Infinitely Divisible Noise for Differential Privacy: Nearly Optimal Error in the High ε Regime

Charlie Harrison
Google, Austin, TX, USA

Pasin Manurangsi
Google Research, Bangkok, Thailand

Abstract

Differential privacy (DP) can be achieved in a distributed manner, where multiple parties add independent noise such that their sum protects the overall dataset with DP. A common technique here is for each party to sample their noise from the decomposition of an infinitely divisible distribution. We analyze two mechanisms in this setting: 1) the generalized discrete Laplace (GDL) mechanism, whose distribution (which is closed under summation) follows from differences of i.i.d. negative binomial shares, and 2) the multi-scale discrete Laplace (MSDLap) mechanism, a novel mechanism following the sum of multiple i.i.d. discrete Laplace shares at different scales. For $\varepsilon \geq 1$, our mechanisms can be parameterized to have $O\left(\Delta^3 e^{-\varepsilon}\right)$ and $O\left(\min\left(\Delta^3 e^{-\varepsilon}, \Delta^2 e^{-2\varepsilon/3}\right)\right)$ MSE, respectively, where Δ denote the sensitivity; the latter bound matches known optimality results. Furthermore, the MSDLap mechanism has the optimal MSE *including constants* as $\varepsilon \to \infty$. We also show a transformation from the discrete setting to the continuous setting, which allows us to transform both mechanisms to the continuous setting and thereby achieve the optimal $O\left(\Delta^2 e^{-2\varepsilon/3}\right)$ MSE. To our knowledge, these are the first infinitely divisible additive noise mechanisms that achieve order-optimal MSE under pure DP for either the discrete or continuous setting, so our work shows formally there is no separation in utility when query-independent noise adding mechanisms are restricted to infinitely divisible noise. For the continuous setting, our result improves upon Pagh and Stausholm's Arete distribution which gives an MSE of $O\left(\Delta^2 e^{-\varepsilon/4}\right)$ [39]. Furthermore, we give an exact sampler tuned to efficiently implement the MSDLap mechanism, and we apply our results to improve a state of the art multi-message shuffle DP protocol from [7] in the high ε regime.

2012 ACM Subject Classification Mathematics of computing → Distribution functions; Security and privacy

Keywords and phrases Differential Privacy, Distributed Noise Addition

Digital Object Identifier 10.4230/LIPIcs.FORC.2025.12

Related Version *Full Version*: https://arxiv.org/abs/2504.05202

Acknowledgements Thanks to Peter Kairouz for useful discussions and pointing out the prior work, especially [6]. This work was supported by Google.

1 Introduction

Differential Privacy (DP) [15] is a formal notion of privacy which bounds the sensitive information revealed by an algorithm. While there are many "flavors" of DP, most relevant to this work is so-called pure-DP or ε-DP which bounds the privacy loss via the parameter ε.

▶ **Definition 1** ([15]). *A randomized mechanism $M : \mathcal{X}^d \to \mathcal{Y}$ is ε-DP if, for all $x, x' \in \mathcal{X}^d$ differing*[1] *in a single entry, $\Pr[M(x) \in S] \leq e^\varepsilon \cdot \Pr[M(x') \in S]$ for all measurable $S \subseteq \mathcal{Y}$.*

[1] For this work, we may consider any neighboring notion. We use the substitution notion for simplicity.

We focus on the so-called *low-privacy regime* where $\varepsilon \geq 1$. Despite its name, this regime still provides meaningful privacy protection and is the setting most often employed in practical applications of DP (e.g. [1, 4, 44]). The bounds we state throughout will focus on this regime.

A challenge in deploying DP is doing so while also producing *useful* results. In this work, we focus on the mean squared error (MSE) of a query q subject to a query-independent additive noise mechanism, i.e. $M(x) = q(x) + Z$ where Z is a random variable. There is a rich body of work on optimizing MSE in this setting. Notably, the staircase mechanism ([18, 17, 19]) was shown to have the optimal MSE of all ε-DP, query-independent additive noise mechanisms. In the continuous setting, it achieves a MSE of $O\left(\Delta^2 e^{-2\varepsilon/3}\right)$. In the discrete setting, its MSE interpolates between $O\left(\Delta^3 e^{-\varepsilon}\right)$ and $O\left(\Delta^2 e^{-2\varepsilon/3}\right)$. For $\Delta = 1$, the optimal discrete staircase mechanism is the discrete Laplace (aka Geometric) mechanism [25].

A probability distribution $\mathcal{D}$ is *infinitely divisible* iff for every positive integer n, there exists a distribution $\mathcal{D}_{/n}$ such that, when we sample n i.i.d. random variables $Z_1, \cdots, Z_n \sim \mathcal{D}_{/n}$, their sum $Z = \sum_{i=1}^{n} Z_i$ is distributed as $\mathcal{D}$. In distributed DP, a common technique (see [26] for an overview) is for n parties to each sample Z_i such that the sum is distributed according to $\mathcal{D}$, which can be shown to protect the dataset with DP. The infinite divisibility property of $\mathcal{D}$ allows for distributed protocols where an arbitrary $n \geq 1$ number of parties can participate. Under the more restrictive setting where the additive noise mechanism M must sample the noise Z from an *infinitely divisible* distribution, there was previously no known mechanism in either the discrete or continuous settings which matched the MSE of the staircase mechanism. We resolve this gap in this paper for both settings.

1.1 Related work

Distributed noise generation for differential privacy is well studied even for distributions that are not infinitely divisible. In fact, the idea dates back to the very early days of DP [14]. Moreover, several works have studied the setting where $Z_1, \ldots, Z_n$ samples from some distribution $\tilde{\mathcal{D}}$ and directly argue about the distribution of their summation $Z = Z_1 + \cdots + Z_n$. Examples include the case where $\tilde{\mathcal{D}}$ is a Bernoulli distribution [13, 20], for which Z is a Binomial random variable, and the case where $\tilde{\mathcal{D}}$ is a discrete Gaussian distribution [33], for which Z is "close" to discrete Gaussian random variable. The drawback here is that the distribution of Z are different for different values of n, meaning that the privacy analysis often requires n to be sufficiently large (e.g. [20]) or sufficiently small (e.g. [33]). Using infinitely divisible distribution overcomes this issue since the distribution of the total noise Z is always $\mathcal{D}$ regardless of the value of n, leading to a privacy analysis that works for all regimes of n. Due to this, infinitely divisible noise distributions have gained popularity in distributed settings of differential privacy (e.g. [26, 7, 22, 23, 2, 12, 6, 21]).

As discrete distributions are typically easier to embed in multi-party cryptographic protocols and avoid implementation issues [36] with floating point representations, they tend to be more well-studied in distributed DP. The infinite divisibility of the discrete Laplace into negative binomial[2] shares has been studied in [26, 7, 6]. In [6] the authors explicitly analyzed privacy in the face of *dropouts*, or parties that fail to properly add their noise share.

In the continuous setting, the infinitely divisible Arete distribution was introduced in [39] specifically to target the low-privacy, high ε pure-DP regime. It was designed to match the performance of the (continuous) staircase mechanism which is not infinitely divisible

[2] Throughout this work, we use the term negative binomial to refer to the distribution generalized to a real valued stopping parameter r. In other works, this is sometimes called the Pólya distribution.

and therefore unusable in the distributed setting. While the continuous staircase mechanism achieves $O(\Delta^2 e^{-2\varepsilon/3})$ MSE, the authors only proved the Arete mechanism has an MSE of $O(\Delta^2 e^{-\varepsilon/4})$, though we believe this is not tight (Conjecture 30).

Distributed noise is relevant in other notions of differential privacy as well. The aforementioned discrete Gaussian mechanism [11, 33] satisfies zero-concentrated DP [10], and the Skellam mechanism [2, 43, 41] satisfies Rényi DP [37].[3] Both mechanisms were proposed in the context of federated learning via a secure aggregation multi-party protocol [34, 9, 8].

The GDL mechanism was explored informally in a prior blog post by the first author [27]. Concurrent to this work, the GDL distribution was studied in the context of (ε, δ) shuffle DP in [5]. There the authors analyzed a *shifted and truncated* GDL distribution that achieves an approximate DP bound guarantee, vs. the pure DP one in this work which does not perform any truncation. Their primary result in the shuffle setting involving nearly matching the utility of the central discrete Laplace mechanism. On the other hand, as explained below, we *improve on* the central discrete Laplace mechanism for sufficiently large ε.

1.2 Our contributions

Table 1 Summary of our results (bold) compared to known noise distributions satisfying ε-DP. MSEs exclude constant factors.

Discrete Distributions

Distribution	MSE	Inf. Div.	Reference
Discrete Laplace	$e^{-\varepsilon/\Delta}$	✓	[25]
Discrete Staircase	$\min\{\Delta^3 e^{-\varepsilon}, \Delta^2 e^{-2\varepsilon/3}\}$	✗	[18, 17, 19]
Generalized Discrete Laplace (GDL)	$\Delta^3 e^{-\varepsilon}$ for $\varepsilon > 2 + \log(\Delta)$	✓	Theorem 15
Multi-Scale Discrete Laplace (MSDLap)	$\min\{\Delta^3 e^{-\varepsilon}, \Delta^2 e^{-2\varepsilon/3}\}$	✓	Corollary 19

Continuous Distributions

Distribution	MSE	Inf. Div.	Reference
Laplace	Δ^2/ε^2	✓	[15]
Staircase	$\Delta^2 e^{-2\varepsilon/3}$	✗	[18, 17, 19]
Arete	$\Delta^2 e^{-\varepsilon/4}$	✓	[39]
Continuous MSDLap	$\Delta^2 e^{-2\varepsilon/3}$	✓	Theorem 21

In Section 3, we introduce the GDL and MSDLap mechanisms, two discrete noise-adding mechanisms having optimal $O(\Delta^3 e^{-\varepsilon})$ MSE for fixed Δ and any sufficiently large ε. Inspired by the discrete staircase mechanism, we also extend the MSDLap mechanism with a parameter optionally allowing it to satisfy $O(\Delta^2 e^{-2\varepsilon/3})$ MSE, allowing it to achieve asymptotically-optimal error for any fixed Δ and $\varepsilon \geq 1$. Notably, the MSDLap mechanism matches the MSE of the discrete staircase *including constants* as $\varepsilon \to \infty$.

[3] Mechanisms satisfying Rényi DP or zero-concentrated DP naturally also satisfy approximate DP.

The GDL mechanism, as the difference of two i.i.d. negative binomial noise shares, generalizes the distributed discrete Laplace mechanisms in [26, 7, 6] in the face of *unexpected* dropouts when a larger fraction of parties than expected fail to add their noise shares, providing a "smooth" closed-form privacy guarantee.

In Section 4, we introduce a method to transform a discrete infinitely divisible additive mechanism into a continuous one up to a small loss in parameters. This allows us to achieve the asymptotically-optimal $O\left(\Delta^2 e^{-2\varepsilon/3}\right)$ MSE by transforming either the GDL or the MSDlap mechanisms, improving on the Arete mechanism's bound of $O\left(\Delta^2 e^{-\varepsilon/4}\right)$ in [39].

Our noise distributions and previously known distributions are summarized in Table 1.

While its utility exceeds the GDL's, the MSDLap mechanism naively requires sampling from $O(\Delta)$ independent negative binomial random variables. In Section 5 we outline an improved exact sampling algorithm which runs in only $O(1)$ steps in expectation for relevant regimes. This algorithm may be of independent interest for general purpose multivariate negative binomial sampling in the sparse regime where most samples are 0.

Finally, in Section 6, we improve the multi-message "split and mix" real summation shuffle protocol of [7] with our results, attaining the $O\left(e^{-2\varepsilon/3}\right)$ bound on MSE where previous results could only achieve $O\left(1/\varepsilon^2\right)$ from the discrete Laplace.

2 Preliminaries

For any $n \in \mathbb{N}$, we write $[n]$ as a shorthand for $\{1, \ldots, n\}$.

Function identities. We introduce a few functions and identities used in our proofs. Let $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$ be the gamma function. We denote the rising factorial (aka the Pochhammer symbol) by $(x)_n := (x)(x+1)\cdots(x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)}$. Denote the hypergeometric function by ${}_2F_1\left[a, b; c; z\right]$ [38], where

$$ {}_2F_1\left[a, b; c; z\right] = \sum_{s=0}^{\infty} \frac{(a)_s (b)_s}{(c)_s s!} z^s = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{s=0}^{\infty} \frac{\Gamma(a+s)\Gamma(b+s)}{\Gamma(c+s)s!} z^s. \tag{1} $$

Let $\psi(z)$ denote the digamma function, where $\psi(z) = \frac{\mathrm{d}}{\mathrm{d}z}\log(\Gamma(z)) = \frac{\Gamma'(z)}{\Gamma(z)}$. The following observation is well-known (see e.g. [3]):

▶ **Observation 2.** *ψ is increasing and ψ' is decreasing on $(0, \infty)$.*

Finally, we state the so-called "hockey-stick identity" (e.g. [32]):

▶ **Lemma 3.** *For any non-negative integers ℓ, m, $\sum_{j=\ell}^{m} \binom{j}{\ell} = \binom{m+1}{\ell+1}$*

Distributions. For convenience, we write a random variable (r.v.) and its distribution interchangeably.

