Poisson Subsampled Renyi Differential Privacy: Supplementary Materials

A. Proof of Theorem 5 and Theorem 8

Recall that we will denote the density of $\mathcal{M} \circ \mathsf{PoissonSample}(X')$ by q and that of $\mathcal{M} \circ \mathsf{PoissonSample}(X)$ by p. Let's first make a few observations.

1. There is a natural change of measure that we can do:

$$\mathbb{E}_q e^{\alpha \log(p/q)} = \mathbb{E}_q[(p/q)^\alpha] = \mathbb{E}_p[(p/q)^{\alpha-1}] = \mathbb{E}_p[e^{(\alpha-1)\log(p/q)}].$$

   This relates RDP to the moment generating function of the log-odds ratio random variable, or the privacy random variable $\log(p/q)$.

2. With our loss of generality, we can assume $X' = X \cup \{x\}$. In order to bound RDP with order α, it suffices to bound the moments $\mathbb{E}_p[(q/p)^\alpha]$ and $\mathbb{E}_q[(p/q)^\alpha]$ then take the bigger of the two bounds.

3. Both p and q are mixture distributions. Let $|X| = n-1$ and $|X'| = n$. p has 2^{n-1} mixture components and q has 2^n mixture components. Each component corresponds to a particular subset of the data set.

4. If we condition on condition on $J = (\sigma_1, ..., \sigma_{n-1}) \in \{0,1\}^{n-1}$, we get

$$\mathbb{E}_p[(q/p)^\alpha] = \int \frac{\left(\sum_J \mathbb{P}(J)\left[(1-\gamma)\mu_0(J) + \gamma\mu_1(J)\right]\right)^\alpha}{(\sum_J \mathbb{P}(J)\mu_0(J))^{\alpha-1}}$$

   By Lemma 23 of (Wang et al., 2019), $f(x,y) = x^\alpha/y^{\alpha-1}$ is jointly convex on $\mathbb{R}_+^2$ for all $\alpha \in (1, +\infty)$, which allows us to apply Jensen's inequality to get

$$\mathbb{E}_p[(q/p)^\alpha] \le \sum_J \mathbb{P}(J)\mathbb{E}_{\mu_0(J)}\left(\frac{(1-\gamma)\mu_0(J) + \gamma\mu_1(J)}{\mu_0(J)}\right)^\alpha$$

   and similarly

$$\mathbb{E}_q[(p/q)^\alpha] \le \sum_J \mathbb{P}(J)\mathbb{E}_{(1-\gamma)\mu_0(J)+\gamma\mu_1(J)}\left(\frac{\mu_0(J)}{(1-\gamma)\mu_0(J) + \gamma\mu_1(J)}\right)^\alpha.$$

   where μ_0 is the distribution of $\mathcal{M}(X_J)$ and μ_1 is the distribution of $\mathcal{M}(X_j \cup \{x\})$. [4]

Denote $\mu_0 := \mu_0(J)$ and $\mu_1 := \mu_1(J)$ as short hands. What matters is that μ_0 and μ_1 are distributions induced by the application of our base mechanism $\mathcal{M}$ to two adjacent data sets.

The fourth observation reduces the problem to bounding $A_1 := \mathbb{E}_{\mu_0}\left[\left(\frac{(1-\gamma)\mu_0+\gamma\mu_1}{\mu_0}\right)^\alpha\right]$ and $A_2 := \mathbb{E}_{(1-\gamma)\mu_0+\gamma\mu_1}\left[\left(\frac{\mu_0}{(1-\gamma)\mu_0+\gamma\mu_1}\right)^\alpha\right]$ using RDP of $\mathcal{M}$.

Let's start with A_1 and consider only the case when $\alpha \ge 1$ is an integer. Also, without loss of generality, we assume $\gamma < 1$.

[4] Note that the arguments used by Abadi et al. (2016) based on the quasi-convexity of Renyi-divergence will give a slightly weaker result but with the expectation over J replaced with the maximum over J, which will be sufficient for our purpose too in this paper.

