On the Rényi Differential Privacy of the Shuffle Model

Antonious M. Girgis
UCLA

Deepesh Data
UCLA

Suhas Diggavi
UCLA

Ananda Theertha Suresh
Google

Peter Kairouz
Google

ABSTRACT

The central question studied in this paper is Rényi Differential Privacy (RDP) guarantees for general discrete local randomizers in the *shuffle* privacy model. In the shuffle model, each of the n clients randomizes its response using a local differentially private (LDP) mechanism and the untrusted server only receives a random permutation (shuffle) of the client responses without association to each client. The principal result in this paper is the *first* direct RDP bounds for general discrete local randomization in the shuffle privacy model, and we develop new analysis techniques for deriving our results which could be of independent interest. In applications, such an RDP guarantee is most useful when we use it for composing several private interactions. We numerically demonstrate that, for important regimes, with composition our bound yields an improvement in privacy guarantee by a factor of 8× over the state-of-the-art approximate Differential Privacy (DP) guarantee (with standard composition) for shuffle models. Moreover, combining with Poisson subsampling, our result leads to at least 10× improvement over subsampled approximate DP with standard composition.

CCS CONCEPTS

• **Security and privacy** → *Privacy-preserving protocols.*

KEYWORDS

Differential privacy; Rényi divergence; distributed learning; privacy amplification via shuffling; privacy composition.

1 INTRODUCTION

Differential privacy (DP) [16] gives a principled and rigorous framework for data privacy by giving guarantees on the information leakage for individual data points from the output of an algorithm. Algorithmically, a standard method is to randomize the output of

an algorithm to enable such privacy. Originally DP was studied in the *centralized* context, where the privacy from queries to a *trusted server* holding the data was the objective [16]. However, in distributed applications, such as federated learning [30], two significant aspects need to be accommodated: (i) data is held locally at clients and needs to be used for computation with an *untrusted* server; and (ii) to build good learning models, one might need repeated interactions (*e.g.,* through distributed gradient descent).

To accommodate privacy of *locally* held data, a more appropriate notion is that of local differential privacy (LDP) [15, 34]. In the LDP framework, each (distributed) client holding local data, individually randomizes its interactions with the (untrusted) server.[1] Recently, such LDP mechanisms have been deployed by companies such as Google [23], Apple [29], and Microsoft [14]. However, LDP mechanisms suffer from poor performance in comparison with the centralized DP mechanisms, making their applicability limited [15, 31, 34]. To address this, a new privacy framework using anonymization has been proposed in the so-called *shuffle model* [7, 13, 22], where each client sends her (randomized) interaction message to a secure shuffler that randomly permutes all the received messages before forwarding them to the server. Such a shuffling can be enabled through anonymization techniques [10, 20, 22]. This model enables significantly better privacy-utility performance by amplifying LDP through this mechanism.

For the second aspect, where there are repeated interactions (*e.g.,* through distributed gradient descent), one needs privacy composition [9]. In other words, we want to compute the overall privacy budget under the composition of multiple iterations. Clearly, from an optimization viewpoint, we might need to run these interactions longer for better models, but these also result in privacy leakage. Though the privacy leakage can be quantified using advanced composition theorems for DP (*e.g.,* [19, 33]), these might be loose. To address this, Abadi *et al.* [1] developed a "moments accountant" framework, which enabled a much tighter composition. This is enabled by providing the composition privacy guarantee in terms of Rényi Differential Privacy [36], and then mapping it back to the DP guarantee [37]. It is known [1] that the moments accountant provides a significant saving in the total privacy budget in comparison with using the strong composition theorems [19, 33]. Therefore, developing the RDP privacy guarantee can enable stronger composition privacy results. Analyzing the RDP of the shuffle model could have several applications such as private statistics using interactive schemes for heavy hitters, mean estimation, federated learning, and distributed differentially private stochastic gradient descent (DPSGD). This leads us to the central question addressed in this paper:

---

[1]The mechanisms used have a long history including Randomized Response [41], but were recently studied through the lens of local differential privacy (LDP).

Can we develop strong RDP privacy guarantees for general local mechanisms in the shuffle privacy model?

The principal result in this paper is the *first* direct RDP guarantee for general discrete local randomization mechanisms in the shuffle privacy model. In particular, given an *arbitrary discrete* local mechanism with ϵ_0-LDP guarantee, we provide an RDP guarantee for the shuffle model, as a function of ϵ_0 and the number of users n; see Theorem 3.1. This can be seen as an amplification by shuffling result for amplifying pure LDP guarantee to RDP guarantee via shuffling. In contrast, the existing amplification by shuffling results [7, 22, 24] amplify pure LDP guarantee to approximate DP guarantee.

When numerically evaluating our bound, we save a factor of $8\times$ compared to the state-of-the-art approximate DP guarantee for shuffle models in [24][2] combined with strong composition, with the number of iterations $T = 10^5$, LDP parameter $\epsilon_0 = 0.5$, and number of clients $n = 10^6$; see Figure 4a in Section 4 for such example regimes. Furthermore, characterizing the RDP of the shuffle model enables us to compute the RDP of shuffling with Poisson sub-sampling by using the results in [43]. We numerically show that this approach can lead to at least $10\times$ improvement in privacy guarantee. This is for $T = 10^4$, $\epsilon_0 = 3$, and $n = 10^6$. The comparison is with applying the strong composition theorem [33] after getting the state-of-the-art approximate DP of the shuffle model given in [24] with Poisson sub-sampling [35] (see Figure 5a in Section 4 for more such regimes). This in turn implies that we can accommodate at least $10\times$ more interactions for the same privacy budget in these cases. Moreover, our upper bounds also give several orders of magnitude improvement over the simple RDP bound stated in [21, Remark 1] (also stated in (9)) in several regimes (see, for example, Figure 2e in Section 4). We also develop a lower bound for the RDP for the shuffle model and numerically demonstrate that the gap is small for many parameter regimes of interest.

In order to obtain our upper bound result, we develop new analysis techniques which could be of independent interest. In particular, we develop a novel RDP analysis for neighboring datasets with a special structure (see Theorem 3.7), in which one of the datasets has all the data points to be the same (see the definition in (12)). A key technical result is then to relate the RDP of general neighboring datasets to those with special structure (see Theorem 3.6).

- For the RDP analysis of neighboring datasets with the above-mentioned special structure, we first observe that the output distribution of the shuffling mechanism is the multinomial distribution. Using this observation, then we show that the ratio of the distributions of the mechanism on special structure neighboring datasets is a sub-Gaussian random variable (r.v.), and we can write the Rényi divergence of the shuffle mechanism in terms of the moments of this r.v. Bounding the moments of this r.v. then gives an upper bound on the RDP for the special neighboring datasets. See the proof-sketch of Theorem 3.7 in Section 3.3.2 and its complete proof in Section 6.
- We next connect the above analysis to the RDP computation for general neighboring datasets $\mathcal{D} = (d_1, \ldots, d_n)$ and

$\mathcal{D}' = (d_1, \ldots, d_{n-1}, d'_n)$. To do so, a crucial observation is to write the output distribution $\boldsymbol{p}_i$ of the local randomizer $\mathcal{R}$ on the i'th client's data point d_i (for any $i \in [n-1]$) as a mixture distribution $\boldsymbol{p}_i = e^{-\epsilon_0} \boldsymbol{p}'_n + (1 - e^{-\epsilon_0}) \tilde{\boldsymbol{p}}_i$ for some $\tilde{\boldsymbol{p}}_i$.[3] So, the number of clients that sample according to $\boldsymbol{p}'_n$ is concentrated around $e^{-\epsilon_0} n$. Therefore, if we restrict the dataset to these clients only, the resulting datasets will have the special structure, and the size of that dataset will be concentrated around $e^{-\epsilon_0} n$. Finally, in order to be able to reduce the problem to the special case, we remove the effect of the clients that do not sample according to $\boldsymbol{p}'_n$ without affecting the Rényi divergence. See the proof-sketch of Theorem 3.6 in Section 3.3.1 and its complete proof in Section 5.

Related Work

We give the most relevant work related to the paper and put our contributions in the context of these works.

shuffle privacy model: As mentioned, the shuffle model of privacy has been of significant recent interest [5–8, 13, 22, 25, 26]. However, all the existing works in literature [7, 22, 24] only characterize the approximate DP of the shuffle model – among these, [24] is the state-of-the-art, but as we show in our experiments, it yields weaker results when combined with composition. To the best of our knowledge, there is no bound on RDP of the shuffle model in the literature except for the one mentioned briefly in a remark in [22, Remark 1] (which is obtained by the standard conversion results from DP to RDP) and we state it in (9) for comparison. However, this bound is loose (*e.g.*, see Figure 2e) and not useful for conversion to approximate DP (*e.g.*, see Figures 3a, 3c), as well as for composition (*e.g.*, see Figure 4e). Thus, our work makes progress on this important open question of analyzing the RDP of the shuffle model. Both [20] and [27] used advanced composition to analyze privacy of shuffle models in federated learning; our results could be adapted to enhance their privacy guarantees.

Rényi differential privacy: The work of Abadi *et al.* [1] provided a methodology to get stronger composition results. Inherently, this used Rényi divergence, and was later formalized in [36] which defined Rényi differential privacy (RDP). RDP presents a unified definition for several kinds of privacy notions including pure differential privacy (ϵ-DP), approximate differential privacy ((ϵ, δ)-DP), and concentrated differential privacy (CDP) [11, 18]. As mentioned earlier, RDP enables a stronger result for composition, through the "moment accounting" idea. Similarly, several works [37, 40, 43] have shown that analyzing the RDP of subsampled mechanisms provides a tighter bound on the total privacy loss than the bound that can be obtained using the standard strong composition theorems. However, to the best of our knowledge, RDP analysis of the shuffle model and its use for composition in the shuffle model has not been studied. In this paper, we analyze the RDP of the shuffle model, where we can bound the approximate DP of a sequence of shuffle models using the transformation from RDP to approximate DP [1, 2, 12, 40]. We show that our RDP analysis provides a better bound on the total

---

[2]We used the open source implementation for the privacy analysis in [24] available from https://github.com/apple/ml-shuffling-amplification.

[3]The idea of writing the distribution of the output of an LDP mechanism as a mixture distribution was previously proposed in [7, 24]. However, the way these mixture distributions are used in our RDP analysis is different from these works, including what mixtures we create and how we use them.

privacy loss of composition than that can be obtained using the standard strong composition theorems (see Section 4).

Discrete mechanisms: Many of the works in DP use specific randomization mechanisms, adding noise using the Laplace or Gaussian distributions. However, in many situations the data is inherently discrete (*e.g.,* see [12] and references therein) or compression causes it to be so (*e.g.,* see [27, 32] and references therein). It is therefore of interest to directly analyze privacy of discrete randomization mechanisms. Such discrete mechanisms have been studied extensively in shuffle models [5, 25], but for approximate DP. To the best of our knowledge, RDP for general discrete mechanisms in the shuffle privacy framework is new to our work.

Paper Organization

The paper is organized as follows. In Section 2, we give some preliminary definitions and results from the literature and also formulate our problem. In Section 3, we present our main results (two upper bounds and one lower bound on RDP), along with a proof sketch of the first upper bound. We also describe the two main ingredients in its proof – first is the reduction of computing RDP for the arbitrary pairs of neighboring datasets to computing RDP for the special pairs of neighboring datasets, and the second is computing RDP for the special pairs of neighboring datasets. In Section 4, we present several numerical results to demonstrate the advantages of our bounds compared to the state-of-the-art. The rest of the sections are devoted to the full proofs of our main results: Section 5 shows the reduction of our general problem to the special case; Section 6 proves the RDP for the special case; Section 7 proves both our upper bounds; and Section 8 proves our lower bound. In Section 9, we conclude with a short discussion. Omitted details from the proofs are provided in the appendices.

2 PRELIMINARIES AND PROBLEM FORMULATION

We give different privacy definitions that we use in Section 2.1, some existing results on RDP to DP conversion and RDP composition in Section 2.2, and give our problem formulation in Section 2.3.

2.1 Privacy Definitions

In this subsection, we define different privacy notions that we will use in this paper: local differential privacy (LDP), central differential privacy (DP), and Rényi differential privacy (RDP).

Definition 1 (Local Differential Privacy - LDP [34]). For $\epsilon_0 \geq 0$, a randomized mechanism $\mathcal{R} : \mathcal{X} \rightarrow \mathcal{Y}$ is said to be ϵ_0-local differentially private (in short, ϵ_0-LDP), if for every pair of inputs $d, d' \in \mathcal{X}$, we have

$$\Pr[\mathcal{R}(d) \in \mathcal{S}] \leq e^{\epsilon_0} \Pr[\mathcal{R}(d') \in \mathcal{S}], \qquad \forall \mathcal{S} \subseteq \mathcal{Y}. \tag{1}$$

Let $\mathcal{D} = \{d_1, \dots, d_n\}$ denote a dataset comprising n points from $\mathcal{X}$. We say that two datasets $\mathcal{D} = \{d_1, \dots, d_n\}$ and $\mathcal{D}' = \{d'_1, \dots, d'_n\}$ are neighboring (and denoted by $\mathcal{D} \sim \mathcal{D}'$) if they differ in one data point, i.e., there exists an $i \in [n]$ such that $d_i \neq d'_i$ and for every $j \in [n], j \neq i$, we have $d_j = d'_j$.

Definition 2 (Central Differential Privacy - DP [16, 17]). For $\epsilon, \delta \geq 0$, a randomized mechanism $\mathcal{M} : \mathcal{X}^n \rightarrow \mathcal{Y}$ is said to be (ϵ, δ)-differentially private (in short, (ϵ, δ)-DP), if for all neighboring datasets $\mathcal{D} \sim \mathcal{D}' \in \mathcal{X}^n$ and every subset $\mathcal{S} \subseteq \mathcal{Y}$, we have

$$\Pr[\mathcal{M}(\mathcal{D}) \in \mathcal{S}] \leq e^{\epsilon_0} \Pr[\mathcal{M}(\mathcal{D}') \in \mathcal{S}] + \delta. \tag{2}$$

Definition 3 (Rényi Differential Privacy - RDP [36]). A randomized mechanism $\mathcal{M} : \mathcal{X}^n \rightarrow \mathcal{Y}$ is said to have $\epsilon(\lambda)$-Rényi differential privacy of order $\lambda \in (1, \infty)$ (in short, $(\lambda, \epsilon(\lambda))$-RDP), if for any neighboring datasets $\mathcal{D} \sim \mathcal{D}' \in \mathcal{X}^n$, the Rényi divergence between $\mathcal{M}(\mathcal{D})$ and $\mathcal{M}(\mathcal{D}')$ is upper-bounded by $\epsilon(\lambda)$, i.e.,

$$D_\lambda(\mathcal{M}(\mathcal{D})||\mathcal{M}(\mathcal{D}')) = \frac{1}{\lambda - 1} \log \left(\mathbb{E}_{\theta \sim \mathcal{M}(\mathcal{D}')} \left[\left(\frac{\mathcal{M}(\mathcal{D})(\theta)}{\mathcal{M}(\mathcal{D}')(\theta)} \right)^\lambda \right] \right)$$
$$\leq \epsilon(\lambda),$$

where $\mathcal{M}(\mathcal{D})(\theta)$ denotes the probability that $\mathcal{M}$ on input $\mathcal{D}$ generates the output θ. For convenience, instead of $\epsilon(\lambda)$ being an upper bound, we define it as $\epsilon(\lambda) = \sup_{\mathcal{D} \sim \mathcal{D}'} D_\lambda(\mathcal{M}(\mathcal{D})||\mathcal{M}(\mathcal{D}'))$.

Our objective in this paper is to characterize the Rényi differential privacy of a shuffling mechanism $\mathcal{M}$ (see Section 2.3 for details) for different values of order λ.

2.2 RDP to DP Conversion and RDP Composition

In this subsection, we state some preliminary results from the literature that we will use. Though our main objective in this paper is to derive RDP guarantees of a shuffling mechanism, we also give the central privacy guarantees of that mechanism. For that purpose, we use the following result for converting the RDP guarantees of a mechanism to its DP guarantees. To the best of our knowledge, this result gives the best conversion.[4]

LEMMA 2.1 (FROM RDP TO DP [4, 12]). *Suppose for any $\lambda > 1$, a mechanism $\mathcal{M}$ is $(\lambda, \epsilon(\lambda))$-RDP. Then, the mechanism $\mathcal{M}$ is (ϵ, δ)-DP, where ϵ, δ are define below:*

For a given $\delta \in (0, 1)$:

$$\epsilon = \min_\lambda \epsilon(\lambda) + \frac{\log(1/\delta) + (\lambda - 1)\log(1 - 1/\lambda) - \log(\lambda)}{\lambda - 1}$$

For a given $\epsilon > 0$:

$$\delta = \min_\lambda \frac{\exp[(\lambda - 1)(\epsilon(\lambda) - \epsilon)]}{\lambda - 1} \left(1 - \frac{1}{\lambda} \right)^\lambda.$$

As mentioned in Section 1, the main strength of RDP in comparison to other privacy notions comes from composition. The following result states that if we adaptively compose two RDP mechanisms with the same order, their privacy parameters add up in the resulting mechanism.

LEMMA 2.2 (ADAPTIVE COMPOSITION OF RDP [36, PROPOSITION 1]). *For any $\lambda > 1$, let $\mathcal{M}_1 : \mathcal{X} \rightarrow \mathcal{Y}_1$ be a $(\lambda, \epsilon_1(\lambda))$-RDP mechanism and $\mathcal{M}_2 : \mathcal{Y}_1 \times \mathcal{X} \rightarrow \mathcal{Y}$ be a $(\lambda, \epsilon_2(\lambda))$-RDP mechanism. Then, the mechanism defined by $(\mathcal{M}_1, \mathcal{M}_2)$ satisfies $(\lambda, \epsilon_1(\lambda) + \epsilon_2(\lambda))$-RDP.*

---

[4]An optimal conversion from RDP to approximate DP was studied in [2]; however, we observed numerically, that it does not give better performance as compared to the conversion presented above.
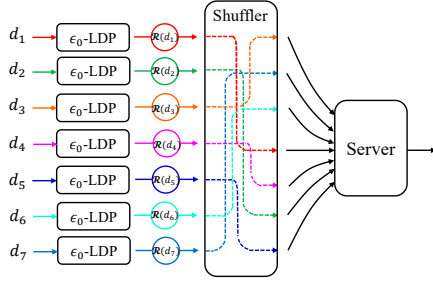
Figure 1: In the shuffle model, clients apply the local randomizer $\mathcal{R}$ on their data points and send them to a secure shuffler that randomly permutes clients' messages before passing them to the server.

By combining the above two lemmas, we can give the DP privacy guarantee of a sequence of T adaptive RDP mechanisms. Let $\mathcal{M}_1(\mathcal{I}_1, \mathcal{D}), \ldots, \mathcal{M}_T(\mathcal{I}_T, \mathcal{D})$ be a sequence of T adaptive mechanisms, where $\mathcal{I}_t$ denotes the auxiliary input to the t'th mechanism, which may depend on the previous mechanisms' outputs and the auxiliary inputs $\{(\mathcal{I}_i, \mathcal{M}_i(\mathcal{I}_i, \mathcal{D})) : i < t\}$, and the t'th mechanism $\mathcal{M}_t$ is $(\lambda, \epsilon_t(\lambda))$-RDP. Thus, from Lemma 2.2, we get that the mechanism defined by $(\mathcal{M}_1, \ldots, \mathcal{M}_T)$ is $\left(\lambda, \sum_{t=1}^{T} \epsilon_t(\lambda)\right)$-RDP, and then from Lemma 2.1 that the mechanism $(\mathcal{M}_1, \ldots, \mathcal{M}_T)$ satisfies (ϵ, δ)-DP, where ϵ, δ are given below:

For a given $\delta \in (0, 1)$:

$$\epsilon = \min_{\lambda} \sum_{t=1}^{T} \epsilon_t(\lambda) + \frac{\log(1/\delta) + (\lambda - 1)\log(1 - 1/\lambda) - \log(\lambda)}{\lambda - 1}$$

For a given $\epsilon > 0$:

$$\delta = \min_{\lambda} \frac{\exp\left[(\lambda - 1)\left(\sum_{t=1}^{T} \epsilon_t(\lambda) - \epsilon\right)\right]}{\lambda - 1}\left(1 - \frac{1}{\lambda}\right)^{\lambda}.$$

2.3 Problem Formulation

Let $\mathcal{D} = (d_1, \ldots, d_n)$ be a dataset consisting of n data points, where d_i is a data point at the i'th client that takes values from a set $\mathcal{X}$. Let $\mathcal{R} : \mathcal{X} \to \mathcal{Y}$ be a local randomizer that satisfies the following two properties:

(1) $\mathcal{R}$ is an ϵ_0-LDP mechanism (see Definition 1).
(2) The range of $\mathcal{R}$ is a discrete set, i.e., the output of $\mathcal{R}$ takes values in a discrete set $[B] = \{1, \ldots, B\}$ for some $B \in \mathbb{N} := \{1, 2, 3, \ldots\}$. Here, $[B]$ could be the whole of $\mathbb{N}$.

