Journal of Privacy and Confidentiality
Vol. 10 (1) 2020 TPDP 2018
Submitted Mar 2019
Published Jan 2020

PRIVACY PROFILES AND AMPLIFICATION BY SUBSAMPLING

BORJA BALLE, GILLES BARTHE, AND MARCO GABOARDI

Unaffiliated
e-mail address: borja.balle@gmail.com

MPI for Security and Privacy and IMDEA Software Institute
e-mail address: gilles.barthe@imdea.org

University at Buffalo
e-mail address: gaboardi@buffalo.edu

Abstract. Differential privacy provides a robust quantifiable methodology to measure and control the privacy leakage of data analysis algorithms. A fundamental insight is that by forcing algorithms to be randomized, their privacy leakage can be characterized by measuring the dissimilarity between output distributions produced by applying the algorithm to pairs datasets differing in one individual. After the introduction of differential privacy, several variants of the original definition have been proposed by changing the measure of dissimilarity between distributions, including concentrated, zero-concentrated and Rényi differential privacy.

The first contribution of this paper is to introduce the notion of privacy profile of a mechanism. This profile captures all valid (ε, δ) differential privacy parameters satisfied by a given mechanism, and contrasts with the usual approach of providing guarantees in terms of a single point in this curve. We show that knowledge of this curve is equivalent to knowledge of the privacy guarantees with respect to the alternative definitions listed above. This sheds further light into the connections among multiple privacy definitions, and suggests that these should be considered alternative but otherwise equivalent points of view.

The second contribution of this paper is to apply the privacy profiles machinery to study the so-called "privacy amplification by subsampling" principle, which ensures that a differentially private mechanism run on a random subsample of a population provides higher privacy guarantees than when run on the entire population. Several instances of this principle have been studied for different random subsampling methods, each with an ad hoc analysis. In this paper we set out to study this phenomenon in detail with the aim to provide a general method capable of recovering prior analyses in a streamlined fashion. Our method makes extensive use of coupling argument, and introduces a new tool to analyse differential privacy for mixture distributions.

Key words and phrases: Differential Privacy, Concentrated Differential Privacy, Rényi Differential Privacy, Privacy Amplification, Subsampling.

* A preliminary version of this work appeared in Balle et al. (2018). Balle was unaffiliated at the time of writing this article. He is now with Deepmind.

JPrivacy Confidentiality
www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.726

1. Introduction

The success of differential privacy (Dwork et al., 2006) in providing a rigorous methodology to design privacy-preserving data analysis algorithms has sparkled an interest in variants of the original definition adapted to particular circumstances. A salient example of this phenomenon is the multiple definitions that arose simultaneously around the concentration property of privacy loss random variables, which play a crucial role in the classical analyses of the Gaussian mechanism (Dwork et al., 2006) and the advanced composition theorem (Dwork et al., 2010). These definitions include concentrated differential privacy (Dwork and Rothblum, 2016), zero-concentrated differential privacy (Bun and Steinke, 2016), Rényi differential privacy (Mironov, 2017), truncated-concentrated differential privacy (Bun et al., 2018), as well as the implicit usage made of the concentration property in the moments accountant technique (Abadi et al., 2016). This profusion of definitions, together with a practical need to explain the resulting privacy guarantees to regulators and decision-makers, resulted in a number of results allowing one to translate the privacy parameters used in one definition into the parameters of another definition. For example, all the definitions based on the concentration of privacy loss random variables imply a certain level of approximate differential privacy, and pure differential privacy implies a certain level of privacy with respect to the concentration-based definitions.

The first goal of this paper is to take a deeper look at these conversion results from a mathematical standpoint. Our starting point is to associate with each mechanism $\mathcal{M}$ a *privacy profile* $\delta_{\mathcal{M}}(\varepsilon)$ such that the mechanism satisfies $(\varepsilon, \delta_{\mathcal{M}}(\varepsilon))$-DP for each $\varepsilon \geq 0$. To do this, we leverage the connection between differential privacy and a family of divergences in the sense of Csiszár, which was first observed in the context of formal verification for differential privacy (Barthe et al., 2012; Barthe and Olmedo, 2013; Barthe et al., 2016). By studying the properties of these profiles we conclude that knowledge of the privacy profile of a mechanism is mathematically equivalent to knowledge of its privacy guarantees with respect to the concentration-based definitions listed above. In other words, we show that moving from the pointwise approach – which focuses on (ε, δ)-DP guarantees for a single setting of the privacy parameters — to the functional approach – which views one parameter as a function of another — provides a deeper understanding of the privacy properties of a given mechanism, and, in particular, that these different privacy definitions are just different points of view of the same phenomenon. This is the content of Section 3, which is preceded by a number of preliminaries discussed in Section 2.

The second part of the paper is devoted to the study of privacy amplification by subsampling. Subsampling is a fundamental tool in the design and analysis of differentially private mechanisms. Broadly speaking, the intuition behind the "privacy amplification by subsampling" principle is that the privacy guarantees of a differentially private mechanism can be amplified by applying it to a small random subsample of records from a given dataset. In machine learning, many classes of algorithms involve sampling operations, e.g., stochastic optimization methods and Bayesian inference algorithms, and it is not surprising that results quantifying the privacy amplification obtained via subsampling play a key role in designing differentially private versions of these learning algorithms (Bassily et al., 2014; Wang et al., 2015; Abadi et al., 2016; Jälkö et al., 2017; Park et al., 2016b,a). Additionally, from a practical standpoint, subsampling provides a straightforward method to obtain privacy amplification when the final mechanism is only available as a black-box. For example, in Apple's iOS and Google's Chrome deployments of differential privacy for data collection the

privacy parameters are hard-coded into the implementation and cannot be modified by the user. In these types of settings, if the default privacy parameters are not satisfactory, one could achieve a stronger privacy guarantee by devising a strategy that only submits to the mechanism a random sample of the data.

Despite the practical importance of subsampling, existing tools to bound privacy amplification work only for specific forms of subsampling and typically come with cumbersome proofs providing no information about the tightness of the resulting bounds. The goal of Section 4 is to remedy this situation by providing a general framework for deriving tight privacy amplification results that can be applied to any of the subsampling strategies considered in the literature. Our framework builds on the privacy profiles machinery developed in the first part of the paper, and includes a novel analytical tool – advanced joint convexity – which is used to analyze the privacy guarantees of mixture distributions and might be of independent interest.

One of our motivations to initiate a systematic study of privacy amplification by subsampling is that this is an important primitive for the design of differentially private algorithms that has received less attention than other building blocks like composition theorems (Dwork et al., 2010; Kairouz et al., 2017; Murtagh and Vadhan, 2016). Given the relevance of sampling operations in machine learning, it is important to understand what are the limitations of privacy amplification and develop a fine-grained understanding of its theoretical properties. Our results provide a first step in this direction by showing how privacy amplification resulting from different sampling techniques can be analyzed by means of single set of tools, and by showing how these tools can be used for proving lower bounds. Our analyses also highlight the importance of choosing a sampling technique that is well-adapted to the notion of neighboring datasets under consideration.

A second motivation is that subsampling provides a natural example of mechanisms where the output distribution is a mixture. Because mixtures have an additive structure and differential privacy is defined in terms of a multiplicative guarantee, analyzing the privacy guarantees of mechanisms whose output distribution is a mixture is in general a challenging task. Although our analyses are specialized to mixtures arising from subsampling, we believe the tools we develop in terms of couplings and divergences will also be useful to analyze other types of mechanisms involving mixture distributions. Furthermore, amplification by subsampling is just one of the many privacy amplification phenomena identified so far. Others include amplification by iteration (Feldman et al., 2018) and by shuffling (Erlingsson et al., 2019; Cheu et al., 2018; Balle et al., 2019). Studying these amplification phenomena is important because they enable finer privacy analysis of useful data analysis pipelines, and show that operations like subsampling and shuffling, which by themselves provide no meaningful differential privacy guarantees, can in fact amplify the privacy guarantees of existing mechanism. Finally, we want to remark that privacy amplification results also play a role in analyzing the generalization and sample complexity properties of private learning algorithms (Kasiviswanathan et al., 2011; Beimel et al., 2013; Bun et al., 2015; Wang et al., 2016); an in-depth understanding of the interplay between sampling and differential privacy might also have applications in this direction.

2. Preliminaries

2.1. **Distributions, Densities and Divergences.** Let $\mathbb{P}(Z, \Sigma)$ denote the set of probability measures on a measurable space Z equipped with a σ-algebra Σ. We will just write $\mathbb{P}(Z)$ when the σ-algebra is clear from the context. For example, Σ might be the collection of all possible subsets of Z when the space is discrete, or the collection of all Lebesgue-measurable subsets of Z when the space is a (subset of) Euclidean space. Recall that given measures $\mu, \nu \in \mathbb{P}(Z, \Sigma)$ we say that μ is absolutely continuous with respect to ν (and we write $\mu \ll \nu$) if $\nu(E) = 0$ implies $\mu(E) = 0$ for any measurable set $E \in \Sigma$. The Radon-Nikodym theorem (see, e.g. Pollard, 2002) says that if $\mu \ll \nu$ then there exists a measurable function $f : Z \to [0, \infty)$ such that $d\mu = f d\nu$ in the sense that for any measurable E we have $\mu(E) = \int_E d\mu = \int_E f d\nu$. Such function is known as the Radon-Nikodym derivative of μ with respect to ν and is usually denoted as $f = d\mu/d\nu$. Another usual name for f is the *density* of μ with respect to ν.

A standard way to measure the similarity between two probability distributions is to use divergences. A divergence in the sense of Csiszár is obtained from a convex function $\phi : [0, \infty] \to \mathbb{R} \cup \{\infty\}$ such that $\phi(1) = 0$. Given such a function ϕ and two probability measures $\mu \ll \nu$, the ϕ-divergence between μ and ν is defined as

$$D_\phi(\mu \| \nu) = \int \phi\left(\frac{d\mu}{d\nu}\right) d\nu \ . \tag{2.1}$$

When μ is not absolutely continuous with respect to ν, the divergence D_ϕ takes a slightly different definition. Let λ be a probability measure such that $\mu \ll \lambda$ and $\nu \ll \lambda$ (e.g., $\lambda = (\mu + \nu)/2$, and define the respective densities $p = d\mu/d\lambda$ and $q = dnu/d\lambda$. Then the divergence between μ and ν is given by

$$D_\phi(\mu \| \nu) = \int \phi\left(\frac{p}{q}\right) d\nu \ . \tag{2.2}$$

Noting that the function $p/q : X \to [0, \infty]$ is independent of the choice of λ, one sees that $D_\phi(\mu \| \nu)$ is well-defined.

The assumptions on ϕ imply that all ϕ-divergences satisfy a number of interesting properties (see, e.g., (Liese and Vajda, 2006)). Here we recall the following:

(1) (Nonnegativity) $D_\phi(\mu \| \nu) \geq 0$, $D_\phi(\mu \| \mu) = 0$, and if additionally ϕ is strictly convex at 1 then $D_\phi(\mu \| \nu) = 0$ implies $\mu = \nu$.
(2) (Joint convexity) $D_\phi((1-\gamma)\mu + \gamma\mu' \| (1-\gamma)\nu + \gamma\nu') \leq (1-\gamma) D_\phi(\mu \| \nu) + \gamma D_\phi(\mu' \| \nu')$ for any $\gamma \in (0, 1)$.
(3) (Processing inequality) If K is a Markov kernel, then $D_\phi(\mu K \| \nu K) \leq D_\phi(\mu \| \nu)$.

A particular family of ϕ-divergences that plays a central role in differential privacy are *hockey-stick divergences* (Sason and Verdú, 2016)[1]. For $\beta \geq 1$, the hockey-stick divergence of order β is the divergence defined by $\phi_\beta(u) = [u - \beta]_+$, where $[a]_+ = \max\{0, a\}$. Throughout the paper we use the shorthand notation D_β to denote the divergence D_{ϕ_β}.

Another family of divergences that plays an important role in the theory of differential privacy are Rényi divergences. Given distributions $\mu \ll \nu$ and $\alpha > 1$, the Rényi divergence

[1] Also known in the literature as elementary divergences (Österreicher, 2002) or alpha-divergences (Barthe and Olmedo, 2013)

between μ and ν is defined as[2]

$$R_\alpha(\mu\|\nu) = \frac{1}{\alpha-1}\log\left(\int\left(\frac{d\mu}{d\nu}\right)^\alpha d\nu\right) \ .$$

Rényi divergences are not divergences in the sense of Csiszár, and in particular they fail to satisfy the joint convexity property, although they satisfy a weaker joint quasi-convexity property (Liese and Vajda, 2006). Nonetheless, Rényi divergences can be directly related to ϕ-divergences by taking $\tilde{\phi}_\alpha(u) = u^\alpha - 1$ and noting that

$$R_\alpha(\mu\|\nu) = \frac{1}{\alpha-1}\log\left(D_{\tilde{\phi}_\alpha}(\mu\|\nu)+1\right) \ .$$

2.2. **Differential Privacy.** A *mechanism* $\mathcal{M}: X \to \mathbb{P}(Z)$ with input space X and output space Z is a randomized algorithm that on input x outputs a sample from the distribution $\mathcal{M}(x)$ over Z. We assume the input space X is equipped with a binary *symmetric* relation $\simeq_X$ defining a notion of neighboring (or adjacent) inputs. When the input space X is clear from the context we shall just write $\simeq$.

Suppose that $\varepsilon \geq 0$ and $\delta \in [0,1]$. A mechanism $\mathcal{M}$ is said to be (ε,δ)-*DP* with respect to $\simeq$ if for every pair of inputs $x \simeq x'$ and every (measurable) subset $E \subseteq Z$ we have

$$\mathsf{Pr}[\mathcal{M}(x) \in E] \leq e^\varepsilon \mathsf{Pr}[\mathcal{M}(x') \in E] + \delta \ . \tag{2.3}$$

For our purposes, it will sometimes be more convenient to express differential privacy in terms of the hockey-stick divergence D_{e^ε}. The exact characterization is given by the following result.

Theorem 1 (Barthe and Olmedo, 2013)**.** *A mechanism $\mathcal{M}$ is (ε,δ)-DP with respect to $\simeq$ if and only if $D_{e^\varepsilon}(\mathcal{M}(x)\|\mathcal{M}(x')) \leq \delta$ for every x and x' such that $x \simeq x'$.*

It is an instructive exercise to see why this characterization holds, so we provide a proof sketch of this result for completeness.

Proof. Fix $x \simeq x'$ and start by simply re-writing the condition that (2.3) holds for every (measurable) subset $E \subseteq Z$ as

$$\sup_E \left(\mathsf{Pr}[\mathcal{M}(x) \in E] - e^\varepsilon \mathsf{Pr}[\mathcal{M}(x') \in E]\right) \leq \delta \ .$$

Writing μ and ν for the distributions of $\mathcal{M}(x)$ and $\mathcal{M}(x')$ respectively, we have

$$\begin{aligned}\mathsf{Pr}[\mathcal{M}(x) \in E] - e^\varepsilon \mathsf{Pr}[\mathcal{M}(x') \in E] &= \mu(E) - e^\varepsilon \nu(E) \\ &= \int_E d\mu - e^\varepsilon \int_E d\nu \\ &= \int_E \left(\frac{d\mu}{d\lambda} - e^\varepsilon \frac{d\nu}{d\lambda}\right) d\lambda \ ,\end{aligned}$$

where λ is any probability measure such that $\mu \ll \lambda$ and $\nu \ll \lambda$. Denote by $p = d\mu/d\lambda$ and $q = d\nu/d\lambda$ the densities of μ and ν with respect to λ. It is easy to check that the set $E \subseteq Z$ that maximizes the integral expression above is given by

$$E^* = E^*(x,x') = \{z \in Z \mid p(z) > e^\varepsilon q(z)\} \ .$$

[2]If $\mu \ll \nu$ is not satisfied we take $R_\alpha(\mu\|\nu) = \infty$.