Let $\alpha \geq 1$ be an integer, and assume $\gamma < 1$. By the Binomial theorem:

$$\begin{aligned}
A_1 =&\mathbb{E}_{\mu_0}\left[\left((1-\gamma)+\gamma\frac{\mu_1}{\mu_0}\right)^{\alpha}\right]\\
=&\sum_{\ell=0}^{\alpha}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^{\ell}\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^{\ell}\\
=&(1-\gamma)^{\alpha-1}(\alpha\gamma-\gamma+1)+\sum_{\ell=2}^{\alpha}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^{\ell}\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^{\ell}\\
\leq&(1-\gamma)^{\alpha-1}(\alpha\gamma-\gamma+1)+\sum_{\ell=2}^{\alpha}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^{\ell}e^{(\ell-1)\epsilon(\ell)}.
\end{aligned}$$

A_2 is tricky as we cannot always calculate it explicitly or approximate efficiently with Renyi-divergence. By a change of measure, we can write A_2 as moments of a negative order.

$$A_2 = \mathbb{E}_{\mu_0}\left[\left(1-\gamma+\gamma\frac{\mu_1}{\mu_0}\right)^{-(\alpha-1)}\right].$$

Trivially, we have two somewhat trivial upper bounds

$$A_2 \leq (1-\gamma)^{-(\alpha-1)}. \tag{5}$$

When $\mathcal{M}$ is ϵ-DP,

$$A_2 \leq (1-\gamma(1-e^{-\epsilon}))^{-(\alpha-1)}.$$

Other than these two, the expression does not seem to give us a more meaningful bound. It might be tempted to use Binomial series expansion (now an infinite series). However, it is not guaranteed to converge for some μ_0, μ_1. Even in cases when it converges, we will have positive and negative terms that we could not construct a tight expression with RDP.

A.1. A novel alternative decomposition.

Let us try to bound $\mathbb{E}_q[(p/q)^{\alpha}]$ through an alternative means. We will redefine the index set $J = (\sigma_1, ..., \sigma_n) \subset \{0,1\}^n$.

Define $q' = \sum_J \mathbb{P}(J)q'(J)$ such that $q'(J) = q((\sigma_1, ..., \sigma_{n-1}, 1))$. Define $p' = \sum_J \mathbb{P}(J)p'(J)$ such that $p'(J) = q((\sigma_1, ..., \sigma_{n-1}, 0))$. Note that $p = p'$, $q = (1-\gamma)p + \gamma q'$ and therefore $p = q + \gamma p' - \gamma q'$.

It follows from Jensen's inequality and the joint convexity that

$$\begin{aligned}
\mathbb{E}_q[(p/q)^{\alpha}] =&\mathbb{E}_q\left[\left(\frac{q+\gamma p'-\gamma q'}{q}\right)^{\alpha}\right] \leq \mathbb{E}_J\mathbb{E}_q\left[\left(\frac{q+\gamma p'-\gamma q'}{q}\right)^{\alpha}\right]\\
\leq&\mathbb{E}_{\sigma_1,...,\sigma_{n-1}}\mathbb{E}_{\sigma_n}\mathbb{E}_{q(J)}\left[\left(\frac{q(J)+\gamma p'(J)-\gamma q'(J)}{q(J)}\right)^{\alpha}\right]\\
=&\mathbb{E}_{\sigma_1,...,\sigma_{n-1}}\Big\{\gamma\mathbb{E}_{q'(J)}\left[\left(\frac{q'(J)+\gamma p'(J)-\gamma q'(J)}{q'(J)}\right)^{\alpha}\Big|\sigma_n=1\right]\\
&+(1-\gamma)\mathbb{E}_{p'(J)}\left[\left(\frac{p'(J)+\gamma p'(J)-\gamma q'(J)}{p'(J)}\right)^{\alpha}\Big|\sigma_n=0\right]\Big\}\\
=&\gamma\mathbb{E}_{\mu_1}\left[\left(\frac{(1-\gamma)\mu_1+\gamma\mu_0}{\mu_1}\right)^{\alpha}\right]+(1-\gamma)\mathbb{E}_{\mu_0}\left[\left(\frac{(1+\gamma)\mu_0-\gamma\mu_1}{\mu_0}\right)^{\alpha}\right] \qquad (6)\\
=&\gamma\mathbb{E}_{\mu_1}\left[\left((1-\gamma)+\gamma\frac{\mu_0}{\mu_1}\right)^{\alpha}\right]+(1-\gamma)\mathbb{E}_{\mu_0}\left[\left(1-\gamma+\gamma(2-\frac{\mu_1}{\mu_0})\right)^{\alpha}\right]\\
=&\sum_{\ell=0}^{\alpha}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^{\ell}\left\{\gamma\mathbb{E}_{\mu_1}\left(\frac{\mu_0}{\mu_1}\right)^{\ell}+(1-\gamma)\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\right\} \qquad (7)
\end{aligned}$$

There are two interesting things about the above chain of derivation. (6) really allows us to evaluate the quantity for any pair of μ_0 and μ_1. However, we cannot really easily upper bound it for all μ_1, μ_2 easily since some of the terms are negative.