Client i applies $\mathcal{R}$ on d_i (each client uses independent randomness for computing $\mathcal{R}(d_i)$) and sends $\mathcal{R}(d_i)$ to the shuffler, who shuffles the received n inputs and outputs the result; see Figure 1. To formalize this, let $\mathcal{H}_n : \mathcal{Y}^n \to \mathcal{Y}^n$ denote the shuffling operation that takes n inputs and outputs their uniformly random permutation. We define the shuffling mechanism as

$$\mathcal{M}(\mathcal{D}) := \mathcal{H}_n(\mathcal{R}(d_1), \ldots, \mathcal{R}(d_n)). \tag{3}$$

Our goal is to characterize the Rényi differential privacy of $\mathcal{M}$.

Since the output of $\mathcal{M}$ is a random permutation of the n outputs of $\mathcal{R}$, the server cannot associate the n messages to the clients; and

the only information it can use from the messages is the histogram, i.e., the number of messages that give any particular output in $[B]$. We define a set $\mathcal{A}_B^n$ as follows

$$\mathcal{A}_B^n = \left\{\boldsymbol{h} = (h_1, \ldots, h_B) : \sum_{j=1}^{B} h_j = n\right\}, \tag{4}$$

to denote the set of all possible histograms of the output of the shuffler with n inputs. Therefore, we can assume, without loss of generality (w.l.o.g.), that the output of $\mathcal{M}$ is a distribution over $\mathcal{A}_B^n$ for input dataset $\mathcal{D}$ of n data points.

A succinct summary of the notation used throughout the paper is given in Table 1.

3 MAIN RESULTS

This section is dedicated to presenting the main results of this paper, along with their implications and comparisons with related work. We state two different upper bounds on the RDP of the shuffle model in Section 3.1 and a lower bound in Section 3.2. We present a detailed proof-sketch of our first upper bound in Section 3.3, along with all the main ingredients required to prove the upper bound.

3.1 Upper Bounds

In this subsection, we will provide two upper bounds.

THEOREM 3.1 (UPPER BOUND 1). *For any $n \in \mathbb{N}$, $\epsilon_0 \geq 0$, and any integer $\lambda \geq 2$, the RDP of the shuffle model is upper-bounded by*

$$\epsilon(\lambda) \leq \frac{1}{\lambda - 1} \log\left(1 + \binom{\lambda}{2}\frac{(e^{\epsilon_0} - 1)^2}{\bar{n}e^{\epsilon_0}}\right.$$
$$\left. + \sum_{i=3}^{\lambda} \binom{\lambda}{i} i\Gamma(i/2)\left(\frac{(e^{2\epsilon_0} - 1)^2}{2e^{2\epsilon_0}\bar{n}}\right)^{i/2} + e^{\epsilon_0\lambda - \frac{n-1}{8e^{\epsilon_0}}}\right), \tag{5}$$

where $\bar{n} = \lfloor\frac{n-1}{2e^{\epsilon_0}}\rfloor + 1$ and $\Gamma(z) = \int_0^{\infty} x^{z-1}e^{-x}dx$ is the Gamma function.

We give a proof sketch of Theorem 3.1 in Section 3.3 and provide its complete proof in Section 7.1.

When n, ϵ_0, λ satisfy a certain condition, we can simplify the bound in (5) to the following:

COROLLARY 3.2 (SIMPLIFIED UPPER BOUND 1). *For any $n \in \mathbb{N}$, $\epsilon_0 \geq 0$, and any integer $\lambda \geq 2$ that satisfy $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$, we can simplify the bound in (5) to the following:*

$$\epsilon(\lambda) \leq \frac{1}{\lambda - 1} \log\left(1 + \binom{\lambda}{2}\frac{4(e^{\epsilon_0} - 1)^2}{n}\right). \tag{6}$$

We prove Corollary 3.2 in Appendix A.

Note that the upper bounds in Theorem 3.1 and Corollary 3.2 hold for any ϵ_0-LDP mechanism.

Remark 1. Note that any λ, ϵ_0, n that satisfy $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$ lead to the bound in (6). For example, we can take $\epsilon_0 = c \ln n$ and $\lambda < \frac{n^{(1-5c)/4}}{2}$ for any $c < \frac{1}{5}$, and it will satisfy the condition. In particular, taking $\epsilon_0 = \frac{1}{25} \ln n$ and $\lambda < \frac{n^{1/5}}{2}$ will also give the bound in (6).

Remark 2 (Generalization to real orders λ). Theorem 3.1 provides an upper bound on the RDP of the shuffle model for only integer orders $\lambda \geq 2$. However, the result can be generalized to real orders

Symbol	Description
$[B]$	$\{1, 2, \ldots, B\}$ for any $B \in \mathbb{N}$
ϵ_0	LDP parameter (see Definition 1)
(ϵ, δ)	Approximate DP parameters (see Definition 2)
$(\lambda, \epsilon(\lambda))$	RDP parameters (see Definition 3)
$\mathcal{R} : \mathcal{X} \to [B]$	A discrete ϵ_0-LDP mechanism at clients for mapping their data points to elements in $[B]$
$\boldsymbol{p} = (p_1, \ldots, p_B)$	The output distribution of $\mathcal{R}$ when the data point is d
$\boldsymbol{p}' = (p'_1, \ldots, p'_B)$	The output distribution of $\mathcal{R}$ when the data point is d'
$\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iB})$ for $i \in [n]$	The output distribution of $\mathcal{R}$ when the data point is d_i
$\boldsymbol{p}'_n = (p'_{n1}, \ldots, p'_{nB})$	The output distribution of $\mathcal{R}$ when the data point is d'_n
$\mathcal{P}$	A collection of n distributions $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\}$
$\mathcal{P}_{-i}$	A collection of $(n-1)$ distributions $\mathcal{P} \setminus \{\boldsymbol{p}_i\}$
$\mathcal{P}_C$, where $C \subseteq [n-1]$	A collection of n distributions, where clients in the set C map according to $\boldsymbol{p}'_n$, clients in the set $[n-1] \setminus C$ map according to $\tilde{\boldsymbol{p}}_i$ (see (18)), and client n maps according to $\boldsymbol{p}_n$ (see (19)-(21))
$\mathcal{A}^n_B$	A set of all possible histograms with B bins and n elements (see (4))
$\boldsymbol{h}$	$\boldsymbol{h} = (h_1, \ldots, h_B)$ with $\sum_{i=1}^B h_i = n$ is an element of $\mathcal{A}^n_B$
$\mathcal{M}(\mathcal{D})$	The shuffle mechanism $\mathcal{M}$ on the dataset $\mathcal{D} \in \mathcal{X}^n$; $\mathcal{M}(\mathcal{D})$ is a distribution over $\mathcal{A}^n_B$ (see (3))
$F(\mathcal{P})$	Distribution over $\mathcal{A}^n_B$ when client i maps its data point according to the distribution $\boldsymbol{p}_i$ (see (17))

Table 1: Notation used throughout the paper

λ using convexity of the function $(\lambda - 1) \epsilon(\lambda)$ as follows. From [39, Corollary 2], the function $(\lambda - 1) D_\lambda (\mathbf{P} || \mathbf{Q})$ is convex in λ for any given two distributions $\mathbf{P}$ and $\mathbf{Q}$. Thus, for any real order $\lambda > 1$, we can bound the RDP of the shuffle model by

$$\epsilon(\lambda) \le \frac{a \cdot (\lfloor \lambda \rfloor - 1) \cdot \epsilon(\lfloor \lambda \rfloor) + (1 - a) \cdot (\lceil \lambda \rceil - 1) \cdot \epsilon(\lceil \lambda \rceil)}{\lambda - 1}, \quad (7)$$

where $a = \lceil \lambda \rceil - \lambda$, since $\lambda = a\lfloor \lambda \rfloor + (1 - a)\lceil \lambda \rceil$ for any real λ. Here, $\lfloor \lambda \rfloor$ and $\lceil \lambda \rceil$ respectively denote the largest integer smaller than or equal to λ and the smallest integer bigger than or equal to λ.

In the following theorem, we also present another bound on RDP that readily holds for all $\lambda \ge 1$.

THEOREM 3.3 (UPPER BOUND 2). *For any $n \in \mathbb{N}$, $\epsilon_0 \ge 0$, and any $\lambda \ge 1$ (including the non-integral λ), the RDP of the shuffle model is upper-bounded by*

$$\epsilon(\lambda) \le \frac{1}{\lambda - 1} \log \left(e^{\lambda^2 \frac{(e^{\epsilon_0} - 1)^2}{\bar{n}}} + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}} \right), \quad (8)$$

where $\bar{n} = \lfloor \frac{n-1}{2e^{\epsilon_0}} \rfloor + 1$.

We prove Theorem 3.3 in Section 7.2.

Remark 3 (Improved Upper Bounds – Saving a Factor of 2). The exponential term $e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}}$ in both the upper bounds stated in (5) and (8) comes from the Chernoff bound, where we naively choose the factor $\gamma = 1/2$ instead of optimizing it; see the proof of Theorem 3.1 in Section 7.1. If we instead had optimized γ and chosen it to be, for example, $\gamma = \sqrt{\frac{2\epsilon_0 e^{\epsilon_0}}{\sqrt{n} \log(n)}}$ (which goes to 0 when, say, $\epsilon_0 \le \frac{1}{4} \log(n)$), we would have asymptotically saved a multiplicative factor of 2 in the leading term in both upper bounds, because

in this case we have $\bar{n} = \lfloor (1 - \gamma) \frac{n-1}{e^{\epsilon_0}} \rfloor + 1 \to \lfloor \frac{n-1}{e^{\epsilon_0}} \rfloor + 1$ as $n \to \infty$. We chose to evaluate our bound with $\gamma = 1/2$ because of two reasons: first, it gives a simpler expression to compute; and second, the evaluated bound does not give good results (as compared to the ones with $\gamma = 1/2$) for the parameter ranges of interest.

Remark 4 (Difference in Upper Bounds). Since the quadratic term in λ inside the log in (8) has an extra multiplicative factor of e^{ϵ_0} in comparison with the corresponding term in (5), our first upper bound presented in Theorem 3.1 is better than our second upper bound presented in Theorem 3.3 for all parameter ranges of interest; see also Figure 2 in Section 4. However, the expression in (8) is much cleaner to state as well as to compute as compared to that in (5). As we will see later, the techniques required to prove both upper bounds are different.

Remark 5 (Potentially Better Upper Bounds for Specific Mechanisms). Since both our upper bounds are worse-case bounds that hold for *all* ϵ_0-LDP mechanisms, it is possible that for specific mechanisms, we may be able to exploit their structure for potentially better bounds. See Remark 8 on this just after (32).

The upper bounds on the RDP of the shuffle model presented in (5) and (8) are general and hold for any discrete ϵ_0-LDP mechanism. Furthermore, these bounds are in closed form expressions that can be easily implemented. To the best of our knowledge, there is no bound on RDP of the shuffle model in literature except for the one given in [21, Remark 1], which we provide below[5] in (9). For the LDP parameter ϵ_0 and number of clients n, they showed that

---

[5] As mentioned in Section 1, this was obtained by the standard conversion results from DP to RDP, which could be loose.

for any $\lambda > 1$, the shuffle mechanism $\mathcal{M}$ is $(\lambda, \epsilon(\lambda))$-RDP, where

$$\epsilon(\lambda) = \lambda \frac{2e^{4\epsilon_0}(e^{\epsilon_0} - 1)^2}{n}. \tag{9}$$

In Section 4, we evaluate numerically the performance of both our bounds (from Theorems 3.1 and 3.3) against the above bound in (9). We demonstrate that both our bounds outperform the above bound in all cases; and in particular, the gap is significant when $\epsilon_0 > 1$ – note that the bound in [22] is worse than our simplified bound given in Corollary 3.2 by a multiplicative factor of $e^{4\epsilon_0}$.

3.2 Lower Bound

In this subsection, we provide a lower bound on the RDP for any integer order λ satisfying $\lambda \geq 2$.

THEOREM 3.4 (LOWER BOUND). *For any $n \in \mathbb{N}$, $\epsilon_0 \geq 0$, and any integer $\lambda \geq 2$, the RDP of the shuffle model is lower-bounded by:*

$$\epsilon(\lambda) \geq \frac{1}{\lambda - 1} \log \left(1 + \binom{\lambda}{2} \frac{(e^{\epsilon_0} - 1)^2}{ne^{\epsilon_0}} \right.$$
$$\left. + \sum_{i=3}^{\lambda} \binom{\lambda}{i} \left(\frac{(e^{2\epsilon_0} - 1)}{ne^{\epsilon_0}} \right)^i \mathbb{E}\left[\left(k - \frac{n}{e^{\epsilon_0} + 1} \right)^i \right] \right), \tag{10}$$

where expectation is taken w.r.t. the binomial random variable $k \sim Bin(n, p)$ with parameter $p = \frac{1}{e^{\epsilon_0}+1}$.

We give a proof-sketch of Theorem 3.4 in Section 8 and provide its complete proof in Appendix E.

When i is an even integer, then the expectation term in (10) is positive. When $i \geq 3$ is an odd integer, then using the convexity of function $f(x) = x^i$, it follows from the Jensen's inequality (i.e., $\mathbb{E}f(X) \geq f(\mathbb{E}X)$) and $\mathbb{E}[k] = \frac{n}{e^{\epsilon_0}+1}$, that $\mathbb{E}\left[\left(k - \frac{n}{e^{\epsilon_0}+1} \right)^i \right] \geq \left(\mathbb{E}\left[k - \frac{n}{e^{\epsilon_0}+1} \right] \right)^i = 0$. Using these observations, we can safely ignore the summation term from (10) and obtain the following simplified lower bound.

COROLLARY 3.5 (SIMPLIFIED LOWER BOUND). *For any $n \in \mathbb{N}$, $\epsilon_0 \geq 0$, and integer $\lambda \geq 2$, the RDP of the shuffle model is lower-bounded by:*

$$\epsilon(\lambda) \geq \frac{1}{\lambda - 1} \log \left(1 + \binom{\lambda}{2} \frac{(e^{\epsilon_0} - 1)^2}{ne^{\epsilon_0}} \right). \tag{11}$$

Remark 6 (Upper and Lower Bound Proofs). Both our upper bounds stated in Theorems 3.1 and 3.3 hold for any ϵ_0-LDP mechanism. In other words, they are the worst case privacy bounds, in the sense that there is no ϵ_0-LDP mechanism for which the associated shuffle model gives a higher RDP parameter than those stated in (5) and (8). Therefore, the lower bound that we derive should serve as the lower bound on the RDP privacy parameter of the mechanism that achieves the largest privacy bound (i.e., worst privacy).

We prove our lower bound result (stated in Theorem 3.4) by showing that a specific mechanism (in particular, the binary Randomized response (RR)) on a specific pair of neighboring datasets yields the RDP privacy parameter stated in the right hand side (RHS) of (10). This implies that RDP privacy bound (which is the supremum over all neighboring datasets) of binary RR for the shuffle model is at least the bound stated in (10), which in turn implies that the lower bound (which is the tightest bound for any ϵ_0-LDP mechanism) is also at least that.

Remark 7 (Gap in Upper and Lower Bounds). When comparing our simplified upper and lower bounds from Corollaries 3.2 and 3.5, respectively, we observe that when $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$, our upper and lower bounds differ by a multiplicative factor of $4e^{\epsilon_0}$. In our generic upper bound (5), note that when n is large, only the term corresponding to λ^2 matters, and with our improved upper bound (which saves a factor of 2 in that term asymptotically – see Remark 3), the upper and lower bounds are away by the factor of e^{ϵ_0}, which tends to 1 as $\epsilon_0 \to 0$. Thus, in the regime of large n and small ϵ_0, our upper and lower bounds coincide. Without any constraints on n, ϵ_0, we believe that our lower bound is tight. Closing this gap by showing a tighter upper bound is an interesting and important open problem.

3.3 Proof Sketch of Theorem 3.1

The proof has two main steps. In the first step, we reduce the problem of deriving RDP for arbitrary neighboring datasets to the problem of deriving RDP for specific neighboring datasets, $\mathcal{D}, \mathcal{D}'$, where all elements in $\mathcal{D}$ are the same and $\mathcal{D}'$ differs from $\mathcal{D}$ in one entry. In the second step, we derive RDP for the special neighboring datasets. Details follow:

The specific neighboring datasets to which we reduce our general problem have the following form:

$$\mathcal{D}_{\text{same}}^m = \Big\{ (\mathcal{D}_m, \mathcal{D}'_m) : \mathcal{D}_m = (d, \dots, d, d) \in \mathcal{X}^m,$$
$$\mathcal{D}'_m = (d, \dots, d, d') \in \mathcal{X}^m, \text{ where } d, d' \in \mathcal{X} \Big\}. \tag{12}$$

Consider arbitrary neighboring datasets $\mathcal{D} = (d_1, \dots, d_n) \in \mathcal{X}^n$ and $\mathcal{D}' = (d_1, \dots, d_{n-1}, d'_n) \in \mathcal{X}^n$. For any $m \in \{0, \dots, n-1\}$, define new neighboring datasets $\mathcal{D}_{m+1}^{(n)} = (d'_n, \dots, d'_n, d_n) \in \mathcal{X}^{m+1}$ and $\mathcal{D}_{m+1}'^{(n)} = (d'_n, \dots, d'_n, d'_n) \in \mathcal{X}^{m+1}$, each having $(m+1)$ elements. Observe that $\left(\mathcal{D}_{m+1}'^{(n)}, \mathcal{D}_{m+1}^{(n)} \right) \in \mathcal{D}_{\text{same}}^{m+1}$. The first step of our proof is summarized in the following theorem.

THEOREM 3.6 (REDUCTION TO THE SPECIAL CASE). *Let $q = \frac{1}{e^{\epsilon_0}}$ and $m \sim Bin(n-1, q)$ be a binomial random variable. We have:*

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')} \left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})} \right)^{\lambda} \right]$$
$$\leq \mathbb{E}_{m \sim Bin(n-1,q)} \left[\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})} \left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})} \right)^{\lambda} \right] \right]. \tag{13}$$

We give a proof-sketch of Theorem 3.6 in Section 3.3.1 and provide its complete proof in Section 5.

We know (by Chernoff bound) that the binomial random variable is concentrated around its mean, which implies that the terms in the RHS of (13) that correspond to $m < (1-\gamma)q(n-1)$ (we will take $\gamma = 1/2$) will contribute in a negligible amount. Then we show in Lemma D.1 (on page 20) that $E_m := \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})} \left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})} \right)^{\lambda} \right]$ is a non-increasing function of m. These observations together imply that the RHS in (13) is approximately upper bounded by $E_{(1-\gamma)q(n-1)}$ via Chernoff bound.

Since E_m is precisely what is required to bound the RDP for the specific neighboring datasets, we have reduced the problem of computing RDP for arbitrary neighboring datasets to the problem of

computing RDP for specific neighboring datasets. The second step of the proof bounds $E_{(1-\gamma)q(n-1)}$, which follows from the result below that holds for any $m \in \mathbb{N}$.

Theorem 3.7 (RDP for the Special Case). *Let $m \in \mathbb{N}$ be arbitrary. For any integer $\lambda \geq 2$, we have*

$$\sup_{(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}^m_{\text{same}}} \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} \right)^{\lambda} \right]$$

$$\leq 1 + \binom{\lambda}{2} \frac{(e^{\epsilon_0} - 1)^2}{m e^{\epsilon_0}} + \sum_{i=3}^{\lambda} \binom{\lambda}{i} i \Gamma(i/2) \left(\frac{(e^{2\epsilon_0} - 1)^2}{2m e^{2\epsilon_0}} \right)^{i/2} . \quad (14)$$

We give a proof-sketch of Theorem 3.7 in Section 3.3.2 and provide its complete proof in Section 6.

Substituting $m = (1 - \gamma)q(n - 1) + 1$ in (14) yields the bound in Theorem 3.1.

