Journal of Privacy and Confidentiality
Vol. 14 (1) 2024

Submitted June 2022
Published Feb 2024

EXACT PRIVACY ANALYSIS OF THE GAUSSIAN SPARSE HISTOGRAM MECHANISM

BRIAN KARRER, DANIEL KIFER, ARJUN WILKINS, AND DANFENG ZHANG

FAIR, Meta, Menlo Park, CA

Pennsylvania State University, University Park, PA

Central Applied Science, Meta, Menlo Park, CA

Pennsylvania State University, University Park, PA

Abstract. Sparse histogram methods can be useful for returning differentially private counts of items in large or infinite histograms or large group-by queries, and more generally, releasing a set of statistics with sufficient item counts. We consider the Gaussian version of the sparse histogram mechanism and study the exact ϵ, δ differential privacy guarantees satisfied by this mechanism. We compare these exact ϵ, δ parameters to the simpler overestimates used in prior work to quantify the impact of looser privacy bounds.

1. Introduction

Consider a dataset with a very large, or possibly infinite domain, such as a dataset of user interactions with a URL—every time a user from country C performs an action A (e.g., share, like) on a url U, the record (U, C, A) is added to the data. It is natural to ask group-by queries, such as

```
SELECT COUNT(*) FROM table WHERE Action=‘‘share’’
    GROUP BY URL, COUNTRY,
```

which counts the number of "shares" a URL has in each country. Answering this type of query under pure differential privacy would essentially require enumerating every *possible* URL and country combination: this is called a *Cartesian expansion* (of the grouping columns URL and Country). Note that under pure differential privacy, a Cartesian expansion also includes combinations that have zero counts in the dataset, since adding or removing an individual in the dataset might change those counts. Clearly, computing a Cartesian expansion is infeasible for extremely large (or, as in this case, infinite) domains.[1]

In such cases, one settles for approximate differential privacy and thresholding schemes [Korolova et al., 2009, Balcer and Vadhan, 2019, Wilson et al., 2020, Google Anonymization Team, 2020, Gotz et al., 2012, Bun et al., 2016]: one first filters out items whose true counts

Key words and phrases: Differential privacy.

[1]Another example where the domain is too large for this to be practical is returning counts of appearances of n-grams (n consecutive words) from a text corpus.

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.823

are 0, adds noise to the remaining items, then returns the noisy counts for items whose noisy counts meet or exceed some threshold τ^*. This parameter τ^* should be set high enough so that even if noise were added to a count whose true value is 0, the noisy value will, with overwhelming probability, still be less than τ^*. This mechanism is known in the literature as a *sparse histogram* or stability histogram mechanism [Balcer and Vadhan, 2019].

We consider a generalization of this approach that can accommodate common situations in which data have been collected and pre-processed before being sent to a privacy expert. Specifically, infrequent data items may have been removed (e.g., URLs with few user interactions are removed) to save space. Thus, taking account of this pre-processing, the mechanism effectively becomes the following: first, filter out items whose true counts are less than some threshold τ, then add noise to the remaining items, and then return those noisy counts for items whose noisy counts meet or exceed a second threshold τ^*. Normally, the initial pre-processing would be considered bad for privacy, since the deletion or addition of one item could cause an entire group of size τ to appear/disappear before noise injection. Under pure differential privacy, such a situation would typically require the noise injection phase to use τ times as much noise. However, for the privacy properties of the sparse histogram mechanism, all that matters is the difference $\tau^* - \tau$, thus providing an example of how one can recover from pre-processing steps over which a privacy expert may have no control.

In this setting, we study the sparse histogram mechanism and derive an exact ϵ, δ curve for the case that the noise used is Gaussian. The reason for emphasis on the Gaussian is that many end users are more comfortable with this distribution for their subsequent statistical analyses. We compare the exact ϵ, δ curve to the approach of Wilson et al. [2020] and Google Anonymization Team [2020], which provides an over-approximation of the privacy parameters. (Originally they derived their results for Laplace noise and $\tau = 1$, thus filtering out exactly those cells with 0 counts; later they extended the work to Gaussian noise, but still with $\tau = 1$.) One goal of this paper is to quantify the impact of this over-approximation and to identify when a more exact privacy loss accounting is necessary.

Our contributions are the following:

- We derive the exact ϵ, δ curve for the Gaussian noise-based sparse histogram mechanism. In the database setting, this is equivalent to a group-by query that returns group sizes along with other aggregations for each group, but filters out small groups.
- We provide a case study that allows us to analyze the impact of the conservative ϵ, δ calculations used in prior work.

This paper is organized as follows. In Section 2, we present relevant background material on differential privacy and the Gaussian mechanism. In Section 3 we review the Gaussian sparse histogram mechanism (GSHM). Notation introduced in these sections is summarized in Table 1. Next, in Section 4, we present related work on sparse histogram mechanisms, prior to deriving our exact privacy analysis of the Gaussian sparse histogram mechanism in Section 5. In Section 6 and 7, we compare our results against privacy accounting approaches in prior work on a case study and against an alternative f-DP [Dong et al., 2022] analysis of the Gaussian sparse histogram mechanism. Section 8 contains conclusions.

2. Background

Differential privacy is an emerging standard for settings where a data release mechanism must process private data and produce publicly shareable information while each individual's

Data	X	Dataset
	X_{-j}	Dataset X with user j removed
Mechanism	d	Number of potential rows in output ($1 \leq d \leq \infty$)
	m	Number of columns per output row ($m \geq 1$)
	$\emptyset$	Null output value; simply not returned
	$M(X)$	random mechanism applied to dataset X returning $\{\emptyset, \mathbb{R}\}^{d \times m}$
	$M(X)_i$	ith output row of mechanism $\in \{\emptyset, \mathbb{R}\}^m$
Parameters	τ	Low nonnegative threshold
	τ^*	High nonnegative threshold ($\tau^* > \tau$)
	σ	Standard deviation of noise for user count column
	Σ	Covariance matrix of noise for remaining $m-1$ columns
Privacy analysis	(ϵ, δ)	Approximate differential privacy parameters
	C_u	Maximum number of rows a user can affect
	a_+	Number of rows affected by user j with user count above τ
	$a_=$	Number of rows affected by user j with user count equal to τ
	a_-	Number of rows affected by user j with user count below τ
	F_j	A row affected by user j with user count equal to τ is not $\emptyset^m$
	μ_o	μ contribution from remaining columns when $m > 1$

TABLE 1. Table of notation for Gaussian sparse histogram mechanism

privacy is protected. Differential privacy is a set of restrictions on the behavior of the data release mechanism. Roughly speaking, a privacy mechanism is differentially private if the probability distribution of the output of the mechanism is fairly insensitive to any individual's contribution to the input dataset. The probability is with respect to the randomness in the mechanism, not the randomness in the data.

Definition 2.1 (Neighbors)**.** Datasets X and X' are neighbors if one can be obtained from the other by adding records from one individual. (An individual can contribute multiple records to a dataset.)

Definition 2.2 (Approximate differential privacy [Dwork et al., 2006b,a])**.** Suppose that $\epsilon \geq 0$ and $\delta \in [0, 1]$. A randomized mechanism M satisfies (ϵ, δ)-DP if for every pair of neighbors X and X' , and every output set S,

$$\mathbb{P}(M(X) \in S) \leq e^{\epsilon}\, \mathbb{P}(M(X') \in S) + \delta. \tag{2.1}$$

The ϵ, δ parameters of a mechanism M are typically derived with the help of a mathematical construct called the *privacy loss random variable* (PLRV), which is defined as follows.