Meanwhile, (7) is a slightly prettier form. If we can show that $\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^\ell \le \mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^\ell$, then we are done. In fact, for $\ell = 0, 1, 2$, it is straightforward to show that $\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^\ell = \mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^\ell$. For $\ell \ge 3$, it becomes quite a deep question whether it is true that $\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^\ell \le \mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^\ell$.

Our first attempt establishes that this is related to the sign of Pearson-Vajda pseudo-divergences of odd orders.

Lemma 13. *For any pairs of distribution μ_0, μ_1 such that the Renyi-divergence $D_\alpha(\mu_1, \mu_0)$ exists up to order ℓ.*

$$\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^\ell = \mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^\ell - 2\sum_{j \text{ is odd}, j\le\ell}\binom{\ell}{j}\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}-1\right)^j,$$

where $\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}-1\right)^j$ is the Pearson-Vajda χ^j-pseudo-divergence of μ_1 and μ_2.

Proof. Observe that $2-\mu_1/\mu_0 = 1-(\mu_1/\mu_0-1)$ and that $\mu_1/\mu_0 = 1+(\mu_1/\mu_0-1)$. It follows that

$$\begin{aligned}
&\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^\ell - \mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^\ell \\
&= \sum_{j=0}^{\ell}\binom{\ell}{j}\left\{((-1)^j-1)\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}-1\right)^j\right\} \\
&= -2\sum_{j \text{ is odd}, j\le\ell}\binom{\ell}{j}\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}-1\right)^j
\end{aligned}$$

□

Proof of Theorem 8. Note that Condition (4) implies

$$\sum_{j \text{ is odd}, j\le\ell}\binom{\ell}{j}\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}-1\right)^j \ge 0 \tag{8}$$

as a result, Lemma 13 implies that

$$\mathbb{E}_{\mu_0}\left(2-\frac{\mu_1}{\mu_0}\right)^\ell \le \mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}\right)^\ell \le e^{(\ell-1)\epsilon_{\mathcal{M}}(\ell)}.$$

Substitute the above into (7), then we can obtain a bound identical to the lower bound in the Theorem 6. □

A bigger question is that what if the condition above is not true? Can we obtain a general-purpose bound that applies to all $\mathcal{M}$ without needing to worry about whether Condition (4) is true.

One idea is to directly evaluate $\sum_{j \text{ is odd}, j\le\ell}\binom{\ell}{j}\mathbb{E}_{\mu_0}\left(\frac{\mu_1}{\mu_0}-1\right)^j$ and replace all Renyi-divergences of μ_1, μ_0 with the corresponding RDP bound. This is not really a valid argument. We cannot directly evaluate irwith RDP because it is not straightforward how we can take supremum over μ_1, μ_0 (two neighboring data sets). Substituting the RDP into it is not really correct, because there might be some pair of μ_1, μ_0 that do not match the RDP bound.

Can we still obtain a bound that is quantitatively the same as Theorem 8?

In the following we write two lemmas that allow us to prove such a general purpose bound (Theorem 5).

A.2. Approximation upper bound for $\ell \geq 3$.

Lemma 14 (Relax order of RDP)**.**

$$\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\right] \leq \begin{cases} e^{\ell\epsilon(\ell+1)} & \textit{if } \ell \textit{ is odd} \\ e^{(\ell-1)\epsilon(\ell)} + e^{\ell\epsilon(\ell+1)} & \textit{if } \ell \textit{ is even.} \end{cases}$$

Proof. We consider decomposing the expression to several pieces.

$$\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\right] = \mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}\leq 2)\right] \tag{9}$$

$$+\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}> 2)\right] \tag{10}$$

In the first term, we use the basic inequality that $\frac{\mu_1}{\mu_0}+\frac{\mu_0}{\mu_1}\geq 2$, which implies that

$$0 \leq \mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}\leq 2)\right] \leq \mathbb{E}_{\mu_0}\left[\left(\frac{\mu_0}{\mu_1}\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}\leq 2)\right]$$
$$\leq \mathbb{E}_{\mu_0}\left[\left(\frac{\mu_0}{\mu_1}\right)^{\ell}\right] \leq e^{\ell\epsilon_{\mathcal{M}}(\ell+1)}.$$

The second term is negative if ℓ is an odd number. Moreover, we can bound its absolute value:

$$\left|\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}> 2)\right]\right| = \mathbb{E}_{\mu_0}\left[\left(\frac{\mu_1}{\mu_0}-2\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}> 2)\right]$$
$$\leq \mathbb{E}_{\mu_0}\left[\left(\frac{\mu_1}{\mu_0}\right)^{\ell}\right] \leq e^{(\ell-1)\epsilon_{\mathcal{M}}(\ell)}.$$