For a discrete distribution $\mathcal{D}$ with support on $\mathcal{X}$, denote its probability mass function (PMF) as $f_{\mathcal{D}}(k)$ for $k \in \mathcal{X}$. When we say that a discrete distribution is infinitely divisible, we assume implicitly that $\mathcal{D}_{/n}$ are also discrete. Relevant to this work are the following well-known discrete distributions (where $a, r \in \mathbb{R}_{>0}, p \in (0,1)$):

- The negative binomial (NB) distribution, denoted $\mathrm{NB}(r,p)$ and with support on $\mathbb{Z}_{\geq 0}$, has PMF $f_{\mathrm{NB}(r,p)}(k) = (1-p)^k p^r \frac{\Gamma(k+r)}{\Gamma(r)\Gamma(k+1)}$. It is infinitely divisible, as $\sum_{i=1}^{n} \mathrm{NB}(r/n, p) \sim \mathrm{NB}(r,p)$. Its variance is $\mathrm{Var}(\mathrm{NB}(r,p)) = \frac{(1-p)r}{p^2}$. For $r \in \mathbb{N}$, the negative binomial distribution models the number of failures before the first r successes in a sequence of i.i.d. Bernoulli trials with success probability p.

- The geometric distribution, denoted $\mathrm{Geo}(p)$ is a special case of NB with $r = 1$.
- The discrete Laplace distribution, denoted $\mathrm{DLap}(a)$ and with support on $\mathbb{Z}$, has PMF $f_{\mathrm{DLap}(a)}(k) = \tanh(a/2)e^{-a|k|}$. Since $\mathrm{DLap}(a) \sim \mathrm{NB}(1, 1-e^{-a}) - \mathrm{NB}(1, 1-e^{-a})$, it inherits infinite divisibility from NB. Its variance is $\mathrm{Var}(\mathrm{DLap}(a)) = \frac{1}{\cosh(a)-1}$.
- The Bernoulli distribution, denoted $\mathrm{Ber}(p)$, has PMF $f_{\mathrm{Ber}(p)}(k) = \begin{cases} p & k = 1 \\ 1-p & k = 0 \end{cases}$.

For a continuous distribution $\mathcal{D}$ on $\mathcal{X}$, let $f_{\mathcal{D}}(x)$ for $x \in \mathcal{X}$ be its probability density function (PDF) at x. Recall the following continuous distributions (where $k, \theta, b \in \mathbb{R}_{>0}$):

- The gamma distribution, denoted $\Gamma(k, \theta)$, with support on $\mathbb{R}_+$ has PDF $f_{\Gamma(k,\theta)}(x) = \frac{1}{\Gamma(k)\theta^k} \cdot x^{k-1}e^{-x/\theta}$. It is infinitely divisible as $\sum_{i=1}^{n} \Gamma(k/n, \theta) \sim \Gamma(k, \theta)$.
- The Laplace distribution, denoted $\mathrm{Lap}(b)$, with support on $\mathbb{R}$ has PDF $f_{\mathrm{Lap}(b)}(x) = \frac{1}{2b}e^{-|x|/b}$. Its variance is $2b^2$. It is infinitely divisible as $\sum_{i=1}^{n} (\Gamma(1/n, b) - \Gamma(1/n, b)) \sim \mathrm{Lap}(b)$.

We will also use the following simple observations (where Z, Z_1, Z_2 are r.v.s):

▶ **Observation 4.** *If Z is infinitely divisible, $c \cdot Z$ is infinitely divisible for any constant c.*

▶ **Observation 5.** *If Z_1, Z_2 are infinitely divisible and independent, $Z_1 + Z_2$ is infinitely divisible.*

We say that a distribution $\mathcal{D}$ is *closed under summation* if $\mathcal{D}$ is infinitely divisible and additionally, $\mathcal{D}_{/n}$ follows the same distribution family as $\mathcal{D}$ for all $n \in \mathbb{N}$. This additional property provides benefits in the distributed setting as it ensures the mechanism's privacy is well-understood even as parties drop out or join the protocol.

Max Divergence and DP. Let $\mathrm{D}_\infty(P \parallel Q)$ denote the max divergence between two distributions P, Q, i.e., $\sup_{x \in \mathrm{supp}(P)} \frac{f_P(x)}{f_Q(x)}$. We state the following well-known properties.

▶ **Lemma 6** (Post-Processing). *For any (possibly randomized) function f and any random variables U, V, we have $\mathrm{D}_\infty(f(U) \parallel f(V)) \le \mathrm{D}_\infty(U \parallel V)$.*

▶ **Lemma 7** (Triangle Inequality). *For any distributions P, Q, R, $\mathrm{D}_\infty(P \parallel Q) \le \mathrm{D}_\infty(P \parallel R) + \mathrm{D}_\infty(R \parallel Q)$.*

For a query function $q : X^d \to \mathcal{Y}$, we let $\Delta(q) = \max_{x,x'} |q(x) - q(x')|$ where the maximum is over all pairs x and x' differing on one entry. The $\mathcal{D}$-noise addition mechanism for a query function q is the mechanism $M(x)$ that outputs $q(x) + Z$ where Z is drawn from $\mathcal{D}$. For a discrete (resp. continuous) distribution $\mathcal{D}$ and $\Delta \in \mathbb{N}$ (resp. $\Delta \in \mathbb{R}_{>0}$), we say the *$\mathcal{D}$-noise addition mechanism is ε-DP for sensitivity Δ* iff the $\mathcal{D}$ noise addition mechanism is ε-DP for all queries $q : \mathcal{X}^d \to \mathbb{Z}$ (resp. $q : \mathcal{X}^d \to \mathbb{R}$) such that $\Delta(q) \le \Delta$. It follows from the definition of DP and max divergence that this condition translates to the following (see e.g. [18]):

▶ **Lemma 8.** *For a discrete (resp. continuous) distribution $\mathcal{D}$, the $\mathcal{D}$-noise addition mechanism is ε-DP for sensitivity Δ iff $\mathrm{D}_\infty(\mathcal{D} + \xi \parallel \mathcal{D}) \le \varepsilon$ for all $\xi \in \{-\Delta, -(\Delta-1), \ldots, \Delta\}$ (resp. if[4] $\mathrm{D}_\infty(\mathcal{D} + \xi \parallel \mathcal{D}) \le \varepsilon$ for all $\xi \in [-\Delta, \Delta]$).*

[4] The other direction of the implication for the continuous case does not hold since e.g. the set of x where $f_{\mathcal{D}+\xi}(x) > e^{\varepsilon} \cdot f_{\mathcal{D}}(x)$ might have measure zero.

3 Discrete mechanisms

We present two infinitely divisible discrete noises with order-optimal MSEs.

3.1 The generalized discrete Laplace mechanism

This section introduces the generalized discrete Laplace distribution and associated mechanism. We start by the description of the distribution and its PMF:

▶ **Definition 9** ([35]). *For $\beta, a > 0$, let $\mathrm{GDL}(\beta, a)$ denote the distribution of $Z_1 - Z_2$ where $Z_1, Z_2 \overset{i.i.d.}{\sim} \mathrm{NB}\left(\beta, 1 - e^{-a}\right)$. For all $x \in \mathbb{Z}$, the PMF of $\mathrm{GDL}(\beta, a)$ at x, i.e. $f_{\mathrm{GDL}(\beta,a)}(x)$, is*

$$e^{-a|x|}\left(1-e^{-a}\right)^{2\beta} {}_2F_1\left[\beta, \beta+|x|; 1+|x|; e^{-2a}\right] \frac{\Gamma(\beta+|x|)}{\Gamma(1+|x|)\Gamma(\beta)}. \tag{2}$$

The discrete Laplace distribution is a special case of the GDL with $\beta = 1$.

The infinite divisibility of NB immediately implies that GDL is also infinitely divisible:

▶ **Observation 10.** *GDL is infinitely divisible and closed under summation. In particular, for independent $X_1 \sim \mathrm{GDL}(\beta_1, a), \cdots, X_n \sim \mathrm{GDL}(\beta_n, a)$, we have $\sum_{i=1}^n X_i \sim \mathrm{GDL}(\sum_{i=1}^n \beta_i, a)$.*

It will be convenient to consider the extension of the PMF to all real numbers; let $f^{\mathbb{R}}_{\mathrm{GDL}(\beta,a)} : \mathbb{R} \to \mathbb{R}$ be the function defined by (2). We start with the following lemma.

▶ **Lemma 11.** *For $\beta \in (0, 1)$, $f^{\mathbb{R}}_{\mathrm{GDL}(\beta,a)}$ is decreasing and log convex on $[0, \infty)$.*

Proof. Because the product of log convex and decreasing functions is log convex and decreasing, and e^{-ax} is both log convex and decreasing, we focus on remaining relevant term.

$$\begin{aligned}{}_2F_1\left[\beta, \beta+|x|; 1+|x|; e^{-2a}\right] \frac{\Gamma(\beta+|x|)}{\Gamma(1+|x|)\Gamma(\beta)} &\overset{(1)}{=} \sum_{s=0}^{\infty} \frac{\Gamma(\beta+s)\Gamma(\beta+x+s)}{\Gamma(\beta)^2\Gamma(1+x+s)s!} e^{-2as} \\ &= \sum_{s=0}^{\infty} g(x) \frac{\Gamma(\beta+s)e^{-2as}}{\Gamma(\beta)^2 s!},\end{aligned}$$

where $g(x) = \frac{\Gamma(\beta+x+s)}{\Gamma(1+x+s)}$. First we show that $g'(x) = g(x)(\psi(x+s+\beta) - \psi(x+s+1)) < 0$ and that $\frac{\mathrm{d}^2}{\mathrm{d}x^2} \log g(x) = \psi'(\beta+x+s) - \psi'(1+x+s) \geq 0$. The derivatives follow from the definition of ψ, and the inequalities follow from Observation 2. Finally, the result follows as the sum of decreasing and log convex functions is decreasing and log convex. ◀

We also need the following technical lemma which, together with Lemma 11, allows us to consider only $f(0)/f(\Delta)$. Its proof is deferred to Section A.1.

▶ **Lemma 12.** *Let $f : \mathbb{R} \to \mathbb{R}$ be symmetric about 0, and decreasing and log convex on $[0, \infty)$. Then for any x, x' such that $|x - x'| \leq \Delta$, we have $\frac{f(x)}{f(x')} \leq \frac{f(0)}{f(\Delta)}$.*

We are now ready to state the privacy guarantee of the GDL-noise addition mechanism.

▶ **Theorem 13.** *For any $\Delta \in \mathbb{N}, \beta, a > 0$, the $\mathrm{GDL}(\beta, a)$-noise addition mechanism is ε-DP for sensitivity Δ iff*

$$\varepsilon \leq \begin{cases} a\Delta + \log \frac{{}_2F_1\left[\beta,\beta;1;e^{-2a}\right]}{{}_2F_1\left[\beta,\beta+\Delta;1+\Delta;e^{-2a}\right]} \frac{\Gamma(\Delta+1)\Gamma(\beta)}{\Gamma(\beta+\Delta)} & 0 < \beta < 1 \\ a\Delta & \beta \geq 1 \end{cases}$$

Proof. For the $0 < \beta < 1$ case, by Lemmas 8 and 12 along with the fact that $f^{\mathbb{R}}_{\mathrm{GDL}(\beta,a)}$ is log-convex on $[0,\infty)$ and symmetric about 0, the $\mathrm{GDL}(\beta,a)$-noise addition mechanism is ε-DP for sensitivity Δ iff $\varepsilon \le \max_{\xi\in\{-\Delta,\dots,\Delta\}} \max_{x\in\mathbb{Z}} \frac{f_{\mathrm{GDL}(\beta,a)}(x-\xi)}{f_{\mathrm{GDL}(\beta,a)}(x)} \overset{\text{Lemma 12}}{=} \log \frac{f_{\mathrm{GDL}(\beta,a)}(0)}{f_{\mathrm{GDL}(\beta,a)}(\Delta)}$, which is exactly the claimed bound.

For $\beta \ge 1$, we decompose the mechanism by observing (see Observation 10) that $Z \sim \mathrm{GDL}(\beta,a)$ can be sampled as $Z = Z_1 + Z_2$ where $Z_1 \sim \mathrm{DLap}(a), Z_2 \sim \mathrm{GDL}(\beta-1,a)$ are independent. By post-processing (Lemma 6), the $\mathrm{GDL}(\beta,a)$-noise addition mechanism is at least as private as the $\mathrm{DLap}(a)$-noise addition mechanism, which is $a\Delta$-DP for sensitivity Δ.

