right)^{t+1}\right]\right] \qquad \text{(Jensen)}
\end{aligned}$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{\hat{Y}\leftarrow\hat{Q}}\left[\left(\frac{\hat{P}(\hat{Y})}{\hat{Q}(\hat{Y})}\right)^{t+1}\right] \\
&= \mathop{\mathbb{E}}_{\hat{Y}\leftarrow\hat{Q}}\left[\exp((t+1)\cdot f_{\hat{P}\|\hat{Q}}(\hat{Y}))\right] \\
&= \mathop{\mathbb{E}}_{\hat{Y}\leftarrow\hat{P}}\left[\exp(t\cdot f_{\hat{P}\|\hat{Q}}(\hat{Y}))\right] && \text{(Lemma 3.6)} \\
&= \mathop{\mathbb{E}}_{\hat{Z}\leftarrow\mathrm{PrivLoss}\left(\hat{P}\|\hat{Q}\right)}\left[\exp(t\hat{Z})\right],
\end{aligned}
$$

where the inequality follows from Jensen's inequality and the convexity of the function $v \mapsto v^{t+1}$. □

Corollary 3.21. *Let $\hat{M} : \mathcal{X}^n \to \hat{\mathcal{Y}}$ satisfy ρ-zCDP. Let $g : \hat{\mathcal{Y}} \to \mathcal{Y}$ be an arbitrary function. Define $M : \mathcal{X}^n \to \mathcal{Y}$ by $M(x) = g(\hat{M}(x))$. Then M is also ρ-zCDP.*

Proof. Fix neighboring inputs $x, x' \in \mathcal{X}^n$. Let $P = M(x)$, $Q = M(x')$, $\hat{P} = \hat{M}(x)$, and $\hat{Q} = \hat{M}(x')$. By Lemma 3.20 and the assumption that $\hat{M}$ is ρ-zCDP, for all $t \geq 0$,

$$
\begin{aligned}
\mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(M(x)\|M(x'))}\left[\exp(tZ)\right] &= \mathop{\mathbb{E}}_{Z\leftarrow\mathrm{PrivLoss}(P\|Q)}\left[\exp(tZ)\right] \\
&\leq \mathop{\mathbb{E}}_{\hat{Z}\leftarrow\mathrm{PrivLoss}\left(\hat{P}\|\hat{Q}\right)}\left[\exp(t\hat{Z})\right] \\
&= \mathop{\mathbb{E}}_{\hat{Z}\leftarrow\mathrm{PrivLoss}\left(\hat{M}(x)\|\hat{M}(x')\right)}\left[\exp(t\hat{Z})\right] \\
&\leq \exp(t(t+1)\rho),
\end{aligned}
$$

which implies that M is also ρ-zCDP. □

3.4.3 Composition of Approximate (ε, δ)-DP

Thus far we have only considered the composition of pure DP mechanisms (Theorems 3.1 & 3.18) and the Gaussian mechanism (Corollary 3.10). What about approximate (ε, δ)-DP?

We have the following result which extends Theorems 3.1 & 3.18 to approximate DP and to adaptive composition.

Theorem 3.22 (Advanced Composition Starting with Approximate DP)**.** *For $j \in [k]$, let $M_j : \mathcal{X}^n \times \mathcal{Y}_{j-1} \to \mathcal{Y}_j$ be randomized algorithms. Suppose M_j is $(\varepsilon_j, \delta_j)$-DP for each $j \in [k]$. For $j \in [k]$, inductively define $M_{1\cdots j} : \mathcal{X}^n \to \mathcal{Y}_j$ by $M_{1\cdots j}(x) =$*

$M_j(x, M_{1\cdots(j-1)}(x))$, where each algorithm is run independently and $M_{1\cdots 0}(x) = y_0$ for some fixed $y_0 \in \mathcal{Y}_0$. Then $M_{1\cdots k}$ is (ε, δ)-DP for any $\delta > \sum_{j=1}^{k} \delta_j$ with

$$\varepsilon = \min\left\{\sum_{j=1}^{k} \varepsilon_j, \frac{1}{2}\sum_{j=1}^{k} \varepsilon_j^2 + \sqrt{2\log(1/\delta')\sum_{j=1}^{k} \varepsilon_j^2}\right\},$$

where $\delta' = \delta - \sum_{j=1}^{k} \delta_j$.

Intuitively, if you consider the privacy loss $\text{PrivLoss}\left(M(x)\middle\|M(x')\right)$ (where $x, x' \in \mathcal{X}^n$ are arbitrary neighboring inputs), then M being (ε, δ)-DP is equivalent to the privacy loss being in $[-\varepsilon, +\varepsilon]$ with probability at least $1 - \delta$; otherwise the privacy loss can be arbitrary (including possibly infinite). Informally, the proof of Theorem 3.22 uses a union bound to show that with probability at least $1 - \sum_{j=1}^{k} \delta_j$ all of the privacy losses of the k algorithms are bounded by their respective ε_js. Once we condition on this event, the proof proceeds as before.

Formally, rather than reasoning about possibly infinite privacy losses, we use the following decomposition result.

Lemma 3.23. *Let P and Q be probability distributions over $\mathcal{Y}$. Fix $\varepsilon, \delta \geq 0$. Suppose that, for all measurable $S \subset \mathcal{Y}$, we have $P(S) \leq e^{\varepsilon} \cdot Q(S) + \delta$ and $Q(S) \leq e^{\varepsilon} P(S) + \delta$.*

Then there exist distributions P', Q', P'', Q'' over $\mathcal{Y}$ with the following properties. We can express P and Q as convex combinations of these distributions, namely $P = (1-\delta)P' + \delta P''$ and $Q = (1-\delta)Q' + \delta Q''$. And, for every measurable $S \subset \mathcal{Y}$, we have $e^{-\varepsilon} \cdot Q'(S) \leq P'(S) \leq e^{\varepsilon} \cdot Q'(S)$.

Proof. Fix $\varepsilon_1, \varepsilon_2 \in [0, \varepsilon]$ to be determined later. Define distributions P', P'', Q', and Q'' as follows.[vi] For all points $y \in \mathcal{Y}$,

$$\begin{aligned}
P'(y) &= \frac{\min\{P(y), e^{\varepsilon_1} \cdot Q(y)\}}{1-\delta_1}, \\
P''(y) &= \frac{P(y) - (1-\delta_1)P'(y)}{\delta_1} = \frac{\max\{0, P(y) - e^{\varepsilon_1} \cdot Q(y)\}}{\delta_1}, \\
Q'(y) &= \frac{\min\{Q(y), e^{\varepsilon_2} \cdot P(y)\}}{1-\delta_2}, \\
Q''(y) &= \frac{Q(y) - (1-\delta_2)Q'(y)}{\delta_2} = \frac{\max\{0, Q(y) - e^{\varepsilon_2} \cdot P(y)\}}{\delta_2},
\end{aligned}$$

vi. Formally, $P(y)$, $P'(y)$, $P''(y)$, $Q(y)$, $Q'(y)$, and $Q''(y)$ denote the Radon-Nikodym derivative of these distributions with respect to some base measure – usually either the counting measure (in which case these quantities are probability mass functions) or Lebesgue measure (in which case these quantities are probability density functions) – in any case, we can take $P + Q$ to be the base measure.

where δ_1 and δ_2 are appropriate normalizing constants.

By construction, $(1-\delta_1)P' + \delta_1 P'' = P$ and $(1-\delta_2)Q' + \delta_2 Q'' = Q$.

If $\delta_1 = \delta_2 = \delta$, then we have the appropriate decomposition and, for all $y \in \mathcal{Y}$, we have

$$e^{-\varepsilon} \leq e^{-\varepsilon_2} \leq \frac{P'(y)}{Q'(y)} = \frac{\min\{P(y), e^{\varepsilon_1} \cdot Q(y)\}}{\min\{Q(y), e^{\varepsilon_2} \cdot P(y)\}} \leq e^{\varepsilon_1} \leq e^{\varepsilon},$$

as required. If $\delta_1 = \delta_2 < \delta$, we can change the decomposition to

$$P = (1-\delta)P' + (\delta - \delta_1)P' + (1-\delta_1)P'' = (1-\delta)P' + \delta \cdot \left(\frac{\delta - \delta_1}{\delta} P' + \frac{\delta_1}{\delta} P''\right),$$

and likewise for Q, which also yields the result.

It only remains to show that we can ensure that $\delta_1 = \delta_2 \leq \delta$ by appropriately setting $\varepsilon_1, \varepsilon_2 \in [0, \varepsilon]$. We have

$$\delta_1 = \int_{\mathcal{Y}} \max\{0, P(y) - e^{\varepsilon_1} \cdot Q(y)\} \mathrm{d}y = \int_S P(y) - e^{\varepsilon_1} \cdot Q(y) \mathrm{d}y = P(S) - e^{\varepsilon_1} Q(S),$$

where $S = \{y \in \mathcal{Y} : P(y) \geq e^{\varepsilon_1} \cdot Q(y)\}$. If $\varepsilon_1 = \varepsilon$, then $\delta_1 \leq \delta$ by the assumptions of the Lemma. If $\varepsilon_1 = 0$, then $\delta_1 = \mathrm{d_{TV}}(P, Q)$. By decreasing ε_1, we continuously increase δ_1. Thus we can pick $\varepsilon_1 \in [0, \varepsilon]$ such that $\delta_1 = \min\{\delta, \mathrm{d_{TV}}(P, Q)\}$. Similarly, we can pick $\varepsilon_2 \in [0, \varepsilon]$, such that $\delta_2 = \min\{\delta, \mathrm{d_{TV}}(P, Q)\}$. □

The proof of Theorem 3.22 is, unfortunately, quite technical. Most of the steps are the same as we have seen in the pure DP case. The only novelty is applying the decomposition of Lemma 3.23 inductively; this requires cumbersome notation, but is otherwise straightforward.

Proof of Theorem 3.22. Fix neighboring datasets $x, x' \in \mathcal{X}^n$. We inductively define distributions P_j and Q_j on $\mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_j$ as follows. For $j \in [k]$, $P_j = (Y_0, Y_1, \cdots, Y_{j-1}, M_j(x, Y_{j-1}))$, where $(Y_1, \cdots, Y_{j-1}) \leftarrow P_{j-1}$, and $Q_j = (Y_0, Y_1, \cdots, Y_{j-1}, M_j(x', Y_{j-1}))$, where $(Y_1, \cdots, Y_{j-1}) \leftarrow Q_{j-1}$. We define $P_0 = Q_0$ to be the point mass on y_0.

We will prove by induction that, for each $j \in [k]$, there exist distributions P'_j, P''_j, Q'_j, and Q''_j on $\mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_j$ such that

$$P_j = \prod_{\ell=1}^{j} (1-\delta_\ell) P'_j + \left(1 - \prod_{\ell=1}^{j} (1-\delta_\ell)\right) P''_j$$

and

$$Q_j = \prod_{\ell=1}^{j}(1-\delta_\ell)Q'_j + \left(1 - \prod_{\ell=1}^{j}(1-\delta_\ell)\right) Q''_j$$

and, for all $t \geq 0$,

$$\mathop{\mathbb{E}}_{Z'_j \leftarrow \mathrm{PrivLoss}\left(P'_j \middle\| Q'_j\right)} \left[\exp(tZ'_j)\right] \leq \exp\left(\frac{t(t+1)}{2}\sum_{\ell=1}^{j}\varepsilon_\ell^2\right)$$

and, for all measurable $S \subset \mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_j$, $P'_j(S) \leq \exp\left(\sum_{\ell=1}^{j}\varepsilon_\ell\right) \cdot Q'_j(S)$.

Before proving the inductive claim, we show that it suffices to prove the result. Fix an arbitrary measurable $S \subset \mathcal{Y}_k$ and let $\tilde{S} = \mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{k-1} \times S$. We have

$$\begin{aligned}
\mathbb{P}\left[M(x) \in S\right] &= P_k(\tilde{S}) && \text{(Postprocessing)}\\
&= \prod_{\ell=1}^{k}(1-\delta_\ell)P'_k(\tilde{S}) + \left(1 - \prod_{\ell=1}^{k}(1-\delta_\ell)\right)P''_k(\tilde{S})\\
&\leq \prod_{\ell=1}^{k}(1-\delta_\ell)P'_k(\tilde{S}) + \sum_{j=1}^{k}\delta_j\\
&\qquad (P''_k(\tilde{S}) \leq 1 \text{ and } 1 - \textstyle\prod_{\ell=1}^{k}(1-\delta_\ell) \leq \sum_{j=1}^{k}\delta_j)\\
&\leq \prod_{\ell=1}^{k}(1-\delta_\ell)\left(e^{\varepsilon}\cdot Q'_k(\tilde{S}) + \delta'\right) + \sum_{j=1}^{k}\delta_j && (*)\\
&\leq e^{\varepsilon}\cdot\prod_{\ell=1}^{k}(1-\delta_\ell)\cdot Q'_k(\tilde{S}) + \delta && (\delta = \delta' + \textstyle\sum_{j=1}^{k}\delta_j)\\
&\leq e^{\varepsilon}\cdot Q_k(\tilde{S}) + \delta\\
&\qquad (Q_k = \textstyle\prod_{\ell=1}^{k}(1-\delta_\ell)Q'_k + \left(1 - \prod_{\ell=1}^{k}(1-\delta_\ell)\right)Q''_k)\\
&= e^{\varepsilon}\cdot\mathbb{P}\left[M(x') \in S\right] + \delta.
\end{aligned}$$

The inequality $P'_k(\tilde{S}) \leq e^{\varepsilon}\cdot Q'_k(\tilde{S}) + \delta'$ (*) follows the proof we have seen before. Our inductive conclusion includes a pure DP result – $P'_j(\tilde{S}) \leq \exp\left(\sum_{\ell=1}^{j}\varepsilon_\ell\right)\cdot Q'_j(\tilde{S})$ – and a concentrated DP result – for all $t \geq 0$, we have

$\mathop{\mathbb{E}}_{Z'_j \leftarrow \mathrm{PrivLoss}\left(P'_j \middle\| Q'_j\right)} \left[\exp(tZ'_j)\right] \le \exp\left(\frac{t(t+1)}{2} \sum_{\ell=1}^{j} \varepsilon_\ell^2\right)$, which implies

$$
\begin{aligned}
P'_k(\tilde{S}) &\le e^{\varepsilon} \cdot Q'_k(\tilde{S}) + \mathop{\mathbb{P}}_{Z'_k \leftarrow \mathrm{PrivLoss}(P'_k \| Q'_k)} \left[Z'_k > \varepsilon\right] && \text{(Proposition 3.7)} \\
&\le e^{\varepsilon} \cdot Q'_k(\tilde{S}) + \mathop{\mathbb{E}}_{Z'_k \leftarrow \mathrm{PrivLoss}(P'_k \| Q'_k)} \left[\exp(t(Z'_k - \varepsilon))\right] \\
& && (\mathbb{I}[Z'_k > \varepsilon] \le \exp(t(Z'_k - \varepsilon))) \\
&\le e^{\varepsilon} \cdot Q'_k(\tilde{S}) + \exp\left(\frac{t(t+1)}{2} \sum_{j=1}^{k} \varepsilon_j^2\right) \cdot \exp(-t\varepsilon) \\
& && \text{(Induction conclusion)} \\
&\le e^{\varepsilon} \cdot Q'_k(\tilde{S}) + \delta',
\end{aligned}
$$

where the final inequality holds for the case $\varepsilon = \frac{1}{2}\sum_{j=1}^{k} \varepsilon_j^2 + \sqrt{2\log(1/\delta') \sum_{j=1}^{k} \varepsilon_j^2}$ and requires setting $t = \frac{\varepsilon}{\sum_{j=1}^{k} \varepsilon_j^2} - \frac{1}{2} = \sqrt{\frac{2\log(1/\delta')}{\sum_{j=1}^{k} \varepsilon_j^2}}$.

It only remains for us to perform the induction. The base case ($j = 0$) is trivial.

Fix $j \in [k]$ and assume the induction hypothesis holds for $j-1$. The distribution P_j is defined as a mixture (i.e., convex combination) of $P_j|_Y$ for $Y \leftarrow P_{j-1}$, where $P_j|_Y := (Y, M_j(x, Y_{j-1}))$. For every y, we apply Lemma 3.23 to the conditional distribution $P_j|_y$ and then we take the convex combination of these decompositions to obtain a decomposition of P_j. Of course, we must also decompose Q_j at the same time.

