A Additional Experimental Results

In this section, we present some additional experimental results. For completeness we also include the subset of experiments already described in Section 4. We consider three popular randomized privacy mechanisms: (1) Gaussian mechanism (2) Laplace mechanism, and (3) randomized response mechanism, and investigate the amplification effect of subsampling with these mechanisms on RDP. The evaluation relies on the following closed-form expressions for the RDP guarantees of the Gaussian, Laplace, and randomized response mechanism (See, Mironov, 2017, Table II):

$$\epsilon_{\text{Gaussian},\sigma}(\alpha) = \frac{\alpha}{2\sigma^2},$$
$$\epsilon_{\text{Laplace},b}(\alpha) = \frac{1}{\alpha-1}\log\left(\left(\frac{\alpha}{2\alpha-1}\right)e^{(\alpha-1)/b} + \left(\frac{\alpha-1}{2\alpha-1}\right)e^{-\alpha/b}\right) \text{ for } \alpha > 1,$$
$$\epsilon_{\text{RandResp},p}(\alpha) = \frac{1}{\alpha-1}\log\left(p^\alpha(1-p)^{1-\alpha} + (1-p)^\alpha p^{1-\alpha}\right) \text{ for } \alpha > 1.$$

Here σ^2 represents the variance of the Gaussian perturbation, $2b^2$ the variance of the Laplace perturbation, and p the probability of replying truthfully in randomized response. We considered two groups of parameters σ, b, p for the three base mechanisms $\mathcal{M}$. The subsampling ratio γ is taken to be 0.001 for both regimes.

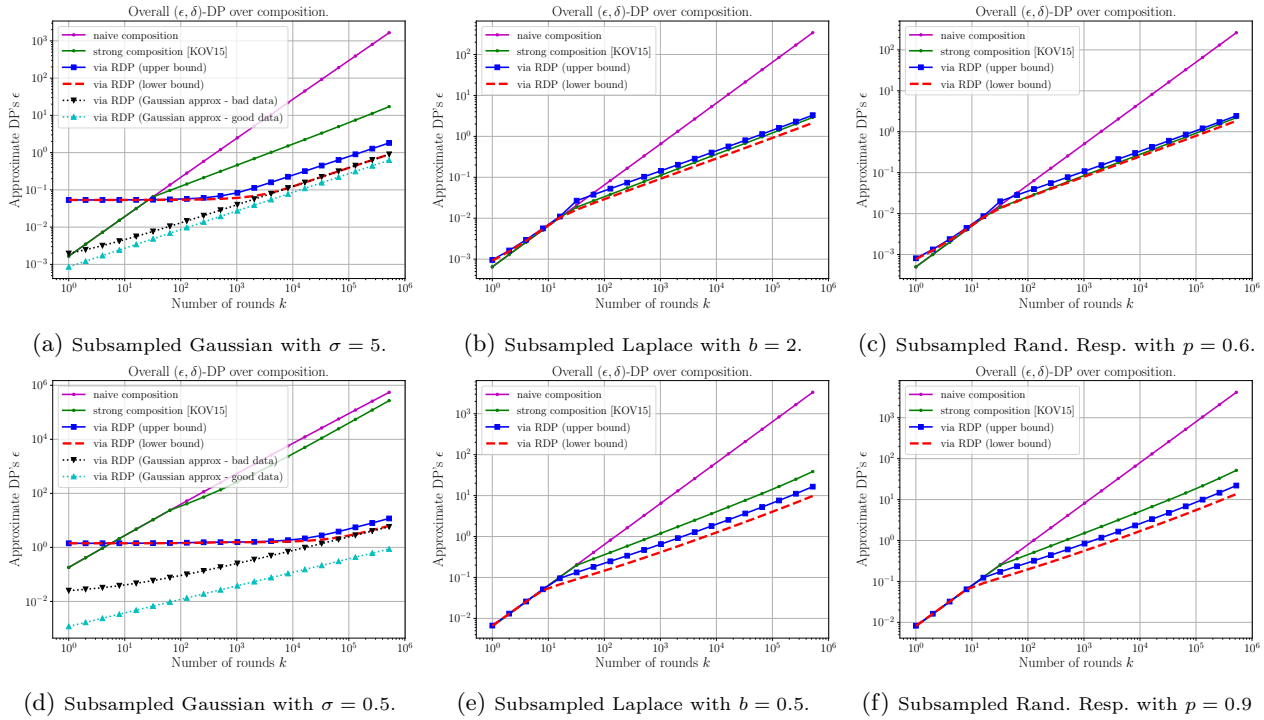
High Privacy Regime. We set $\sigma = 5$, $b = 2$ and $p = 0.6$. These correspond to $(0.2\sqrt{2\log(1.25/\delta)}, \delta)$-DP, $(0.5, 0)$-DP, and approximately $(0.41, 0)$-DP for the Gaussian, Laplace, and Randomized response mechanisms, respectively.

Low Privacy Regime. We set $\sigma = 1$, $b = 0.5$ and $p = 0.9$. These correspond to $(\sqrt{2\log(1.25/\delta)}, \delta)$-DP, $(2, 0)$-DP, and approximately $(2.2, 0)$-DP for the Gaussian, Laplace, and Randomized response mechanisms, respectively.

First we compare the effect of composition on the three subsampled mechanisms for both privacy regimes. The results are given in Figure 2, where we plot the overall (ϵ, δ)-DP for $\delta = 10^{-8}$ as we compose each of the three subsampled mechanisms for $600,000$ times. The ϵ is obtained as a function of δ for each k separately by calling the $\delta \Rightarrow \epsilon$ query in our analytical moments accountant. Our results are compared to the algorithm-independent techniques for differential privacy including naïve composition and strong composition. The strong composition baseline is carefully calibrated for each k by choosing an appropriate pair of $(\tilde{\epsilon}, \tilde{\delta})$ for $\mathcal{M}$ such that the overall (ϵ, δ)-DP guarantee that comes from composing k rounds of $\mathcal{M} \circ \mathsf{subsample}$ using Kairouz et al. (2015) obeys that $\delta < 10^{-8}$ and ϵ is minimized. Each round is described by the $(\log(1+\gamma(e^{\tilde{\epsilon}}-1)), \gamma\tilde{\delta})$-DP guarantee using the standard subsampling lemma (Lemma 3) and $\tilde{\epsilon}$ is obtained as a function of $\tilde{\delta}$ via (2).

We first observe that in the high privacy regime the Gaussian mechanism has a qualitatively similar behavior as the one observed in Section 4 for the low privacy regime. In particular, as the number of rounds of composition continues to grow, we end up having about an order of magnitude smaller ϵ than the baseline approaches in the high privacy regime (see Figure 2d). The results for composing subsampled Laplace mechanisms and subsampled randomized response mechanisms are shown in Figures 2b, 2c, 2e, and 2f. In this case, the RDP approach achieves about the same or better ϵ bound for all k when compared to what is obtained using the subsampling lemma and strong composition.

Next, we compare the upper and lower bounds for $\epsilon'(\alpha)$ given by Theorem 9 and Proposition 11 respectively. Essentially, our evaluation shows these bounds are tight up to constant factors for the three mechanisms considered in Section 4. We also observe that the phase transition observed in Section 4 for the subsampled Gaussian mechanism in the low privacy regime occurs earlier in the high privacy regime. This is consistent with the claim that the phase transition occurs whenever $\gamma\alpha e^{\epsilon(\alpha)} \approx 1$, since $\epsilon(\alpha)$ is inversely proportional to σ^2 in this case. Note also that our finite sample bound (blue curve) matches the lower bound up to a multiplicative constant throughout the considered regimes. For subsampled Gaussian mechanism in Figures 3a and 3d, the RDP parameter matches up to an (not visible in log scale) additive factor for large α. The RDP parameter for subsampled Laplace and subsampled randomized response (in the second and third columns) are both linear in α at the beginning, then they flatten as $\epsilon(\alpha)$ approaches $\epsilon(\infty)$, as we can clearly see in plots Plots 3e and 3f.



(a) Subsampled Gaussian with $\sigma = 5$. (b) Subsampled Laplace with $b = 2$. (c) Subsampled Rand. Resp. with $p = 0.6$.

(d) Subsampled Gaussian with $\sigma = 0.5$. (e) Subsampled Laplace with $b = 0.5$. (f) Subsampled Rand. Resp. with $p = 0.9$

Figure 2: Composition of (ϵ, δ)-DP over 600,000 data accesses with three different subsampled mechanisms. We plot ϵ as a function of the number of rounds of composition k with $\delta = 10^{-8}$ (note that smaller ϵ is better). The top row illustrates the high privacy regime where the base mechanism has $\epsilon \approx 0.5$, while the bottom row shows the low privacy regime with $\epsilon \approx 2$. We consider two baselines: the naïve composition that simply adds up (ϵ, δ) and the strong composition through the result of (Kairouz et al., 2015) with an optimal choice of "per-round δ" computed for every k. The blue curve is based on the composition applied to the RDP upper bound obtained through Theorem 9, and the red dashed curve is based on the composition applied to the lower bound on RDP obtained through Proposition 11. For the Gaussian case, we also present the curves based on applying the composition on the RDP bound obtained through the Gaussian approximation idea explained in Appendix C.7.