Definition 2.3 (PLRV)**.** For a randomized mechanism M, two neighboring inputs X and X', and an output ω, let $l_{M,X,X'}(\omega) = \log\left(\mathbb{P}(M(X) = \omega)/\,\mathbb{P}(M(X') = \omega)\right)$. Then $L_{M,X,X'}$ is the *privacy loss random variable* defined as the distribution of $l_{M,X,X'}(\omega)$ when ω is sampled from the distribution $\mathbb{P}(M(X))$.

Approximate differential privacy can then be written in terms of PLRV's.

Theorem 2.1 (Theorem 5 from Balle and Wang [2018])**.** A randomized mechanism M is (ϵ, δ)-DP if and only if for every pair of neighboring datasets X and X' the following holds

for the associated PLRVs:

$$\mathbb{P}(L_{M,X,X'} \geq \epsilon) - e^{\epsilon}\,\mathbb{P}(L_{M,X',X} \leq -\epsilon) \leq \delta. \tag{2.2}$$

One important noise distribution for mechanisms satisfying approximate differential privacy is the Gaussian distribution, which leads to the concept of a Gaussian mechanism.

Definition 2.4. Let f be a function (known as a "query") whose input is a database and output is a vector in $\mathbb{R}^m$. The Gaussian mechanism with covariance Σ is the mechanism that outputs $f(X) + Z$, where $Z \in \mathbb{R}^m$ is drawn from $\mathbb{N}(0, \Sigma)$.

The exact ϵ, δ parameters for this mechanism can be computed using the following theorem.

Theorem 2.2 (Analytic Gaussian mechanism privacy [Balle and Wang, 2018, Dong et al., 2021, Xiao et al., 2021])**.** The Gaussian mechanism is (ϵ, δ)-DP if and only if

$$\Phi\left(\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - e^{\epsilon}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) \leq \delta \tag{2.3}$$

where Φ is the CDF of the standard normal distribution and μ is

$$\mu = \max_{\text{neighboring } X,X'} \sqrt{(f(X) - f(X'))^T\,\Sigma^{-1}\,(f(X) - f(X'))}. \tag{2.4}$$

In particular, when Σ is a diagonal matrix with diagonals $\sigma_1^2, \sigma_2^2, \ldots$ then

$$\mu = \max_{\text{neighboring} X,X'} \sqrt{\sum_i (f(X)_i - f(X')_i)^2 / \sigma_i^2}.$$

Furthermore, the quantity in Equation 2.3 is a monotonically increasing function of μ.

Privacy interpretations require ϵ and δ be nonnegative; however, Definition 2.2 is still mathematically valid even when $\epsilon < 0$ or $\delta < 0$. Furthermore, the proofs of Theorem 2.2 [Balle and Wang, 2018, Dong et al., 2021, Xiao et al., 2021] also make no assumptions on ϵ, nor do they require that Equation 2.3 be positive. This observation turns out to be useful for the results of this paper. A similar observation that negative privacy parameters are useful for privacy analysis occurred in Zhu et al. [2022].

Observation 2.1. The explicit PLRV expressions in Equation 2.3 for the Gaussian mechanism are mathematically correct for any value of ϵ including $\epsilon < 0$.

3. Gaussian sparse histogram mechanism

In the introduction, we briefly described a generalization of the sparse histogram mechanism that avoids Cartesian expansion through a combination of two thresholds: τ for cell suppression and τ^* for noisy thresholding with Gaussian noise. Here we introduce this Gaussian sparse histogram mechanism in detail. Let $X = \{x_i\}_{i=1}^n$ be a dataset of n records, $x_i \in \mathbb{X}$.

We are interested in queries that partition the records in X into d groups. For every group, m statistics are computed, one of which is the number of records in the group. The statistics should only be reported for groups that are large enough, having at least τ records. For simplicity of presentation, we consider the setting where a user can contribute to at

most C_u records, with each one belonging to a different group. Thus each user affects the counts in at most C_u groups by at most one per group. Note that this is the same setting as studied by Wilson et al. [2020] and Google Anonymization Team [2020], reusing the same notation for C_u.

This is a natural setting for group-by queries. Consider the URL example from the introduction. Each record has the form (*user id*, *URL*, *country*, *view*, *like*, *share*); it records which actions (view, like, share) a user from the country has ever performed on the URL. Here view, like, and share are Boolean (0/1-valued) attributes. Note that a user can only share or like a URL if viewed. Each user is limited to C_u records and we are interested in group-by queries such as

```
SELECT COUNT(*) AS cnt,
   SUM(likes) AS likes,
   SUM(shares) AS shares
FROM user_url_country_table
GROUP BY url, country
HAVING cnt >= tau.
```

Note that each user contributes a count of one to each of at most C_u groups and the number of views is actually the number of records in each group.

Here d, the number of groups, is equal to the number of countries times the number of URLs (which may be infinite) times the number of aggregates per group, m, is three. This output can be represented as a table with at most d rows and m columns.

The Gaussian sparse histogram mechanism introduces a second threshold $\tau^* > \tau$ and release noisy group statistics for all groups whose noisy counts are greater than or equal to τ^*. In this running example, it would look like the following SQL query:

```
SELECT noisy_cnt, noisy_likes, noisy_shares FROM (
    SELECT
      COUNT(*) AS user_cnt,
      COUNT(*) + GaussianNoise_1 AS noisy_cnt,
      SUM(likes) + GaussianNoise_2 AS noisy_likes,
      SUM(shares) + GaussianNoise_3 AS noisy_shares
    FROM user_url_country_table
    GROUP BY url, country
    HAVING user_cnt >= tau AND noisy_cnt >= tau_star.
)
```

Formally, the mechanism is denoted as M and its goal is to privately answer a group-by aggregation query that groups the records of X into d groups and computes m noisy aggregates for each group. One of the aggregates must be count, and the rest can be arbitrary (as long as their sensitivity is known). Its pseudocode is shown in Algorithm 1.

Thus, conceptually, the output can be organized as a $d \times m$ matrix, where each entry comes from the domain $\{\emptyset, \mathbb{R}\}^{d \times m}$. Groups that are filtered out are represented as rows full of $\emptyset$. We let $M(X)_i \in \{\emptyset, \mathbb{R}\}^m$ denote the ith row of the output (i.e., aggregations over the ith group of records).

When analyzing the privacy properties of M, we will make use of the following notation. Let $G_i \subseteq \mathbb{X}$ be the set of possible records corresponding to the ith group (groups are disjoint). Without loss of generality, we assume the dataset X has been aggregated per-user and per-group such that each user has at most one record per group and each record is

Algorithm 1: Gaussian sparse histogram mechanism (GSHM)

Input: User-group aggregated dataset $X = \{x_i\}_{i=1}^n$ and groups $\{G_1, ...G_d\}$ where each user has at most one record in X per group and affects at most C_u groups. Optional aggregation function A. Parameters τ, τ^*, σ, and Σ if A is provided.