3.3.1 Proof Sketch of Theorem 3.6. For $i \in [n]$, let $\boldsymbol{p}_i$ denote the distribution of the ϵ_0-LDP mechanism $\mathcal{R}$ when the input data point is d_i, and $\boldsymbol{p}'_n$ denote the distribution of $\mathcal{R}$ when the input data point is d'_n. The main idea of the proof is the observation that each distribution $\boldsymbol{p}_i$ can be written as the following mixture distribution: $\boldsymbol{p}_i = \frac{1}{e^{\epsilon_0}} \boldsymbol{p}'_n + \left(1 - \frac{1}{e^{\epsilon_0}}\right) \tilde{\boldsymbol{p}}_i$, where $\tilde{\boldsymbol{p}}_i$ is a certain distribution associated with $\boldsymbol{p}_i$. So, instead of client $i \in [n - 1]$ mapping its data point d_i according to $\boldsymbol{p}_i$, we can view it as the client i maps d_i according to $\boldsymbol{p}'_n$ with probability $\frac{1}{e^{\epsilon_0}}$ and according to $\tilde{\boldsymbol{p}}_i$ with probability $(1 - \frac{1}{e^{\epsilon_0}})$. Thus the number of clients that sample from the distribution $\boldsymbol{p}'_n$ follows a binomial distribution $\text{Bin}(n - 1, q)$ with parameter $q = \frac{1}{e^{\epsilon_0}}$. This allows us to write the distribution of $\mathcal{M}$ when clients map their data points according to $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n, \boldsymbol{p}'_n$ as a convex combination of the distribution of $\mathcal{M}$ when clients map their data points according to $\tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_{n-1}, \boldsymbol{p}_n, \boldsymbol{p}'_n$; see Lemma 5.1. Then using a joint convexity argument (see Lemma 5.2), we write the Rényi divergence between the original pair of distributions of $\mathcal{M}$ in terms of the same convex combination of the Rényi divergence between the resulting pairs of distributions of $\mathcal{M}$ as in Lemma 5.1. Using a monotonicity argument (see Lemma 5.3), we can remove the effect of clients that do not sample from the distribution $\boldsymbol{p}'_n$ without decreasing the Rényi divergence. By this chain of arguments, we have reduced the problem to the one involving the computation of Rényi divergence only for the special form of neighboring datasets, which proves Theorem 3.6. Details can be found in Section 5.

3.3.2 Proof Sketch of Theorem 3.7. Consider any pair of special neighboring datasets $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}^m_{\text{same}}$ for any $m \in \mathbb{N}$. Using the polynomial expansion, we get

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} \right)^{\lambda} \right]$$

$$= \sum_{i=0}^{\lambda} \binom{\lambda}{i} \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} - 1 \right)^{i} \right] . \quad (15)$$

Let $X : \mathcal{A}^m_B \to \mathbb{R}$ denote a random variable (r.v.) associated with the distribution $\mathcal{M}(\mathcal{D}_m)$, and for every $\boldsymbol{h} \in \mathcal{A}^m_B$, is defined as $X(\boldsymbol{h}) = m \left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} - 1 \right)$. With this, we can rewrite (15) in terms of the moments of X. Then we show that X is a sub-Gaussian r.v. that

has zero-mean and bounded variance. Using the sub-Gaussianity of X, we bound its higher moments (see Lemma 6.1). Substituting these bounds in (15) proves Theorem 3.7. Details can be found in Section 6.

4 NUMERICAL RESULTS

In this section, we present numerical experiments to show the performance of our bounds on the RDP of the shuffle model and its usage for getting approximate DP and composition results.

RDP of the shuffle model: In Figure 2, we plot several bounds on the RDP of the shuffle model in different regimes. In particular, we compare between the first upper bound on the RDP given in Theorem 3.1, the second upper bound on the RDP given in Theorem 3.3, the lower bound on the RDP given in Theorem 3.4, and the upper bound on the RDP given in [22, Remark 1] and stated in (9).[6] It is clear that our first upper bound (5) gives a tighter bound on the RDP in comparison with the second bound (8) and the upper bound given in [22]. Furthermore, the first upper bound is close to the lower bound for small values of the LDP parameter ϵ_0 and for high orders λ. In addition, the gap between our proposed bound in Theorem 3.1 and the bound given in [22] increases as the LDP parameter ϵ_0 increases. We also observe that the curves of the lower and upper bounds on the RDP of the shuffle model saturate close to ϵ_0 when the order λ approaches to infinity. This indicates that the pure DP of the shuffle model is bounded below by ϵ_0, an observation made in literature [3, 22]. As can be seen in Figures 2d and 2e, the RDP obtained by standard approximate DP to RDP conversion in [22, Remark 1], can be several orders of magnitude loose in comparison to our analysis.

Approximate DP of the shuffle model: Analyzing RDP of the shuffle model provides a bound on the approximate DP of the shuffle model from the relation between the RDP and approximate DP as shown in Lemma 2.1. In Figure 3, we plot several bounds on the approximate (ϵ, δ)-DP of the shuffle model for fixed $\delta = 10^{-6}$. In Figures 3d and 3b, we do not plot the results given in [22], since their bounds are quite loose and are far from the plotted range when $\epsilon_0 > 1$. We can see that our analysis of the RDP of the shuffle model provides a tighter bound on the approximate DP of the shuffle model in comparison with the bound given in [7] in some regimes. However, our RDP analysis performs worse than the best known bound given in [24], when used without composition. This might be due to the gap between our upper and lower bound on the RDP of the shuffle model as the lower bound provides better performance than the bound given in [24] for all values of LDP parameter ϵ_0. Note that the main use case for converting our RDP analysis to approximate DP is after composition rather than in the single-shot conversion illustrated in Figure 3.

Composition of a sequence of shuffle models: We now numerically evaluate the privacy parameters of the approximate (ϵ, δ)-DP for a composition of T mechanisms $(\mathcal{M}_1, \ldots, \mathcal{M}_T)$, where $\mathcal{M}_t$ is a shuffle mechanism for all $t \in [T]$. In Figure 4, we plot three different bounds on the overall privacy parameter ϵ for fixed $\delta = 10^{-8}$ for a composition of T identical shuffle models. The first bound on

---

[6]The results in [24] are for approximate DP (not for RDP), that is why we did not compare with them in Figure 2.

(a) RDP as a function of λ for $\epsilon_0 = 0.1$ and $n = 10^4$

(b) RDP as a function of n for $\epsilon_0 = 0.1$ and $\lambda = 100$

(c) RDP as a function of ϵ_0 for $n = 10^4$ and $\lambda = 100$

(d) RDP as a function of λ for $\epsilon_0 = 3$ and $n = 10^4$

(e) RDP as a function of n for $\epsilon_0 = 3$ and $\lambda = 100$

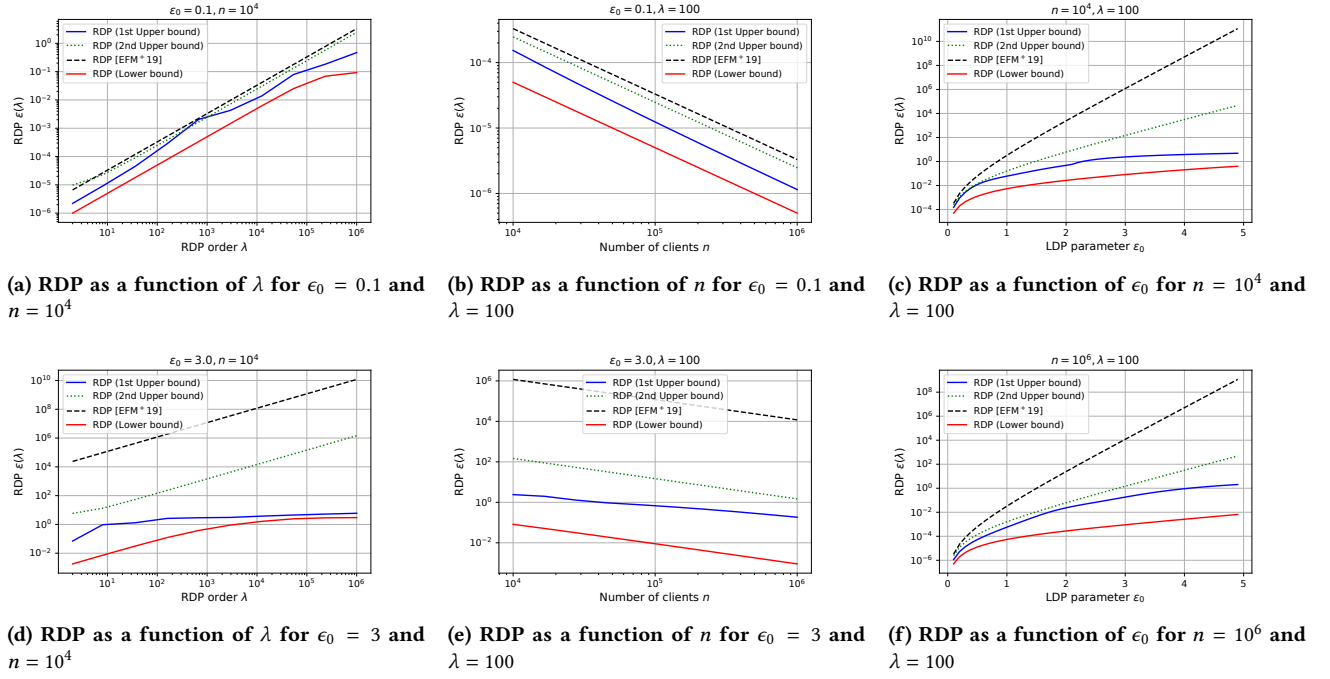**(f) RDP as a function of ϵ_0 for $n = 10^6$ and $\lambda = 100$**

Figure 2: Comparison of several bounds on the RDP of the shuffle model: (i) Our first upper bound (5) in Theorem 3.1; (ii) Our second upper bound (8) in Theorem 3.1; (iii) Our lower bound proposed in Theorem 3.4; and (iv) The upper bound on the RDP of the shuffle model given in [22, Remark 1].
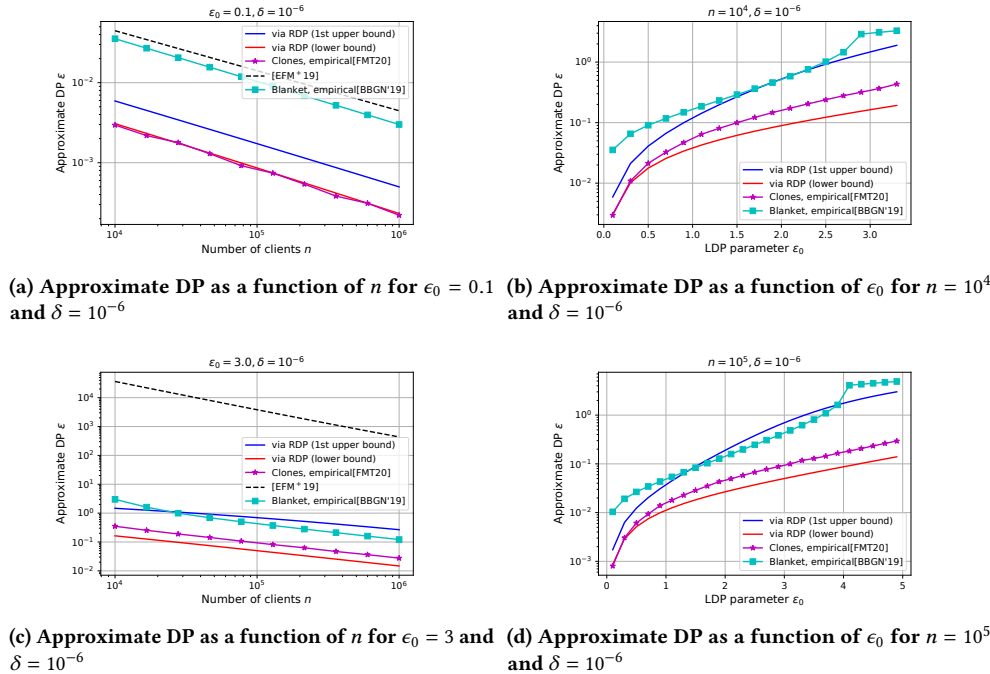

(a) Approximate DP as a function of n for $\epsilon_0 = 0.1$ and $\delta = 10^{-6}$

(b) Approximate DP as a function of ϵ_0 for $n = 10^4$ and $\delta = 10^{-6}$

(c) Approximate DP as a function of n for $\epsilon_0 = 3$ and $\delta = 10^{-6}$

(d) Approximate DP as a function of ϵ_0 for $n = 10^5$ and $\delta = 10^{-6}$

Figure 3: Comparison of several bounds on the Approximate (ϵ, δ)-DP of the shuffle model for $\delta = 10^{-6}$: (i) Approximate DP obtained from our first upper bound (5) of the RDP in Theorem 3.1; (ii) Approximate DP obtained from our lower bound on the RDP proposed in Theorem 3.4; (iii) The empirical upper bound on the approximate DP given in [24]; (iv) The theoretical bound on the approximate DP given in [22]; and (v) The generic bound on the approximate DP given in [7].

the overall privacy parameter ϵ is obtained as a function of δ and the number of iterations T by optimizing over the RDP order λ using our upper bound on the RDP of the shuffle model given in

Theorem 3.1. The second bound is obtained by optimizing over the RDP order λ using the upper bound on the RDP of the shuffle model given in [22]. The third bound is obtained by first computing the
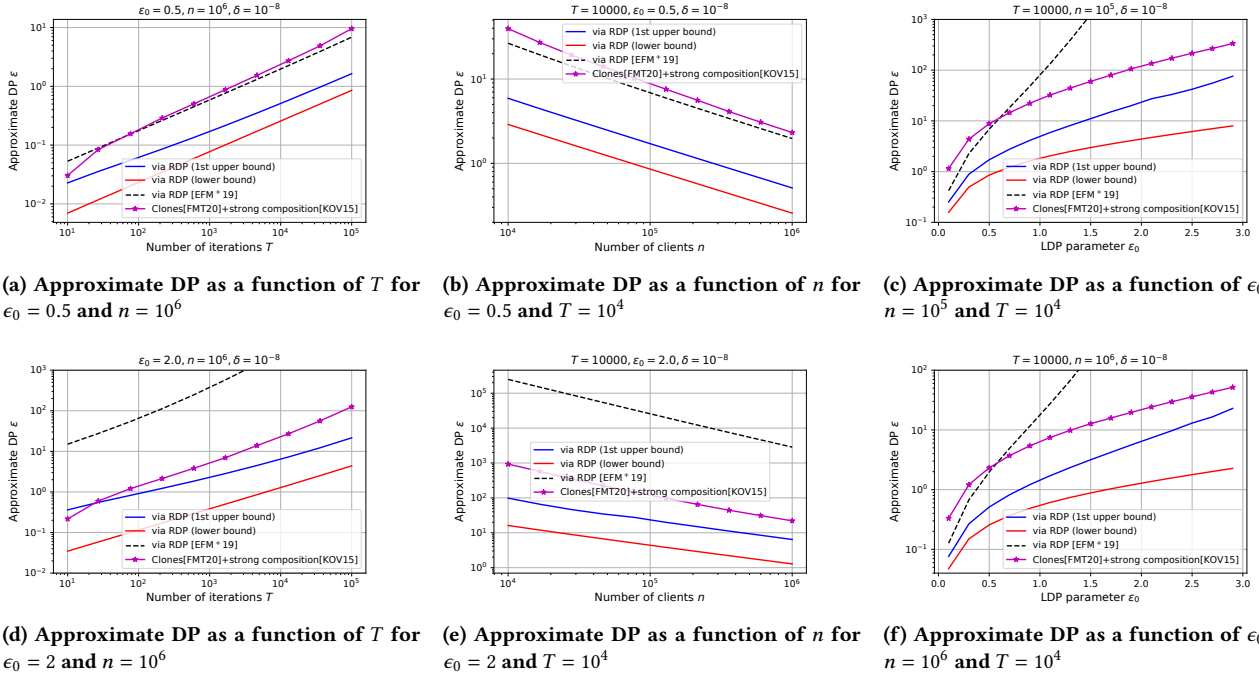
Figure 4: Comparison of several bounds on the Approximate (ϵ, δ)-DP for composition of a sequence of shuffle models for $\delta = 10^{-8}$: (i) Approximate DP obtained from our first upper bound (5) on the RDP; (ii) Approximate DP obtained from our lower bound on the RDP proposed in Theorem 3.4; (iii) Approximate DP obtained from the upper bound on the RDP given in [22]; and (iv) Applying the strong composition theorem [33] after getting the approximate DP of the shuffle model given in [24].

privacy parameters $(\tilde{\epsilon}, \tilde{\delta})$ of the shuffle model given in [24]. Then, we use the strong composition theorem given in [33] to obtain the overall privacy loss ϵ. We observe that there is a significant saving in the overall privacy parameter ϵ-DP using our bound on RDP in comparison with using the bound on DP [24] with the strong composition theorem [33]. For example, we save a factor of 8× in computing the overall privacy parameter ϵ for number of iterations $T = 10^5$, LDP parameter $\epsilon_0 = 0.5$, and number of clients $n = 10^6$. We observe that the bound given in [24] with the strong composition theorem [33] behaves better for small number of iterations $T < 10$ and large LDP parameter $\epsilon_0 = 2$. However, the typical number of iterations T in the standard SGD algorithm is usually larger. Therefore, this demonstrates the significance of our RDP analysis for composition in the regimes of interest.

Privacy amplification by shuffling and Poisson sub-sampling: In the Differentially Private Stochastic Gradient Descent (DP-SGD), shuffling and sampling the dataset at each iteration are important tools to provide a strong privacy guarantee [20, 28]. In these frameworks, the further advantage of sampling with shuffling[7] can be analyzed by standard combination of approximate DP with Poisson subsampling [35]. The resulting approximate DP along with the strong composition theorem given in [33] gives the overall privacy loss ϵ. An alternate path we use is to combine our RDP analysis

---

[7]In this framework we assume that the sampling and shuffling is done by a secure mechanism which is separated from the server, *i.e.*, the server does not know which clients are participating.
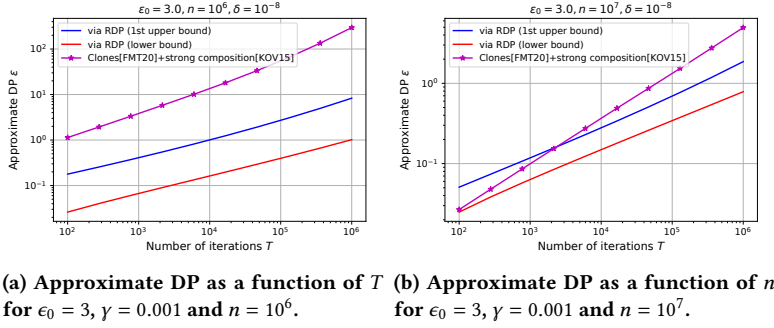
with sampling of RDP mechanisms using [40, 43]. This enables us to get an RDP guarantee with sampling, which we can then compose using properties of RDP. We can use the conversion from RDP to approximate DP to obtain a bound on the overall privacy loss of multiple iterations. In Figure 5, we compare our results of amplifying the RDP of the shuffle model by Poisson sub-sampling to the strong composition [33] after getting the approximate DP of the shuffle model given in [24] with Poisson sub-sampling given in [35]. We observe that we save a factor of 11× by using our RDP bound for $n = 10^6$ and $\gamma = 0.001$. However, we can see that the gap between our (lower/upper) bounds and the strong composition decreases when $n = 10^7$. This could be due to the simplistic combination of our analysis with the RDP subsampling of [43].

5 PROOF OF THE REDUCTION TO THE SPECIAL CASE

In this section, we will prove Theorem 3.6 by reducing the problem of computing RDP for the arbitrary pairs of neighboring datasets to the problem of computing RDP for the neighboring datasets with the special structure.

Recall that the LDP mechanism $\mathcal{R} : \mathcal{X} \to \mathcal{Y}$ has a discrete range $\mathcal{Y} = [B]$ for some $B \in \mathbb{N}$. Let $\boldsymbol{p}_i := (p_{i1}, \dots, p_{iB})$ and $\boldsymbol{p}'_n := (p'_{n1}, \dots, p'_{nB})$ denote the probability distributions over $\mathcal{Y}$ when the input to $\mathcal{R}$ is d_i and d'_n, respectively, where $p_{ij} = \Pr[\mathcal{R}(d_i) = j]$ and $p'_{nj} = \Pr[\mathcal{R}(d'_n) = j]$ for all $j \in [B]$ and $i \in [n]$. Let $\mathcal{P} = \{\boldsymbol{p}_i : i \in [n]\}$ and $\mathcal{P}' = \{\boldsymbol{p}_i : i \in [n-1]\} \bigcup \{\boldsymbol{p}'_n\}$. For $i \in [n-1]$, let $\mathcal{P}_{-i} =$

(a) Approximate DP as a function of T for $\epsilon_0 = 3, \gamma = 0.001$ and $n = 10^6$.