For each $y \in \mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \mathcal{Y}_{j-1}$, the conditional distributions satisfy $\forall S\ P_j|_y(S) \le e^{\varepsilon_j} Q_j|_y(S) + \delta_j$ and vice versa. Thus Lemma 3.23 allows us to decompose the conditional distributions $P_j|_y$ and $Q_j|_y$ as $P_j|_y = (1-\delta_j)P'_j|_y + \delta_j P''_j|_y$ and $Q_j|_y = (1-\delta_j)Q'_j|_y + \delta_j Q''_j|_y$ where $e^{-\varepsilon_j} \cdot Q'_j|_y(S) \le P'_j|_y(S) \le e^{\varepsilon_j} \cdot Q'_j|_y(S)$ for all S. This gives us the desired decomposition:

$$
\begin{aligned}
P_j &= \mathop{\mathbb{E}}_{Y \leftarrow P_{j-1}} \left[P_j|_Y\right] \\
&= \mathop{\mathbb{E}}_{Y \leftarrow P_{j-1}} \left[(1-\delta_j)P'_j|_Y + \delta_j P''_j|_Y\right] \\
&= \prod_{\ell=1}^{j-1}(1-\delta_\ell) \mathop{\mathbb{E}}_{Y \leftarrow P'_{j-1}} \left[(1-\delta_j)P'_j|_Y + \delta_j P''_j|_Y\right] + \left(1 - \prod_{\ell=1}^{j-1}(1-\delta_\ell)\right) \\
&\quad \times \mathop{\mathbb{E}}_{Y \leftarrow P''_{j-1}} \left[(1-\delta_j)P'_j|_Y + \delta_j P''_j|_Y\right]
\end{aligned}
$$

$$= \prod_{\ell=1}^{j}(1-\delta_\ell) \mathop{\mathbb{E}}_{Y \leftarrow P'_{j-1}} \left[P'_j|_Y \right] + \delta_j \prod_{\ell=1}^{j-1}(1-\delta_\ell) \mathop{\mathbb{E}}_{Y \leftarrow P'_{j-1}} \left[P''_j|_Y \right]$$
$$+ \left(1 - \prod_{\ell=1}^{j-1}(1-\delta_\ell) \right) (1-\delta_j) \mathop{\mathbb{E}}_{Y \leftarrow P''_{j-1}} \left[P'_j|_Y \right]$$
$$+ \left(1 - \prod_{\ell=1}^{j-1}(1-\delta_\ell) \right) \delta_j \mathop{\mathbb{E}}_{Y \leftarrow P''_{j-1}} \left[P''_j|_Y \right].$$

Thus we define the new decomposition as $P'_j = P'_j|_Y$ for $Y \leftarrow P'_{j-1}$ and $Q'_j = Q'_j|_Y$ for $Y \leftarrow Q'_{j-1}$. The "weight" of P'_j is the product of the weight of P'_{j-1} (i.e., $\prod_{\ell=1}^{j-1}(1-\delta_\ell)$) and the weight of $P'_j|_Y$ (i.e., $1-\delta_j$), as required. The remaining parts of the decomposition are combined to define $P''_j =$ and Q''_j; note that P''_j includes both $P'_j|_Y$ for $Y \leftarrow P''_{j-1}$ and $P''_j|_Y$ for $Y \leftarrow P_{j-1}$. It is easy to verify that this decomposition satisfies the requirements of the induction:

$$\mathop{\mathbb{E}}_{Z'_j \leftarrow \mathrm{PrivLoss}\left(P'_j \middle\| Q'_j\right)} \left[\exp(tZ'_j) \right]$$
$$= \mathop{\mathbb{E}}_{Y'_j \leftarrow P'_j} \left[\exp\left(t \cdot f_{P'_j \| Q'_j}(Y'_j) \right) \right]$$
$$= \mathop{\mathbb{E}}_{Y'_{j-1} \leftarrow P'_{j-1}} \left[\mathop{\mathbb{E}}_{Y'_j \leftarrow P'_j|_{Y_{j-1}}} \left[\exp\left(t \cdot \left(f_{P'_{j-1} \| Q'_{j-1}}(Y'_{j-1}) + f_{P'_j|_{Y_{j-1}} \| Q'_j|_{Y_{j-1}}}(Y'_j) \right) \right) \right] \right]$$
$$= \mathop{\mathbb{E}}_{Y'_{j-1} \leftarrow P'_{j-1}} \left[\exp\left(t \cdot f_{P'_{j-1} \| Q'_{j-1}}(Y'_{j-1}) \right) \right.$$
$$\left. \cdot \mathop{\mathbb{E}}_{Y'_j \leftarrow P'_j|_{Y_{j-1}}} \left[\exp\left(t \cdot f_{P'_j|_{Y_{j-1}} \| Q'_j|_{Y_{j-1}}}(Y'_j) \right) \right] \right]$$
$$= \mathop{\mathbb{E}}_{Y'_{j-1} \leftarrow P'_{j-1}} \left[\exp\left(t \cdot f_{P'_{j-1} \| Q'_{j-1}}(Y'_{j-1}) \right) \right.$$
$$\left. \cdot \mathop{\mathbb{E}}_{Z'_j \leftarrow \mathrm{PrivLoss}\left(P'_j|_{Y_{j-1}} \middle\| Q'_j|_{Y_{j-1}}\right)} \left[\exp\left(t \cdot Z'_j \right) \right] \right]$$
$$\leq \mathop{\mathbb{E}}_{Y'_{j-1} \leftarrow P'_{j-1}} \left[\exp\left(t \cdot f_{P'_{j-1} \| Q'_{j-1}}(Y'_{j-1}) \right) \cdot \exp\left(t(t+1)\frac{1}{2}\varepsilon_j^2 \right) \right]$$
(Proposition 3.16 & $|Z'_j| \leq \varepsilon_j$)

$$\leq \exp\left(\frac{t(t+1)}{2}\sum_{\ell=1}^{j-1}\varepsilon_\ell^2\right)\cdot\exp\left(t(t+1)\frac{1}{2}\varepsilon_j^2\right) \qquad \text{(Induction hypothesis)}$$

$$= \exp\left(\frac{t(t+1)}{2}\sum_{\ell=1}^{j}\varepsilon_\ell^2\right).$$

And, for pure DP, we have $P'_j(S) = \mathop{\mathbb{E}}_{Y'_{j-1}\leftarrow P'_{j-1}}\left[P'_j|_{Y_{j-1}}(S)\right] \leq \mathop{\mathbb{E}}_{Y'_{j-1}\leftarrow P'_{j-1}}\left[e^{\varepsilon_j} Q'_j|_{Y_{j-1}}(S)\right] \leq \exp\left(\sum_{\ell=1}^{j-1}\varepsilon_\ell\right)\cdot \mathop{\mathbb{E}}_{Y'_{j-1}\leftarrow Q'_{j-1}}\left[e^{\varepsilon_j} Q'_j|_{Y_{j-1}}(S)\right] = \exp\left(\sum_{\ell=1}^{j}\varepsilon_\ell\right)\cdot Q'_j(S)$ for all measurable S. □

3.5 Asymptotic Optimality of Composition

Is the advanced composition theorem optimal? That is, could we prove a result that is stronger? This is an important question, but we first need to think about what optimality even means. Recall that, in Section 3.2.1, we proved that basic composition is optimal, but then we showed that we could do better by relaxing the requirement from pure DP to approximate DP or concentrated DP. To prove asymptotic optimality of advanced composition, we will show that no algorithm can provide better accuracy than advanced composition gives (except for constant factors) subject to approximate DP. Furthermore, we will see that the analysis is not specific to approximate DP.

Combining advanced composition (Theorem 3.18 or 3.22) with Laplace noise addition shows that we can answer k bounded sensitivity queries (e.g., counting queries) with noise scale $\Theta(\sqrt{k/\rho})$ for each query, where ρ only depends on the privacy parameters, e.g., $\rho = \Theta(\varepsilon^2/\log(1/\delta))$ for (ε,δ)-DP. (Gaussian noise addition also gives the same asymptotics, per Corollary 3.10.)

We can prove that this asymptotics – average error per query $\Omega(\sqrt{k})$ – is optimal. Formally, we have the following result.

Theorem 3.24 (Negative Result for Error of Private Mean Estimation)**.** *Let $\mathcal{X} = \{0,1\}^k$ and $\mathcal{Y} = [0,1]^k$. Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ε,δ)-DP. If $\delta \leq 1/100n$ and $k \geq 200(e^\varepsilon - 1)^2 n$, then there exists some $x \in \mathcal{X}^n$ such that*

$$\sqrt{\mathbb{E}\left[\frac{1}{k}\|M(x) - \overline{x}\|_2^2\right]} \geq \min\left\{\frac{\sqrt{k}}{16\cdot n\cdot(e^\varepsilon - 1)}, \frac{1}{10}\right\},$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i \in [0,1]^k$ is the mean of input dataset.

Theorem 3.24 shows that any DP algorithm answering k queries must have error per query scaling with $\Omega(\sqrt{k})$, which matches the guarantees of the advanced composition theorem. We briefly remark on some of the properties of this theorem:

First, M could just output $\frac{1}{2}$ for each coordinate; this is trivially private and has root mean square error at most $\frac{1}{2}$. The theorem must apply to such an algorithm too, which is the fundamental reason why the lower bound in the conclusion of Theorem 3.24 cannot be larger than a constant $\frac{1}{10}$.

Second, the assumption $\delta \leq 1/100n$ is also necessary, up to constant factors. If $\delta \gg 1/n$, then M could sample $n\delta$ of the inputs and return the sample mean. This would be $(0, \delta)$-DP and would give accuracy $\sqrt{\mathbb{E}\left[\frac{1}{k}\|M(x) - \bar{x}\|_2^2\right]} \leq \frac{1}{\sqrt{n\delta}} \ll 1$. Note that the advanced composition theorem includes a $\sqrt{\log(1/\delta)}$ term. It is possible to extend the negative results to include such a term too [SU15] (see also Lemma 2.3.6 of Bun [Bun16]), but we do not do this here for simplicity.

Third, the assumption $k \geq 200(e^{\varepsilon} - 1)^2 n$ is not really necessary; it is an artifact of our analysis. If $k \ll \varepsilon^2 n$, then the privacy error is lower than the sampling error (if we think of x as consisting of n samples from some distribution). A different analysis is possible in this case.

Fourth, Theorem 3.24 has $e^{\varepsilon} - 1$ in the denominator, where our positive results have ε. For small ε, we have $e^{\varepsilon} - 1 \approx \varepsilon$. But, for large ε, there is an exponential difference. Surprisingly, this is inherent; by using discrete noise [CKS20] in place of continuous Laplace noise it is possible to improve the positive results to yield this asymptotic behavior. However, we are generally not interested in the large ε setting.

Finally, fifth, this theorem is not merely an esoteric impossibility result. It corresponds to realistic attacks, which are known as "tracing attacks" [DSSU17] or "membership inference attacks" [SSSS17], which are the subject of Chapter 5 of this book.

Proof of Theorem 3.24. The theorem guarantees that there exists a specific input x on which M has high error. In general, x must depend on M. To prove this we show that, for a random input from a carefully chosen distribution, any M must have high error. It follows that for each specific M there must exist some fixed input with high error.

For $p \in [0, 1]^k$, let $\mathcal{D}_p$ be the product distribution over $\{0, 1\}^k$ with mean p. Our random input $X \in \mathcal{X}^n$ will consist of n independent draws from $\mathcal{D}_p$. Furthermore, we select the mean parameter randomly too. That is, $P \in [0, 1]^d$ is uniformly random and X consists of n conditionally independent draws from $\mathcal{D}_P$.

We analyze the quantity

$$Z := \sum_{i=1}^{n} \langle M(X) - P, X_i - P \rangle .$$

Applying Lemma 3.25 with $f(x) = \mathbb{E}\left[M(X)_j | X_j = x\right]$ and summing over the coordinates $j \in [k]$ shows that

$$\mathop{\mathbb{E}}_{\substack{P \leftarrow \text{Uniform}([0,1]^k) \\ X \leftarrow \mathcal{D}_P^n}} \left[Z + \|M(X) - \overline{X}\|_2^2\right]$$

$$= \sum_{j=1}^{k} \mathop{\mathbb{E}}_{\substack{P \leftarrow \text{Uniform}([0,1]^k) \\ X \leftarrow \mathcal{D}_P^n}} \left[(M(X)_j - P_j) \cdot \sum_{i=1}^{n} (X_{i,j} - P_j)\right] \geq \frac{k}{12}.$$

Denoting $\alpha^2 k = \mathbb{E}\left[\|M(X) - \overline{X}\|_2^2\right]$, we have $\mathbb{E}[Z] \geq \frac{k}{12} - \alpha^2 k$. Intuitively, Z measures the total correlation between the output of M and its inputs. What Lemma 3.25 shows is that, if M is accurate – i.e., $\mathbb{E}\left[\|M(X) - \overline{X}\|_2^2\right] \leq o(k)$ – then this correlation must be large.

The punchline of the proof is that we show that differential privacy means the correlation must be small, which conflicts with the fact that we have proven it must be large. Ergo, we will obtain the desired impossibility result.

For $i \in [n]$, define

$$Z_i = \langle M(X) - P, X_i - P \rangle,$$

so that $Z = \sum_{i=1}^{n} Z_i$. Let X_0 be a fresh sample from $\mathcal{D}_P$ that is (conditionally) independent from $X_1, \cdots, X_n$. Let $M(X_0, X_{-i})$ denote running M on the dataset X where X_i has been replaced by X_0 and define

$$\tilde{Z}_i = \langle M(X_0, X_{-i}) - P, X_i - P \rangle.$$

By differential privacy, $M(X_0, X_{-i})$ is indistinguishable from $M(X)$, even if we condition on $X_0, X_1, \cdots, X_n$. Thus the distributions of $\tilde{Z}_i$ and Z_i are also indistinguishable.

Since $M(X_0, X_{-i})$ and X_i are independent (conditioned on P) and $\mathbb{E}[X_i - P] = \vec{0}$, $\mathbb{E}\left[\tilde{Z}_i\right] = 0$ and

$$\mathop{\mathbb{E}}_{P,X,M}\left[\tilde{Z}_i^2\right] = \sum_{j=1}^{k} \mathop{\mathbb{E}}_{P}\left[P_j(1 - P_j) \cdot \mathop{\mathbb{E}}_{X,M}\left[(M(X_0, X_{-i})_j - P_j)^2\right]\right]$$

$$\leq \frac{1}{4} \mathop{\mathbb{E}}_{P,X,M}\left[\|M(X) - P\|_2^2\right].$$

Now $\mathbb{E}\left[\|M(X) - P\|_2^2\right] \le 2\mathbb{E}\left[\|M(X) - \overline{X}\|_2^2\right] + 2\mathbb{E}\left[\|\overline{X} - P\|_2^2\right] \le 2\alpha^2 k + \frac{k}{3n}$.[vii]

Lemma 3.26, $|Z_i| \le k$, $|\tilde{Z}_i| \le k$, and $\mathbb{E}\left[|\tilde{Z}_i|\right] \le \sqrt{\mathbb{E}\left[\tilde{Z}_i^2\right]}$ (i.e., Jensen's inequality) gives

$$\mathbb{E}[Z_i] \le \mathbb{E}\left[\tilde{Z}_i\right] + (e^\varepsilon - 1)\mathbb{E}\left[|\tilde{Z}_i|\right] + 2\delta k \le \frac{e^\varepsilon - 1}{2}\sqrt{2\alpha^2 k + \frac{k}{3n}} + 2\delta k.$$

Putting things together, we have

$$\frac{k}{12} - \alpha^2 k \le \mathbb{E}[Z] = \sum_{i=1}^{n} \mathbb{E}[Z_i] \le n \cdot \left(\frac{e^\varepsilon - 1}{2}\sqrt{2\alpha^2 k + \frac{k}{3n}} + 2\delta k\right).$$

Ignoring terms that are (hopefully) low order, this is $\Omega(k) \le O(n \cdot \varepsilon\sqrt{\alpha^2 k})$, which rearranges to $\alpha = \sqrt{\mathbb{E}\left[\frac{1}{k}\|M(X) - \overline{X}\|_2^2\right]} \ge \Omega\left(\frac{\sqrt{k}}{n\varepsilon}\right)$, which is the desired asymptotic result. To be precise, this rearranges to

$$\alpha \ge \sqrt{\left(\frac{1}{6} - 2\alpha^2 - 4n\delta\right)^2 \cdot \frac{k}{2n^2 \cdot (e^\varepsilon - 1)^2} - \frac{1}{6n}}.$$

If $\alpha \le 1/10$ and $\delta \le 1/100n$, then $\frac{1}{6} - 2\alpha^2 - 4n\delta \ge \frac{1}{10}$. If $k \ge 200(e^\varepsilon - 1)^2 n$, then $\left(\frac{1}{10}\right)^2 \cdot \frac{k}{2n^2(e^\varepsilon - 1)^2} \ge \frac{1}{n}$. If all three of these conditions hold, then

$$\sqrt{\mathbb{E}\left[\frac{1}{k}\|M(X) - \overline{X}\|_2^2\right]} = \alpha \ge \sqrt{\frac{k}{200 \cdot n^2 \cdot (e^\varepsilon - 1)^2}\left(1 - \frac{1}{6}\right)} \ge \frac{\sqrt{k}}{16n(e^\varepsilon - 1)}.$$

Hence, if $\delta \le 1/100n$ and $k \ge 200(e^\varepsilon - 1)^2 n$, then either $\alpha > 1/10$ or $\alpha \ge \sqrt{k}/16n(e^\varepsilon - 1)$, as required. □

Now we prove the two lemma that were used to prove Theorem 3.24. We begin with the lemma showing that the correlation Z must be large if M is accurate.

The lemma only contemplates one coordinate and then we sum over the k coordinates in the proof of Theorem 3.24. That is, the function f in the theorem is simply one coordinate of M and we average out the randomness of M and the other coordinates.

Lemma 3.25. *Let $f : \{0,1\}^d \to [0,1]$ be an arbitrary function. Let $P \in [0,1]$ be uniformly random and, conditioned on P, let $X_1, \cdots, X_n \in \{0,1\}$ be independent*

vii. $\mathbb{E}\left[\|\overline{X} - P\|_2^2\right] = \sum_{j=1}^{k} \mathop{\mathbb{E}}_{P_j \leftarrow \text{Uniform}([0,1])}\left[\mathop{\mathbb{E}}_{Y_j \leftarrow \text{Binomial}(n,P_j)}\left[(\frac{1}{n}Y_j - P_j)^2\right]\right] = k \cdot \int_0^1 \frac{p(1-p)}{n}\mathrm{d}p = \frac{k}{6n}$.

with $\mathbb{E}[X_i] = P$ *for each* $i \in [n]$. *Then*

$$\mathop{\mathbb{E}}_{X,P}\left[(f(X)-P)\cdot\sum_{i=1}^{n}(X_i-P)\right]+\mathop{\mathbb{E}}_{P}\left[\mathop{\mathbb{E}}_{X}\left[f(X)-\overline{X}\right]^2\right]\geq\frac{1}{12}.$$

By Jensen's inequality $\mathop{\mathbb{E}}_{P}\left[\mathop{\mathbb{E}}_{X}\left[f(X)-\overline{X}\right]^2\right]\leq\mathop{\mathbb{E}}_{P,X}\left[\left(f(X)-\overline{X}\right)^2\right]$. Thus

$$\mathop{\mathbb{E}}_{X,P}\left[(f(X)-P)\cdot\sum_{i=1}^{n}(X_i-P)+\left(f(X)-\overline{X}\right)^2\right]\geq\frac{1}{12}.$$

To gain some intuition for the lemma statement, suppose $f(x) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Then

$$\begin{aligned}\mathbb{E}\left[\left(f(X)-P, X_i-P\right)\right] &= \mathbb{E}\left[\left(\overline{X}-P\right)\cdot\left(\sum_{i=1}^{n}X_i-P\right)\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i-P)^2\right]=\frac{1}{6}.\end{aligned}$$

The constant $\frac{1}{6} = \int_0^1 p(1-p)\mathrm{d}p$ in this example is slightly better than the constant $\frac{1}{12}$ in the general result. However, if $f(x) = \frac{1}{2}$ is a constant function, then the constant is tight, as $\mathop{\mathbb{E}}_{P}\left[\mathop{\mathbb{E}}_{X}\left[f(X)-\overline{X}\right]^2\right] = \mathop{\mathbb{E}}_{P}\left[\left(\frac{1}{2}-P\right)^2\right] = \frac{1}{12}$.