B Composition of Differentially Private Mechanisms

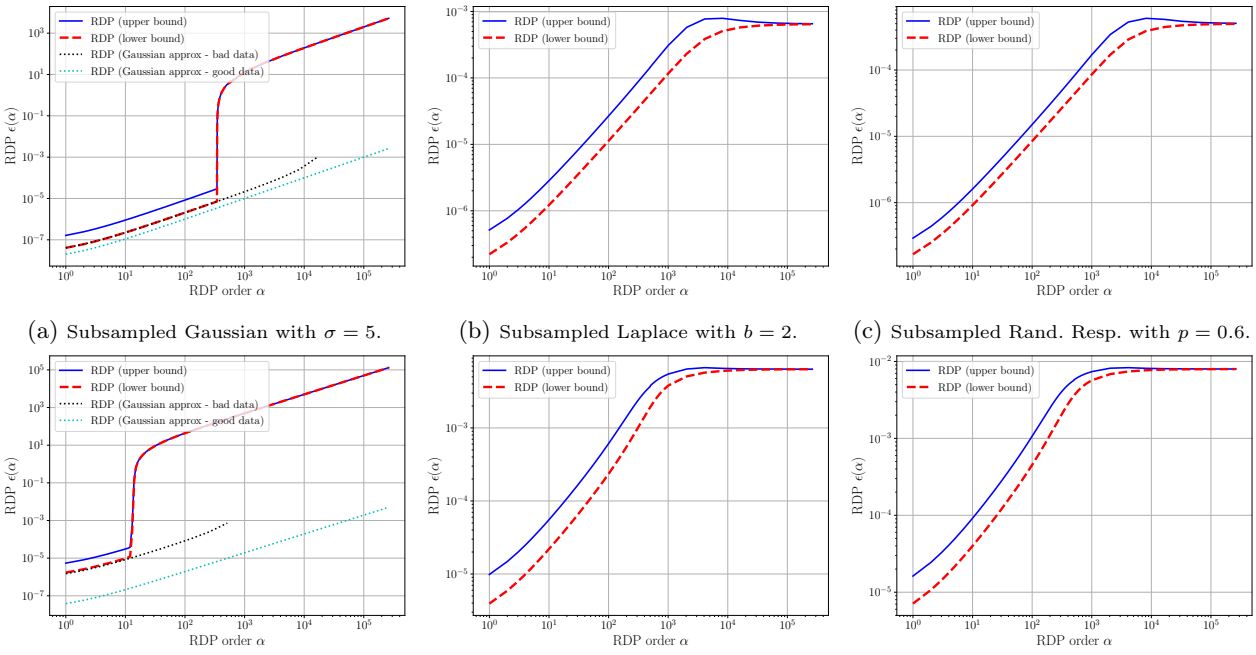
Composition theorems for differential privacy allow a modular design of privacy preserving mechanisms based on mechanisms for simpler sub tasks:

Theorem 12 (Naïve composition, Dwork et al. (2006a)). *A mechanism that permits k adaptive interactions with mechanisms that preserves (ϵ, δ)-differential privacy (and does not access the database otherwise) ensures $(k\epsilon, k\delta)$-differential privacy.*

A stronger composition is also possible as shown by Dwork et al. (2010).

Theorem 13 (Strong composition, Dwork et al. (2010)). *Let $\epsilon, \delta, \delta^* > 0$ and $\epsilon \leq 1$. A mechanism that permits k adaptive interactions with mechanisms that preserves (ϵ, δ)-differential privacy ensures $(\epsilon\sqrt{2k\ln(1/\delta^*)} + 2k\epsilon^2, k\delta + \delta^*)$-differential privacy.*

Kairouz et al. (2015) recently gave an optimal composition theorem for differential privacy, which provides an exact characterization of the best privacy parameters that can be guaranteed when composing a number of (ϵ, δ)-differentially private mechanisms. Unfortunately, the resulting optimal composition bound is quite complex to state exactly, and indeed is even #P-complete to compute exactly when composing mechanisms with different (ϵ_i, δ_i) parameters (Murtagh and Vadhan, 2016).



(a) Subsampled Gaussian with $\sigma = 5$. (b) Subsampled Laplace with $b = 2$. (c) Subsampled Rand. Resp. with $p = 0.6$.

(d) Subsampled Gaussian with $\sigma = 0.5$. (e) Subsampled Laplace with $b = 0.5$. (f) Subsampled Rand. Resp. with $p = 0.9$

Figure 3: The RDP parameter ($\epsilon(\alpha)$) of the three subsampled mechanisms as a function of order α, with subsampling rate $\gamma = 0.001$ in all the experiments. The top row illustrates the case where the base mechanism $\mathcal{M}$ (before amplification using subsampling) is in a relatively high privacy regime (with $\epsilon \approx 0.5$) and the bottom row shows the low privacy regime with $\epsilon \approx 2$. RDP upper bound obtained through Theorem 9 is represented as the blue curve, and the corresponding lower bound obtained through Proposition 11 is represented as the red dashed curve. For the Gaussian case, we also present the RDP bound obtained through the Gaussian approximation idea explained in Appendix C.7.

C Missing Details from Section 3.1

In this section, we fill in the missing details and proofs from Section 3.1. We first define a few quantities needed to establish our results.

Pearson-Vajda Divergence and the Moments of Linearized Privacy Random Variable. The Pearson-Vajda Divergence (or $|\chi|^\alpha$-divergence) of order α is defined as follows (Vajda, 1973):

$$D_{|\chi|^\alpha}(p\|q) := \mathbb{E}_q\left[\left|\frac{p}{q} - 1\right|^\alpha\right]. \tag{4}$$

This is closely related to the moment of the privacy random variable in that $(p/q - 1)$ is the linearized version of $\log(p/q)$. More interestingly, the αth moment of the privacy random variable is the αth derivate of the MGF evaluated at 0:

$$\mathbb{E}[\log(p/q)^\alpha] = \frac{\partial^\alpha}{\partial t^\alpha}[e^{K_{\mathcal{M}}(t)}](0),$$

while at least for the even order, the $|\chi|^\alpha$-divergence is the αth order *forward finite difference* of the MGF evaluated at 0:

$$\mathbb{E}[(p/q - 1)^\alpha] = \Delta^{(\alpha)}[e^{K_{\mathcal{M}}(\cdot)}](0). \tag{5}$$

In the above expression, the αth order *forward difference operator* $\Delta^{(\alpha)}$ is defined recursively with

$$\Delta^{(\alpha)} := \underbrace{\Delta \circ \ldots \circ \Delta}_{\alpha\text{-times}}, \tag{6}$$

where Δ denote the first order forward difference operator such that $\Delta[f](x) = f(x+1) - f(x)$ for any function $f : \mathbb{R} \to \mathbb{R}$. See Appendix D for more information on $\Delta^{(\alpha)}$ and its connection to binomial numbers.

C.1 A Sketch of the Proof of Theorem 9

In this section, we present a sketch of the proof of our main theorem. The arguments are divided into three parts. In the first part, we define a new family of privacy definitions called *ternary-$|\chi|^\alpha$-differential privacy* and show that it handles subsampling naturally. In the second part, we bound the Rényi DP using the ternary-$|\chi|^\alpha$-differential privacy and apply their subsampling lemma. In the third part, we propose several different ways of converting the expression stated as ternary-$|\chi|^\alpha$-differential privacy back to that of RDP, hence giving rise to the stated results in the remarks following Theorem 9.

Part 1: Ternary-$|\chi|^\alpha$-divergence and Natural Subsampling. Ternary-$|\chi|^\alpha$-divergence is a novel quantity that measures the discrepancy of three distributions instead of two. Let p, q, r be three probability distributions[10], we define

$$D_{|\chi|^\alpha}(p, q\|r) := \mathbb{E}_r\left[\left|\frac{p-q}{r}\right|^\alpha\right].$$

Using, this ternary-$|\chi|^\alpha$-divergence notion, we define ζ-ternary-$|\chi|^\alpha$-differential privacy as follows. Analogously with RDP where we considered ϵ as a function of α, we consider ζ as a function of α.

Definition 14 (Ternary-$|\chi|^\alpha$-differential privacy)**.** *We say that a randomized mechanism $\mathcal{M}$ is ζ-ternary-$|\chi|^\alpha$-DP if for all $\alpha \geq 1$:*

$$\sup_{X,X',X'' \text{ mutually adjacent}} \left(D_{|\chi|^\alpha}(\mathcal{M}(X), \mathcal{M}(X')\|\mathcal{M}(X''))\right)^{1/\alpha} \leq \zeta(\alpha).$$

Here, the *mutually adjacent* condition means $d(X, X'), d(X', X''), d(X, X'') \leq 1$, and $\zeta(\alpha)$ is a function from $\mathbb{R}^+$ to $\mathbb{R}^+$. Note that the above definition is a general case of the following binary-$|\chi|^\alpha$-differential privacy definition that works with the standard Person-Vajda $|\chi|^\alpha$-divergences (as defined in (4)).

Definition 15 (Binary-$|\chi|^\alpha$-differential privacy)**.** *We say that a randomized mechanism $\mathcal{M}$ is ξ-binary-$|\chi|^\alpha$-DP if for all $\alpha \geq 1$:*

$$\sup_{X,X':d(X,X')\leq 1} \left(D_{|\chi|^\alpha}(\mathcal{M}(X)\|\mathcal{M}(X'))\right)^{1/\alpha} \leq \xi(\alpha).$$

Again, $\xi(\alpha)$ is a function from $\mathbb{R}^+$ to $\mathbb{R}^+$.

As we described earlier, this notion of privacy shares many features of RDP and could have independent interest. It subsumes $(\epsilon, 0)$-DP (for $\alpha \to \infty$) and implies an entire family of $(\epsilon(\delta), \delta)$-DP through Markov's inequality. We provide additional details on this point in Appendix F.

For our ternary-$|\chi|^\alpha$-differential privacy, what makes it stand out relative to Rényi DP is how it allows privacy amplification to occur in an extremely clean fashion, as the following proposition states:

Proposition 16 (Subsampling Lemma for Ternary-$|\chi|^\alpha$-DP)**.** *Let a mechanism $\mathcal{M}$ obey ζ-ternary-$|\chi|^\alpha$-DP, then the algorithm $\mathcal{M} \circ \mathsf{subsample}$ obeys $\gamma\zeta$-ternary-$|\chi|^\alpha$-DP.*

The entire proof is presented in Appendix C.3. The key idea involves using conditioning on subsampling events, constructing dummy random variables to match up each of these events, and the use of Jensen's inequality to convert the intractable ternary-$|\chi|^\alpha$-DP of a mixture distribution to that of three simple distributions that come from mutually adjacent datasets.

Part 2: Bounding RDP with Ternary-$|\chi|^\alpha$-DP. We will now show that (a transformation of) the quantity of interest — RDP of the subsampled mechanism — can be expressed as a linear combination of a sequence of binary-$|\chi|^\alpha$-DP parameters $\xi(\alpha)$ for integer $\alpha = 2, 3, ...$ through Newton's series expansion of the moment generating function:

$$\mathbb{E}_q\left[\left(\frac{p}{q}\right)^\alpha\right] = 1 + \binom{\alpha}{1}\mathbb{E}_q\left[\frac{p}{q} - 1\right] + \sum_{j=2}^{\alpha}\binom{\alpha}{j}\mathbb{E}_q\left[\left(\frac{p}{q} - 1\right)^j\right]. \tag{7}$$

Observe that $\mathbb{E}_q\left[\frac{p}{q} - 1\right] = 0$, so it suffices to bound $\mathbb{E}_q\left[\left(\frac{p}{q} - 1\right)^j\right]$ for $j \geq 2$.

[10] We think of p, q, r as the distributions $\mathcal{M} \circ \mathsf{subsample}(X), \mathcal{M} \circ \mathsf{subsample}(X'), \mathcal{M} \circ \mathsf{subsample}(X'')$, respectively, for mutually adjacent datasets X, X', X''.