For the tightness claim, first note that Equation (1) yields

$$\log\left(\frac{f_{\mathrm{GDL}(\beta,a)}(k)}{f_{\mathrm{GDL}(\beta,a)}(k+\Delta)}\right) = a\Delta + \log\left(\frac{\sum_{s=0}^{\infty}\frac{(\beta)_s(\beta+k)_s e^{-2as}}{(1+k)_s s!}}{\sum_{s=0}^{\infty}\frac{(\beta)_s(\beta+k+\Delta)_s e^{-2as}}{(1+k+\Delta)_s s!}} \cdot \frac{\Gamma(\beta+k)\Gamma(1+k+\Delta)}{\Gamma(1+k)\Gamma(\beta+k+\Delta)}\right).$$

Observe that $\frac{\frac{(\beta)_s(\beta+k)_s e^{-2as}}{(1+k)_s s!}}{\frac{(\beta)_s(\beta+k+\Delta)_s e^{-2as}}{(1+k+\Delta)_s s!}} = \prod_{i=0}^{s-1}\frac{\beta+k+i}{1+k+i}\cdot\frac{1+k+\Delta+i}{\beta+k+\Delta+i} \ge 1$, because every term in the product is at least one. Plugging this into the above, we get

$$\log\left(\frac{f_{\mathrm{GDL}(\beta,a)}(k)}{f_{\mathrm{GDL}(\beta,a)}(k+\Delta)}\right) \ge a\Delta + \log\left(\frac{\Gamma(\beta+k)\Gamma(1+k+\Delta)}{\Gamma(1+k)\Gamma(\beta+k+\Delta)}\right) = a\Delta + \log\left(\frac{(1+k)_\Delta}{(\beta+k)_\Delta}\right).$$

Thus, $\lim_{k\to\infty}\log\left(\frac{f_{\mathrm{GDL}(\beta,a)}(k)}{f_{\mathrm{GDL}(\beta,a)}(k+\Delta)}\right) \ge a\Delta$, meaning that our bound is tight. ◀

The exact privacy bound for GDL above (for $0 < \beta < 1$) is fairly unwieldy. We show a simplified bound below, as well as a tighter, slightly more complex bound in Section A.2.

▶ **Corollary 14.** *For any $\Delta \in N, a > 0, \beta \in (0,1)$, the $\mathrm{GDL}(\beta,a)$-noise addition mechanism is ε-DP for sensitivity Δ where $\varepsilon \le a\Delta + \log\frac{\Delta}{\beta}$.*

Proof. Since $\frac{(\beta)_s}{s!} \le \frac{(\beta+\Delta)_s}{(1+\Delta)_s}$, we can bound the ratio of the hypergeometric functions by

$$\frac{{}_2F_1\left[\beta,\beta;1;e^{-2a}\right]}{{}_2F_1\left[\beta,\beta+\Delta;1+\Delta;e^{-2a}\right]} = \frac{\sum_{s=0}^{\infty}\frac{(\beta)_s(\beta)_s}{s!\cdot s!}e^{-2as}}{\sum_{s=0}^{\infty}\frac{(\beta)_s(\beta+\Delta)_s}{(1+\Delta)_s s!}e^{-2as}} \le 1.$$

Thus, from Theorem 13, the mechanism is ε-DP for all $\varepsilon \le a\Delta + \log\frac{\Gamma(\Delta+1)\Gamma(\beta)}{\Gamma(\beta+\Delta)}$. Observe also that $\frac{\Gamma(\Delta+1)\Gamma(\beta)}{\Gamma(\beta+\Delta)} = \frac{\Delta}{\beta}\cdot\frac{(\Delta-1)!}{(\beta+1)_{\Delta-1}} \le \frac{\Delta}{\beta}$. Combining these yields the desired claim. ◀

Finally, we prove our accuracy guarantee for the GDL mechanism in the high ε regime.

▶ **Theorem 15.** *For any $\Delta \in \mathbb{N}$ and $\varepsilon > 2 + \log\Delta$, the $\mathrm{GDL}\left(\Delta e^{2-\varepsilon}, \frac{2}{\Delta}\right)$-noise addition mechanism is ε-DP for sensitivity Δ and has MSE $O(\Delta^3 e^{-\varepsilon})$.*

Proof. Since $\varepsilon > 2 + \log\Delta$, we have $\beta = \Delta e^{2-\varepsilon} < 1$. Thus, Corollary 14 implies the privacy guarantee. Finally, the MSE is $2\,\mathrm{Var}(\mathrm{NB}(\beta, 1-e^{-a})) = \frac{\Delta e^{2-\varepsilon}}{\cosh(2/\Delta)-1} = O(\Delta^3 e^{-\varepsilon})$. ◀

3.2 The multi-scale discrete Laplace mechanism

The (ε,Δ)-*multi-scale discrete Laplace ((ε,Δ)-MSDLap) distribution* with parameter $\varepsilon > 0, \Delta \in \mathbb{N}$ is defined as the distribution of $\sum_{i=1}^{\Delta} i \cdot X_i$ where $X_1,\dots,X_\Delta \overset{\text{i.i.d.}}{\sim} \mathrm{DLap}(\varepsilon)$. From Observation 4 and Observation 5, the (ε,Δ)-MSDLap distribution is infinitely divisible.

This mechanism is ε-DP, and its accuracy guarantee matches that of Theorem 15:

▶ **Theorem 16.** *For any $\varepsilon > 0, \Delta \in \mathbb{N}$, the (ε, Δ)-MSDLap-noise addition mechanism is ε-DP for sensitivity Δ. Furthermore, for $\varepsilon \geq 1$, the MSE is $O(\Delta^3 \cdot e^{-\varepsilon})$.*

Proof. *(Privacy)* From Lemma 8 and the symmetry of the noise around 0, it suffices to show $\mathrm{D}_\infty\left(\xi + \sum_{i=1}^{\Delta} i \cdot X_i \,\middle\|\, \sum_{i=1}^{\Delta} i \cdot X_i\right) \leq \varepsilon$ for all $\xi \in [\Delta]$. From Lemma 6, we have

$$\mathrm{D}_\infty\left(\xi + \sum_{i=1}^{\Delta} i \cdot X_i \,\middle\|\, \sum_{i=1}^{\Delta} i \cdot X_i\right) \leq \mathrm{D}_\infty\left(\xi + \xi \cdot X_\xi \parallel \xi \cdot X_\xi\right) = \mathrm{D}_\infty\left(1 + X_\xi \parallel X_\xi\right) \leq \varepsilon,$$

where the last inequality follows from $X_i \sim \mathrm{DLap}(\varepsilon)$.

(Accuracy) MSE is $\mathrm{Var}\left(\sum_{i=1}^{\Delta} i \cdot X_i\right) = \sum_{i=1}^{\Delta} i^2 \cdot \mathrm{Var}(X_i) = O(\Delta^3 \cdot e^{-\varepsilon})$. ◀

Theorem 16 implies that, for fixed Δ and sufficiently large ε, the (ε, Δ)-MSDLap-noise achieves asymptotically optimal MSE. In fact, below we prove a stronger statement: When $\varepsilon \to \infty$, the ratio between MSE of MSDLap and the optimal MSE [17, 19] approaches 1.

▶ **Corollary 17.** *For a fixed Δ, the ratio of the MSE of the (ε, Δ)-MSDLap-noise addition mechanism and that of the ε-DP discrete staircase mechanism for sensitivity Δ [17, 19] approaches 1 as $\varepsilon \to \infty$.*

Proof. As we show in Observation 35 (in Section A.3), the MSE of the ε-DP discrete staircase mechanism for sensitivity Δ is at least $\frac{1}{e^\varepsilon + (2\Delta - 1)} \cdot \frac{\Delta(\Delta+1)(2\Delta+1)}{3}$ for any sufficiently large ε. Thus, in this regime, the ratio between the two MSEs is at most $\frac{\frac{\Delta(\Delta+1)(2\Delta+1)}{6(\cosh(\varepsilon)-1)}}{\frac{1}{e^\varepsilon+(2\Delta-1)} \cdot \frac{\Delta(\Delta+1)(2\Delta+1)}{3}} = \frac{1+(2\Delta-1)e^{-\varepsilon}}{1-2e^{-\varepsilon}+e^{-2\varepsilon}}$. The RHS approaches 1 as $\varepsilon \to \infty$ as claimed. ◀

While the naive approach to sample from the MSDLap distribution requires sampling $O(\Delta)$ random variables, we show in Section 5 an efficient algorithm for the high ε regime.

3.2.1 Generalizing the MSDLap mechanism

We can also generalize the MSDLap mechanism to match the error in [18] for every setting of parameters Δ, ε. We state this below where $r \in \{0, \ldots, \Delta\}$ is a parameter so that it matches the "r" parameter in the discrete staircase mechanism in [18]. In our results henceforth, we assume that $\varepsilon \geq 2$ for simplicity; this 2 can be changed to any constant[5] but we keep it for simplicity of the distribution description.

▶ **Theorem 18.** *For any $\varepsilon \geq 2, \Delta \in \mathbb{N}$ and $r \in \{0, \ldots, \Delta\}$, there is an infinitely divisible discrete noise-addition mechanism that is ε-DP for sensitivity Δ and has MSE $O\left(r^2 + \frac{e^{-\varepsilon}\Delta^3}{r+1}\right)$.*

Plugging in $r = 0, \lceil e^{-\varepsilon/3}\Delta \rceil$ gives the following, which will be useful later in Section 6.

▶ **Corollary 19.** *For any $\varepsilon \geq 2$, there exists an infinitely divisible discrete noise-addition mechanism that is ε-DP for sensitivity Δ with MSE $O\left(\Delta^2 \min\{e^{-\varepsilon}\Delta, e^{-2\varepsilon/3}\}\right)$.*

We are now ready to prove Theorem 18. The rough idea is that, instead of using only the (ε, Δ)-MSDLap noise, we first add a scaled-up $(\varepsilon - 1, \Delta_0)$-MSDLap noise where $\Delta_0 < \Delta$. The scaling up leaves us with "holes" in the noise distribution. To fix this, we additionally add another DLap noise to "smoothen out the holes". This idea is formalized below.

[5] By changing the distribution of Y in the proof of Theorem 18 to $\mathrm{DLap}(c/r)$ for some larger $c > 1$.

Proof of Theorem 18. The case $r = 0$ follows from Theorem 16.

For $r \geq 1$, let $\Delta_0 = \lfloor \Delta/r \rfloor$. Let the noise distribution $\mathcal{D}$ be the distribution of $Z = r \cdot X + Y$ where $X \sim (\varepsilon - 1, \Delta_0)-\text{MSDLap}, Y \sim \text{DLap}(1/r)$ are independent. The infinite divisibility of $\mathcal{D}$ follows from the infinite divisibility of MSDLap, DLap and Observations 4 and 5.

(Privacy) From Lemma 8, it suffices to show $\mathrm{D}_\infty(\xi + Z \parallel Z) \leq \varepsilon$ for all $\xi \in \{-\Delta, \ldots, \Delta\}$. Let $i^* = \lfloor \xi/r \rfloor$ and $j^* = \xi - r \cdot i^*$. Note that $i^* \in [\Delta_0]$ and $j^* \in [r]$. From Lemmas 6 and 7,

$$\begin{aligned}
&\mathrm{D}_\infty(\xi + Z \parallel Z) = \mathrm{D}_\infty(r \cdot i^* + j^* + r \cdot X + Y \parallel r \cdot X + Y) \\
&\leq \mathrm{D}_\infty(r \cdot i^* + j^* + r \cdot X + Y \parallel r \cdot i^* + r \cdot X + Y) + \mathrm{D}_\infty(r \cdot i^* + r \cdot X + Y \parallel r \cdot X + Y) \\
&\leq \mathrm{D}_\infty(j^* + Y \parallel Y) + \mathrm{D}_\infty(i^* + X \parallel X) \leq 1 + (\varepsilon - 1) = \varepsilon,
\end{aligned}$$

where the last inequality follows from $Y \sim \text{DLap}(1/r)$ and $X \sim (\varepsilon - 1, \Delta_0)$-MSDLap (together with Theorem 16 and Lemma 8).