Output: Sparse dictionary mapping group index to noisy aggregates

```
SparseGroupAggregates = {}
for each nonempty group G_i do
    Sample v ~ N(0, σ^2)
    C = (number of records from X in group G_i)
    if C ≥ τ and C + v ≥ τ* then
        Sample m ~ N(0, Σ)
        SparseGroupAggregates[i] = [C + v, A(X, G_i) + m]
return SparseGroupAggregates
```

(userID, groupID, otherInfo). Let $C(X, G_i)$ be the number of records from X in group G_i (i.e., the count) and let A be an optional aggregation function that returns a vector of $m-1$ real values for a group (i.e., $A(X, G_i) \in \mathbb{R}^{m-1}$). Examples of such an A include the number of shares and likes in a group, but in general, could be arbitrary as long as its privacy impact μ (see Theorem 2.2), after adding $N(0, \Sigma)$ noise can be calculated. With this notation, row i in the output of the Gaussian sparse histogram mechanism M can be written as

$$M(X)_i = \begin{cases} \emptyset^m & \text{if } C(X, G_i) < \tau \text{ or } C(X, G_i) + v_i < \tau^* \\ \{C(X, G_i) + v_i, A(X, G_i) + m_i\} & \text{otherwise,} \end{cases}$$

where v_i is univariate Gaussian noise with standard deviation σ and m_i is multivariate Gaussian noise with covariance matrix $\Sigma \in \mathbb{R}^{(m-1)\times(m-1)}$. The noise of M is independent across all rows i. If $m = 1$ (i.e., the only aggregation is the count), then there is no $A(X, G_i)$ part.

We summarize relevant notation introduced so far for the Gaussian sparse histogram mechanism in Table 1 within the data, mechanism, and parameters sections. Additional terms defined for our upcoming privacy analysis are also listed there for convenience.

4. Related work

Sparse histogram methods using Laplace noise were proposed for releasing click and search logs in Korolova et al. [2009], Gotz et al. [2012] and also analyzed in Bun et al. [2016]. An overview of such sparse histogram approaches, including error bounds can be found in Balcer and Vadhan [2019]. To our knowledge, this past research has not specifically considered Gaussian noise.

Accounting for unknown or large domains has also been considered for the related context of top-k selection, in Durfee and Rogers [2019]. Like sparse histograms, this research involves a data-dependent pruning of outputs, in this case returning at most k items whose noisy counts are large compared to the noisy k'th element (with $k' > k$). It is worth noting that top-k algorithms return the identities of large items, but not an estimate of their counts.

Returning to group-by queries, instead of thresholding small groups first to achieve sparsity and then adding noise, one could consider a postprocessing approach that first adds noise to each group and then removes cells with with noisy counts less than a threshold τ^*. This approach would satisfy pure differential privacy and could even be implemented efficiently (without enumerating all groups in the Cartesian expansion) when Laplace noise is used [Cormode et al., 2012]. However, to achieve a desired level of sparsity, the threshold τ^* has to increase with the (logarithm of the) size of the Cartesian expansion, indicating dataset utility could be reduced by post-processing approaches in high-dimensional or infinite settings.

Sparse histogram methods with Laplace noise were applied by Wilson et al. [2020] as part of a differentially private SQL system, where avoiding Cartesian expansion was helpful for implementing group-by operations efficiently. They additionally propose composing the count part of the query with other aggregations using the Laplace mechanism. They extend the previous Laplace approaches of Korolova et al. [2009] and Gotz et al. [2012] to return multiple aggregations for each group, i.e., $m > 1$. In a later unpublished technical report [Google Anonymization Team, 2020], they derived an (ϵ, δ)-DP guarantee for sparse histogram mechanisms with a wide range of noise distributions (including Gaussian noise) for the single count output ($m = 1$).[2] Using the notation of our paper, their main results on the (ϵ, δ) privacy parameters can be expressed as follows. We refer to their technique as "add the deltas." (Google Anonymization Team [2020])

Theorem 4.1 (Add the deltas from Wilson et al. [2020], Google Anonymization Team [2020])**.** Let C_u be the maximum number of rows affected by a user. Algorithm 1 with $\tau = 1$, τ^*, d, $m = 1$, and σ, satisfies $(\epsilon, \delta_{\text{Gaussian}} + \delta_{\text{infinite}})$-DP, where

$$\delta_{\text{Gaussian}} = \Phi\left(\frac{\sqrt{C_u}}{2\sigma} - \frac{\epsilon\sigma}{\sqrt{C_u}}\right) - e^{\epsilon}\Phi\left(-\frac{\sqrt{C_u}}{2\sigma} - \frac{\epsilon\sigma}{\sqrt{C_u}}\right)$$
$$\delta_{\text{infinite}} = 1 - \Phi\left(\frac{\tau^* - 1}{\sigma}\right)^{C_u}. \tag{4.1}$$

We recognize δ_{Gaussian} from the Gaussian mechanism in Theorem 2.2 (where $\mu = \sqrt{C_u}/\sigma$), plus another contribution due to thresholding, δ_{infinite}. We refer to the contribution from thresholding as δ_{infinite} because it corresponds to the worst-case probability of infinite privacy loss under the mechanism. In particular, the privacy loss random variable $l_{M,X,X'}$ is infinite when $M(X)$ returns a row that cannot be returned by $M(X')$ due to the deterministic τ threshold. As we shall see, the worst-case probability of at least one such row being returned under the mechanism for two neighboring datasets X and X' is given by this expression for δ_{infinite}.

The result in Theorem 4.1 is overly conservative. Next, we will derive our exact result and compare it to this theorem. (In Section 7 we will also consider an analysis based on the tools of f-DP [Dong et al., 2022].) In addition to tighter accounting, our result is also applicable to arbitrary $\tau \geq 1$ and $m \geq 1$ for Gaussian noise.

[2]This derivation also includes an extension to thresholding on non-count columns with bounded positive contributions. We do not consider this non-count extension here, but believe our results would extend to this setting.

5. Exact Privacy analysis

In this section, we analyze the privacy guarantees provided by the Gaussian sparse histogram mechanism. Recall that, as in prior work [Wilson et al., 2020, Google Anonymization Team, 2020], each user contributes at most one record to up to C_u groups. We also define

$$\mu_o^2 = \max_{i \text{ and neighboring } X,X'} (A(X, G_i) - A(X', G_i))\Sigma^{-1}(A(X, G_i) - A(X', G_i)), \tag{5.1}$$

which summarizes the contribution of A and noise covariance Σ to the ϵ, δ curve in Theorem 2.2 (and indirectly in Theorem 2.1).

We use privacy loss random variables (Definition 2.3) and Theorem 2.1 to obtain the exact ϵ, δ curve for the Gaussian sparse histogram mechanism. So we begin by setting up the relevant privacy loss random variables. Without loss of generality, the target person j we consider for analyzing DP properties is the first person and the output rows affected are the first $C \leq C_u$ rows. Among those C rows, we use a_+, $a_=$, a_- to denote the number of rows whose true user count (when X is the input) is above, equal to, below the threshold τ respectively. Note that $a_+ + a_= + a_- = C \leq C_u$.

There are two types of privacy loss random variables (dependence on $M, X, X_{-j}, a_+, a_-, a_=$ omitted from the notation) for our mechanism M:

- L_+ is defined as the distribution of $\log(\mathbb{P}(M(X) = \omega)/\,\mathbb{P}(M(X_{-j}) = \omega))$, where ω is sampled from the distribution $\mathbb{P}(M(X))$ and X_{-j} is the dataset with user j removed.
- L_- is defined as the distribution of $\log(\mathbb{P}(M(X_{-j}) = \omega)/\,\mathbb{P}(M(X) = \omega))$, where ω is sampled from the distribution $\mathbb{P}(M(X_{-j}))$.

The rows to which the target person does not contribute (i.e., rows after row C) do not affect the privacy loss random variable. The same is true with the rows where the count is below the threshold (when X is the input). Therefore the privacy loss random variables are only affected by a_+ and $a_=$ and the condition that $a_+ + a_= \leq C_u$.