Note that $\frac{p}{q} - 1$ is a special case of $(p-q)/r$ with $q = r$, therefore,

$$\max_{p,q} \mathbb{E}_q\left[\left(\frac{p-q}{q}\right)^j\right] \le \max_{p,q,r} \mathbb{E}_r\left[\left(\frac{p-q}{r}\right)^j\right] \le \max_{p,q,r} D_{|\chi|^j}(p,q\|r).$$

The same holds if we write $\mathcal{M}' = \mathcal{M} \circ \mathsf{subsample}$ and restrict the maximum on the left to $p = \mathcal{M}'(X)$ and $q = \mathcal{M}'(X')$ with X, X' adjacent, and the maximum on the right to $p = \mathcal{M}'(X)$, $q = \mathcal{M}'(X')$ and $r = \mathcal{M}'(X')$ with mutually adjacent X, X' and X''. For the subsampled mechanism, the right-hand side of the above equation can be bounded by Proposition 16. Putting these together, we can bound (7) as

$$\mathbb{E}_q\left[\left(\frac{p}{q}\right)^\alpha\right] \le 1 + \sum_{j=2}^{\alpha}\binom{\alpha}{j}\gamma^j\zeta(j)^j,$$

where mechanism $\mathcal{M}$ satisfies ζ-ternary-$|\chi|^\alpha$-DP and p, q denote the distributions $\mathcal{M} \circ \mathsf{subsample}(X), \mathcal{M} \circ \mathsf{subsample}(X')$, respectively, for adjacent datasets X, X'. Using this result along with the definition of Rényi differential privacy (from Definition 4) implies the RDP parameter following bound,

$$\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha) \le \frac{1}{\alpha-1}\log\left(1 + \sum_{j=2}^{\alpha}\binom{\alpha}{j}\gamma^j\zeta(j)^j\right), \tag{8}$$

Part 3: Bounding Ternary-$|\chi|^\alpha$-DP using RDP. It remains to bound $\zeta(j)^j := \sup_{p,q,r} \mathbb{E}_r[\frac{|p-q|^j}{r^j}]$ using RDP. We provide several ways of doing so and plugging them into (8) show how the various terms in the bound of Theorem 9 arise. Missing proofs are presented in Appendix C.4.

(a) The $4(e^{\epsilon(2)} - 1)$ Term. To begin with, we show that the binary-$|\chi|^\alpha$-DP and ternary-$|\chi|^\alpha$-DP are equivalent up to a constant of 4.

Lemma 17. *If a randomized mechanism $\mathcal{M}$ is ξ-binary-$|\chi|^\alpha$-DP, then it is ζ-ternary-$|\chi|^\alpha$-DP for some ζ satisfying $\xi(\alpha)^\alpha \le \zeta(\alpha)^\alpha \le 4\xi(\alpha)^\alpha$.*

For the special case of $j = 2$, we have

$$\mathbb{E}_q[|p/q - 1|^2] = \mathbb{E}_q[(p/q)^2] - 2\mathbb{E}_q[p/q] + 1 = e^{\epsilon(2)} - 1.$$

Using the bound from Lemma 17 relating the binary and ternary-$|\chi|^\alpha$-DP, gives that $\zeta(2) \le 4(e^{\epsilon(2)} - 1)$.

(b) The $e^{(j-1)\epsilon(j)}\min\{2, (e^{\epsilon(\infty)} - 1)^j\}$ Term. Now, we provide a bound for $j \ge 2$. We start with the following simple lemma.

Lemma 18. *Let X, Y be nonnegative random variables, for any $j \ge 1$*

$$\mathbb{E}[|X - Y|^j] \le \mathbb{E}[X^j] + \mathbb{E}[Y^j].$$

This "triangular inequality"-like result exploits the nonnegativity of X, Y and captures the intrinsic cancellations of the 2^j terms of a Binomial expansion. If we do not have non-negativity, the standard expansion will have a 2^j factor rather than 2 (see e.g., Proposition 3.2 of Bobkov et al. (2016)).
An alternative bound that is tighter in cases when X and Y is related to each other with a multiplicative bound. Note that this bound is only going to be useful when $\mathcal{M}$ has a bounded $\epsilon(\infty)$, such as when $\mathcal{M}$ satisfies $(\epsilon, 0)$-DP guarantee.

Lemma 19. *Let X, Y be nonnegative random variables and with probability 1, $e^{-\varepsilon}Y \le X \le e^{\varepsilon}Y$. Then for any $j \ge 1$*

$$\mathbb{E}[|X - Y|^j] \le \mathbb{E}[Y^j](e^{\varepsilon} - 1)^j.$$

Take $X = p/r$ and $Y = q/r$. Applying Lemma 18 gives $\zeta(j) \le 2e^{(j-1)\epsilon(j)}$. Using Lemma 19 instead with $\varepsilon = \epsilon(\infty)$ provided by the mechanism $\mathcal{M}$, we have $\zeta(j) \le e^{(j-1)\epsilon(j)}(e^{\epsilon(\infty)} - 1)^j$. Using these bounds together, we get the overall bound of,

$$\zeta(j) \le e^{(j-1)\epsilon(j)}\min\{2, (e^{\epsilon(\infty)} - 1)^j\}.$$

Note that at $j = 2$, $e^{(j-1)\epsilon(j)}\min\{2, (e^{\epsilon(\infty)} - 1)^j\}$ simplifies to $e^{\epsilon(2)}\min\{2, (e^{\epsilon(\infty)} - 1)^2\}$.

C.2 Improving the Bound in Theorem 9

We note that we can improve the bound in Theorem 9 under some additional assumptions on the RDP guarantee. We formalize this idea in this section. We use $d(X, X') \leq 1$ to represent neighboring datasets. We start with some additional conditions on the mechanism $\mathcal{M}$ as defined below.

Definition 20 (Tightness and Self-consistency). *We say a mechanism $\mathcal{M}$ and its corresponding RDP privacy guarantee $\epsilon_{\mathcal{M}}(\cdot)$ are* tight *if $\max_{X,X':d(X,X')\leq 1} D_\ell(\mathcal{M}(X)\|\mathcal{M}(X')) = \epsilon_{\mathcal{M}}(\ell)$ for every $\ell = 1, 2, 3, \ldots$ We say that a tight pair $(\mathcal{M}, \epsilon_{\mathcal{M}}(\cdot))$ is* self-consistent *with respect to $|\chi|^\alpha$-divergence, if*

$$\left(\cap_{\ell=1,2,\ldots,\alpha} \operatorname*{argmax}_{X,X':d(X,X')\leq 1} D_\ell(\mathcal{M}(X)\|\mathcal{M}(X'))\right) \cap \operatorname*{argmax}_{X,X':d(X,X')\leq 1} D_{|\chi|^\alpha}(\mathcal{M}(X)\|\mathcal{M}(X')) \neq \emptyset.$$

The tightness condition requires that the RDP function $\epsilon_{\mathcal{M}}(\cdot)$ to be attainable by two distributions induced by a pair of adjacent datasets and the self-consistency condition requires that *the same* pair of distributions attains the maximal $|\chi|^\alpha$-divergence for a given range of parameters. Self-consistency is a non-trivial condition in general but it is true in most popular cases such as the Gaussian mechanism, Laplace mechanism, etc., where we know the Rényi divergence analytically and the difference of two datasets are characterized by one numerical number, e.g., sensitivity. (See Appendix E for a discussion.)

Define,

$$B(\epsilon, l) := \Delta^{(l)}\left[e^{(\cdot-1)\epsilon(\cdot)}\right](0) = \sum_{i=0}^{l}(-1)^i\binom{l}{i}e^{(i-1)\epsilon(i)},$$

as the lth order forward finite difference (see (6)) of the functional $e^{(\cdot-1)\epsilon(\cdot)}$ evaluated at 0.

Theorem 21 (Tighter RDP Parameter Bounds). *Given a dataset of n points drawn from a domain $\mathcal{X}$ and a (randomized) mechanism $\mathcal{M}$ that takes an input from $\mathcal{X}^m$ for $m \leq n$, let the randomized algorithm $\mathcal{M} \circ$ subsample be defined as: (1)* subsample*: subsample without replacement m datapoints of the dataset (sampling parameter $\gamma = m/n$), and (2) apply $\mathcal{M}$: a randomized algorithm taking the subsampled dataset as the input. If $\mathcal{M}$ obeys $(\alpha, \epsilon(\alpha))$-RDP and additionally the RDP guarantee is tight and $(\alpha+1)$-self-consistent as per Definition 20, then for all integer $\alpha \geq 2$, this new randomized algorithm $\mathcal{M} \circ$ subsample obeys $(\alpha, \epsilon'(\alpha))$-RDP where,*

$$\epsilon'(\alpha) \leq \frac{1}{\alpha-1}\log\left(1 + \gamma^2\binom{\alpha}{2}\min\left\{4(e^{\epsilon(2)}-1), e^{\epsilon(2)}\min\{2, (e^{\epsilon(\infty)}-1)^2\}\right\} + 4\sum_{j=3}^{\alpha}\gamma^j\binom{\alpha}{j}\sqrt{B(\epsilon, 2\lfloor j/2\rfloor))\cdot B(\epsilon, 2\lceil j/2\rceil)}\right).$$

Proof Idea. The proof is identical to that of Theorem 9 as laid out in Appendix C.1. The part where it differs is in Part 3, i.e., bounding $\zeta(j)^j$ using RDP. As a result of the assumptions in Definition 20, we know that there exist a pair of adjacent data sets, which give rise to a pair of distribution p and q, that simultaneously achieves the upper bound in the definition of both $\xi(j)$ and $\epsilon(j)$ divergences for all j of interest. For even j, the χ^j-divergence can be written in an analytical form as a Rényi divergence (Nielsen and Nock, 2014) using a binomial expansion. Using Lemma 17 along with this expansion, gives rise to the $4\Delta^{(j)}[e^{(\cdot-1)\epsilon(\cdot)}](0) = 4B(\epsilon, j)$ bound for even j. For odd j, we reduce it to the even j case through the Cauchy-Schwartz inequality

$$\mathbb{E}_q[|p/q-1|^j] = \mathbb{E}_q[|p/q-1|^{(j-1)/2}|p/q-1|^{(j+1)/2}] \leq \sqrt{\mathbb{E}_q[(p/q-1)^{j-1}]\mathbb{E}_q[(p/q-1)^{j+1}]},$$

where each of the term in the square root can now be bounded by the binomial expansion. Putting these together, one notices that one can replace $e^{(j-1)\epsilon(j)}\min\{2, (e^{\epsilon(\infty)}-1)^j\}$ with a more exact evaluation given by $4\sqrt{B(\epsilon, 2\lfloor j/2\rfloor))\cdot B(\epsilon, 2\lceil j/2\rceil)}$ in the bound of Theorem 9. We use this bound only for $j \geq 3$ because for $j = 2$, as discussed in Appendix C.1, we have an alternative way of bounding $\zeta(2)$ that does not require these additional assumptions.