(b) Approximate DP as a function of n for $\epsilon_0 = 3, \gamma = 0.001$ and $n = 10^7$.

Figure 5: Comparison of several bounds on the Approximate (ϵ, δ)-DP for composition a sequence of shuffle models with Poisson sub-sampling for $\delta = 10^{-8}$ and $\gamma = 0.001$: (i) Approximate DP obtained from our first upper bound (5) on the RDP in Theorem 3.1 with RDP amplification by Poisson sub-sampling [43]; (ii) Approximate DP obtained from our lower bound on the RDP proposed in Theorem 3.4 with RDP amplification by Poisson sub-sampling [43]; and (iii) Applying the strong composition theorem [33] after getting the approximate DP of the shuffle model given in [24] with Poisson sub-sampling [35].

$\mathcal{P} \setminus \{\boldsymbol{p}_i\}, \mathcal{P}'_{-i} = \mathcal{P}' \setminus \{\boldsymbol{p}_i\}$, and also $\mathcal{P}_{-n} = \mathcal{P} \setminus \{\boldsymbol{p}_n\}, \mathcal{P}'_{-n} = \mathcal{P}' \setminus \{\boldsymbol{p}'_n\}$. Here, $\mathcal{P}, \mathcal{P}'$ correspond to the datasets $\mathcal{D} = \{d_1, \ldots, d_n\}, \mathcal{D}' = \{d_1, \ldots, d_{n-1}, d'_n\}$, respectively, and for any $i \in [n]$, $\mathcal{P}_{-i}$ and $\mathcal{P}'_{-i}$ correspond to the datasets $\mathcal{D}_{-i} = \{d_1, \ldots, d_{i-1}, d_{i+1}, \ldots, d_n\}$ and $\mathcal{D}'_{-i} = \{d_1, \ldots, d_{i-1}, d_{i+1}, \ldots, d_{n-1}, d'_n\}$, respectively.

For any collection $\mathcal{P} = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\}$ of n distributions, we define $F(\mathcal{P})$ to be the distribution over $\mathcal{A}_B^n$ (which is the set of histograms on B bins with n elements as defined in (4)) that is induced when every client i (independent to the other clients) samples an element from $[B]$ accordingly to the probability distribution $\boldsymbol{p}_i$. Formally, for any $\boldsymbol{h} \in \mathcal{A}_B^n$, define

$$\mathcal{U}_{\boldsymbol{h}} := \left\{ (\mathcal{U}_1, \ldots, \mathcal{U}_B) : \mathcal{U}_1, \ldots, \mathcal{U}_B \subseteq [n] \right.$$
$$\left. \text{s.t.} \bigcup_{j=1}^{B} \mathcal{U}_j = [n] \text{ and } |\mathcal{U}_j| = h_j, \forall j \in [B] \right\}. \quad (16)$$

Note that for each $(\mathcal{U}_1, \ldots, \mathcal{U}_B) \in \mathcal{U}_{\boldsymbol{h}}$, $\mathcal{U}_j$ for $j = 1, \ldots, B$ denotes the identities of the clients that map to the j'th element in $[B]$, where $\mathcal{U}_j$'s are disjoint for all $j \in [B]$. Note also that $|\mathcal{U}_{\boldsymbol{h}}| = \binom{n}{\boldsymbol{h}} = \frac{n!}{h_1! h_2! \ldots h_B!}$. It is easy to verify that for any $\boldsymbol{h} \in \mathcal{A}_B^n$, $F(\mathcal{P})(\boldsymbol{h})$ is equal to

$$F(\mathcal{P})(\boldsymbol{h}) = \sum_{(\mathcal{U}_1, \ldots, \mathcal{U}_B) \in \mathcal{U}_{\boldsymbol{h}}} \prod_{j=1}^{B} \prod_{i \in \mathcal{U}_j} p_{ij} \quad (17)$$

Similarly, we can define $F(\mathcal{P}'), F(\mathcal{P}_{-i}), F(\mathcal{P}'_{-i})$. Note that $F(\mathcal{P})$ and $F(\mathcal{P}')$ are distributions over $\mathcal{A}_B^n$, whereas, $F(\mathcal{P}_{-i})$ and $F(\mathcal{P}'_{-i})$ are distributions over $\mathcal{A}_B^{n-1}$. It is easy to see that $F(\mathcal{P}) = \mathcal{M}(\mathcal{D})$ and $F(\mathcal{P}') = \mathcal{M}(\mathcal{D}')$. Similarly, $F(\mathcal{P}_{-i}) = \mathcal{M}(\mathcal{D}_{-i})$ and $F(\mathcal{P}'_{-i}) = \mathcal{M}(\mathcal{D}'_{-i})$. Now we are ready to prove Theorem 3.6.

Since $\mathcal{R}$ is an ϵ_0-LDP mechanism, we have

$$e^{-\epsilon_0} \leq \frac{p_{ij}}{p'_{nj}} \leq e^{\epsilon_0}, \qquad \forall j \in [B], i \in [n].$$

As mentioned in Section 3.3.1, a crucial observation is that any distribution $\boldsymbol{p}_i$ can be written as the following mixture distribution:

$$\boldsymbol{p}_i = q \boldsymbol{p}'_n + (1 - q) \tilde{\boldsymbol{p}}_i, \quad (18)$$

where $q = \frac{1}{e^{\epsilon_0}}$. The distribution $\tilde{\boldsymbol{p}}_i = [\tilde{p}_{i1}, \ldots, \tilde{p}_{iB}]$ is given by $\tilde{p}_{ij} = \frac{p_{ij} - q p'_{nj}}{1 - q}$, where it is easy to verify that $\tilde{p}_{ij} \geq 0$ and $\sum_{j=1}^{B} \tilde{p}_{ij} = 1$.

This idea of writing the distribution of the output of an LDP mechanism as a mixture distribution is inspired from [7, 24]. However, we create different mixtures and use them in a distinct way to reduce the Renyi divergence calculation to those distributions with a certain neighborhood structure using Lemma 5.3.

Now we show that since each $\boldsymbol{p}_i = q \boldsymbol{p}'_n + (1 - q) \tilde{\boldsymbol{p}}_i$ is a mixture distribution, we can write $F(\mathcal{P})$ and $F(\mathcal{P}')$ as certain convex combinations. Before stating the result, we need some notation.

For any $C \subseteq [n-1]$, define two sets $\mathcal{P}_C, \mathcal{P}'_C$, having n distributions each, as follows:

$$\mathcal{P}_C = \{\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_{n-1}\} \bigcup \{\boldsymbol{p}_n\}, \quad (19)$$

$$\mathcal{P}'_C = \{\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_{n-1}\} \bigcup \{\boldsymbol{p}'_n\}, \quad (20)$$

where, for every $i \in [n-1]$, $\hat{\boldsymbol{p}}_i$ is defined as follows:

$$\hat{\boldsymbol{p}}_i = \begin{cases} \boldsymbol{p}'_n & \text{if } i \in C, \\ \tilde{\boldsymbol{p}}_i & \text{if } i \in [n-1] \setminus C. \end{cases} \quad (21)$$

Note that $\mathcal{P}_C$ and $\mathcal{P}'_C$ differ only in one distribution, where $\mathcal{P}_C$ contains $\boldsymbol{p}_n$ whereas $\mathcal{P}'_C$ contains $\boldsymbol{p}'_n$. In words, if clients map their data points according to the distributions in either $\mathcal{P}_C$ or $\mathcal{P}'_C$ for any $C \subseteq [n-1]$, then for all clients $i \in C$, the i'th client maps its data point according to $\boldsymbol{p}'_n$ (which is the distribution of $\mathcal{R}$ on input d'_n), and for all clients $i \in [n-1] \setminus C$, the i'th client maps its data point according to $\tilde{\boldsymbol{p}}_i$. The last client maps its data point according to $\boldsymbol{p}_n$ or $\boldsymbol{p}'_n$ depending on whether the set is $\mathcal{P}_C$ or $\mathcal{P}'_C$.

In the following lemma, we show that $F(\mathcal{P})$ and $F(\mathcal{P}')$ can be written as convex combinations of $\{F(\mathcal{P}_C) : C \subseteq [n-1]\}$ and $\{F(\mathcal{P}'_C) : C \subseteq [n-1]\}$, respectively, where for any $C \subseteq [n-1]$, both $F(\mathcal{P}_C)$ and $F(\mathcal{P}'_C)$ can be computed analogously as in (17).

LEMMA 5.1 (MIXTURE INTERPRETATION). $F(\mathcal{P})$ and $F(\mathcal{P}')$ can be written as the following convex combinations:

$$F(\mathcal{P}) = \sum_{C \subseteq [n-1]} q^{|C|} (1-q)^{n-|C|-1} F(\mathcal{P}_C), \quad (22)$$

$$F(\mathcal{P}') = \sum_{C \subseteq [n-1]} q^{|C|} (1-q)^{n-|C|-1} F(\mathcal{P}'_C), \quad (23)$$

where $\mathcal{P}_C, \mathcal{P}'_C$ are defined in (19)-(21).

We prove Lemma 5.1 in Appendix B.1.

Now, using Lemma 5.1, in the following lemma we show that the Rényi divergence between $F(\mathcal{P})$ and $F(\mathcal{P}')$ can be upper-bounded by a convex combination of the Rényi divergence between $F(\mathcal{P}_C)$ and $F(\mathcal{P}'_C)$ for $C \subseteq [n-1]$.

LEMMA 5.2 (JOINT CONVEXITY). *For any $\lambda > 1$, the function* $\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(\boldsymbol{h})}{F(\mathcal{P}')(\boldsymbol{h})}\right)^\lambda\right]$ *is jointly convex in $(F(\mathcal{P}), F(\mathcal{P}'))$, i.e.,*

$$\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(\boldsymbol{h})}{F(\mathcal{P}')(\boldsymbol{h})}\right)^\lambda\right]$$

$$\leq \sum_{C \subseteq [n-1]} q^{|C|}(1-q)^{n-|C|-1} \mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_C)}\left[\left(\frac{F(\mathcal{P}_C)(\boldsymbol{h})}{F(\mathcal{P}'_C)(\boldsymbol{h})}\right)^\lambda\right]. \tag{24}$$

We prove Lemma 5.2 in Appendix B.2. For any $C \subseteq [n-1]$, let $\widetilde{\mathcal{P}}_{[n-1]\backslash C} = \{\tilde{p}_i : i \in [n-1] \backslash C\}$. With this notation, note that $\mathcal{P}_C \backslash \widetilde{\mathcal{P}}_{[n-1]\backslash C} = \{p'_n, \ldots, p'_n\} \bigcup \{p_n\}$ and $\mathcal{P}'_C \backslash \widetilde{\mathcal{P}}_{[n-1]\backslash C} = \{p'_n, \ldots, p'_n\} \bigcup \{p'_n\}$ is a pair of specific neighboring distributions, each containing $|C| + 1$ distributions. In other words, if we define $\mathcal{D}^{(n)}_{|C|+1} = (d'_n, \ldots, d'_n, d_n)$ and $\mathcal{D}'^{(n)}_{|C|+1} = (d'_n, \ldots, d'_n, d'_n)$, each having $(|C| + 1)$ data points, then the mechanisms $\mathcal{M}(\mathcal{D}^{(n)}_{|C|+1})$ and $\mathcal{M}(\mathcal{D}'^{(n)}_{|C|+1})$ will have distributions $F(\mathcal{P}_C \backslash \widetilde{\mathcal{P}}_{[n-1]\backslash C})$ and $F(\mathcal{P}'_C \backslash \widetilde{\mathcal{P}}_{[n-1]\backslash C})$, respectively.

Now, since $(\mathcal{D}'^{(n)}_{|C|+1}, \mathcal{D}^{(n)}_{|C|+1}) \in \mathcal{D}^{|C|+1}_{\text{same}}$, if we remove the effect of distributions in $\widetilde{\mathcal{P}}_{[n-1]\backslash C}$ in the RHS of (24), we would be able to bound the RHS of (24) using the RDP for the special neighboring datasets in $\mathcal{D}^{|C|+1}_{\text{same}}$. This is precisely what we will do in the following lemma and the subsequent corollary, where we will eliminate the distributions in $\widetilde{\mathcal{P}}_{[n-1]\backslash C}$ in the RHS (24).

The following lemma holds for arbitrary pairs $(\mathcal{P}, \mathcal{P}')$ of neighboring distributions $\mathcal{P} = \{p_1, \ldots, p_n\}$ and $\mathcal{P}' = \{p_1, \ldots, p_{n-1}, p'_n\}$, where we show that $\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(\boldsymbol{h})}{F(\mathcal{P}')(\boldsymbol{h})}\right)^\lambda\right]$ does not decrease when we eliminate a distribution p_i (i.e., remove the data point d_i from the datasets) for any $i \in [n-1]$. We need this general statement as it will be required in the proof of Theorem 3.1 later.

LEMMA 5.3 (MONOTONICITY). *For any $i \in [n-1]$, we have*

$$\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(\boldsymbol{h})}{F(\mathcal{P}')(\boldsymbol{h})}\right)^\lambda\right] \leq \mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_{-i})}\left[\left(\frac{F(\mathcal{P}_{-i})(\boldsymbol{h})}{F(\mathcal{P}'_{-i})(\boldsymbol{h})}\right)^\lambda\right], \tag{25}$$

where, for $i \in [n-1]$, $\mathcal{P}_{-i} = \mathcal{P} \backslash \{p_i\}$ and $\mathcal{P}'_{-i} = \mathcal{P}' \backslash \{p_i\}$. Note that in the left hand side (LHS) of (25), $F(\mathcal{P}), F(\mathcal{P}')$ are distributions over $\mathcal{A}^n_B$, whereas, in the RHS, $F(\mathcal{P}_{-i}), F(\mathcal{P}'_{-i})$ for any $i \in [n-1]$ are distributions over $\mathcal{A}^{n-1}_B$.

We prove Lemma 5.3 in Appendix B.3. Note that Lemma 5.3 is a general statement that holds for arbitrary pairs $(\mathcal{P}, \mathcal{P}')$ of neighboring distributions. For our purpose, we apply Lemma 5.3 with $(\mathcal{P}_C, \mathcal{P}'_C)$ for any $C \subseteq [n-1]$ and then eliminate the distributions in $\widetilde{\mathcal{P}}_{[n-1]\backslash C}$ one by one. The result is stated in the following corollary.

COROLLARY 5.4. *Consider any $m \in \{0, 1, \ldots, n-1\}$. Let $\mathcal{D}^{(n)}_{m+1} = (d'_n, \ldots, d'_n, d_n)$ and $\mathcal{D}'^{(n)}_{m+1} = (d'_n, \ldots, d'_n)$. Then, for any $C \in \binom{[n-1]}{m}$ (i.e., $C \subseteq [n-1]$ such that $|C| = m$), we have*

$$\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_C)}\left[\left(\frac{F(\mathcal{P}_C)(\boldsymbol{h})}{F(\mathcal{P}'_C)(\boldsymbol{h})}\right)^\lambda\right] \leq \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}'^{(n)}_{m+1})}\left[\left(\frac{\mathcal{M}(\mathcal{D}^{(n)}_{m+1})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}'^{(n)}_{m+1})(\boldsymbol{h})}\right)^\lambda\right]. \tag{26}$$

We prove Corollary 5.4 in Appendix B.4. Substituting from (26) into (24) and noting that for every $\boldsymbol{h} \in \mathcal{A}^n_B$, $F(\mathcal{P})(\boldsymbol{h})$ and $F(\mathcal{P}')(\boldsymbol{h})$ are distributionally equal to $\mathcal{M}(\mathcal{D})(\boldsymbol{h})$ and $\mathcal{M}(\mathcal{D}')(\boldsymbol{h})$, respectively, we get

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')}\left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})}\right)^\lambda\right]$$

$$\overset{(a)}{\leq} \sum_{m=0}^{n-1} \sum_{C \in \binom{[n-1]}{m}} q^m(1-q)^{n-m-1} \mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_C)}\left[\left(\frac{F(\mathcal{P}_C)(\boldsymbol{h})}{F(\mathcal{P}'_C)(\boldsymbol{h})}\right)^\lambda\right]$$

$$\overset{(b)}{\leq} \sum_{m=0}^{n-1} \sum_{C \in \binom{[n-1]}{m}} q^m(1-q)^{n-m-1} \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}'^{(n)}_{m+1})}\left[\left(\frac{\mathcal{M}(\mathcal{D}^{(n)}_{m+1})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}'^{(n)}_{m+1})(\boldsymbol{h})}\right)^\lambda\right]$$

$$\overset{(c)}{=} \sum_{m=0}^{n-1} \binom{n-1}{m} q^m(1-q)^{n-m-1} \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}'^{(n)}_{m+1})}\left[\left(\frac{\mathcal{M}(\mathcal{D}^{(n)}_{m+1})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}'^{(n)}_{m+1})(\boldsymbol{h})}\right)^\lambda\right]$$

$$= \mathbb{E}_{m \sim \text{Bin}(n-1,q)}\left[\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}'^{(n)}_{m+1})}\left[\left(\frac{\mathcal{M}(\mathcal{D}^{(n)}_{m+1})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}'^{(n)}_{m+1})(\boldsymbol{h})}\right)^\lambda\right]\right].$$

The inequality (a) is the same as (24), just writing it differently. In (b) we used (26) and in (c) we used the fact that number of m-sized subsets of $[n-1]$ is equal to $\binom{n-1}{m}$. This completes the proof of Theorem 3.6.

6 PROOF OF RDP FOR THE SPECIAL FORM

Fix an arbitrary $m \in \mathbb{N}$ and consider any pair of neighboring datasets $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}^m_{\text{same}}$. Let $\mathcal{D}_m = (d, \ldots, d) \in \mathcal{X}^m$ and $\mathcal{D}'_m = (d, \ldots, d, d') \in \mathcal{X}^m$. Let $\boldsymbol{p} = (p_1, \ldots, p_B)$ and $\boldsymbol{p}' = (p'_1, \ldots, p'_B)$ be the probability distributions of the discrete ϵ_0-LDP mechanism $\mathcal{R} : \mathcal{X} \to \mathcal{Y} = [B]$ when its inputs are d and d', respectively, where $p_j = \Pr[\mathcal{R}(d) = j]$ and $p'_j = \Pr[\mathcal{R}(d') = j]$ for all $j \in [B]$. Since $\mathcal{R}$ is ϵ_0-LDP, we have

$$e^{-\epsilon_0} \leq \frac{p_j}{p'_j} \leq e^{\epsilon_0}, \qquad \forall j \in [B]. \tag{27}$$

Since $\mathcal{M}$ is a shuffle mechanism, it induces a distribution on $\mathcal{A}^m_B$ for any input dataset. So, for any $\boldsymbol{h} \in \mathcal{A}^m_B$, $\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})$ and $\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})$ are equal to the probabilities of seeing $\boldsymbol{h}$ when the inputs to $\mathcal{M}$ are $\mathcal{D}_m$ and $\mathcal{D}'_m$, respectively. Thus, for a given histogram $\boldsymbol{h} = (h_1, \ldots, h_B) \in \mathcal{A}^m_B$ with m elements and B bins, we have

$$\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h}) = MN(m, \boldsymbol{p}, \boldsymbol{h}) = \binom{m}{\boldsymbol{h}} \prod_{j=1}^B p_j^{h_j}, \tag{28}$$

where $MN(m, \boldsymbol{p}, \boldsymbol{h})$ denotes the Multinomial distribution with $\binom{m}{\boldsymbol{h}} = \frac{m!}{h_1! \cdots h_B!}$. Note that (28) can be obtained as a special case of the general distribution in (17) by putting $\boldsymbol{p}_j = \boldsymbol{p}$ for each client j.