Because each output row is independent of the others, we can write

$$\begin{aligned} L_+ &= L_+^+ + L_+^= \\ L_- &= L_-^+ + L_-^=, \end{aligned} \tag{5.2}$$

where L_+^+ is the PLRV over the a_+ rows (rows containing user j and above the threshold) and $L_+^=$ is the PLRV over the $a_=$ rows (containing user j and at the threshold τ), similarly for L_-^+ (PLRV for the same a_+ rows, but now user j is removed) and $L_-^=$.

For our mechanism to be (ϵ, δ)-DP per Theorem 2.1, we require that the following two expressions hold for any values of $a_+ + a_= \leq C_u$.

$$\begin{aligned} \mathbb{P}(L_+ \geq \epsilon) - e^\epsilon\, \mathbb{P}(L_- \leq -\epsilon) \leq \delta \\ \mathbb{P}(L_- \geq \epsilon) - e^\epsilon\, \mathbb{P}(L_+ \leq -\epsilon) \leq \delta \end{aligned} \tag{5.3}$$

for every X and j. The first expression corresponds to X containing j and X' not containing j, and vice versa.

Next, we evaluate PLRVs under two cases (depending on whether $a_+ = 0$). Proofs appear in the Appendix.

Lemma 5.1 (Case $a_+ = 0$)**.** If $a_+ = 0$, Equation 5.3 is satisfied when

$$1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u} \leq \delta. \tag{5.4}$$

Lemma 5.2 (Case $a_+ > 0$)**.** Define

$$\epsilon_2 = \epsilon - a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right),$$

and

$$\epsilon_3 = \epsilon + a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)$$

If $a_+ > 0$, Equation 5.3 is satisfied when the following two conditions hold:

$$\begin{aligned} \max_{a_+ + a_= \le C_u, a_+ > 0} & 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \\ & + \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \left[\mathbb{P}(L_+^+ \ge \epsilon_2) - e^{\epsilon_2}\,\mathbb{P}(L_-^+ \le -\epsilon_2)\right] \le \delta, \\ \max_{a_+ + a_= \le C_u, a_+ > 0} & \mathbb{P}(L_-^+ \ge \epsilon_3) - e^{\epsilon_3}\,\mathbb{P}(L_+^+ \le -\epsilon_3) \le \delta. \end{aligned} \tag{5.5}$$

To simplify Lemma 5.2 further, we work out the remaining PLRV terms that correspond to rows above the threshold τ. For these a_+ rows, the Gaussian sparse histogram mechanism behaves identically to the Gaussian mechanism with a post-processing threshold τ^* applied to the count column. Utilizing Observation 2.1 to account for possibly negative ϵ_3, we can then claim where the right-hand side is the evaluation for the Gaussian mechanism without post-processing:

Lemma 5.3. Define

$$\mu(a_+) = \sqrt{\frac{a_+}{\sigma^2} + a_+ \mu_o^2}. \tag{5.6}$$

Then,

$$\mathbb{P}(L_+^+ \ge \epsilon_2) - e^{\epsilon_2}\,\mathbb{P}(L_-^+ \le -\epsilon_2) \le \Phi\left(\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right) - e^{\epsilon_2}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right)$$

and

$$\mathbb{P}(L_-^+ \ge \epsilon_3) - e^{\epsilon_3}\,\mathbb{P}(L_+^+ \le -\epsilon_3) \le \Phi\left(\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right) - e^{\epsilon_3}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right), \tag{5.7}$$

where the functions ϵ_2 and ϵ_3 are defined in Lemma 5.2. Without further assumptions about A and the groups, these inequalities are tight.

Combining the above lemmas, and that the quantity in Equation 2.3 is a monotonically increasing function of μ, gives our final result.

Theorem 5.4. Recall our previous definitions that

$$\epsilon_2(a_=) = \epsilon - a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right),$$

$$\epsilon_3(a_=) = \epsilon + a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right),$$

and

$$\mu(a_+) = \sqrt{\frac{a_+}{\sigma^2} + a_+ \mu_o^2}.$$

Then Algorithm 1 with parameters τ^*, τ, σ, and Σ satisfies (ϵ, δ)-DP with $\epsilon \geq 0$ and $\delta \in [0, 1]$ if the following condition holds

$$\max\Bigg[1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u},$$
$$\max_{a_+ + a_= = C_u, a_+ > 0} 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} + \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=}\left[\Phi\left(\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right) - e^{\epsilon_2}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right)\right],$$
$$\max_{a_+ + a_= = C_u, a_+ > 0} \Phi\left(\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right) - e^{\epsilon_3}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right)\Bigg] \leq \delta. \tag{5.8}$$

Without further assumptions on A and the groups, this privacy accounting is exact.

Because the optimization is over $a_+ + a_= = C_u$, this expression can be evaluated in linear time with respect to C_u. Now let us compare our result in Theorem 5.4 directly to "add the deltas" (Google Anonymization Team [2020]) Theorem 4.1.

Corollary 5.4.1. Let $\mu(C_u)$ be Equation 5.6 evaluated at C_u and define $m \geq 1$ generalizations of Eq. 4.1:

$$\delta_{\text{Gaussian}} = \Phi\left(\frac{\mu(C_u)}{2} - \frac{\epsilon}{\mu(C_u)}\right) - e^{\epsilon}\Phi\left(-\frac{\mu(C_u)}{2} - \frac{\epsilon}{\mu(C_u)}\right)$$
$$\delta_{\text{infinite}} = 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u}. \tag{5.9}$$

Algorithm 1 with parameters τ^*, τ, σ, and Σ has a minimal δ at a given $\epsilon \geq 0$, given by equality in Eq. 5.8, where

$$\max(\delta_{\text{infinite}}, \delta_{\text{Gaussian}}) \leq \delta < \delta_{\text{infinite}} + \delta_{\text{Gaussian}}. \tag{5.10}$$

For $C_u = 1$, the lower bound is an equality.

With realistic parameters, the minimal δ is often equal to the lower-bound in this corollary. Equality with the lower-bound both implies no additional privacy cost for thresholding over the Gaussian mechanism with the same noise and ϵ when $\delta_{\text{infinite}} \leq \delta_{\text{Gaussian}}$, and a separation from the upper-bound of $\delta_{\text{infinite}} + \delta_{\text{Gaussian}}$, the bound for $m = 1$ in Theorem 4.1 derived by Wilson et al. [2020].

Because our analysis simplifies at $C_u = 1$ for which $\delta = \max(\delta_{\text{infinite}}, \delta_{\text{Gaussian}})$, we can derive a precise comparison between the minimum noisy threshold $\tau^* - \tau$ required between using "add the deltas" (Google Anonymization Team [2020]) versus our improved accounting here equivalent to "max the deltas." Recall that we want to use the smallest $\tau^* - \tau$ to preserve utility.

Corollary 5.4.2. Let $C_u = 1$ and suppose that $\delta \geq \delta_{\text{Gaussian}}$. Then the ratio of the minimal $\tau^* - \tau$ difference that satisfies (ϵ, δ)-DP for Algorithm 1 with other parameters σ and Σ under "add the deltas" (Google Anonymization Team [2020]) and exact accounting is given by

$$\frac{\Phi^{-1}(1 - \delta + \delta_{\text{Gaussian}})}{\Phi^{-1}(1 - \delta)}. \tag{5.11}$$

This ratio is always greater than one and can be arbitrarily large, implying arbitrarily large gains in utility due to smaller noisy thresholds are possible via the exact accounting for fixed privacy parameters. We shall see similar behavior when $C_u > 1$ in our case study.

After the case study, we will then examine if there is anything to be gained from an f-DP analysis in Section 7.