C.3 Proof of the Subsampling Lemma for Ternary-$|\chi|^\alpha$-DP

In this section, we prove Proposition 16. The proof uses the following simple lemma.

Lemma 22. *Bivariate function $f(x,y) = x^j/y^{j-1}$ is jointly convex on $\mathbb{R}^2_+$ for $j > 1$.*

Proof. Note that the function is continuously differentiable on $\mathbb{R}^2_+$. The two eigenvalues of the Hessian matrix

$$0 \quad \text{and} \quad (j^2-j)\frac{x^j}{y^{j+1}}\left(1+\frac{y^2}{x^2}\right)$$

and both are nonnegative in the first quadrant. □

Proposition 23 (Proposition 16 Restated)**.** *Let a mechanism $\mathcal{M}$ obey ζ-ternary-$|\chi|^\alpha$-DP, then the algorithm $\mathcal{M} \circ \mathsf{subsample}$ obeys $\gamma\zeta$-ternary-$|\chi|^\alpha$-DP.*

Proof. If three datasets X, X', X'' of size n are mutually adjacent, they must differ on the same data point (w.l.o.g., let it be the nth), and the remaining $n-1$ data points are the same. Let p, q, r denote the distributions $\mathcal{M} \circ \mathsf{subsample}(X), \mathcal{M} \circ \mathsf{subsample}(X'), \mathcal{M} \circ \mathsf{subsample}(X'')$, respectively.

Let E be the event such that the subsample includes the nth item (and E^c be complement event), we have

$$p = \gamma p(\cdot|E) + (1-\gamma)p(\cdot|E^c)$$
$$q = \gamma q(\cdot|E) + (1-\gamma)q(\cdot|E^c).$$

and by construction, $p(\cdot|E^c) = q(\cdot|E^c)$.

Substituting the observation into the ternary-$|\chi|^j$-divergence, we get γ^j to show up.

$$\begin{aligned} D_{|\chi|^j}(p,q\|r) &= \mathbb{E}_r\left[\left(\frac{|p-q|}{r}\right)^j\right] = \gamma^j \mathbb{E}_r\left[\left(\frac{|p(\cdot|E) - q(\cdot|E)|}{r}\right)^j\right] \\ &= \gamma^j D_{|\chi|^j}(p(\cdot|E), q(\cdot|E)\|r). \end{aligned} \tag{9}$$

Note that $p(\cdot|E), q(\cdot|E)$ and r are mixture distributions with combinatorially many mixing components.

Let J be a random subset of size γn chosen by the $\mathsf{subsample}$ operator. In addition, we define an auxiliary dummy variable $i \sim \text{Unif}(1, ..., \gamma n)$. Let i be independent to everything else, so it is clear that $r(\theta|J) = r(\theta|J,i)$. In other words,

$$r(\theta) = \mathbb{E}_{J,i}[q(\theta|J,i)] = \frac{1}{\gamma n \binom{n}{\gamma n}} \sum_{J\subset[n], i\in[\gamma n]} r(\theta|J).$$

Now, define functions g and g' on index set J, i such that:

$$g(J,i) = \begin{cases} p(\theta|J) & \text{if } n \in J \\ p(\theta|J\cup\{n\}\backslash J[i]) & \text{otherwise,} \end{cases} \qquad g'(J,i) = \begin{cases} q(\theta|J) & \text{if } n \in J \\ q(\theta|J\cup\{n\}\backslash J[i]) & \text{otherwise.} \end{cases}$$

Check that $p(\theta|E) = \mathbb{E}_{J,i} g(J,i)$ and $q(\theta|E) = \mathbb{E}_{J,i} g'(J,i)$.

The above definitions and the introduction of the dummy random variable i may seem mysterious. Let us explain the rationale behind them. Note that mixture distributions $p(\theta|E), q(\theta|E)$ have a different number of mixture components comparing to $q(\theta)$. $q(\theta)$ has $\binom{n}{\gamma n}$ components while $p(\theta|E)$ and $q(\theta|E)$ only have $\binom{n-1}{\gamma n-1}$ components due to the conditioning on the event E that fixes the differing (say the nth) datapoint in the sampled set.

The dummy random variable i allows us to define a new σ-field to redundantly represent both subsampling over $[n-1]$ and $[n]$ under the same uniform probability measure while establishing a one-to-one mapping between pairs of events such that the corresponding index of the subsample differs by only one datapoint.

This trick allows us to write:

$$\begin{aligned}\mathbb{E}_q\left(\frac{|p(\theta|E)-q(\theta|E)|}{q(\theta)}\right)^j &= \int \frac{[p(\theta|E)-q(\theta|E)]^j}{q(\theta)^{j-1}}d\theta \\ &\underset{\text{Jensen}}{\uparrow} \leq \int \mathbb{E}_{J,i}\left[\frac{|g(J,i)-g'(J,i)|^j}{q(\theta|J)^{j-1}}\right]d\theta \\ &\underset{\text{Fubini}}{\uparrow} = \mathbb{E}_{J,i}\mathbb{E}_q\left[\left(\frac{|g(J,i)-g'(J,i)|}{q(\theta|J)}\right)^j \,\middle|\, J,i\right] \leq \zeta(j)^j. \end{aligned} \tag{10}$$

The second but last line uses Jensen's inequality and Lemma 22, which proves the joint convexity of function $x^j/y(j-1)$ on $\mathbb{R}_+^2$. In the last line, we exchange the order of the integral, from which we get the expression for the ternary DP directly. Combining (9) with (10) gives the claimed result because the definitions of g and g' ensure that each inner expectation is a ternary Liese–Vajda divergence of the original mechanism on a triple of mutually adjacent datasets. □

C.4 Missing Proofs on Bounding Ternary-$|\chi|^\alpha$-DP using RDP

Lemma 24 (Lemma 17 Restated). *If a randomized mechanism $\mathcal{M}$ is ξ-binary-$|\chi|^\alpha$-DP, then it is ζ-ternary-$|\chi|^\alpha$-DP for some ζ satisfying $\xi(\alpha)^\alpha \leq \zeta(\alpha)^\alpha \leq 4\xi(\alpha)^\alpha$.*

Proof. The first inequality follows trivially by definition. We now prove the second. Let p, q, r be three probability distributions. Consider four events:

$$\{x|p \geq q, q \geq r\}, \{x|p \geq q, q < r\}, \{x|p < q, p \geq r\}, \{x|p < q, p \geq r\}$$

Under the first event $|p-q|^j/r^{j-1} = (p-q)^j/r^{j-1} \leq (p-r)^j/r^{j-1}$. Under the second event $|p-q|^j/r^{j-1} \leq (p-q)^j/q^j$.Similarly, under the third and fourth event, $|p-q|^j/r^{j-1}$ is bounded by $(q-r)^j/r^{j-1}$ and $(q-p)^j/p^{j-1}$ respectively. It then follows that:

$$\begin{aligned}&\mathbb{E}_r[|p-q|^j/r^j] \\ =&\mathbb{E}_r[|p-q|^j/r^j\mathbf{1}_{\{E_1\}}] + \mathbb{E}_r[|p-q|^j/r^j\mathbf{1}_{\{E_2\}}] + \mathbb{E}_r[|p-q|^j/r^j\mathbf{1}_{\{E_3\}}] + \mathbb{E}_r[|p-q|^j/r^j\mathbf{1}_{\{E_4\}}] \\ \leq&\mathbb{E}_r[|p-r|^j/r^j\mathbf{1}_{\{E_1\}}] + \mathbb{E}_q[|p-q|^j/q^j\mathbf{1}_{\{E_2\}}] + \mathbb{E}_r[|q-r|^j/r^j\mathbf{1}_{\{E_3\}}] + \mathbb{E}_p[|q-p|^j/p^j\mathbf{1}_{\{E_4\}}] \\ \leq&D_{|\chi|^j}(p\|r) + D_{|\chi|^j}(p\|q) + D_{|\chi|^j}(q\|r) + D_{|\chi|^j}(q\|p) \leq 4\xi(j).\end{aligned}$$

□

Lemma 25 (Lemma 18 Restated). *Let X, Y be nonnegative random variables, for any $j \geq 1$*

$$\mathbb{E}[|X-Y|^j] \leq \mathbb{E}[X^j] + \mathbb{E}[Y^j].$$

Proof. Using that the $X, Y \geq 0$

$$\begin{aligned}\mathbb{E}[|X-Y|^j] &= \mathbb{E}[(X-Y)^j\mathbf{1}(X \geq Y)] + \mathbb{E}[(X-Y)^j\mathbf{1}(X < Y)] \\ &\leq \mathbb{E}[X^j \cdot \mathbf{1}(X \geq Y)] + \mathbb{E}\left[Y^j \cdot \mathbf{1}(X < Y)\right] \leq \mathbb{E}[X^j] + \mathbb{E}[Y^j]\end{aligned}$$

□

Lemma 26 (Lemma 19 Restated). *Let X, Y be nonnegative random variables and with probability 1, $e^{-\varepsilon}Y \leq X \leq e^\varepsilon Y$. Then for any $j \geq 1$*

$$\mathbb{E}[|X-Y|^j] \leq \mathbb{E}[Y^j](e^\varepsilon - 1)^j$$

Proof. The multiplicative bound implies that: $-Y(1-e^{-\varepsilon}) \leq X - Y \leq Y(e^\varepsilon - 1)$, which gives that with probability 1

$$|X-Y| \leq \max\{e^\varepsilon - 1, 1 - e^{-\varepsilon}\}Y = (e^\varepsilon - 1)Y,$$

and the claimed result follows. □

C.5 Proof of Corollary 10

Corollary 27 (Corollary 10 Restated). *Let $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denotes the floor and ceiling operators*

$$K_{\mathcal{M}}(\lambda) \leq (1-\lambda+\lfloor\lambda\rfloor)K_{\mathcal{M}}(\lfloor\lambda\rfloor) + (\lambda-\lfloor\lambda\rfloor)K_{\mathcal{M}}(\lceil\lambda\rceil).$$

Proof. The result is a simple corollary of the convexity of the CGF. Specifically, take $\lambda_1 = \lfloor\lambda\rfloor$, $\lambda_2 = \lceil\lambda\rceil$ and $v := \lambda - \lfloor\lambda\rfloor$. Note that $\lambda = (1-v)\lfloor\lambda\rfloor + v\lceil\lambda\rceil$. The result follows from the definition of convexity. □