Note that because p and q are measurable functions it follows that E^* is a measurable set. Now observe that from the definition of E^* and the identity $[a]_+ = a\mathbb{I}[a > 0]$ it follows that

$$\begin{aligned}\int_{E^*} (p(z) - e^\varepsilon q(z))\, d\lambda(z) &= \int_Z [p(z) - e^\varepsilon q(z)]_+ \, d\lambda(z) \\ &= \int_Z \left[\frac{p(z)}{q(z)} - e^\varepsilon\right]_+ q(z) d\lambda(z) \\ &= D_{e^\varepsilon}(\mu\|\nu) \ ,\end{aligned}$$

where we used the definition of hockey-stick divergence in terms of (2.2). Thus, the derivation above yields the following identity, from which the characterization follows:

$$\sup_E \left(\mathsf{Pr}[\mathcal{M}(x) \in E] - e^\varepsilon \mathsf{Pr}[\mathcal{M}(x') \in E]\right) = D_{e^\varepsilon}(\mathcal{M}(x)\|\mathcal{M}(x')) \ .$$

□

The main advantage of the divergence point of view is that it allows one to move from the usual characterization of (ε, δ)-DP in terms of events to a characterization involving the integral of a quantity defined on individual outputs. Reasoning about individual outputs is usually easier than reasoning about all possible events, so this approach sometimes leads to simpler or tighter privacy proofs. Furthermore, this characterization immediately makes the properties of ϕ-divergences available in the analysis of differentially private algorithms; we shall return to this point in Section 3.3.

We conclude this section by recalling the group privacy property of differential privacy. Given an integer $k \geq 1$ we write $\simeq^k$ for the k-fold transitive extension of the neighboring relation $\simeq$ defined as

$$x \simeq^k x' \quad \Leftrightarrow \quad \exists x_1, \ldots, x_{k-1} : x \simeq x_1, x_1 \simeq x_2, \ldots, x_{k-1} \simeq x' \ .$$

The relation $\simeq^k$ captures the notion that sometimes we might want to protect the privacy of a dataset with respect to k changes instead of just one; for example, when a single individual contributes a maximum of k records to a database, protecting that individual's data will require hiding up to k potential changes in the database. The *group privacy* property states that if $\mathcal{M}$ is an (ε, δ)-DP mechanism with respect to $\simeq$, then $\mathcal{M}$ is also $(k\varepsilon, \delta')$-DP with respect to $\simeq^k$, where $\delta' = \frac{e^{k\varepsilon}-1}{e^\varepsilon - 1}\delta$ (Vadhan, 2017, Lemma 2.2). For further reference, we recall that the relation $\simeq^k$ can also be defined in terms of the *path-metric* $d(x, x')$ induced by $\simeq$ on X:

$$d(x, x') = \min\{k : \exists x_1, \ldots, x_{k-1}, x \simeq x_1, x_1 \simeq x_2, \ldots, x_{k-1} \simeq x'\} \ .$$

This metric satisfies $x \simeq^k x'$ if and only if $d(x, x') \leq k$.

2.3. **Rényi Differential Privacy and Related Variants.** Since the introduction of differential privacy, a number of variants of the definition have been proposed. A fruitful way to obtain interesting variants is to change the way in which the dissimilarity between the distributions of $\mathcal{M}(x)$ and $\mathcal{M}(x')$ is measured. In particular, several such definitions have been inspired by the connections between the privacy loss random variable, the advanced composition theorem, and the analysis of the Gaussian mechanism. We now recall the definition of the privacy loss random variable, show how it is related to the definition of Rényi differential privacy, and sketch the connections between this definition and other notions of concentrated differential privacy.

The *privacy loss random variable* of a mechanism $\mathcal{M}$ on a pair of inputs $x \simeq x'$ is defined as $\mathsf{L}_{\mathcal{M}}^{x,x'} = \log\left(\frac{d\mu}{d\nu}(\mathsf{Z})\right)$, where $\mu = \mathcal{M}(x)$, $\nu = \mathcal{M}(x')$, and $\mathsf{Z} \sim \mu$. In the case where μ is not absolutely continuous with respect to ν we can still define the privacy loss random variable by following the same idea in the definition (2.2) of ϕ-divergences. In particular, taking a probability measure λ such that $\mu \ll \lambda$ and $\nu \ll \lambda$, we take $p = d\mu/d\lambda$ and $q = d\nu/d\lambda$, and define $\mathsf{L}_{\mathcal{M}}^{x,x'} = \log(p(\mathsf{Z})/q(\mathsf{Z}))$, with Z as above. We note, however, that these technicalities have typically no effect on the type of results we are interested in in this paper, and we shall ignore them from now onward unless explicitly stated.

The privacy loss random variable provides a well-known sufficient condition for differential privacy, stating that if $\mathcal{M}$ satisfies $\mathsf{Pr}[\mathsf{L}_{\mathcal{M}}^{x,x'} \geq \varepsilon] \leq \delta$ for any $x \simeq x'$, then $\mathcal{M}$ is (ε, δ)-DP (e.g., see (Dwork and Roth, 2014)). Since tail inequalities of this type are often established by applying Chernoff's method to Markov's inequality in order to control the concentration of a random variable in terms of its moment generating function (e.g., (Boucheron et al., 2013)), this has lead to the realization that interesting privacy definitions can be phrased in terms of the moment generating function of the privacy loss random variable.

One such definition is Rényi differential privacy (Mironov, 2017). Suppose that $\alpha \in (1, \infty)$ and $\varepsilon \geq 0$. A mechanism $\mathcal{M}$ is said to be (α, ϵ)-*RDP* with respect to $\simeq$ if for every pair of inputs $x \simeq x'$ we have

$$R_\alpha(\mathcal{M}(x) \| \mathcal{M}(x')) \leq \epsilon \ . \tag{2.4}$$

To understand the connection between the privacy loss random variable and RDP one needs to note that the moment generating function $\varphi(s)$, $s > 0$, of $\mathsf{L} = \mathsf{L}_{\mathcal{M}}^{x,x'}$ can be written as

$$\begin{aligned}
\varphi(s) &= \mathsf{E}\left[e^{s\mathsf{L}}\right] && (2.5)\\
&= \mathsf{E}\left[\left(\frac{d\mu}{d\nu}(\mathsf{Z})\right)^s\right] && (2.6)\\
&= \int \left(\frac{d\mu}{d\nu}\right)^s d\mu \\
&= \int \left(\frac{d\mu}{d\nu}\right)^{s+1} d\nu && (2.7)\\
&= e^{sR_{s+1}(\mathcal{M}(x)\|\mathcal{M}(x'))} \ . && (2.8)
\end{aligned}$$

Hence, upper bounding the Rényi divergence between $\mathcal{M}(x)$ and $\mathcal{M}(x')$ for any pair of inputs $x \simeq x'$ is equivalent to upper bounding a particular value of the moment generating function of the privacy loss random variables on every pair of neighboring inputs.

In addition to the pointwise bound on the Rényi divergence assumed by RDP, one can assume a *parametric* bound holding for all (or a subset of) values of α. This leads to the following variants of differential privacy:

- (Concentrated Differential Privacy (Dwork and Rothblum, 2016)) A mechanism $\mathcal{M}$ is (μ, τ)-CDP if $R_\alpha(\mathcal{M}(x)\|\mathcal{M}(x')) \leq \mu + (\alpha - 1)\frac{\tau^2}{2}$ for all $\alpha \in (1, \infty)$ and $x \simeq x'$.
- (Zero-Concentrated Differential Privacy (Bun and Steinke, 2016)) A mechanism $\mathcal{M}$ is (ξ, ρ)-zCDP if $R_\alpha(\mathcal{M}(x)\|\mathcal{M}(x')) \leq \xi + \alpha\rho$ for all $\alpha \in (1, \infty)$ and $x \simeq x'$.
- (Truncated-Concentrated Differential Privacy (Bun et al., 2018)) A mechanism $\mathcal{M}$ is (ρ, ω)-tCDP if $R_\alpha(\mathcal{M}(x)\|\mathcal{M}(x')) \leq \alpha\rho$ for all $\alpha \in (1, \omega)$ and $x \simeq x'$.

2.4. **Couplings and Total Variation Distance.** Couplings are a standard tool in probability theory, and are useful for several tasks, including deriving upper bounds for the total variation distance between distributions. A coupling between two distributions $\mu, \nu \in \mathbb{P}(Y)$ is a distribution $\pi \in \mathbb{P}(Y \times Y)$ whose marginals along the projections $(y, y') \mapsto y$ and $(y, y') \mapsto y'$ are μ and ν respectively. Couplings always exist, and furthermore, there exists a *maximal coupling*, which exactly characterizes the total variation distance between μ and ν.

Recall that the *total variation distance* $\mathsf{TV}(\mu, \nu)$ between two probability distributions admits a number of characterizations, including being a divergence in the sense of Csiszár obtained from $\phi(u) = [u-1]_+$ and $\phi(u) = \frac{1}{2}|u-1|$, as well as the supremum characterizations

$$\mathsf{TV}(\mu, \nu) = \sup_E (\mu(E) - \nu(E)) = \sup_E |\mu(E) - \nu(E)| \ .$$

It is also well-known that the total variation distance between two distributions $\mu, \nu \in \mathbb{P}(Y)$ satisfies $\mathsf{TV}(\mu, \nu) \leq \mathsf{Pr}_\pi[y \neq y']$ for any coupling π, where equality is attained by taking the maximal coupling, which we show how to construct next.

Since throughout the paper we will only need couplings for distributions with finite support, we give the construction of the maximal coupling only for distributions assigning mass to atomic sets. The extension to the general case where atoms have measure zero is straightforward but requires introducing densities. Suppose $\nu, \nu' \in \mathbb{P}(Y)$ are distributions with finite support. Let $\nu_0(y) = \min\{\nu(y), \nu'(y)\}$ and let $\eta = \mathsf{TV}(\nu, \nu') = 1 - \sum_{y \in Y} \nu_0(y)$, where TV denotes the total variation distance. The maximal coupling between ν and ν' is defined as the mixture $\pi = (1-\eta)\pi_0 + \eta\pi_1$, where $\pi_0(y, y') = \nu_0(y)\mathbb{1}[y = y']/(1-\eta)$, and $\nu_1(y, y') = (\nu(y) - \nu_0(y))(\nu'(y') - \nu_0(y'))/\eta$. Projecting the maximal coupling along the marginals yields the overlapping mixture decompositions $\nu = (1-\eta)\nu_0 + \eta\nu_1$ and $\nu' = (1-\eta)\nu_0 + \eta\nu'_1$.

3. Privacy Profiles

In this section we introduce the first contribution of this paper: the identification of *privacy profiles* as an object that fully captures the privacy properties of a given mechanism. The profile of a mechanism is defined as the curve of all (ε, δ)-DP guarantees satisfied by the mechanism. After formally defining this object and establishing some of its basic properties, we calculate the privacy profile of some well-known mechanisms. Then we derive a connection between privacy profiles and other definitions of differential privacy, and provide an application to bounding the RDP guarantees of any pure DP mechanism.

3.1. **Definition of Privacy Profiles.** Let $\mathcal{M} : X \to \mathbb{P}(Z)$ be a mechanism over an input set X equipped with a neighboring relation $\simeq$. The *privacy profile* of $\mathcal{M}$ is the function $\delta_{\mathcal{M}} : [0, \infty) \to [0, 1]$ given by

$$\delta_{\mathcal{M}}(\varepsilon) = \sup_{x \simeq x'} D_{e^\varepsilon}(\mathcal{M}(x) \| \mathcal{M}(x')) \ . \tag{3.1}$$

By the connection between differential privacy and hockey-stick divergences we see that the privacy profile $\delta_{\mathcal{M}}$ is a function associating to each privacy parameter ε the best possible δ that can be achieved under this ε. In particular, we have that $\mathcal{M}$ is $(\varepsilon, \delta_{\mathcal{M}}(\varepsilon))$-DP for any $\varepsilon \geq 0$. When the mechanism is clear from the context, we shall just write $\delta(\varepsilon)$.

The notion of privacy profile emphasizes a functional view on the privacy guarantees of a mechanism, and it helps highlight the fact that any mechanism satisfies a full curve of privacy

guarantees. This in contrast with the traditional pointwise approach to quantify the privacy of a mechanism, which usually provides a single point in this curve. We can also see this curve as separating the privacy parameters which are valid for a given mechanism from those which are not. In particular, recall that an (ε, δ)-DP mechanism $\mathcal{M}$ is also (ε', δ')-DP for any $\varepsilon' \geq \varepsilon$ and any $\delta' \geq \delta$. Thus, the privacy profile $\delta_{\mathcal{M}}(\varepsilon)$ defines a curve in $[0, \infty) \times [0, 1]$ that separates the space of privacy parameters into two regions: the ones above the curve, for which $\mathcal{M}$ satisfies differential privacy, and the ones below it, for which it does not. This curve exists for every mechanism $\mathcal{M}$, even for mechanisms that satisfy pure DP for some value of ε.

The idea that it is possible to trade-off between the ε and δ guarantees in differentially private mechanisms is not new. For example, this phenomenon can be observed in the analysis of advanced composition and the Gaussian mechanism. These two results have in common that they are both usually proved using an argument based on the privacy loss random variable, which is tightly connected to Rényi differential privacy. We will see in Section 3.4 that this is not a coincidence, and that in fact the points of view of privacy loss random variables and Rényi DP are in some sense equivalent to privacy profiles.

By replacing the standard neighboring relation $\simeq$ in the definition of privacy profile with its k-fold extension $\simeq^k$ we can also define *group privacy profiles* for a mechanism $\mathcal{M}$. For $k \geq 1$, the k-group privacy profile $\delta_{\mathcal{M},k}(\varepsilon)$ is defined as

$$\delta_{\mathcal{M},k}(\varepsilon) = \sup_{x \simeq^k x'} D_{e^\varepsilon}(\mathcal{M}(x) \| \mathcal{M}(x')) \ . \tag{3.2}$$

Note that for $k = 1$ we recover the standard privacy profile, i.e.,$\delta_{\mathcal{M},1}(\varepsilon) = \delta_{\mathcal{M}}(\varepsilon)$. In addition, the standard analysis of group privacy in the context of differential privacy yields a bound for group privacy profiles in terms of the standard privacy profile:

$$\delta_{\mathcal{M},k}(\varepsilon) \leq \frac{e^\varepsilon - 1}{e^{\varepsilon/k} - 1} \delta_{\mathcal{M}}(\varepsilon/k) \ . \tag{3.3}$$

We will see in the examples below that this black-box bound can be improved to get tighter group privacy profiles when considering specific mechanisms. Group privacy profiles will also play an important role on the results on privacy amplification by subsampling with replacement that we study in Section 4.4.

3.2. **Examples of Privacy Profiles.** We proceed to illustrate the concept of privacy profile by computing it for three well-known mechanism: randomized response, Laplace output perturbation and Gaussian output perturbation. The resulting profiles are plotted in Figure 1 for some choice of each mechanism's parameters.

We start with the following result, which gives an expression for the privacy profile of a binary randomized response mechanism.