In addition, in the case of pure DP with $\epsilon \leq \log(2)$, we have that $\mu_1 \leq e^{\epsilon}\mu_0 \leq 2\mu_0$, which implies that the second term is 0. □

Lemma 15 (Relax multiplicative constant)**.**

$$\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\right] \leq \begin{cases} 2e^{(\ell-1)\epsilon(\ell)} & \textit{if } \ell \textit{ is odd} \\ 3e^{(\ell-1)\epsilon(\ell)} & \textit{if } \ell \textit{ is even.} \end{cases}$$

Proof. We start with the case when ℓ is even.

$$\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\right]$$
$$= \mathbb{E}_{\mu_0}\left[(2-\frac{\mu_1}{\mu_0})\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell-1}\right]$$
$$= 2\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell-1}\right] - \mathbb{E}_{\mu_1}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell-1}\right] \tag{11}$$
$$= 2\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell-1}\right] + \mathbb{E}_{\mu_1}\left[\left(\frac{\mu_1}{\mu_0}-2\right)^{\ell-1}\right]$$

Note that we used the fact that $\ell - 1$ is odd in the last line. By Lemma 14 we can bound the first term by $e^{(\ell-1)\epsilon(\ell)}$. Now by the fact that $x^{\ell-1}$ is a monotonically increasing function, we can bound the second term by $\mathbb{E}_{\mu_1}\left[\left(\frac{\mu_1}{\mu_0}\right)^{\ell-1}\right]$, which also is smaller than $e^{(\ell-1)\epsilon(\ell)}$. That gives us the constant multiplicative factor of 3.

Now consider the case when ℓ is odd. Decompose the expression by (10) and drop the second term since it is negative, we can write

$$\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\right] \le \mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell}\mathbf{1}(\frac{\mu_1}{\mu_0}\le 2)\right].$$

Now apply the same trick as we did to get (11), we can rewrite the above as

$$2\mathbb{E}_{\mu_0}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell-1}\mathbf{1}(\frac{\mu_1}{\mu_0}\le 2)\right] - \mathbb{E}_{\mu_1}\left[\left(2-\frac{\mu_1}{\mu_0}\right)^{\ell-1}\mathbf{1}(\frac{\mu_1}{\mu_0}\le 2)\right].$$

Again note that the second term is negative, and by $\frac{\mu_1}{\mu_0}+\frac{\mu_0}{\mu_1}\ge 2$, we can bound the first term by

$$\begin{aligned}
2\mathbb{E}_{\mu_0}\left[\left(\frac{\mu_0}{\mu_1}\right)^{\ell-1}\mathbf{1}(\frac{\mu_1}{\mu_0}\le 2)\right] &\le 2\mathbb{E}_{\mu_0}\left[\left(\frac{\mu_0}{\mu_1}\right)^{\ell-1}\right] \\
&= 2\mathbb{E}_{\mu_1}\left[\left(\frac{\mu_0}{\mu_1}\right)^{\ell}\right] \\
&\le 2e^{(\ell-1)\epsilon(\ell)}.
\end{aligned}$$

□

Now we are ready to present the main theorem.

Proof of Theorem 5. Substituting the results in Lemma 15 to (7), relax the constant to 3 and then apply the RDP upper bound of the Renyi-divergence. □

B. Tight bounds for Gaussian and Laplace mechanism

In this section, we prove Proposition 10 and also that our tight bound Theorem 8 applies to the Gaussian mechanism and Laplace mechanism. In particular, we will show that the condition (4) in Theorem 8 that requires the Pearson-Vajda χ^α divergences to be nonnegative for the π, μ that come running either the Gaussian mechanism or the Laplace mechanism on any two adjacent data sets.

The proof for the Gaussian mechanism uses a novel inductive argument, while the proof for the Laplace mechanism directly proves that moving $f(X')$ away from $f(X)$ strictly increases the odd-order Pearson-Vajda χ^α divergence using tools from convex optimization.

These calculations are possible because the discrepancy of two data sets can be fully described by a single parameter. The general recipe used in this section can also be applied to other cases where only a small number of parameters can be used to avoid the intractable search over any pair of data sets to find the worst pair.