C.6 Proof of the Lower Bound (Proposition 11)

Proposition 28 (Proposition 11 Restated). *Let $\mathcal{M}$ be a randomized algorithm that takes a dataset in $\mathcal{X}^{\gamma n}$ as an input. If $\mathcal{M}$ obeys $(\alpha, \epsilon(\alpha))$-RDP for a function $\epsilon : \mathbb{R}_+ \to \mathbb{R}_+$ and that there exists $x, x' \in \mathcal{X}$ such that $\epsilon(\alpha) = D_\alpha(\mathcal{M}([x,x,...,x,x'])\|\mathcal{M}([x,x,...,x,x]))$ for all integer $\alpha \geq 1$ (e.g., this condition is true for all output perturbation mechanisms for counting queries), then the RDP function ϵ' for $\mathcal{M} \circ \mathsf{subsample}$ obeys that for all integer $\alpha \geq 1$*

$$\epsilon'(\alpha) \geq \frac{\alpha}{\alpha-1}\log(1-\gamma) + \frac{1}{\alpha-1}\log\left(1 + \alpha\frac{\gamma}{1-\gamma} + \sum_{j=2}^{\alpha}\binom{\alpha}{j}\left(\frac{\gamma}{1-\gamma}\right)^j e^{(j-1)\epsilon(j)}\right).$$

Proof. Consider two datasets $X, X' \in \mathcal{X}^n$ where X' contains n data points that are identically x and X is different from X' only in its last data point. By construction, $\mathsf{subsample}(X') \equiv [x,x,...,x]$, $\Pr[\mathsf{subsample}(X) = [x,x,...,x]] = 1-\gamma$ and $\Pr[\mathsf{subsample}(X) = [x,x,...,x,x'] = \gamma$. In other words, $\mathcal{M} \circ \mathsf{subsample}(X') = \mathcal{M}([x,x,...,x]) := p$ and $\mathcal{M} \circ \mathsf{subsample}(X) = (1-\gamma)p + \gamma\mathcal{M}([x,x,...,x,x']) := (1-\gamma)p + \gamma q$. It follows that

$$\begin{aligned}\mathbb{E}_q\left[\left(\frac{(1-\gamma)q+\gamma p}{q}\right)^\alpha\right] &= \mathbb{E}_q\left[\left(1-\gamma+\gamma\frac{p}{q}\right)^\alpha\right] = (1-\gamma)^\alpha\mathbb{E}_q\left[\left(1+\frac{\gamma}{1-\gamma}\frac{p}{q}\right)^\alpha\right] \\ &= (1-\gamma)^\alpha\left(1+\alpha\frac{\gamma}{1-\gamma} + \sum_{j=2}^{\alpha}\binom{\alpha}{j}\left(\frac{\gamma}{1-\gamma}\right)^j\mathbb{E}_q\left[\left(\frac{p}{q}\right)^j\right]\right).\end{aligned}$$

When we take x, x' to be the one in the assumption that attains the RDP $\epsilon(\cdot)$ upper bound, then we can replace $\mathbb{E}_q\left[(p/q)^j\right]$ in the above bound with $e^{(j-1)\epsilon(j)}$ as claimed. □

C.7 Asymptotic Approximation of Rényi Divergence for Subsampled Gaussian Mechanism

In this section, we present an asymptotic upper bound on the Rényi divergence for the subsampled Gaussian mechanism. The results from this section are also used in our numerical experiments detailed in Section 4 and Appendix A.

Let $\mathcal{X}$ denote the input domain. Let $f : \mathcal{X} \to \Theta$ be some statistical query. We consider a subsampled Gaussian mechanism which releases the answers to f by adding Gaussian noise to the mean of a subsampled dataset. In this case, the output θ of the subsampled Gaussian mechanism is a sample from $\mathcal{N}(\mu_J, \sigma^2/|J|^2)$ where μ_J is short for $\mu(X_J) := \frac{1}{|J|}\sum_{i\in J} f(x_i)$ and J is a random subset of size γn. The distribution of J induces a discrete prior distribution of μ_J. Without loss of generality, we assume that $f(x_i) \leq 1/2$, which implies that the global sensitivity of μ is $1/|J|$. By the sampling without replacement version of the central limit theorem[11], $\sqrt{|J|}(\mu(X_J) - \frac{1}{n}\sum_{i=1}^n f(x_i))$ converges in distribution to $\mathcal{N}(0, \frac{1}{n}\sum_{i=1}^n (f(x_i)-\mu(X))^2)$. In other words, the distribution of θ asymptotically converges to

$$\mathcal{N}\left(\frac{1}{n}\sum_{i=1}^n f(x_i), \frac{1}{n|J|}\sum_{i=1}^n (f(x_i)-\mu(X))^2 + \frac{\sigma^2}{|J|^2}\right).$$

This allows us to use the analytical formula of the Rényi divergence between two Gaussians (see Appendix I) as an asymptotic approximation of the Rényi divergence between the more complex mixture distributions. We disclaim

[11] Under boundedness of $f(x_i)$, the regularity conditions holds.

that this is a truly asymptotic approximation and should only be true when $|J|, n \to \infty$ and $\gamma = |J|/n \to 0$, but it is nevertheless interesting as it allows us to understand the dependence of different parameters in the bound. One important observation is that the part of the variance due to the dataset can be either bigger or smaller than that of the added noise, and this could imply a vastly different Rényi divergence. We give examples here of two contrasting situations.

Example 29 (Gaussian approximation - a "bad" data case)**.** *Let $f(x_1) = f(x_2) = ... = f(x_{n-1}) = f(x_n) = -1/2$ for the elements in X', and for X the only difference (from X') is that in X we have $f(x_n) = 1/2$. Then the two asymptotic distributions are $p = \mathcal{N}(-\frac{1}{2} + \frac{1}{n}, \frac{n-1}{n^2|J|} + \frac{\sigma^2}{|J|^2})$ and $q = \mathcal{N}(-\frac{1}{2}, \frac{\sigma^2}{|J|^2})$, and the corresponding Rényi divergence equals*

$$D_\alpha(p\|q) = \begin{cases} +\infty & \text{if } \alpha \geq \frac{\sigma^2}{\gamma}\frac{n}{n-1} + 1, \\ \frac{\alpha\gamma^2}{2\sigma^2}\left(\frac{\alpha^*}{\alpha^*-\alpha}\right) + \frac{1}{2}\log\left(\frac{\alpha^*-1}{\alpha^*}\right) + \frac{1}{2(\alpha-1)}\log(\frac{\alpha^*}{\alpha^*-\alpha}) & \text{otherwise.} \end{cases}$$

Example 30 (Gaussian approximation - a "good" data case)**.** *Let n be an odd number, and let X' be such that $f(x_i) = 1/2$ for $i \leq \lfloor n/2 \rfloor$ and $f(x_i) = -1/2$ otherwise, and for X the only difference (from X') is that in X we have $f(x_n) = 1/2$. The two asymptotic distributions are $p = \mathcal{N}(\frac{1}{2n}, \frac{\sigma^2}{|J|^2} + \frac{1}{4|J|} - \frac{1}{4n^2|J|})$ and $q = \mathcal{N}(-\frac{1}{2n}, \frac{\sigma^2}{|J|^2} + \frac{1}{4|J|} - \frac{1}{4n^2|J|})$, and the corresponding Rényi divergence equals*

$$D_\alpha(p\|q) = \frac{\alpha\gamma^2}{2\sigma^2 + \gamma(n - n^{-1})/2}.$$

The first example (a "bad" data case) is closely related to our construction in the proof of Proposition 11. For $\alpha \ll \sigma^2/\gamma$, the example shows an $O(\alpha\gamma^2/\sigma^2)$ rate, matching our upper bound from Theorem 9 (see Remark "Bound under Additional Assumptions" in Section 3.1) in the small α, large σ regime. The second example corresponds to a "good" data case where the dataset has a variety of different datapoints, and as we can see, the variance of the asymptotic distribution that comes from subsampling the dataset dominates the noise from Gaussian mechanism and the per-instance RDP loss for this particular pair of X and X' can be γn times smaller than the bad case.

D Discrete Difference Operators and Newton's Series Expansion

In this section, we provide more details of the discrete calculus objects that we used in the proof, and also illustrate how the interesting identity (5) comes about.

Discrete Difference Operators. Discrete difference operators are linear operators that transform a function into its discrete derivatives. Let f be a function $\mathbb{R} \to \mathbb{R}$, the first order forward difference operator of f is a function such that

$$\Delta[f](x) = f(x+1) - f(x).$$

The αth order forward difference operator $\Delta^{(\alpha)}$ can be constructed recursively by

$$\Delta^{(\alpha)} = \Delta \circ \Delta^{(\alpha-1)}$$

for all $\alpha = 1, 2, 3, ...$ with $\Delta^{(1)} := \mathrm{Id}$.

The forward difference operators are linear transformation of functions that can be thought of as a convolution (denoted by $\star$) with a linear combination of Dirac-delta functions (δ_{dirac}), which we call filters.

$$\Delta[f] = f \star (\delta_{\text{dirac}}(x-1) - \delta_{\text{dirac}}(x)).$$

From the linear combination point of view, the first order forward difference operator is the linear combination of the (infinite) basis functions of Dirac-delta functions supported on all integers with coefficient sequence $[..., 0, -1, 1, 0, ...]$. This sequence of coefficients uniquely defines the difference operators. For example, when $\alpha = 2$, the coefficients that construct operator $\Delta^{(\alpha)}$ are

$$\ldots, 0, 0, 1, -2, 1, 0, 0 \ldots$$

and when $\alpha = 3$ and $\alpha = 4$, we get

$$\ldots, 0, 0, -1, 3, -3, 1, 0, 0 \ldots$$

and

$$\ldots, 0, 0, 1, -4, 6, -4, 1, 0, 0 \ldots$$

respectively. In general, these convolution operators can be constructed by Pascal's triangle of the αth order, or simply the binomial coefficients with alternating signs.

When computing the bound in Theorem 9 we need to calculate $\Delta^{(\ell)}[f](0)$ for all integer $\ell \leq \alpha$. The recursive definition of the bound above allows us to compute all finite differences up to order α by $O(\alpha^2)$ evaluation of f rather than the naïve direct calculation of $O(\alpha^3)$. In Appendix G we will describe further speed-ups with approximate evaluation.

Newton Series Expansion. Newton series expansion is the discrete analogue of the continuous Taylor series expansion, with all derivatives replaced with discrete difference operators and all monomials replaced with falling factorials.