Theorem 2 . *Let $X = \{0, 1\}$ be equipped with the trivial relation $0 \simeq 1$. Given $p \in [1/2, 1]$, let $\mathcal{M} : X \to \mathbb{P}(X)$ be the randomized response mechanism that on input x returns x with probability p and $1 - x$ with probability $1 - p$. The privacy profile of $\mathcal{M}$ is given by*

$$\delta_{\mathcal{M}}(\varepsilon) = [p - e^\varepsilon (1 - p)]_+ \ .$$

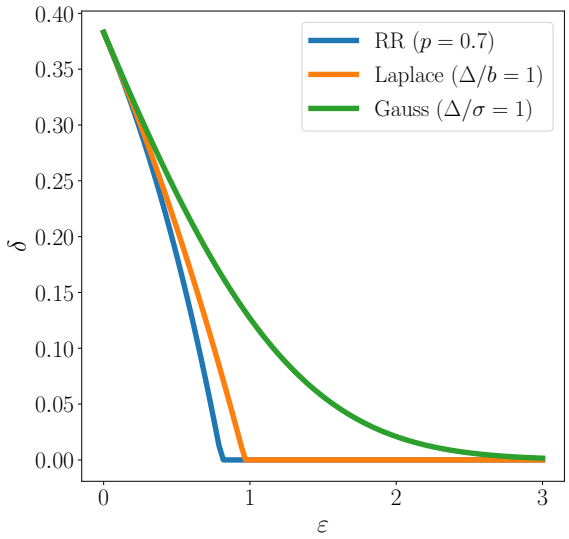


FIGURE 1. Privacy profiles with mechanisms calibrated to provide the same δ at $\varepsilon = 0$. Profile expressions are given in Theorem 2 (Randomized Response), Theorem 3 (Laplace), and Theorem 4 (Gauss).

Proof. By symmetry, we just need to compute $D_{e^\varepsilon}(\mathcal{M}(0)\|\mathcal{M}(1))$. Expanding the definition of hockey-stick divergence we get

$$\begin{aligned} D_{e^\varepsilon}(\mathcal{M}(0)\|\mathcal{M}(1)) &= [\mathsf{Pr}[\mathcal{M}(0)=0] - e^\varepsilon \mathsf{Pr}[\mathcal{M}(1)=0]]_+ + [\mathsf{Pr}[\mathcal{M}(0)=1] - e^\varepsilon \mathsf{Pr}[\mathcal{M}(1)=1]]_+ \\ &= [p - e^\varepsilon(1-p)]_+ + [(1-p) - e^\varepsilon p]_+ \\ &= [p - e^\varepsilon(1-p)]_+ \ , \end{aligned}$$

where the last step used that $(1-p) - e^\varepsilon p \leq 0$ since $p \geq 1/2$. □

Note that the profile of the randomized response mechanism is linear in e^ε in the regime $0 \leq \varepsilon \leq \log \frac{1-p}{p}$, and it saturates to zero for $\varepsilon \geq \log \frac{1-p}{p}$, which is expected since we know that the mechanism satisfies local $(\varepsilon, 0)$-DP for such privacy parameters.

Next we perform the calculations to obtain the privacy profile of the 1-dimensional Laplace output perturbation mechanism.

Theorem 3 . *Let $f : X \to \mathbb{R}$ be a function with global sensitivity $\Delta = \sup_{x \simeq x'} |f(x) - f(x')|$. Suppose $\mathcal{M}(x) = f(x) + \mathsf{Lap}(b)$ is a Laplace output perturbation mechanism with noise parameter b. The privacy profile of $\mathcal{M}$ is given by*

$$\delta_{\mathcal{M}}(\varepsilon) = \left[1 - e^{\frac{\varepsilon}{2} - \frac{\Delta}{2b}}\right]_+ \ .$$

Proof. Suppose $x \simeq x'$ and assume without loss of generality that $y = f(x) = 0$ and $y' = f(x) = \Delta > 0$. Plugging the density of the Laplace distribution in the definition of α-divergence we get

$$D_{e^\varepsilon}(\mathsf{Lap}(b)\|\Delta + \mathsf{Lap}(b)) = \frac{1}{2b}\int_{\mathbb{R}} \left[e^{-\frac{|z|}{b}} - e^\varepsilon e^{-\frac{|z-\Delta|}{b}}\right]_+ dz \ .$$

Now we observe that the quantity inside the integral above is positive if and only if $|z-\Delta| - |z| \geq \varepsilon b$. Since $||z+\Delta| - |z|| \leq \Delta$, we see that the divergence is zero for $\varepsilon > \Delta/b$. On the

other hand, for $\varepsilon \in [0, \Delta/b]$ we have $\{z : |z-\Delta| - |z| \geq \varepsilon b\} = (-\infty, (\Delta - \varepsilon b)/2]$. Thus, we have

$$\frac{1}{2b}\int_{\mathbb{R}} \left[e^{-\frac{|z|}{b}} - e^{\varepsilon} e^{-\frac{|z-\Delta|}{b}}\right]_+ dz = \frac{1}{2b}\int_{-\infty}^{(\Delta-\varepsilon b)/2} e^{-\frac{|z|}{b}} dz - \frac{e^{\varepsilon}}{2b}\int_{-\infty}^{(\Delta-\varepsilon b)/2} e^{-\frac{|z-\Delta|}{b}} dz \ .$$

Now we can compute both integrals as probabilities under the Laplace distribution:

$$\begin{aligned}
\frac{1}{2b}\int_{-\infty}^{(\Delta-\varepsilon b)/2} e^{-\frac{|z|}{b}} dz &= \mathsf{Pr}\left[\mathsf{Lap}(b) \leq \frac{\Delta - \varepsilon b}{2}\right] \\
&= 1 - \frac{1}{2}\exp\left(\frac{\varepsilon b - \Delta}{2b}\right) \ , \\
\frac{e^{\varepsilon}}{2b}\int_{-\infty}^{(\Delta-\varepsilon b)/2} e^{-\frac{|z-\Delta|}{b}} dz &= e^{\varepsilon}\mathsf{Pr}\left[\mathsf{Lap}(b) \leq \frac{-\Delta - \varepsilon b}{2}\right] \\
&= \frac{e^{\varepsilon}}{2}\exp\left(\frac{-\varepsilon b - \Delta}{2b}\right) \ .
\end{aligned}$$

Putting these two quantities together we finally get, for $\varepsilon \leq \Delta/b$:

$$D_{e^{\varepsilon}}(\mathsf{Lap}(b) \| \Delta + \mathsf{Lap}(b)) = 1 - \exp\left(\frac{\varepsilon}{2} - \frac{\Delta}{2b}\right) \ .$$

□

The well-known fact that the Laplace mechanism with $b \geq \Delta/\varepsilon$ is $(\varepsilon, 0)$-DP follows from this result by noting that $\delta_{\mathcal{M}}(\varepsilon) = 0$ for any $\varepsilon \geq \theta$. However, Theorem 3 also provides more information: it shows that for $\varepsilon < \Delta/b$ the Laplace mechanism with noise parameter b satisfies (ε, δ)-DP with $\delta = \delta_{\mathcal{M}}(\varepsilon)$. In this regime the profile is linear in $\sqrt{e^{\varepsilon}}$.

For mechanisms that only satisfy approximate DP, the privacy profile provides information about the behaviour of $\delta_{\mathcal{M}}(\varepsilon)$ as we increase $\varepsilon \to \infty$. The classical analysis for the Gaussian output perturbation mechanism provides some information in this respect. Recall that for a function $f : X \to \mathbb{R}^d$ with L_2 global sensitivity $\Delta = \sup_{x \simeq x'} \|f(x) - f(x)\|_2$ the mechanism $\mathcal{M}(x) = f(x) + \mathcal{N}(0, \sigma^2 I)$ satisfies (ε, δ)-DP if $\sigma^2 \geq 2\Delta^2 \log(1.25/\delta)/\varepsilon^2$ and $\varepsilon \in (0, 1)$ (*cf.* (Dwork and Roth, 2014, Theorem A.1)). This can be rewritten as $\delta_{\mathcal{M}}(\varepsilon) \leq 1.25 e^{-\varepsilon^2\sigma^2/2\Delta^2}$ for $\varepsilon \in (0, 1)$. Recently, (Balle and Wang, 2018) gave a tight analysis of the Gaussian mechanism that is valid for all values of ε. Their analysis can be interpreted as providing an expression for the privacy profile of the Gaussian mechanism in terms of the CDF of a standard normal distribution $\Phi(t) = \mathsf{Pr}[\mathcal{N}(0,1) \leq t] = (1/sqrt(2\pi)) \int_{-\infty}^{t} e^{-r^2/2} dr$.

Theorem 4 (Balle and Wang, 2018)**.** *Let $f : X \to \mathbb{R}^d$ be a function with L_2 global sensitivity Δ. The privacy profile of the Gaussian mechanism $\mathcal{M}(x) = f(x) + \mathcal{N}(0, \sigma^2 I)$ is given by*

$$\delta_{\mathcal{M}}(e^{\varepsilon}) = \Phi\left(\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) - e^{\varepsilon}\Phi\left(-\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) \ .$$

Although the form of the exact privacy profile of the Gaussian mechanism is not very appealing from the point of view of asymptotic bounds, accurate implementations of the CDF of standard normal distributions are widely available. Such implementations can be used not only to evaluate the profile, but also to find the smallest value of σ providing a desired (ε, δ)-DP guarantee (see (Balle and Wang, 2018) for more details). Here we illustrate the importance of using the exact profile of the Gaussian mechanism versus the one obtained

from the bound in (Dwork and Roth, 2014, Theorem A.1) by plotting the exact (analytic) profile and its classical approximation for two values of Δ/σ in Figure 2. We observe that the difference between the exact and approximate profiles is larger for smaller values of ε and larger values of Δ/σ.

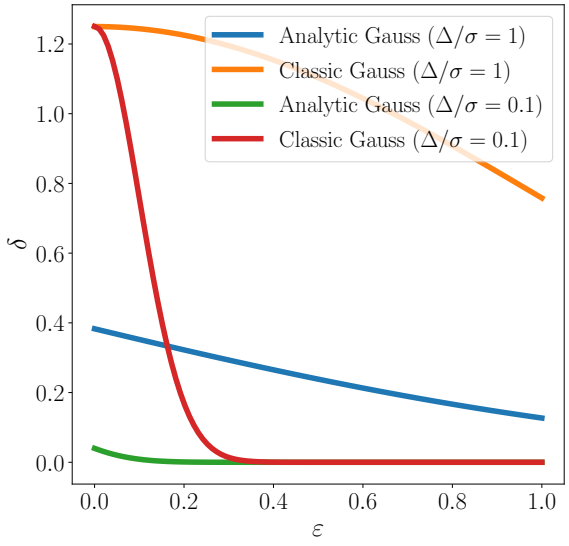
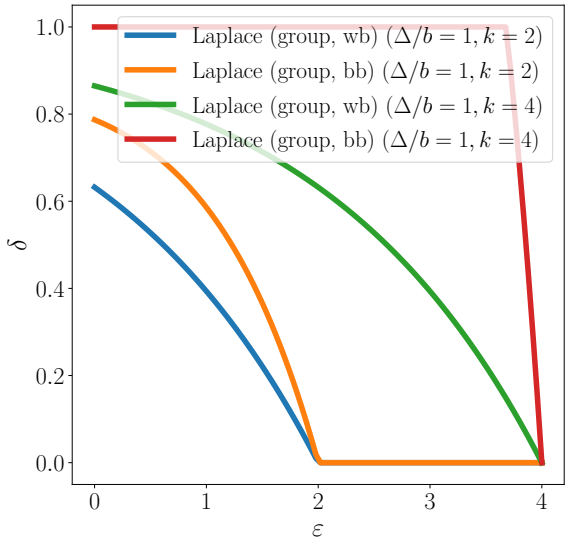


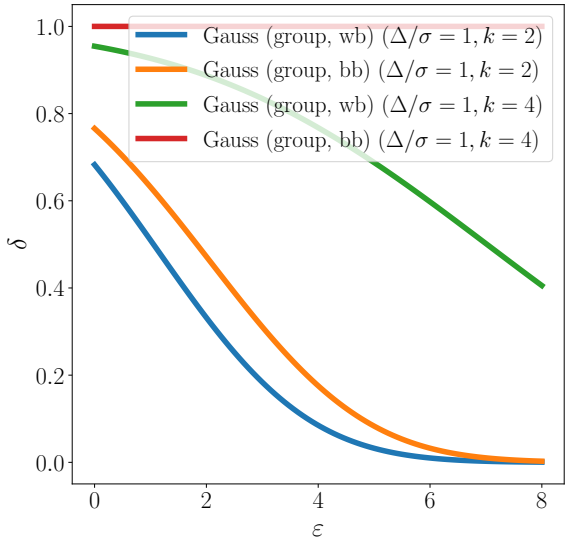
FIGURE 2. Comparison between exact and approximate privacy profiles for the Gaussian mechanism.

To conclude this section we will show that, at least in the case of output perturbation mechanisms, having access to the exact profile allows one to obtain better bounds on group privacy profiles than the bound provided in (3.3). We call group privacy profiles obtained with the former method *white-box*, while the ones obtained by the latter method we shall call *black-box*. To obtain white-box bounds on the group privacy profiles of output perturbation mechanism all we need is to observe that by the triangle inequality, the global sensitivity of a function f with respect to the k-fold relation $\simeq_k$ is upper bounded by $k\Delta$, where Δ is the standard global sensitivity. Thus, substituting Δ by $k\Delta$ in the profiles given in Theorem 3 and 4 we obtain upper bounds on the group privacy profiles of the Laplace and Gaussian mechanisms. The plots in Figure 3 compare the white-box and black-box bounds for these two mechanisms for different values of k, where we observe that white-box bounds are uniformly better than black-box ones, and that the discrepancies between both bounds increase with k.

3.3. **Properties of Privacy Profiles.** Privacy profiles enjoy a number of properties. On the one hand, standard properties of differential privacy can be translated into properties of privacy profiles. The bound (3.3) on group privacy profiles is an example; the same exercise can be performed for other composition properties. Other properties of privacy profiles follow directly from its definition as a supremum over the ϕ-divergences D_{e^ε}. For example, the processing inequality for ϕ-divergences provides a direct way to establish the post-processing property of differential privacy. The joint convexity property of ϕ-divergences also implies



(A) Laplace mechanism.



(B) Gaussian mechanism.

FIGURE 3. Comparison between white-box and black-box group privacy profiles.

the bound

$$\delta_{\mathcal{M}}(\varepsilon) \leq (1-\gamma)\delta_{\mathcal{M}_1}(\varepsilon) + \gamma\delta_{\mathcal{M}_2}(\varepsilon)$$

for a mixture mechanism $\mathcal{M}$ that on input x returns $\mathcal{M}_1(x)$ with probability γ and $\mathcal{M}_2(x)$ with probability $1-\gamma$, where $\mathcal{M}_1$ and $\mathcal{M}_2$ are two arbitrary mechanisms. We shall see in Section 4 that a generalization of this property specific to hockey-stick divergences plays a crucial role in obtaining tight bounds for privacy amplification under subsampling. Finally, another set of properties of privacy profiles relate to the shape of the profile as we change

the privacy parameter ε. These properties arise from the fact that we are jointly considering a family of ϕ-divergences, and are therefore specific to privacy profiles. In the remaining of this section we establish an important property of this type.