B.1. Qualifying Gaussian Mechanism

Lemma 16. *For any π, μ that are absolutely continuous, and an odd $\alpha \ge 3$,*

$$E_\mu(\frac{\pi}{\mu})^2(\frac{\pi}{\mu}-1)^{\alpha-2} \ge E_\mu[(\frac{\pi}{\mu}-1)^{\alpha-2}]$$

Proof.

$$\begin{aligned} E_\mu(\frac{\pi}{\mu})^2(\frac{\pi}{\mu}-1)^{\alpha-2} &= E_\pi(\frac{\pi}{\mu})(\frac{\pi}{\mu}-1)^{\alpha-2} \\ &= E_\pi(\frac{\pi}{\mu}-1)^{\alpha-1} + E_\pi(\frac{\pi}{\mu}-1)^{\alpha-2} \\ &\geq E_\pi(\frac{\pi}{\mu}-1)^{\alpha-2} \end{aligned}$$

Since $\alpha-1$ is even, $E_\mu(\frac{\pi}{\mu}-1)^{\alpha-1} \geq 0$. $E_\pi(\frac{\pi}{\mu}-1)^{\alpha-2}$ could be rewritten as $E_\mu(\frac{\pi}{\mu})(\frac{\pi}{\mu}-1)^{\alpha-2}$

$$\begin{aligned} E_\mu(\frac{\pi}{\mu})(\frac{\pi}{\mu}-1)^{\alpha-2} &= E_\mu(\frac{\pi}{\mu}-1)^{\alpha-1} + E_\mu(\frac{\pi}{\mu}-1)^{\alpha-2} \\ &\geq E_\mu[(\frac{\pi}{\mu}-1)^{\alpha-2}] \end{aligned}$$

□

Theorem 17. *Let π, μ be two gaussian distriutions, $\pi \sim \mathcal{N}(\sqrt{t}, 1)$ and $\mu \sim \mathcal{N}(0,1)$, for $\forall t \geq 0, \forall$odd $\alpha \geq 1$, we have $E_\mu(\frac{\pi}{\mu}-1)^\alpha \geq 0$.*

Proof. **Base case**: The statement holds when $\alpha = 1$

$$\forall t, E_\mu(\frac{\pi}{\mu}-1) = 0$$

Inductive step: Show that if $\alpha = \tilde{\alpha}$, we have $E_\mu(\frac{\pi}{\mu}-1)^{\tilde{\alpha}} \geq 0$ for all t, then the statement holds for $\alpha = \tilde{\alpha}+2$. This can be done as follows:
We first write an expansion of $E_\mu(\frac{\pi}{\mu}-1)^\alpha$ as :

$$\begin{aligned} E_\mu(\frac{\pi}{\mu}-1)^\alpha &= \sum_{\ell=0}^{\alpha} \binom{\alpha}{\ell} (-1)^{\alpha-\ell} E_\mu(\frac{\pi}{\mu})^\ell \\ &= \alpha - 1 + \sum_{\ell=2}^{\alpha} \binom{\alpha}{\ell} (-1)^{\alpha-\ell} e^{\frac{\ell(\ell-1)t}{2}} \end{aligned}$$

When $t=0$, $E_\mu(\frac{\pi}{\mu}-1)^\alpha = 0$ holds for all α. We then take the derivative of t on the above expansion.

$$\frac{\partial E_\mu(\frac{\pi}{\mu}-1)^\alpha}{\partial t} = \sum_{\ell=2}^{\alpha} \binom{\alpha}{\ell} (-1)^{\alpha-\ell} e^{\frac{\ell(\ell-1)t}{2}} \frac{\ell(\ell-1)}{2}$$

Define $\tilde{\ell} = \ell - 2$ and rewrite the above equation as

$$\begin{aligned} &\frac{\alpha(\alpha-1)}{2} \sum_{\tilde{\ell}=0}^{\alpha-2} \binom{\alpha-2}{\tilde{\ell}} (-1)^{\alpha-2-\tilde{\ell}} E_\mu(\frac{\pi}{\mu})^{\tilde{\ell}+2} \\ &= \frac{\alpha(\alpha-1)}{2} E_\mu[(\frac{\pi}{\mu})^2 \sum_{\tilde{\ell}=0}^{\alpha-2} \binom{\alpha-2}{\tilde{\ell}} (-1)^{\alpha-2-\tilde{\ell}} (\frac{\pi}{\mu})^\ell] \\ &= \frac{\alpha(\alpha-1)}{2} E_\mu[(\frac{\pi}{\mu})^2 (\frac{\pi}{\mu}-1)^{\alpha-2}] \end{aligned}$$

By applying lemma 16, we have $E_\mu[(\frac{\pi}{\mu})^2(\frac{\pi}{\mu}-1)^{\alpha-2}] \geq E_\mu[(\frac{\pi}{\mu}-1)^{\alpha-2}]$, where $\alpha-2 = \tilde{\alpha}$ and $E_\mu[(\frac{\pi}{\mu}-1)^{\tilde{\alpha}}] \geq 0$ from assumption. So the derivative is greater than 0 for all non-negative t. Combined with $E_\mu(\frac{\pi}{\mu}-1)^\alpha = 0$ when $t=0$, we have $E_\mu(\frac{\pi}{\mu}-1)^\alpha \geq 0$ hold for all t.
Since both the base case and the inductive step have been performed, by mathematical induction the statement holds for all odd $\alpha \geq 1$. □