Consider infinitely differentiable function $f : \mathbb{R} \to \mathbb{R}$. The Taylor series expansion of f at 0 and the Newton series expansion of f at 0 are respectively:

$$f(x) = f(0) + \frac{\partial}{\partial x}[f](0)x + \frac{\partial^2}{\partial x^2}[f](0)\frac{x^2}{2!} + ... + \frac{\partial^k}{\partial x^k}[f](0)\frac{x^k}{k!} + ...$$
$$f(x) = f(0) + \Delta^{(1)}[f](0)x + \Delta^{(2)}[f](0)\frac{x(x-1)}{2!} + ... + \Delta^{(k)}[f](0)\frac{(x)_k}{k!} + ...$$

where $(x)_k$ denotes the falling factorials $x(x-1)(x-2)...(x-k+1)$. For integer x, it is clear that the Newton's series expansion has a finite number of terms.

E On Tightness and Self-consistency Guarantees

When specifying a sequence of RDP guarantees for $\mathcal{M}$ in terms of $\sup_{X,X':d(X,X')\leq 1} D_\alpha(\mathcal{M}(X)\|\mathcal{M}(X')) \leq \epsilon(\alpha)$ it really matters whether $\epsilon(\alpha)$ is the exact analytical form of some underlying pairs of distributions induced by a pair of adjacent datasets X, X' or just a sequence of conservative estimates. If it is the latter, then it is unclear at which α the slacks are bigger and at which α the slacks are smaller. And the sequence of $\epsilon(\cdot)$ might not be realizable by any pairs distributions. For example, if we use a polynomial upper bound of $\epsilon(\cdot)$, we know from the theory of CGF that no distribution have a CGF of polynomial order higher than 2 and the only distribution that has polynomial order exactly two is the Gaussian distribution (Lukacs, 1970).

In this section, we provide an example proof that the analytical Rényi DP bound of the Gaussian mechanisms (defined in Section 2) are self-consistent. Again for simplicity, for the Gaussian mechanism, we assume that the sensitivity of function f is 1.

Lemma 31. *For the Gaussian mechanism, $\epsilon(\alpha) = \alpha/(2\sigma^2)$ is tight and self-consistent.*

Proof. The Gaussian mechanism with variance σ^2 has a tight RDP parameter bound $\epsilon(\alpha) = \frac{\alpha}{2\sigma^2}$ (Gil et al., 2013). This is achieved by the distributions $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(1, \sigma^2)$.

For self-consistency, it suffices to show that the $|\chi|^\alpha$-divergence's maximum for every even α are also achieved by the same pair of distributions. Consider $q = \mathcal{N}(0, \sigma^2)$ and $p = \mathcal{N}(\mu, \sigma^2)$ for $0 \leq \mu \leq 1$

$$D_{|\chi|^\alpha}(p\|q) = \mathbb{E}_q[(p/q - 1)^\alpha] = \mathbb{E}_q[(e^{-\frac{-2x\mu+\mu^2}{2\sigma^2}} - 1)^\alpha] = \Delta^{(\alpha)}[e^{(\ell^2-\ell)\mu^2}](0)$$

Take derivative w.r.t. μ, we get

$$2\mu(\ell^2 - \ell)\Delta^{(\alpha)}[\mathbb{E}_q[e^{(\ell^2-\ell)\mu^2}]](0) \geq 0$$

for $\mu > 0$. In other words, the divergence is monotonically increasing in μ. □

In general, verifying the self-consistency is not straightforward, but since $|\chi|^\alpha$-divergence is a proper f-divergence, it is jointly convex in its arguments. When the set of distributions is a convex polytope, it suffices to check for this condition at all the vertices of the polytope.

F Other Properties of Ternary-$|\chi|^\alpha$-DP

When $\alpha = 1$, both the binary- and ternary-$|\chi|^\alpha$-divergence reduces to the total variation distance. When $\alpha = 2$ the binary-$|\chi|^\alpha$-divergence become the χ^2-distance.

The following lemma shows that we can convert binary-$|\chi|^\alpha$-DP (and therefore, ternary-$|\chi|^\alpha$-DP) to the more standard (ϵ, δ)-DP using the tail bound of a privacy random variable.

Lemma 32 ($|\chi|^\alpha$-differential privacy $\Rightarrow$ (ϵ, δ)-DP)**.** *If an algorithm is ξ-binary-$|\chi|^\alpha$-DP, then it is also $\left(\epsilon, \left(\frac{\xi(\alpha)}{e^\epsilon - 1}\right)^\alpha\right)$-DP for all $\epsilon > 0$ and equivalently, $(\log \xi(\alpha) - 1 + \frac{\log(1/\delta)}{\alpha}, \delta)$ for all $\delta > 0$.*

Proof. By Markov's inequality,

$$\Pr[|p/q - 1| > t] \le \mathbb{E}[|p/q - 1|^\alpha]/t^\alpha = \left(\frac{\xi(\alpha)}{t}\right)^\alpha.$$

The results follows from changing the variable from p/q to $e^{\log(p/q)}$. □

The following lemma shows that we can bound the above by a quantity that depends on the Rényi divergence and the Pearson-Vajda divergence. It also generalizes Lemma 19 that we used in the proof of Theorem 9.

Lemma 33. *Let p, q, r are three distributions. For all conjugate pair $u, v \ge 1$ such that $1/u + 1/v = 1$, and all integer $j \ge 2$ we have that*

$$\mathbb{E}_r\left[\left(\frac{|p-q|}{r}\right)^j\right] \le e^{(j-1)D_{(j-1)v+1}(q\|r)} D_{|\chi|^{ju}}(p\|q)^{1/u}.$$

Proof. The proof is a straightforward application of the Hölder's inequality.

$$\begin{aligned}
\mathbb{E}_r\left[\left(\frac{|p-q|}{r}\right)^j\right] &= \int r\left(\frac{q}{r}\right)^j \left|\frac{p}{q} - 1\right|^j d\theta \underset{\text{Change of measure}}{=} \int q\left(\frac{q}{r}\right)^{j-1} \left|\frac{p}{q} - 1\right|^j d\theta \\
&\underset{\text{Hölder}}{\le} \left(\mathbb{E}_q\left[\left(\frac{q}{r}\right)^{(j-1)v}\right]\right)^{1/v} \left(\mathbb{E}_q\left[\left(\frac{p}{q} - 1\right)^{ju}\right]\right)^{1/u} \\
&= e^{(j-1)D_{(j-1)v+1}(q\|r)} D_{|\chi|^{ju}}(p\|q)^{1/u}.
\end{aligned}$$

□

Remark 34. *When we take $v = \infty$ and $u = 1$, we recover the result from Lemma 19. When we take $u = v = 2$, this guarantees that ju is an even number and the above results becomes*

$$\mathbb{E}_r\left[\left(\frac{|p-q|}{r}\right)^j\right] \le e^{(j-1)D_{2j-1}(q\|r)} \sqrt{\Delta^{(2j)}[e^{(\cdot-1)D_{(\cdot)}(p\|q)}](0)},$$

where $\Delta^{(2j)}$ is the finite difference operator of order $2j$. Note that $e^{(\cdot-1)D_{(\cdot)}(q\|r)}$ can be viewed as the moment generating function of the random variable $\log(p(\theta)/q(\theta))$ induced by $\theta \sim q$. The $2j$th order discrete derivative of the MGF at 0 is $\mathbb{E}_q[(\frac{p}{q} - 1)^{2j}]$, which very nicely mirrors the corresponding $2j$th order continuous derivative of the MGF evaluated at 0, which by the property of an MGF is $\mathbb{E}_q[\log(p/q)^{2j}]$.

G Analytical Moments Accountant and Numerically Stable Computation

In this section, we provide more details on the *analytical moments accountant* that we described briefly in Section 3.2. Recall that the analytical moments accountant is a data structure that one can attach to a dataset to keep track of the privacy loss over a sequence of differentially private data accesses. The data structure caches the CGF of the privacy random variables in symbolic form and permits efficient (ϵ, δ)-DP calculations for any desired δ or ϵ. Here is how it works.

Let $\mathcal{M}_1, \mathcal{M}_2, .., \mathcal{M}_k$ be a sequence of (possibly adaptively chosen) randomized mechanisms that one applies to the dataset and the $K_{\mathcal{M}_1}, ..., K_{\mathcal{M}_k}$ be the corresponding CGF. The analytical moments accountant maintains $K = K_{\mathcal{M}_1} + ... + K_{\mathcal{M}_k}$ in symbolic forms and it can evaluate $K(\lambda)$ at any $\lambda > 0$. The two main usage of the analytical moments accountant are for keeping track of: (a) RDP parameter $\epsilon(\alpha)$ for all α, and (b) $(\epsilon(\delta), \delta)$-DP for all $0 \leq \delta < 1$, for a heterogeneous sequence of adaptively chosen randomized mechanisms. The conversion to RDP is straightforward using the one-to-one relationship between CGF and RDP (see Remark 7) with the exception of RDP at $\alpha = 1$ (Kullback-Leibler privacy) and $\alpha = +\infty$ (pure DP), which we keep track of separately. The conversion to (ϵ, δ)-DP is obtained by solving the univariate optimization problems described in (2) and (3).

We note that our analytical moments accountant is conceptually the same as the moments accountant used by Abadi et al. (2016) and the RDP composition used by Mironov (2017). Both prior work however considered only a predefined discrete list of orders λ (or α's). Our main difference is that, for every mechanism, we keep track of the CGF for all $\lambda \in \mathbb{R}_+$ at the same time.

In the remainder of the section, we will describe specific designs of this data structure and substantiate our claims described earlier in Section 3.2.

Space and Time Complexity for Tracking Mechanisms and for (ϵ, δ)-DP Query. We start by analyzing the space and time complexity of basic operations of this data structure.

Proposition 35. *The analytical moments accountant takes $O(1)$ time to compose a new mechanism. At any point in time after the analytical moments accountant has been declared and in operation, let the total number of unique mechanisms that it has seen so far be L. Then the analytical moments accountant takes $O(L)$ space . The CGF queries (at a given λ) takes time $O(L)$. (ϵ, δ)-DP query to accuracy τ (in terms of absolute difference in the argument $|\lambda - \lambda^*|$) takes time $O(L)$ and $O(L \log(\lambda^*)/\tau)$ CGF evaluation calls respectively, where λ^* is the corresponding minimizer in (2) or (3).*

Proof. We keep track of a dictionary of λ functions where the (key,value)-pair is effectively $(\mathcal{M}, (K_{\mathcal{M}}, c_{\mathcal{M}}))$ where $K_{\mathcal{M}}$ is a function that returns the CGF given any positive input, and $c_{\mathcal{M}}$ is the coefficient denoting how many times $\mathcal{M}$ appeared. This naturally allows $O(1)$ time to add a new mechanism and $O(L)$ space.