The property we want to show is a convexity-like restriction on the shape of the privacy profile $\delta_{\mathcal{M}}(\varepsilon)$ of any mechanism $\mathcal{M}$. Before stating and proving this property we recall that we already know another restriction that privacy profiles must satisfy: they are monotonically decreasing in ε. This is a quite intuitive fact: since (ε, δ)-DP mechanism is also (ε', δ')-DP for any $\varepsilon' \geq \varepsilon$ and $\delta' \geq \delta$, it must be the case that for any mechanism $\mathcal{M}$ and $\varepsilon' \geq \varepsilon$ we have $\delta_{\mathcal{M}}(\varepsilon) \geq \delta_{\mathcal{M}}(\varepsilon')$. This property is easily visualized in all the profiles plotted in Section 3.2. A straightforward analytical proof of this fact can be obtained by plugging the inequality $[u - e^{\varepsilon'}]_+ \leq [u - e^{\varepsilon}]_+$ for $\varepsilon' \geq \varepsilon$ in the definition of hockey-stick divergence.

As the examples with the Laplace and randomized response show, this monotonicity is not strict, as the profile of a pure DP mechanism will plateau at $\delta = 0$. For other mechanisms like Gaussian output perturbation, the profile monotonically decreases towards zero but does not attain the limit for any finite ε. Furthermore, one can construct mechanisms where the privacy privacy plateaus at a value $\delta > 0$; e.g., the mechanism that given a database with n outputs a record selected uniformly at random has constant profile $\delta(\varepsilon) = 1/n$.

The main result of this section shows that the structure of privacy profiles exhibits some additional rigidity beyond monotonicity. To state and prove this result it will be convenient to re-parametrize the privacy profile in terms of e^{ε} instead of ε. In consequence, we define the re-parametrized profile $\tilde{\delta}_{\mathcal{M}}(\beta) = \delta_{\mathcal{M}}(\log \beta)$, where now $\tilde{\delta}_{\mathcal{M}}$ is a function with inputs in $[1, \infty)$.

Theorem 5 . *The re-parametrized profile $\tilde{\delta}_{\mathcal{M}} : [1, \infty) \to [0, 1]$ of any mechanism $\mathcal{M}$ is a convex function. In particular, for any β_1, β_2 and $\gamma \in [0, 1]$ we have*

$$\tilde{\delta}_{\mathcal{M}}(\gamma\beta_1 + (1-\gamma)\beta_2) \leq \gamma\tilde{\delta}_{\mathcal{M}}(\beta_1) + (1-\gamma)\tilde{\delta}_{\mathcal{M}}(\beta_2) \ .$$

Proof. First we use that the function $[\bullet]_+$ is subadditive and positive-homogeneous to show:

$$\left[\frac{p(z)}{q(z)} - (\gamma\beta_1 + (1-\gamma)\beta_2)\right]_+ \leq \gamma\left[\frac{p(z)}{q(z)} - \beta_1\right]_+ + (1-\gamma)\left[\frac{p(z)}{q(z)} - \beta_2\right]_+ \ .$$

Let $\beta = \gamma\beta_1 + (1-\gamma)\beta_2$. Plugging this bound into the definition of hockey-stick divergences and re-parametrized privacy profiles we get

$$\sup_{x \simeq x'} D_{\beta}(\mathcal{M}(x) \| \mathcal{M}(x')) \leq \gamma \sup_{x \simeq x'} D_{\beta_1}(\mathcal{M}(x) \| \mathcal{M}(x')) + (1-\gamma) \sup_{x \simeq x'} D_{\beta_2}(\mathcal{M}(x) \| \mathcal{M}(x')) \ .$$

□

In terms of standard differential privacy guarantees, the above theorem provides information allowing us to interpolate between two distinct privacy guarantees. In particular, the theorem can be rephrased as follows: if a mechanism is $(\varepsilon_1, \delta_1)$-DP and $(\varepsilon_2, \delta_2)$-DP, then the mechanism is also (ε', δ')-DP with $\varepsilon' = \log(\gamma e^{\varepsilon_1} + (1-\gamma)e^{\varepsilon_2})$ and $\delta' = \gamma\delta_1 + (1-\gamma)\delta_2$ for any $\gamma \in (0, 1)$. We observe that this bound is tight since for a randomized response mechanism that replies truthfully with probability p we get equality for any $\varepsilon_1, \varepsilon_2 \in [0, \log \frac{1-p}{p}]$ (*cf.* Theorem 2).

3.4. **Unification Theorems.** The purpose of this section is to sketch a unified view of differential privacy by showing that, in some sense, privacy profiles, privacy loss random variables, and Rényi DP all contain the same information about the privacy provided by a mechanism.

The first observation in this respect connects privacy profiles and privacy loss random variables. Incidentally, this connection was obtained by Balle and Wang (2018) in their proof of Theorem 4, albeit the original formulation was not in terms of privacy profiles. Re-formulated in our terms, the result reads as follows.

Theorem 6 (Balle and Wang, 2018)**.** *The privacy profile of a mechanism $\mathcal{M}$ satisfies*

$$\delta_{\mathcal{M}}(\varepsilon) = \sup_{x \simeq x'} \left(\mathsf{Pr}[\mathsf{L}_{\mathcal{M}}^{x,x'} > \varepsilon] - e^{\varepsilon} \mathsf{Pr}[\mathsf{L}_{\mathcal{M}}^{x',x} < -\varepsilon] \right) \ .$$

The characterization above generalizes the well-known sufficient condition for differential privacy in terms of the tail of the privacy loss random variable, which in our notation is expressed by the inequality $\delta_{\mathcal{M}}(\varepsilon) \leq \sup_{x \simeq x'} \mathsf{Pr}[\mathsf{L}_{\mathcal{M}}^{x,x'} > \varepsilon]$ (see Section 2.3). For the sake of completeness we now present a proof of this theorem.

Proof. Let μ and ν be probability distributions with respective densities p and q with respect to some base measure λ. The result will follow if we show the identity

$$D_{e^\varepsilon}(\mu \| \nu) = \mathsf{Pr}[\mathsf{L} > \varepsilon] - e^{\varepsilon} \mathsf{Pr}[\mathsf{L}' < -\varepsilon] \ ,$$

where $\mathsf{L} = p(\mathsf{Z})/q(\mathsf{Z})$ with $\mathsf{Z} \sim \mu$ and $\mathsf{L}' = q(\mathsf{Z}')/p(\mathsf{Z}')$ with $\mathsf{Z}' \sim \nu$. By expanding the definition of hockey-stick divergence and the identity $[u]_+ = u\mathbb{I}[u > 0]$, we get

$$\begin{aligned}
D_{e^\varepsilon}(\mu \| \nu) &= \int [p(z) - e^{\varepsilon} q(z)]_+ d\lambda(z) \\
&= \int (p(z) - e^{\varepsilon} q(z)) \mathbb{I}[p(z) > e^{\varepsilon} q(z)] d\lambda(z) \\
&= \int \mathbb{I}[p(z) > e^{\varepsilon} q(z)] p(z) d\lambda(z) - e^{\varepsilon} \int \mathbb{I}[p(z) > e^{\varepsilon} q(z)] q(z) d\lambda(z) \\
&= \int \mathbb{I}[p(z) > e^{\varepsilon} q(z)] d\mu(z) - e^{\varepsilon} \int \mathbb{I}[p(z) > e^{\varepsilon} q(z)] d\nu(z) \\
&= \mathsf{Pr}\left[\frac{p(\mathsf{Z})}{q(\mathsf{Z})} > e^{\varepsilon} \right] - e^{\varepsilon} \mathsf{Pr}\left[\frac{q(\mathsf{Z}')}{p(\mathsf{Z}')} < e^{-\varepsilon} \right] \ .
\end{aligned}$$

□

We see from the proof of Theorem 6 that, for any value of $\varepsilon \geq 0$, the divergence $D_{e^\varepsilon}(\mathcal{M}(x) \| \mathcal{M}(x'))$ can be recovered from the distributions of the privacy loss random variables $\mathsf{L}_{\mathcal{M}}^{x,x'}$ and $\mathsf{L}_{\mathcal{M}}^{x',x}$. In particular, full knowledge of the distributions of these random variables is enough to compute the privacy profile of $\mathcal{M}$. The next result shows that the divergences $D_{e^\varepsilon}(\mathcal{M}(x) \| \mathcal{M}(x'))$ also contain full knowledge about the distribution of the privacy loss random variables, thus showing that both points of view are equivalent.

Theorem 7 . *Let μ and ν be probability distributions with respective densities p and q with respect to some base measure λ and let $\mathsf{Z}' \sim \nu$. The right-derivative*[3] *of $D_\beta(\mu \| \nu)$ with respect*

[3]The right-derivative of a real-valued function $f(t)$ at t_0 is defined as $\partial_+ f(t_0) = \lim_{t \downarrow t_0} \frac{f(t) - f(t_0)}{t - t_0}$, where $t \downarrow t_0$ denotes limit of $t \in (t_0, \infty)$ approaching t_0. Right-derivatives exist for any convex function (see, e.g., (Liese and Vajda, 2006)).

to β satisfies

$$\frac{\partial_+}{\partial\beta} D_\beta(\mu\|\nu) = -\mathsf{Pr}\left[\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')} > \beta\right] \ .$$

Proof. For any fixed $u \in \mathbb{R}$, we have $\frac{\partial_+}{\partial\beta}[u-\beta]_+ = -\mathbb{I}[u > \beta]$. Therefore,

$$\begin{aligned}
\frac{\partial_+}{\partial\beta} D_\beta(\mu\|\nu) &= \frac{\partial_+}{\partial\beta} \int \left[\frac{p(z)}{q(z)} - \beta\right]_+ d\nu(z) \\
&= \int \frac{\partial_+}{\partial\beta} \left[\frac{p(z)}{q(z)} - \beta\right]_+ d\nu(z) \\
&= -\int \mathbb{I}\left[\frac{p(z)}{q(z)} > \beta\right]_+ d\nu(z) \\
&= -\mathsf{Pr}\left[\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')} > \beta\right] \ .
\end{aligned}$$

□

Now we turn to the connections with Rényi differential privacy. Since the moment generating function of a random variable completely characterizes its distribution, we see that knowing the divergences $R_\alpha(\mathcal{M}(x)\|\mathcal{M}(x'))$ for all $\alpha \in (1,\infty)$ and $x \simeq x'$ is equivalent to knowing the distributions of $\mathsf{L}_{\mathcal{M}}^{x,x'}$. In this sense, we already see that the Rényi divergences used to define RDP contain the same information as the hockey-stick divergences used to define DP. The following result provides an integral representation which makes this relation more tangible by expressing the moment generating function in terms of hockey-stick divergences.

Theorem 8 . *Let μ and ν be probability distributions with respective densities p and q with respect to some base measure λ. Assume that $\mathsf{Z} \sim \mu$ and let $\mathsf{L} = \log(p(\mathsf{Z})/q(\mathsf{Z}))$. For $s > 0$, let $\varphi(s) = \mathsf{E}[e^{s\mathsf{L}}]$ be the moment generating function of L. Then we have*

$$\varphi(s) = 1 + s(s+1)\int_0^\infty \left(e^{s\varepsilon} D_{e^\varepsilon}(\mu\|\nu) + e^{-(s+1)\varepsilon} D_{e^\varepsilon}(\nu\|\mu)\right) d\varepsilon \ .$$

Proof. Recall that for any non-negative random variable X one has $\mathsf{E}[\mathsf{X}] = \int_0^\infty \mathsf{Pr}[\mathsf{X} > t]dt$. We use this to write the moment generating function of the corresponding privacy loss random variable for $s \geq 0$ as follows:

$$\varphi(s) = \int_0^\infty \mathsf{Pr}[e^{s\mathsf{L}} > t]dt = \int_0^\infty \mathsf{Pr}\left[\frac{p(\mathsf{Z})}{q(\mathsf{Z})} > t^{1/s}\right] dt \ .$$

Next we observe the probability inside the integral above can be decomposed in terms of a divergence and a second integral with respect to ν:

$$\begin{aligned}
\mathsf{Pr}\left[\frac{p(\mathsf{Z})}{q(\mathsf{Z})} > t^{1/s}\right] &= \mathsf{Pr}[p(\mathsf{Z}) > t^{1/s} q(\mathsf{Z})] \\
&= \int \mathbb{I}[p(z) > t^{1/s} q(z)] d\mu(z) \\
&= \int \mathbb{I}[p(z) > t^{1/s} q(z)] p(z) d\lambda(z) \\
&= \int \mathbb{I}[p(z) > t^{1/s} q(z)](p(z) - t^{1/s} q(z)) d\lambda(z) + t^{1/s} \int \mathbb{I}[p(z) > t^{1/s} q(z)] q(z) d\lambda(z) \\
&= \int [p(z) - t^{1/s} q(z)]_+ d\lambda(z) + t^{1/s} \int \mathbb{I}[p(z) > t^{1/s} q(z)] q(z) d\lambda(z) \\
&= D_{t^{1/s}}(\mu \| \mu') + t^{1/s} \int \mathbb{I}[p(z) > t^{1/s} q(z)] d\nu(z) \\
&= D_{t^{1/s}}(\mu \| \mu') + t^{1/s} \mathsf{Pr}\left[\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')} > t^{1/s}\right] \ ,
\end{aligned}$$

where $\mathsf{Z}' \sim \nu$. Note the term $D_{t^{1/s}}(\mu \| \mu')$ above is not a divergence in the sense of Csiszár when $t^{1/s} < 1$. Nonetheless, integrating with respect to t we get an expression for $\varphi(s)$ involving two terms that we will need to massage further:

$$\varphi(s) = \int_0^\infty D_{t^{1/s}}(\mu \| \mu') dt + \int_0^\infty t^{1/s} \mathsf{Pr}\left[\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')} > t^{1/s}\right] dt \ .$$

To compute the second integral in the RHS above we perform the change of variables $dt' = t^{1/s} dt$, which comes from taking $t' = t^{1+1/s}/(1+1/s)$, or, equivalently, $t = ((1+1/s)t')^{1/(1+1/s)}$. This allows us to relate this integral to $\varphi(s)$ as follows:

$$\begin{aligned}
\int_0^\infty t^{1/s} \mathsf{Pr}\left[\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')} > t^{1/s}\right] dt &= \int_0^\infty \mathsf{Pr}\left[\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')} > ((1+1/s)t')^{1/(s+1)}\right] dt' \\
&= \int_0^\infty \mathsf{Pr}\left[\frac{s}{s+1}\left(\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')}\right)^{s+1} > t'\right] dt' \\
&= \frac{s}{s+1} \mathsf{E}\left[\left(\frac{p(\mathsf{Z}')}{q(\mathsf{Z}')}\right)^{s+1}\right] \\
&= \frac{s}{s+1} \int \left(\frac{p(z)}{q(z)}\right)^{s+1} d\nu(z) \\
&= \frac{s}{s+1} \int \left(\frac{p(z)}{q(z)}\right)^{s} d\mu(z) \\
&= \frac{s}{s+1} \varphi(s) \ .
\end{aligned}$$

Putting the derivations above together we see that

$$\varphi(s) = (s+1) \int_0^\infty D_{t^{1/s}}(\mu \| \nu) dt \ .$$

Now we observe that some terms in the integral above do not correspond to a hockey-divergence between μ and ν, e.g., for $t \in (0,1)$ the term $D_{t^{1/s}}(\mu\|\nu)$ is not a divergence. Instead, using the definition of $D_{t^{1/s}}(\mu\|\nu)$ we can see that these terms are equal to $1 - t^{1/s} + t^{1/s} D_{t^{-1/s}}(\nu\|\mu)$, where the last term is now a divergence. Thus, we split the integral in the expression for $\varphi(s)$ into two parts and obtain

$$\begin{aligned}\varphi(s) &= (s+1)\int_0^1 \left(1 - t'^{1/s} + t'^{1/s} D_{t'^{-1/s}}(\nu\|\mu)\right) dt' + (s+1)\int_1^\infty D_{t^{1/s}}(\mu\|\nu)dt \\ &= 1 + (s+1)\int_0^1 t'^{1/s} D_{t'^{-1/s}}(\nu\|\mu)dt' + (s+1)\int_1^\infty D_{t^{1/s}}(\mu\|\nu)dt \ .\end{aligned}$$

Finally, we can obtain the desired equation by performing a series of simple changes of variables $t' = 1/t$, $\beta = t^{1/s}$, and $\beta = e^{\varepsilon}$:

$$\begin{aligned}\varphi(s) &= 1 + (s+1)\int_1^\infty t^{-2-1/s} D_{t^{1/s}}(\nu\|\mu)dt + (s+1)\int_1^\infty D_{t^{1/s}}(\mu\|\nu)dt \\ &= 1 + s(s+1)\int_1^\infty \left(\beta^{s-1} D_\beta(\mu\|\nu) + \beta^{-s-2} D_\beta(\nu\|\mu)\right) d\beta \\ &= 1 + s(s+1)\int_0^\infty \left(e^{s\varepsilon} D_{e^\varepsilon}(\mu\|\nu) + e^{-(s+1)\varepsilon} D_{e^\varepsilon}(\nu\|\mu)\right) d\varepsilon \ .\end{aligned}$$

□

As a direct consequence of this result we observe that the RDP guarantees of any mechanism $\mathcal{M}$ can be bounded in terms of its privacy profile. In particular, we can express the full *Rényi privacy profile*

$$\epsilon_{\mathcal{M}}(\alpha) = \sup_{x \simeq x'} R_\alpha(\mathcal{M}(x)\|\mathcal{M}(x'))$$

as a function of the privacy profile as follows.