Proof of Lemma 3.25. Define $g : [0,1] \to [0,1]$ by $g(p) = \mathop{\mathbb{E}}_{X\leftarrow\mathcal{D}_p^n}\left[f(X)\right]$, where $\mathcal{D}_p^n$ denotes the product distribution over $\{0,1\}^n$ with each coordinate having mean p. Then

$$\begin{aligned}g'(p) &= \frac{\mathrm{d}}{\mathrm{d}p}\mathop{\mathbb{E}}_{X\leftarrow\mathcal{D}_p^n}\left[f(X)\right]\\ &= \sum_{x\in\{0,1\}^n} f(x)\frac{\mathrm{d}}{\mathrm{d}p}\prod_{\ell=1}^{n}\left(x_\ell\cdot p+(1-x_\ell)\cdot(1-p)\right)\\ &= \sum_{x\in\{0,1\}^n} f(x)\prod_{\ell=1}^{n}\left(x_\ell\cdot p+(1-x_\ell)\right.\\ &\quad\left.\cdot(1-p)\right)\sum_{i=1}^{n}\frac{\frac{\mathrm{d}}{\mathrm{d}p}(px_i\cdot p+(1-x_i)\cdot(1-p))}{x_i\cdot p+(1-x_i)\cdot(1-p)} \qquad \text{(Product rule)}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x\in\{0,1\}^n} f(x) \prod_{\ell=1}^{n} \big(x_\ell \cdot p + (1-x_\ell) \\
&\quad \cdot (1-p)\big) \sum_{i=1}^{n} \frac{2x_i - 1}{x_i \cdot p + (1-x_i)\cdot(1-p)} \\
&= \sum_{x\in\{0,1\}^n} f(x) \prod_{\ell=1}^{n} \big(x_\ell \cdot p + (1-x_\ell)\cdot(1-p)\big) \sum_{i=1}^{n} \frac{x_i - p}{p(1-p)} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(Case analysis for } x_i \in \{0,1\}\text{)} \\
&= \mathop{\mathbb{E}}_{X\leftarrow\mathcal{D}_p^n}\left[f(X) \cdot \sum_{i=1}^{n} \frac{X_i - p}{p(1-p)} \right].
\end{aligned}
$$

Now we apply integration by parts to this derivative:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{P\leftarrow[0,1]}\left[\mathop{\mathbb{E}}_{X\leftarrow\mathcal{D}_P^n}\left[f(X)\cdot \sum_{i=1}^{n}(X_i - P)\right]\right] \\
&= \int_0^1 g'(p)\cdot p(1-p)\mathrm{d}p \\
&= \int_0^1 \left(\frac{\mathrm{d}}{\mathrm{d}p} g(p)\cdot p(1-p)\right) - g(p)\cdot(1-2p)\mathrm{d}p \\
&= g(1)\cdot 1(1-1) - g(0)\cdot 0(1-0) + \int_0^1 g(p)\cdot(2p-1)\mathrm{d}p \\
&= 2 \mathop{\mathbb{E}}_{P\leftarrow[0,1]}\left[g(P)\cdot\left(P - \frac{1}{2}\right)\right].
\end{aligned}
$$

Using the fact that $\mathop{\mathbb{E}}_{P\leftarrow[0,1]}\left[P - \frac{1}{2}\right] = 0$ and $\mathop{\mathbb{E}}_{X\leftarrow\mathcal{D}_p^n}\left[X_i - p\right] = 0$, we can center these expressions:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{P\leftarrow[0,1]\\ X\leftarrow\mathcal{D}_P^n}}\left[(f(X) - P)\cdot \sum_{i=1}^{n}(X_i - P)\right] \\
&= 2 \mathop{\mathbb{E}}_{P\leftarrow[0,1]}\left[\left(g(P) - \frac{1}{2}\right)\cdot\left(P - \frac{1}{2}\right)\right] \\
&= \mathop{\mathbb{E}}_{P\leftarrow[0,1]}\left[\left(g(P) - \frac{1}{2}\right)^2 + \left(P - \frac{1}{2}\right)^2 - \left(\left(g(P) - \frac{1}{2}\right) - \left(P - \frac{1}{2}\right)\right)^2\right] \\
&\geq \mathop{\mathbb{E}}_{P\leftarrow[0,1]}\left[0 + \left(P - \frac{1}{2}\right)^2 - \big(g(P) - P\big)^2\right]
\end{aligned}
$$

$$= \frac{1}{12} - \mathop{\mathbb{E}}_{P \leftarrow [0,1]} \left[\left(g(P) - P \right)^2 \right]$$
$$= \frac{1}{12} - \mathop{\mathbb{E}}_{P \leftarrow [0,1]} \left[\mathop{\mathbb{E}}_{X \leftarrow \mathcal{D}_P^n} \left[f(X) - \overline{X} \right]^2 \right].$$

□

Lemma 3.26. *Let X and Y be random variables supported on $[-\Delta, \Delta]$ satisfying $\mathbb{P}[X \in S] \leq e^{\varepsilon} \cdot \mathbb{P}[Y \in S] + \delta$ and $\mathbb{P}[Y \in S] \leq e^{\varepsilon} \cdot \mathbb{P}[X \in S] + \delta$ for all measurable S. Then*

$$\mathbb{E}[X] \leq \mathbb{E}[Y] + (e^{\varepsilon} - 1)\mathbb{E}[|Y|] + 2\delta\Delta.$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{\Delta} \mathbb{P}[X > t] - \mathbb{P}[X < -t] \mathrm{d}t \\
&\leq \int_0^{\Delta} e^{\varepsilon} \cdot \mathbb{P}[Y > t] + \delta - e^{-\varepsilon} \cdot (\mathbb{P}[Y < -t] - \delta) \mathrm{d}t \\
&= \int_0^{\Delta} (\mathbb{P}[Y > t] - \mathbb{P}[Y < -t]) + (e^{\varepsilon} - 1) \cdot \mathbb{P}[Y > t] + (1 - e^{-\varepsilon}) \\
&\quad \cdot \mathbb{P}[Y < -t] + (1 + e^{-\varepsilon})\delta \mathrm{d}t \\
&= \mathbb{E}[Y] + (e^{\varepsilon} - 1)\mathbb{E}[\max\{0, Y\}] + (1 - e^{-\varepsilon})\mathbb{E}[\max\{0, -Y\}] \\
&\quad + (1 + e^{-\varepsilon})\delta\Delta \\
&\leq \mathbb{E}[Y] + (e^{\varepsilon} - 1)\mathbb{E}[|Y|] + 2\delta\Delta,
\end{aligned}$$

as $1 - e^{-\varepsilon} \leq e^{\varepsilon} - 1$ and $1 + e^{-\varepsilon} \leq 2$. □

Remark 3.27. *The only part of the proof of Theorem 3.24 that uses differential privacy is Lemma 3.26. Thus, if we were to consider a different definition of differential privacy, as long as an analog of Lemma 3.26 holds for this alternative definition, an analog of Theorem 3.24 would still apply. That is to say, this negative result is robust to our choice of privacy definition (unlike the the negative result in Section 3.2.1).*

3.6 Privacy Amplification by Subsampling

Thus far we have considered the composition of Gaussian mechanisms, and generic mechanisms satisfying pure or approximate DP. We now turn our attention to subsampled privacy mechanisms. These mechanisms introduce some additional quirks into the picture, which will force us to develop new tools.

The premise of privacy amplification by subsampling is that we run a DP algorithm on some random subset of the data. The subset introduces additional uncertainty, which benefits privacy. In particular, there is some probability that your data is not included in the analysis, which can only enhance your privacy. Furthermore, a potential attacker does not know whether or not your data was dropped; this uncertainty can benefit your privacy even when your data is included. Privacy amplification by subsampling theorems make this intuition precise.

Subsampling arises naturally. We often would like to collect the data of the entire population, but this is impractical. Thus we collect the data of a subset of the population and use statistical methods to generalize from this sample to the entire population. In particular, in deep learning applications, we will use stochastic gradient descent. That is, we choose a random subset of our training data (called a minibatch) and compute the gradient of the loss function with respect to this subset, rather than the entire dataset. This method reduces the computational cost for training. If we want to make deep learning differentially private, then we will add noise to the gradients and we should exploit privacy amplification by subsampling to analyze the privacy properties of this algorithm.

In this section we will analyze subsampling precisely and we will show how it interacts with composition.

3.6.1 Subsampling for Pure or Approximate DP

We begin by analyzing privacy amplification by subsampling under pure or approximate differential privacy. This is a relatively simple result, but it will be instructive as we later attempt to derive more precise bounds.

Theorem 3.28 (Privacy Amplification by Subsampling for Approximate DP)**.** *Let $U \subset [n]$ be a random subset. For a dataset $x \in \mathcal{X}^n$, let $x_U \in \mathcal{X}^n$ denote the entries of x indexed by U. That is, $(x_U)_i = x_i$ if $i \in U$ and $(x_U)_i = \perp$ if $i \notin U$, where $\perp \in \mathcal{X}$ is some null value.*

Assume that, for all $i \in [n]$, we can define $U_{-i} \subset [n] \setminus \{i\}$ such that the following two conditions hold.

- *For all $x \in \mathcal{X}^n$ and $i \in [d]$, x_U and $x_{U_{-i}}$ are always neighboring datasets.*
- *For all $i \in [n]$, the marginal distribution of U_{-i} conditioned on $i \in U$ is equal to the marginal distribution of U conditioned on $i \notin U$.*

Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ε, δ)-DP. Define $M^U : \mathcal{X}^n \to \mathcal{Y}$ by $M^U(x) = M(x_U)$.

Let $p = \max_{i \in [n]} \mathop{\mathbb{P}}_{U}[i \in U]$. Then M^U is (ε', δ')-DP for $\varepsilon' = \log(1 + p(e^\varepsilon - 1))$ and $\delta' = p \cdot \delta$.

For small values of ε, we have $\varepsilon' = \log(1 + p(e^\varepsilon - 1)) \approx p \cdot \varepsilon$. More precisely, $\varepsilon' = \log(1+p(e^\varepsilon-1)) \le p\cdot(e^\varepsilon-1)$ and, for $\varepsilon \le 1$, we have $e^\varepsilon - 1 \le \varepsilon+\varepsilon^2 \le 2\varepsilon$.

The technical assumption about U in the theorem statement is satisfied by many natural subsampling distributions: If U is a uniformly random subset of $[n]$ of a fixed size m, then U_{-i} can be obtained by replacing i with a a uniformly random element that is not in U. If U is Poisson subsampled – i.e., each $i \in [n]$ is independently included in U with probability p – then, by independence, we can simply remove i, namely $U_{-i} = U \setminus \{i\}$.

The technical assumption should be thought of as an independence assumption. For example, it rules out distributions of the form $\mathbb{P}_U[U = [n]] = p$ and $\mathbb{P}[U = \emptyset] = 1 - p$, which do not yield meaningful privacy amplification.

Proof of Theorem 3.28. Fix neighboring inputs $x, x' \in \mathcal{X}^n$ and some measurable $S \subset \mathcal{Y}$. Let $i \in [n]$ be the index on which they differ (i.e., $x_j = x'_j$ for all $j \in [n]\setminus\{i\}$) and let $p_i = \mathbb{P}_U[i \in U]$. We have

$$\begin{aligned}
\mathbb{P}_{M^U}\left[M^U(x) \in S\right] &= \mathbb{E}_U\left[\mathbb{P}_M[M(x_U) \in S]\right] \\
&= (1-p_i)\cdot \mathbb{E}_U\left[\mathbb{P}_M[M(x_U) \in S] \mid i \notin U\right] \\
&\quad + p_i \cdot \mathbb{E}_U\left[\mathbb{P}_M[M(x_U) \in S] \mid i \in U\right] \\
&= (1-p_i)\cdot \mathbb{E}_U\left[\mathbb{P}_M\left[M(x'_U) \in S\right] \mid i \notin U\right] \\
&\quad + p_i \cdot \mathbb{E}_U\left[\mathbb{P}_M[M(x_U) \in S] \mid i \in U\right] \\
&= (1-p_i)\cdot a + p_i \cdot b, \\
\mathbb{P}_{M^U}\left[M^U(x') \in S\right] &= (1-p_i)\cdot \mathbb{E}_U\left[\mathbb{P}_M\left[M(x'_U) \in S\right] \mid i \notin U\right] \\
&\quad + p_i \cdot \mathbb{E}_U\left[\mathbb{P}_M\left[M(x'_U) \in S\right] \mid i \in U\right] \\
&= (1-p_i)\cdot a + p_i \cdot b',
\end{aligned}$$

where $a = \mathbb{E}_U\left[\mathbb{P}_M[M(x_U) \in S] \mid i \notin U\right] = \mathbb{E}_U\left[\mathbb{P}_M\left[M(x'_U) \in S\right] \mid i \notin U\right]$, $b = \mathbb{E}_U\left[\mathbb{P}_M[M(x_U) \in S] \mid i \in U\right]$, and $b' = \mathbb{E}_U\left[\mathbb{P}_M\left[M(x'_U) \in S\right] \mid i \in U\right]$.

Note that x_U and x'_U are always neighboring datasets. And, if $i \notin U$, then $x_U = x'_U$. Since M is (ε, δ)-DP, we have $\mathop{\mathbb{P}}_M[M(x_U) \in S] \leq e^\varepsilon \cdot \mathop{\mathbb{P}}_M\left[M(x'_U) \in S\right] + \delta$ for all values of U; thus

$$\begin{aligned} b &= \mathop{\mathbb{E}}_U\left[\mathop{\mathbb{P}}_M[M(x_U) \in S] \mid i \in U\right] \leq \mathop{\mathbb{E}}_U\left[e^\varepsilon \cdot \mathop{\mathbb{P}}_M\left[M(x'_U) \in S\right] + \delta \mid i \in U\right] \\ &= e^\varepsilon \cdot b' + \delta. \end{aligned}$$

However, this inequality alone is not sufficient to prove the claim. We also need to show that $b \leq e^\varepsilon \cdot a + \delta$. Using our technical assumption, we have

$$\begin{aligned} b &= \mathop{\mathbb{E}}_U\left[\mathop{\mathbb{P}}_M[M(x_U) \in S] \mid i \in U\right] \\ &\leq \mathop{\mathbb{E}}_U\left[e^\varepsilon \cdot \mathop{\mathbb{P}}_M\left[M(x_{U_{-i}}) \in S\right] + \delta \mid i \in U\right] \qquad (x_{U_{-i}} \text{ is a neighbour of } x_U) \\ &= \mathop{\mathbb{E}}_U\left[e^\varepsilon \cdot \mathop{\mathbb{P}}_M[M(x_U) \in S] + \delta \mid i \notin U\right] \\ &\qquad\qquad (U_{-i}|_{i\in U} \text{ has the same distribution as } U|_{i\notin U}) \\ &= e^\varepsilon \cdot a + \delta. \end{aligned}$$

Now we can complete the proof: For any $\lambda \in [0, 1]$,

$$\begin{aligned} \mathop{\mathbb{P}}_{M^U}\left[M^U(x) \in S\right] &= (1 - p_i) \cdot a + p_i \cdot b \\ &\leq (1 - p_i) \cdot a + p_i \cdot ((1 - \lambda) \cdot (e^\varepsilon \cdot a + \delta) + \lambda \cdot (e^\varepsilon \cdot b' + \delta)) \\ &= (1 - p_i + e^\varepsilon \cdot (1 - \lambda) \cdot p_i) \cdot a + p_i \cdot e^\varepsilon \cdot \lambda \cdot b' + p_i \cdot \delta. \end{aligned}$$

Set $\lambda = p_i + (1 - p_i) \cdot e^{-\varepsilon}$ to obtain

$$\begin{aligned} \mathop{\mathbb{P}}_{M^U}\left[M^U(x) \in S\right] &\leq (1 - p_i + e^\varepsilon \cdot (1 - \lambda) \cdot p_i) \cdot a + p_i \cdot e^\varepsilon \cdot \lambda \cdot b' + p_i \cdot \delta \\ &= \left(1 + p_i \cdot (e^\varepsilon - 1)\right) \cdot \left((1 - p_i) \cdot a + p_i \cdot b'\right) + p_i \cdot \delta \\ &= \left(1 + p_i \cdot (e^\varepsilon - 1)\right) \cdot \mathop{\mathbb{P}}_{M^U}\left[M_U(x') \in S\right] + p_i \cdot \delta \\ &\leq e^{\varepsilon'} \cdot \mathop{\mathbb{P}}_{M^U}\left[M_U(x') \in S\right] + \delta'. \end{aligned}$$

□

Theorem 3.28 is tight: Consider an algorithm $M : \{0, 1, \perp\}^n \to \{0, 1\}$ that sums its input (excluding $\perp$ values) and performs randomized response on whether

or not the sum is 0.[viii] That is, if $y \in \{0,1,\perp\}^n$ satisfies $\sum_{i:y_i \neq \perp} y_i = 0$, then $\mathbb{P}\left[M(y)=0\right] = \frac{e^\varepsilon}{e^\varepsilon+1}$ and $\mathbb{P}\left[M(y)=1\right] = \frac{1}{e^\varepsilon+1}$ and, if $y \in \{0,1,\perp\}^n$ satisfies $\sum_{i:y_i \neq \perp} y_i > 0$, then $\mathbb{P}\left[M(y)=1\right] = \frac{e^\varepsilon}{e^\varepsilon+1}$ and $\mathbb{P}\left[M(y)=0\right] = \frac{1}{e^\varepsilon+1}$. This algorithm satisfies ε-DP.

Let $U \subset [n]$ be random and let $M^U : \{0,1\}^n \to \{0,1\}$ be as in Theorem 3.28. Consider neighboring datasets $x = (0,0,\cdots,0)$ and $x' = (1,0,0,\cdots,0)$. We have

$$\begin{aligned}
\mathbb{P}\left[M^U(x)=0\right] &= \frac{e^\varepsilon}{e^\varepsilon+1},\\
\mathbb{P}\left[M^U(x)=1\right] &= \frac{1}{e^\varepsilon+1},\\
\mathbb{P}\left[M^U(x')=1\right] &= \frac{\mathbb{P}\left[1 \in U\right]\cdot e^\varepsilon + \mathbb{P}\left[1 \notin U\right]}{e^\varepsilon+1},\\
\mathbb{P}\left[M^U(x')=0\right] &= \frac{\mathbb{P}\left[1 \in U\right] + \mathbb{P}\left[1 \notin U\right]\cdot e^\varepsilon}{e^\varepsilon+1},\\
e^{\varepsilon'} \geq \frac{\mathbb{P}\left[M^U(x')=1\right]}{\mathbb{P}\left[M^U(x)=1\right]} &= 1 + \mathbb{P}\left[1 \in U\right]\cdot(e^\varepsilon - 1),
\end{aligned}$$

where ε' is the pure DP parameter satisfied by M^U. We can assume without loss of generality that $p = \max_i \mathbb{P}\left[i \in U\right] = \mathbb{P}\left[1 \in U\right]$. Thus this bound matches the guarantee of Theorem 3.28. (This example can be extended to approximate DP too.)

3.6.2 Addition or Removal versus Replacement for Neighboring Datasets

For this discussion of subsampling, we need to be careful about what it means for datasets to be neighboring. There are three common definitions of what qualifies as neighboring datasets: (i) addition or removal of one person's data, (ii) replacement of one person's data, or (iii) both. Each of these three options is a reasonable choice. Work on differential privacy often glosses over this choice – often the choice is irrelevant. But it becomes relevant if we want sharp analyses of privacy amplification by subsampling.

For the discussion of composition so far in this chapter, it does not matter at all how we define neighboring datasets, as long as we are consistent. In general, it only matters slightly which we choose: A replacement can be accomplished by a

viii. Alternatively (and equivalently), consider an algorithm that adds discrete Laplace noise to the sum of its non-null inputs.

combination of a removal and an addition. Thus, by group privacy, if the algorithm is (ε, δ)-DP with respect to addition or removal, then it is $(2\varepsilon, (1 + e^{\varepsilon})\delta)$-DP with respect to replacement. Conversely, we can simulate a removal or addition by replacing the record with a "null" value ($\perp$ in the formalism of Theorem 3.28). Thus DP with respect to replacement entails DP with respect to addition or removal with the same parameters, unless the semantics of the algorithm forbids null values.

This subtlety already arises in Theorem 3.28. Let's take a close look at the technical assumption: Theorem 3.28 assumes that, for all $i \in [n]$, we can define $U_{-i} \subset [n] \setminus \{i\}$ such that the following two conditions hold.