For $\mathcal{M}(\mathcal{D}'_m)$, note that the last client (independent of the other clients) maps its input data point d' to the j'th bin with probability p'_j, and the remaining $(m-1)$ clients' mappings induce a distribution on $\mathcal{A}_B^{m-1}$. Thus, $\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})$ for any $\boldsymbol{h} \in \mathcal{A}_B^m$ can be written as

$$\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h}) = \sum_{j=1}^{B} p'_j MN\left(m - 1, \boldsymbol{p}, \widetilde{\boldsymbol{h}}_j\right), \qquad (29)$$

where $\widetilde{\boldsymbol{h}}_j = (h_1, \ldots, h_{j-1}, h_j - 1, h_{j+1}, \ldots, h_B) \in \mathcal{A}_B^{m-1}$. We implicitly assume that if $h_j = 0$ for some $j \in [B]$, then $MN\left(m - 1, \boldsymbol{p}, \widetilde{\boldsymbol{h}}_j\right) = 0$ as one of the elements is negative. Note that similar to (28), (29) can also be obtained from (17) as a special case. Using the polynomial expansion $(1 + x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i$ (with $x = \frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} - 1$ in the following), we have:

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} \right)^\lambda \right]$$
$$= \sum_{i=0}^{\lambda} \binom{\lambda}{i} \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} - 1 \right)^i \right]. \qquad (30)$$

Let $X : \mathcal{A}_B^m \to \mathbb{R}$ be a random variable associated with the distribution $\mathcal{M}(\mathcal{D}_m)$ on $\mathcal{A}_B^m$, and for any $\boldsymbol{h} \in \mathcal{A}_B^m$, define $X(\boldsymbol{h}) := m \left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} - 1 \right)$. Substituting this in (30) gives:

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} \right)^\lambda \right] = 1 + \sum_{i=1}^{\lambda} \binom{\lambda}{i} \frac{\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[(X(\boldsymbol{h}))^i \right]}{m^i}.$$
$$\qquad (31)$$

The RHS of (31) is in terms of the moments of X, which we bound in the following lemma. Before that, first we simplify the expression for $X(\boldsymbol{h})$ by computing the ratio $\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})}$ for any $\boldsymbol{h} \in \mathcal{A}_B^m$:

$$\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} = \sum_{j=1}^{B} p'_j \frac{MN(m-1, \boldsymbol{p}, \widetilde{\boldsymbol{h}}_j)}{MN(m, \boldsymbol{p}, \boldsymbol{h})} = \sum_{j=1}^{B} \frac{p'_j}{p_j} \frac{h_j}{m}. \qquad (32)$$

Thus, we get $X(\boldsymbol{h}) = m \left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} - 1 \right) = \left(\sum_{j=1}^{B} \frac{p'_j}{p_j} h_j \right) - m$.

Remark 8. As mentioned in Remark 5, we could tighten our upper bounds for specific mechanisms. As shown in (31) above, the Rényi divergence of a mechanism between two neighboring datasets can be written in terms of the moments of a r.v. X, which is defined as the ratio of distributions of the mechanism on these two neighboring datasets. However, since our goal is to bound RDP for all ϵ_0-LDP mechanisms, we prove the worse-case bound on the moments of X that holds for all mechanisms; see (34) in Lemma 6.1 for bound on the $i \geq 3$'rd moments of X and (38) in Lemma 6.2 for bound on the variance of X.

LEMMA 6.1. *The random variable X has the following properties:*

(1) *X has zero mean, i.e., $\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} [X(\boldsymbol{h})] = 0$.*

(2) *The variance of X is equal to*

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[X(\boldsymbol{h})^2 \right] = m \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right). \qquad (33)$$

(3) *For $i \geq 3$, the i'th moment of X is bounded by*

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[(X(\boldsymbol{h}))^i \right] \leq i\Gamma\left(i/2\right) \left(2mv^2\right)^{i/2}, \qquad (34)$$

where $v^2 = \frac{(e^{\epsilon_0} - e^{-\epsilon_0})^2}{4}$ and $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ is the Gamma function.

A proof of Lemma 6.1 is presented in Appendix C.1. Substituting the bounds from Lemma 6.1 into (31), we get

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} \right)^\lambda \right] \leq 1 + \binom{\lambda}{2} \frac{1}{m} \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right)$$
$$+ \sum_{i=3}^{\lambda} \binom{\lambda}{i} i\Gamma\left(i/2\right) \left(\frac{(e^{\epsilon_0} - e^{-\epsilon_0})^2}{2m} \right)^{i/2} \qquad (35)$$

Note that $p_1, \ldots, p_m, p'_1, \ldots, p'_m$ are defined for the fixed pair of datasets $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m$ that we started with. So, the term containing $\left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right)$ in the RHS of (35) depends on $(\mathcal{D}_m, \mathcal{D}'_m)$, and that is the only term in (35) that depends on $(\mathcal{D}_m, \mathcal{D}'_m)$. Since Theorem 3.7 requires us to bound (35) for any pair of neighboring datasets $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m$, so, in order to prove Theorem 3.7, we need to compute $\sup_{(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m} \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right)$. We bound this in the following.

Define a set $\mathcal{T}_{\epsilon_0}$ consisting of all pairs of B-dimensional probability vectors satisfying the ϵ_0-LDP constraints as follows:

$$\mathcal{T}_{\epsilon_0} = \left\{ (\boldsymbol{p}, \boldsymbol{p}') \in \mathbb{R}^B \times \mathbb{R}^B : p_j, p'_j \geq 0, \forall j \in [B], \sum_{j=1}^{B} p_j = \sum_{j=1}^{B} p'_j = 1, \right.$$
$$\left. \text{and } e^{-\epsilon_0} \leq \frac{p'_j}{p_j} \leq e^{\epsilon_0}, \forall j \in [B] \right\}. \qquad (36)$$

Note that $\mathcal{T}_{\epsilon_0}$ contains *all* pairs of the output probability distributions $(\boldsymbol{p}, \boldsymbol{p}')$ of *all* ϵ_0-LDP mechanisms $\mathcal{R}$ on *all* neighboring data points $d, d' \in \mathcal{X}$. Since any $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m$ generates a pair of probability distributions $(\boldsymbol{p}, \boldsymbol{p}') \in \mathcal{T}_{\epsilon_0}$ (because $\mathcal{D}_m = (d, \ldots, d)$ and $\mathcal{D}'_m = (d, \ldots, d, d')$ together contain only two distinct data points d, d'), we have

$$\sup_{(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m} \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right) \leq \sup_{(\boldsymbol{p}, \boldsymbol{p}') \in \mathcal{T}_{\epsilon_0}} \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right). \qquad (37)$$

In the following lemma, we bounds the RHS of (37).

LEMMA 6.2. *We have the following bound:*

$$\sup_{(\boldsymbol{p}, \boldsymbol{p}') \in \mathcal{T}_{\epsilon_0}} \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right) = \frac{(e^{\epsilon_0} - 1)^2}{e^{\epsilon_0}}. \qquad (38)$$

We prove Lemma 6.2 in Appendix C.2. Taking supremum over $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m$ in (35) and then using (37) and (38), we get the bound in Theorem 3.7.

7 PROOFS OF THE UPPER BOUNDS

In this section, we will prove our upper bounds stated in Theorems 3.1 and 3.3 in Sections 7.1 and 7.2, respectively.

7.1 Proof of Theorem 3.1

Consider arbitrary neighboring datasets $\mathcal{D} = (d_1, \ldots, d_n) \in \mathcal{X}^n$ and $\mathcal{D}' = (d_1, \ldots, d_{n-1}, d'_n) \in \mathcal{X}^n$. As mentioned in Section 3.3, for any $m \in \{0, \ldots, n-1\}$, we define new neighboring datasets $\mathcal{D}_{m+1}^{(n)} = (d'_n, \ldots, d'_n, d_n)$ and $\mathcal{D}_{m+1}'^{(n)} = (d'_n, \ldots, d'_n, d'_n)$, each having $(m+1)$ elements. Observe that $\left(\mathcal{D}_{m+1}'^{(n)}, \mathcal{D}_{m+1}^{(n)}\right) \in \mathcal{D}_{\text{same}}^{m+1}$.

Recall from Theorem 3.6, we have

$$
\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')} \left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})} \right)^\lambda \right]
$$
$$
\leq \sum_{m=0}^{n-1} q_m \left[\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})} \left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})} \right)^\lambda \right] \right], \tag{39}
$$

where $q_m := \binom{n-1}{m} q^m (1-q)^{n-m-1}$. For simplicity of notation, for any $m \in \{0, 1, \ldots, n-1\}$, define

$$
E_m := \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})} \left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})} \right)^\lambda \right].
$$

We show in Appendix D.1 that E_m is a non-increasing function of m. Using this and concentration properties of the Binomial r.v., we get (details are in Appendix D.1):

$$
\mathbb{E}\left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})} \right)^\lambda \right] \leq e^{\epsilon_0 \lambda} e^{-\frac{q(n-1)\gamma^2}{2}} + E_{(1-\gamma)q(n-1)}, \tag{40}
$$

where $\gamma > 0$ is arbitrary, and expectation is taken w.r.t. $\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')$. Note that we have already bounded E_m for all m in Theorem 3.7. By setting $\gamma = \frac{1}{2}$ and $\overline{n} = \lfloor (1-\gamma)q(n-1) \rfloor + 1 = \lfloor \frac{n-1}{2e^{\epsilon_0}} \rfloor + 1$, we get from Theorem 3.7, that:

$$
\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')} \left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})} \right)^\lambda \right] \leq E_{\overline{n}-1} + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}} \tag{41}
$$

$$
\leq 1 + \binom{\lambda}{2} \frac{(e^{\epsilon_0} - 1)^2}{\overline{n} e^{\epsilon_0}} + \sum_{i=3}^{\lambda} \binom{\lambda}{i} i \Gamma(i/2) \left(\frac{(e^{2\epsilon_0} - 1)^2}{2\overline{n} e^{2\epsilon_0}} \right)^{i/2} + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}}.
$$

Since the above bound holds for arbitrary pairs of neighboring datasets $\mathcal{D}$ and $\mathcal{D}'$, this completes the proof of Theorem 3.1.

7.2 Proof of Theorem 3.3

The proof of Theorem 3.3 follows the same steps as that of the proof of Theorem 3.1 that we outlined in Section 3.3 and also gave formally in Section 7.1, except for the following change. Instead of using Theorem 3.7 for bounding the RDP for specific neighboring datasets, we will use the following theorem.

THEOREM 7.1. *Let $m \in \mathbb{N}$ be arbitrary. For any $\lambda \geq 2$ (including the non-integral λ) and any $(\mathcal{D}_m, \mathcal{D}'_m) \in \mathcal{D}_{\text{same}}^m$, we have*

$$
\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)} \left[\left(\frac{\mathcal{M}(\mathcal{D}'_m)(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})} \right)^\lambda \right] \leq \exp\left(\lambda^2 \frac{(e^{\epsilon_0} - 1)^2}{m} \right). \tag{42}
$$

We prove Theorem 7.1 in Appendix D.2. Note that Theorem 7.1 implies that $E_{m-1} \leq \exp\left(\lambda^2 \frac{(e^{\epsilon_0}-1)^2}{m} \right)$ holds for every integer $m \geq 2$. Substituting this in (41) (by putting $m = \overline{n} = \lfloor \frac{n-1}{2e^{\epsilon_0}} \rfloor + 1$), we get

$$
\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')} \left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})} \right)^\lambda \right] \leq e^{\lambda^2 \frac{(e^{\epsilon_0}-1)^2}{\overline{n}}} + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}}.
$$

This proves Theorem 3.3.

8 PROOF SKETCH OF THE LOWER BOUND

Consider the binary case, where each data point d can take a value from $\mathcal{X} = \{0, 1\}$. Let the local randomizer $\mathcal{R}$ be the binary randomized response (2RR) mechanism, where $\Pr[\mathcal{R}(d) = d] = \frac{e^{\epsilon_0}}{e^{\epsilon_0}+1}$ for $d \in \mathcal{X}$. It is easy to verify that $\mathcal{R}$ is an ϵ_0-LDP mechanism. For simplicity, let $p = \frac{1}{e^{\epsilon_0}+1}$. Consider two neighboring datasets $\mathcal{D}, \mathcal{D}' \in \{0, 1\}^n$, where $\mathcal{D} = (0, \ldots, 0, 0)$ and $\mathcal{D}' = (0, \ldots, 0, 1)$. Let $k \in \{0, \ldots, n\}$ denote the number of ones in the output of the shuffler. As argued in Section 2.3 on page 4, since the output of the shuffle mechanism $\mathcal{M}$ can be thought of as the distribution of the number of ones in the output, we have that $k \sim \mathcal{M}(\mathcal{D})$ is distributed as a Binomial random variable $\text{Bin}(n, p)$. The proof uses some properties of the Binomial r.v., which are provided in Appendix E.

9 CONCLUSION

The analysis of the RDP for the shuffle model presented in this paper was based on some new analysis techniques that may be of independent interest. The utility of these bounds were also demonstrated numerically, where we saw that in important regimes of interest, we get 8× improvement over the state-of-the-art without sampling and at least 10× improvement with sampling (see Section 4 for more details).

A simple extension of the results would be to work with local approximate DP guarantees instead of pure LDP. This can be seen by using the tight conversion between approximate DP and pure DP given in [24]. However, there are several open problems of interest. Our upper bounds hold for general discrete local mechanisms. The extension to continuous distributions requires careful technical analysis as the histogram used for RDP analysis would need to approximate continuous distributions via discretization. We leave the analysis of continuous distributions as a future work. Perhaps the most important one is mentioned in Remark 7. There is a multiplicative gap of the order e^{ϵ_0} in our upper and lower bounds, and closing this gap is an important open problem. We believe that our lower bound is tight (at least for the first order term) and the upper bound is loose. Showing this or getting a tighter upper bound may require new proof techniques. A second question could be how to get an overall RDP guarantee if we are given local RDP guarantees instead of local LDP guarantees.

REFERENCES

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 308–318.

[2] Shahab Asoodeh, Jiachun Liao, Flavio P Calmon, Oliver Kosut, and Lalitha Sankar. 2021. Three variants of differential privacy: Lossless conversion and applications. *IEEE Journal on Selected Areas in Information Theory* 2, 1 (2021), 208–222.

[3] Victor Balcer and Albert Cheu. 2020. Separating Local & Shuffled Differential Privacy via Histograms. In *1st Conference on Information-Theoretic Cryptography, ITC 2020, June 17-19, 2020, Boston, MA, USA (LIPIcs, Vol. 163)*, Yael Tauman Kalai, Adam D. Smith, and Daniel Wichs (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 1:1–1:14. https://doi.org/10.4230/LIPIcs.ITC.2020.1

[4] Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. 2020. Hypothesis Testing Interpretations and Renyi Differential Privacy. In *International Conference on Artificial Intelligence and Statistics (AISTATS) (Proceedings of Machine Learning Research, Vol. 108)*, Silvia Chiappa and Roberto Calandra (Eds.). PMLR, 2496–2506.

[5] Borja Balle, James Bell, Adria Gascon, and Kobbi Nissim. 2019. Differentially private summation with multi-message shuffling. *arXiv preprint arXiv:1906.09116* (2019).

[6] Borja Balle, James Bell, Adria Gascón, and Kobbi Nissim. 2019. Improved summation from shuffling. *arXiv preprint arXiv:1909.11225* (2019).

[7] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. 2019. The privacy blanket of the shuffle model. In *Annual International Cryptology Conference*. Springer, 638–667.

[8] Borja Balle, James Bell, Adria Gascón, and Kobbi Nissim. 2020. Private summation in the multi-message shuffle model. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. 657–676.

[9] Raef Bassily, Adam Smith, and Abhradeep Thakurta. 2014. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 464–473.

[10] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnés, and Bernhard Seefeld. 2017. Prochlo: Strong Privacy for Analytics in the Crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles (SOSP)*. ACM, 441–459.

[11] Mark Bun and Thomas Steinke. 2016. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference (TCC)*. Springer, 635–658.

[12] Clément L. Canonne, Gautam Kamath, and Thomas Steinke. 2020. The Discrete Gaussian for Differential Privacy. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). https://proceedings.neurips.cc/paper/2020/hash/b53b3a3d6ab90ce0268229151c9bde11-Abstract.html

[13] Albert Cheu, Adam D. Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. 2019. Distributed Differential Privacy via Shuffling. In *Advances in Cryptology - EUROCRYPT 2019*, Vol. 11476. Springer, 375–403.

[14] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. 2017. Collecting Telemetry Data Privately. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 3574–3583.

[15] John C Duchi, Michael I Jordan, and Martin J Wainwright. 2013. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 429–438.

[16] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography Conference (TCC)*. 265–284.

[17] Cynthia Dwork and Aaron Roth. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9, 3-4 (2014), 211–407.

[18] Cynthia Dwork and Guy N. Rothblum. 2016. Concentrated Differential Privacy. *CoRR* abs/1603.01887 (2016). arXiv:1603.01887 http://arxiv.org/abs/1603.01887

[19] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. 2010. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 51–60.

[20] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Shuang Song, Kunal Talwar, and Abhradeep Thakurta. 2020. Encode, Shuffle, Analyze Privacy Revisited: Formalizations and Empirical Evaluation. *CoRR* abs/2001.03618 (2020). https://arxiv.org/abs/2001.03618

[21] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2018. Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity. *arXiv preprint arXiv:1811.12469* (2018).

[22] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2019. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2468–2479.

[23] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 1054–1067.

[24] Vitaly Feldman, Audra McMillan, and Kunal Talwar. 2020. Hiding Among the Clones: A Simple and Nearly Optimal Analysis of Privacy Amplification by Shuffling. *arXiv preprint arXiv:2012.12803* (2020). Open source implementation of privacy https://github.com/apple/ml-shuffling-amplification.

[25] Badih Ghazi, Noah Golowich, Ravi Kumar, Rasmus Pagh, and Ameya Velingker. 2019. On the Power of Multiple Anonymous Messages. *IACR Cryptol. ePrint Arch.* 2019 (2019), 1382.

[26] Badih Ghazi, Rasmus Pagh, and Ameya Velingker. 2019. Scalable and differentially private distributed aggregation in the shuffled model. *arXiv preprint arXiv:1906.08320* (2019).

[27] Antonious Girgis, Deepesh Data, Suhas N. Diggavi, Peter Kairouz, and Ananda Theertha Suresh. 2021. Shuffled Model of Differential Privacy in Federated Learning. In *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS (Proceedings of Machine Learning Research, Vol. 130)*. PMLR, 2521–2529.

[28] Antonious M. Girgis, Deepesh Data, Suhas Diggavi, Peter Kairouz, and Ananda Theertha Suresh. 2021. Shuffled Model of Federated Learning: Privacy, Accuracy and Communication Trade-Offs. *IEEE Journal on Selected Areas in Information Theory* 2, 1 (2021), 464–478.

[29] Andy Greenberg. 2016. Apple's 'differential privacy'is about collecting your data—but not your data. *Wired, June* 13 (2016).

[30] Peter Kairouz et al. 2019. Advances and Open Problems in Federated Learning. *CoRR* abs/1912.04977 (2019).

[31] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. 2016. Discrete Distribution Estimation under Local Privacy. In *International Conference on Machine Learning, ICML*. 2436–2444.

[32] Peter Kairouz, Ziyu Liu, and Thomas Steinke. 2021. The Distributed Discrete Gaussian Mechanism for Federated Learning with Secure Aggregation. *CoRR* abs/2102.06387 (2021). arXiv:2102.06387 https://arxiv.org/abs/2102.06387

[33] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2015. The composition theorem for differential privacy. In *International conference on machine learning*. PMLR, 1376–1385.

[34] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What can we learn privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.

[35] Ninghui Li, Wahbeh Qardaji, and Dong Su. 2012. 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*. 32–33.

[36] Ilya Mironov. 2017. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, 263–275.

[37] Ilya Mironov, Kunal Talwar, and Li Zhang. 2019. R\'enyi Differential Privacy of the Sampled Gaussian Mechanism. *arXiv preprint arXiv:1908.10530* (2019).

[38] Phillippe Rigollet and Jan-Christian Hütter. 2015. High dimensional statistics. *Lecture notes for course 18S997* 813 (2015), 814.

[39] Tim van Erven and Peter Harremoës. 2014. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Trans. Inf. Theory* 60, 7 (2014), 3797–3820. https://doi.org/10.1109/TIT.2014.2320500

[40] Yu-Xiang Wang, Borja Balle, and Shiva Prasad Kasiviswanathan. 2019. Subsampled Rényi differential privacy and analytical moments accountant. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 1226–1235.

[41] Stanley L Warner. 1965. Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* 60, 309 (1965), 63–69.

[42] Wikipedia. [n.d.]. Gamma function. https://en.wikipedia.org/wiki/Gamma_function.

[43] Yuqing Zhu and Yu-Xiang Wang. 2019. Poission subsampled rényi differential privacy. In *International Conference on Machine Learning*. PMLR, 7634–7642.