6. Case study on URL dataset

Differential privacy implementations for count datasets with grouping columns typically require constructing a Cartesian expansion across all combinations of values in the grouping columns that are not structural zeros (e.g., impossible combinations, such as 98-year-old infants). This can require inclusion of a very large number of rows in a private dataset that are "sampling zeroes:" counts that happen to be zero in the dataset but are not structural zeros, which become indistinguishable from small positive values after the addition of noise. Let us consider an example implementation of the Gaussian sparse histogram method using the exact accounting in this paper, as compared to the "add the deltas" (Google Anonymization Team [2020]) accounting.

We consider differences using the "Facebook Privacy-Protected Full URLs Data Set," which we will refer to as the Facebook URL Shares dataset (for more details on this dataset, see Messing et al. [2020]). The Facebook URL Shares dataset contains aggregated and de-identified information about exposure to and engagement with URLs that were shared on Facebook. The key table of data in this dataset is called the "URL Breakdowns" table, which has columns recording the number of users who viewed, clicked, liked, reacted, commented, or shared any URL that had been posted to Facebook, provided that URL had been shared publicly at least 100 times.[3] Gaussian noise was added to each of the count columns in order to satisfy action-level and user-level differential privacy (the former protecting user interactions with a particular URL in the dataset and the latter protecting a user's cumulative interactions with URLs in the dataset). The differential privacy implementation was set such that the 99th percent(ile) most active user would receive a specified (ϵ, δ) privacy guarantee.

The URL Breakdowns table groups URL engagement data columns based on: (1) year and month when the interaction took place (2) six user age brackets plus a `NULL` category (3) user gender (4) user country of residence and (5) a 5 category user "political page affinity" categorization, plus a `NULL` category, for U.S. users only. The privacy implementation for the URL Breakdowns table was not via a sparse histogram method, and required constructing a Cartesian expansion across all five aggregation columns. That meant that in the initial dataset covering 31 year-months and 46 countries, every URL included in the dataset would have 33,201 rows in the breakdowns table: 29,295 rows for all non-U.S. countries (45 countries, 31 year-months, 7 age categories, and 3 gender categories) and 3,906 for the U.S. These rows need to be included for each URL in the dataset, even if a given URL only received engagement in one country across one year month.

A sparse histogram mechanism would allow us to exclude all rows with true values of zero, but at the cost of setting a noisy threshold that would filter out some non-zero values. The URL views column would present a logical choice as a filter column, because other types of interactions can only occur if a URL has been viewed (i.e. users can't click or like a URL they have not seen). In theory, a URL in this dataset could have zero views, but this would be highly unlikely for any URL that received over 100 public shares. As discussed in the dataset codebook, each interaction column limits users to contributing one interaction per

[3]Note that Laplace noise was added to the public share counts for each URL prior to implementing the 100 public shares threshold, so this was only post-processing.

column per row, so the data are already structured in a manner that would make it well suited to implement our mechanism.

The codebook notes that the 99th percent most active user contributed $51,914$ URL views. (A procedure was used to compute a noisy version of this statistic, see Messing et al. [2020] for more details.) Following the privacy guarantee aimed at the 99th percent most active user, we set $C_u = 51,914$ for our case study.[4] Because "add the deltas" (Google Anonymization Team [2020]) was previously derived for the $m = 1$ setting, we limit our case study to just considering the views column. Incorporating the other columns (via a non-zero μ_o^2) only increases differences between accounting methods. The codebook notes that the standard deviation of the Gaussian noise added to the views column was $\sigma = 2228$, which as a Gaussian mechanism satisfies $(\epsilon = 0.349, \delta = 10^{-5})$-DP according to Theorem 2.2.

Suppose that we are interested in the Gaussian sparse histogram mechanism for implementing differential privacy for the views column in the Facebook URL Shares dataset, with $\tau = 1$. After fixing τ, the Gaussian sparse histogram mechanism has two remaining parameters τ^*, σ. We will consider two scenarios; the minimal τ^* versus σ that satisfies a given (ϵ, δ)-DP constraint, and (ϵ, δ)-DP curves for a fixed σ and τ^*. In both cases, we will see a separation between the curves produced by the exact and "add the deltas" (Google Anonymization Team [2020]) accounting.[5]

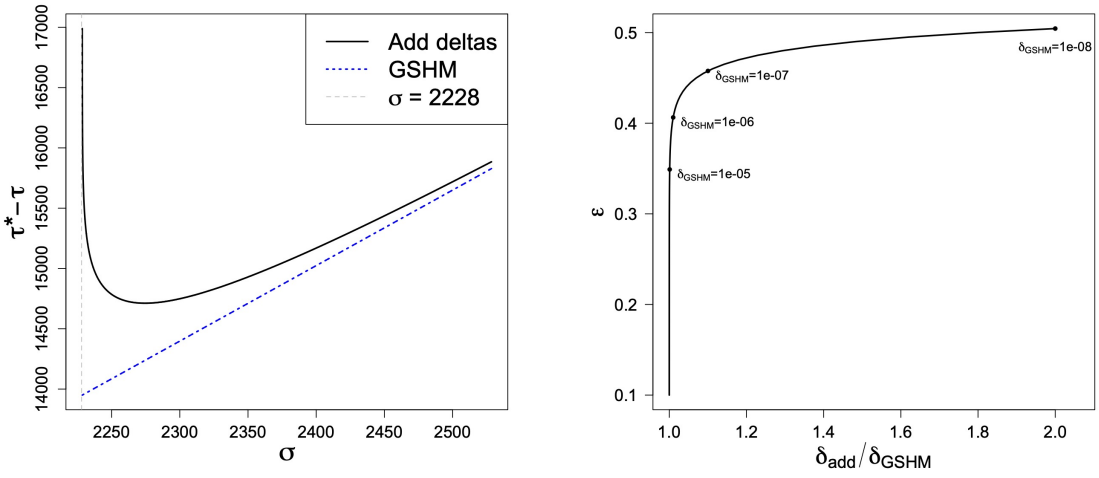
6.1. **Scenario 1: Comparison of minimal τ^* versus σ.** We fix $(\epsilon = 0.349, \delta = 10^{-5})$, the same privacy parameters implied by the Gaussian mechanism for the views column. As the Gaussian sparse histogram mechanism cannot release a lower σ than the Gaussian mechanism at the same privacy, we therefore consider $\sigma \geq 2228$. At $\sigma = 2228$, the Gaussian sparse histogram mechanism can use $\tau^* = 13948$. [6] For each σ, we compute the minimum τ^* that satisfy $(\epsilon = 0.349, \delta = 10^{-5})$-DP from "add the deltas" in Theorem 4.1 and our exact accounting in Theorem 5.4. We show the resulting curves in Figure 1a.

The exact accounting curve produces strictly lower thresholds τ^* than "add the deltas". The difference is greatest as we approach $\sigma = 2228$ where "add the deltas" cannot produce a threshold τ^* that meets the criteria at this lower bound. As in Corollary 5.4.2, the difference at $\sigma = 2228$ is unbounded, and it is precisely these lowest σ and lowest τ^* values that are of primary interest as they provide the maximum utility. The shape of the exact accounting curve requires no tradeoff between the two objectives, as we can choose both the lowest σ and the lowest τ^*. However, for "add the deltas" we are required to use a higher σ and

[4] Messing et al. [2020] note that users who have a total number of actions in the top one percent in any engagement category will suffer privacy loss greater than the specified (ϵ,δ) privacy guarantee. As we follow Messing et al. [2020], this also applies to our hypothetical case study as the actual number of URL views may be greater than C_u for these users.