- For all $x \in \mathcal{X}^n$ and $i \in [d]$, x_U and $x_{U_{-i}}$ are always neighboring datasets.
- For all $i \in [n]$, the marginal distribution of U_{-i} conditioned on $i \in U$ is equal to the marginal distribution of U conditioned on $i \notin U$.

Suppose $U \subset [n]$ is a uniformly random subset of a fixed size $|U| = m$. Then we would define U_{-i} to be U with i replaced by a uniformly random element that is not already in U. Thus, for x_U and $x_{U_{-i}}$ to be neighboring datasets, our neighboring relation must allow replacement, not just addition or removal.

However, if U corresponds to Poisson subsampling (i.e., each $i \in [n]$ is included in U independently with probability p), then U_{-i} would just correspond to removing i. In that case, for x_U and $x_{U_{-i}}$ to be neighboring datasets, our neighboring relation must allow addition and removal.

It turns out to be easier to work with Poisson subsampling and assuming the neighboring relation is addition or removal. In this case, the proof of Theorem 3.28 simplifies to the following.

Proof of Theorem 3.28 for the special case of Poisson sampling and addition or removal. Let $U \subset [n]$ independently include each element with probability p. Let $x, x' \in \mathcal{X}^n$ be neighboring datasets in terms of addition or removal. Without loss of generality, assume x' is x with x_i removed (or, rather, replaced by $x'_i = \perp$).[ix] For any measurable $S \subset \mathcal{Y}$,

$$\begin{aligned}
\mathop{\mathbb{P}}_{M^U}\left[M^U(x) \in S\right] &= \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x_U) \in S\right]\right] \\
&= (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x_U) \in S\right] \mid i \notin U\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x_U) \in S\right] \mid i \in U\right]
\end{aligned}$$

ix. To be formal, we assume $\perp \in \mathcal{X}$ is a null value that is equivalent to removing the item. In particular, for $x \in \mathcal{X}^n$ and $U \subset [n]$ we can define $x_U \in \mathcal{X}^n$ such that $(x_U)_i = x_i$ if $i \in U$ and $(x_U)_i = \perp$ if $i \in [n] \setminus U$.

$$
\begin{aligned}
&= (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right] \mid i \notin U\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x_U) \in S\right] \mid i \in U\right] \\
&= (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right]\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x_U) \in S\right] \mid i \in U\right] \\
&\leq (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right]\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right] + \delta \mid i \in U\right] \\
&= (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right]\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[e^{\varepsilon} \cdot \mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right] + \delta\right] \\
&= (1-p+p \cdot e^{\varepsilon}) \cdot \mathop{\mathbb{P}}_{M^U}\left[M^U(x') \in S\right] + p \cdot \delta
\end{aligned}
$$

and, by the same calculation,

$$
\begin{aligned}
\mathop{\mathbb{P}}_{M^U}\left[M^U(x) \in S\right] &= (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right]\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x_U) \in S\right] \mid i \in U\right] \\
&\geq (1-p) \cdot \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right]\right] \\
&\quad + p \cdot \mathop{\mathbb{E}}_{U}\left[e^{-\varepsilon} \cdot (\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right] - \delta) \mid i \in U\right] \\
&= (1-p+p \cdot e^{-\varepsilon}) \cdot \mathop{\mathbb{P}}_{M^U}\left[M^U(x') \in S\right] - p \cdot e^{-\varepsilon} \cdot \delta \\
&\geq \frac{1}{1-p+p \cdot e^{\varepsilon}} \cdot (\mathop{\mathbb{P}}_{M^U}\left[M^U(x') \in S\right] - p \cdot \delta).
\end{aligned}
$$

(Lemma 3.35)

The key step in the proof is the equality $\mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right] \mid i \notin U\right] = \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right] \mid i \in U\right] = \mathop{\mathbb{E}}_{U}\left[\mathop{\mathbb{P}}_{M}\left[M(x'_U) \in S\right]\right]$. This holds because $x'_i =$

$\perp$, so whether or not $i \in U$ is irrelevant for x'_U, and because the event $i \in U$ is independent from $U \setminus \{i\}$. □

For the rest of this section, we will restrict our attention to Poisson subsampling and assume that the neighboring relation corresponds to addition or removal of one individual's data.

3.6.3 Subsampling & Composition

How does composition work with subsampling? Of course, we can combine the advanced composition theorem (Theorem 3.22) with our privacy amplification by subsampling result (Theorem 3.28). However, it turns out this is not the best way to analyze many realistic systems.

Consider the following algorithm (which arises in differentially private deep learning applications). Let $x \in \mathcal{X}^n$ be the private input. Iteratively, for $t = 1, \cdots, T$, we pick some function $q_t : \mathcal{X}^n \to \mathbb{R}^d$ and randomly sample a subset $U_t \subset [n]$; then we reveal $\mathcal{N}(q_t(x_{U_t}), \sigma^2 I_d)$.

This algorithm interleaves composition with privacy amplification by subsampling. That is, we combine multivariate Gaussian noise addition (which is a form of composition over the d coordinates) with subsampling and then we compose over the T iterations.

We can use Corollary 3.10 to show that releasing $\mathcal{N}(q_t(x), \sigma^2 I_d)$ satisfies $(\varepsilon_0, \delta_0)$-DP for $\varepsilon_0 = O\left(\sqrt{\frac{\Delta_2^2}{\sigma^2}\log(1/\delta_0)}\right)$, where $\Delta_2 = \sup_{\substack{x,x' \in \mathcal{X}^n \\ \text{neighboring}}} \|q_t(x) - q_t(x')\|_2$ is the sensitivity of q_t. Then we can use Theorem 3.28 to show that, if U_t is a Poisson sample which contains each element with probability p, then $\mathcal{N}(q_t(x_{U_t}), \sigma^2 I_d)$ is $(\varepsilon_1, \delta_1)$-DP where $\varepsilon_1 = \log(1 + p \cdot (e^{\varepsilon_0} - 1)) = O(p \cdot \varepsilon_0)$ and $\delta_1 = p \cdot \delta_0$. Finally, Theorem 3.22 tells us that the composition over T iterations satisfies (ε, δ)-DP with $\varepsilon = O\left(\varepsilon_1 \cdot \sqrt{T\log(1/\delta_2)}\right)$ and $\delta = \delta_2 + T \cdot \delta_1$. Overall, we have

$$\varepsilon = O\left(\frac{\Delta_2}{\sigma} \cdot p \cdot \sqrt{T} \cdot \log(T/\delta)\right).$$

This result is asymptotically suboptimal because we have picked up two $\sqrt{\log(1/\delta)}$ terms. We obtained one from the Gaussian noise addition (Corollary 3.10) and another from the composition (Theorem 3.22). Both arise from bounding the tails of the privacy loss distribution. This is redundant; we should only need to bound the tails of the privacy loss distribution once.

Intuitively, we started with a Gaussian privacy loss; then we applied a tail bound to obtain a $(\varepsilon_0, \delta_0)$-DP guarantee to which we applied the subsampling theorem; and then we converted this back into a concentrated DP guarantee to apply

advanced composition and finally we applied a tail bound to convert this back to (ε, δ)-DP.

We are going to avoid this redundancy by analyzing privacy amplification by subsampling directly in terms of the privacy loss distribution, rather than needing to go via approximate DP. To do so, we need to introduce a new tool.

3.6.4 Rényi Differential Privacy

Rényi differential privacy was introduced by Mironov [Mir17] and was motivated by analyzing privacy amplification by subsampling interleaved with composition, which arises in differentially private deep learning [Aba+16].

Definition 3.29 (Rényi Differential Privacy)**.** *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized algorithm. We say that M satisfies (α, ε)-Rényi differential privacy ((α, ε)-RDP) if, for all neighboring inputs $x, x' \in \mathcal{X}^n$, the privacy loss distribution* $\mathrm{PrivLoss}\big(M(x)\|M(x')\big)$ *is well-defined (see Definition 3.2) and*

$$\mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(M(x)\|M(x'))}\Big[\exp((\alpha-1)Z)\Big] \le \exp((\alpha-1)\cdot\varepsilon).$$

Rényi DP is closely related to concentrated DP (Definition 3.11). Specifically, ρ-zCDP is equivalent to satisfying $(\alpha, \alpha\cdot\rho)$-RDP for all $\alpha \in (1, \infty)$. Rényi DP inherits the nice composition properties of concentrated DP:

Lemma 3.30. *Let $M_1 : \mathcal{X}^n \to \mathcal{Y}_1$ be (α, ε_1)-RDP. Let $M_2 : \mathcal{X}^n \times \mathcal{Y}_1 \to \mathcal{Y}_2$ be such that, for all $y_1 \in \mathcal{Y}_1$, the algorithm $x \mapsto M(x, y_1)$ is (α, ε_2)-RDP. Define $M : \mathcal{X}^n \to \mathcal{Y}_2$ by $M(x) = M_2(x, M_1(x))$. Then M is $(\alpha, \varepsilon_1 + \varepsilon_2)$-RDP.*

The proof of Lemma 3.30 is identical to that of Proposition 3.19. Note that, while the ε parameter adds up, the α parameter does not change. More generally, composing an $(\alpha_1, \varepsilon_1)$-RDP algorithm with an $(\alpha_2, \varepsilon_2)$-RDP algorithm yields $(\min\{\alpha_1, \alpha_2\}, \varepsilon_1 + \varepsilon_2)$-RDP.

It is helpful to think of ε in (α, ε)-RDP as a function of α, rather than a single number. This function can encode a rich variety of privacy guarantees. (Concentrated DP corresponds to a linear function.) In particular, it allows us to more precisely represent the kinds of guarantees obtained by subsampling.

Concentrated DP corresponds to the privacy loss being subgaussian (i.e., ρ-zCDP implies $\mathop{\mathbb{P}}_{Z \leftarrow \mathrm{PrivLoss}(M(x)\|M(x'))}[Z > \tilde{\varepsilon}] \le \exp(-(\tilde{\varepsilon}-\rho)^2/4\rho)$ for all $\tilde{\varepsilon} \ge \rho$ and all neighboring inputs x and x'), whereas Rényi DP corresponds to the privacy loss being subexponential (i.e., (α, ε)-RDP implies $\mathop{\mathbb{P}}_{Z \leftarrow \mathrm{PrivLoss}(M(x)\|M(x'))}[Z > \tilde{\varepsilon}] \le \exp(-(\alpha-1)(\tilde{\varepsilon}-\varepsilon))$). That is, Rényi DP is more appropriate for analyzing privacy loss distributions with slightly heavier tails

than Gaussian. In contrast, pure DP corresponds to the privacy loss being bounded (i.e., ε-DP implies $\mathop{\mathbb{P}}_{Z \leftarrow \text{PrivLoss}(M(x)\|M(x'))}[Z > \varepsilon] = 0$). So we can view concentrated DP as a relaxation of pure DP and, in turn, Rényi DP is a relaxation of concentrated DP.

Rényi DP is typically formulated in terms of Rényi divergences [Rén61], which were studied in the information theory literature long before differential privacy was discovered.

Definition 3.31 (Rényi Divergences)**.** *Let P and Q be distributions over $\mathcal{Y}$.*[x] *For $\alpha \in (1, \infty)$, define*

$$\begin{aligned}
\mathrm{D}_\alpha(P\|Q) &= \frac{1}{\alpha - 1} \log \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P\|Q)} \left[\exp((\alpha - 1)Z)\right] \\
&= \frac{1}{\alpha - 1} \log \mathop{\mathbb{E}}_{Y \leftarrow P} \left[\left(\frac{P(Y)}{Q(Y)}\right)^{\alpha - 1}\right] \\
&= \frac{1}{\alpha - 1} \log \mathop{\mathbb{E}}_{Y \leftarrow Q} \left[\left(\frac{P(Y)}{Q(Y)}\right)^{\alpha}\right].
\end{aligned}$$

Also, define

$$\begin{aligned}
\mathrm{D}_1(P\|Q) &= \lim_{\alpha \to 1} \mathrm{D}_\alpha(P\|Q) \\
&= \mathop{\mathbb{E}}_{Z \leftarrow \text{PrivLoss}(P\|Q)}[Z] \\
&= \mathop{\mathbb{E}}_{Y \leftarrow P}\left[\log\left(\frac{P(Y)}{Q(Y)}\right)\right], \\
\mathrm{D}_\infty(P\|Q) &= \lim_{\alpha \to \infty} \mathrm{D}_\alpha(P\|Q) \\
&= \sup\left\{\log\left(\frac{P(S)}{Q(S)}\right) : S \subset \mathcal{Y}, Q(S) > 0\right\}.
\end{aligned}$$

Thus an equivalent definition of M satisfying (α, ε)-RDP is that $\mathrm{D}_\alpha\left(M(x)\middle\|M(x')\right) \le \varepsilon$ for all neighboring x, x'.

We now state several key properties of Rényi divergences; most of these are properties we have proved earlier, but we now restate them in a new language.

x. We make the usual measure theoretic disclaimers: We assume the P and Q have the same sigma-algebra. We assume P is absolutely continuous with respect to Q – i.e., $\forall S \;\; Q(S) = 0 \implies P(S) = 0$ – so that the Radon-Nikodym derivative is well-defined; we denote the Radon-Nikodym derivative of P with respect to Q evaluated at y by $P(y)/Q(y)$. More generally, if the absolute continuity assumption does not hold, then we define $\mathrm{D}_\alpha(P\|Q) = \infty$ for all $\alpha \in [1, \infty]$.

Lemma 3.32. *Let P, Q be probability distributions over $\mathcal{Y}$ with a common sigma-algebra such that P is absolutely continuous with respect to Q.*

1. ***Post-processing (a.k.a. data processing inequality) & non-negativity:*** *Let $f : \mathcal{Y} \to \mathcal{Z}$ be a measurable function. Let $f(P)$ denote the distribution on $\mathcal{Z}$ obtained by applying f to a sample from P; define $f(Q)$ similarly. Then $0 \leq \mathrm{D}_\alpha\left(f(P)\middle\|f(Q)\right) \leq \mathrm{D}_\alpha\left(P\middle\|Q\right)$ for all $\alpha \in [1,\infty]$.*
2. ***Composition:*** *If $P = P' \times P''$ and $Q = Q' \times Q''$ are product distributions, then $\mathrm{D}_\alpha\left(P\middle\|Q\right) = \mathrm{D}_\alpha\left(P'\middle\|Q'\right) + \mathrm{D}_\alpha\left(P''\middle\|Q''\right)$ for all $\alpha \in [1,\infty]$.*

   More generally, suppose P and Q are distributions on $\mathcal{Y} = \mathcal{Y}' \times \mathcal{Y}''$. Let P' and Q' be the marginal distributions on $\mathcal{Y}'$ induced by P and Q respectively. For $y' \in \mathcal{Y}'$, let $P''_{y'}$ and $Q''_{y'}$ be the conditional distributions on $\mathcal{Y}''$ induced by P and Q respectively. That is, we can generate a sample $Y = (Y', Y'') \leftarrow P$ by first sampling $Y' \leftarrow P'$ and then sampling $Y'' \leftarrow P''_{Y'}$ and similarly for Q. Then $\mathrm{D}_\alpha\left(P\middle\|Q\right) \leq \mathrm{D}_\alpha\left(P'\middle\|Q'\right) + \sup_{y' \in \mathcal{Y}'} \mathrm{D}_\alpha\left(P''_{y'}\middle\|Q''_{y'}\right)$ for all $\alpha \in [1,\infty]$.
3. ***Monotonicity:*** *For all $1 \leq \alpha \leq \alpha' \leq \infty$, $\mathrm{D}_\alpha\left(P\middle\|Q\right) \leq \mathrm{D}_{\alpha'}\left(P\middle\|Q\right)$.*
4. ***Gaussian divergence:*** *For all $\mu, \mu', \sigma \in \mathbb{R}$ with $\sigma > 0$ and all $\alpha \in [1,\infty)$,*

$$\mathrm{D}_\alpha\left(\mathcal{N}(\mu,\sigma^2)\middle\|\mathcal{N}(\mu',\sigma^2)\right) = \alpha \cdot \frac{(\mu-\mu')^2}{2\sigma^2}.$$

5. ***Pure DP to Concentrated DP:*** *For all $\alpha \in [1,\infty)$,*

$$\mathrm{D}_\alpha\left(P\middle\|Q\right) \leq \frac{\alpha}{8} \cdot \left(\mathrm{D}_\infty\left(P\middle\|Q\right) + \mathrm{D}_\infty\left(Q\middle\|P\right)\right)^2.$$

6. ***Quasi-convexity:*** *Let P' and Q' be probability distributions over $\mathcal{Y}$ such that P' is absolutely continuous with respect to Q'. For $s \in [0,1]$, let $(1-s)\cdot P + s\cdot P'$ denote the convex combination of the distributions P and P' with weighting s. For all $\alpha \in (1,\infty)$ and all $s \in [0,1]$,*

$$\begin{aligned}
&\mathrm{D}_\alpha\left((1-s)\cdot P + s\cdot P'\middle\|(1-s)\cdot Q + s\cdot Q'\right) \\
&\leq \frac{1}{\alpha-1}\log\Big((1-s)\cdot\exp\left((\alpha-1)\mathrm{D}_\alpha\left(P\middle\|Q\right)\right) \\
&\quad + s\cdot\exp\left((\alpha-1)\mathrm{D}_\alpha\left(P'\middle\|Q'\right)\right)\Big) \\
&\leq \max\left\{\mathrm{D}_\alpha\left(P\middle\|Q\right), \mathrm{D}_\alpha\left(P'\middle\|Q'\right)\right\}
\end{aligned}$$

   and $\mathrm{D}_1\left((1-s)\cdot P + s\cdot P'\middle\|(1-s)\cdot Q + s\cdot Q'\right) \leq (1-s)\cdot\mathrm{D}_1\left(P\middle\|Q\right) + s\cdot\mathrm{D}_1\left(P'\middle\|Q'\right)$.
7. ***Triangle-like inequality (a.k.a. group privacy):*** *Let R be a distribution on $\mathcal{Y}$ and assume that Q is absolutely continuous with respect to R. For all $1 < \alpha <$*

$\alpha' < \infty$,

$$\mathrm{D}_\alpha(P\|R) \le \frac{\alpha'}{\alpha'-1} \cdot \mathrm{D}_{\alpha\cdot\frac{\alpha'-1}{\alpha'-\alpha}}(P\|Q) + \mathrm{D}_{\alpha'}(Q\|R).$$

In particular, if $\mathrm{D}_\alpha(P\|Q) \le \rho_1 \cdot \alpha$ *and* $\mathrm{D}_\alpha(Q\|R) \le \rho_2 \cdot \alpha$ *for all* $\alpha \in (1,\infty)$, *then* $\mathrm{D}_\alpha(P\|R) \le (\sqrt{\rho_1}+\sqrt{\rho_2})^2 \cdot \alpha$ *for all* $\alpha \in (1,\infty)$.