Corollary 9 . *For any mechanism $\mathcal{M}$ we have*

$$\epsilon_{\mathcal{M}}(\alpha) \le \frac{1}{\alpha - 1} \log\left(1 + \alpha(\alpha-1)\int_0^\infty \left(e^{(\alpha-1)\varepsilon} + e^{-\alpha\varepsilon}\right)\delta_{\mathcal{M}}(\varepsilon)d\varepsilon\right) \ . \tag{3.4}$$

We note that a weaker result for converting privacy profiles with a fixed parametric form into zCDP guarantees was given in (Bun and Steinke, 2016, Lemma 3.7).

3.5. **Application: RDP of Pure DP Mechanisms.** We now provide an application of the results from the last two sections to bound the RDP guarantees of any pure DP mechanism. The interest of this result resides in obtaining tight quantitative bounds on RDP guarantees of a wide family of mechanisms, which can then be used in composition calculations involving a large number of mechanisms to obtain tight privacy guarantees in the spirit of the moments accountant technique of Abadi et al. (2016) (see also (Wang et al., 2019)).

Theorem 10 . *Let $\mathcal{M}$ be a mechanism with privacy profile $\delta_{\mathcal{M}}$. Let $\theta = \delta_{\mathcal{M}}(0) = \sup_{x \simeq x'} \mathsf{TV}(\mathcal{M}(x), \mathcal{M}(x'))$ and suppose there exists $\varepsilon_* > 0$ such that $\delta_{\mathcal{M}}(\varepsilon) = 0$ for $\varepsilon \ge \varepsilon_*$.*

Then the RDP profile of $\mathcal{M}$ satisfies

$$\epsilon_{\mathcal{M}}(\alpha) \leq \frac{1}{\alpha-1}\log\left(1+\theta\left(\frac{e^{\varepsilon_*}+1}{e^{\varepsilon_*}-1}\right)(e^{(\alpha-1)\varepsilon_*}-1)\right) \ .$$

Proof. By the convexity of the re-parametrized privacy profile (Theorem 5) we can assume that $\delta_{\mathcal{M}}(\varepsilon) \leq [a - be^{\varepsilon}]_+$ with $a = \frac{\theta}{1-e^{-\varepsilon_*}}$ and $b = e^{-\varepsilon_*}a$. Now we just need to plug this upper bound into (3.4) and compute the integral to obtain:

$$\begin{aligned} e^{(\alpha-1)\epsilon_{\mathcal{M}}(\alpha)} &\leq \alpha(\alpha-1)\int_0^\infty \left(e^{(\alpha-1)\varepsilon}+e^{-\alpha\varepsilon}\right)\delta_{\mathcal{M}}(\varepsilon)d\varepsilon \\ &\leq \frac{\theta\alpha(\alpha-1)}{1-e^{-\varepsilon_*}}\int_0^{\varepsilon_*}\left(e^{(\alpha-1)\varepsilon}+e^{-\alpha\varepsilon}\right)(1-e^{\varepsilon-\varepsilon_*})d\varepsilon \\ &= \frac{\theta}{1-e^{-\varepsilon_*}}(1+e^{-\varepsilon_*})(e^{(\alpha-1)\varepsilon_*}-1) \ . \end{aligned}$$

□

Note that the bound above satisfies $\lim_{\alpha\to\infty}\epsilon_{\mathcal{M}}(\alpha) = \varepsilon_*$, so the bound is tight for large values of α. In fact, one can also check by comparison with (Mironov, 2017, Proposition 5) that this bound is achieved with equality for randomized response. Other results for converting the guarantees of pure DP mechanism into concentrated-like notions of DP can be found in (Dwork and Rothblum, 2016, Theorem 3.5), (Bun and Steinke, 2016, Proposition 3.3), and (Mironov, 2017, Lemma 1). These bounds generally show that $\epsilon_{\mathcal{M}}(\alpha) = O(\alpha\varepsilon_*^2)$, which is the case when $\alpha\varepsilon_*$ is small. Our bound, albeit more cumbersome, is more accurate and valid for all ranges of parameters, which makes it more suitable for numerical privacy calibration.

4. Privacy Amplification by Subsampling

A well-known method for increasing privacy of a mechanism is to apply the mechanism to a random subsample of the input database, rather than on the database itself. Intuitively, the method decreases the chances of leaking information about a particular individual because nothing about that individual can be leaked in the cases where the individual is not included in the subsample. The question addressed in this section is to devise methods for quantifying amplification and for proving optimality of the bounds. This turns out to be a surprisingly subtle problem.

Formally, a *subsampling mechanism* is a randomized algorithm $\mathcal{S} : X \to \mathbb{P}(Y)$ that takes as input a database x and outputs a finitely supported distribution over datasets. The most common forms of subsampling methods are subsampling with and without replacement and Poisson subsampling. We assume that both X and Y contain databases (modelled as sets, multisets, or tuples) over a universe $\mathcal{U}$ that represents all possible records contained in a database. However, we distinguish between X and Y because the input database $x \in X$ and a possible output $y \in Y$ might not always have the same type. For example, sampling with replacement from a set x yields a multiset y.

Now, consider a mechanism $\mathcal{M} : Y \to \mathbb{P}(Z)$, and define the subsampled mechanism $\mathcal{M}^{\mathcal{S}} : X \to \mathbb{P}(Z)$ by the clause $\mathcal{M}^{\mathcal{S}}(x) = \mathcal{M}(\mathcal{S}(x))$, where the composition notation means we feed a sample from $\mathcal{S}(x)$ into $\mathcal{M}$. Furthermore, assume a neighboring relation $\simeq_Y$ on Y such that $\mathcal{M}$ is (ε, δ)-differentially private with respect to $\simeq_Y$, and a neighboring relation $\simeq_X$ on X. What are the possible values ε' and δ' such that $\mathcal{M}^{\mathcal{S}}$ is (ε', δ')-differentially private

with respect to $\simeq_X$? We are specifically interested in the case where $\varepsilon' \leq \varepsilon$ and $\delta' \leq \delta$. In such cases, the subsampled mechanism has better privacy parameters than the original one, i.e., subsampling amplifies privacy.

We formalize the problem of privacy amplification using privacy profiles. Let X and Y be two sets equipped with neighboring relation $\simeq_X$ and $\simeq_Y$, respectively. Let $\mathcal{M} : Y \to \mathbb{P}(Z)$ be a mechanism with privacy profile $\delta_{\mathcal{M}}$ with respect to $\simeq_Y$, and let $\mathcal{S}$ be a subsampling mechanism. The goal is to relate the privacy profiles of $\mathcal{M}$ and $\mathcal{M}^{\mathcal{S}}$, via an inequality of the form: for every $\varepsilon \geq 0$, there exists $0 \leq \varepsilon' \leq \varepsilon$ such that $\delta_{\mathcal{M}^{\mathcal{S}}}(\varepsilon') \leq h(\delta_{\mathcal{M}}(\varepsilon))$, where h is some function to be determined.

The main challenge in the analysis of privacy amplification by subsampling resides in the fact that the output distribution of the subsampled mechanism $\mu = \mathcal{M}^{\mathcal{S}}(x) \in \mathbb{P}(Z)$ is a *mixture distribution*. In particular, writing $\mu_y = \mathcal{M}(y) \in \mathbb{P}(Z)$ for any $y \in Y$ and taking $\omega = \mathcal{S}(x) \in \mathbb{P}(Y)$ to be the (finitely supported) distribution over subsamples from x produced by the subsampling mechanism, we can write $\mu = \sum_y \omega(y)\mu_y = \omega M$, where M denotes the Markov kernel operating on measures defined by $\mathcal{M}$. Consequently, proving privacy amplification requires reasoning about the mixtures obtained when sampling from two neighboring datasets $x \simeq_X x'$, and how the privacy parameters are affected by the mixture.

Our contribution is to provide a unified method for deriving privacy amplification by subsampling bounds. Our method recovers all existing results in the literature and is useful to derive novel amplification bounds. In most cases our method also provides optimal constants which are shown to be tight by a generic lower bound. Our analysis relies on properties of divergences and privacy profiles, together with two additional ingredients.

The first ingredient is a novel *advanced joint convexity* property that uses ideas from probabilistic couplings, and more specifically the maximal coupling construction, to provide upper bounds on the hockey-stick divergence between overlapping mixture distributions. The second ingredient is a (rather specialized) notion of *distance-compatible coupling*, which we use to establish an upper bound for the divergences obtained by advanced joint convexity in terms of group-privacy profiles.

The combination of these results yields a bound of the privacy profile of $\mathcal{M}^{\mathcal{S}}$ as a function of the group-privacy profiles of $\mathcal{M}$. Based on this inequality, we will establish several privacy amplification result and prove tightness results. This methodology can be applied to any of the settings discussed above in terms of dataset representation, neighboring relation, and type of subsampling. Table 1 summarizes several results that can be obtained with our method, where some of the notation is defined below.

We focus on order-independent representations of datasets without repetitions, i.e., sets or multisets. This is mostly for technical convenience, since all our results also hold if one considers datasets represented as tuples. More specifically, we assume a universe of records $\mathcal{U}$ and let $2 = \{0, 1\}$. We write $2^{\mathcal{U}}$ and $\mathbb{N}^{\mathcal{U}}$ for the spaces of all sets and multisets with records from $\mathcal{U}$. Note every set is also a multiset. For $n \geq 0$ we also write $2_n^{\mathcal{U}}$ and $\mathbb{N}_n^{\mathcal{U}}$ for the spaces of all sets and multisets containing exactly n records[4] from $\mathcal{U}$. Given $x \in \mathbb{N}^{\mathcal{U}}$ we write x_u for the number of occurrences of $u \in \mathcal{U}$ in x. The support of a multiset x is the defined as the set $\mathsf{supp}(x) = \{u \in \mathcal{U} : x_u > 0\}$ of elements that occur at least once in x. Given multisets $x, x' \in \mathbb{N}^{\mathcal{U}}$ we write $x' \subseteq x$ to denote that $x'_u \leq x_u$ for all $u \in \mathcal{U}$.

[4]In the case of multisets records are counted with multiplicity.

Subsampling	$\simeq_Y$	$\simeq_X$	η	δ'	Theorem
Poisson(γ)	R	R	γ	$\gamma\delta$	13
WOR(n,m)	S	S	$\frac{m}{n}$	$\frac{m}{n}\delta$	14
WR(n,m)	S	S	$1-\left(1-\frac{1}{n}\right)^m$	$\sum_{k=1}^m \binom{m}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{m-k}\delta_k$	15
WR(n,m)	S	R	$1-\left(1-\frac{1}{n}\right)^m$	$\sum_{k=1}^m \binom{m}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{m-k}\delta_k$	16

TABLE 1. Summary of privacy amplification bounds. Amplification parameter η: $e^{\varepsilon'} = 1+\eta(e^\varepsilon-1)$. Types of subsampling: without replacement (WOR) and with replacement (WR). Neighboring relations: remove/add-one (R) and substitute one (S).

For multisets, we consider the two following neighboring relations. The *remove/add-one* relation is obtained by letting $x \simeq_r x'$ hold whenever $x \subseteq x'$ with $|x| = |x'|-1$ or $x' \subseteq x$ with $|x| = |x'|+1$; i.e., x' is obtained by removing or adding a single element to x. The *substitute-one* relation is obtained by letting $x \simeq_s x'$ hold whenever $\|x - x'\|_1 = 2$ and $|x| = |x'|$; i.e., x' is obtained by replacing an element in x with a different element from $\mathcal{U}$. Note how $\simeq_r$ relates pairs of datasets with different sizes, while $\simeq_s$ only relates pairs of datasets with the same size.

4.1. **Technical Tools.** We introduce two main technical tools used in our analysis. The first technical tool is a strengthening of joint convexity, which we call advanced joint convexity. This result may be of independent interest.

4.1.1. *Advanced Joint Convexity.* The privacy amplification phenomenon is tightly connected to an interesting new form of joint convexity for hockey-stick divergences, which we call advanced joint convexity.

Theorem 11 Advanced Joint Convexity of D_β**.** *Let $\mu, \mu' \in \mathbb{P}(Z)$ be measures satisfying $\mu = (1-\eta)\mu_0 + \eta\mu_1$ and $\mu' = (1-\eta)\mu_0 + \eta\mu_1'$ for some η, μ_0, μ_1, and μ_1'. Given $\beta \geq 1$, let $\beta' = 1+\eta(\beta-1)$ and $\theta = \beta'/\beta$. Then the following holds:*

$$D_{\beta'}(\mu\|\mu') = \eta D_\beta(\mu_1\|(1-\theta)\mu_0 + \theta\mu_1') \ . \tag{4.1}$$

Proof. Let p and p' be the densities of μ and μ' with respect to some base measure λ. By linearity of the Radon-Nikodym derivative, we have $p = (1-\eta)p_0 + \eta p_1$ and $p' = (1-\eta)p_0 + \eta p_1'$, where p_0, p_1 and p_1' are the respective densities of μ_0, μ_1 and μ_1' with respect to λ. Now note that for every measurable $z \in Z$,

$$[p(z) - \beta' p'(z)]_+ = \eta[p_1(z) - \beta((1-\theta)p_0(z) + \theta p_1'(z))]_+ \ .$$

Plugging this identity in the definition of $D_{\beta'}$, we get the desired equality. □

Note that writing $\beta = e^{\varepsilon}$ and $\beta' = e^{\varepsilon'}$ in the above lemma, we get the relation $\varepsilon' = \log(1 + \eta(e^{\varepsilon} - 1))$. Applying standard joint convexity to the right hand side above we conclude: $D_{\beta'}(\mu\|\mu') \leq (1-\theta)\eta D_{\beta}(\mu_1\|\mu_0) + \theta\eta D_{\beta}(\mu_1\|\mu_1')$. On the other hand, applying joint convexity directly on $D_{\beta'}(\mu\|\mu')$ instead of advanced joint complexity yields a weaker bound which implies amplification for the δ privacy parameter, but not for the ε privacy parameter.