[5] Code to replicate the analyses in this section can be found at
https://github.com/facebookresearch/gaussian_sparse_histogram_mechanism

[6] That the Gaussian sparse histogram mechanism improves upon simply post-processing the existing Facebook URL Shares dataset released via the Gaussian mechanism is unlikely in this circumstance. Consider dropping rows with noisy view counts less than some desired sparsity threshold τ_{post}. If $\tau_{mboxpost} = \tau^* = 13948$, the probability of a given zero row remaining after post-processing is roughly 10^{-10}. This probability is extremely small indicating a smaller τ_{post} would likely suffice for sparsity. On the other hand, if a desired $\tau_{\text{post}} \geq \tau^*$, the Gaussian sparse histogram mechanism is preferred as it performs the same filtering on non-zero count rows, while removing the zero count rows. Our emphasis in this case study is to understand the effects of privacy accounting, not to determine whether applying the Gaussian sparse histogram method would have produced a more useful dataset.



(A) (B)

FIGURE 1. (A). Scenario 1 showing the minimal τ^* versus σ satisfying $(\epsilon = 0.349, \delta = 10^{-5})$-DP for both "add the deltas" (Google Anonymization Team [2020]) and exact accounting. (B). Scenario 2 showing a $(\delta_{\text{add}}/\delta_{\text{GSHM}}, \epsilon)$ curve for $\sigma = 2228$ and $\tau^* - \tau = 16176$, where δ_{add} is from "add the deltas" and δ_{GSHM} is the exact accounting. The "add the deltas" and exact accounting are in Theorem 4.1 and Theorem 5.4 respectively.

higher τ^* to satisfy the criteria. We can further quantify these differences in noisy threshold in terms of the number of additional non-zero rows removed due to using a higher threshold:[7]

- With $\sigma = 2400$, we will lose about 1.1% more non-zero rows;
- With $\sigma = 2300$, we will lose about 2.8% more non-zero rows;
- With $\sigma = 2240$, we will lose about 6.9% more non-zero rows.

These differences directly reflect upon the improved precision provided by our privacy accounting over "add the deltas" on identifying non-zero rows.

The two curves converge as $\sigma \to \infty$. "Add the deltas" (Google Anonymization Team [2020]) gets within 1% of Theorem 5.4 at $\sigma = 2396$ (the necessary value of $\tau^* - \tau$ is $15,148$ under "add the deltas" and $14,998$ under Theorem 5.4). "Add the deltas" gets within 0.1%

[7]For these calculations, we use the breakdowns table in the Facebook URL Shares dataset that covers a period from January 2017 to February 2021, for users living in the U.S. The URL Shares dataset is updated periodically as new data become available. An exact computation of the expected fraction of rows lost as a function of σ and τ^* would require access to data not included in the Facebook URL Shares dataset. But in practice, the expected fraction of rows lost should be almost identical when computed using the privacy-protected version of the Facebook URL Shares dataset that is available to researchers via Social Science One. This is because after dropping rows of the privacy-protected data where noisy views are smaller than τ^*, the number of zero-valued rows is vanishingly small. The privacy-protected dataset has on the order of 2×10^{11} rows, of which over 38 million rows have more than $14,022$ noisy views (where country is U.S. and where the views occurred between January 2017 and February 2021). If all the rows had true values of zero (with 2×10^{11} rows, $\sigma = 2240$ and $\tau^* = 14,022$), the expected number of zero-valued rows with noisy views greater than τ^* is 31, which is vanishingly small considering the over 38 million rows with noisy views greater than $14,022$.

of Theorem 5.4 at $\sigma = 2699$ (the necessary value of $\tau^* - \tau$ is about 16,910 under "add the deltas" and 16,894 Theorem 5.4). At the points where the curves converge though, we are adding much more noise and thresholding out far more rows than would be necessary to satisfy the desired differential privacy target. Further, this convergence occurs only because we examined just the views column (i.e. $m = 1$ and $\mu_o = 0$). If additional columns were included, the curves may converge to a constant factor as in Corollary 5.4.2, considered as σ varies.

6.2. **Scenario 2: Comparison of (ϵ, δ) curves.** We can also fix σ and τ^* and examine the (ϵ, δ)-DP curves produced by "add the deltas" (Google Anonymization Team [2020]) and our exact accounting. For this example, say we use $\sigma = 2228$ and $\tau^* - \tau = 16176$ which meets $\delta_{\text{infinite}} = 10^{-8}$. For our curves, we know that $\delta \geq \delta_{\text{infinite}}$.

With these parameters set, we can examine $\delta(\epsilon)$ or $\epsilon(\delta)$. We consider the former in Figure 1b which displays how $\delta(\epsilon)$ varies over $\epsilon \in [0.1, 0.504]$ under Theorem 5.4 (GSHM) and "add the deltas" under Theorem 4.1. As seen in Figure 1b, the final δ returned by "add the deltas" is double that of GSHM at $\delta_{\text{GSHM}} = 10^{-8}$, 10% greater at $\delta_{\text{GSHM}} = 10^{-7}$, 1% greater at $\delta_{\text{GSHM}} = 10^{-6}$, and 0.1% greater at $\delta_{\text{GSHM}} = 10^{-5}$. This aligns with our expectations from Corollary 5.4.1. When δ_{infinite} is very small with respect to δ_{Gaussian}, the lower and upper bounds in Equation 5.10 become closer and δ produced from both accounting approaches will become similar. However when δ_{infinite} is non-trivial compared to δ_{Gaussian}, the exact accounting produces a smaller δ, by up to a factor of two.

7. f-DP analysis of Gaussian sparse histogram mechanism

In this section, we analyze the Gaussian sparse histogram mechanism using the f-DP framework [Dong et al., 2021, 2022], a recent generalization of differential privacy. In particular, we prove the following theorem in Appendix A.3 by following a similar novel argument as in the main text, but using f-DP composition instead of PLRV's.

Theorem 7.1. Define

$$\mu(a_+) = \sqrt{\frac{a_+}{\sigma^2} + a_+ \mu_o^2}.$$

Then using f-DP composition, Algorithm 1 with parameters τ^*, τ, σ, and Σ satisfies (ϵ, δ)-DP with $\epsilon \geq 0$ and $\delta \in [0, 1]$ if

$$\max_{a_+ + a_= = C_u} 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} + \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \left[\Phi\left(\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - e^{\epsilon}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right)\right] \leq \delta. \quad (7.1)$$

This f-DP result is similar but in general not identical to the results from our exact analysis. In Eq. 7.1 we have ϵ while the middle term of Eq. 5.8 uses $\epsilon_2 \leq \epsilon$. Because the Gaussian mechanism term in brackets is monotone decreasing with respect to ϵ, we have that $\delta_{\text{GSHM}} \leq \delta_{f\text{-DP}}$. Examining the expression, we also have that f-DP provides a substantially tighter result than "add the deltas," i.e., $\delta_{f\text{-DP}} \leq \delta_{\text{add}}$.

Our exact analysis and f-DP results are actually the same when $C_u = 1$ as can be seen by comparing Theorem 7.1 with Corollary 5.4.1, but possibly different when $C_u > 1$. If the maximum for the f-DP analysis has $a_+ = 0$ or $a_+ = C_u$, then the results are also the same but there is a gap when the optimal a_+ satisfies $0 < a_+ < C_u$. Further, we point out there is no practical difference in complexity between computing the privacy parameters using

the exact versus the f-DP analysis. So we conclude that for the Gaussian sparse histogram mechanism, the exact analysis is preferable over using f-DP composition.

8. Conclusion

Applications of differential privacy to count datasets traditionally require constructing a Cartesian expansion across all possible combinations of values in grouping columns. Constructing such a Cartesian expansion can be difficult or impossible for multiple reasons, especially when the domains are large or even infinite. In these cases, sparse histogram methods provide reasonable alternatives to Cartesian expansion.