8. ***Conversion to approximate DP:*** *For all measurable* $S \subset \mathcal{Y}$, *all* $\alpha \in (1,\infty)$, *and all* $\tilde{\varepsilon} \ge \mathrm{D}_\alpha(P\|Q)$,

$$\begin{aligned} P(S) &\le e^{\tilde{\varepsilon}} \cdot Q(S) + e^{-(\alpha-1)(\tilde{\varepsilon}-\mathrm{D}_\alpha(P\|Q))} \cdot \frac{1}{\alpha} \cdot \left(1-\frac{1}{\alpha}\right)^{\alpha-1} \\ &\le e^{\tilde{\varepsilon}} \cdot Q(S) + e^{-(\alpha-1)(\tilde{\varepsilon}-\mathrm{D}_\alpha(P\|Q))}. \end{aligned}$$

Proof. 1. *Post-processing (a.k.a. data processing inequality) & non-negativity:* See Lemma 3.20. Non-negativity follows from setting f to be a constant function and noting that the divergence between two point masses is zero.

2. *Composition:* See Proposition 3.19.

3. *Monotonicity:* Let $1 < \alpha \le \alpha' < \infty$. (The cases where $\alpha = 1$ and $\alpha' = \infty$ follow from continuity.) Let $f(x) = x^{\frac{\alpha'-1}{\alpha-1}}$. Then f is convex and, by Jensen's inequality,

$$\begin{aligned} e^{(\alpha'-1)\mathrm{D}_\alpha(P\|Q)} &= f\left(\mathop{\mathbb{E}}_{Y\leftarrow P}\left[\left(\frac{P(Y)}{Q(Y)}\right)^{\alpha-1}\right]\right) \\ &\le \mathop{\mathbb{E}}_{Y\leftarrow P}\left[f\left(\left(\frac{P(Y)}{Q(Y)}\right)^{\alpha-1}\right)\right] = e^{(\alpha'-1)\mathrm{D}_{\alpha'}(P\|Q)}, \end{aligned}$$

which implies $\mathrm{D}_\alpha(P\|Q) \le \mathrm{D}_{\alpha'}(P\|Q)$.

4. *Gaussian divergence:* See Lemma 3.12.

5. *Pure DP to Concentrated DP:* See Proposition 3.16.

6. *Quasi-convexity:* See Lemma B.6 of Bun and Steinke [BS16].

7. *Triangle-like inequality (a.k.a. group privacy):* Let $\alpha \in (1,\infty)$. Let $p, q \in (1,\infty)$ satisfy $\frac{1}{p}+\frac{1}{q}=1$. By Hölder's inequality,

$$\begin{aligned} e^{(\alpha-1)\mathrm{D}_\alpha(P\|R)} &= \mathop{\mathbb{E}}_{Y\leftarrow P}\left[\left(\frac{P(Y)}{R(Y)}\right)^{\alpha-1}\right] = \mathop{\mathbb{E}}_{Y\leftarrow R}\left[\left(\frac{P(Y)}{R(Y)}\right)^{\alpha}\right] \\ &= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\frac{P(Y)}{Q(Y)} \cdot \left(\frac{P(Y)}{Q(Y)} \cdot \frac{Q(Y)}{R(Y)}\right)^{\alpha-1}\right] \end{aligned}$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^{\alpha}\cdot\left(\frac{Q(Y)}{R(Y)}\right)^{\alpha-1}\right] \\
&\le \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^{\alpha p}\right]^{1/p}\cdot\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{Q(Y)}{R(Y)}\right)^{(\alpha-1)q}\right]^{1/q} \\
&= e^{\frac{\alpha p-1}{p}\mathrm{D}_{\alpha p}(P\|Q)}\cdot e^{\frac{(\alpha-1)q}{q}\mathrm{D}_{(\alpha-1)q+1}(Q\|R)}.
\end{aligned}
$$

This rearranges to

$$
\begin{aligned}
\mathrm{D}_{\alpha}(P\|R) &\le \frac{\alpha p-1}{(\alpha-1)p}\mathrm{D}_{\alpha p}(P\|Q)+\mathrm{D}_{(\alpha-1)q+1}(Q\|R) \\
&= \left(1+\frac{1}{(\alpha-1)q}\right)\cdot\mathrm{D}_{\alpha p}(P\|Q)+\mathrm{D}_{(\alpha-1)q+1}(Q\|R) \\
&= \frac{\alpha'}{\alpha'-1}\cdot\mathrm{D}_{\alpha\cdot\frac{\alpha'-1}{\alpha'-\alpha}}(P\|Q)+\mathrm{D}_{\alpha'}(Q\|R),
\end{aligned}
$$

where the final equality sets $p=\frac{\alpha'-1}{\alpha'-\alpha}$ and $q=\frac{\alpha'-1}{\alpha-1}$

Now assume $\mathrm{D}_{\alpha}(P\|Q)\le\rho_1\cdot\alpha$ and $\mathrm{D}_{\alpha}(Q\|R)\le\rho_2\cdot\alpha$ for all $\alpha\in(1,\infty)$. Then

$$
\begin{aligned}
\mathrm{D}_{\alpha}(P\|R) &\le \inf_{\alpha'>\alpha}\frac{\alpha'}{\alpha'-1}\cdot\mathrm{D}_{\alpha\cdot\frac{\alpha'-1}{\alpha'-\alpha}}(P\|Q)+\mathrm{D}_{\alpha'}(Q\|R) \\
&\le \inf_{\alpha'>\alpha}\frac{\alpha'}{\alpha'-1}\cdot\alpha\cdot\frac{\alpha'-1}{\alpha'-\alpha}\cdot\rho_1+\alpha'\cdot\rho_2 \\
&= \inf_{x>0}\alpha\cdot\frac{x+1}{x}\cdot\rho_1+\alpha\cdot(x+1)\cdot\rho_2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(Reparameterize } \alpha'=(x+1)\cdot\alpha) \\
&= \alpha\cdot\inf_{x>0}\rho_1+\rho_2+\frac{1}{x}\rho_1+x\cdot\rho_2 \\
&= \alpha\cdot(\rho_1+\rho_2+2\sqrt{\rho_1\cdot\rho_2}) \qquad\qquad (x=\sqrt{\rho_1/\rho_2}) \\
&= \alpha\cdot(\sqrt{\rho_1}+\sqrt{\rho_2})^2.
\end{aligned}
$$

8. *Conversion to approximate DP:* See Proposition 3.14.

□

3.6.5 Sharp Privacy Amplification by Poisson Subsampling for Rényi DP

Now we analyze privacy amplification by subsampling under Rényi DP. We start with a Rényi DP guarantee and we obtain an amplified Rényi DP guarantee. The

goal is to obtain a sharp analysis that avoids converting to approximate DP and back.

For mathematical simplicity, we restrict our attention to Poisson subsampling and assume that neighboring datasets correspond to addition or removal of one person's data.

Theorem 3.33 (Tight Privacy Amplification by Subsampling for Rényi DP)**.** *Let $U \subset [n]$ be a random set that contains each element independently with probability p. For $x \in \mathcal{X}^n$ let $x_U \in \mathcal{X}^n$ be given by $(x_U)_i = x_i$ if $i \in U$ and $(x_U)_i = \perp$ if $i \notin U$, where $\perp \in \mathcal{X}$ is some fixed value.*

Let $\varepsilon : \mathbb{N}_{\geq 2} \to \mathbb{R} \cup \{\infty\}$ be a function. Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy $(\alpha, \varepsilon(\alpha))$-RDP for all $\alpha \in \mathbb{N}_{\geq 2}$ with respect to addition or removal – i.e., $x, x' \in \mathcal{X}^n$ are neighboring if, for some $i \in [n]$, we have $x_i = \perp$ or $x'_i = \perp$, and $\forall j \neq i \;\; x_j = x'_j$.

Define $M^U : \mathcal{X}^n \to \mathcal{Y}$ by $M^U(x) = M(x_U)$. Then M^U satisfies $(\alpha, \varepsilon'_p(\alpha))$-RDP for all $\alpha \in \mathbb{N}_{\geq 2}$ where

$$\varepsilon'_p(\alpha) = \frac{1}{\alpha - 1} \log \Bigg((1-p)^{\alpha-1}(1 + (\alpha - 1)p) + \sum_{k=2}^{\alpha} \binom{\alpha}{k} (1-p)^{\alpha-k} p^k \cdot e^{(k-1)\varepsilon(k)} \Bigg).$$

Note that $(1-p)^{\alpha-1}(1 + (\alpha-1)p) \leq 1$. It is easy to see from the proof that this analysis is tight. That is, if the assumption that M satisfies $(\alpha, \varepsilon(\alpha))$-RDP for all α is tight for some fixed pair of neighboring inputs, then the conclusion that M^U satisfies $(\alpha, \varepsilon'_p(\alpha))$-RDP is also tight.

Theorem 3.33 only considers Rényi DP with orders $\alpha \in \mathbb{N}_{\geq 2} = \{2, 3, 4, \cdots\}$. This restriction arises because the proof uses a binomial expansion, which only works for integer exponents. In certain cases, it is possible to obtain an expression all $\alpha \in (1, \infty)$ using an infinite binomial series [MTZ19]. In general, we can use Monotonicity (part 3 of Lemma 3.32) to bound non-integer α, namely for all $\alpha \in (1, \infty)$, M^U satisfies $(\alpha, \varepsilon'_p(\lceil \alpha \rceil))$-RDP.

Proof of Theorem 3.33. Fix neighboring datasets $x, x' \in \mathcal{X}^n$. Without loss of generality, assume that x' is x with one element removed – i.e., $\exists i \in [n] \;\; (x'_i = \perp) \wedge (\forall j \in [n] \setminus \{i\} \;\; x_j = x'_j)$. Fix this i.

Let $Q = M(x'_U) = M^U(x')$. Let $P = M(x_U)|_{i \in U}$ be the conditional distribution of $M(x_U)$ with $i \in U$. Note that $M(x_U)|_{i \notin U} = Q$ because $x_U = x'_U$ when $i \notin U$ and the event $i \in U$ is independent from $U \setminus \{i\}$. (This is where we use the Poisson subsampling assumption.)

Thus we can express the distribution of $M^U(x)$ as a convex combination: $M(x_U) = p \cdot P + (1-p) \cdot Q$, since $p = \mathbb{P}[i \in U]$.

For all $\alpha \in \mathbb{N}_{\geq 2}$, M is assumed to be $(\alpha, \varepsilon(\alpha))$-RDP, so we have $\mathrm{D}_\alpha(P\|Q) \leq \varepsilon(\alpha)$ and $\mathrm{D}_\alpha(Q\|P) \leq \varepsilon(\alpha)$.

To complete the proof we must show that

$$\mathrm{D}_\alpha\left(p \cdot P + (1-p) \cdot Q \middle\| Q\right) \leq \varepsilon'_p(\alpha)$$

and

$$\mathrm{D}_\alpha\left(Q \middle\| p \cdot P + (1-p) \cdot Q\right) \leq \varepsilon'_p(\alpha)$$

for all $\alpha \in \mathbb{N}_{\geq 2}$.

Fix $\alpha \in \mathbb{N}_{\geq 2}$. We have

$$\begin{aligned}
e^{(\alpha-1)\mathrm{D}_\alpha\left(p\cdot P+(1-p)\cdot Q\|Q\right)} &= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{p\cdot P(Y)+(1-p)\cdot Q(Y)}{Q(Y)}\right)^\alpha\right] \\
&= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(1-p+p\cdot\frac{P(Y)}{Q(Y)}\right)^\alpha\right] \\
&= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\sum_{k=0}^{\alpha}\binom{\alpha}{k}(1-p)^{\alpha-k}p^k\left(\frac{P(Y)}{Q(Y)}\right)^k\right] \\
&\qquad\qquad \text{(Binomial expansion)} \\
&= \sum_{k=0}^{\alpha}\binom{\alpha}{k}(1-p)^{\alpha-k}p^k \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^k\right] \\
&= (1-p)^\alpha + \alpha(1-p)^{\alpha-1}p \\
&\quad + \sum_{k=2}^{\alpha}\binom{\alpha}{k}(1-p)^{\alpha-k}p^k \cdot e^{(k-1)\mathrm{D}_k(P\|Q)} \\
&\qquad\qquad \left(\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\tfrac{P(Y)}{Q(Y)}\right] = 1\right) \\
&\leq (1-p)^{\alpha-1}(1+(\alpha-1)p) \\
&\quad + \sum_{k=2}^{\alpha}\binom{\alpha}{k}(1-p)^{\alpha-k}p^k \cdot e^{(k-1)\varepsilon(k)} \\
&\qquad\qquad (\mathrm{D}_k(P\|Q) \leq \varepsilon(k)) \\
&= e^{(\alpha-1)\varepsilon'_p(\alpha)}.
\end{aligned}$$

Note that $(1-p)^{\alpha-1}(1+(\alpha-1)p) \leq (e^{-p})^{\alpha-1}e^{(\alpha-1)p} = 1$.

An identical calculation shows that

$$\mathrm{D}_\alpha\left(p\cdot Q+(1-p)\cdot P\middle\|P\right)\le\varepsilon'_p(\alpha).$$

Finally, Theorem 3.34 gives

$$\mathrm{D}_\alpha\left(Q\middle\|pP+(1-p)Q\right)\le\max\left\{\begin{array}{c}\mathrm{D}_\alpha\left(pP+(1-p)Q\middle\|Q\right),\\ \mathrm{D}_\alpha\left(pQ+(1-p)P\middle\|P\right)\end{array}\right\}\le\varepsilon'_p(\alpha).$$

□

The following result shows that, in terms of subsampling for Rényi DP, it suffices to analyze one side of the add/remove neighboring relation.

Theorem 3.34. *Let P and Q be probability distributions that are absolutely continuous with respect to each other. Let $p\in[0,1]$ and $\alpha\in(1,\infty)$. Set $\lambda=\frac{(2\alpha-1)p}{(2\alpha-1)p+3(1-p)}$. Then*

$$\begin{aligned}e^{(\alpha-1)\mathrm{D}_\alpha\left(Q\middle\|pP+(1-p)Q\right)}\le{}&(1-\lambda)\cdot e^{(\alpha-1)\mathrm{D}_\alpha\left(pP+(1-p)Q\middle\|Q\right)}\\&+\lambda\cdot e^{(\alpha-1)\mathrm{D}_\alpha\left(pQ+(1-p)P\middle\|P\right)}.\end{aligned}$$

Since $\lambda\in[0,1]$, this implies

$$\mathrm{D}_\alpha\left(Q\middle\|pP+(1-p)Q\right)\le\max\left\{\begin{array}{c}\mathrm{D}_\alpha\left(pP+(1-p)Q\middle\|Q\right),\\ \mathrm{D}_\alpha\left(pQ+(1-p)P\middle\|P\right)\end{array}\right\}.$$

Proof. Define $f:(0,\infty)\to\mathbb{R}$ by

$$f(x)=(1-\lambda)(1-p+p\cdot x)^\alpha+\lambda\cdot x\cdot\left(1-p+\frac{p}{x}\right)^\alpha-\left(1-p+p\cdot x\right)^{1-\alpha}.$$

We have

$$\begin{aligned}&(1-\lambda)\cdot e^{(\alpha-1)\mathrm{D}_\alpha\left(pP+(1-p)Q\middle\|Q\right)}+\lambda\cdot e^{(\alpha-1)\mathrm{D}_\alpha\left(pQ+(1-p)P\middle\|P\right)}\\&\quad-e^{(\alpha-1)\mathrm{D}_\alpha\left(Q\middle\|pP+(1-p)Q\right)}\\&=(1-\lambda)\cdot\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{pP(Y)+(1-p)Q(Y)}{Q(Y)}\right)^\alpha\right]\\&\quad+\lambda\cdot\mathop{\mathbb{E}}_{Y\leftarrow P}\left[\left(\frac{pQ(Y)+(1-p)P(Y)}{P(Y)}\right)^\alpha\right]\\&\quad-\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{Q(Y)}{pP(Y)+(1-p)Q(Y)}\right)^{\alpha-1}\right]\\&=(1-\lambda)\cdot\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(p\frac{P(Y)}{Q(Y)}+(1-p)\right)^\alpha\right]\end{aligned}$$

$$
\begin{aligned}
&+\lambda\cdot\mathop{\mathbb{E}}_{Y\leftarrow P}\left[\left(p\left(\frac{P(Y)}{Q(Y)}\right)^{-1}+1-p\right)^{\alpha}\right]\\
&-\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(p\frac{P(Y)}{Q(Y)}+(1-p)\right)^{1-\alpha}\right]\\
&=(1-\lambda)\cdot\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(p\frac{P(Y)}{Q(Y)}+(1-p)\right)^{\alpha}\right]\\
&+\lambda\cdot\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\frac{P(Y)}{Q(Y)}\cdot\left(p\left(\frac{P(Y)}{Q(Y)}\right)^{-1}+1-p\right)^{\alpha}\right]\\
&-\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(p\frac{P(Y)}{Q(Y)}+(1-p)\right)^{1-\alpha}\right]\\
&\qquad\qquad\text{(For any } g,\ \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\tfrac{P(Y)}{Q(Y)}g(Y)\right]=\mathop{\mathbb{E}}_{Y\leftarrow P}\left[g(Y)\right])\\
&=\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[(1-\lambda)\cdot\left(1-p+p\cdot\frac{P(Y)}{Q(Y)}\right)^{\alpha}\right.\\
&\left.+\lambda\cdot\frac{P(Y)}{Q(Y)}\cdot\left(1-p+\frac{p}{\frac{P(Y)}{Q(Y)}}\right)^{\alpha}-\left(1-p+p\cdot\frac{P(Y)}{Q(Y)}\right)^{1-\alpha}\right]\\
&=\mathop{\mathbb{E}}_{X}\left[f(X)\right],
\end{aligned}
$$

where $X=\frac{P(Y)}{Q(Y)}$ for $Y\leftarrow Q$. Thus our objective is to show that $\mathop{\mathbb{E}}_{X}\left[f(X)\right]\geq 0$.[xi]

We claim that f is convex. Convexity implies $f(x)\geq f(1)+f'(1)\cdot(x-1)$ for all $x\in(0,\infty)$. Since $f(1)=0$ and $\mathbb{E}[X]=\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\frac{P(Y)}{Q(Y)}\right]=1$, this implies $\mathbb{E}\left[f(X)\right]\geq f(1)+f'(1)\mathbb{E}[X-1]=0$, as required.