(Accuracy) MSE is $r^2 \cdot \mathrm{Var}(X) + \mathrm{Var}(Y) \leq O(r^2 \cdot \Delta_0^3 \cdot e^{-\varepsilon}) + O(r^2) = O(r^2 + e^{-\varepsilon}\Delta^3/r)$ ◀

▶ Remark 20. We can generalize the poof of Theorem 18 further by considering the $(\mathcal{D}_1, \mathcal{D}_2)$-generalized-multi-scale mechanism that adds noise from two distributions $\mathcal{D}_1, \mathcal{D}_2$ where

- $\mathcal{D}_1$-noise addition mechanism is ε_1-DP for $\Delta = 1$, and is added at multiple scales
- $\mathcal{D}_2$-noise addition mechanism is ε_2-DP for $\Delta = r$, and is used to "smoothen out the holes"

Specifically, the noise is $r \cdot \left(\sum_{i=1}^{\Delta_0} i \cdot X_i\right) + Y$ where $X_1, \ldots, X_{\Delta_0} \overset{\text{i.i.d.}}{\sim} \mathcal{D}_1, Y \sim \mathcal{D}_2$ are independent. The privacy proof proceeds identically to Theorem 18 yielding an $(\varepsilon_1 + \varepsilon_2)$-DP mechanism. This allows us to consider $\mathcal{D}_1, \mathcal{D}_2 \sim \text{GDL}$, which gives us the multi-scale version of the GDL mechanism; this noise distribution is additionally closed under summation.

We plot the MSE of our new mechanisms and the established baselines in Figure 1.

4 From Discrete to Continuous

We next show simple methods to transform discrete noises to continuous ones. The approach is similar to Theorem 18, except that we use Laplace noise to "smoothen out the holes".

▶ **Theorem 21.** *For $\varepsilon \geq 2$ and $\Delta > 0$, there exists a continuous infinitely divisible noise-addition mechanism that is ε-DP for sensitivity Δ with MSE $O(\Delta^2 \cdot e^{-2\varepsilon/3})$.*

Proof. By scaling, we may assume w.l.o.g. that $\Delta = 1$.

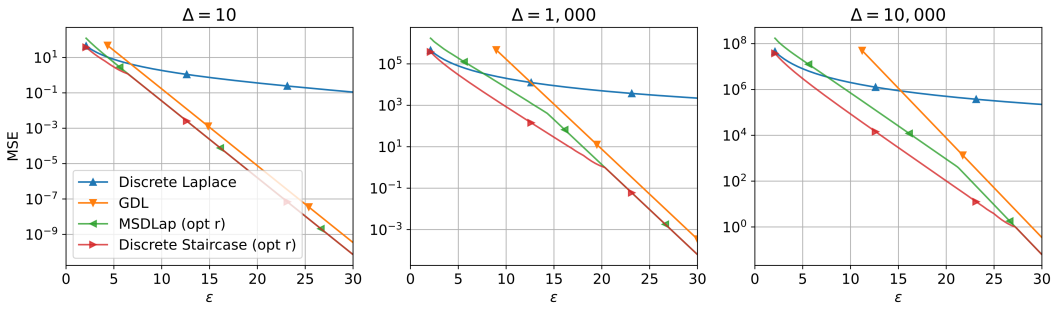

Figure 1 The MSE of the GDL mechanism (Theorem 15) and MSDLap mechanism (Theorem 16) with optimized r. We include the discrete Laplace and staircase (Section A.3) baselines. In the high ε regime our mechanisms closely track the MSE of the discrete staircase.

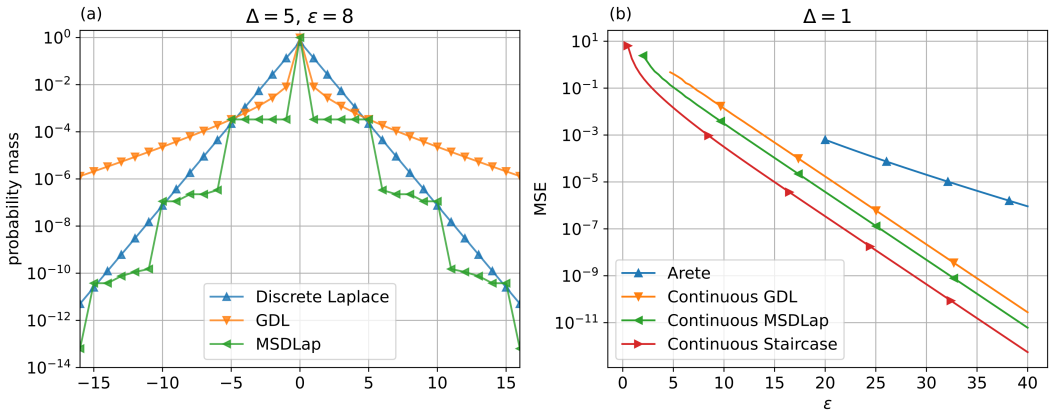

Figure 2 **(a)** The PMF of the GDL distribution parameterized by Theorem 15, the MSDLap distribution (Theorem 16), and the discrete Laplace distribution with $a = \varepsilon/\Delta$. GDL has a much sharper peak around 0, before flattening out and decreasing slower than discrete Laplace. MSDLap has a "staircase" shaped distribution with sharp drops at Δ-width intervals. Its PMF appears fully dominated by the discrete Laplace's, except at multiples of Δ. **(b)** The MSE of the continuous GDL (Theorem 15) and MSDLap (Theorem 16) after the continuous transformation of Theorem 21 is applied to them. We also plot the Arete [39, Lemma 3] and continuous staircase [18, eq. 51] mechanisms as baselines.

Let $\varepsilon_{\mathrm{d}} = \varepsilon - 1, \Delta_{\mathrm{d}} = \lceil e^{\varepsilon/3} \rceil$ and $r = \frac{1}{\Delta_{\mathrm{d}}}$. Let $\mathcal{D}_{\mathrm{d}}$ be any infinitely divisible discrete distribution such that the $\mathcal{D}_{\mathrm{d}}$-noise addition mechanism is ε_{d}-DP for sensitivity Δ_{d} with MSE $O(\Delta_{\mathrm{d}}^3 \cdot e^{-\varepsilon_{\mathrm{d}}})$. (Such a distribution exists due to Theorem 16.[6])

Let $\mathcal{D}$ be the distribution of $Z = r \cdot X + Y$ where $X \sim \mathcal{D}_{\mathrm{d}}, Y \sim \mathrm{Lap}\,(r/2)$ are independent. Since X, Y are infinitely divisible, Observations 4 and 5 imply that $\mathcal{D}$ is infinitely divisible.

(Privacy) From Lemma 8, it suffices to show $\mathrm{D}_\infty\,(\xi + Z \parallel Z) \le \varepsilon$ for all $\xi \in [-1, 1]$. Let i^* be the closest integer to ξ/r and let $j^* = \xi - r \cdot i^*$. Note that $i^* \in \{-\Delta_{\mathrm{d}}, \ldots, \Delta_{\mathrm{d}}\}$ and[7] $j^* \in [-r/2, r/2]$. From Lemmas 6 and 7, we have

$$\begin{aligned}
&\mathrm{D}_\infty\,(\xi + Z \parallel Z) = \mathrm{D}_\infty\,(r \cdot i^* + j^* + r \cdot X + Y \parallel r \cdot X + Y) \\
&\le \mathrm{D}_\infty\,(r \cdot i^* + j^* + r \cdot X + Y \parallel r \cdot i^* + r \cdot X + Y) + \mathrm{D}_\infty\,(r \cdot i^* + r \cdot X + Y \parallel r \cdot X + Y) \\
&\le \mathrm{D}_\infty\,(j^* + Y \parallel Y) + \mathrm{D}_\infty\,(i^* + X \parallel X) \le 1 + (\varepsilon - 1) = \varepsilon,
\end{aligned}$$

where the last inequality follows from $Y \sim \mathrm{Lap}(r/2)$ and X-noise addition mechanism is ε-DP for sensitivity Δ_{d} (and Lemma 8).

(Accuracy) MSE is $\mathrm{Var}\,(r \cdot X + Y) = r^2 \cdot \mathrm{Var}(X) + \mathrm{Var}(Y) \le O\,(\Delta_{\mathrm{d}} \cdot e^{-\varepsilon}) + O\left(\left(\frac{1}{\Delta_{\mathrm{d}}}\right)^2\right) \le O(e^{-2\varepsilon/3})$, where the last inequality is from our choice $\Delta_{\mathrm{d}} = \Theta(e^{\varepsilon/3})$. ◀

Similar to Remark 20, we may also consider $\mathcal{D}_{\mathrm{d}}$ apart from the MSDLap distribution, as long as it satisfies ε_{d}-DP for sensitivity Δ_{d}. In particular, we may use the GDL distribution. We plot the results of transforming the discrete mechanisms from Section 3 in Figure 2.

[6] GDL also meets the requirements, but only for a specific regime of ε. Specifically, Theorem 15 requires $\varepsilon_{\mathrm{d}} > 2 + \log(\Delta_{\mathrm{d}}) > 2 + \varepsilon/3$. For $\varepsilon_{\mathrm{d}} = \varepsilon - 1$, the continuous transformation of GDL is valid when $\varepsilon > 4.5$.

[7] Note that the choice of i^* that halves the maximum value of $|j^*|$ cannot be directly applied to Theorem 18. In that theorem, we take $\Delta_0 = \lfloor \Delta/r \rfloor$, while the approach here only works when $\Delta_0 \ge \Delta/r$.

5 Efficiently and exactly sampling from the distributed MSDLap

This section outlines an algorithm to efficiently sample from the MSDLap mechanism described in Theorem 16 in the distributed setting over n parties. Recall that the (ε, Δ)-MSDLap noise is defined as $\sum_{i=1}^{\Delta} i \cdot X_i$ where $X_1, \ldots, X_\Delta \overset{i.i.d.}{\sim} \text{DLap}(\varepsilon) = \text{NB}(1, 1-e^{-\varepsilon}) - \text{NB}(1, 1-e^{-\varepsilon})$. In other words, each of the n parties need to sample $\sum_{i=1}^{\Delta} i \cdot (U_i - V_i)$ where $U_1, \ldots, U_\Delta, V_1, \ldots, V_\Delta \overset{i.i.d.}{\sim} \text{NB}(1/n, 1-e^{-\varepsilon})$. Thus, the naive algorithm requires each party to sample from $k = 2\Delta$ negative binomial random variables. For large Δ, or for $\Delta = O(e^{\varepsilon}/3)$ as in Corollary 19, this may be computationally expensive.

Our algorithm resolves this issue, allowing us to sample from exponentially many (in ε) negative binomial random variables in time polynomial in ε.

▶ **Theorem 22.** *For input $k \in \mathbb{N}, r, \gamma \in \mathbb{Q}_{>0}$, and $p = e^{-\gamma}$, Algorithm 3 returns non-zero samples from k i.i.d. samples of $\text{NB}(r, 1-e^{-\varepsilon})$ and completes in $\widetilde{O}\left(1 + k \cdot r \cdot e^{-\varepsilon}\right)$ steps in expectation, where $\varepsilon = \log \frac{1}{1-e^{-\gamma}}$ and $\widetilde{O}$ hides polynomial factors in ε.*

We leverage the fact that in the high ε regime, most of these NB random variables will be 0. We re-frame the problem of sampling many negative binomials into two separate problems:

- Sampling from the *sum* of many i.i.d. NBs.
- Fairly *allocating* the result across each r.v.

While sampling from the sum of many NBs is simple on its face given their infinite divisibility, standard samplers for $\text{NB}(r, p)$ (e.g. [28]) take time linear in r which is not desirable. Below we will describe an algorithm (Algorithm 1) whose (expected) running time only scales with the mean of $\text{NB}(r, p)$, which is only $O(r \cdot e^{-\varepsilon})$ in our setting.

To fairly allocate across the random variables, we leverage the fact that the conditional distribution of the sequence of NB random variables given their sum follows the *Dirichlet multinomial* distribution denoted $\text{DirM}(n, \boldsymbol{\alpha})$ where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_k\}$ and $\alpha_0 = \sum_{i=1}^{k} \alpha_i$. Its PMF is $f_{\text{DirM}(n,\boldsymbol{\alpha})}(x) = \frac{\Gamma(\alpha_0)\Gamma(n+1)}{\Gamma(n+\alpha_0)} \prod_{i=1}^{k} \frac{\Gamma(x_i+\alpha_i)}{\Gamma(\alpha_i)\Gamma(x_i+1)}$.

▶ **Lemma 23** (e.g. [45, 42]). *Let $X = \{X_1, \ldots, X_k\}$ be such that $X_i \sim \text{NB}(\alpha_i, p)$ are independent. Let $T = \sum_{i=1}^{k} X_i$. Then the conditional distribution $X | T = t \sim \text{DirM}(t, \boldsymbol{\alpha})$.*

Our sampler can be implemented on a finite computer in the Word RAM model avoiding any real-arithmetic operations. Following [11, Section 5], we focus on the runtime in the *expected* number of arithmetic operations, which take only polynomial time in the bit complexity of the parameters. We make the following assumptions on the availability of sampling primitives requiring only $O(1)$ arithmetic operations in expectation:

- A uniform sampler to draw $D \in \{1, 2, \ldots, d\}$ for $d \in \mathbb{N}$.
- A $\text{Ber}(n/d)$ sampler for $n, d \in \mathbb{N}$, which trivially follows from the uniform sampler.
- A $\text{Geo}(1 - e^{-\gamma})$ sampler for $\gamma \in \mathbb{Q}$ from [11].