In this paper, we have provided an exact privacy loss analysis of the Gaussian sparse histogram mechanism and demonstrated that our exact accounting is feasible in practice. On our URL case study, our comparison against past research demonstrated that in practical circumstances our more precise privacy accounting can increase utility by a significant amount, primarily in situations where it is desirable to set a low enough noisy threshold such that δ_{infinite} is comparable to δ_{Gaussian}. On the other hand, when δ_{infinite} can be made vanishing through use of a large noisy threshold, our accounting matches those from "add the deltas" (Google Anonymization Team [2020]). Given that the implementation of exact accounting is simple and that smaller noisy thresholds are of primary concern when using a sparse histogram method, we believe our improved accounting should be useful in practice.

The exactness of our privacy analysis relies upon uniform sensitivity across groups for $m > 1$ and unbounded group counts. If the count of users per group is bounded or non-uniformity is of interest, future research could improve upon our privacy analysis via revisiting Lemma 5.3 with additional assumptions.

Acknowledgement

This research was supported by funding from Meta.

References

V. Balcer and S. Vadhan. Differential privacy on finite computers. *Journal of Privacy and Confidentiality*, 9(2), Sep. 2019. doi: 10.29012/jpc.679. URL https://journalprivacyconfidentiality.org/index.php/jpc/article/view/679.

B. Balle and Y.-X. Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *35th International Conference on Machine Learning (ICML)*, 2018. URL https://arxiv.org/abs/1805.06530.

M. Bun, K. Nissim, and U. Stemmer. Simultaneous private learning of multiple concepts. ITCS '16, page 369–380, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340571. doi: 10.1145/2840728.2840747. URL https://doi.org/10.1145/2840728.2840747.

G. Cormode, C. M. Procopiuc, D. Srivastava, and T. T. L. Tran. Differentially private summaries for sparse data. In *ICDT '12*, 2012.

J. Dong, A. Roth, and W. J. Su. Gaussian differential privacy. *Journal of the Royal Statistical Society, Series B*, 2021. URL https://arxiv.org/abs/1905.02383.

J. Dong, A. Roth, and W. J. Su. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37, 2022.

D. Durfee and R. M. Rogers. Practical differentially private top-k selection with pay-what-you-get composition. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/b139e104214a08ae3f2ebcce149cdf6e-Paper.pdf.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006a.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, 2006b.

Google Anonymization Team. Delta for thresholding. *github.com/google/differential_privacy*, 2020. URL github.com/google/differential_privacy.

M. Gotz, A. Machanavajjhala, G. Wang, X. Xiao, and J. Gehrke. Publishing search logs—a comparative study of privacy guarantees. *IEEE Transactions on Knowledge and Data Engineering*, 24(3):520–532, 2012. doi: 10.1109/TKDE.2011.26.

A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas. Releasing search queries and clicks privately. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, page 171–180, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584874. doi: 10.1145/1526709.1526733. URL https://doi.org/10.1145/1526709.1526733.

S. Messing, C. DeGregorio, B. Hillenbrand, G. King, N. Persily, B. State, and A. Wilkins. Facebook Privacy-Protected Full URLs Data Set. 2020. URL https://doi.org/10.7910/DVN/TDOAPG.

R. J. Wilson, C. Y. Zhang, W. Lam, D. Desfontaines, D. Simmons-Marengo, and B. Gipson. Differentially private SQL with bounded user contribution. *Proceedings on Privacy Enhancing Technologies*, 2020(2):230–250, 2020. URL https://doi.org/10.2478/popets-2020-0025.

Y. Xiao, Z. Ding, Y. Wang, D. Zhang, and D. Kifer. Optimizing fitness-for-use of differentially private linear queries. *Proc. VLDB Endow.*, 14(10):1730–1742, 2021.

Y. Zhu, J. Dong, and Y.-X. Wang. Optimal accounting of differential privacy via characteristic function. In *International Conference on Artificial Intelligence and Statistics*, pages 4782–4817. PMLR, 2022.

Appendix A.

A.1. **Additional preliminaries for privacy analysis.** We partition the output space Ω into events for which L_+ (through $L_+^=$) is finite and infinite. Let F_j be all events where some subset of the rows $a_=$ are not $\emptyset^m$. $M(X) \in F_j$ if and only if L_+ is infinite. We have these facts, where we use $\sim F$ to represent the event that F did not happen:

- $\mathbb{P}_{M(X_{-j})}(F_j) = 0$;
- $\mathbb{P}_{M(X_{-j})}(\sim F_j) = 1$;
- $\mathbb{P}_{M(X)}(F_j) = 1 - \Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=}$;
- $\mathbb{P}_{M(X)}(\sim F_j) = \Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=}$;
- $L_+^= = \infty$ with probability $1-\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=}$ and equals $\log\left(\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=}/1\right) = a_= \log(\Phi\left(\frac{\tau^*-\tau}{\sigma}\right))$ with probability $\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=}$
- $L_-^= = \log\left(1/\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=}\right) = -a_= \log(\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)) > 0$ with probability 1.

A.2. **Proofs using above preliminaries.**

Lemma 5.1 (Case $a_+ = 0$)**.** If $a_+ = 0$, Equation 5.3 is satisfied when

$$1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u} \leq \delta. \tag{5.4}$$

Proof of Lemma 5.1. In this case, we have the following facts:

- $L_+^+ = 0$;
- $L_-^+ = 0$;
- For any $\epsilon > 0$, $\mathbb{P}(L_+^= \geq \epsilon) = 1 - \Phi\left((\tau^* - \tau)/\sigma\right)^{a_=}$ and $\mathbb{P}(L_-^= \leq -\epsilon) = 0$;
- For any $\epsilon > 0$, $\mathbb{P}(L_-^= \geq \epsilon) = 1$ if $\epsilon \leq -a_= \log(\Phi\left((\tau^* - \tau)/\sigma\right))$ and is 0 otherwise. $\mathbb{P}(L_+^= \leq -\epsilon) = \Phi\left((\tau^* - \tau/\sigma)\right)^{a_=}$ if $-\epsilon \geq a_= \log(\Phi\left((\tau^* - \tau)/\sigma\right))$, and is 0 otherwise.

Therefore

$$\mathbb{P}(L_+^= \geq \epsilon) - e^\epsilon \mathbb{P}(L_-^= \leq -\epsilon) = 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=}$$

and

$$\mathbb{P}(L_-^= \geq \epsilon) - e^\epsilon \mathbb{P}(L_+^= \leq -\epsilon) = \begin{cases} 0 & \text{if } \epsilon > -a_= \log(\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)) \\ 1 - e^\epsilon \Phi\left(\frac{\tau^*-\tau}{\sigma}\right)^{a_=} & \text{if } \epsilon \leq -a_= \log(\Phi\left(\frac{\tau^*-\tau}{\sigma}\right)). \end{cases}$$

Note that the maximum of these is $1 - \Phi\left((\tau^* - \tau)/\sigma\right)^{a_=}$; maximizing over $a_= \leq C_u$, we get

$$1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u} \leq \delta. \tag{A.1}$$

□

Lemma 5.2 (Case $a_+ > 0$)**.** Define

$$\epsilon_2 = \epsilon - a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right),$$

and

$$\epsilon_3 = \epsilon + a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)$$