B.2. Qualifying Laplace mechanism

Theorem 18. *Let π, μ Laplace density functions obeying $\mu(x) = \frac{1}{\lambda}e^{\frac{|x|}{\lambda}}$, and $\pi(x) = \frac{1}{\lambda}e^{\frac{|x+t|}{\lambda}}$. For all $\lambda > 0$, all natural number α, function $f(t) := \mathbb{E}_\mu\left[\left(\frac{\pi}{\mu} - 1\right)^\alpha\right]$ obeys that*

1. *$f(t) \geq 0$ for any $t \in \mathbb{R}$.*
2. *$f(t)$ monotonically increases for $t > 0$.*
3. *$f(t)$ monotonically decreases for $t < 0$.*

Proof. When $t = 0$, $\pi/\mu = 1$ and trivially $\mathbb{E}_\mu[(\pi/\mu - 1)^\alpha] = 0$ for any α. We will show that this is actually the minimizer for all $t \in \mathbb{R}$ by proving that the subdifferential $\partial_t f(t) \geq 0$ for $t > 0$ and $\partial_t f(t) \leq 0$ for $t < 0$.

$$\begin{aligned}
\partial_t f(t) =& \partial_t \left[\int \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \left(e^{\frac{-|x+t|+|x|}{\lambda}} - 1\right)^\alpha dx\right] \\
=& \int \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \cdot \alpha \left(e^{\frac{-|x+t|+|x|}{\lambda}} - 1\right)^{\alpha-1} \cdot e^{\frac{-|x+t|+|x|}{\lambda}} \frac{(-1)\cdot \partial_t |x+t|}{\lambda} dx \\
=& \int -\frac{\alpha}{2\lambda^2} e^{-\frac{|x+t|}{\lambda}} \left(e^{\frac{-|x+t|+|x|}{\lambda}} - 1\right)^{\alpha-1} \partial_t |x+t| dx \\
\underset{u:=x+t}{\uparrow}=& \int -\frac{\alpha}{2\lambda^2} e^{-\frac{|u|}{\lambda}} \left(e^{\frac{-|u|+|u-t|}{\lambda}} - 1\right)^{\alpha-1} \partial_u |u| du.
\end{aligned}$$

Note that

$$\partial_u |u| = \begin{cases} [-1, 1] & \text{if } u = 0; \\ \{\text{sign}(u)\} & \text{otherwise,} \end{cases}$$

which implies that we can write

$$\begin{aligned}
\partial_t f(t) =& \int_0^{+\infty} -\frac{\alpha}{2\lambda^2} e^{-\frac{|u|}{\lambda}} \left(e^{\frac{-|u|+|u-t|}{\lambda}} - 1\right)^{\alpha-1} du \\
&+ \int_{-\infty}^0 -\frac{\alpha}{2\lambda^2} e^{-\frac{|u|}{\lambda}} \left(e^{\frac{-|u|+|u-t|}{\lambda}} - 1\right)^{\alpha-1} du \\
=& \int_0^{+\infty} \frac{\alpha}{2\lambda^2} e^{-\frac{|u|}{\lambda}} \left[-\left(e^{\frac{-|u|+|u-t|}{\lambda}} - 1\right)^{\alpha-1} + \left(e^{\frac{-|u|+|u+t|}{\lambda}} - 1\right)^{\alpha-1}\right] du
\end{aligned}$$

For positive t, we can decompose the integral into

$$\begin{aligned}
&\int_t^{+\infty} \frac{\alpha}{2\lambda^2} e^{-\frac{|u|}{\lambda}} \left[-\left(e^{\frac{-t}{\lambda}} - 1\right)^{\alpha-1} + \left(e^{\frac{t}{\lambda}} - 1\right)^{\alpha-1}\right] du \\
&+ \int_0^t \frac{\alpha}{2\lambda^2} e^{-\frac{|u|}{\lambda}} \left[-\left(e^{\frac{t-2u}{\lambda}} - 1\right)^{\alpha-1} + \left(e^{\frac{t}{\lambda}} - 1\right)^{\alpha-1}\right] du.
\end{aligned}$$

For even $\alpha \geq 2$, $\alpha - 1$ is an even number and above expression is trivially nonnegative.

For odd $\alpha \geq 3$, that $\alpha - 1$ is even. By the inequality that $e^t - 1 \geq 1 - e^{-t}$ for any t, therefore the first term is nonnegative.