Finally, we assume a map data structure with $O(1)$ accesses and updates in expectation, and a vector data structure with $O(1)$ random access and append operations.

Algorithm 1 Fast NB sampler for $p > \frac{1}{2}$.

Input: $r \in \mathbb{Q}_{>0}$, $p = e^{-\gamma}$ for $\gamma \in \mathbb{Q}_{>0}$
Output: A sample from $\mathrm{NB}(r, p)$

loop ▷ $\mathrm{NB}(\lceil r \rceil, p)$ rejection sampler
  Sample $w \leftarrow \textsc{IntSample}(\lceil r \rceil, p)$
  $A_w \leftarrow (r)_w / (\lceil r \rceil)_w$
  Sample $accept \leftarrow \mathrm{Ber}(A_w)$
  if $accept$ **then return** w
procedure $\textsc{IntSample}(r,\ p = e^{-\gamma})$
  $failures \leftarrow 0, successes \leftarrow 0$
  loop ▷ Use [11] for Geo sampling
    Sample $s \leftarrow \mathrm{Geo}(1 - e^{-\gamma})$
    $successes \leftarrow successes + s$
    if $successes \geq r$ **then**
      return $failures$
    $failures \leftarrow failures + 1$

Algorithm 2 Dirichlet multinomial sampler.

Input: $n \in \mathbb{N}$, $k \in \mathbb{N}$, $\alpha = a/b \in \mathbb{Q}_{>0}$
Output: A sample from $\mathrm{DirM}(n, \{\alpha, \ldots, \alpha\})$, encoded as a sparse map from variate index to count, with zero variates removed.

$picked \leftarrow []$ ▷ Vector data structure
$initialsize \leftarrow k \cdot a$
for i from 0 to $n - 1$ **do**
  $size \leftarrow initialsize + i \cdot b$
  Sample $U \leftarrow \mathrm{Unif}(\{1, \ldots, size\})$
  if $U < initialsize$ **then**
    $idx \leftarrow \lceil U/a \rceil$
    Append idx to $picked$
  else
    $idx \leftarrow \lceil (U - initialsize)/b \rceil$
    Append $picked[idx]$ to $picked$
$counter \leftarrow \{\}$ ▷ Map data structure
for p in $picked$ **do**
  if p in $counter$ **then**
    $counter[p] \leftarrow counter[p] + 1$
  else
    $counter[p] \leftarrow 1$
return $counter$

Algorithm 3 Sparse NB sampler.

Input: $k, r \in \mathbb{N}$, $p = e^{-\gamma}$ for $\gamma \in \mathbb{Q}_{>0}$
Output: Non-zero samples $X_1, \ldots, X_k$ where $X_i \sim NB(r, p)$

Sample $T \leftarrow \mathrm{NB}(k \cdot r, p)$
Sample $counter \leftarrow \mathrm{DirM}(T, k, r)$
return $counter$

Our Algorithm 3 is straightforward. It consists of

1. Sampling from $\mathrm{NB}(r, p)$ with Algorithm 1 to learn the sum of all the terms, handling rational values of r following the approach in [28] using a simple rejection sampler.
2. Sampling from the Dirichlet multinomial distribution with Algorithm 2, which uses a version of the Pólya urn process [31] modified to handle rational fractions of balls. We sparsely encode the output to avoid storing zero entries, as MSDLap sampling only requires summing non-zero random variables.

▶ **Proposition 24.** *For input $r, \gamma \in \mathbb{Q}_{>0}$ and $p = e^{-\gamma}$, Algorithm 1 returns one sample from $NB(r, 1 - e^{-\varepsilon})$ and completes in $\widetilde{O}\left(1 + r \cdot e^{-\varepsilon}\right)$ arithmetic operations in expectation, where $\varepsilon = \log \frac{1}{1-e^{-\gamma}}$ and $\widetilde{O}$ hides polynomial factors in ε.*

Proof. We begin by analyzing the case $r \in \mathbb{N}$, i.e., $\textsc{IntSample}$ subroutine. Let K be the r.v. describing the output, and X_i be the ith $\mathrm{Geo}(1-p)$ r.v. Denote $Z \sim \sum_{i=1}^{k} X_i \sim \mathrm{NB}(k, 1-p)$. Note that $\mathbb{P}[K = 0] = \mathbb{P}[X_1 \geq r] = p^r = f_{\mathrm{NB}(r,p)}(0)$. For $k \geq 1$,

$$\mathbb{P}[K = k] = \mathbb{P}[Z < r \leq Z + X_{k+1}] = \sum_{z=0}^{r-1} \sum_{x=r-z}^{\infty} f_{\mathrm{NB}(k,1-p)}(z) f_{\mathrm{Geo}(1-p)}(x)$$
$$= \sum_{z=0}^{r-1} \sum_{x=r-z}^{\infty} (1-p)^k p^z \binom{k+z-1}{z} \cdot (1-p) p^x$$

$$= (1-p)^{k+1} \sum_{z=0}^{r-1} \binom{k+z-1}{k-1} \frac{p^r}{1-p}$$

$$\text{Lemma 3} = (1-p)^k p^r \binom{k+r-1}{k} = f_{\mathrm{NB}(r,p)}(k).$$

Thus, the output K follows the $\mathrm{NB}(r,p)$ distribution as desired.

The expected number of iterations of the loop is exactly $1+\mathbb{E}[\mathrm{NB}(r,p)]$. Each iteration of the loop takes expected $\widetilde{O}(1)$ time, as the geometric sampler requires arithmetic operations polynomial in the bit complexity of γ (which is $O(\varepsilon)$).

Next we analyze the outer loop which handles $r \in \mathbb{Q}_{>0}$ using the accept-reject approach in [28], ensuring that in each iteration, for a proposed sample W and acceptance event A:

$$\mathbb{P}[W = w \wedge A] = \mathbb{P}[A|W=w]\,\mathbb{P}[W=w] = \frac{(r)_w}{(\lceil r \rceil)_w} f_{\mathrm{NB}(\lceil r \rceil,p)}(w) = p^{r-\lceil r \rceil} \cdot f_{\mathrm{NB}(r,p)}(w).$$

This implies that the output follows $\mathrm{NB}(r,p)$ distribution as desired, and that $\mathbb{P}[A] = p^{r-\lceil r \rceil}$. The latter in turn implies that the number of trials follows a geometric distribution with success probability $p^{\lceil r \rceil - r}$. Therefore the expected number of trials is $O(1/p^{\lceil r \rceil - r}) = O(1/p)$, and the result follows as $\widetilde{O}\left(\frac{1}{1-e^{-\varepsilon}}\left(1+\mathbb{E}[\mathrm{NB}(r,1-e^{-\varepsilon})]\right)\right) = \widetilde{O}(1+r\cdot e^{-\varepsilon})$. ◀

▶ **Proposition 25.** *For input $n,k \in \mathbb{N}$, and $\alpha \in \mathbb{Q}$, Algorithm 2 returns one sample from $\mathrm{DirM}(n,\{\alpha,\cdots,\alpha\})$ and requires $O(n)$ arithmetic operations in expectation.*

Proof. We first map Algorithm 2 to the Pólya urn model:

- To initialize the urn contents, a balls are added to the urn for each of the k colors.
- The algorithm proceeds over n steps. At each step, a ball is chosen uniformly at random from the urn. When a ball is selected, it is replaced, along with b other balls of that color.
- For each color, the algorithm returns the count of how many times that color was chosen.

Algorithm 2 implements this, as *idx* maps to the *idx*th color c_{idx}, and *picked* is the list of selected colors. Note that the return value is *sparsely encoded* to avoid storing zero entries, or colors that have never been picked. For purposes of the proof of correctness we assume an unrolled output of *counter* equal to $X = \{x_1,\cdots,x_k\}$, where each x_i counts the number of times the c_i color was picked. This can be obtained with a simple post-processing step.

Let S_m be the color of the ball chosen at iteration m. The probability that $S_m = c$ given the previous draws is: $\mathbb{P}[S_m = c | S_1 = s_1,\ldots,S_{m-1}=s_{m-1}] = \frac{a+\sum_{i=1}^{m-1} b\cdot s_i}{k\cdot a+b(m-1)} = \frac{a+b\cdot z}{k\cdot a+b(m-1)}$, where $z = \sum_{i=1}^{m-1} \mathbb{1}(s_i = c)$ denote the number of previous draws of color c.

Note that S_m only depends about the current state of the urn, not the order in which balls are picked. Similarly, note that the denominator of the fraction is independent of any information about z. From this we can show that the probability of seeing any one *particular sequence* $S = \{s_1,\cdots,s_n\}$ which has color counts $X = \{x_1,\ldots,x_k\}$ is

$$\mathbb{P}[S_1 = s_1,\cdots,S_n = s_n] = \left(\prod_{i=1}^{n} \frac{1}{k\cdot a + b(i-1)}\right) \prod_{j=1}^{k}\prod_{t=1}^{x_i} a + b(t-1)$$

$$= \frac{\prod_{j=1}^{k} b^{x_i}(a/b)_{x_i}}{b^n (k\cdot a/b)_n} = \frac{\Gamma(k\cdot\alpha)}{\Gamma(n+k\cdot\alpha)} \prod_{j=1}^{k} \frac{\Gamma(\alpha+x_i)}{\Gamma(\alpha)}.$$

Since there are $\binom{n}{x_1,x_2,\ldots,x_k} = \frac{\Gamma(n+1)}{\prod_{i=1}^{k}\Gamma(x_i+1)}$ sequences S with color count x, we have $\mathbb{P}[X=x] = \frac{\Gamma(k\cdot\alpha)\Gamma(n+1)}{\Gamma(n+k\cdot\alpha)} \prod_{j=1}^{k} \frac{\Gamma(\alpha+x_i)}{\Gamma(x_i+1)\Gamma(\alpha)} = f_{\mathrm{DirM}(n,\{\alpha,\ldots,\alpha\})}(x)$, proving the correctness.

For the run time analysis, both loops in the algorithm iterate n times, and each require only constant arithmetic operations in expectation, assuming $O(1)$ map and vector operations. ◀

Lemma 23, Proposition 24, and Proposition 25 immediately show Theorem 22.

6 Order optimal MSE in the multi-message shuffle model

We next apply the noises from Section 3 to protocols in the shuffle model of DP.

First, we recall the definition of the shuffle model [13, 16]. An m-message protocol in the shuffle model consists of a *randomizer* $\mathcal{R} : \mathcal{X} \to \mathcal{Y}^m$ where $\mathcal{Y}$ denote the set of all possible messages and an *analyzer* $\mathcal{A} : \mathcal{Y}^{nm} \to \mathcal{O}$ where $\mathcal{O}$ denotes the output domain. In the shuffle model, the analyst does not see the output of each randomizer directly but only the randomly shuffled messages. We write $\mathcal{S}_{\mathcal{R}}(x_1, \ldots, x_n)$ to denote the output of randomly shuffling nm (random) messages produced by $\mathcal{R}(x_1), \ldots, \mathcal{R}(x_n)$ for $x_1, \ldots, x_n \in \mathcal{X}$. The shuffle model required that these shuffled messages have to satisfy DP, as stated more formally below.

▶ **Definition 26** ([13, 16]). *An m-message protocol $(\mathcal{R}, \mathcal{A})$ is (ε, δ)-shuffle-DP if, for every $x, x' \in \mathcal{X}^n$ differing in a single entry, $\Pr[\mathcal{S}_{\mathcal{R}}(x) \in S] \le e^{\varepsilon} \cdot \Pr[\mathcal{S}_{\mathcal{R}}(x') \in S] + \delta$ for all $S \subseteq \mathcal{Y}^{nm}$.*

In the real summation problem, each x_i is a real number in $[0, 1]$ and the goal is to compute $\sum_{i\in[n]} x_i$. Our main result of this section can be stated as follows:

▶ **Theorem 27.** *For $\varepsilon \ge 2, \delta \in (0, 1/n)$, there is an (ε, δ)-shuffle-DP, $O\left(\frac{\varepsilon+\log(1/\delta)}{\log(n)}\right)$-message protocol for real summation with MSE $O\left(e^{-2\varepsilon/3}\right)$ where each message is $O\left(\varepsilon + \log n\right)$ bit.*

Prior to this work, the best known protocol has MSE of $O(1/\varepsilon^2)$ (even for large ε) [7, 24, 23][8] and thus our result provides a significant improvement in the large ε regime.