A PROOF OF COROLLARY 1

In this section, we prove the simplified bound (stated in (6)) on the RDP of the shuffle model, provided that λ, ϵ_0, n satisfy a certain condition. In particular, we will show that if λ, ϵ_0, n satisfy $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$, then

$$\epsilon(\lambda) \leq \frac{1}{\lambda - 1} \log \left(1 + \frac{\lambda^2 \left(e^{\epsilon_0} - 1 \right)^2}{\overline{n} e^{\epsilon_0}} \right), \tag{43}$$

where $\overline{n} = \frac{n-1}{2e^{\epsilon_0}} + 1$. In order to show (43), it suffices to prove the following (using which in (5) will yield (6)):

$$\sum_{i=3}^{\lambda} \binom{\lambda}{i} i \Gamma \left(i/2 \right) \left(\frac{\left(e^{2\epsilon_0} - 1 \right)^2}{2e^{2\epsilon_0} \overline{n}} \right)^{i/2} + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}} \leq \binom{\lambda}{2} \frac{\left(e^{\epsilon_0} - 1 \right)^2}{\overline{n} e^{\epsilon_0}}. \tag{44}$$

First notice that $\binom{\lambda}{i} i \Gamma \left(i/2 \right) \leq \lambda^i$ (see Claim 1 on page 15). In order to show (44), it suffices to show

$$\sum_{i=3}^{\lambda} \left(\frac{\lambda \left(e^{2\epsilon_0} - 1 \right)}{(2e^{\epsilon_0} \overline{n})^{1/2}} \right)^i + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}} \leq \binom{\lambda}{2} \frac{\left(e^{\epsilon_0} - 1 \right)^2}{\overline{n} e^{\epsilon_0}}. \tag{45}$$

Note that there are $(\lambda - 2)$ terms inside the summation. If we show that each of those terms is smaller than 1 (which would imply that the term corresponding to $i = 3$ is the largest one), then the summation is at most $(\lambda - 2)$ times the term with $i = 3$. Further, if the additional exponential term in the LHS is upper-bounded by the term with $i = 3$, then we can prove (45) by showing that $(\lambda - 1)$ times the term with $i = 3$ is upper-bounded by the RHS. These arguments are summarized in the following set of three inequalities:

$$\frac{\lambda \left(e^{2\epsilon_0} - 1 \right)}{(2e^{2\epsilon_0} \overline{n})^{1/2}} < 1 \tag{46}$$

$$e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}} \leq \left(\frac{\lambda \left(e^{2\epsilon_0} - 1 \right)}{(2e^{2\epsilon_0} \overline{n})^{1/2}} \right)^3 \tag{47}$$

$$(\lambda - 1) \left(\frac{\lambda \left(e^{2\epsilon_0} - 1 \right)}{(2e^{2\epsilon_0} \overline{n})^{1/2}} \right)^3 \leq \binom{\lambda}{2} \frac{\left(e^{\epsilon_0} - 1 \right)^2}{\overline{n} e^{\epsilon_0}} \tag{48}$$

In the rest of this proof, we will derive the condition on ϵ_0, λ, n such that (48) is satisfied. As we see later, the values of ϵ_0, λ thus obtained will automatically satisfy (46) and (47).

By canceling same terms from both sides of (48), we get

$$\frac{\lambda^2 \left(e^{2\epsilon_0} - 1 \right)^3}{(2e^{\epsilon_0} \overline{n})^{3/2} e^{3\epsilon_0/2}} \leq \frac{\left(e^{\epsilon_0} - 1 \right)^2}{2\overline{n} e^{\epsilon_0}}$$

$$\Longleftrightarrow \lambda^2 (e^{2\epsilon_0} - 1)(e^{\epsilon_0} + 1)^2 \leq \sqrt{2\overline{n} e^{\epsilon_0}} e^{3\epsilon_0/2} \tag{49}$$

For the LHS and the RHS, we respectively have

$$(e^{2\epsilon_0} - 1)(e^{\epsilon_0} + 1)^2 = (e^{2\epsilon_0} - 1)(e^{2\epsilon_0} + 2e^{\epsilon_0} + 1)$$

$$\leq e^{4\epsilon_0} + 2e^{3\epsilon_0} \leq 3e^{4\epsilon_0} \tag{50}$$

$$2\overline{n} e^{\epsilon_0} = n - 1 + 2e^{\epsilon_0} \geq n. \tag{51}$$

Therefore, in order to show (49), it suffices to show $3\lambda^2 e^{4\epsilon_0} \leq \sqrt{n e^{3\epsilon_0}}$, which is equivalent to $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$. Thus, we have shown that $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$ implies (48).

Now we show that when $\lambda^4 e^{5\epsilon_0} < \frac{n}{9}$, (46) and (47) are automatically satisfied:

(1) Proof of (46):

$$\frac{\lambda \left(e^{2\epsilon_0} - 1 \right)}{\sqrt{2e^{2\epsilon_0} \overline{n}}} \leq \frac{\lambda e^{2\epsilon_0}}{\sqrt{2e^{\epsilon_0} \overline{n}}} \leq \sqrt{\frac{\lambda^4 e^{5\epsilon_0}}{2e^{\epsilon_0} \overline{n}}} \leq \sqrt{\frac{n/9}{n}} < 1.$$

In the second inequality we used $\lambda \geq 1$ and in the penultimate inequality we used $2e^{\epsilon_0} \overline{n} \geq n$ from (51).

(2) Proof of (47): For this, first we upper-bound the LHS and lower-bound the RHS, and then note that the upper-bound is smaller than the lower-bound. For the upper-bound on $\exp(\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}})$, note that $\epsilon_0 \lambda \leq e^{5\epsilon_0/4} \lambda = \left(e^{5\epsilon_0} \lambda^4 \right)^{1/4} < \left(\frac{n}{9} \right)^{1/4} = \frac{n^{1/4}}{\sqrt{3}}$. Also note that $e^{\epsilon_0} \leq e^{5\epsilon_0/4} \lambda < \frac{n^{1/4}}{\sqrt{3}}$, which implies $\frac{n-1}{8e^{\epsilon_0}} = \frac{\sqrt{3}}{8} \frac{n-1}{n^{1/4}} \geq \frac{\sqrt{3}}{16} n^{3/4}$. Substituting these bounds in the exponent of $\exp(\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}})$, we get:

$$\exp \left(\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}} \right) \leq \exp \left(\frac{n^{1/4}}{\sqrt{3}} - \frac{\sqrt{3}}{16} n^{3/4} \right)$$

$$= \exp \left(-n^{3/4} \left(\frac{\sqrt{3}}{16} - \frac{1}{\sqrt{3n}} \right) \right) \tag{52}$$

$$\leq \exp \left(-c' n^{3/4} \right),$$

where $c' > 0$ is a constant even for small values of n. For example, for $n = 100$, we get $c' \geq \frac{1}{20}$.

For the lower-bound on $\left(\frac{\lambda (e^{2\epsilon_0} - 1)}{(2e^{2\epsilon_0} \overline{n})^{1/2}} \right)^3$, note that $2e^{\epsilon_0} \overline{n} = n - 1 + 2e^{\epsilon_0} \leq n - 1 + 2 \left(\frac{n}{9} \right)^{1/5} \leq 2n$, where $e^{\epsilon_0} \leq \left(\frac{n}{9} \right)^{1/5}$ follows from $e^{5\epsilon_0} \leq \lambda^4 e^{5\epsilon_0} < \frac{n}{9}$. Now we show the lower bound:

$$\frac{\lambda^3 (e^{2\epsilon_0} - 1)^3}{(2e^{2\epsilon_0} \overline{n})^{3/2}} \geq \frac{(e^{\epsilon_0} - 1)^3 (e^{\epsilon_0} + 1)^3}{(2e^{\epsilon_0} \overline{n})^{3/2} e^{3\epsilon_0/2}}$$

$$\geq \frac{(e^{\epsilon_0} - 1)^3 e^{3\epsilon_0}}{(2n)^{3/2} e^{3\epsilon_0/2}} \tag{53}$$

$$\geq \frac{(e^{\epsilon_0} - 1)^3}{(2n)^{3/2}} \geq \frac{\epsilon_0^3}{(2n)^{3/2}}$$

Note that the upper-bound on $\exp(\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}})$ is exponentially small in $n^{3/4}$, whereas, the lower-bound on $\frac{\lambda^3 (e^{2\epsilon_0} - 1)^3}{(2e^{2\epsilon_0} \overline{n})^{3/2}}$ is inverse-polynomial in n. So, for sufficiently large n, (47) will be satisfied.

This completes the proof of Corollary 3.2

Claim 1 (An Inequality for the Gamma Function). *For any $\lambda \in \mathbb{N}$ and $k \geq 3$, we have $\binom{\lambda}{k} k \Gamma(k/2) \leq \lambda^k$.*

PROOF. Note that for any $\lambda \in \mathbb{N}$ and $k \leq \lambda$, we have $\binom{\lambda}{k} = \frac{\lambda(\lambda-1)(\lambda-2)...(\lambda-k+1)}{k!}$.

We show the claim separately for the cases when k is an even integer or not.

(1) *When k is an even integer:* Since for any integer $n \in \mathbb{N}$, $\Gamma(n) = (n - 1)!$, so when k is an even integer, we have

$$\binom{\lambda}{k} k \Gamma(k/2) = \frac{\lambda(\lambda - 1)(\lambda - 2) \dots (\lambda - k + 1)}{k!} \times k \times \left(\frac{k}{2} - 1 \right)!$$

$$\leq \lambda(\lambda - 1)(\lambda - 2) \dots (\lambda - k + 1)$$
$$\leq \lambda^k.$$

(2) *When k is an odd integer:* Note that for any integer $n \in \mathbb{N}$, we have $\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$; see [42]. Let $k = 2a + 1$. Then

$$\binom{\lambda}{k} k \Gamma(k/2) = \binom{\lambda}{k} k \Gamma(a + \frac{1}{2})$$

$$= \frac{\lambda(\lambda - 1)(\lambda - 2) \dots (\lambda - k + 1)}{k!} \times k \times \frac{(2a)!}{4^a a!}\sqrt{\pi}$$

$$= \lambda(\lambda - 1)(\lambda - 2) \dots (\lambda - k + 1)\frac{\sqrt{\pi}}{4^a a!}$$

$$\overset{(a)}{\leq} \lambda(\lambda - 1)(\lambda - 2) \dots (\lambda - k + 1)$$

$$\leq \lambda^k$$

where (a) follows because $\frac{\sqrt{\pi}}{4^a a!} \leq 1$ when $a \geq 1 \iff k \geq 3$. This proves Claim 1. ∎

B OMITTED DETAILS FROM SECTION 5

B.1 Proof of Lemma 5.1

LEMMA (RESTATING LEMMA 5.1). *$F(\mathcal{P})$ and $F(\mathcal{P}')$ can be written as the following convex combinations:*

$$F(\mathcal{P}) = \sum_{C \subseteq [n-1]} q^{|C|}(1 - q)^{n-|C|-1} F(\mathcal{P}_C), \qquad (54)$$

$$F(\mathcal{P}') = \sum_{C \subseteq [n-1]} q^{|C|}(1 - q)^{n-|C|-1} F(\mathcal{P}'_C), \qquad (55)$$

where $\mathcal{P}_C, \mathcal{P}'_C$ are defined in (19)-(21).

PROOF. We only show (54); (55) can be shown similarly.
For convenience, for any $C \subseteq [n-1]$, define

$$\mathcal{P}'_{|C|,n} = \{\boldsymbol{p}'_n, \dots, \boldsymbol{p}'_n\} \text{ with } |\mathcal{P}'_{|C|,n}| = |C|,$$

$$\widetilde{\mathcal{P}}_{[n-1]\backslash C} = \{\tilde{\boldsymbol{p}}_i : i \in [n-1] \backslash C\}.$$

With these notations, we can write $\mathcal{P}_C = \mathcal{P}'_{|C|,n} \bigcup \widetilde{\mathcal{P}}_{[n-1]\backslash C} \bigcup \{\boldsymbol{p}_n\}$ and $\mathcal{P}'_C = \mathcal{P}'_{|C|,n} \bigcup \widetilde{\mathcal{P}}_{[n-1]\backslash C} \bigcup \{\boldsymbol{p}'_n\}$.

Note that $\boldsymbol{p}_i = q\boldsymbol{p}'_n + (1 - q)\tilde{\boldsymbol{p}}_i$ for all $i \in [n-1]$. For any $i \in [n-1]$, define the following random variable $\widehat{\boldsymbol{p}}_i$:

$$\widehat{\boldsymbol{p}}_i = \begin{cases} \boldsymbol{p}'_n & \text{w.p. } q, \\ \tilde{\boldsymbol{p}}_i & \text{w.p. } 1 - q. \end{cases}$$

Note that $\mathbb{E}[\widehat{\boldsymbol{p}}_i] = \boldsymbol{p}_i$.
For any subset $C \subseteq [n-1]$, define an event $\mathcal{E}_C := \{\widehat{\boldsymbol{p}}_i = \boldsymbol{p}'_n \text{ for } i \in C \text{ and } \widehat{\boldsymbol{p}}_i = \tilde{\boldsymbol{p}}_i \text{ for } i \in [n-1] \backslash C\}$. Since $\widehat{\boldsymbol{p}}_1, \dots, \widehat{\boldsymbol{p}}_{n-1}$ are independent random variables, we have $\Pr[\mathcal{E}_C] = q^{|C|}(1 - q)^{n-|C|-1}$.

Consider an arbitrary $\boldsymbol{h} \in \mathcal{A}_B^n$. Define a random variable $U(\mathcal{P})$ over $\mathcal{A}_B^n$ whose distribution is equal to $F(\mathcal{P})$.

$$F(\mathcal{P})(\boldsymbol{h}) = \Pr[U(\mathcal{P}) = \boldsymbol{h}]$$

$$= \Pr[U(\boldsymbol{p}_1, \dots, \boldsymbol{p}_{n-1}, \boldsymbol{p}_n) = \boldsymbol{h}]$$

$$= \Pr\left[U\left(\mathbb{E}[\widehat{\boldsymbol{p}}_1], \dots, \mathbb{E}[\widehat{\boldsymbol{p}}_{n-1}], \boldsymbol{p}_n\right) = \boldsymbol{h}\right]$$

$$= \sum_{C \subseteq [n-1]} \Pr[\mathcal{E}_C] \Pr\left[U\left(\mathbb{E}[\widehat{\boldsymbol{p}}_1], \dots, \mathbb{E}[\widehat{\boldsymbol{p}}_{n-1}], \boldsymbol{p}_n\right) = \boldsymbol{h} \mid \mathcal{E}_C\right]$$

$$\overset{(e)}{=} \sum_{C \subseteq [n-1]} \Pr[\mathcal{E}_C] \Pr\left[U\left(\mathcal{P}'_{|C|,n} \bigcup \widetilde{\mathcal{P}}_{[n-1]\backslash C} \bigcup \{\boldsymbol{p}_n\}\right) = \boldsymbol{h}\right]$$

$$= \sum_{C \subseteq [n-1]} \Pr[\mathcal{E}_C] \Pr\left[U(\mathcal{P}_C) = \boldsymbol{h}\right]$$

$$= \sum_{C \subseteq [n-1]} q^{|C|}(1 - q)^{n-|C|-1} \Pr\left[U(\mathcal{P}_C) = \boldsymbol{h}\right],$$

$$= \sum_{C \subseteq [n-1]} q^{|C|}(1 - q)^{n-|C|-1} F(\mathcal{P}_C)(\boldsymbol{h}) \qquad (56)$$

where, $\mathcal{P}'_{|C|,n}$ and $\widetilde{\mathcal{P}}_{[n-1]\backslash C}$ in the RHS of (e) are defined in the statement of the claim.

Since the above calculation holds for every $\boldsymbol{h} \in \mathcal{A}_B^n$, we have proved (54). ∎

B.2 Proof of Lemma 5.2

LEMMA (RESTATING LEMMA 5.2). *For any $\lambda > 1$, the function $\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(\boldsymbol{h})}{F(\mathcal{P}')(\boldsymbol{h})}\right)^\lambda\right]$ is jointly convex in $(F(\mathcal{P}), F(\mathcal{P}'))$, i.e.,*

$$\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(\boldsymbol{h})}{F(\mathcal{P}')(\boldsymbol{h})}\right)^\lambda\right]$$

$$\leq \sum_{C \subseteq [n-1]} q^{|C|}(1 - q)^{n-|C|-1} \mathbb{E}_{\boldsymbol{h} \sim F\left(\mathcal{P}'_C\right)}\left[\left(\frac{F(\mathcal{P}_C)(\boldsymbol{h})}{F\left(\mathcal{P}'_C\right)(\boldsymbol{h})}\right)^\lambda\right].$$
$$(57)$$

PROOF. For simplicity of notation, let $P = F(\mathcal{P})$ and $Q = F(\mathcal{P}')$. Note that $\mathbb{E}_Q\left[\left(\frac{P}{Q}\right)^\lambda\right] = \int P^\lambda Q^{1-\lambda} d\mu$, which is also called the Hellinger integral. In order to prove the lemma, it suffices to show that $\int P^\lambda Q^{1-\lambda} d\mu$ is jointly convex in (P, Q), i.e., if $P_\alpha = \alpha P_0 + (1 - \alpha)P_1$ and $Q_\alpha = \alpha Q_0 + (1 - \alpha)Q_1$ for some $\alpha \in [0, 1]$, then the following holds

$$\int P_\alpha^\lambda Q_\alpha^{1-\lambda} d\mu \leq \alpha \int P_0^\lambda Q_0^{1-\lambda} d\mu + (1 - \alpha) \int P_1^\lambda Q_1^{1-\lambda} d\mu. \quad (58)$$

Proof of (58) is implicit in the proof of [39, Theorem 13]. However, for completeness, we prove (58) in Lemma B.1 below.

Since $P = F(\mathcal{P})$ and $Q = F(\mathcal{P}')$ are convex combinations of $P_C = F(\mathcal{P}_C)$ and $Q_C = F(\mathcal{P}'_C)$, respectively, with same coefficients, repeated application of (58) implies (57). ∎

LEMMA B.1. *For $\lambda \geq 1$, the Hellinger integral $\int P^\lambda Q^{1-\lambda} d\mu$ is jointly convex in (P, Q), i.e., if $P_\alpha = \alpha P_0 + (1 - \alpha)P_1$ and $Q_\alpha = \alpha Q_0 + (1 - \alpha)Q_1$ for some $\alpha \in [0, 1]$, then we have*

$$\int P_\alpha^\lambda Q_\alpha^{1-\lambda} d\mu \leq \alpha \int P_0^\lambda Q_0^{1-\lambda} d\mu + (1 - \alpha) \int P_1^\lambda Q_1^{1-\lambda} d\mu. \quad (59)$$

PROOF. Let $f(x) = x^\lambda$. It is easy to show that for any $\lambda \geq 1$, $f(x)$ is a convex function when $x > 0$. This implies that for any point $\omega \in \Omega$ in the sample space, we have

$$f\left(\frac{P_\alpha(\omega)}{Q_\alpha(\omega)}\right) = f\left(\frac{\alpha P_0(\omega)}{Q_\alpha(\omega)} + \frac{(1 - \alpha)P_1(\omega)}{Q_\alpha(\omega)}\right)$$

$$= f\left(\frac{\alpha Q_0(\omega)}{Q_\alpha(\omega)}\frac{P_0(\omega)}{Q_0(\omega)} + \frac{(1-\alpha)Q_1(\omega)}{Q_\alpha(\omega)}\frac{P_1(\omega)}{Q_1(\omega)}\right)$$
$$\leq \frac{\alpha Q_0(\omega)}{Q_\alpha(\omega)}f\left(\frac{P_0(\omega)}{Q_0(\omega)}\right) + \frac{(1-\alpha)Q_1(\omega)}{Q_\alpha(\omega)}f\left(\frac{P_1(\omega)}{Q_1(\omega)}\right),$$

where the last inequality follows from the convexity of $f(x)$. By multiplying both sides with $Q_\alpha(\omega)$ and substituting the definition of $f(x) = x^\lambda$, we get

$$P_\alpha^\lambda(\omega)Q_\alpha^{1-\lambda}(\omega) \leq \alpha P_0^\lambda(\omega)Q_0^{1-\lambda}(\omega) + (1-\alpha)P_1^\lambda(\omega)Q_1^{1-\lambda}(\omega).$$

By integrating this equality, we get (59). ∎

B.3 Proof of Lemma 5.3

LEMMA (RESTATING LEMMA 5.3). *For any $i \in [n-1]$, we have*

$$\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right] \leq \mathbb{E}_{h\sim F(\mathcal{P}'_{-i})}\left[\left(\frac{F(\mathcal{P}_{-i})(h)}{F\left(\mathcal{P}'_{-i}\right)(h)}\right)^\lambda\right],$$

where, for $i \in [n-1]$, $\mathcal{P}_{-i} = \mathcal{P} \setminus \{p_i\}$ and $\mathcal{P}'_{-i} = \mathcal{P}' \setminus \{p_i\}$. Note that in the LHS, $F(\mathcal{P}), F(\mathcal{P}')$ are distributions over $\mathcal{A}_B^n$, whereas, in the RHS, $F(\mathcal{P}_{-i}), F(\mathcal{P}'_{-i})$ for any $i \in [n-1]$ are distributions over $\mathcal{A}_B^{n-1}$.