Now we address the second term. For $u \in [0, t/2]$, $0 \leq t - 2u \leq t$ and the nonnegativity follows directly from the monotonicity of $(e^\cdot - 1)$ on $[0, +\infty)$. For $u \in (t/2, t]$, $-t \leq t - 2u \leq 0$, and the nonnegativity follows from the fact that

$$e^t - 1 \geq 1 - e^{-t} \geq 1 - e^{-v}$$

for all $0 \leq v \leq t$. This concludes the proof for the positive t.

The results that the subgradient is positive for negative t follows naturally by symmetry. □

Remark 19 (Handling Laplace Mechanism in higher dimension)**.** *The generalization to higher dimension is trivial. The perturbation t is now a vector, but since the noise is added independently for each coordinate, we can work out the monotonicity for each coordinate separately.*

C. Proofs related to efficient approximation

Proof of Theorem 11. Apply $\epsilon(\ell) \le \epsilon(\alpha)$ for all $\ell = \tau+1, ..., \alpha$, we have:

$$\begin{aligned}
&\epsilon_{\mathcal{M}\circ\mathsf{PoissonSample}}(\alpha) \le \frac{1}{\alpha-1}\log\Big\{(1-\gamma)^{\alpha-1}(\alpha\gamma-\gamma+1) + \sum_{\ell=2}^{\tau}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell e^{(\ell-1)\epsilon(\ell)} \\
&\quad + \sum_{\ell=\tau+1}^{\alpha}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell e^{(\ell-1)\epsilon(\alpha)}\Big\} \\
=&\frac{1}{\alpha-1}\log\Big\{(1-\gamma)^{\alpha-1}(\alpha\gamma-\gamma+1) + \sum_{\ell=2}^{\tau}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell e^{(\ell-1)\epsilon(\ell)} \\
&\quad - \sum_{\ell=0}^{\tau}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell e^{(\ell-1)\epsilon(\alpha)} + e^{-\epsilon(\alpha)}(1-\gamma+\gamma e^{\epsilon(\alpha)})^\alpha\Big\} \\
=&\frac{1}{\alpha-1}\log\Big\{(1-\gamma)^{\alpha-1}(\alpha\gamma-\gamma+1) + \sum_{\ell=2}^{\tau}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell (e^{(\ell-1)\epsilon(\ell)} - e^{(\ell-1)\epsilon(\alpha)}) \\
&\quad - (1-\gamma)^\alpha e^{-\epsilon(\alpha)} - \alpha(1-\gamma)^{\alpha-1}\gamma + e^{-\epsilon(\alpha)}(1-\gamma+\gamma e^{\epsilon(\alpha)})^\alpha\Big\} \\
=&\frac{1}{\alpha-1}\log\Big\{(1-\gamma)^\alpha(1-e^{-\epsilon(\alpha)}) + e^{-\epsilon(\alpha)}(1-\gamma+\gamma e^{\epsilon(\alpha)})^\alpha \\
&\quad - \sum_{\ell=2}^{\tau}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell (e^{(\ell-1)\epsilon(\alpha)} - e^{(\ell-1)\epsilon(\ell)})\Big\}
\end{aligned}$$

□

Theorem 20 (Fast approximation for general upper bound)**.**

$$\begin{aligned}
&\epsilon_{\mathcal{M}\circ\mathsf{PoissonSample}}(\alpha) \le \frac{1}{\alpha-1}\log\Big\{(1-\gamma)^\alpha(1-3e^{-\epsilon(\alpha)}) + 3e^{-\epsilon(\alpha)}(1-\gamma+\gamma e^{\epsilon(\alpha)})^\alpha \\
&- 3\sum_{\ell=3}^{\tau}\binom{\alpha}{\ell}(1-\gamma)^{\alpha-\ell}\gamma^\ell (e^{(\ell-1)\epsilon(\alpha)} - e^{(\ell-1)\epsilon(\ell)}) \\
&- 2\gamma\alpha(1-\gamma)^{\alpha-1} + \binom{\alpha}{2}\gamma^2(1-\gamma)^{\alpha-2}(e^{\epsilon(2)} - 3e^{\epsilon(\alpha)})\Big\}.
\end{aligned}$$

Proof is similar to that of Theorem 11 thus omitted.