To prove this result, it is convenient to define the $\mathbb{Z}_q$-summation problem: The input $x_1, \ldots, x_n$ belongs to $\mathbb{Z}_q$ and the goal is to compute $\sum_{i\in[n]} \mathbb{Z}_q$. Similar to before, an m-message protocol consists of a randomizer $\mathcal{R}$ and an analyzer $\mathcal{A}$. The protocol is *exact* if the analyzer always output the correct answer. The protocol is *σ-secure* if, for all $x, x' \in \mathbb{Z}_q^n$ such that $\sum_{i\in[n]} x_i = \sum_{i\in[n]} x'_i$, we have $\mathrm{D}_{\mathrm{TV}}(\mathcal{S}_{\mathcal{R}}(x), \mathcal{S}_{\mathcal{R}}(x')) \le 2^{-\sigma}$. Building on the "split and mix" protocol of [30], Balle at al. [7] and Ghazi et al. [24] gave the following protocol:

▶ **Theorem 28** ([7, 24]). *For $\sigma \in (0, 1/2)$ and $q \in \mathbb{N}$, there is an σ-secure $O\left(1 + \frac{\sigma}{\log n}\right)$-message protocol for $\mathbb{Z}_q$-summation in the shuffle model where each message is from $\mathbb{Z}_q$.*

Balle et al. [7] presented an elegant method to translate a secure $\mathbb{Z}_q$-summation protocol to a shuffle-DP real summation protocol. Below, we provide a slight generalization and improvement[9] of their result, which eventually allows us to achieve Theorem 27.

▶ **Lemma 29** (generalization of [7] Lemma 5.2). *Suppose that, for some $\varepsilon > 0, \Delta \in \mathbb{N}$, there is a zero-mean discrete infinitely divisible distribution $\mathcal{D}$ such that the $\mathcal{D}$-noise addition mechanism is ε-DP for sensitivity Δ. Furthermore, suppose that, for some $q, n \in \mathbb{N}$ with*

[8] Pagh and Stausholm [39, Corollary 23] claim that their result implies a protocol with absolute error $\frac{1}{e^{\Omega(\varepsilon)}-1}$, but, to our knowledge, this is not the case. See the discussion at the end of this section.

[9] Originally, their analysis has an additional error term due to "overflow / underflow". We observe that this is in fact unnecessary, which reduces the claimed error and also simplifies the analysis.

$q \geq 2\Delta n$ and $\sigma \in (0, 1/2)$, there exists an n-party σ-secure m-message exact $\mathbb{Z}_q$-summation protocol in the shuffle model. Then, there is an m-message $(\varepsilon, (1+e^{\varepsilon})2^{-\sigma})$-shuffle-DP protocol for real summation with MSE $\frac{\mathrm{Var}(\mathcal{D})}{\Delta^2} + \frac{n}{4\Delta^2}$.

Moreover, the message length is the same as in the $\mathbb{Z}_q$-summation protocol.

Proof. Let $\Pi = (\mathcal{R}_\Pi, \mathcal{A}_\Pi)$ be the σ-secure exact $\mathbb{Z}_q$-summation protocol. Our protocol $\mathcal{P} = (\mathcal{R}_\Pi \circ \mathcal{R}, \mathcal{A} \circ \mathcal{A}_\Pi)$ where $\mathcal{R}, \mathcal{A}$ are defined as follows[10]:

1. $\mathcal{R} : [0,1] \to \mathbb{Z}_q$ on input x_i works by first computing a randomized encoding $y_i = \begin{cases} 1 + \lfloor \Delta x_i \rfloor & \text{w.p. } \Delta x_i - \lfloor \Delta x_i \rfloor \\ \lfloor \Delta x_i \rfloor & \text{w.p. } 1 - (\Delta x_i - \lfloor \Delta x_i \rfloor) \end{cases}$. It then outputs $(y_i + Z_i) \mod q$ where $Z_i \sim \mathcal{D}_{/n}$.
2. $\mathcal{A}$ decodes the result r by returning $r' = \begin{cases} r/\Delta & \text{if } 0 \leq r \leq n\Delta, \\ n & \text{if } n\Delta + 1 \leq r \leq 2n\Delta, \\ 0 & \text{otherwise.} \end{cases}$

(Privacy) The proof of privacy proceeds identically to [7, Lemma 5.2] and is omitted here.

(Accuracy) Since Π is an exact $\mathbb{Z}_q$-summation protocol, its output r is exactly equal to $\tilde{u} \mod q$ where $\tilde{u} = \sum_{i\in[n]}(y_i + z_i) = Z + \sum_{i\in[n]} y_i$ where $Z = z_1 + \cdots + z_n$ is distributed as $\mathcal{D}$.

Let $u = \sum_{i\in[n]} x_i$ be the true (unnoised) sum. The first step in our accuracy analysis is a claim[11] that $|r' - u| \leq |\tilde{u}/\Delta - u|$. To see that this is true, let us consider the following cases:

- Case I: $\tilde{u} \notin (-n\Delta, 2n\Delta)$. In this case, $|\tilde{u}/\Delta - u| \geq n \geq |r' - u|$.
- Case II: $\tilde{u} \in [0, n\Delta]$. We have $r' = \tilde{u}/\Delta$ and thus the inequality holds as an equality.
- Case III: $\tilde{u} \in (-n\Delta, 0)$. We set $r' = 0$. Thus, $|\tilde{u}/\Delta - u| = u - \tilde{u}/\Delta \geq u - r' = |r' - u|$
- Case IV: $\tilde{u} \in (n\Delta, 2n\Delta)$. We set $r' = n$. Thus, $|\tilde{u}/\Delta - u| = \tilde{u}/\Delta - u \geq r' - u = |r' - u|$.

Thus, in all cases, we have $|r' - u| \leq |\tilde{u}/\Delta - u|$. Therefore, the MSE is at most

$$\mathbb{E}\left[(\tilde{u}/\Delta - u)^2\right] = \mathbb{E}\left[(Z/\Delta)^2\right] + \sum_{i\in[n]} \mathbb{E}\left[(y_i/\Delta - x_i)^2\right] \leq \frac{\mathrm{Var}(\mathcal{D})}{\Delta^2} + \frac{n}{4\Delta^2}. \quad \blacktriangleleft$$

We can now prove Theorem 27 by plugging our noise to the above lemma.

Proof of Theorem 27. Let $\Delta = \left\lceil e^{\varepsilon/3}\sqrt{n} \right\rceil, q = 2n\Delta$ and $\sigma = \log_2\left(\frac{e^{\varepsilon}+1}{\delta}\right)$. From Corollary 19, there is a zero-mean discrete infinitely divisible distribution $\mathcal{D}$ such that $\mathcal{D}$-noise addition mechanism is ε-DP for sensitivity Δ where $\mathrm{Var}(\mathcal{D}) \leq O(\Delta^2 \cdot e^{-2\varepsilon/3})$. Furthermore, Theorem 28 ensures that there exists an n-party σ-secure m-message exact $\mathbb{Z}_q$-summation protocol where $m = O\left(1 + \frac{\sigma}{\log n}\right) = O\left(1 + \frac{\varepsilon + \log(1/\delta)}{\log n}\right)$. Thus, Lemma 29 implies that there is an (ε, δ)-shuffle-DP m-message protocol for real summation where each message is of length $O(\log q) = O(\varepsilon + \log n)$ bits and MSE $\frac{\mathrm{Var}(\mathcal{D})}{\Delta^2} + \frac{n}{4\Delta^2} \leq O(e^{-2\varepsilon/3})$. $\blacktriangleleft$

In the above proof, we use the noise from Corollary 19, which is a modified version of MSDLap. We note that the continuous noises in Section 4 or [39] cannot be used here, as the protocol must round each contribution to a finite group $\mathbb{Z}_q$ prior to summing. Neither the privacy nor the infinite divisibility of the resulting sum distribution is clear in this case.

[10] Note that the final protocol $\mathcal{P}$ is done by composing the randomizer $\mathcal{R}$ / analyzer $\mathcal{A}$ with those of Π.
[11] This claim is indeed our improvement over the error analysis of [7, Lemma 5.2].

7 Conclusion

This work closes the utility gap for infinitely divisible mechanisms in the high ε pure DP regime. We find no separation in either the discrete or continuous settings by restricting the private mechanism to infinitely divisible noise addition. The "staircase-like" MSE of both GDL and MSDLap in the low-privacy regime make them a natural replacement for the staircase mechanism in the discrete distributed setting, and we hope the results introduced here can be of practical value. We show one such practical application by extending the [30] "split and mix" protocol under shuffle DP, resolving the open question posed in [23].

Beyond improving utility, we believe GDL is of independent interest for distributed private mechanism design due to the fact that it is closed under summation. This makes it well-suited for cases where "smooth" privacy guarantees are needed for multiple outcomes, or for a single deployed system (like a secure aggregation MPC protocol over n clients e.g. [9]) where different people can make *different* assumptions about what the honest fraction of the involved clients are. Additionally, this property is useful in *post-hoc* privacy loss analysis when honesty assumptions are broken in a production system, and where otherwise an analyst must resort to numerical approximation of realized privacy loss.

In the continuous setting, our continuous transformations of all of the mechanisms in Section 3 outperform the existing Arete mechanism (Figure 2). We also note a strong resemblance between the Arete and the GDL distributions, as the NB distribution converges to the gamma distribution under certain conditions. See Section A.4 for more detail.

Finally, we hope our optimized MSDLap sampler Algorithm 3 or its constituent parts can be useful in any context where sparse, exact multivariate NB random generation is needed, or even where a general $\mathrm{NB}(r, p)$ sampler needs to be sublinear in r (Algorithm 1). For multivariate sampling when p is very close to 1, our approach should be a large improvement over standard methods. We note that Algorithm 2 can be extended in a straightforward way (due to Lemma 23) to support NB variates with different r values. The only change to the algorithm is initializing the urn with varying numbers of balls.

Open questions. In preparing this paper, we studied the Arete mechanism [39] extensively and are convinced it can also achieve the order optimal MSE of $O\left(\Delta^2 e^{-2\varepsilon/3}\right)$, but the formal proof eluded us. We present the following conjecture[12] as an open question:

▶ **Conjecture 30.** *Let $\Delta \geq 1$, $\varepsilon > 10$. The $\mathrm{Arete}(\alpha, \theta, \lambda)$-noise addition mechanism with $\alpha = e^{-2\varepsilon/3}$, $\theta = (1+t)\frac{1}{\log 2}$, and $\lambda = (1+t)e^{-\varepsilon/3}$ where $t = o(1)$ is ε-DP for sensitivity Δ and has MSE $O\left(\Delta^2 e^{-2\varepsilon/3}\right)$.*

Our work also raises the following questions:

- How close to the optimal MSE *including constants* can be achieved in the for continuous and infinite divisible mechanisms? While we match the constants of the discrete staircase for large ε, there is still a sizable gap in the continuous case.
- Can the *number of messages* required for our shuffle-DP protocol in Theorem 27 be improved? Recall that we use the approach of [7] to translate a secure summation protocol to a shuffle-DP one. In fact, there is a more direct approach by [23] that achieves a smaller number of messages. Since their proof also uses infinitely divisible noises, it is plausible that our noise distribution can be used there. However, their proof is considerably more involved compared to [7] and does not use the noise distributions in a black-box manner.

[12] Even if this conjecture were proven, the constant factors for the Arete's MSE still underperform the continuous multi-scale discrete Laplace.

References

1 John M Abowd. The US Census Bureau adopts differential privacy. In *KDD*, pages 2867–2867, 2018.

2 Naman Agarwal, Peter Kairouz, and Ziyu Liu. The skellam mechanism for differentially private federated learning. *NeurIPS*, pages 5052–5064, 2021.

3 Horst Alzer and Graham Jameson. A harmonic mean inequality for the digamma function and related results. *Rendiconti del Seminario Matematico della Università di Padova*, 137:203–209, 2017.

4 Apple Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 2017.

5 Andreas Athanasiou, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Enhancing metric privacy with a shuffler. In *PETS 2025 - 25th Privacy Enhancing Technologies Symposium*, 2025.

6 Eugene Bagdasaryan, Peter Kairouz, Stefan Mellem, Adrià Gascón, Kallista Bonawitz, Deborah Estrin, and Marco Gruteser. Towards sparse federated analytics: Location heatmaps under distributed differential privacy with secure aggregation. *Proceedings on Privacy Enhancing Technologies*, 4:162–182, 2022. doi:10.56553/POPETS-2022-0104.

7 Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Private summation in the multi-message shuffle model. In *CCS*, pages 657–676, 2020. doi:10.1145/3372297.3417242.

8 James Henry Bell, Kallista A. Bonawitz, Adrià Gascón, Tancrède Lepoint, and Mariana Raykova. Secure single-server aggregation with (poly)logarithmic overhead. In *CCS*, pages 1253–1269, 2020. doi:10.1145/3372297.3417885.

9 Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *CCS*, pages 1175–1191, New York, NY, USA, 2017. doi:10.1145/3133956.3133982.

10 Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *TCC*, pages 635–658, 2016. doi:10.1007/978-3-662-53641-4_24.