PROOF. First we show that $\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right]$ is convex in p_i for any $i \in [n-1]$.

Note that due to the independence of $\mathcal{R}$ on different data points, for any $h = (h_1, \ldots, h_B) \in \mathcal{A}_B^n$, we can recursively write the distributions $F(\mathcal{P})(h)$ and $F(\mathcal{P}')(h)$ (which are defined in (17)) as follows:

$$F(\mathcal{P})(h) = \sum_{j=1}^B p_{ij}F(\mathcal{P}_{-i})(\widetilde{h}_j), \qquad \forall i \in [n] \tag{60}$$

$$F(\mathcal{P}')(h) = \sum_{j=1}^B p_{ij}F(\mathcal{P}'_{-i})(\widetilde{h}_j) = \sum_{j=1}^B p'_{nj}F(\mathcal{P}'_{-n})(\widetilde{h}_j), \ \forall i \in [n-1], \tag{61}$$

where $\widetilde{h}_j = (h_1, \ldots, h_{j-1}, h_j - 1, h_{j+1}, \ldots, h_B)$ for any $j \in [B]$. Here, $F(\mathcal{P}_{-i}), F(\mathcal{P}'_{-i})$ are distributions over $\mathcal{A}_B^{n-1}$.[8]

Fix any $i \in [n-1]$ and also fix arbitrary $p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n, p'_n$. Take any $\alpha \in [0,1]$, and consider $p_i^\alpha = \alpha p_i^0 + (1-\alpha)p_i^1$. Let $\mathcal{P}_\alpha = (p_1, \ldots, p_i^\alpha, \ldots, p_n)$, $\mathcal{P}_0 = (p_1, \ldots, p_i^0, \ldots, p_n)$, and $\mathcal{P}_1 = (p_1, \ldots, p_i^1, \ldots, p_n)$. Similarly, let $\mathcal{P}'_\alpha = (p_1, \ldots, p_i^\alpha, \ldots, p'_n)$, $\mathcal{P}'_0 = (p_1, \ldots, p_i^0, \ldots, p'_n)$, and $\mathcal{P}'_1 = (p_1, \ldots, p_i^1, \ldots, p'_n)$. With these definitions, we have $\mathcal{P}_\alpha = \alpha\mathcal{P}_0 + (1-\alpha)\mathcal{P}_1$. Note that $(\mathcal{P}_\alpha)_{-i} = (\mathcal{P}_0)_{-i} = (\mathcal{P}_1)_{-i}$.

Then, from the recursive definitions of $F(\mathcal{P})$ and $F(\mathcal{P}')$ (given in (60) and (61), respectively), for any $h \in \mathcal{A}_B^n$, we get

$$F(\mathcal{P}_\alpha)(h) = \sum_{j=1}^B p_{ij}^\alpha F((\mathcal{P}_\alpha)_{-i})(\widetilde{h}_j)$$

---
[8]We assume that $F(\mathcal{P}_{-i})(\widetilde{h}_j) = 0$ and $F(\mathcal{P}'_{-i})(\widetilde{h}_j) = 0$ if $h_j - 1 < 0$.

$$= \alpha \sum_{j=1}^B p_{ij}^0 F((\mathcal{P}_\alpha)_{-i})(\widetilde{h}_j) + (1-\alpha)\sum_{j=1}^B p_{ij}^1 F((\mathcal{P}_\alpha)_{-i})(\widetilde{h}_j)$$
$$\text{(since } p_i^\alpha = \alpha p_i^0 + (1-\alpha)p_i^1)$$
$$= \alpha \sum_{j=1}^B p_{ij}^0 F((\mathcal{P}_0)_{-i})(\widetilde{h}_j) + (1-\alpha)\sum_{j=1}^B p_{ij}^1 F((\mathcal{P}_1)_{-i})(\widetilde{h}_j)$$
$$\text{(since } (\mathcal{P}_\alpha)_{-i} = (\mathcal{P}_0)_{-i} = (\mathcal{P}_1)_{-i})$$
$$= \alpha F(\mathcal{P}_0)(h) + (1-\alpha)F(\mathcal{P}_1)(h).$$

Similarly, we can show that $F(\mathcal{P}'_\alpha)(h) = \alpha F(\mathcal{P}'_0)(h) + (1-\alpha)F(\mathcal{P}'_1)(h)$.
Thus we have shown that

$$F(\mathcal{P}_\alpha) = \alpha F(\mathcal{P}_0) + (1-\alpha)F(\mathcal{P}_1)$$
$$F(\mathcal{P}'_\alpha) = \alpha F(\mathcal{P}'_0) + (1-\alpha)F(\mathcal{P}'_1).$$

From Lemma B.1, we have that $\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right]$ is jointly convex in $F(\mathcal{P})$ and $F(\mathcal{P}')$. As a result, we get

$$\mathbb{E}_{h\sim F(\mathcal{P}'_\alpha)}\left[\left(\frac{F(\mathcal{P}_\alpha)(h)}{F(\mathcal{P}'_\alpha)(h)}\right)^\lambda\right] \leq \alpha\mathbb{E}_{h\sim F(\mathcal{P}'_0)}\left[\left(\frac{F(\mathcal{P}_0)(h)}{F(\mathcal{P}'_0)(h)}\right)^\lambda\right]$$
$$+ (1-\alpha)\,\mathbb{E}_{h\sim F(\mathcal{P}'_1)}\left[\left(\frac{F(\mathcal{P}_1)(h)}{F(\mathcal{P}'_1)(h)}\right)^\lambda\right] \tag{62}$$

Thus, we have shown that $\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right]$ is convex in p_i for any $i \in [n-1]$.

Now we are ready to prove Lemma 5.3.

The LDP constraints put some restrictions on the set of values that the distribution p_i can take; however, the maximum value that $\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right]$ takes can only increase when we remove those constraints. We instead maximize it w.r.t. p_i over the simplex $\Delta_B := \{(p_{i1}, \ldots, p_{iB}) : p_{ij} \geq 0 \text{ for } j \in [B] \text{ and } \sum_{j=1}^B p_{ij} = 1\}$. This implies

$$\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right] \leq \max_{p_i\in\Delta_B}\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right] \tag{63}$$

Substituting from (60) and (61) into (63), we get

$$\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{F(\mathcal{P})(h)}{F(\mathcal{P}')(h)}\right)^\lambda\right] \leq \tag{64}$$

$$\max_{p_i\in\Delta_B}\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{\sum_{j=1}^B p_{ij}F(\mathcal{P}_{-i})(\widetilde{h}_j)}{\sum_{j=1}^B p_{ij}F(\mathcal{P}'_{-i})(\widetilde{h}_j)}\right)^\lambda\right] \tag{65}$$

Since maximizing a convex function over a polyhedron attains its maximum value at one of its vertices, and there are B vertices in the simplex Δ_B, which are of the form $p_{ij^*} = 1$ for some $j^* \in [B]$ and $p_{ik} = 0$ for all $k \neq j^*$, we have

$$\max_{p_i\in\Delta_B}\mathbb{E}_{h\sim F(\mathcal{P}')}\left[\left(\frac{\sum_{j=1}^B p_{ij}F(\mathcal{P}_{-i})(\widetilde{h}_j)}{\sum_{j=1}^B p_{ij}F(\mathcal{P}'_{-i})(\widetilde{h}_j)}\right)^\lambda\right]$$

$$\overset{(a)}{=} \mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}')} \left[\left(\frac{F(\mathcal{P}_{-i})(\widetilde{\boldsymbol{h}}_{j^*})}{F(\mathcal{P}'_{-i})(\widetilde{\boldsymbol{h}}_{j^*})} \right)^\lambda \right]$$

$$\overset{(b)}{=} \mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_{-i})} \left[\left(\frac{F(\mathcal{P}_{-i})(\boldsymbol{h})}{F(\mathcal{P}'_{-i})(\boldsymbol{h})} \right)^\lambda \right]$$

Since the i'th data point deterministically maps to the j^*'th output by the mechanism $\mathcal{R}$, the expectation term in the RHS of (a) has no dependence on the i'th data point, so we can safely remove that, which gives (b). This proves Lemma 5.3. ∎

B.4 Proof of Corollary 5.4

Corollary (Restating Corollary 5.4). Consider any $m \in \{0, 1, \ldots, n-1\}$. Let $\mathcal{D}_{m+1}^{(n)} = (d'_n, \ldots, d'_n, d_n)$ and $\mathcal{D}_{m+1}'^{(n)} = (d'_n, \ldots, d'_n)$. Then for any $C \in \binom{[n-1]}{m}$, we have

$$\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_C)} \left[\left(\frac{F(\mathcal{P}_C)(\boldsymbol{h})}{F(\mathcal{P}'_C)(\boldsymbol{h})} \right)^\lambda \right] \leq \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})} \left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})} \right)^\lambda \right].$$

PROOF. Recall from Lemma 5.1 and the notation defined in Appendix B, that for any $C \subseteq [n-1]$, we have $\mathcal{P}_C = \mathcal{P}'_{|C|,n} \cup \widetilde{\mathcal{P}}_{[n-1] \setminus C} \cup \{\boldsymbol{p}_n\}$ and $\mathcal{P}'_C = \mathcal{P}'_{|C|,n} \cup \widetilde{\mathcal{P}}_{[n-1] \setminus C} \cup \{\boldsymbol{p}'_n\}$, where $\mathcal{P}'_{|C|,n} = \{\boldsymbol{p}'_n, \ldots, \boldsymbol{p}'_n\}$ with $|\mathcal{P}'_{|C|,n}| = |C|$ and $\widetilde{\mathcal{P}}_{[n-1] \setminus C} = \{\tilde{\boldsymbol{p}}_i : i \in [n-1] \setminus C\}$.

Now, repeatedly applying Lemma 5.3 over the set of distributions $\tilde{\boldsymbol{p}}_i \in \widetilde{\mathcal{P}}_{[n-1] \setminus C}$, we get that

$$\mathbb{E}_{\boldsymbol{h} \sim F(\mathcal{P}'_C)} \left[\left(\frac{F(\mathcal{P}_C)(\boldsymbol{h})}{F(\mathcal{P}'_C)(\boldsymbol{h})} \right)^\lambda \right]$$

$$\leq \mathbb{E}_{\boldsymbol{h} \sim F\left(\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}'_n\} \right)} \left[\left(\frac{F\left(\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}_n\} \right)(\boldsymbol{h})}{F\left(\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}'_n\} \right)(\boldsymbol{h})} \right)^\lambda \right]$$

$$= \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})} \left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})} \right)^\lambda \right]$$

In the last equality, we used that $\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}_n\}$ has $|C| + 1 = m + 1$ distributions which are associated with the $(m + 1)$ data points $\{d'_n, \ldots, d'_n, d_n\}$ (m of them are equal to d'_n); similarly, $\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}'_n\}$ also has $|C| + 1 = m + 1$ distributions which are associated with the $(m + 1)$ data points $\{d'_n, \ldots, d'_n, d'_n\}$ (all of them are equal to d'_n). This implies that for every $\boldsymbol{h} \in \mathcal{A}_B^{m+1}$, $F\left(\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}_n\} \right)(\boldsymbol{h})$ and $F\left(\mathcal{P}'_{|C|,n} \cup \{\boldsymbol{p}'_n\} \right)(\boldsymbol{h})$ are distributionally equal to $\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})$ and $\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})$, respectively.

This proves Corollary 5.4. ∎

C OMITTED DETAILS FROM SECTION 6

C.1 Proof of Lemma 6.1

LEMMA (RESTATING LEMMA 6.1). *The random variable X has the following properties:*

(1) *X has zero mean, i.e., $\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}[X(\boldsymbol{h})] = 0$.*

(2) *The variance of X is equal to*

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[X(\boldsymbol{h})^2\right] = m\left(\sum_{j=1}^B \frac{p_j'^2}{p_j} - 1\right).$$

(3) *For $i \geq 3$, the ith moment of X is bounded by*

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[(X(\boldsymbol{h}))^i\right] \leq \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[|X(\boldsymbol{h})|^i\right] \leq i\Gamma\,(i/2)\left(2mv^2\right)^{i/2},$$

where $v^2 = \frac{(e^{\epsilon_0} - e^{-\epsilon_0})^2}{4}$ and $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ is the Gamma function.

PROOF. For simplicity of notation, let μ_0, μ_1 denote the distributions $\mathcal{M}(\mathcal{D}_m), \mathcal{M}(\mathcal{D}_m')$, respectively. As shown in (32), for any $\boldsymbol{h} \in \mathcal{A}_B^m$, we have

$$X(\boldsymbol{h}) = m\left(\frac{\mu_1(\boldsymbol{h})}{\mu_0(\boldsymbol{h})} - 1\right) = \left(\sum_{j=1}^B a_j h_j\right) - m,$$

where $a_j = \frac{p_j'}{p_j} \in [e^{-\epsilon_0}, e^{\epsilon_0}]$ for all $j \in [B]$.

Now we show the three properties.

(1) The mean of the random variable X is given by

$$\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[X(\boldsymbol{h})] = m\mathbb{E}_{\boldsymbol{h} \sim \mu_0}\left[\frac{\mu_1(\boldsymbol{h})}{\mu_0(\boldsymbol{h})} - 1\right]$$

$$= m\sum_{\boldsymbol{h} \in \mathcal{A}_B^m} \mu_0(\boldsymbol{h})\left(\frac{\mu_1(\boldsymbol{h})}{\mu_0(\boldsymbol{h})} - 1\right)$$

$$= m\sum_{\boldsymbol{h} \in \mathcal{A}_B^m} (\mu_1(\boldsymbol{h}) - \mu_0(\boldsymbol{h})) = 0$$

(2) The variance of the random variable X is given by

$$\mathbb{E}_{\boldsymbol{h} \sim \mu_0}\left[X(\boldsymbol{h})^2\right] = \mathbb{E}_{\boldsymbol{h} \sim \mu_0}\left[\left(\sum_{j=1}^B a_j h_j - m\right)^2\right]$$

$$= m^2\mathbb{E}_{\boldsymbol{h} \sim \mu_0}\left[\sum_{j=1}^B\sum_{l=1}^B a_j a_l \frac{h_j h_l}{m^2} - 2\sum_{j=1}^B a_j\frac{h_j}{m} + 1\right]$$

$$= m^2\mathbb{E}_{\boldsymbol{h} \sim \mu_0}\left[\sum_{j=1}^B a_j^2\frac{h_j^2}{m^2} + \sum_{j=1}^B\sum_{l \neq j} a_j a_l\frac{h_j h_l}{m^2} - 2\sum_{j=1}^B a_j\frac{h_j}{m} + 1\right]$$

$$= m^2\Big[\sum_{j=1}^B \frac{(p_j')^2}{p_j^2}\frac{\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[h_j^2]}{m^2} + \sum_{j=1}^B\sum_{l \neq j}\frac{p_j'p_l'}{p_j p_l}\frac{\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[h_j h_l]}{m^2}$$

$$- 2\sum_{j=1}^B \frac{p_j'}{p_j}\frac{\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[h_j]}{m} + 1\Big]$$

$$\overset{(b)}{=} m^2\Big[\sum_{j=1}^B \frac{(p_j')^2}{p_j^2}\frac{(mp_j(1-p_j) + m^2 p_j^2)}{m^2}$$

$$+ \sum_{j=1}^B\sum_{l \neq j}\frac{p_j'p_l'}{p_j p_l}\frac{(-mp_j p_l + m^2 p_j p_l)}{m^2} - 2\sum_{j=1}^B \frac{p_j'}{p_j}\frac{p_j m}{m} + 1\Big]$$

$$= m^2\Big[\sum_{j=1}^B \left(\frac{(p_j')^2(1-p_j)}{p_j m} + (p_j')^2\right)$$

$$+ \sum_{j=1}^{B} \sum_{l \neq j} \left(-\frac{p'_j p'_l}{m} + p'_j p'_l \right) - 1 \Bigg]$$

$$= m^2 \Bigg[\frac{1}{m} \left(\sum_{j=1}^{B} \frac{(p'_j)^2 (1 - p_j)}{p_j} - \sum_{j=1}^{B} \sum_{l \neq j} p'_j p'_l \right)$$

$$+ \sum_{j=1}^{B} (p'_j)^2 + \sum_{j=1}^{B} \sum_{l \neq j} p'_j p'_l - 1 \Bigg]$$

$$= m^2 \Bigg[\frac{1}{m} \left(\sum_{j=1}^{B} \frac{(p'_j)^2}{p_j} - \sum_{j=1}^{B} (p'_j)^2 - \sum_{j=1}^{B} \sum_{l \neq j} p'_j p'_l \right)$$

$$+ \sum_{j=1}^{B} (p'_j)^2 + \sum_{j=1}^{B} \sum_{l \neq j} p'_j p'_l - 1 \Bigg]$$

$$\overset{(c)}{=} m \left(\sum_{j=1}^{B} \frac{(p'_j)^2}{p_j} - 1 \right).$$

Here, step (b) uses properties of multinomial distribution: $\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[h_j] = mp_j$, $\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[h_j^2] = mp_j(1 - p_j) + m^2 p_j^2$, and $\mathbb{E}_{\boldsymbol{h} \sim \mu_0}[h_j h_l] = -mp_j p_l + m^2 p_j p_l$ for $j \neq l$. Step (c) follows because $\sum_{j=1}^{B} (p'_j)^2 + \sum_{j=1}^{B} \sum_{l \neq j} p'_j p'_l = \left(\sum_{j=1}^{B} p'_j \right)^2 = 1$, as $\boldsymbol{p}' = (p'_1, \ldots, p'_B)$ is a probability distribution.

(3) Let Y_i denote the random variable associated with the output of the local randomizer at the i'th client. So, $\Pr[Y_i = j] = p_j$ for $j \in [B]$. Recall that h_j denote the number of clients that map to the j'th element from $[B]$. This implies that for any $j \in [B]$, we have $h_j = \sum_{i=1}^{m} \mathbf{1}_{\{Y_i = j\}}$. For any $i \in [m]$, define a random variable $X_i = \left(\sum_{j=1}^{B} a_j \mathbf{1}_{\{Y_i = j\}} \right) - 1$, where $a_j = \frac{p'_j}{p_j}$. Observe that $X_1, \ldots, X_m$ are zero mean i.i.d. random variables, because for any $i \in [m]$, we have $\mathbb{E}[X_i] = \left(\sum_{j=1}^{B} a_j p_j \right) - 1 = 0$. With these definitions, we can equivalently represent $X(\boldsymbol{h}) = \left(\sum_{j=1}^{B} a_j h_j \right) - m$ as $X(\boldsymbol{h}) = \sum_{i=1}^{m} X_i$, which is the sum of m zero mean i.i.d. r.v.s. Furthermore, since $a_j \in [e^{-\epsilon_0}, e^{\epsilon_0}]$ for any $j \in [B]$, we have $X_i \in [e^{-\epsilon_0} - 1, e^{\epsilon_0} - 1]$. Since any bounded r.v. $Z \in [a, b]$ is a sub-Gaussian r.v. with parameter $\frac{(b-a)^2}{4}$ (see [38, Lemma 1.8])), we have that X_i is a sub-Gaussian r.v. with parameter $v^2 = \frac{(e^{\epsilon_0} - e^{-\epsilon_0})^2}{4}$, i.e.,

$$\mathbb{E}\left[e^{sX_i} \right] \leq e^{\frac{s^2 v^2}{2}}, \qquad \forall s \in \mathbb{R}.$$

It follows that $X(\boldsymbol{h}) = \sum_{i=1}^{m} X_i$ is also a sub-Gaussian random variable with parameter mv^2. The remaining steps are similar to bound the moments of a sub-Gaussian random variable. We write them here for completeness. From Chernoff bound we get

$$\Pr[X \geq t] \leq \min_{s \geq 0} \frac{\mathbb{E}\left[e^{sX} \right]}{e^{st}}$$

$$\leq \min_{s \geq 0} \frac{e^{\frac{s^2 m v^2}{2}}}{e^{st}}$$

$$\overset{(b)}{\leq} e^{-\frac{t^2}{2mv^2}}$$

where (b) follows by setting $s = \frac{t}{mv^2}$. Similarly, we can bound the term $\Pr[-X \geq t]$. Thus, we get

$$\Pr[|X| \geq t] \leq 2e^{-\frac{t^2}{2mv^2}}$$

Hence, the i'th moment of the random variable X can be bounded by

$$\mathbb{E}\left[X^i \right] \leq \mathbb{E}\left[|X|^i \right]$$

$$= i \int_0^\infty t^{i-1} \Pr[|X| \geq t] \, dt$$

$$\leq 2i \int_0^\infty t^{i-1} e^{-\frac{t^2}{2mv^2}} \, dt$$

$$\overset{(b)}{=} i \left(2mv^2 \right)^{i/2} \int_0^\infty u^{i/2 - 1} e^{-u} \, du$$

$$= i \left(2mv^2 \right)^{i/2} \Gamma(i/2),$$

where step (b) follows by setting $u = \frac{t^2}{2mv^2}$ (change of variables). In the last step, $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ denotes the Gamma function. Thus, we conclude that for every $i \geq 3$, we have $\mathbb{E}\left[|X|^i \right] \leq i\Gamma(i/2) \left(2mv^2 \right)^{i/2}$, where $v^2 = \frac{(e^{\epsilon_0} - e^{-\epsilon_0})^2}{4}$.