When using advanced joint convexity to analyze privacy amplification we consider two elements x and x' and fix the following notation. Let $\omega = \mathcal{S}(x)$ and $\omega' = \mathcal{S}(x')$ and $\mu = \omega M$ and $\mu' = \omega' M$, where we use the notation M to denote the Markov kernel associated with mechanism $\mathcal{M}$ operating on measures over Y. We then consider the mixture factorization of ω and ω' obtained by taking the decompositions induced by projecting the *maximal coupling* $\pi = (1-\eta)\pi_0 + \eta\pi_1$ on the first and second marginals: $\omega = (1-\eta)\omega_0 + \eta\omega_1$ and $\omega' = (1-\eta)\omega_0 + \eta\omega_1'$, where $\eta = \mathsf{TV}(\mathcal{S}(x), \mathcal{S}(x'))$. It is easy to see from the construction of the maximal coupling that ω_1 and ω_1' have disjoint supports. In this way we obtain the canonical mixture decompositions $\mu = (1-\eta)\mu_0 + \eta\mu_1$ and $\mu' = (1-\eta)\mu_0 + \eta\mu_1'$, where $\mu_0 = \omega_0 M$, $\mu_1 = \omega_1 M$ and $\mu_1' = \omega_1' M$.

In the specific context of differential privacy this result yields for every $x \simeq_X x'$:

$$D_{e^{\varepsilon'}}(\mathcal{M}^{\mathcal{S}}(x)\|\mathcal{M}^{\mathcal{S}}(x')) \leq \eta \cdot \left((1-\theta)D_{e^{\varepsilon}}(\mu_1\|\mu_0) + \theta D_{e^{\varepsilon}}(\mu_1\|\mu_1')\right) \tag{4.2}$$

for $e^{\varepsilon'} = 1 + \eta(e^{\varepsilon} - 1)$, some $\theta \in [0,1]$, and $\eta = \mathsf{TV}(\mathcal{S}(x), \mathcal{S}(x'))$ being the total variation distance between the distributions over subsamples. It is interesting to note that the nonlinear relation $\varepsilon' = \log(1 + \eta(e^{\varepsilon} - 1))$ already appears in some existing privacy amplification results (e.g., Li et al. (2012)). Although for small ε and η this relation yields $\varepsilon' = O(\eta\varepsilon)$, our results show that the more complicated nonlinear relation is in fact a fundamental aspect of privacy amplification by subsampling. In particular, the relation $\varepsilon' = \log(1 + \eta(e^{\varepsilon} - 1))$ displays a phase transition: for small ε the result behaves like $\eta\varepsilon$, while for large ε the result behaves like ε (i.e., the amplification effect vanishes). This is illustrated in Figure 4.

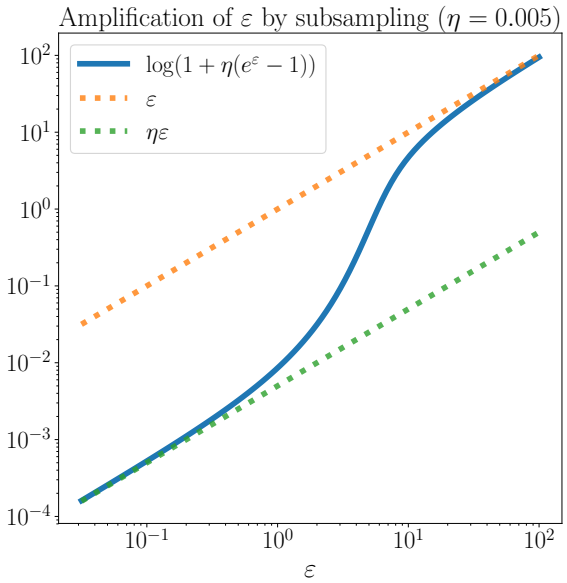


FIGURE 4. Phase transition in privacy amplification by subsampling.

4.1.2. *Distance-compatible Coupling.* The other tool we use to prove general privacy amplification bounds based on hockey-stick divergences is the existence of a certain type of couplings between two distributions like the ones occurring in the right hand side of (4.1). Recall that any coupling π between two distributions $\nu, \nu' \in \mathbb{P}(Y)$ can be used to rewrite the mixture distributions $\tilde{\mu} = \nu M$ and $\tilde{\mu}' = \nu' M$ as $\tilde{\mu} = \sum_{y,y'} \pi_{y,y'} \mathcal{M}(y)$ and $\tilde{\mu}' = \sum_{y,y'} \pi_{y,y'} \mathcal{M}(y')$. Using the joint convexity of D_{e^ε} and the definition of group-privacy profiles yields the bound

$$D_{e^\varepsilon}(\tilde{\mu} \| \tilde{\mu}') \leq \sum_{y,y'} \pi_{y,y'} D_{e^\varepsilon}(\mathcal{M}(y) \| \mathcal{M}(y')) \leq \sum_{y,y'} \pi_{y,y'} \delta_{\mathcal{M},d_Y(y,y')}(\varepsilon) \ . \tag{4.3}$$

Since this bound holds for any coupling π, one can set out to optimize it by finding a coupling the minimizes the right hand side of (4.3). We show that the existence of couplings whose support is contained inside a certain subset of $Y \times Y$ is enough to obtain an optimal bound. Furthermore, we show that when this condition is satisfied the resulting bound depends only on ν and the group-privacy profiles of $\mathcal{M}$.

Let d_Y be the path-distance induced by $\simeq_Y$. We say that two distributions $\nu, \nu' \in \mathbb{P}(Y)$ are d_Y*-compatible* if there exists a coupling π between ν and ν' such for any $(y, y') \in \mathsf{supp}(\pi)$ we have $d_Y(y, y') = d_Y(y, \mathsf{supp}(\nu'))$, where the distance between a point y and the set $\mathsf{supp}(\nu')$ is defined as the distance between y and the closest point in $\mathsf{supp}(\nu')$.

Theorem 12 . *Let $C(\nu, \nu')$ be the set of all couplings between ν and ν' and for $k \geq 1$ let $Y_k = \{y \in \mathsf{supp}(\nu) : d_Y(y, \mathsf{supp}(\nu')) = k\}$. If ν and ν' are d_Y-compatible, then the following holds:*

$$\min_{\pi \in C(\nu,\nu')} \sum_{y,y'} \pi_{y,y'} \delta_{\mathcal{M},d_Y(y,y')}(\varepsilon) = \sum_{k \geq 1} \nu(Y_k) \delta_{\mathcal{M},k}(\varepsilon) \ . \tag{4.4}$$

Proof. The result follows from a few simple observations. The first observation is that for any coupling $\pi \in C(\nu, \nu')$ and $y \in \mathsf{supp}(\nu')$ we have

$$\sum_{y'} \pi_{y,y'} \delta_{\mathcal{M},d_Y(y,y')}(\varepsilon) \geq \sum_{y'} \pi_{y,y'} \delta_{\mathcal{M},d_Y(y,\mathsf{supp}(\nu'))}(\varepsilon) = \sum_{y} \nu_y \delta_{\mathcal{M},d_Y(y,\mathsf{supp}(\nu'))}(\varepsilon) \ ,$$

where the first inequality follows from $d_Y(y, y') \geq d_Y(y, \mathsf{supp}(\nu'))$ and the fact that $\delta_{\mathcal{M},k}(\varepsilon)$ is monotonically increasing with k. Thus the RHS of (4.4) is always a lower bound for the LHS. Now let π be a d_Y-compatible coupling. Since the support of π only contains pairs (y, y') such that $d_Y(y, y') = d_Y(y, \mathsf{supp}(\nu'))$, we see that

$$\sum_{y,y'} \pi_{y,y'} \delta_{\mathcal{M},d_Y(y,y')}(\varepsilon) = \sum_{y,y'} \pi_{y,y'} \delta_{\mathcal{M},d_Y(y,\mathsf{supp}(\nu'))}(\varepsilon) = \sum_{y} \nu_y \delta_{\mathcal{M},d_Y(y,\mathsf{supp}(\nu'))}(\varepsilon) \ .$$

□

Applying this result to the bound resulting from the right hand side of (4.1) yields most of the concrete privacy amplification results presented in the next paragraphs.

4.2. **Poisson Subsampling.** We first analyze privacy amplification of Poisson subsampling with respect to the remove/add-one relation. In this case the subsampling mechanism $\mathcal{S}_\gamma^{\mathsf{po}} : 2^{\mathcal{U}} \to \mathbb{P}(2^{\mathcal{U}})$ takes a set x and outputs a sample y from the distribution $\omega = \mathcal{S}_\gamma^{\mathsf{po}}(x)$ supported on all sets $y \subseteq x$ given by $\omega(y) = \gamma^{|y|}(1-\gamma)^{|x|-|y|}$. This corresponds to independently adding to y each element from x with probability γ. Now, given a mechanism $\mathcal{M} : 2^{\mathcal{U}} \to \mathbb{P}(Z)$ with privacy profile $\delta_{\mathcal{M}}$ with respect to $\simeq_r$, we are interested in bounding the privacy profile of the subsampled mechanism $\mathcal{M}^{\mathcal{S}_\gamma^{\mathsf{wo}}}$ with respect to $\simeq_r$.

Theorem 13 . *Let $\mathcal{M}' = \mathcal{M}^{\mathcal{S}_\gamma^{\mathsf{po}}}$. For any $\varepsilon \geq 0$ we have $\delta_{\mathcal{M}'}(\varepsilon') \leq \gamma\delta_{\mathcal{M}}(\varepsilon)$, where $\varepsilon' = \log(1+\gamma(e^\varepsilon - 1))$.*

Proof. Given $x, x' \in 2^{\mathcal{U}}$ with $x \simeq_r x'$, we write $\omega = \mathcal{S}_\eta^{\mathsf{wo}}(x)$ and $\omega' = \mathcal{S}_\eta^{\mathsf{wo}}(x')$ and note that $\mathsf{TV}(\omega, \omega') = \eta$. Next we define $x_0 = x \cap x'$ and observe that either $x_0 = x$ or $x_0 = x'$ by the definition of $\simeq_r$. Let $\omega_0 = \mathcal{S}_\eta^{\mathsf{po}}(x_0)$. Then the decompositions of ω and ω' induced by their maximal coupling have either $\omega_1 = \omega_0$ when $x = x_0$ or $\omega_1' = \omega_0$ when $x' = x_0$. Noting that applying advanced joined convexity in the former case leads to an additional cancellation we see that the maximum will be attained when $x' = x_0$. In this case, the distribution ω_1 is given by $\omega_1(y \cup \{v\}) = \omega_0(y)$. This observation yields an obvious $d_{\simeq_r}$-compatible coupling between ω_1 and $\omega_0 = \omega_1'$: first sample y' from ω_0, and then build y by adding v to y'. Since every pair of datasets generated by this coupling has distance one with respect to $d_{\simeq_r}$, Theorem 12 yields the bound $\delta_{\mathcal{M}'}(\varepsilon') \leq \eta\delta_{\mathcal{M}}(\varepsilon)$. □

Privacy amplification with Poisson sampling was used in (Chaudhuri and Mishra, 2006; Beimel et al., 2010; Kasiviswanathan et al., 2011; Beimel et al., 2014), which considered loose bounds. A proof of this tight result in terms of (ε, δ)-DP was first given in (Li et al., 2012). In the context of the moments accountant technique based on the moment generating function of the privacy loss random variable, (Abadi et al., 2016) provide an amplification result for Gaussian output perturbation mechanisms under Poisson subsampling.

4.3. **Subsampling Without Replacement.** Another known results on privacy amplification corresponds to the analysis of sampling without replacement with respect to the substitution relation. In this case one considers the subsampling mechanism $\mathcal{S}_m^{\mathsf{wo}} : 2_n^{\mathcal{U}} \to \mathbb{P}(2_m^{\mathcal{U}})$ that given a set $x \in 2_n^{\mathcal{U}}$ of size n outputs a sample from the uniform distribution $\omega = \mathcal{S}_m^{\mathsf{wo}}(x)$ over all subsets $y \subseteq x$ of size $m \leq n$. Then, for a given a mechanism $\mathcal{M} : 2_m^{\mathcal{U}} \to \mathbb{P}(Z)$ with privacy profile $\delta_{\mathcal{M}}$ with respect to the substitution relation $\simeq_s$ on sets of size m, we are interested in bounding the privacy profile of the mechanism $\mathcal{M}^{\mathcal{S}_m^{\mathsf{wo}}}$ with respect to the substitution relation on sets of size n.

Theorem 14 . *Let $\mathcal{M}' = \mathcal{M}^{\mathcal{S}_m^{\mathsf{wo}}}$. For any $\varepsilon \geq 0$ we have $\delta_{\mathcal{M}'}(\varepsilon') \leq (m/n)\delta_{\mathcal{M}}(\varepsilon)$, where $\varepsilon' = \log(1+(m/n)(e^\varepsilon - 1))$.*

Proof. The analysis proceeds along the lines of the previous proof. First we note that for any $x, x' \in 2_n^{\mathcal{U}}$ with $x \simeq_s x'$, the total variation distance between $\omega = \mathcal{S}_m^{\mathsf{wo}}(x)$ and $\omega' = \mathcal{S}_m^{\mathsf{wo}}(x')$ is given by $\eta = \mathsf{TV}(\omega, \omega') = m/n$. Applying advanced joint convexity (Theorem 11) with the decompositions $\omega = (1-\eta)\omega_0 + \eta\omega_1$ and $\omega' = (1-\eta)\omega_0 + \eta\omega_1'$ given by the maximal coupling, the analysis of $D_{e^{\varepsilon'}}(\omega M \| \omega' M)$ reduces to bounding the divergences $D_{e^\varepsilon}(\omega_1 M \| \omega_0 M)$ and $D_{e^\varepsilon}(\omega_1 M \| \omega_1' M)$. In this case both quantities can be bounded by $\delta_{\mathcal{M}}(\varepsilon)$ by constructing appropriate $d_{\simeq_s}$-compatible couplings and combining (4.3) with Theorem 12.

We construct the couplings as follows. Suppose $v, v' \in \mathcal{U}$ are the elements where x and x' differ: $x_v = x_v' + 1$ and $x_{v'}' = x_{v'} + 1$. Let $x_0 = x \cap x'$. Then we have $\omega_0 = \mathcal{S}_m^{\mathsf{wo}}(x_0)$. Furthermore, writing $\tilde{\omega}_1 = \mathcal{S}_{m-1}^{\mathsf{wo}}(x_0)$ we have $\omega_1(y) = \tilde{\omega}_1(y \cap x_0)$ and $\omega_1'(y) = \tilde{\omega}_1(y \cap x_0)$. Using these definitions we build a coupling $\pi_{1,1}$ between ω_1 and ω_1' through the following generative process: sample y_0 from $\tilde{\omega}_1$ and then let $y = y_0 \cup \{v\}$ and $y' \cup \{v'\}$. Similarly, we build a coupling $\pi_{1,0}$ between ω_1 and ω_0 as follows: sample y_0 from $\tilde{\omega}_1$, sample u uniformly from $x_0 \setminus y_0$, and then let $y = y_0 \cup \{v\}$ and $y' = y_0 \cup \{u\}$. It is obvious from these constructions that $\pi_{1,1}$ and $\pi_{0,1}$ are both $d_{\simeq_s}$-compatible. Plugging these observations together, we get $\delta_{\mathcal{M}'}(\varepsilon') \leq (m/n)\delta_{\mathcal{M}}(\varepsilon)$. □

This setting has been used in (Beimel et al., 2013; Bassily et al., 2014; Wang et al., 2016) with non-tight bounds. A proof of this tight bound formulated in terms of (ε, δ)-DP can be directly recovered from Ullman's class notes (Ullman, 2017), although the stated bound is weaker. Rényi DP amplification bounds for subsampling without replacement were recently developed by Wang et al. (2019), who also emphasized the importance of Rényi privacy profiles in their analysis.