11 Clément L Canonne, Gautam Kamath, and Thomas Steinke. The discrete gaussian for differential privacy. *NeurIPS*, pages 15676–15688, 2020.

12 Clément L. Canonne and Hongyi Lyu. Uniformity testing in the shuffle model: Simpler, better, faster. In *SOSA*, pages 182–202, 2022. doi:10.1137/1.9781611977066.13.

13 Albert Cheu, Adam D. Smith, Jonathan R. Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *EUROCRYPT*, pages 375–403, 2019. doi:10.1007/978-3-030-17653-2_13.

14 Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006. doi:10.1007/11761679_29.

15 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006. doi:10.1007/11681878_14.

16 Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *SODA*, pages 2468–2479, 2019. doi:10.1137/1.9781611975482.151.

17 Quan Geng, Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The staircase mechanism in differential privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1176–1184, 2015. doi:10.1109/JSTSP.2015.2425831.

18 Quan Geng and Pramod Viswanath. The optimal mechanism in differential privacy. In *ISIT*, pages 2371–2375, 2014. doi:10.1109/ISIT.2014.6875258.

19 Quan Geng and Pramod Viswanath. The optimal noise-adding mechanism in differential privacy. *IEEE Transactions on Information Theory*, 62(2):925–951, 2016. doi:10.1109/TIT.2015.2504967.

20 Badih Ghazi, Noah Golowich, Ravi Kumar, Rasmus Pagh, and Ameya Velingker. On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy. In *EUROCRYPT*, pages 463–488, 2021. doi:10.1007/978-3-030-77883-5_16.

21 Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Pure-dp aggregation in the shuffle model: Error-optimal and communication-efficient. In *ITC*, pages 4:1–4:13, 2024. doi:10.4230/LIPICS.ITC.2024.4.

22 Badih Ghazi, Ravi Kumar, Pasin Manurangsi, and Rasmus Pagh. Private counting from anonymous messages: Near-optimal accuracy with vanishing communication overhead. In *ICML*, pages 3505–3514, 2020. URL: http://proceedings.mlr.press/v119/ghazi20a.html.

23 Badih Ghazi, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Amer Sinha. Differentially private aggregation in the shuffle model: Almost central accuracy in almost a single message. In *ICML*, pages 3692–3701, 2021. URL: https://proceedings.mlr.press/v139/ghazi21a.html.

24 Badih Ghazi, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Private aggregation from fewer anonymous messages. In Anne Canteaut and Yuval Ishai, editors, *EUROCRYPT*, pages 798–827, 2020. doi:10.1007/978-3-030-45724-2_27.

25 Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM J. Comput.*, 41(6):1673–1693, 2012. doi:10.1137/09076828X.

26 Slawomir Goryczka and Li Xiong. A comprehensive comparison of multiparty secure additions with differential privacy. *IEEE transactions on dependable and secure computing*, 14(5):463–477, 2015.

27 Charlie Harrison. Distributed noise addition and infinite divisibility. https://charlieharrison.xyz/2024/08/05/distributed-noise.html, 2024. Accessed: 2025-02-11.

28 Creighton Heaukulani and Daniel M Roy. Black-box constructions for exchangeable sequences of random multisets. *arXiv preprint arXiv:1908.06349*, 2019.

29 Wolfram Research, Inc. Mathematica, Version 12.0, 2019. Champaign, IL, 2019.

30 Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Cryptography from anonymity. In *FOCS*, pages 239–248, USA, 2006. doi:10.1109/FOCS.2006.25.

31 Norman L. Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Discrete Multivariate Distributions*, chapter 40, pages 200–203. Wiley, 1997.

32 Charles H Jones. Generalized hockey stick identities and n-dimensional blockwalking. *The Fibonacci Quarterly*, 34(3):280–288, 1996.

33 Peter Kairouz, Ziyu Liu, and Thomas Steinke. The distributed discrete gaussian mechanism for federated learning with secure aggregation. In *ICML*, pages 5201–5212, 2021. URL: http://proceedings.mlr.press/v139/kairouz21a.html.

34 Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu. Practical and private (deep) learning without sampling or shuffling. In *ICML*, pages 5213–5225, 2021. URL: http://proceedings.mlr.press/v139/kairouz21b.html.

35 Seetha Lekshmi V and Simi Sebastian. A skewed generalized discrete laplace distribution. *IJMSI*, 2:95–102, 2014.

36 Ilya Mironov. On significance of the least significant bits for differential privacy. In *CCS*, pages 650–661, 2012. doi:10.1145/2382196.2382264.

37 Ilya Mironov. Rényi Differential Privacy. In *CSF*, pages 263–275, 2017. doi:10.1109/CSF.2017.11.

38 F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. V. Saunders B. R. Mille and, H. S. Cohl, and eds. M. A. McClain. *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/15.2, Release 1.2.2 of 2024-09-15, 2024. URL: https://dlmf.nist.gov/15.2.

39 Rasmus Pagh and Nina Mesing Stausholm. Infinitely divisible noise in the low privacy regime. In *ALT*, pages 881–909, 2022. URL: https://proceedings.mlr.press/v167/pagh22a.html.

40 Feng Qi. Bounds for the ratio of two gamma functions. *Journal of Inequalities and Applications*, 2010:1–84, 2010.

41 Aaron Schein, Zhiwei Steven Wu, Alexandra Schofield, Mingyuan Zhou, and Hanna Wallach. Locally private bayesian inference for count models. In *ICML*, pages 5638–5648, 2019. URL: http://proceedings.mlr.press/v97/schein19a.html.

42 F William Townes. Review of probability distributions for modeling count data. *arXiv preprint arXiv:2001.04343*, 2020.

43 Filipp Valovich and Francesco Alda. Computational differential privacy from lattice-based cryptography. In *International Conference on Number-Theoretic Methods in Cryptology*, pages 121–141, 2017.

44 Zheng Xu, Yanxiang Zhang, Galen Andrew, Christopher A. Choquette-Choo, Peter Kairouz, H. Brendan McMahan, Jesse Rosenstock, and Yuanbo Zhang. Federated learning of gboard language models with differential privacy. In *ACL: Industry Track*, pages 629–639, 2023. doi:10.18653/V1/2023.ACL-INDUSTRY.60.

45 Mingyuan Zhou. Nonparametric bayesian negative binomial factor analysis. *Bayesian Analysis*, 13, April 2016. doi:10.1214/17-BA1070.

A Appendix

A.1 Deferred Proof of Lemma 12

Before we prove Lemma 12, it will be convenient to state the following simple lemma.

▶ **Lemma 31.** *Let $f(x)$ be log convex on the interval $[a,b]$. Then for any $x \in [a,b]$ and $\Delta \in \mathbb{R}^+$ such that $x+\Delta \in [a,b]$: $\frac{f(0)}{f(\Delta)} \geq \frac{f(x)}{f(x+\Delta)}$.*

Proof. From log-convexity, we have $f(0)^{\frac{\Delta}{x+\Delta}} f(x+\Delta)^{\frac{x}{x+\Delta}} \geq f(x)$, and $f(0)^{\frac{x}{x+\Delta}} f(x+\Delta)^{\frac{\Delta}{x+\Delta}} \geq f(\Delta)$. Multiplying the two yields the claimed inequality. ◀

We are now ready to prove Lemma 12.

Proof of Lemma 12. Assume w.l.o.g. that $x \leq x'$. There are three cases to consider.

1. $x < 0$ and $x' < 0$. We have $\frac{f(x)}{f(x')} \leq \frac{f(-x)}{f(-x')}$.
2. $x < 0$ and $'x \geq 0$. We have $\frac{f(x)}{f(x')} \leq \frac{f(0)}{f(x')}$
3. $x \geq 0$ and $x' \geq 0$. By f decreasing on $[0,\infty)$, we have $\frac{f(x)}{f(x')} \leq \frac{f(x)}{f(x+\Delta)}$. Applying Lemma 31 concludes the proof. ◀

A.2 Alternative simplified GDL privacy bound

In this section we will outline a tighter version of Corollary 14, which only well-approximates the privacy loss in the small β regime. This bound well-approximates the privacy loss in all β regimes, at the cost of some added complexity in the expression.

▶ **Corollary 32** (to Theorem 13). *For any $\Delta \in N, a > 0, \beta \in (0,1)$, the GDL$(\beta, a)$-noise addition mechanism is ε-DP for sensitivity Δ where $\varepsilon \leq a\Delta + (1-\beta)\log(\beta+\Delta) + \log(\Gamma(\beta))$*

Proof. The proof is identical to Corollary 14, except in the last step where we use the following Wendel's double inequality [40, eq. 2.6]:

$$\log\left(\frac{\Gamma(\Delta+1)\Gamma(\beta)}{\Gamma(\beta+\Delta)}\right) \leq \log\left(\Gamma(\beta)(\Delta+\beta)^{1-\beta}\right) = (1-\beta)\log(\beta+\Delta) + \log(\Gamma(\beta)).$$ ◀

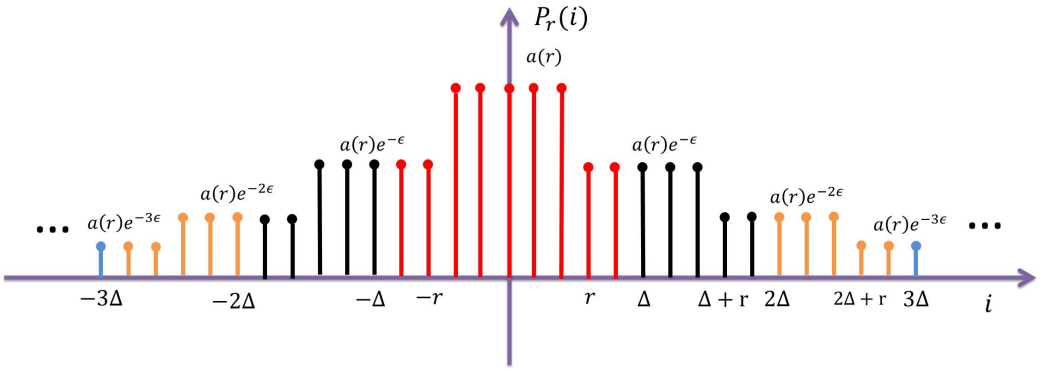

Figure 3 The Discrete Staircase PMF from [18, Figure 5].

A.3 Analytical variance of the discrete staircase distribution

In this section we derive the analytical variance of the discrete staircase using the Mathematica software [29], for the purposes of generating Figure 1.

▶ **Definition 33** ([18]). *The discrete staircase distribution with parameters* $1 \leq r \leq \Delta$, $\varepsilon > 0$, *and* $\Delta \in \mathbb{N}$ *is defined as*

$$f_{\mathrm{DStair}_{r,\varepsilon,\Delta}}(i) = \begin{cases} a(r) & 0 \leq i < r \\ a(r)e^{-\epsilon} & r \leq i < \Delta \\ e^{-k\epsilon} f_{\mathrm{DStair}_{r,\varepsilon,\Delta}}(i - k\Delta) & k\Delta \leq i < (k+1)\Delta \\ f_{\mathrm{DStair}_{r,\varepsilon,\Delta}}(-i) & i < 0 \end{cases}$$

where $a(r) = \frac{1-b}{2r+2b(\Delta-r)-(1-b)}$, $b = e^{-\epsilon}$, *and* $k \in \mathbb{N}$.

▶ **Lemma 34.** *Let* $z = e^{\varepsilon} - 1$, *then*

$$\mathrm{Var}(\mathrm{DStair}_{r,\varepsilon,\Delta}) = \frac{x_1 + x_2 + x_3}{3z^2\left(1 - 2r + e^{\varepsilon}(2r-1) + 2\Delta\right)}$$

Where

- $x_1 = 2r^3z^3 - 3r^2z^2(z - 2\Delta)$
- $x_2 = rz(1 + e^{2\varepsilon} + 6\Delta(1+\Delta) + e^{\varepsilon}(6\Delta(\Delta-1) - 2))$
- $x_3 = 2e^{\varepsilon}\Delta(-1 + 4\Delta^2 + \cosh(\varepsilon) + 2\Delta^2\cosh(\varepsilon) - 3\Delta\sinh(\varepsilon))$

Proof. Let $X \sim \mathrm{DStair}(r, \varepsilon, \Delta)$. We will first compute the variance from just the central "stair".

$$C = \sum_{x=-r+1}^{r-1} x^2 a(r) = \frac{1}{3} r(r-1)(2r-1)a(r)$$

Finally we compute the variance from the full support of the distribution.