Since CGFs composes by simply adding up the functions, the overall CGF is $\sum_{i=1}^{L} c_{\mathcal{M}_i} K_{\mathcal{M}_i}$. Evaluating this function takes L CGF queries. We think of the problems of solving for ϵ given δ and solving for δ given ϵ as zeroth order optimization problem using these queries. These problems are efficiently solvable due to the geometric properties of CGFs that we mention in Section 2 and Appendix H.

When solving for ϵ given δ, we keep doubling the candidate $\lambda_{\max}$ and calculating $\frac{1/\delta + K_{\mathcal{M}}(\lambda_{\max})}{(\lambda_{\max})} - \frac{1/\delta + K_{\mathcal{M}}(\lambda_{\max}-1)}{(\lambda_{\max}-1)}$ until we find that it is positive. This procedure is guaranteed to detect a bounded interval that guarantees to contain λ^* in $O(\log \lambda^*)$ time thanks to the monotonicity of RDP. Then we do bisection to find the optimal λ^*, using the unimodal property of the objective function. Note that $\lambda_{\max} \leq 2\lambda^*$. This ensures that the oracle evaluation complexity to find a τ-optimal solution (i.e., to within accuracy τ) of λ^* is $O(\log(\lambda^*/\tau)$. We can solve for δ given ϵ using the same bisection algorithm with the same time complexity, by using the fact that (3) is a log-convex problem. □

The results are compared to a naïve implementation of the standard moments accountant that keeps track of an array of size $\lambda_{\max}$ and handles $\delta \Rightarrow \epsilon$ queries without regarding the geometry of CGFs. The latter will take $O(\lambda_{\max})$ time and space for tracking a new mechanism, and $O(\lambda_{\max})$ time to find an 1-suboptimal solution. In addition, it does not allow a dynamic choice of $\lambda_{\max}$. The analytical moments accountant described here, despite its simplicity, is an exponential improvement over the naïve version, besides being more flexible and adaptive.

There are still several potential problems. First, the input could be an upper bound which may not be an actual CGF function of any random variable, therefore breaking the computational properties. Secondly, when we need to handle subsampled mechanisms, even just evaluating the RDP bound in Theorem 9 for once at α will cost $O(\alpha^2)$ (therefore $O(\lambda^2)$). Lastly, the quantities in the bound of Theorem 9 could be exponentially large and dealing them naïvely will cause floating point numbers overflow or underflow. We address these problems below.

"Projecting" a CGF Upper Bound into a Feasible Set. Note that an upper bound of the CGF does not necessarily have the standard properties associated with CGF that we note in Appendix H, however, we can "project" it to another valid upper bound using the proposition below so that it satisfies the properties from

Appendix H.

Proposition 36. *Let $\bar{K}_{\mathcal{M}}$ be an upper bound of $K_{\mathcal{M}}$, there is a functional F such that $F[\bar{K}_{\mathcal{M}}] \leq K_{\mathcal{M}}$ and $F[\bar{K}_{\mathcal{M}}]$ obeys that $F[\bar{K}_{\mathcal{M}}]$ is convex, monotonically increasing, evaluates to 0 at 0, and $\frac{1}{\lambda}F[\bar{K}_{\mathcal{M}}](\lambda)$ is monotonically increasing on $\lambda \geq 0$.*

Proof. We prove by constructing such an F explicitly. First define $g := \text{convexhull}(\bar{K}_{\mathcal{M}})$. By definition, g is the pointwise largest convex function that satisfies the given upper bound. Secondly, we find the largest β such that $\beta\lambda \leq g(\lambda),\ \forall\lambda$. Let the smallest λ such that $g(\lambda) = \beta\lambda$ be $\tilde{\lambda}$. Then, we define

$$F[\bar{K}_{\mathcal{M}}](\lambda) = \begin{cases} 0 & \text{when } \lambda \leq 0, \\ \beta\lambda & \text{when } 0 < \lambda \leq \tilde{\lambda}, \\ g(\lambda) & \text{when } \lambda > \tilde{\lambda}. \end{cases}$$

Clearly, this is the largest function that satisfy the shape constraints, and therefore must be an upper bound of the actual true CGF of interest. □

This ensures that if we replace $K_{\mathcal{M}}$ with $F[\bar{K}_{\mathcal{M}}]$ for any upper bound $\bar{K}_{\mathcal{M}}$, the computational properties of (2) and (3) remain unchanged.

Approximate Computation of Theorem 9. The evaluation of the RDP itself for a subsampled mechanism according to our bounds in Theorem 9 could still depend polynomially in α. We resolve this by only calculating the bound exactly up to a reasonable α_{thresh} and then for $\alpha > \alpha_{thresh}$, we use an optimization based-upper bound.

Noting that the expression in Theorem 9 can be written as a log-sum-exp or softmax function of $\alpha + 1$ items, where the jth item corresponds to:

$$\log \binom{\alpha}{j} + j \log \gamma + j \log \zeta(j).$$

Here, $\zeta(j)$ is the smallest of the upper bounds that we have of the ternary $|\mathcal{X}|^j$-privacy of order j using RDP.

For any vector x of length $\alpha + 1$ we can use the following approximation:

$$\max(x) \leq \text{softmax}(x) \leq \max(x) + \log(\alpha).$$

When $\exp(x - \max(x))$ is dominated by a geometric series (which it often is for most mechanism $\mathcal{M}$ of interest), then we can further improve $\log(\alpha)$ by something independent to α.

The $\max(x)$ can be solved efficiently in $O(\log(\alpha))$ time as the function can have at most two local minima. This observation follows from the fact that $\log \zeta(j)$ (or any reasonable upper bound of it) is monotonically increasing, $j \log \gamma$ is monotonically decreasing, and that $\log \binom{\alpha}{j}$ is unimodal. Furthermore, we use the Stirling approximation for $\log \binom{\alpha}{j}$ when α is large.

Numerical Stability in Computing the bound in Theorem 9. Since log-sum-exp is involved, we use the standard numerically stable implementation of the log-sum-exp function via: $\log(\sum_i \exp(x_i)) = \max_j x_j + \log(\sum_i \exp(x_i - \max_j(x_j)))$.

We also run into new challenges. For instance, the $\sum_{\ell=0}^{j} \binom{j}{\ell}(-1)^{j-\ell} e^{(\ell-1)\epsilon(\ell)}$ term involves taking structured differences of very large numbers that ends up being very small. We find that the alternative higher order finite difference operator representation $\Delta^{(j)}[e^{(\cdot-1)\epsilon(\cdot)}](0)$ and a polar representation of real numbers with a sign and log absolute value allows us to avoid floating point number overflow. However, the latter approach still suffers from the problem of error propagation and does not accurately compute the expression for large j.

To the best of our knowledge, the numerical considerations and implementation details of the moments accountant have not been fully investigated before, and accurately computing the closed form expression of χ^j-divergences using Rényi Divergences for large j remains an open problem of independent interest.

H Properties of Cumulant Generating Functions and Rényi Divergence

In this section, we highlight some interesting properties of CGF, which in part enables our analytical moments accountant data structure described in Appendix G.

Lemma 37. *The CGF of a random variable (if finite for $\lambda \in \mathbb{R}$), obeys that:*

(a) It is infinitely differentiable.

(b) $\frac{\partial}{\partial\lambda} K_{\mathcal{M}}(\lambda)$ monotonically increases from the infimum to the supremum of the support of the random variable.

(c) It is convex (and strictly convex for all distributions that is not a single point mass).

(d) $K_{\mathcal{M}}(0) = 0$, e.g., it passes through the origin.

(e) The CGF of a privacy loss random variable further obeys that $K_{\mathcal{M}}(-1) = 0$.

These properties are used in establishing the computational properties of the analytical moments accountant as we have seen before.

We provide a first-principle proof of convexity (c), which is elementary and does not use a variational characterization of the Rényi divergence as in the Corollary 2 of Van Erven and Harremos (2014).

Proof. We use the definition of convex functions. By definition, for all $\lambda \geq 0$, we have

$$K_{\mathcal{M}}(\lambda) = \log \mathbb{E}_p[e^{\lambda \log \frac{p(\theta)}{q(\theta)}}] = \log \mathbb{E}_p \left[\left(\frac{p(\theta)}{q(\theta)} \right)^{\lambda} \right].$$

Let $\lambda_1, \lambda_2 \geq 0$ and $v \in [0,1]$. Take $\lambda = (1-v)\lambda_1 + v\lambda_2)/2$ and apply Hölder's inequality with the exponents being the conjugate pair $1/(1-v)$ and $1/v$:

$$\begin{aligned}
\mathbb{E}_p \left[\left(\frac{p(\theta)}{q(\theta)} \right)^{\lambda} \right] &= \mathbb{E}_p \left[\left(\frac{p(\theta)}{q(\theta)} \right)^{(1-v)\lambda_1 + v\lambda_2} \right] = \mathbb{E}_p \left[\left(\frac{p(\theta)}{q(\theta)} \right)^{(1-v)\lambda_1} \left(\frac{p(\theta)}{q(\theta)} \right)^{v\lambda_2} \right] \\
&\leq \mathbb{E}_p \left[\left(\frac{p(\theta)}{q(\theta)} \right)^{\lambda_1} \right]^{1-v} \mathbb{E}_p \left[\left(\frac{p(\theta)}{q(\theta)} \right)^{\lambda_2} \right]^{v} \\
&= \exp[K_{\mathcal{M}}(\lambda_1)]^{1-v} \exp[K_{\mathcal{M}}(\lambda_2)]^{v}.
\end{aligned}$$

Take logarithm on both sides, we get

$$K_{\mathcal{M}}((1-v)\lambda_1 + v\lambda_2) \leq (1-v)K_{\mathcal{M}}(\lambda_1) + vK_{\mathcal{M}}(\lambda_2)$$

and the proof is complete. □

Corollary 38. *Optimization problem (3) is log-convex. Optimization problem (2) is unimodal / quasi-convex.*

Proof. To see the first claim, check that the logarithm of (3) is the sum of a convex function and an affine function, which is convex. To see the second claim, first observe $1/\lambda$ is monotonically decreasing in $\mathbb{R}_+$. It suffices to show that $\frac{K_{\mathcal{M}}(\lambda)}{\lambda}$ (this is RDP!) is monotonically increasing. Let $\partial K_{\mathcal{M}}(\lambda)$ be a subgradient of $K_{\mathcal{M}}(\lambda)$, we can take the "derivative" of the function

$$\lim_{\delta \to 0} \frac{1}{\delta} \left(\frac{K_{\mathcal{M}}(\lambda+\delta)}{\lambda+\delta} - \frac{K_{\mathcal{M}}(\lambda)}{\lambda} \right) \geq \frac{\partial K_{\mathcal{M}}(\lambda)}{\lambda} - \frac{K_{\mathcal{M}}(\lambda)}{\lambda^2} \geq 0$$

The last inequality follows from the first order condition of a convex function

$$K_{\mathcal{M}}(0) \geq K_{\mathcal{M}}(\lambda) + (0-\lambda) \cdot \partial K_{\mathcal{M}}(\lambda)$$

and that $K_{\mathcal{M}}(0) = 0$. □

The corollary implies that optimization problems defined in (2) and (3) have unique minimizers and they can be solved efficiently using bisection or convex optimization to arbitrary precision even if all we have is (possibly noisy) blackbox access to $K_{\mathcal{M}}(\cdot)$ or its derivative.