This completes the proof of Lemma 6.1. ∎

C.2 Proof of Lemma 6.2

LEMMA (RESTATING LEMMA 6.2). *We have the following bound:*

$$\sup_{(\boldsymbol{p}, \boldsymbol{p}') \in \mathcal{T}_{\epsilon_0}} \left(\sum_{j=1}^{B} \frac{p'^2_j}{p_j} - 1 \right) = \frac{(e^{\epsilon_0} - 1)^2}{e^{\epsilon_0}}.$$

PROOF. For any $(\boldsymbol{p}, \boldsymbol{p}') \in \mathcal{T}_{\epsilon_0}$, define $f(\boldsymbol{p}, \boldsymbol{p}') = \sum_{j=1}^{B} \frac{(p'_j)^2}{p_j}$. Since the function $g(x, y) = \frac{x^2}{y}$ is convex in (x, y) for $y > 0$, it implies that the objective function $f(\boldsymbol{p}, \boldsymbol{p}')$ is also convex in $(\boldsymbol{p}, \boldsymbol{p}')$. It is easy to verify that $\mathcal{T}_{\epsilon_0}$ is a polytope.

Since we maximize a convex function $f(\boldsymbol{p}, \boldsymbol{p}')$ over a polytope $\mathcal{T}_{\epsilon_0}$, the optimal solution is one of the vertices of the polytope. Note that any vertex $(\boldsymbol{p}, \boldsymbol{p}')$ of the polytope in B dimensions satisfies all the B LDP constraints (i.e., $e^{-\epsilon_0} \leq \frac{p_j}{p'_j} \leq e^{\epsilon_0}$, $j = 1, \ldots, B$) with equality. Without loss of generality, assume that the optimal solution $(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{p}}')$ is a vertex such that $\frac{\tilde{p}'_j}{\tilde{p}_j} = e^{\epsilon_0}$ for $j = 1, \ldots, l$ and $\frac{\tilde{p}'_j}{\tilde{p}_j} = e^{-\epsilon_0}$ for $j = l+1, \ldots, B$, for some $l \in [B]$. Thus, we have

$$1 = \sum_{j=1}^{B} \tilde{p}'_j = e^{\epsilon_0} \sum_{j=1}^{l} \tilde{p}_j + e^{-\epsilon_0} \sum_{j=l+1}^{B} \tilde{p}_j$$

$$= e^{\epsilon_0} \sum_{j=1}^{l} \tilde{p}_j + e^{-\epsilon_0} \left(1 - \sum_{j=1}^{l} \tilde{p}_j \right) = e^{-\epsilon_0} + (e^{\epsilon_0} - e^{-\epsilon_0}) \sum_{j=1}^{l} \tilde{p}_j$$

Rearranging the above gives $\sum_{j=1}^{l} \tilde{p}_j = \frac{1}{e^{\epsilon_0}+1}$. This implies $\sum_{j=1}^{l} \tilde{p}_j' = \frac{e^{\epsilon_0}}{e^{\epsilon_0}+1}$, which in turn implies $\sum_{j=l+1}^{B} \tilde{p}_j' = \frac{1}{e^{\epsilon_0}+1}$. Now the result follows from the following set of equalities:

$$f\left(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{p}}'\right) = \sum_{j=1}^{B} \frac{(\tilde{p}_j')^2}{\tilde{p}_j} = \sum_{j=1}^{l} \frac{\tilde{p}_j'}{\tilde{p}_j}\tilde{p}_j' + \sum_{j=l+1}^{B} \frac{\tilde{p}_j'}{\tilde{p}_j}\tilde{p}_j'$$

$$= e^{\epsilon_0} \sum_{j=1}^{l} \tilde{p}_j' + e^{-\epsilon_0} \sum_{j=l+1}^{B} \tilde{p}_j'$$

$$= \frac{e^{2\epsilon_0}}{e^{\epsilon_0}+1} + \frac{1}{e^{\epsilon_0}(e^{\epsilon_0}+1)} = \frac{(e^{\epsilon_0})^3+1}{e^{\epsilon_0}(e^{\epsilon_0}+1)} = \frac{(e^{\epsilon_0}-1)^2}{e^{\epsilon_0}} + 1,$$

where the last equality uses the identity $x^3 + 1 = (x+1)(x^2-x+1)$. This completes the proof of Lemma 6.2. ∎

D OMITTED DETAILS FROM SECTION 6

D.1 Omitted Details from Section 7.1

Before proving (40), first we show an important property of E_m that we will use in the proof.

LEMMA D.1. E_m is a non-increasing function of m, i.e.,

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_{m+1}'^{(n)})}\left[\left(\frac{\mathcal{M}(\mathcal{D}_{m+1}^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_{m+1}'^{(n)})(\boldsymbol{h})}\right)^\lambda\right]$$

$$\leq \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m'^{(n)})}\left[\left(\frac{\mathcal{M}(\mathcal{D}_m^{(n)})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m^{(n)})(\boldsymbol{h})}\right)^\lambda\right],$$
(66)

where, for any $k \in \{m, m+1\}$, $\mathcal{D}_k^{(n)} = (d_n', \ldots, d_n', d_n)$ and $\mathcal{D}_k'^{(n)} = (d_n', \ldots, d_n', d_n')$ with $|\mathcal{D}_k| = |\mathcal{D}_k'| = k$.

PROOF. Lemma D.1 follows from Lemma 5.3 in a straightforward manner, as, unlike Lemma D.1, in Lemma 5.3 we consider arbitrary pairs of neighboring datasets. ∎

Now we can prove (40).

PROOF OF (40).

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}')}\left[\left(\frac{\mathcal{M}(\mathcal{D})(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}')(\boldsymbol{h})}\right)^\lambda\right] \leq \sum_{m=0}^{n-1} q_m E_m$$

$$= \sum_{m < \lfloor(1-\gamma)q(n-1)\rfloor} q_m E_m + \sum_{m \geq \lfloor(1-\gamma)q(n-1)\rfloor} q_m E_m$$

$$\overset{(a)}{\leq} E_0 \sum_{m < \lfloor(1-\gamma)q(n-1)\rfloor} q_m + \sum_{m \geq \lfloor(1-\gamma)q(n-1)\rfloor} q_m E_m$$

$$\overset{(b)}{\leq} E_0 e^{-\frac{q(n-1)\gamma^2}{2}} + \sum_{m \geq \lfloor(1-\gamma)q(n-1)\rfloor} q_m E_m$$

$$\overset{(c)}{\leq} e^{\epsilon_0 \lambda} e^{-\frac{q(n-1)\gamma^2}{2}} + \sum_{m \geq \lfloor(1-\gamma)q(n-1)\rfloor} q_m E_m$$

$$\overset{(d)}{\leq} e^{\epsilon_0 \lambda} e^{-\frac{q(n-1)\gamma^2}{2}} + E_{(1-\gamma)q(n-1)}.$$

Here, steps (a) and (d) follow from the fact that E_m is a non-increasing function of m (see Lemma D.1). Step (b) follows from the Chernoff

bound. In step (c), we used that $\mathcal{M}(d_n) = \mathcal{R}(d_n)$ and $\mathcal{M}(d_n') = \mathcal{R}(d_n')$, which together imply that

$$E_0 = \mathbb{E}\left[\left(\frac{\mathcal{M}(d_n)}{\mathcal{M}(d_n')}\right)^\lambda\right] = \mathbb{E}\left[\left(\frac{\mathcal{R}(d_n)}{\mathcal{R}(d_n')}\right)^\lambda\right] \leq e^{\epsilon_0 \lambda},$$

where the inequality follows because $\mathcal{R}$ is an ϵ_0-LDP mechanism. ∎

D.2 Proof of Theorem 7.1

THEOREM (RESTATING THEOREM 7.1). Let $m \in \mathbb{N}$ be arbitrary. For any $\lambda \geq 2$ (including the non-integral λ), we have

$$\sup_{(\mathcal{D}_m, \mathcal{D}_m') \in \mathcal{D}_{\text{same}}^m} \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[\left(\frac{\mathcal{M}(\mathcal{D}_m')(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})}\right)^\lambda\right]$$
$$\leq \exp\left(\lambda^2 \frac{(e^{\epsilon_0}-1)^2}{m}\right).$$
(67)

PROOF. Fix an arbitrary $m \in \mathbb{N}$. Let $(\mathcal{D}_m, \mathcal{D}_m') \in \mathcal{D}_{\text{same}}^m$ and $\boldsymbol{p} = (p_1, \ldots, p_B)$, $\boldsymbol{p}' = (p_1', \ldots, p_B')$ be the same as defined in the proof of Theorem 3.7 in Section 6.

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[\left(\frac{\mathcal{M}(\mathcal{D}_m')(\boldsymbol{h})}{\mathcal{M}(\mathcal{D}_m)(\boldsymbol{h})}\right)^\lambda\right] = \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[\left(\sum_{j=1}^{B} \frac{p_j'}{p_j}\frac{h_j}{m}\right)^\lambda\right]$$

$$= \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[\left(1 + \sum_{j=1}^{B} \frac{p_j'}{p_j}\frac{h_j}{m} - 1\right)^\lambda\right]$$

$$\leq \mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[\exp\left(\lambda\left(\sum_{j=1}^{B} \frac{p_j'}{p_j}\frac{h_j}{m} - 1\right)\right)\right],$$
(68)

where the first equality uses (32) and the last inequality follows from $1 + x \leq e^x$.

In (68), $\boldsymbol{h}$ is distributed according to $\mathcal{M}(\mathcal{D}_m) = \mathcal{H}_m(\mathcal{R}(d), \ldots, \mathcal{R}(d))$, where $\mathcal{H}_m$ denotes the shuffling operation on m elements and range of $\mathcal{R}$ is equal to $[B]$. Since all the m data points are identical, and all clients use independent randomness for computing $\mathcal{R}(d)$, we can assume, w.l.o.g., that $\mathcal{M}(\mathcal{D}_m)$ is a collection of m i.i.d. random variables $X_1, \ldots, X_m$, where $\Pr[X_i = j] = p_j$ for $j \in [B]$. Thus, we have (in the following, note that $\boldsymbol{h} = (h_1, \ldots, h_B)$ is a r.v.)

$$\frac{1}{m}\sum_{j=1}^{B} \frac{p_j'}{p_j}h_j = \frac{1}{m}\sum_{j=1}^{B} \frac{p_j'}{p_j}\sum_{i=1}^{m} \mathbf{1}_{\{X_i=j\}}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{B} \frac{p_j'}{p_j}\mathbf{1}_{\{X_i=j\}} = \frac{1}{m}\sum_{i=1}^{m} \frac{p_{X_i}'}{p_{X_i}},$$
(69)

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator r.v. Substituting from (69) into (68), we get

$$\mathbb{E}_{\boldsymbol{h} \sim \mathcal{M}(\mathcal{D}_m)}\left[\exp\left(\lambda\left(\sum_{j=1}^{B} \frac{p_j'}{p_j}\frac{h_j}{m} - 1\right)\right)\right]$$

$$= \mathbb{E}_{X_1,\ldots,X_m}\left[\exp\left(\frac{\lambda}{m}\sum_{i=1}^{m}\left(\frac{p_{X_i}'}{p_{X_i}} - 1\right)\right)\right]$$

$$= \prod_{i=1}^{m} \mathbb{E}_{X_i} \left[\exp \left(\frac{\lambda}{m} \left(\frac{p'_{X_i}}{p_{X_i}} - 1 \right) \right) \right]$$

$$= \left(\mathbb{E}_{X \sim p} \left[e^{\frac{\lambda}{m} \left(\frac{p'_X}{p_X} - 1 \right)} \right] \right)^{m} \tag{70}$$

where $p = [p_1, \ldots, p_B]$. From Taylor expansion of $e^x = 1 + \sum_{k=1}^{\infty} \frac{x^k}{k!}$, we get

$$\mathbb{E}_{X \sim p} \left[e^{\frac{\lambda}{m} \left(\frac{p'_X}{p_X} - 1 \right)} \right] = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbb{E}_{X \sim p} \left[\left(\frac{\lambda}{m} \left(\frac{p'_X}{p_X} - 1 \right) \right)^{k} \right]$$

$$= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{j=1}^{B} p_j \left(\frac{\lambda}{m} \left(\frac{p'_j}{p_j} - 1 \right) \right)^{k}$$

$$= 1 + \sum_{k=2}^{\infty} \frac{1}{k!} \sum_{j=1}^{B} p_j \left(\frac{\lambda}{m} \left(\frac{p'_j}{p_j} - 1 \right) \right)^{k}$$

$$\leq 1 + \sum_{k=2}^{\infty} \frac{1}{k!} \sum_{j=1}^{B} p_j \left(\frac{\lambda (e^{\epsilon_0} - 1)}{m} \right)^{k}$$

$$= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\lambda (e^{\epsilon_0} - 1)}{m} \right)^{k} - \frac{\lambda (e^{\epsilon_0} - 1)}{m}$$

$$= e^{\frac{\lambda (e^{\epsilon_0} - 1)}{m}} - \frac{\lambda (e^{\epsilon_0} - 1)}{m}, \tag{71}$$

where the inequality follows from $\frac{p'_j}{p_j} \leq e^{\epsilon_0}$, which holds for all $j \in [B]$. Substituting from (71) into (70), we get

$$\mathbb{E}_{h \sim M(\mathcal{D}_m)} \left[\left(\frac{M(\mathcal{D}'_m)(h)}{M(\mathcal{D}_m)(h)} \right)^{\lambda} \right] \leq \left(e^{\frac{\lambda (e^{\epsilon_0} - 1)}{m}} - \frac{\lambda (e^{\epsilon_0} - 1)}{m} \right)^{m}$$

$$= e^{\lambda (e^{\epsilon_0} - 1)} \left[1 - \frac{\lambda (e^{\epsilon_0} - 1)}{m} e^{\frac{-\lambda (e^{\epsilon_0} - 1)}{m}} \right]^{m}$$

$$\leq e^{\lambda (e^{\epsilon_0} - 1)} e^{-\lambda (e^{\epsilon_0} - 1) e^{\frac{-\lambda (e^{\epsilon_0} - 1)}{m}}} \quad \text{(since } 1 - x \leq e^{-x}\text{)}$$

$$= e^{\lambda (e^{\epsilon_0} - 1) \left[1 - e^{\frac{-\lambda (e^{\epsilon_0} - 1)}{m}} \right]}$$

$$\leq e^{\frac{\lambda^2 (e^{\epsilon_0} - 1)^2}{m}}. \quad \text{(since } 1 - e^{-x} \leq x\text{)}$$

This completes the proof of Theorem 7.1. ∎

E OMITTED DETAILS FROM SECTION 8

In this section, we provide a complete proof of Theorem 3.4.

Consider the binary case, where each data point d can take a value from $\mathcal{X} = \{0, 1\}$. Let the local randomizer $\mathcal{R}$ be the binary randomized response (2RR) mechanism, where $\Pr[\mathcal{R}(d) = d] = \frac{e^{\epsilon_0}}{e^{\epsilon_0} + 1}$ for $d \in \mathcal{X}$. It is easy to verify that $\mathcal{R}$ is an ϵ_0-LDP mechanism. For simplicity, let $p = \frac{1}{e^{\epsilon_0} + 1}$. Consider two neighboring datasets $\mathcal{D}, \mathcal{D}' \in \{0, 1\}^n$, where $\mathcal{D} = (0, \ldots, 0, 0)$ and $\mathcal{D}' = (0, \ldots, 0, 1)$. Let $k \in \{0, \ldots, n\}$ denote the number of ones in the output of the shuffler. As argued in Section 2.3 on page 4, since the output of the shuffle mechanism $\mathcal{M}$ can be thought of as the distribution of the number of ones in the output, we have that $k \sim \mathcal{M}(\mathcal{D})$ is

distributed as a Binomial random variable $\text{Bin}(n, p)$. Thus, we have

$$\mathcal{M}(\mathcal{D})(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\mathcal{M}(\mathcal{D}')(k) = (1 - p) \binom{n-1}{k-1} p^{k-1} (1 - p)^{n-k}$$

$$+ p \binom{n-1}{k} p^k (1 - p)^{n-k-1}.$$

It will be useful to compute $\frac{\mathcal{M}(\mathcal{D})(k)}{\mathcal{M}(\mathcal{D}')(k)} - 1$ for the calculations later.

$$\frac{\mathcal{M}(\mathcal{D}')(k)}{\mathcal{M}(\mathcal{D})(k)} - 1 = \frac{k}{n} \frac{(1 - p)}{p} + \frac{(n - k)}{n} \frac{p}{(1 - p)} - 1$$

$$= \frac{k}{n} e^{\epsilon_0} + \frac{(n - k)}{n} e^{-\epsilon_0} - 1$$

$$= \frac{k}{n} \left(e^{\epsilon_0} - e^{-\epsilon_0} \right) + e^{-\epsilon_0} - 1$$

$$= \frac{k}{n} \left(\frac{e^{2\epsilon_0} - 1}{e^{\epsilon_0}} \right) - \left(\frac{e^{\epsilon_0} - 1}{e^{\epsilon_0}} \right)$$

$$= \left(\frac{e^{2\epsilon_0} - 1}{n e^{\epsilon_0}} \right) \left(k - \frac{n}{e^{\epsilon_0} + 1} \right) \tag{72}$$

Thus, we have that

$$\mathbb{E}_{k \sim \mathcal{M}(\mathcal{D})} \left[\left(\frac{\mathcal{M}(\mathcal{D}')(k)}{\mathcal{M}(\mathcal{D})(k)} \right)^{\lambda} \right] = \mathbb{E} \left[\left(1 + \frac{\mathcal{M}(\mathcal{D}')(k)}{\mathcal{M}(\mathcal{D})(k)} - 1 \right)^{\lambda} \right]$$

$$\stackrel{(a)}{=} 1 + \sum_{i=1}^{\lambda} \binom{\lambda}{i} \mathbb{E} \left[\left(\frac{\mathcal{M}(\mathcal{D}')(k)}{\mathcal{M}(\mathcal{D})(k)} - 1 \right)^{i} \right]$$

$$\stackrel{(b)}{=} 1 + \sum_{i=2}^{\lambda} \binom{\lambda}{i} \mathbb{E} \left[\left(\frac{\mathcal{M}(\mathcal{D}')(k)}{\mathcal{M}(\mathcal{D})(k)} - 1 \right)^{i} \right]$$

$$= 1 + \sum_{i=2}^{\lambda} \binom{\lambda}{i} \left(\frac{(e^{2\epsilon_0} - 1)}{n e^{\epsilon_0}} \right)^{i} \mathbb{E} \left[\left(k - \frac{n}{e^{\epsilon_0} + 1} \right)^{i} \right] \quad \text{(from (72))}$$

$$\stackrel{(c)}{=} 1 + \binom{\lambda}{2} \frac{(e^{\epsilon_0} - 1)^2}{n e^{\epsilon_0}} + \sum_{i=3}^{\lambda} \binom{\lambda}{i} \left(\frac{(e^{2\epsilon_0} - 1)}{n e^{\epsilon_0}} \right)^{i} \mathbb{E} \left[\left(k - \frac{n}{e^{\epsilon_0} + 1} \right)^{i} \right].$$

Here, step (a) from the polynomial expansion $(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$, step (b) follows because the term corresponding to $i = 1$ is zero (i.e., $\mathbb{E}_{k \sim \mathcal{M}(\mathcal{D})} \left[\left(\frac{\mathcal{M}(\mathcal{D}')(k)}{\mathcal{M}(\mathcal{D})(k)} - 1 \right) \right] = 0$), and step (c) from the from the fact that $\mathbb{E}_{k \sim \mathcal{M}(\mathcal{D})} \left[\left(k - \frac{n}{e^{\epsilon_0} + 1} \right)^2 \right] = np(1 - p) = \frac{n e^{\epsilon_0}}{(e^{\epsilon_0} + 1)^2}$, which is equal to the variance of the Binomial random variable.

In view of Remark 6, this completes the proof of Theorem 3.4.