4.4. **Subsampling With Replacement.** Next we consider the case of sampling with replacement with respect to the substitution relation $\simeq_s$. The subsampling with replacement mechanism $\mathcal{S}^{\mathsf{wr}}_m : 2^{\mathcal{U}}_n \to \mathbb{P}(\mathbb{N}^{\mathcal{U}}_m)$ takes a set x of size n and outputs a sample from the multinomial distribution $\omega = \mathcal{S}^{\mathsf{wr}}_m(x)$ over all multisets y of size $m \leq n$ with $\mathsf{supp}(y) \subseteq x$, given by $\omega(y) = (m!/n^m) \prod_{u \in \mathcal{U}} x_u/(y_u!)$. In this case we suppose the base mechanism $\mathcal{M} : \mathbb{N}^{\mathcal{U}}_m \to \mathbb{P}(Z)$ is defined on multisets and has privacy profile $\delta_{\mathcal{M}}$ with respect to $\simeq_s$. We are interested in bounding the privacy profile of the subsampled mechanism $\mathcal{M}^{\mathcal{S}^{\mathsf{wr}}_m} : 2^{\mathcal{U}}_n \to \mathbb{P}(Z)$ with respect to $\simeq_s$.

Theorem 15 . *Let $\mathcal{M}' = \mathcal{M}^{\mathcal{S}^{\mathsf{wr}}_m}$. Given $\varepsilon \geq 0$ and $\varepsilon' = \log(1 + (1 - (1 - 1/n)^m)(e^\varepsilon - 1))$ we have*

$$\delta_{\mathcal{M}'}(\varepsilon') \leq \sum_{k=1}^{m} \binom{m}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} \delta_{\mathcal{M},k}(\varepsilon) \ .$$

Proof. To bound the privacy profile of the subsampled mechanism $\mathcal{M}^{\mathcal{S}^{\mathsf{wr}}_m}$ on $2^{\mathcal{U}}_n$ with respect to $\simeq_s$ we start by noting that taking $x, x' \in 2^{\mathcal{U}}_n$, $x \simeq_s x'$, the total variation distance between $\omega = \mathcal{S}^{\mathsf{wr}}_m(x)$ and $\omega' = \mathcal{S}^{\mathsf{wr}}_m(x')$ is given by $\eta = \mathsf{TV}(\omega, \omega') = 1 - (1 - 1/n)^m$. To define appropriate mixture components for applying the advanced joint composition property we write v and v' for the elements where x and x' differ and $x_0 = x \cap x'$ for the common part between both datasets. Then we have $\omega_0 = \mathcal{S}^{\mathsf{wr}}_m(x_0)$. Furthermore, ω_1 is the distribution obtained from sampling $\tilde{y}$ from $\tilde{\omega}_1 = \mathcal{S}^{\mathsf{wr}}_{m-1}(x)$ and building y by adding one occurrence of v to $\tilde{y}$. Similarly, sampling y' from ω'_1 corresponds to adding v' to a multiset sampled from $\mathcal{S}^{\mathsf{wr}}_{m-1}(x')$.

Now we construct appropriate distance-compatible couplings. First we let $\pi_{1,1} \in \mathbb{P}(\mathbb{N}^{\mathcal{U}}_m \times \mathbb{N}^{\mathcal{U}}_m)$ be the distribution given by sampling y from ω_1 as above and outputting the pair (y, y') obtained by replacing each v in y by v'. It is immediate from this construction that $\pi_{1,1}$ is a $d_{\simeq_s}$-compatible coupling between ω_1 and ω'_1. Furthermore, using the notation from Theorem 12 and the construction of the maximal coupling, we see that for $k \geq 1$:

$$\omega_1(Y_k) = \frac{\omega(Y_k) - (1-\eta)\omega_0(Y_k)}{\eta} = \frac{\mathsf{Pr}_{y \sim \omega}[y_v = k]}{\eta} = \frac{1}{\eta}\binom{m}{k}\left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} \ ,$$

where we used $\omega_0(Y_k) = 0$ since ω_0 is supported on multisets that do not include v. Therefore, the distributions $\mu_1 = \omega_1 M$ and $\mu'_1 = \omega'_1 M$ satisfy

$$\eta D_{e^\varepsilon}(\mu_1 \| \mu'_1) \leq \sum_{k=1}^{m} \binom{m}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} \delta_{\mathcal{M},k}(\varepsilon) \ . \tag{4.5}$$

On the other hand, we can build a $d_{\simeq_s}$-compatible coupling between ω_1 and ω_0 by first sampling y from ω_1 and then replacing each occurrence of v by an element picked uniformly at random from x_0. Again, this shows that $D_{e^\varepsilon}(\mu_1 \| \mu_0)$ is upper bounded by the right hand side of (4.5).

Therefore, we conclude that

$$\delta_{\mathcal{M}'}(\varepsilon') \leq \sum_{k=1}^{m} \binom{m}{k} \left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{m-k} \delta_{\mathcal{M},k}(\varepsilon) \ .$$

□

Note that if $m = \gamma n$, then $1-(1-1/n)^m \approx \gamma$. A version of this bound in terms of (ε,δ)-DP that implicitly uses the group privacy property can be found in (Bun et al., 2015). Our bound matches the asymptotics of (Bun et al., 2015) while providing optimal constants and allowing for white-box group privacy bounds.

4.5. **Hybrid Results.** Using our method it is also possible to analyze new settings which have not been considered before. One interesting example occurs when there is a mismatch between the two neighboring relations arising in the analysis. For example, suppose one knows the group-privacy profiles $\delta_{\mathcal{M},k}$ of a base mechanism $\mathcal{M} : \mathbb{N}_m^{\mathcal{U}} \to \mathbb{P}(Z)$ with respect to the substitution relation $\simeq_s$. In this case one could ask whether it makes sense to study the privacy profile of the subsampled mechanism $\mathcal{M}^{\mathcal{S}_m^{\mathsf{wr}}} : 2^{\mathcal{U}} \to \mathbb{P}(Z)$ with respect to the remove/add relation $\simeq_r$. In principle, this makes sense in settings where the size of the inputs to $\mathcal{M}$ is restricted due to implementation constraints (e.g., limited by the memory available in a GPU used to run a private mechanism that computes a gradient on a mini-batch of size m). In this case one might still be interested in analyzing the privacy loss incurred from releasing such stochastic gradients under the remove/add relation. Note that this setting cannot be implemented using sampling without replacement since under the remove/add relation we cannotguarantee *a priori* that the input dataset will have at least size m because the size of the dataset must be kept private (Vadhan, 2017). Furthermore, one cannot hope to get a meaningful result about the privacy profile of the subsampled mechanism across all inputs sets in $2^{\mathcal{U}}$; instead the privacy guarantee will depend on the size of the input dataset as shown in the following result.

Theorem 16 . *Let $\mathcal{M}' = \mathcal{M}^{\mathcal{S}_m^{\mathsf{wr}}}$. For any $\varepsilon \geq 0$ and $n \geq 0$ we have*

$$\sup_{x \in 2_n^{\mathcal{U}}, x \simeq_r x'} D_{e^{\varepsilon'}}(\mathcal{M}'(x) \| \mathcal{M}'(x')) \leq \sum_{k=1}^{m} \binom{m}{k} \left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{m-k} \delta_{\mathcal{M},k}(\varepsilon) \ ,$$

where $\varepsilon' = \log(1+(1-(1-1/n)^m)(e^{\varepsilon}-1))$.

Proof. Suppose $x \simeq_r x'$ with $|x| = n$ and $|x'| = n-1$. This is the worst-case direction for the neighboring relation like in the proof of Theorem 13. Let $\omega = \mathcal{S}_m^{\mathsf{wr}}(x)$ and $\omega = \mathcal{S}_m^{\mathsf{wr}}(x')$. We have $\eta = \mathsf{TV}(\omega,\omega') = 1-(1-1/n)^m$, and the factorization induced by the maximal coupling has $\omega_0 = \omega_1' = \omega'$ and ω_1 is given by first sampling $\tilde{y}$ from $\mathcal{S}_{m-1}^{\mathsf{wr}}(x)$ and then producing y by adding to $\tilde{y}$ a copy of the element v where x and x' differ. This definition of ω_1 suggests the following coupling between ω_1 and ω_0: first sample y from ω_1, then produce y' by replacing each copy of v with a element from x' sampled independently and uniformly. By construction we see that this coupling is $d_{\simeq_s}$-compatible, so we can apply Theorem 12. Using the same argument as in the proof of Theorem 15 we see that $\eta\omega_1(Y_k) = \binom{m}{k}(1/n)^k(1-1/n)^{m-k}$.

Thus, we finally get

$$\begin{aligned} D_{e^{\varepsilon'}}(\mathcal{M}^{\mathcal{S}_m^{\mathsf{wr}}}(x)\|\mathcal{M}^{\mathcal{S}_m^{\mathsf{wr}}}(x')) &= \eta D_{e^{\varepsilon}}(\omega_1 M\|\omega_0 M) \\ &\leq \eta \sum_{k=1}^{m} \omega_1(Y_k)\delta_{\mathcal{M},k}(\varepsilon) \\ &= \sum_{k=1}^{m} \binom{m}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{m-k}\delta_{\mathcal{M},k}(\varepsilon) \ . \end{aligned}$$

□

4.6. **When the Neighboring Relation is "Incompatible".** Now we consider a simple example where distance-compatible couplings are not available: Poisson subsampling with respect to the substitution relation. Suppose $x, x' \in 2_n^{\mathcal{U}}$ are sets of size n related by the substitution relation $\simeq_s$. Let $\omega = \mathcal{S}_\eta^{\mathsf{po}}(x)$ and $\omega' = \mathcal{S}_\eta^{\mathsf{po}}(x')$ and note that $\mathsf{TV}(\omega, \omega') = \eta$. Let $x_0 = x \cap x'$ and $v = x \setminus x_0$, $v' = x' \setminus x_0$. In this case the factorization induced by the maximal coupling is obtained by taking $\omega_0 = \mathcal{S}_\eta^{\mathsf{po}}(x_0)$, $\omega_1(y \cup \{v\}) = \omega_0(y)$, and $\omega_1'(y \cup \{v'\}) = \omega_0(y)$. Now the support of ω_0 contains sets of sizes between 0 and $n-1$, while the supports of ω_1 and ω_1 contain sets of sizes between 1 and n. From this observation one can deduce that ω_1 and ω_0 are not $d_{\simeq_s}$-compatible, and ω_1 and ω_1' are not $d_{\simeq_r}$-compatible.

This argument shows that the method we used to analyze the previous settings cannot be extended to analyze Poisson subsampling under the substitution relation, regardless of whether the privacy profile of the base mechanism is given in terms of the replacement/addition or the substitution relation. This observation is saying that some pairings between subsampling method and neighboring relation are more natural than others. Nonetheless, even without distance-compatible couplings it is possible to provide privacy amplification bounds for Poisson subsampling with respect to the substitution relation, although the resulting bound is quite cumbersome.

Theorem 17 . *Let $\mathcal{M} : 2^{\mathcal{U}} \to \mathbb{P}(Z)$ be a mechanism with privacy profile $\delta_{\mathcal{M}}$ with respect to $\simeq_s$. Then the privacy profile with respect of $\simeq_s$ of the subsampled mechanism $\mathcal{M}' = \mathcal{M}^{\mathcal{S}_\gamma^{\mathsf{po}}} : 2_n^{\mathcal{U}} \to \mathbb{P}(Z)$ on datasets of size n satisfies the following:*

$$\delta_{\mathcal{M}'}(\varepsilon') \leq \gamma\theta\delta_{\mathcal{M}}(\varepsilon) + \gamma(1-\theta)\left(\sum_{k=1}^{n-1}\tilde{\gamma}_k\delta_{\mathcal{M}}(\varepsilon_k) + \tilde{\gamma}_n\right) \ ,$$

where $\varepsilon' = \log(1+\gamma(e^{\varepsilon}-1))$, $\theta = e^{\varepsilon'}/e^{\varepsilon}$, $\varepsilon_k = \varepsilon + \log(\frac{\gamma}{1-\gamma}(\frac{n}{k}-1))$, and $\tilde{\gamma}_k = \binom{n-1}{k-1}\gamma^{k-1}(1-\gamma)^{n-k}$.

Proof. Suppose $x, x' \in 2_n^{\mathcal{U}}$ are sets of size n related by the substitution relation $\simeq_s$. Let $\omega = \mathcal{S}_\eta^{\mathsf{po}}(x)$ and $\omega' = \mathcal{S}_\eta^{\mathsf{po}}(x')$ and note that $\mathsf{TV}(\omega, \omega') = \eta$. Let $x_0 = x \cap x'$ and $v = x \setminus x_0$, $v' = x' \setminus x_0$. In this case the factorization induced by the maximal coupling is obtained by taking $\omega_0 = \mathcal{S}_\eta^{\mathsf{po}}(x_0)$, $\omega_1(y \cup \{v\}) = \omega_0(y)$, and $\omega_1'(y \cup \{v'\}) = \omega_0(y)$. From this factorization we see it is easy to construct a coupling $\pi_{1,1}$ between ω_1 and ω_1' that is $d_{\simeq_s}$-compatible. Therefore we have $D_{e^{\varepsilon}}(\omega_1 M\|\omega_1' M) \leq \delta_{\mathcal{M}}(\varepsilon)$.

Since we have already identified that no $d_{\simeq_s}$-compatible coupling between ω_1 and ω_0 can exist, we shall further decompose these distributions "by hand." Let $\nu_k = \mathcal{S}_k^{\mathsf{wo}}(x_0)$ and note that ν_k corresponds to the distribution ω_0 conditioned on $|y| = k$. Similarly, we define

$\tilde{\nu}_k$ as the distribution corresponding to sampling $\tilde{y}$ from $\mathcal{S}^{\mathsf{wo}}_{k-1}(x_0)$ and outputting the set y obtained by adding v to $\tilde{y}$. Then $\tilde{\nu}_k$ equals the distribution of ω_1 conditioned on $|y| = k$. Now we write $\gamma_k = \mathsf{Pr}_{y\sim\omega_0}[|y| = k] = \binom{n-1}{k}\gamma^k(1-\gamma)^{n-1-k}$ and $\tilde{\gamma}_k = \mathsf{Pr}_{y\sim\omega_1}[|y| = k] = \binom{n-1}{k-1}\gamma^{k-1}(1-\gamma)^{n-k}$. With these notations we can write the decompositions $\omega_0 = \sum_{k=0}^{n-1}\gamma_k\nu_k$ and $\omega_1 = \sum_{k=1}^{n}\tilde{\gamma}_k\tilde{\nu}_k$. Further, we observe that the construction of $\tilde{\nu}_k$ and ν_k shows there exist $d_{\simeq_s}$-compatible couplings between these pairs of distributions when $1 \le k \le n-1$, leading to $D_{e^\varepsilon}(\tilde{\nu}_k M \| \nu_k M) \le \delta_{\mathcal{M}}(\varepsilon)$. To exploit this fact we first write

$$D_{e^\varepsilon}(\omega_1 M \| \omega_0 M) = D_{e^\varepsilon}\left(\sum_{k=1}^{n-1}\tilde{\gamma}_k\tilde{\nu}_k M + \tilde{\gamma}_n\tilde{\nu}_n M \middle\| \gamma_0\nu_0 M + \sum_{k=1}^{n-1}\gamma_k\nu_k M\right) \ .$$

Now we use that hockey-stick divergences can be applied to arbitrary nonnegative measures, which are not necessarily probability measures, using the same definition we have used so far. Under this relaxation, given non-negative measures ν_i, ν_i', $i = 1, 2$, on a measure space Z we have $D_\alpha(\nu_1 + \nu_2 \| \nu_1' + \nu_2') \le D_\alpha(\nu_1\|\nu_1') + D_\alpha(\nu_2\|\nu_2')$, $D_\alpha(a\nu_1\|b\nu_2) = aD_{\alpha b/a}(\nu_1\|\nu_2)$ for $a \ge 0$ and $b > 0$, and $D_\alpha(\nu_1\|0) = \nu_1(Z)$. Using these properties on the decomposition above we see that

$$\begin{aligned} D_{e^\varepsilon}(\omega_1 M\|\omega_0 M) &\le \sum_{k=1}^{n-1}\tilde{\gamma}_k D_{e^{\varepsilon_k}}(\tilde{\nu}_k M\|\nu_k M) + \tilde{\gamma}_n \\ &\le \sum_{k=1}^{n-1}\tilde{\gamma}_k\delta_{\mathcal{M}}(\varepsilon_k) + \tilde{\gamma}_n \ , \end{aligned}$$

where $e^{\varepsilon_k} = (\gamma_k/\tilde{\gamma}_k)e^\varepsilon = (\gamma/(1-\gamma))(n/k-1)e^\varepsilon$. □

4.7. **Lower Bounds.** In this section we show that many of the results given in the previous section are tight by constructing a randomized membership mechanism that attains these upper bounds. For the sake of generality, we state the main construction in terms of tuples instead of multisets. In fact, we prove a general lemma that can be used to obtain tightness results for any subsampling mechanism and any neighboring relation satisfying two natural assumptions.