I Rényi Divergence of Exponential Family Distributions and RDP

Exponential Family Distributions. Let θ be a random variable whose distribution parameterized by ϕ. It is an exponential family distribution if the probability density function can be written as

$$p(\theta;\phi) = h(\theta)\exp(\eta(\phi)^T T(\theta) - F(\phi)).$$

If we re-parameterize, we can rewrite the exponential family distribution as a *natural* exponential family

$$p(\theta;\eta) = h(\theta)\exp(\eta^T T(\theta) - A(\eta))$$

where the normalization constant A is called the log-partition function.

Rényi Divergence of Two Natural Exponential Family Distributions. Let $\mathcal{S}$ be the natural parameter space, i.e., every $\eta \in \mathcal{S}$ defines a valid distribution. Then for $\eta_1, \eta_2 \in \mathcal{S}$, the Rényi divergence between the two exponential family distribution $p_{\eta_1} := p(\theta;\eta_1)$ and $p_{\eta_2} := p(\theta;\eta_2)$ is:

1. If $\alpha \notin \{0,1\}$ and $\alpha\eta_1 + (1-\alpha)\eta_2 \in \mathcal{S}$,

$$D_\alpha(p_{\eta_1}\|p_{\eta_2}) = \frac{1}{\alpha-1}\log\left(\frac{A(\alpha\eta_1+(1-\alpha)\eta_2)}{A(\eta_1)^\alpha A(\eta_2)^{1-\alpha}}\right).$$

2. If $\alpha \notin \{0,1\}$ and $\alpha\eta_1 + (1-\alpha)\eta_1 \notin \mathcal{S}$,

$$D_\alpha(p_{\eta_1}\|p_{\eta_2}) = +\infty$$

3. If $\alpha = 1$,

$$D_\alpha(p_{\eta_1}\|p_{\eta_2}) = D_{KL}(p_{\eta_1}\|p_{\eta_2}) = (\eta_1-\eta_2)^T\nabla_\eta A(\eta_1) + A(\eta_2) - A(\eta_1),$$

   namely, the Kullback-Liebler divergence of the two distributions and also the Bregman divergence with respect to convex function A.

4. If $\alpha = 0$,

$$D_\alpha(p_{\eta_1}\|p_{\eta_2}) = -\log(\Pr_{\eta_2}[p_{\eta_1} > 0]).$$

For example, the Rényi divergence between multivariate normal distributions $\mathcal{N}(\mu_1,\Sigma_1), \mathcal{N}(\mu_2,\Sigma_2)$ equals (Gil et al., 2013)

$$\begin{aligned}&D_\alpha(\mathcal{N}(\mu_1,\Sigma_1)\|\mathcal{N}(\mu_2,\Sigma_2))\\ =&\begin{cases}+\infty, & \text{if } \Sigma_\alpha := \alpha\Sigma_2 + (1-\alpha)\Sigma_1 \text{ is not positive definite.}\\ \frac{\alpha}{2}(\mu_1-\mu_2)^T\Sigma_\alpha^{-1}(\mu_1-\mu_2) - \frac{1}{2(\alpha-1)}\log\left(\frac{|\Sigma_\alpha|}{|\Sigma_1|^{1-\alpha}|\Sigma_2|^\alpha}\right), & \text{otherwise.}\end{cases}\end{aligned}$$

Exponential Family Mechanisms and its Rényi-DP. Let the differentially private mechanism to release θ be sampling from an exponential family. Let

$$p(\theta) = h(\theta)\exp(\eta(X)^T T(\theta) - A(\eta(X)))$$

denote the distribution induced by this differentially private mechanism on dataset X, and similarly let

$$q(\theta) = h(\theta)\exp(\eta(X')^T T(\theta) - A(\eta(X'))).$$

be the corresponding distribution when the dataset is X'.

In this case, the privacy random variable $\log(p/q)$ has a specific form

$$\varphi(\theta) = [\eta(X) - \eta(X')]^T T(\theta) - [A(\eta(X)) - A(\eta(X'))].$$

Using this, it can be shown that the α-Rényi divergence between p and q is

$$\begin{aligned} D_\alpha(p\|q) &= \log \mathbb{E}_q\left[e^{\alpha\varphi(\theta)}\right]^{\frac{1}{\alpha-1}} \\ &= \frac{1}{\alpha-1}\left[A(\alpha\eta(X)+(1-\alpha)\eta(X')) - \alpha A(\eta(X)) - (1-\alpha)A(\eta(X'))\right]. \end{aligned}$$

A special case of the exponential family mechanisms of particular interest is the posterior sampling mechanisms where $\eta(X)$ has a specific form (Geumlek et al., 2017).

To obtain RDP from the above closed-form Rényi divergence, it remains to maximize over two adjacent data sets X, X'. We make a subset of the following three assumptions.

(A) Bounded parameter difference: $\sup_{X,X':d(X,X')\le 1}\|\eta(X)-\eta(X')\| \le \Delta$ with respect a norm $\|\cdot\|$.

(B) (B,κ)-Local Lipschitz: The log-partition function A is (B,κ)-Local Lipschitz with respect to $\|\cdot\|$ if for all data set X and all η such that $\|\eta-\eta(X)\|\le\kappa$, we have

$$A(\eta) \le A(\eta(X)) + B\|\eta-\eta(X)\|.$$

(C) (L,κ)-Local smoothness: The log-partition function A is (L,κ)-smooth with respect to $\|\cdot\|$ if for all data set X and all η such that $\|\eta-\eta(X)\|\le\kappa$, we have

$$A(\eta) \le A(\eta(X)) + \langle\nabla A(\eta(X)), \eta-\eta(X)\rangle + L\|\eta-\eta(X)\|^2.$$

The following proposition refines the results of (Geumlek et al., 2017, Lemma 3).

Proposition 39 (RDP of exponential family mechanisms)**.** *Let $\mathcal{M}$ is an exponential family mechanism that obeys Assumption (A)(B)(C) with parameter Δ, B, L, κ with a common norm $\|\cdot\|$. If in addition, $\kappa \ge \Delta$, then $\mathcal{M}$ obeys $(\alpha,\epsilon(\alpha))$-RDP for all $\alpha \in (1, \kappa/\Delta+1]$ with*

$$\epsilon(\alpha) \le \min\left\{\frac{\alpha L\Delta^2}{2}, 2B\Delta\right\}.$$

Remark 40. *We can view B and L as (nondecreasing) functions of κ. For any fixed α of interest, we can optimize over all feasible choice of κ:*

$$\epsilon(\alpha) \le \min_{\kappa:\alpha\Delta\le\kappa} \min\left\{\alpha L(\kappa)\Delta^2, 2B(\kappa)\Delta\right\} = \min\left\{\alpha L(\alpha\Delta)\Delta^2, 2B(\alpha\Delta)\Delta\right\}.$$

In fact, as can be seen clearly from the proof, $2B(\alpha\Delta)\Delta$ can be improved to $[B((\alpha-1)\Delta)+B(\Delta)]\Delta$.

Proof of Proposition 39. Assumption (A) implies that $\|\eta(X)-\eta(X')\| \le \Delta$. Note that for all $\alpha \le \kappa/\Delta$, $\|\alpha\eta(X)+(1-\alpha)\eta(X')-\eta(X)\| \le \kappa$. Assumption (B) implies that

$$A(\alpha\eta(X)+(1-\alpha)\eta(X')) \le A(\eta(X)) + (\alpha-1)B\|\eta(X')-\eta(X))\| \le A(\eta(X)+(\alpha-1)B\Delta,$$

and that

$$A(\eta(X')) \le A(\eta(X)) + B\Delta.$$

Substitute these into the definition of $D_\alpha(p\|q)$ we get that

$$D_\alpha(p\|q) \le \frac{1}{\alpha-1}[A(\eta(X)) + (\alpha-1)B\Delta - A(\eta(X)) + (\alpha-1)B\Delta] = 2B\Delta. \tag{11}$$

Assumption (C) implies that for all $\alpha \le \kappa/\Delta + 1$

$$\begin{aligned} &A(\alpha\eta(X)+(1-\alpha)\eta(X')) = A(\eta(X)+(\alpha-1)(\eta(X)-\eta(X'))) \\ \le& A(\eta(X)) + (\alpha-1)\langle\nabla A(\eta(X), \eta(X)-\eta(X'))\rangle + \frac{(\alpha-1)^2L}{2}\|\eta(X)-\eta(X')\|^2 \\ \le& A(\eta(X)) + (\alpha-1)\langle\nabla A(\eta(X), \eta(X)-\eta(X'))\rangle + \frac{(\alpha-1)^2L\Delta^2}{2} \end{aligned}$$

where the last step uses Assumption (A). Assumption (C) also implies that

$$\begin{aligned} A(\eta(X')) - A(\eta(X) &\leq \langle \nabla A(\eta(X), \eta(X') - \eta(X)) \rangle + \frac{L\|\eta(X) - \eta(X')\|^2}{2} \\ &\leq \langle \nabla A(\eta(X), \eta(X) - \eta(X')) \rangle + \frac{L\Delta^2}{2}. \end{aligned}$$

Substitute these into the definition of $D_\alpha(p\|q)$ we get that

$$\begin{aligned} D_\alpha(p\|q) \leq \frac{1}{\alpha - 1}\Big[& A(\eta(X)) + (\alpha-1)\langle \nabla A(\eta(X), \eta(X) - \eta(X')) \rangle + \frac{(\alpha-1)^2 L\Delta^2}{2} \\ & - A(\eta(X)) + (\alpha-1)\langle \nabla A(\eta(X), \eta(X') - \eta(X)) \rangle + \frac{(\alpha-1)L\Delta^2}{2} \Big] = \frac{\alpha L\Delta^2}{2}, \end{aligned}$$

which, together with (11), produces the bound as claimed. □