It only remains to prove that f is convex. We have, for all $x>0$,

$$
\begin{aligned}
f(x)&=(1-\lambda)(1-p+p\cdot x)^{\alpha}+\lambda\cdot x\cdot\left(1-p+\frac{p}{x}\right)^{\alpha}-\left(1-p+p\cdot x\right)^{1-\alpha},\\
f'(x)&=(1-\lambda)\alpha p(1-p+p\cdot x)^{\alpha-1}+\lambda\cdot\left(1-p+\frac{p}{x}\right)^{\alpha}\\
&\quad-\lambda\alpha\frac{p}{x}\left(1-p+\frac{p}{x}\right)^{\alpha-1}+(\alpha-1)p\left(1-p+p\cdot x\right)^{-\alpha}\\
f''(x)&=(1-\lambda)\alpha(\alpha-1)p^{2}(1-p+p\cdot x)^{\alpha-2}
\end{aligned}
$$

xi. Note that we assume P and Q are absolutely continuous with respect to each other – i.e., $\forall S\ P(S)=0\iff Q(S)=0$. This ensures that the Radon-Nikodym derivative $\frac{P(y)}{Q(y)}$ is well-defined and, further that $\mathop{\mathbb{P}}_{Y\leftarrow Q}\left[\frac{P(Y)}{Q(Y)}\leq 0\right]=0$. Thus the function f need only be defined on $(0,\infty)$.

$$
\begin{aligned}
&- \lambda\alpha \frac{p}{x^2}\left(1-p+\frac{p}{x}\right)^{\alpha-1} + \lambda\alpha\frac{p}{x^2}\left(1-p+\frac{p}{x}\right)^{\alpha-1} \\
&+ \lambda\alpha(\alpha-1)\frac{p^2}{x^3}\left(1-p+\frac{p}{x}\right)^{\alpha-2} - \alpha(\alpha-1)p^2\left(1-p+p\cdot x\right)^{-\alpha-1} \\
=\ & \alpha(\alpha-1)p^2\Bigg((1-\lambda)(1-p+px)^{\alpha-2} + \lambda\frac{1}{x^3}\left(1-p+\frac{p}{x}\right)^{\alpha-2} \\
&-(1-p+px)^{-\alpha-1}\Bigg) \\
=\ & \frac{\alpha(\alpha-1)p^2}{(1-p+px)^{\alpha+1}}\Bigg((1-\lambda)(1-p+px)^{2\alpha-1} + \lambda\frac{1}{x^3}\left(1-p+\frac{p}{x}\right)^{\alpha-2} \\
&\cdot(1-p+px)^{\alpha+1} - 1\Bigg) \\
=\ & \frac{\alpha(\alpha-1)p^2}{(1-p+px)^{\alpha+1}}\Bigg((1-\lambda)(1-p+px)^{2\alpha-1} + \lambda\left(\frac{1-p+px}{x}\right)^3 \\
&\cdot\left(1-p+\frac{p}{x}\right)^{\alpha-2}\cdot(1-p+px)^{\alpha-2} - 1\Bigg) \\
\geq\ & \frac{\alpha(\alpha-1)p^2}{(1-p+px)^{\alpha+1}} \\
&\times\left((1-\lambda)(1-p+px)^{2\alpha-1} + \lambda\left(\frac{1-p+px}{x}\right)^3\cdot 1 - 1\right) \\
&&\text{(Lemma 3.35)} \\
=\ & \frac{\alpha(\alpha-1)p^2}{(1-p+px)^{\alpha+1}}\left(\frac{\begin{array}{c} 3(1-p)(1-p+px)^{2\alpha-1} \\ +(2\alpha-1)p\left(\frac{1-p}{x}+p\right)^3 \\ -3(1-p)-(2\alpha-1)p \end{array}}{3(1-p)+(2\alpha-1)p}\right) \\
&&\left(\lambda = \tfrac{(2\alpha-1)p}{(2\alpha-1)p+3(1-p)}\right) \\
\geq\ & 0. && \text{(Lemma 3.36)}
\end{aligned}
$$

□

We now give the auxiliary lemmata used in the proof of Theorem 3.34.

Lemma 3.35. *For all* $p \in [0,1]$ *and* $x \in (0,\infty)$*,*

$$
\frac{1}{1-p+p/x} \leq 1-p+p\cdot x.
$$

Proof. Let $f(t) = t + 1/t$. Then $f'(t) = 1 - 1/t^2$ and $f''(t) = 2/t^3 > 0$ for all $t > 0$. Thus $f'(t) = 0 \iff t = 1$ and $f(x) \geq f(1) = 2$. Now

$$\begin{aligned}(1 - p + p \cdot x) \cdot (1 - p + p/x) &= p^2 + (1-p)^2 + p(1-p)(x + 1/x) \\ &\geq p^2 + (1-p)^2 + p(1-p) \cdot 2 \\ &= (p + (1-p))^2 = 1.\end{aligned}$$

Rearranging yields the result. □

Lemma 3.36. *For all $v \geq 1$, $p \in [0,1]$, and $x \in (0,\infty)$,*

$$3(1-p)(1-p+px)^v + vp\left(\frac{1-p}{x} + p\right)^3 \geq 3(1-p) + vp.$$

Proof. Define $f : (0,\infty) \to \mathbb{R}$ by

$$f(x) = 3(1-p)(1-p+px)^v + vp\left(\frac{1-p}{x} + p\right)^3.$$

Our goal is to prove that $f(x) \geq f(1) = 3(1-p) + vp$ for all $x \in (0,\infty)$. It suffices to prove that f is convex and that $f'(1) = 0$. We have

$$f'(x) = 3vp(1-p)\left((1-p+px)^{v-1} - \frac{1}{x^2}\left(\frac{1-p}{x} + p\right)^2\right),$$

$$\begin{aligned}f''(x) = 3vp(1-p)\Bigg((v-1)p(1-p+px)^{v-2} + \frac{2}{x^3} \cdot \left(\frac{1-p}{x} + p\right)^2 \\ + \frac{2}{x^2} \cdot \frac{1-p}{x^2} \cdot \left(\frac{1-p}{x} + p\right)\Bigg).\end{aligned}$$

From these expressions, it is easy to see that $f'(1) = 0$ and $f''(x) \geq 0$ for all $x \in (0,\infty)$. □

3.6.6 Analytic Rényi DP Bound for Privacy Amplification by Poisson Subsampling

Theorem 3.33 gives a tight RDP bound for privacy amplification by Poisson subsampling. However, the bound is in the form of a series. This is adequate for numerical purposes, but it is helpful for our understanding to have a simpler closed-form expression.

In this subsection we will provide a simpler expression and attempt to build some understanding of how privacy amplification by subsampling applies to Rényi DP.

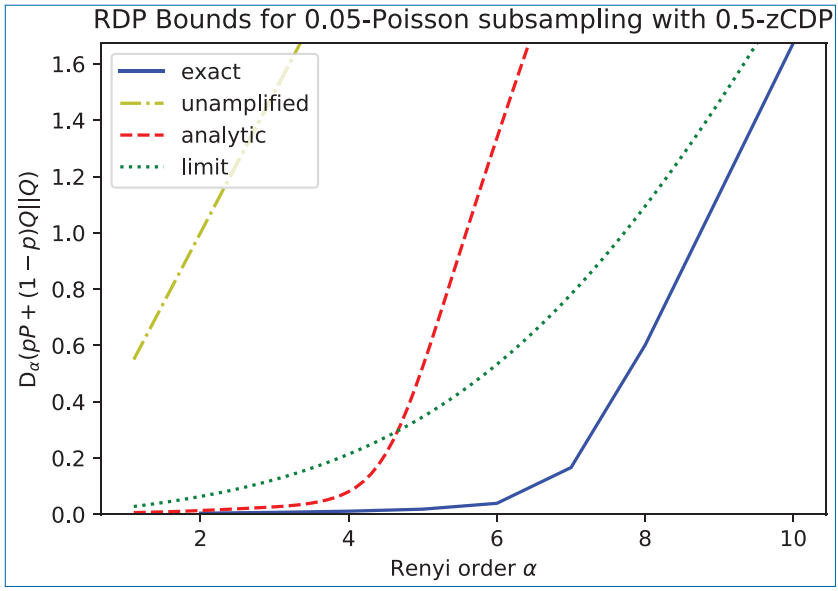

Figure 3.2. Comparison of Rényi divergence guarantees for Poisson subsampling – i.e., including each person with probability $p = 0.05$. The *unamplified* algorithm satisfies 0.5-zCDP. The *exact* bound is given by Theorem 3.33. For comparison, we have the *analytic* upper bound from Proposition 3.40 as well as the behavior in the *limit* given by Proposition 3.41.

Theorem 3.37 (Asymptotic Privacy Amplification by Subsampling for Rényi DP)**.** *Let* $p \in [0, 1/2]$ *and* $\rho \in (0, 1]$. *Define* $\omega = \min\left\{\frac{\log(1/p)}{4\rho}, 1 + p^{-1/4}\right\}$. *Assume* $\omega \geq 3 + 2\frac{\log(1/\rho)}{\log(1/p)}$.

Let $M : \mathcal{X}^n \to \mathcal{Y}$ *satisfy* ρ*-zCDP with respect to addition or removal.*[xii]

Define $M^U : \mathcal{X}^n \to \mathcal{Y}$ *by* $M^U(x) = M(x_U)$, *where* $U \subset [n]$ *be a random set that contains each element independently with probability p and, for* $x \in \mathcal{X}^n$, $x_U \in \mathcal{X}^n$ *is given by* $(x_U)_i = x_i$ *if* $i \in U$ *and* $(x_U)_i = \perp$ *if* $i \notin U$.

Then M^U *satisfies* $(\alpha, 10p^2\rho\alpha)$*-RDP for all* $\alpha \in (1, \omega)$.

There are many caveats in the statement of Theorem 3.37, but the high level message is that Poisson subsampling a p fraction of the dataset amplifies ρ-zCDP to something like $O(p^2 \cdot \rho)$-zCDP. We will discuss these caveats in a moment, but, ignoring these caveats and the constant factor in the guarantee, this is exactly the kind of guarantee we would hope for.

xii. I.e., $x, x' \in \mathcal{X}^n$ are neighboring if, for some $i \in [n]$, we have $x_i = \perp$ or $x'_i = \perp$, and $\forall j \neq i$ $x_j = x'_j$, where $\perp \in \mathcal{X}$ is some fixed value.

Consider the following example, which illustrates what kind guarantee we would hope for. Suppose we have a query $q : \mathcal{X} \to [0,1]$ and a sensitive dataset $x \in \mathcal{X}^n$ and our goal is to estimate $q(x) := \frac{1}{n}\sum_i^n q(x_i)$. We could release a sample from $\mathcal{N}(q(x), \sigma^2)$, which satisfies $\frac{1}{2n^2\sigma^2}$-zCDP and has mean squared error σ^2. However, perhaps due to computational constraints, we might instead select a random p fraction $U \subset [n]$ and instead release a sample from $M^U(x) = \mathcal{N}\left(\frac{1}{pn}\sum_{i\in U} q(x_i), \sigma^2\right)$. We can calculate that the mean squared error of this algorithm is at most $\sigma^2 + \frac{1-p}{pn}$. Without amplification this satisfies $(\rho = \frac{1}{2p^2n^2\sigma^2})$-zCDP. With amplification, Theorem 3.37 tells us that this satisfies $(\alpha, O(p^2\cdot\rho))$-RDP for α not too large. Now $p^2\cdot\rho = \frac{1}{2n^2\sigma^2}$ is exactly the guarantee that we obtained by simply evaluating q on the entire dataset and avoiding subsampling. We cannot hope to do better than this.

The constant factor of 10 in the theorem can be improved, but a constant factor loss is the price we pay for having a simpler expression; if we want tight constants we should apply Theorem 3.33.

The main caveat in Theorem 3.37 is the requirement that $\alpha \le \omega \le \frac{\log(1/p)}{4\rho}$. This is necessary, as the $(\alpha, \varepsilon(\alpha))$-RDP guarantee qualitatively changes when $\alpha \ge O(\log(1/p)/\rho)$. It changes from $\varepsilon(\alpha) = O(p^2\rho\alpha)$ to $\varepsilon(\alpha) = \rho\alpha - O(\log(1/p))$. To see why this is inherent, consider the following lower bound. For all $p \in [0,1]$, all $\alpha \in (1,\infty)$, and all absolutely continuous probability distributions P and Q, we have

$$\begin{aligned} e^{(\alpha-1)\mathrm{D}_\alpha\left(pP+(1-p)Q\middle\|Q\right)} &= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(1-p+p\cdot\frac{P(Y)}{Q(Y)}\right)^\alpha\right] \\ &\ge \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(p\cdot\frac{P(Y)}{Q(Y)}\right)^\alpha\right] = p^\alpha\cdot e^{(\alpha-1)\mathrm{D}_\alpha(P\|Q)}. \end{aligned}$$

Thus $\mathrm{D}_\alpha\left(pP+(1-p)Q\middle\|Q\right) \ge \mathrm{D}_\alpha(P\|Q) - \frac{\alpha}{\alpha-1}\log(1/p)$. This tells us that, for large α, we cannot have more than an additive improvement in the RDP guarantee, whereas for small α we have a multiplicative improvement. Proposition 3.41 shows that this lower bound is tight.

We now proceed to prove Theorem 3.37.

Lemma 3.38. *Let $\alpha \in (1,\infty)$, $p \in \left[0, 1-e^{-1}\right]$, and $x \in [0,\infty)$. If either $\alpha \le 2$ or $\alpha > 2$ and $px \le \max\left\{p, \frac{1-p}{\alpha-2}\right\}$, then*

$$(1-p+p\cdot x)^\alpha \le 1 + \alpha p(x-1) + \frac{e}{2}\alpha(\alpha-1)p^2(x-1)^2.$$

Proof. We assume $p > 0$. Otherwise the result is trivial.

Define $f : [0, \infty) \to \mathbb{R}$ by

$$f(x) = (1 - p + px)^\alpha.$$

For all $x \in [0, \infty)$, we have

$$\begin{aligned} f'(x) &= \alpha p(1 - p + px)^{\alpha - 1}, \\ f''(x) &= \alpha(\alpha - 1)p^2(1 - p + px)^{\alpha - 2}, \\ f'''(x) &= \alpha(\alpha - 1)(\alpha - 2)p^3(1 - p + px)^{\alpha - 3}. \end{aligned}$$

By Taylor's theorem, for all $x \in [0, \infty)$, there exists $\xi_x \in [\min\{1, x\}, \max\{1, x\}]$ such that

$$\begin{aligned} f(x) &= f(1) + f'(1)(x - 1) + \frac{1}{2} f''(\xi_x)(x - 1)^2 \\ &= 1 + \alpha p(x - 1) + \frac{1}{2} f''(\xi_x)(x - 1)^2. \end{aligned}$$

To complete the proof is suffices to show that $f''(\xi) \le e \cdot \alpha(\alpha - 1)p^2$ in two cases: First, for all $\xi \in [0, \infty)$ assuming $\alpha \le 2$. Second, for all $\xi \in \left[0, \max\left\{1, \frac{1-p}{p}\frac{1}{\alpha-2}\right\}\right]$ assuming $\alpha > 2$. (Note that, $\xi_x \in \left[0, \max\left\{1, \frac{1-p}{p}\frac{1}{\alpha-2}\right\}\right]$ is implied by the assumptions $px \le \max\left\{p, \frac{1-p}{\alpha-2}\right\}$ and $p > 0$.)

First, suppose $\alpha \le 2$. Then $f'''(x) \le 0$ for all $x \in [0, \infty)$. Thus f'' is decreasing (or constant) and, for all $\xi \in [0, \infty)$, we have

$$\begin{aligned} f''(\xi) &\le f''(0) \\ &= \alpha(\alpha - 1)p^2(1 - p)^{\alpha - 2} \\ &\le \alpha(\alpha - 1)p^2 \frac{1}{1 - p} && (\alpha > 1) \\ &\le \alpha(\alpha - 1)p^2 \cdot e. && (p \le 1 - e^{-1}) \end{aligned}$$

Second, assume $\alpha > 2$ and $px \le \max\left\{p, \frac{1-p}{\alpha-2}\right\}$, which implies $\xi_x \le \max\left\{1, \frac{1-p}{p}\frac{1}{\alpha-2}\right\}$.

We have $f'''(x) > 0$ for all $x \in [0, \infty)$. Thus f'' is increasing and, for all $\xi \in \left[0, \max\left\{1, \frac{1-p}{p}\frac{1}{\alpha-2}\right\}\right]$, we have

$$\begin{aligned} f''(\xi) &\le f''\left(\max\left\{1, \frac{1 - p}{p}\frac{1}{\alpha - 2}\right\}\right) \\ &= \alpha(\alpha - 1)p^2\left(1 - p + p \cdot \max\left\{1, \frac{1 - p}{p}\frac{1}{\alpha - 2}\right\}\right)^{\alpha - 2} \end{aligned}$$

$$= \alpha(\alpha-1)p^2 \max\left\{1, (1-p)^{\alpha-2} \cdot \left(1+\frac{1}{\alpha-2}\right)^{\alpha-2}\right\}$$
$$\leq \alpha(\alpha-1)p^2 \max\left\{1, (1-p)^{0} \cdot \left(e^{\frac{1}{\alpha-2}}\right)^{\alpha-2}\right\}$$
$$= \alpha(\alpha-1)p^2 \cdot e.$$

□

Lemma 3.39. *Let $\alpha, \omega \in (1,\infty)$ with $\alpha \leq \omega$, $p \in \left[0, 1-e^{-1}\right]$, and $x \in [0,\infty)$. Then*

$$(1-p+p\cdot x)^{\alpha} \leq 1 + \alpha p(x-1) + \frac{e}{2}\alpha(\alpha-1)p^2(x-1)^2 + \left((\alpha-1)px\right)^{\omega}.$$

Proof. We can assume $p > 0$, as otherwise the result is trivial.

If $\alpha \leq 2$ or if $\alpha > 2$ and $x \leq \max\left\{1, \frac{1-p}{p}\frac{1}{\alpha-2}\right\}$, then the result follows from Lemma 3.38, as $\left((\alpha-1)px\right)^{\omega} \geq 0$.

Thus we assume $\alpha > 2$ and $x \geq \max\left\{1, \frac{1-p}{p}\frac{1}{\alpha-2}\right\}$.

Since $x \geq 1$, we have $\alpha p(x-1) + \frac{e}{2}\alpha(\alpha-1)p^2(x-1)^2 \geq 0$. Therefore it suffices to prove that $(1-p+px)^{\alpha} \leq ((\alpha-1)px)^{\omega}$.

The assumption $x \geq 1$ implies $1-p+px \geq 1$ and, hence, that $(1-p+px)^{\alpha} \leq (1-p+px)^{\omega}$, as we have $\alpha \leq \omega$. The assumption $x \geq \frac{1-p}{p}\frac{1}{\alpha-2}$ rearranges to $1-p \leq px(\alpha-2)$, which implies $1-p+px \leq (\alpha-1)px$ and, hence, $(1-p+px)^{\omega} \leq ((\alpha-1)px)^{\omega}$, as required. □

Proposition 3.40 (Analytic Privacy Amplification by Subsampling for Rényi Divergence)**.** *Let P and Q be probability distributions with P absolutely continuous with respect to Q. Let $p \in \left[0, 1-e^{-1}\right]$ and $\alpha, \omega \in (1,\infty)$ with $\alpha \leq \omega$. Then*

$$\begin{aligned}
&\mathrm{D}_{\alpha}\left(pP + (1-p)Q \middle\| Q\right) \\
&\quad \leq \frac{1}{\alpha-1}\log\left(1 + \frac{e}{2}\alpha(\alpha-1)p^2\left(e^{\mathrm{D}_2(P\|Q)}-1\right)\right. \\
&\qquad \left. + ((\alpha-1)p)^{\omega} \cdot e^{(\omega-1)\mathrm{D}_{\omega}(P\|Q)}\right) \\
&\quad \leq \alpha \cdot \frac{e}{2} \cdot p^2 \cdot \left(e^{\mathrm{D}_2(P\|Q)}-1\right) + p \cdot \left((\alpha-1)\cdot p \cdot e^{\mathrm{D}_{\omega}(P\|Q)}\right)^{\omega-1}.
\end{aligned}$$

Proof. We have

$$e^{(\alpha-1)\mathrm{D}_{\alpha}\left(pP+(1-p)Q\middle\|Q\right)}$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(\frac{p\cdot P(Y)+(1-p)\cdot Q(Y)}{Q(Y)}\right)^{\alpha}\right] \\
&= \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(1-p+p\cdot\frac{P(Y)}{Q(Y)}\right)^{\alpha}\right] \\
&\le \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[1+\alpha p\left(\frac{P(Y)}{Q(Y)}-1\right)+\frac{e}{2}\alpha(\alpha-1)p^2\left(\frac{P(Y)}{Q(Y)}-1\right)^2\right. \\
&\quad \left.+\left((\alpha-1)p\frac{P(Y)}{Q(Y)}\right)^{\omega}\right] \qquad \text{(Lemma 3.39)} \\
&= 1+\alpha p\,(1-1)+\frac{e}{2}\alpha(\alpha-1)p^2\left(e^{\mathrm{D}_2(P\|Q)}-1\right) \\
&\quad +\left((\alpha-1)p\right)^{\omega}\cdot e^{(\omega-1)\mathrm{D}_{\omega}(P\|Q)}.
\end{aligned}
$$

The second inequality in the result follows from the fact that $\log(1+u)\le u$ for all $u>-1$. □

We also have the following simpler result that provides better bounds when the Rényi order α is large.