For $p \in [0,1]$ let $\mathcal{R}_p : \{0,1\} \to \mathbb{P}(\{0,1\})$ be the randomized response mechanism that given $b \in \{0,1\}$ returns b with probability p and $1-b$ with probability $1-p$. Note that for $p = (e^\varepsilon + \delta)/(e^\varepsilon + 1)$ this mechanism is (ε,δ)-DP. Let $\nu_0 = \mathcal{R}_p(0)$ and $\nu_1 = \mathcal{R}_p(1)$. For any $\varepsilon \ge 0$ and $p \in [0,1]$ define $\psi_p(\varepsilon) = [p - e^\varepsilon(1-p)]_+$. It is easy to verify that $D_{e^\varepsilon}(\nu_0\|\nu_1) = D_{e^\varepsilon}(\nu_1\|\nu_0) = \psi_p(\varepsilon)$. Now let $\mathcal{U}$ be a universe containing at least two elements. For $v \in \mathcal{U}$ and $p \in [0,1]$ we define the *randomized membership* mechanism $\mathcal{M}_{v,p}$ that given a tuple $x = (u_1, \ldots, u_n) \in \mathcal{U}^\star$ returns $\mathcal{M}_{v,p}(x) = \mathcal{R}_p(\mathbb{I}[v \in x])$. We say that a subsampling mechanism $\mathcal{S} : X \to \mathbb{P}(\mathcal{U}^\star)$ defined on some set $X \subseteq \mathcal{U}^\star$ is *natural* if the following two conditions are satisfied:

(1) For any $x \in X$ and $u \in \mathcal{U}$, if $u \in x$ then there exists $y \in \mathsf{supp}(\mathcal{S}(x))$ such that $u \in y$.
(2) For any $x \in X$ and $u \in \mathcal{U}$, if $u \notin x$ then we have $u \notin y$ for every $y \in \mathsf{supp}(\mathcal{S}(x))$.

Theorem 18 . *Let $X \subseteq \mathcal{U}^\star$ be equipped with a neighboring relation $\simeq_X$ such that there exist $x \simeq_X x'$ with $v \in x$ and $v \notin x'$. Suppose $\mathcal{S} : X \to \mathbb{P}(\mathcal{U}^\star)$ is a natural subsampling mechanism*

and let $\eta = \sup_{x \simeq_X x'} \mathsf{TV}(\mathcal{S}(x), \mathcal{S}(x'))$. *For any* $\varepsilon \geq 0$ *and* $\varepsilon' = \log(1 + \eta(e^\varepsilon - 1))$ *we have*

$$\delta_{\mathcal{M}^{\mathcal{S}}_{v,p}}(\varepsilon') = \sup_{x \simeq_X x'} D_{e^{\varepsilon'}}(\mathcal{M}^{\mathcal{S}}_{v,p}(x) \| \mathcal{M}^{\mathcal{S}}_{v,p}(x')) = \eta \psi_p(\varepsilon) \ .$$

Proof. We start by observing that for any $x \in X$ the distribution $\mu = \mathcal{M}^{\mathcal{S}}_{v,p}(x)$ must be a mixture $\mu = (1-\theta)\nu_0 + \theta\nu_1$ for some $\theta \in [0,1]$. This follows from the fact that there are only two possibilities ν_0 and ν_1 for $\mathcal{M}_{v,p}(y)$ depending on whether $v \notin y$ or $v \in y$. Similarly, taking $x \simeq_X x'$ we get $\mu' = \mathcal{M}^{\mathcal{S}}_{v,p}(x')$ with $\mu' = (1-\theta')\nu_0 + \theta'\nu_1$ for some $\theta' \in [0,1]$. Assuming (without loss of generality) $\theta \geq \theta'$, we use the advanced joint convexity property of D_α to get

$$\begin{aligned} D_{e^{\varepsilon'}}(\mu \| \mu') &= \theta D_{e^\varepsilon}(\nu_1 \| (1-\theta'/\theta)\nu_0 + (\theta'/\theta)\nu_1) \\ &\leq \theta(1-\theta'/\theta) D_{e^\varepsilon}(\nu_1 \| \nu_0) = (\theta - \theta')\psi_p(\varepsilon) \leq \theta\psi_p(\varepsilon) \ , \end{aligned}$$

where $\varepsilon' = \log(1 + \theta(e^\varepsilon - 1))$ and $\theta = e^{\varepsilon'}/e^\varepsilon$, and the inequality follows from joint convexity. Now note the inequalities above are in fact equalities when $\theta' = 0$, which is equivalent to the fact $v \notin x'$ because $\mathcal{S}$ is a natural subsampling mechanism. Thus, observing that the function $\theta \mapsto \theta\psi_p(\log(1 + (e^{\varepsilon'} - 1)/\theta))$ is monotonically increasing, we get

$$\begin{aligned} \sup_{x \simeq_X x'} D_{e^{\varepsilon'}}(\mathcal{M}^{\mathcal{S}}_{v,p}(x) \| \mathcal{M}^{\mathcal{S}}_{v,p}(x')) &= \sup_{x \simeq_X x', v \notin x'} \theta\psi_p(\log(1 + (e^{\varepsilon'} - 1)/\theta)) \\ &= \eta\psi_p(\log(1 + (e^{\varepsilon'} - 1)/\eta)) = \eta\psi_p(\varepsilon) \ . \end{aligned}$$

□

We can now apply this result to show that the first three results from previous section are tight. This requires specializing from tuples to (multi)sets, and plugging in the definitions of neighboring relation, subsampling mechanism, and η used in each of these theorems. Other than that, the proof is a direct calculation.

Corollary 19 . *The mechanism* $\mathcal{M}_{v,p}$ *attains the bounds in Theorems 13, 14, 15 for every* p *and* η.

5. Conclusion

We have introduced and developed the concept of privacy profiles as a method to capture the whole set of privacy guarantees offered by a given mechanism. The results in Section 3 provide explicit examples of privacy profiles for well-known mechanisms and study some geometric properties of privacy profiles. We also showed how the concept of privacy profile is connected to the methods used to bound differential privacy in terms of privacy loss random variables, and to alternative definitions of privacy, including (zero)-concentrated and Rényi differential privacy. These results, we hope, shed further light into the nature of these alternative definitions. In particular, our thesis is that the functional views of differential privacy (i.e., $\delta(\varepsilon)$) and Rényi differential privacy (i.e., $\epsilon(\alpha)$) provide deeper insights into the privacy properties of a given mechanism than what can be achieved by a point-wise guarantee (i.e., (ε, δ) and (α, ϵ)), and that when this lens is considered, both approaches to measure privacy contain essentially the same information as illustrated by the results in Section 3.4.

We have also developed a general method for reasoning about privacy amplification by subsampling. Our method is applicable to many different settings, some which have already been studied in the literature, and others which are new. Technically, our method leverages

a new tool of independent interest: advanced joint convexity. In the future, it would be interesting to apply our tools to more elaborate forms of subsampling, especially those which are widely used in the design of statistical surveys (Särndal et al., 2003).

Acknowledgments. This research was initiated during the 2017 Probabilistic Programming Languages workshop hosted by McGill University's Bellairs Research Institute.

References

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*, pages 308–318. ACM, 2016. https://doi.org/10.1145/2976749.2978318.

B. Balle and Y.-X. Wang. Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *Proceedings of the 35th International Conference on Machine Learning, ICML*, 2018. http://proceedings.mlr.press/v80/balle18a.html.

B. Balle, G. Barthe, and M. Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pages 6280–6290, 2018. http://papers.nips.cc/paper/7865-privacy-amplification-by-subsampling-tight-analyses-via-couplings-and-divergences.

B. Balle, J. Bell, A. Gascón, and K. Nissim. The privacy blanket of the shuffle model. In *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II*, pages 638–667, 2019. https://doi.org/10.1007/978-3-030-26951-7_22.

G. Barthe and F. Olmedo. Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs. In *International Colloquium on Automata, Languages, and Programming*, pages 49–60. Springer, 2013. https://doi.org/10.1007/978-3-642-39212-2_8.

G. Barthe, B. Köpf, F. Olmedo, and S. Z. Béguelin. Probabilistic relational reasoning for differential privacy. In *Symposium on Principles of Programming Languages (POPL)*, pages 97–110, 2012. https://doi.org/10.1145/2492061.

G. Barthe, M. Gaboardi, B. Grégoire, J. Hsu, and P. Strub. Proving differential privacy via probabilistic couplings. In *Symposium on Logic in Computer Science (LICS)*, pages 749–758, 2016. https://doi.org/10.1145/2933575.2934554.

R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 464–473. IEEE, 2014. https://doi.org/10.1109/FOCS.2014.56.

A. Beimel, S. P. Kasiviswanathan, and K. Nissim. Bounds on the sample complexity for private learning and private data release. In *Theory of Cryptography Conference*, pages 437–454. Springer, 2010. https://doi.org/10.1007/978-3-642-11799-2_26.

A. Beimel, K. Nissim, and U. Stemmer. Characterizing the sample complexity of private learners. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 97–110. ACM, 2013. https://doi.org/10.1145/2422436.2422450.

A. Beimel, H. Brenner, S. P. Kasiviswanathan, and K. Nissim. Bounds on the sample complexity for private learning and private data release. *Machine learning*, 94(3):401–437, 2014. https://doi.org/10.1007/s10994-013-5404-1.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I*, pages 635–658, 2016. https://doi.org/10.1007/978-3-662-53641-4_24.

M. Bun, K. Nissim, U. Stemmer, and S. Vadhan. Differentially private release and learning of threshold functions. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 634–649. IEEE, 2015. https://doi.org/10.1109/FOCS.2015.45.

M. Bun, C. Dwork, G. Rothblum, and T. Steinke. Composable and versatile privacy via truncated CDP. In *Symposium on Theory of Computing, STOC*, 2018. https://doi.org/10.1145/3188745.3188946.

K. Chaudhuri and N. Mishra. When random sampling preserves privacy. In *Annual International Cryptology Conference*, pages 198–213. Springer, 2006. https://doi.org/10.1007/11818175_12.

A. Cheu, A. D. Smith, J. Ullman, D. Zeber, and M. Zhilyaev. Distributed differential privacy via Mixnets. *CoRR*, abs/1808.01394, 2018. http://arxiv.org/abs/1808.01394.

C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014. https://doi.org/10.1561/0400000042.

C. Dwork and G. N. Rothblum. Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*, 2016. http://arxiv.org/abs/1603.01887.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pages 265–284. Springer, 2006. https://doi.org/10.1007/11681878_14.

C. Dwork, G. N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 51–60. IEEE, 2010. https://doi.org/10.1109/FOCS.2010.12.

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 Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2468–2479. SIAM, 2019. https://doi.org/10.1137/1.9781611975482.151.

V. Feldman, I. Mironov, K. Talwar, and A. Thakurta. Privacy amplification by iteration. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 521–532. IEEE, 2018. https://doi.org/10.1109/FOCS.2018.00056.

J. Jälkö, A. Honkela, and O. Dikmen. Differentially private variational inference for non-conjugate models. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017*, 2017. http://auai.org/uai2017/proceedings/papers/152.pdf.

P. Kairouz, S. Oh, and P. Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, 2017. https://doi.org/10.1109/TIT.2017.2685505.

S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011. https://doi.org/10.1109/FOCS.2008.27.

N. Li, W. Qardaji, and D. Su. On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, pages 32–33. ACM, 2012. https://doi.org/10.1145/2414456.2414474.

F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, 2006. https://doi.org/10.1109/TIT.2006.881731.

I. Mironov. Rényi differential privacy. In *30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21-25, 2017*, pages 263–275, 2017. https://doi.org/10.1109/CSF.2017.11.

J. Murtagh and S. Vadhan. The complexity of computing the optimal composition of differential privacy. In *Theory of Cryptography Conference*, pages 157–175. Springer, 2016. https://doi.org/10.1007/978-3-662-49096-9_7.

F. Österreicher. Csiszár's f-divergences-basic properties. *RGMIA Res. Rep. Coll*, 2002. https://documents.epfl.ch/groups/i/ip/ipg/www/2010-2011/Measures_of_Information/osterreicher_f_divergences.pdf.

M. Park, J. R. Foulds, K. Chaudhuri, and M. Welling. Private topic modeling. *CoRR*, abs/1609.04120, 2016a. https://arxiv.org/abs/1609.04120.

M. Park, J. R. Foulds, K. Chaudhuri, and M. Welling. Variational bayes in private settings (VIPS). *CoRR*, abs/1611.00340, 2016b. https://arxiv.org/abs/1611.00340.

D. Pollard. *A User's Guide to Measure Theoretic Probability*, volume 8. Cambridge University Press, 2002. https://doi.org/10.1017/CBO9780511811555.

C.-E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*. Springer Science & Business Media, 2003. https://psycnet.apa.org/doi/10.1007/978-1-4612-4378-6.

I. Sason and S. Verdú. f-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, 2016. https://doi.org/10.1109/TIT.2016.2603151.

J. Ullman. Cs7880: Rigorous approaches to data privacy. http://www.ccs.neu.edu/home/jullman/PrivacyS17/HW1sol.pdf, 2017.

S. P. Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. 2017. https://doi.org/10.1007/978-3-319-57048-8_7.

Y.-X. Wang, S. Fienberg, and A. Smola. Privacy for free: Posterior sampling and stochastic gradient Monte Carlo. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 2493–2502, 2015. http://proceedings.mlr.press/v37/wangg15.html.

Y.-X. Wang, J. Lei, and S. E. Fienberg. Learning with differential privacy: Stability, learnability and the sufficiency and necessity of ERM principle. *Journal of Machine Learning Research*, 17(183):1–40, 2016. http://www.jmlr.org/papers/v17/15-313.html.

Y.-X. Wang, B. Balle, and S. Kasiviswanathan. Subsampled Rényi differential privacy and analytical moments accountant. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019. http://proceedings.mlr.press/v89/wang19b.html.