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=-\infty}^{\infty} x^2 f_{\mathrm{DStair}_{r,\varepsilon,\Delta}}(x) \\ &= C + 2\sum_{k=1}^{\infty}\sum_{i=1}^{\Delta}((k-1)\Delta + r + i - 1)^2 f_{\mathrm{DStair}_{r,\varepsilon,\Delta}}(k\Delta - r + i) \\ &= C + 2\sum_{k=1}^{\infty}\sum_{i=1}^{\Delta}((k-1)\Delta + r + i - 1)^2 a(r)e^{-k\epsilon} \end{aligned}$$

The result follows from symbolic simplification in Mathematica, and replacing $e^\varepsilon - 1$ terms with z for ease of presentation. ◀

We also derive the following concrete bound for large ε, which facilitates a comparison with our MSDLap noise.

▶ **Observation 35.** *For any Δ and $\varepsilon \geq \log\left(\frac{\Delta(\Delta+1)(2\Delta+1)}{2}\right)$, the variance of the discrete staircase distribution (for any value of r) is at least $\frac{1}{e^\varepsilon+(2\Delta-1)} \cdot \frac{\Delta(\Delta+1)(2\Delta+1)}{3}$*

Proof. First, notice that, if $r \neq 1$, then $f_{\mathrm{DStair}_{r,\varepsilon,\Delta}}(0) \leq \frac{1}{3}$. Since the noise is zero-mean, the variance is thus at least $\frac{2}{3}$ which is at least $\frac{1}{e^\varepsilon+(2\Delta-1)} \cdot \frac{\Delta(\Delta+1)(2\Delta+1)}{3}$ by our condition on ε.

Next, consider the case $r = 1$. In this case, the variance is exactly

$$\begin{aligned}
&\sum_{i=-\infty}^{\infty} i^2 \cdot f_{\mathrm{DStair}_{1,\varepsilon,\Delta}}(i) \\
&= 2\sum_{i=1}^{\infty} i^2 \cdot f_{\mathrm{DStair}_{1,\varepsilon,\Delta}}(i) \\
&= 2\sum_{j=1}^{\Delta}\sum_{\ell=0}^{\infty} (j+\ell\Delta)^2 \cdot f_{\mathrm{DStair}_{1,\varepsilon,\Delta}}(j+\ell\Delta) \\
&= 2\sum_{j=1}^{\Delta}\sum_{\ell=0}^{\infty} (j+\ell\Delta)^2 \cdot a(1) \cdot e^{-\varepsilon(\ell+1)} \\
&\geq 2\sum_{j=1}^{\Delta}\sum_{\ell=0}^{\infty} j^2 \cdot a(1) \cdot e^{-\varepsilon(\ell+1)} \\
&= 2a(1)\left(\sum_{j=1}^{\Delta} j^2\right)\left(\sum_{\ell}^{\infty} e^{-\varepsilon(\ell+1)}\right) \\
&= 2\left(\frac{1-e^{-\varepsilon}}{1+(2\Delta-1)e^{-\varepsilon}}\right)\left(\frac{\Delta(\Delta+1)(2\Delta+1)}{6}\right)\left(\frac{e^{-\varepsilon}}{1-e^{-\varepsilon}}\right) \\
&= \frac{1}{e^\varepsilon+(2\Delta-1)} \cdot \frac{\Delta(\Delta+1)(2\Delta+1)}{3}
\end{aligned}$$

where $a(1)$ is as defined in Definition 33. ◀

A.4 Arete convergence

In this section we show a link between the GDL distribution and the Arete distribution from [39].

▶ **Definition 36.** *For $k, \theta, \lambda > 0$, let $\mathrm{Arete}(k, \theta, \lambda)$ denote the distribution of $Z_1 - Z_2 + Z_3$ where $Z_1, Z_2 \overset{i.i.d.}{\sim} \Gamma(k, \theta)$ and $Z_3 \sim \mathrm{Lap}(\lambda)$.*

The main technical lemma linking the GDL and Arete follows from showing how the negative binomial distribution converges to the gamma distribution.

▶ **Lemma 37.** *Let $X \sim \mathrm{NB}(k, 1 - e^{-\frac{1}{\theta\Delta_\mathrm{d}}})$ and $Z_{\Delta_\mathrm{d}} \sim X/\Delta_\mathrm{d}$. Then $Z_{\Delta_\mathrm{d}} \xrightarrow{dist} \Gamma(k, \theta)$ as $\Delta_\mathrm{d} \to \infty$.*

Proof. By Lévy's continuity theorem, convergence in distribution follows from pointwise convergence of the characteristic function. Denote $\varphi(\mathcal{D})$ the characteristic function of the distribution $\mathcal{D}$. We first note the following facts:

- $\varphi(\mathrm{NB}(r,p)) = \left(\frac{p}{1-e^{it}(1-p)}\right)^r$
- $\varphi(\mathrm{NB}(r,p)/\Delta_\mathrm{d}) = \left(\frac{p}{1-e^{it/\Delta_\mathrm{d}}(1-p)}\right)^r$
- $\varphi(\Gamma(k,\theta)) = (1-\theta t)^{-k}$

Finally,

$$
\begin{aligned}
\lim_{\Delta_\mathrm{d}\to\infty} \varphi(Z) &= \lim_{\Delta_\mathrm{d}\to\infty} \left(\frac{1-e^{-\frac{1}{\theta\Delta_\mathrm{d}}}}{1-e^{(it-1/\theta)/\Delta_\mathrm{d}}}\right)^k \\
\text{L'Hôpital} &= \left(\lim_{\Delta_\mathrm{d}\to\infty} \frac{\frac{\mathrm{d}}{\mathrm{d}\Delta_\mathrm{d}} 1-e^{-\frac{1}{\theta\Delta_\mathrm{d}}}}{\frac{\mathrm{d}}{\mathrm{d}\Delta_\mathrm{d}} 1-e^{(it-1/\theta)/\Delta_\mathrm{d}}}\right)^k \\
&= \left(\lim_{\Delta_\mathrm{d}\to\infty} -\frac{e^{-\frac{1}{\theta\Delta_\mathrm{d}}}}{\theta\Delta_\mathrm{d}^2} \cdot \frac{\Delta_\mathrm{d}^2}{e^{\frac{it-1/\theta}{\Delta_\mathrm{d}}}(it-1/\theta)}\right)^k \\
&= \left(\lim_{\Delta_\mathrm{d}\to\infty} \frac{e^{-\frac{1}{\theta\Delta_\mathrm{d}}-\frac{it}{\Delta_\mathrm{d}}+\frac{1}{\theta\Delta_\mathrm{d}}}}{-\theta(it-1/\theta))}\right)^k \\
&= \left(\lim_{\Delta_\mathrm{d}\to\infty} \frac{e^{-\frac{it}{\Delta_\mathrm{d}}}}{1-it\theta}\right)^k \\
&= \left(\lim_{\Delta_\mathrm{d}\to\infty} \frac{ie^{\frac{-it\theta}{\Delta_\mathrm{d}\theta}}}{i+t\theta}\right)^k \\
&= \left(\frac{i}{i+t\theta}\right)^k \\
&= \varphi(\Gamma(k,\theta)).
\end{aligned}
$$

◀

▶ **Proposition 38.** *Let* $Z_{\Delta_\mathrm{d}} \sim \mathrm{Lap}(\lambda) + \mathrm{GDL}(k, \frac{1}{\theta\Delta_\mathrm{d}})/\Delta_\mathrm{d}$. *Then* $Z_{\Delta_\mathrm{d}} \xrightarrow{dist} \mathrm{Arete}(k,\theta,\lambda)$ *as* $\Delta_\mathrm{d} \to \infty$.

Proof. This follows from Definition 9, Definition 36, and Lemma 37. ◀

▶ Remark 39. The convergence result in Proposition 38 shows that the Arete mechanism is quite similar to the GDL mechanism transformed with real support via the approach in Theorem 21. The primary difference is how large Δ_d gets. Using Theorem 21, we only set $\Delta_\mathrm{d} = O(e^{\varepsilon/3})$, allowing the resulting (discrete) distribution to have "holes" in its support. The purpose of the Laplace noise in that case is to smooth out the holes and ensure support on $\mathbb{R}$. On the other hand the Arete (via Proposition 38) requires $\Delta_\mathrm{d} \to \infty$ (with no holes in its support), and the purpose of the Laplace noise is to smooth out the resulting *singularity* at 0 for sufficiently small values of k.[13] As such, the proof technique for Theorem 21 cannot be immediately used to help prove (or disprove) Conjecture 30.

[13] See [39, Page 3] for further discussion on this point.

A.5 Parameterized Difference Set and The multi-scale discrete Laplace Mechanism

Recall that, for privacy analysis, we usually only consider the sensitivity of the function q, which is defined as $\Delta(q) = \max_{x,x'} |q(x) - q(x')|$ where the maximum is over all pairs x and x' differing on one entry. In this section, we show that, if we parameterized the potential different $q(x) - q(x')$ values in a more fine-grained manner, we can achieve an improved error in certain cases.

For a given query function $q : X^d \to \mathbb{R}$, we define the *difference set* of q, denoted by $S_{\text{diff}}(q)$, as the set of all possible values of $|q(x) - q(x')|$ for all pairs x and x' differing on one entry. If $S_{\text{diff}}(q)$ is finite, we let the $S_{\text{diff}}(q)$-multi-scale discrete Laplace Mechanism to be the mechanism that outputs $q(x) + \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i$ where $X_i \overset{\text{i.i.d.}}{\sim} \text{DLap}(\varepsilon)$ for all $i \in S_{\text{diff}}(q)$.

▶ **Theorem 40.** *For any query function $q : X^d \to \mathbb{R}$ such that $S_{\text{diff}}(q)$ is finite, the $S_{\text{diff}}(q)$-multi-scale discrete Laplace Mechanism is ε-DP. Furthermore, for $\varepsilon \geq 1$, its MSE is $O\left(e^{-\varepsilon} \cdot \sum_{i \in S_{\text{diff}}(q)} i^2\right)$.*

Proof. *(Privacy)* To show that this mechanism is ε-DP, it suffices to show that $\text{D}_\infty\left(q(x) + \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i \,\middle\|\, q(x') + \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i\right) \leq \varepsilon$ for any pair $x, x' \in X^d$ that differs on one entry. Consider any such fixed pair of x, x'. Let $\xi = q(x) - q(x')$; due to symmetry of the noise around zero, we may assume that $\xi \geq 0$. If $\xi = 0$, the statement is clearly true. Otherwise, from definition of S_{diff}, we have $\xi \in S_{\text{diff}}$.

From Lemma 6, we have

$$\begin{aligned}
&\text{D}_\infty\left(q(x) + \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i \,\middle\|\, q(x') + \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i\right) \\
&= \text{D}_\infty\left(\xi + \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i \,\middle\|\, \sum_{i \in S_{\text{diff}}(q)} i \cdot X_i\right) \\
&\leq \text{D}_\infty\left(\xi + \xi \cdot X_\xi \parallel \xi \cdot X_\xi\right) \\
&= \text{D}_\infty\left(1 + X_\xi \parallel X_\xi\right) \leq \varepsilon,
\end{aligned}$$

where the last inequality follows from $X_\xi \sim \text{DLap}(\varepsilon)$. Thus, the mechanism is ε-DP.

(Accuracy) The MSE of the mechanism is

$$\text{Var}\left(\sum_{i \in S_{\text{diff}}(q)} i \cdot X_i\right) = \sum_{i \in S_{\text{diff}}(q)} i^2 \cdot \text{Var}(X_i) = O\left(e^{-\varepsilon} \cdot \sum_{i \in S_{\text{diff}}(q)} i^2\right). \qquad \blacktriangleleft$$

To see the advantage of the above mechanism, we note a few scenarios where this mechanism has MSE $O(\Delta^2 \cdot e^{-\varepsilon})$, but where approaches which consider the sensitivity alone must have MSE at least $\Omega(\min\{\Delta^3 e^{-\varepsilon}, \Delta^2 e^{-2\varepsilon/3}\})$ [18, 17, 19], which is asymptotically larger. In each example we consider a query q with sensitivity Δ.

- $S_{\text{diff}}(q) \subseteq [\lceil \Delta^{2/3} \rceil] \cup \{\Delta\}$, i.e. there is one large possible difference, but possibly many small ones far from Δ.
- $S_{\text{diff}}(q) = \{n^m : m \leq \log_n(\Delta) \in \mathbb{Z}_{\geq 0}\}$ for fixed $n, m > 0$ i.e. differences are structured to form exponential buckets.

Finally, we note that the setting where $|S_{\text{diff}}(q)|$ is small (even when $\Delta(q)$ is large) can occur in practice. As an example, imagine a simple merchant that sells items whose prices are all in the set $S = \{5, 10, 30, 100\}$ and they want to privately sum a database of sales where each row is the sale price of a single item. If a neighboring dataset adds or removes a row, it is clear the sensitivity of this query is $\Delta = 100$. For $\epsilon = 10$, the continuous staircase will have MSE[14] 8.5, but the MSDLap mechanism with $S_{\text{diff}}(q) = S$ will have MSE 1.0, resulting in nearly an order of magnitude improvement.

[14] MSE results rounded to two significant figures.