Proposition 3.41. *Let P and Q be probability distributions with P absolutely continuous with respect to Q. Let $p\in[0,1]$ and $\alpha\in(1,\infty)$. Then*

$$
\begin{aligned}
&\mathrm{D}_{\alpha}\left(pP+(1-p)Q\middle\|Q\right) \\
&\quad\le \frac{\alpha}{\alpha-1}\log\left(1-p+p\cdot e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right) \\
&\quad= \mathrm{D}_{\alpha}\left(P\|Q\right)-\frac{\alpha}{\alpha-1}\log(1/p)+\frac{\alpha}{\alpha-1}\log\left(1+\frac{1-p}{p}\cdot e^{-\frac{\alpha-1}{\alpha}\mathrm{D}_{\alpha}(P\|Q)}\right) \\
&\quad\le \mathrm{D}_{\alpha}\left(P\|Q\right)-\frac{\alpha}{\alpha-1}\log(1/p)+\frac{\alpha}{\alpha-1}\cdot\frac{1-p}{p}\cdot e^{-\frac{\alpha-1}{\alpha}\mathrm{D}_{\alpha}(P\|Q)}
\end{aligned}
$$

Proof. We assume $0<p<1$, as the result is immediate otherwise. By Jensen's inequality and the convexity of $v\mapsto v^{\alpha}$, for all $x\in[0,\infty)$ and all $\lambda\in(0,1)$,

$$
(1-p+px)^{\alpha}=\left((1-\lambda)\cdot\frac{1-p}{1-\lambda}+\lambda\cdot\frac{px}{\lambda}\right)^{\alpha}\le(1-\lambda)\cdot\left(\frac{1-p}{1-\lambda}\right)^{\alpha}+\lambda\cdot\left(\frac{px}{\lambda}\right)^{\alpha}.
$$

Now, for all $\lambda\in(0,1)$, we have

$$
e^{(\alpha-1)\mathrm{D}_{\alpha}\left(pP+(1-p)Q\|Q\right)}=\mathop{\mathbb{E}}_{Y\leftarrow Q}\left[\left(1-p+p\frac{P(Y)}{Q(Y)}\right)^{\alpha}\right]
$$

$$\leq \mathop{\mathbb{E}}_{Y\leftarrow Q}\left[(1-\lambda)\cdot\left(\frac{1-p}{1-\lambda}\right)^{\alpha}+\lambda\cdot\left(\frac{p}{\lambda}\cdot\frac{P(Y)}{Q(Y)}\right)^{\alpha}\right]$$
$$= (1-\lambda)^{1-\alpha}\cdot(1-p)^{\alpha}+\lambda^{1-\alpha}\cdot p^{\alpha}\cdot e^{(\alpha-1)\mathrm{D}_{\alpha}(P\|Q)}.$$

We can choose λ to minimize this expression. It turns out to be optimal to set $\lambda = \frac{1}{1+\frac{1-p}{p}\cdot e^{-(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}}$. Now we have

$$\begin{aligned}
&e^{(\alpha-1)\mathrm{D}_{\alpha}\left(pP+(1-p)Q\|Q\right)}\\
&\leq (1-\lambda)^{1-\alpha}\cdot(1-p)^{\alpha}+\lambda^{1-\alpha}\cdot p^{\alpha}\cdot e^{(\alpha-1)\mathrm{D}_{\alpha}(P\|Q)}\\
&= \left(1+\frac{p}{1-p}e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right)^{\alpha-1}\cdot(1-p)^{\alpha}\\
&\quad+\left(1+\frac{1-p}{p}\cdot e^{-(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right)^{\alpha-1}\cdot p^{\alpha}\cdot e^{(\alpha-1)\mathrm{D}_{\alpha}(P\|Q)}\\
&= \left(1-p+p\cdot e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right)^{\alpha-1}\cdot(1-p)\\
&\quad+\left(p+(1-p)\cdot e^{-(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right)^{\alpha-1}\cdot p\cdot e^{(\alpha-1)\mathrm{D}_{\alpha}(P\|Q)}\\
&= \left(1-p+p\cdot e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right)^{\alpha-1}\cdot(1-p)\\
&\quad+\left(p\cdot e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}+(1-p)\right)^{\alpha-1}\cdot p\cdot e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\\
&= \left(1-p+p\cdot e^{(1-1/\alpha)\mathrm{D}_{\alpha}(P\|Q)}\right)^{\alpha}.
\end{aligned}$$

Rearranging yields the result. □

Proof of Theorem 3.37. Fix neighboring inputs $x, x' \in \mathcal{X}^n$. Fix some $\alpha \in (1,\omega)$ with $\omega = \min\left\{\frac{\log(1/p)}{4\rho}, 1+p^{-1/4}\right\} \geq 3+2\frac{\log(1/\rho)}{\log(1/p)}$.

Without loss of generality x' is x with some element removed. That is, we can fix some $i \in [n]$ such that $x'_i = \perp$ and $x'_j = x_j$ for all $j \neq i$.

Let $P = M(x_U)|_{i\in U}$ and let $Q = M(x_U)|_{i\notin U}$. Then $M^U(x) = M(x_U) = pP + (1-p)Q$. Also $M(x') = Q$

Thus we must prove that $\mathrm{D}_{\alpha}\left(pP+(1-p)Q\middle\|Q\right) \leq 10p^2\rho\alpha$ and $\mathrm{D}_{\alpha}\left(Q\middle\|pP+(1-p)Q\right) \leq 10p^2\rho\alpha$. Since M is assumed to be ρ-zCDP, we have $\mathrm{D}_{\alpha'}(P\|Q) \leq \rho\alpha'$ and $\mathrm{D}_{\alpha'}(Q\|P) \leq \rho\alpha'$ for all $\alpha' \in (1,\infty)$.

By Proposition 3.40,

$$\begin{aligned}
\mathrm{D}_{\alpha}\left(pP+(1-p)Q\middle\|Q\right) \leq \alpha\cdot\frac{e}{2}\cdot p^2\cdot\left(e^{\mathrm{D}_2(P\|Q)}-1\right)\\
+p\cdot\left((\alpha-1)\cdot p\cdot e^{\mathrm{D}_{\omega}(P\|Q)}\right)^{\omega-1}
\end{aligned}$$

$$
\begin{aligned}
&\le \alpha \cdot \frac{e}{2} \cdot p^2 \cdot \left(e^{2\rho} - 1\right) + p \cdot \left((\alpha - 1) \cdot p \cdot e^{\omega\rho}\right)^{\omega - 1} \\
&\le \alpha \cdot \frac{e}{2} \cdot p^2 \cdot \left(e^{2\rho} - 1\right) + p \cdot \left(p^{-1/4} \cdot p \cdot p^{-1/4}\right)^{\omega - 1} \\
&\qquad\qquad (\alpha \le \omega = \min\{1 + p^{-1/4}, \log(1/p)/4\rho\}) \\
&= \alpha \cdot \frac{e}{2} \cdot p^2 \cdot \left(e^{2\rho} - 1\right) + p^{\frac{1+\omega}{2}} \\
&\le \alpha \cdot \frac{e}{2} \cdot p^2 \cdot \left(e^{2\rho} - 1\right) + p^2 \cdot \rho \\
&\qquad\qquad (\omega \ge 3 + 2\log(1/\rho)/\log(1/p)) \\
&= \alpha \cdot p^2 \cdot \rho \cdot \left(\frac{e}{2} \cdot \frac{e^{2\rho} - 1}{\rho} + \frac{1}{\alpha}\right) \\
&\le \alpha \cdot p^2 \cdot \rho \cdot 10. \qquad\qquad (\rho \in (0,1) \text{ and } \alpha \in (1, \omega))
\end{aligned}
$$

Symmetrically, we have $\mathrm{D}_\alpha\left(pQ + (1-p)P \middle\| P\right) \le \alpha \cdot p^2 \cdot \rho \cdot 10$. By Theorem 3.34,

$$
\mathrm{D}_\alpha\left(Q \middle\| pP + (1-p)Q\right) \le \max\left\{\begin{matrix}\mathrm{D}_\alpha\left(pP + (1-p)Q \middle\| Q\right), \\ \mathrm{D}_\alpha\left(pQ + (1-p)P \middle\| P\right)\end{matrix}\right\} \le \alpha \cdot p^2 \cdot \rho \cdot 10.
$$

□

3.6.7 How to Use Privacy Amplification by Subsampling

The most common use case for privacy amplification by subsampling is analyzing noisy stochastic gradient descent. That is, we repeatedly sample a small subset of the data, compute a function on this subset, and add Gaussian noise. To be precise, let $x \in \mathcal{X}^n$ be the private input. Iteratively, for $t = 1, \cdots, T$, we pick some function $q_t : \mathcal{X}^n \to \mathbb{R}^d$ and randomly sample a subset $U_t \subset [n]$; then we reveal $\mathcal{N}(q_t(x_{U_t}), \sigma^2 I_d)$.

The addition of Gaussian noise satisfies concentrated DP. Specifically, Lemma 3.12 shows that releasing $\mathcal{N}(q_t(x), \sigma^2 I_d)$ satisfies $\frac{\Delta_2^2}{2\sigma^2}$-zCDP, where $\Delta_2 = \sup_{\substack{x,x' \in \mathcal{X}^n \\ \text{neighboring}}} \|q_t(x) - q_t(x')\|_2$ is the sensitivity of q_t. We can thus apply Theorem 3.33 to obtain a tight Rényi DP guarantee for $\mathcal{N}(q_t(x_{U_t}), \sigma^2 I_d)$, where U_t is a Poisson sample. Finally, we can apply the composition property of Rényi DP (Lemma 3.30) over the T rounds and we can convert this final Rényi DP guarantee to approximate DP using Proposition 3.14. This is how differentially private deep learning is analyzed in practice by libraries such as TensorFlow Privacy [Goo18; McM+18].

We can also obtain an asymptotic analysis: Theorem 3.37 shows that $\mathcal{N}(q_t(x_{U_t}), \sigma^2 I_d)$ with $U_t \subset [n]$ including each element independently with probability p satisfies $\left(\alpha, 5\alpha p^2 \Delta_2^2/\sigma^2\right)$-RDP for all $\alpha \in (1, \omega)$. Composition over T rounds yields $\left(\alpha, 5\alpha T p^2 \Delta_2^2/\sigma^2\right)$-RDP for all $\alpha \in (1, \omega)$, which implies (ε, δ)-DP for all $\delta > 0$ and

$$\varepsilon = O\left(\frac{\Delta_2}{\sigma} \cdot p \cdot \sqrt{T \cdot \log(1/\delta)}\right).$$

This bound is directly comparable to the bound from Section 3.6.3, which was derived by converting back and forth between concentrated DP and approximate DP. The difference is that here we have a $\sqrt{\log(1/\delta)}$ whereas there we had a $\log(T/\delta)$ term. This is the asymptotic improvement obtained by keeping the analysis within RDP. This asymptotic improvement also translates into a significant improvement in practice.

We have only analyzed Poisson subsampling, where the size of the sample is random. (Specifically, it follows a binomial distribution.) Naturally, other subsampling schemes may arise in practice. In particular, a fixed size sample is common. As discussed in Section 3.6.2, this corresponds to neighboring datasets allowing the replacement of one individual's data, rather than addition or removal. In terms of Rényi divergences, we must analyze $\mathrm{D}_\alpha\left(pP + (1-p)Q \middle\| pP' + (1-p)Q\right)$, whereas addition and removal correspond to $\mathrm{D}_\alpha\left(pP + (1-p)Q \middle\| Q\right)$ and $\mathrm{D}_\alpha\left(Q \middle\| pP' + (1-p)Q\right)$. However, we can apply group privacy (part 7 of Lemma 3.32) to reduce the analysis to the case we have already analyzed: For all $\alpha' > \alpha$, we have

$$\begin{aligned}&\mathrm{D}_\alpha\left(pP+(1-p)Q \middle\| pP'+(1-p)Q\right)\\ &\quad\le \frac{\alpha'}{\alpha'-1} \cdot \mathrm{D}_{\alpha \cdot \frac{\alpha'-1}{\alpha'-\alpha}}\left(pP+(1-p)Q \middle\| Q\right) + \mathrm{D}_{\alpha'}\left(Q \middle\| pP'+(1-p)Q\right).\end{aligned}$$

3.7 Concluding Remarks

Composition

Differential privacy (specifically, pure DP) was introduced by Dwork, McSherry, Nissim, and Smith [DMNS06].[xiii] The original paper gives a form of basic composition (Theorem 3.1), but does not state it in full generality; rather it states a result specific to Laplace noise addition. Approximate DP was introduced by Dwork,

xiii. The name "differential privacy" does not appear in the original paper. It is attributed to Michael Schroeder [DMNS17] and first appeared in a talk by Dwork [Dwo06].

Kenthapadi, McSherry, Mironov, and Naor [Dwo+06] and this work gave a more general statement of the basic composition result, as well as an analysis of the Gaussian mechanism (although not as tight as Corollary 3.8). The tight analysis of the Gaussian mechanism (Corollaries 3.8 & 3.10) is due to Balle and Wang [BW18].

The advanced composition theorem (Theorem 3.22) was proved by Dwork, Rothblum, and Vadhan [DRV10].[xiv] The key concepts of privacy loss distributions and concentrated DP were implicit in this proof, but they were only made explicit in a separate paper by Dwork and Rothblum [DR16]. Bun and Steinke [BS16] refined the notion of concentrated DP and presented the definition that we use here (Definition 3.11).

Kairouz, Oh, and Viswanathan [KOV15] proved an optimal composition theorem for approximate differential privacy. Specifically, the k-fold composition of (ε, δ)-differential privacy satisfies (ε', δ')-differential privacy if and only if

$$\frac{1}{(1+e^{\varepsilon})^k} \sum_{\ell=0}^{k} \binom{k}{\ell} \cdot e^{\ell\varepsilon} \cdot \max\left\{0, 1 - e^{\varepsilon' - (2\ell - k)\varepsilon}\right\} \leq 1 - \frac{1-\delta'}{(1-\delta)^k}.$$

This expression is rather complex, but the proof is relatively intuitive. The key insight is that we can reduce the analysis to the k-fold composition of a specific worst-case (ε, δ)-DP mechanism. With probability δ, this mechanism has infinite privacy loss. With probability $(1-\delta) \cdot \frac{e^{\varepsilon}}{1+e^{\varepsilon}}$, it has privacy loss ε. And, with probability $(1-\delta) \cdot \frac{1}{1+e^{\varepsilon}}$, it has privacy loss $-\varepsilon$. The privacy loss of the k-fold composition is the convolution of k of these privacy losses. Thus, with probability $1-(1-\delta)^k$ the privacy loss of the composition is infinite. Otherwise – i.e., with probability $(1-\delta)^k$ – the privacy loss has a shifted binomial distribution. Namely, for all $\ell \in [k] \cup \{0\}$,

$$\mathbb{P}\left[Z = \varepsilon \cdot \ell - \varepsilon \cdot (k-\ell)\right] = (1-\delta)^k \cdot \binom{k}{\ell} \cdot \left(\frac{e^{\varepsilon}}{e^{\varepsilon}+1}\right)^{\ell} \cdot \left(\frac{1}{e^{\varepsilon}+1}\right)^{k-\ell},$$

where Z is the privacy loss of the k-fold composition of the worst-case (ε, δ)-DP mechanism. Applying Proposition 3.7 to this distribution yields the expression for the optimal composition theorem.

Kairouz, Oh, and Viswanathan [KOV15] also considered *heterogeneous* optimal composition. That is, the composition of k mechanisms where each mechanism $j \in [k]$ has a different $(\varepsilon_j, \delta_j)$-DP guarantee. However, the expression becomes more complicated. Intuitively, this is because the privacy loss distribution can be

xiv. The original proof showed that the k-fold composition of (ε, δ)-DP algorithms satisfies $(\varepsilon', k\delta + \delta')$-DP with $\delta' > 0$ arbitrary and $\varepsilon' = k\varepsilon(e^{\varepsilon} - 1) + \varepsilon \cdot \sqrt{2k \log(1/\delta')}$. The first term $k\varepsilon(e^{\varepsilon} - 1)$ is slightly worse than Theorem 3.22, which gives $\frac{1}{2}k\varepsilon^2$ instead.

supported on 2^k points in the heterogeneous case, whereas, in the homogeneous case, it is supported on only $k+1$ points. Thus it takes exponential time to compute the privacy loss distribution. To be precise, Murtagh and Vadhan [MV16] showed that exactly computing the optimal composition is #P-complete, even if $\delta_j = 0$ for each $j \in [k]$. However, Murtagh and Vadhan also showed that the optimal composition theorem could be approximated to arbitrary precision in polynomial time.

Although these composition results [KOV15; MV16] are optimal, they are limited in that they begin by assuming some $(\varepsilon_j, \delta_j)$-DP guarantees about the algorithms being composed. However, we usually know more about the algorithms being composed than simply these two parameters. For example, we may know that the algorithms being composed are Gaussian noise addition. Incorporating this additional information allows us to prove even better bounds than optimal composition. This was the main impetus for the development of concentrated DP and Rényi DP, which we have discussed.

A recent line of work [MM18; KJH20; KJPH21; KH21; GLW21; DRS19; ZDW22; CKS20; GKKM22] has explored optimal composition guarantees whilst incorporating additional information about the mechanisms being composed. To make these computations efficient they consider the (discrete) Fourier transform of the privacy loss.[xv] That is, where concentrated DP and Rényi DP consider the moment generating function of the privacy loss $\mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(M(x)\|M(x'))}\left[\exp(tZ)\right]$, these works look at the characteristic function $\mathop{\mathbb{E}}_{Z \leftarrow \mathrm{PrivLoss}(M(x)\|M(x'))}\left[\exp(itZ)\right]$, where $i^2 = -1$. These methods provide composition guarantees which are arbitrarily close to optimal, which are thus better than what is attainable via concentrated DP or Rényi DP.