C.1. experiments on approximate methods

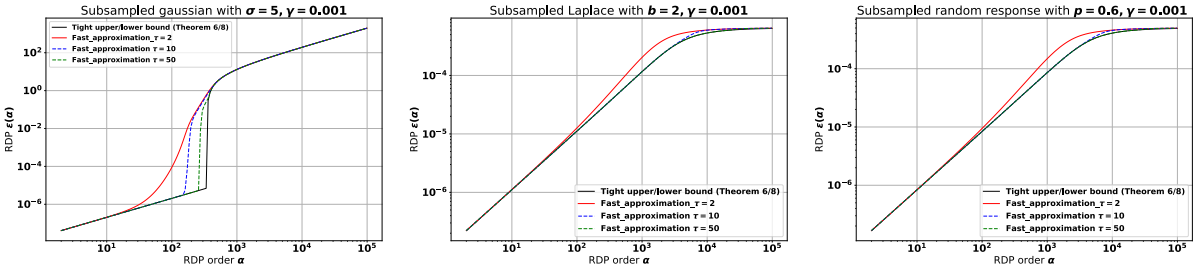
D. Comparison to the implementation of Abadi et al. (2016)

According to Section 3.2 of Abadi et al. (2016), in the implementation of the moments accountant they used numerical integration to compute

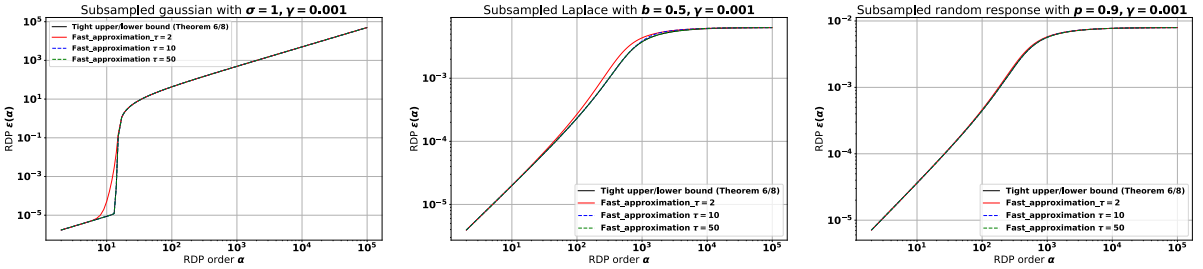
$$\begin{aligned}
E_1 &= \mathbb{E}_{z\sim\mu_0}[(\mu_0(z)/\mu(z))^{\alpha-1}] = \mathbb{E}_{z\sim\mu}[(\mu_0(z)/\mu(z))^\alpha] \\
E_2 &= \mathbb{E}_{z\sim\mu}[(\mu(z)/\mu_0(z))^{\alpha-1}] = \mathbb{E}_{z\sim\mu_0}[(\mu(z)/\mu_0(z))^\alpha]
\end{aligned}$$

where $\mu_0 = \mathcal{N}(0,\sigma^2)$ and $\mu = \gamma\mathcal{N}(1,\sigma^2) + (1-\mu)\mathcal{N}(0,\sigma^2)$ then output

$$\epsilon(\alpha) \le \frac{1}{\alpha-1}\log(\max\{E_1, E_2\}).$$



(a) Subsampled Gaussian with $\sigma = 5$ (b) Subsampled Laplace with $b = 2$ (c) Subsampled Random Response with $p = 0.6$



(d) Subsampled Gaussian with $\sigma = 1$ (e) Subsampled Laplace with $b = 0.5$ (f) Subsampled Random Response with $p = 0.9$

Figure 5. Illustration of the numerical fast τ-term apprixomation results under high privacy and low privacy regimes. The x-axis is the order α, and the y-axis is the RDP parameter $(\epsilon(\alpha))$, the subsampling rate $\gamma = 0.001$ in all the experiments. The Approximate RDP upper bound is obtained through Theorem 11, and the corresponding tight upper bound in possion subsample case is represented as the black curve.

This approach is correct but costly, because a different numerical integration is needed for each α. Our result implies that $E_2 > E_1$ and one never need to numerically simulate E_1.

The most recent update to the moments accountant implementation of the Tensorflow Privacy package is slightly different from the version described in Section 3.2 of Abadi et al. (2016). The new version of their code https://github.com/tensorflow/privacy/blob/master/privacy/analysis/rdp_accountant.py implements an analytical version of E_2 via the Binomial expansion — essentially our tight bound Theorem 6 for Poisson-Sampled Gaussian mechanism verbatim with a prescribed list of αs. The current paper complements this implementation with a proof that $E_2 > E_1$, which justifies that doing this is correct. To the best of our knowledge, the current paper is the first that rigorously establishes $E_2 \geq E_1$ which establishes that this new implementation is correct for Poisson subsampling.

Our implementation of moments accountant in AutoDP (https://github.com/yuxiangw/autodp) is a more flexible framework that allows us to exactly or almost exactly track the RDP of any subsampled differentially private mechanisms provided that the based mechanism's RDP is known.