The optimality of advanced composition (Theorem 3.24) is due to Bun, Ullman, and Vadhan [BUV14]. We present the analysis following Kamath and Ullman [KU20].

The composition results we have presented all assume that the privacy parameters of the algorithms being composed (i.e., $(\varepsilon_j, \delta_j)$ for $j \in [k]$ in the language of Theorem 3.22) are fixed. It is natural to consider the setting where these parameters are chosen adaptively [RRUV16] – i.e., $(\varepsilon_j, \delta_j)$ could depend on the output of M_{j-1}. For the most part, the composition results carry over seamlessly to the setting of adaptively-chosen privacy parameters. In particular, for Concentrated or Rényi DP, as long as the sum of the adaptively-chosen privacy parameters remains

xv. To apply a *discrete* Fourier transform, we must first discretize the privacy loss distribution, e.g., by rounding it to a grid. The choice of discretization determines the tightness of the final guarantee, and the computational complexity of computing it.

bounded, we attain privacy with that bound [FZ21]. Another extension is "concurrent composition" [VW21], which applies when an adversary may simultaneously access multiple interactive DP systems. Fortunately, the standard composition results readily extend to this setting [VZ22; Lyu22].

Privacy Amplification by Subsampling

The first explicit statement of differential privacy amplification by subsampling was in a blog post by Smith [Smi09], although it appeared implicitly earlier [Kas+11] and the privacy effects of sampling on its own had also been studied [CM06].

For approximate DP, Balle, Barthe, and Gaborardi [BBG18] provide a thorough analysis of privacy amplification by subsampling (cf. Theorem 3.28). They present tight results for Poisson sampling (i.e., including each element independently), sampling a subset of a fixed size (without replacement), and also sampling with replacement, which means a person's data may appear *multiple* times in the subsampled dataset.

As discussed in Sections 3.6.3 and 3.6.7, subsampling arises in differentially private versions of stochastic gradient descent [CMS11; BST14]. Abadi, Chu, Goodfellow, McMahan, Mironov, Talwar, and Zhang [Aba+16] applied this in the context of deep learning. To obtain better analyses, they developed the "Moments Accountant" – i.e., Rényi DP (although the connection to Rényi divergences was only made later [Mir17; BS16]).

Abadi et al. [Aba+16] obtained asymptotic Rényi DP bounds for the Poisson subsampled Gaussian mechanism, but they used numerical integration for their implementation. Mironov, Talwar, and Zhang [MTZ19] improved these asymptotic results and gave a better numerical method for exact computation (cf. Theorem 3.33); our presentation in Section 3.6.5 largely follows theirs. Bun, Dwork, Rothblum, and Steinke [BDRS18] prove asymptotic Rényi DP bounds for Poisson subsampling applied to a concentrated DP mechanism (cf. Theorem 3.37). Zhu and Wang [ZW19] gave generic Rényi DP bounds for Poisson subsampling.[xvi]

Moving away from Poisson subsampling, Wang, Balle, and Kasiviswanathan [WBK19] provide Rényi DP results for sampling a fixed-size set without replacement.

Koskela, Jälkö, and Honkela [KJH20] provide expressions for the privacy loss distribution of the subsampled Gaussian (under both Poisson subsampling and sampling a fixed size set with or without replacement) which can be numerically integrated to obtain optimal composition results.

xvi. Mironov, Talwar, and Zhang [MTZ19] and Zhu and Wang [ZW19] both provide analogs of Theorem 3.34. However, to the best of our knowledge, Theorem 3.34 is novel.

Closely related to privacy amplification by subsampling is privacy amplification by *shuffling* [Bit+17; Erl+19; Che+19; BBGN19; FMT21; FMT22]. Privacy amplification by shuffling is usually presented in terms of local differential privacy [Kas+11]. That is, there are n individuals who independently generate random messages that satisfy local ε-DP. Those messages are then "shuffled" so that the potential adversary/attacker cannot identify which message originated from which individual. The additional randomness of the shuffling amplifies the privacy to $\left(O\left(\varepsilon \cdot \sqrt{\frac{\log(1/\delta)}{n}}\right), \delta\right)$-DP.

Intuitively, shuffling is similar to subsampling with composition. Suppose we repeatedly sample one individual uniformly at random and perform an ε-DP computation on their data and the number of repetitions is equal to the number of individuals n. We can analyze this as subsampling a $1/n$ fraction (fixed size set) composed n times. Privacy amplification by subsampling (Theorem 3.28) says each repetition is ε'-DP for $\varepsilon' = \log(1 + \frac{1}{n}(e^{\varepsilon} - 1)) = O(\varepsilon/n)$. Advanced composition (Theorem 3.18) over the n repetitions yields (ε'', δ)-DP for $\varepsilon'' = O(\sqrt{n \log(1/\delta)} \cdot \varepsilon') = O\left(\varepsilon \cdot \sqrt{\frac{\log(1/\delta)}{n}}\right)$.

In contrast, for shuffling, we sample without replacement, so no individual is sampled more than once. This means the samples are not independent, so we cannot appeal to the subsampling plus composition analysis. Nevertheless, this intuition leads to the correct result.

References

[Aba+16] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. "Deep Learning with Differential Privacy". In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. CCS '16. Vienna, Austria: Association for Computing Machinery, 2016, pp. 308–318. ISBN: 9781450341394. URL: https://doi.org/10.1145/2976749.2978318 (cit. on pp. 126, 147).

[BBG18] B. Balle, G. Barthe, and M. Gaboardi. "Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences". In: Advances in Neural Information Processing Systems. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/

file / 3b5020bb891119b9f5130f1fea9bd773 - Paper . pdf (cit. on p. 147).

[BBGN19] B. Balle, J. Bell, A. Gascón, and K. Nissim. "The Privacy Blanket of the Shuffle Model". In: Advances in Cryptology – CRYPTO 2019. Ed. by A. Boldyreva and D. Micciancio. Cham: Springer International Publishing, 2019, pp. 638–667. ISBN: 978-3-030-26951-7. URL: https://arxiv.org/abs/1903.02837 (cit. on p. 148).

[BDRS18] M. Bun, C. Dwork, G. N. Rothblum, and T. Steinke. "Composable and Versatile Privacy via Truncated CDP". In: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 74–86. ISBN: 9781450355599. URL: https://doi.org/10.1145/3188745.3188946 (cit. on p. 147).

[Bit+17] A. Bittau, Ú. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnes, and B. Seefeld. "Prochlo: Strong privacy for analytics in the crowd". In: Proceedings of the 26th symposium on operating systems principles. 2017, pp. 441–459 (cit. on p. 148).

[BS16] M. Bun and T. Steinke. "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: Theory of Cryptography Conference. Springer. 2016, pp. 635–658. URL: https://arxiv.org/abs/1605.02065 (cit. on pp. 96, 129, 145, 147).

[BST14] R. Bassily, A. Smith, and A. Thakurta. "Private empirical risk minimization: Efficient algorithms and tight error bounds". In: 2014 IEEE 55th annual symposium on foundations of computer science. IEEE. 2014, pp. 464–473 (cit. on p. 147).

[Bun16] M. M. Bun. "New Separations in the Complexity of Differential Privacy". PhD thesis. 2016 (cit. on p. 113).

[BUV14] M. Bun, J. Ullman, and S. Vadhan. "Fingerprinting Codes and the Price of Approximate Differential Privacy". In: Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing. STOC '14. New York, New York: Association for Computing Machinery, 2014, pp. 1–10. ISBN: 9781450327107. URL: https://arxiv.org/abs/1311.3158 (cit. on p. 146).

[BW18] B. Balle and Y.-X. Wang. "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising". In: International Conference on Machine Learning. PMLR. 2018, pp. 394–403 (cit. on p. 145).

[Che+19] A. Cheu, A. Smith, J. Ullman, D. Zeber, and M. Zhilyaev. "Distributed differential privacy via shuffling". In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer. 2019, pp. 375–403. URL: https://arxiv.org/abs/1808.01394 (cit. on p. 148).

[CKS20] C. L. Canonne, G. Kamath, and T. Steinke. "The discrete gaussian for differential privacy". In: Advances in Neural Information Processing Systems 33 (2020), pp. 15676–15688. URL: https://journalprivacyconfidentiality.org/index.php/jpc/article/view/784 (cit. on pp. 113, 146).

[CM06] K. Chaudhuri and N. Mishra. "When Random Sampling Preserves Privacy". In: Proceedings of the 26th Annual International Conference on Advances in Cryptology. CRYPTO'06. Santa Barbara, California: Springer-Verlag, 2006, pp. 198–213. ISBN: 3540374329. URL: https://doi.org/10.1007/11818175_12 (cit. on p. 147).

[CMS11] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate. "Differentially Private Empirical Risk Minimization". In: J. Mach. Learn. Res. 12.null (July 2011), pp. 1069–1109. ISSN: 1532-4435 (cit. on p. 147).

[DMNS06] C. Dwork, F. McSherry, K. Nissim, and A. Smith. "Calibrating noise to sensitivity in private data analysis". In: Theory of cryptography conference. Springer. 2006, pp. 265–284 (cit. on pp. 82, 144).

[DMNS17] C. Dwork, F. McSherry, K. Nissim, and A. Smith. "Calibrating Noise to Sensitivity in Private Data Analysis". In: Journal of Privacy and Confidentiality 7.3 (May 2017), pp. 17–51. URL: https://journalprivacyconfidentiality.org/index.php/jpc/article/view/405 (cit. on p. 144).

[DR16] C. Dwork and G. N. Rothblum. "Concentrated differential privacy". In: arXiv preprint arXiv:1603.01887 (2016). URL: https://arxiv.org/abs/1603.01887 (cit. on pp. 96, 145).

[DRS19] J. Dong, A. Roth, and W. J. Su. "Gaussian differential privacy". In: arXiv preprint arXiv:1905.02383 (2019) (cit. on p. 146).

[DRV10] C. Dwork, G. N. Rothblum, and S. Vadhan. "Boosting and differential privacy". In: 2010 IEEE 51st Annual Symposium on Foundations of Computer Science. IEEE. 2010, pp. 51–60 (cit. on pp. 101, 145).

[DSSU17] C. Dwork, A. Smith, T. Steinke, and J. Ullman. "Exposed! a survey of attacks on private data". In: Annu. Rev. Stat. Appl 4.1 (2017), pp. 61–84 (cit. on p. 113).

[Dwo+06] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. "Our data, ourselves: Privacy via distributed noise generation". In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer. 2006, pp. 486–503 (cit. on pp. 82, 145).

[Dwo06] C. Dwork. "Differential Privacy". In: Automata, Languages and Programming. Ed. by M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–12. ISBN: 978-3-540-35908-1 (cit. on p. 144).

[Erl+19] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta. "Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity". In: Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). 2019, pp. 2468–2479. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611975482.151. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611975482.151 (cit. on p. 148).

[FMT21] V. Feldman, A. McMillan, and K. Talwar. "Hiding Among the Clones: A Simple and Nearly Optimal Analysis of Privacy Amplification by Shuffling". In: 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS). 2021, pp. 954–964. URL: https://arxiv.org/abs/2012.12803 (cit. on p. 148).

[FMT22] V. Feldman, A. McMillan, and K. Talwar. "Stronger Privacy Amplification by Shuffling for R\'enyi and Approximate Differential Privacy". In: arXiv preprint arXiv:2208.04591 (2022). URL: https://arxiv.org/abs/2208.04591 (cit. on p. 148).

[FZ21] V. Feldman and T. Zrnic. "Individual Privacy Accounting via a Rényi Filter". In: Advances in Neural Information Processing Systems. Ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan. Vol. 34. Curran Associates, Inc., 2021, pp. 28080–28091. URL: https://proceedings.neurips.cc/paper/2021/file/ec7f346604f518906d35ef0492709f78-Paper.pdf (cit. on p. 147).

[GKKM22] B. Ghazi, P. Kamath, R. Kumar, and P. Manurangsi. "Faster Privacy Accounting via Evolving Discretization". In: International Conference on Machine Learning. PMLR. 2022, pp. 7470–7483 (cit. on p. 146).

[GLW21] S. Gopi, Y. T. Lee, and L. Wutschitz. "Numerical Composition of Differential Privacy". In: Advances in Neural Information Processing Systems. Ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan. Vol. 34. Curran Associates, Inc., 2021, pp. 11631–11642. URL: https://proceedings.neurips.cc/paper/2021/file/6097d8f3714205740f30debe1166744e-Paper.pdf (cit. on p. 146).

[Goo18] Google. TensorFlow Privacy Library. https://github.com/tensorflow/privacy & https://github.com/google/differential-privacy. 2018 (cit. on p. 143).

[Hoe63] W. Hoeffding. "Probability for sums of bounded random variables". In: J. Am. Stat. Assoc 58 (1963) (cit. on p. 100).

[Kas+11] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. "What Can We Learn Privately?" In: SIAM Journal on Computing 40.3 (2011), pp. 793–826. eprint: https://doi.org/10.1137/090756090. URL: https://doi.org/10.1137/090756090 (cit. on pp. 147, 148).

[KH21] A. Koskela and A. Honkela. "Computing differential privacy guarantees for heterogeneous compositions using fft". In: arXiv preprint arXiv:2102.12412 (2021) (cit. on p. 146).

[KJH20] A. Koskela, J. Jälkö, and A. Honkela. "Computing Tight Differential Privacy Guarantees Using FFT". In: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. Ed. by S. Chiappa and R. Calandra. Vol. 108. Proceedings of Machine Learning Research. PMLR, Aug. 2020, pp. 2560–2569. URL: https://proceedings.mlr.press/v108/koskela20b.html (cit. on pp. 146, 147).

[KJPH21] A. Koskela, J. Jälkö, L. Prediger, and A. Honkela. "Tight Differential Privacy for Discrete-Valued Mechanisms and for the Subsampled Gaussian Mechanism Using FFT". In: Proceedings of The 24th International Conference on Artificial Intelligence and Statistics. Ed. by A. Banerjee and K. Fukumizu. Vol. 130. Proceedings of Machine Learning Research. PMLR, Apr. 2021, pp. 3358–3366.

URL: https://proceedings.mlr.press/v130/koskela21a.html (cit. on p. 146).

[KOV15] P. Kairouz, S. Oh, and P. Viswanath. "The composition theorem for differential privacy". In: International conference on machine learning. PMLR. 2015, pp. 1376–1385 (cit. on pp. 102, 145, 146).

[KU20] G. Kamath and J. Ullman. "A primer on private statistics". In: arXiv preprint arXiv:2005.00010 (2020). URL: https://arxiv.org/abs/2005.00010 (cit. on p. 146).

[Lyu22] X. Lyu. "Composition Theorems for Interactive Differential Privacy". In: arXiv preprint arXiv:2207.09397 (2022) (cit. on p. 147).

[McM+18] H. B. McMahan, G. Andrew, U. Erlingsson, S. Chien, I. Mironov, N. Papernot, and P. Kairouz. "A general approach to adding differential privacy to iterative training procedures". In: arXiv preprint arXiv:1812.06210 (2018) (cit. on p. 143).

[Mir17] I. Mironov. "Rényi differential privacy". In: 2017 IEEE 30th computer security foundations symposium (CSF). IEEE. 2017, pp. 263–275 (cit. on pp. 126, 147).

[MM18] S. Meiser and E. Mohammadi. "Tight on Budget? Tight Bounds for r-Fold Approximate Differential Privacy". In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18. Toronto, Canada: Association for Computing Machinery, 2018, pp. 247–264. ISBN: 9781450356930. URL: https://doi.org/10.1145/3243734.3243765 (cit. on p. 146).

[MTZ19] I. Mironov, K. Talwar, and L. Zhang. "R\'enyi differential privacy of the sampled gaussian mechanism". In: arXiv preprint arXiv:1908.10530 (2019) (cit. on pp. 131, 147).

[MV16] J. Murtagh and S. Vadhan. "The complexity of computing the optimal composition of differential privacy". In: Theory of Cryptography Conference. Springer. 2016, pp. 157–175 (cit. on p. 146).

[NP33] J. Neyman and E. S. Pearson. "IX. On the problem of the most efficient tests of statistical hypotheses". In: Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231.694-706 (1933), pp. 289–337 (cit. on p. 86).

[Rén61] A. Rényi. "On measures of entropy and information". In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. Vol. 1. 547-561. Berkeley, California, USA. 1961. URL: https://projecteuclid.org/proceedings/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fourth-Berkeley-Symposium-on-Mathematical-Statistics-and/Chapter/On-Measures-of-Entropy-and-Information/bsmsp/1200512181 (cit. on p. 127).

[RRUV16] R. Rogers, A. Roth, J. Ullman, and S. Vadhan. "Privacy Odometers and Filters: Pay-as-You-Go Composition". In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16. Barcelona, Spain: Curran Associates Inc., 2016, pp. 1929–1937. ISBN: 9781510838819 (cit. on p. 146).

[Smi09] A. Smith. Differential privacy and the secrecy of the sample. https://adamdsmith.wordpress.com/2009/09/02/sample-secrecy/. 2009 (cit. on p. 147).

[SSSS17] R. Shokri, M. Stronati, C. Song, and V. Shmatikov. "Membership inference attacks against machine learning models". In: 2017 IEEE symposium on security and privacy (SP). IEEE. 2017, pp. 3–18 (cit. on p. 113).

[SU15] T. Steinke and J. Ullman. "Between pure and approximate differential privacy". In: arXiv preprint arXiv:1501.06095 (2015) (cit. on p. 113).

[VW21] S. Vadhan and T. Wang. "Concurrent Composition of Differential Privacy". In: Theory of Cryptography Conference. Springer. 2021, pp. 582–604 (cit. on p. 147).

[VZ22] S. Vadhan and W. Zhang. "Concurrent Composition Theorems for all Standard Variants of Differential Privacy". In: arXiv preprint arXiv:2207.08335 (2022) (cit. on p. 147).

[WBK19] Y.-X. Wang, B. Balle, and S. P. Kasiviswanathan. "Subsampled Renyi Differential Privacy and Analytical Moments Accountant". In: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Ed. by K. Chaudhuri and M. Sugiyama. Vol. 89. Proceedings of Machine Learning Research. PMLR, Apr. 2019, pp. 1226–1235. URL: https://proceedings.mlr.press/v89/wang19b.html (cit. on p. 147).

[ZDW22] Y. Zhu, J. Dong, and Y.-X. Wang. "Optimal Accounting of Differential Privacy via Characteristic Function". In: Proceedings of The 25th International Conference on Artificial Intelligence and Statistics. Ed. by G. Camps-Valls, F. J. R. Ruiz, and I. Valera. Vol. 151. Proceedings of Machine Learning Research. PMLR, Mar. 2022, pp. 4782–4817. URL: https://proceedings.mlr.press/v151/zhu22c.html (cit. on p. 146).

[ZW19] Y. Zhu and Y.-X. Wang. "Poission Subsampled Rényi Differential Privacy". In: Proceedings of the 36th International Conference on Machine Learning. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 7634–7642. URL: https://proceedings.mlr.press/v97/zhu19c.html (cit. on p. 147).