If $a_+ > 0$, Equation 5.3 is satisfied when the following two conditions hold:

$$\begin{aligned}
&\max_{a_+ + a_= \le C_u, a_+ > 0} 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \\
&\quad + \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \left[\mathbb{P}(L_+^+ \ge \epsilon_2) - e^{\epsilon_2}\,\mathbb{P}(L_-^+ \le -\epsilon_2)\right] \le \delta, \\
&\max_{a_+ + a_= \le C_u, a_+ > 0} \mathbb{P}(L_-^+ \ge \epsilon_3) - e^{\epsilon_3}\,\mathbb{P}(L_+^+ \le -\epsilon_3) \le \delta.
\end{aligned} \tag{5.5}$$

Proof of Lemma 5.2. Conditioned on event F_j not happening,

$$\begin{aligned}
L_+ | \{M(X) \in \sim F_j\} &= L_+^+ + \log \frac{\mathbb{P}_{M(X)}(\sim F_j)}{\mathbb{P}_{M(X_{-j})}(\sim F_j)} = L_+^+ + a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right) \\
L_- | \{M(X_{-j}) \in \sim F_j\} &= L_-^+ + \log \frac{\mathbb{P}_{M(X_{-j})}(\sim F_j)}{\mathbb{P}_{M(X)}(\sim F_j)} = L_-^+ - a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right).
\end{aligned} \tag{A.2}$$

As mentioned in the main text, for our mechanism to be (ϵ, δ)-DP, we require that the following two expressions hold for any values of $a_+ + a_= \le C_u$:

$$\begin{aligned}
\mathbb{P}(L_+ \ge \epsilon) - e^\epsilon\,\mathbb{P}(L_- \le -\epsilon) \le \delta \\
\mathbb{P}(L_- \ge \epsilon) - e^\epsilon\,\mathbb{P}(L_+ \le -\epsilon) \le \delta
\end{aligned} \tag{A.3}$$

The first expression corresponds to X having j and X' not having j, and the second expression corresponds to X not having j and X' having j.

We first consider the top expression. Dividing into conditioning on $\sim F_j$ and F_j, and remembering that $\sim F_j$ happens with probability one under $M(X_{-j})$,

$$\begin{aligned}
&\mathbb{P}_{M(X)}(F_j)\,\mathbb{P}(L_+ \ge \epsilon | F_j) + \mathbb{P}_{M(X)}(\sim F_j)\,\mathbb{P}(L_+ \ge \epsilon | \sim F_j) - e^\epsilon\,\mathbb{P}(L_- \le -\epsilon | \sim F_j) \\
&= \mathbb{P}_{M(X)}(F_j) + \mathbb{P}_{M(X)}(\sim F_j)\,\mathbb{P}(L_+ \ge \epsilon | \sim F_j) - e^\epsilon\,\mathbb{P}(L_- \le -\epsilon | \sim F_j) \\
&= 1 - \mathbb{P}_{M(X)}(\sim F_j) + \mathbb{P}_{M(X)}(\sim F_j)\left[\mathbb{P}(L_+ \ge \epsilon | \sim F_j) - e^{\epsilon - \log(\mathbb{P}_{M(X)}(\sim F_j))}\,\mathbb{P}(L_- \le -\epsilon | \sim F_j)\right] \\
&= \underbrace{1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=}}_{\delta \text{ for infinite privacy loss}} + \underbrace{\Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \left[\mathbb{P}(L_+^+ \ge \epsilon_2) - e^{\epsilon_2}\,\mathbb{P}(L_-^+ \le -\epsilon_2)\right]}_{\delta \text{ from the numerical } a_+ \text{ rows, no } a_= \text{ rows are output}},
\end{aligned}$$

where $\epsilon_2 = \epsilon - a_= \log \Phi\left((\tau^* - \tau)/\sigma\right)$. Note that $\epsilon > 0$ implies that $\epsilon_2 > 0$.

We return to the bottom expression that considers X without j and X' with j. For this case, we want

$$\mathbb{P}(L_- \ge \epsilon) - e^\epsilon\,\mathbb{P}(L_+ \le -\epsilon) \le \delta. \tag{A.4}$$

Expanding the left-hand side (similarly to before, being careful with the signs and noting that (a) $\mathbb{P}(L_+ \le -\epsilon | F_j) = 0$ and (b) $\mathbb{P}(L_- \ge \epsilon | \sim F_j) = \mathbb{P}(L_- \ge \epsilon)$ (since $\sim F_j$ always

happens under $M(X_{-j})$), we get

$$\begin{aligned}
&\mathbb{P}(L_- \geq \epsilon | \sim F_j) - e^{\epsilon}\Big(\mathbb{P}_{M(X)}(\sim F_j)\, \mathbb{P}(L_+ \leq -\epsilon | \sim F_j) + \mathbb{P}_{M(X)}(F_j)\, \mathbb{P}(L_+ \leq -\epsilon | F_j)\Big) \\
&= \mathbb{P}(L_- \geq \epsilon | \sim F_j) - e^{\epsilon + \log(\mathbb{P}(\sim F_j))}\, \mathbb{P}(L_+ \leq -\epsilon | \sim F_j) \\
&= \mathbb{P}(L_-^+ \geq \epsilon_3) - e^{\epsilon_3}\, \mathbb{P}(L_+^+ \leq -\epsilon_3), \qquad \text{(A.5)}
\end{aligned}$$

where $\epsilon_3 = \epsilon + a_= \log \Phi\left(\tau^* - \tau/\sigma\right)$. □

Lemma 5.3. Define

$$\mu(a_+) = \sqrt{\frac{a_+}{\sigma^2} + a_+ \mu_o^2}. \qquad (5.6)$$

Then,

$$\mathbb{P}(L_+^+ \geq \epsilon_2) - e^{\epsilon_2}\, \mathbb{P}(L_-^+ \leq -\epsilon_2) \leq \Phi\left(\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right) - e^{\epsilon_2} \Phi\left(-\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right)$$

and

$$\mathbb{P}(L_-^+ \geq \epsilon_3) - e^{\epsilon_3}\, \mathbb{P}(L_+^+ \leq -\epsilon_3) \leq \Phi\left(\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right) - e^{\epsilon_3} \Phi\left(-\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right), \qquad (5.7)$$

where the functions ϵ_2 and ϵ_3 are defined in Lemma 5.2. Without further assumptions about A and the groups, these inequalities are tight.

Proof of Lemma 5.3. Let A_+ be the set of a_+ rows containing j with counts greater than the threshold τ. To evaluate the remaining PLRV expressions when $a_+ > 0$, for these A_+ rows, we note that the Gaussian sparse histogram mechanism applied to these rows is identical to the Gaussian mechanism with a post-processing threshold τ^* applied to the noisy counts for each row. Lemma 5.3 then follows from the argument about post-processing in A.2.1 which says we can use the PLRV expressions from the Gaussian mechanism as an upper-bound, regardless of the sign of ϵ. Observation 2.1 says that the Gaussian PLRV expressions contained in Theorem 2.2 are correct, despite possibly negative ϵ_3.

We now address the tightness of these inequalities. Since we are releasing these rows each of which has a μ_i contribution, the total μ^2 for releasing the A_+ rows is given by $\sum_{i \in A_+} \mu_i^2 \leq a_+/\sigma^2 + a_+ \mu_o^2$. Recalling that the Gaussian PLRV expressions are increasing functions of μ, using a larger μ^2 (and therefore larger μ) results in a larger upper bound. Under uniformity (μ_i^2 contributions equal for all rows i), this final inequality for the μ contribution from the A_+ rows is an equality. Without any assumptions on $A(x)$ and the groups, uniformity is possible (and reasonable in many circumstances) and hence this inequality is tight.

We next address tightness of the inequality due to post-processing. Consider a pair of neighboring datasets X and X_{-j}, where for all rows in A_+ the counts in both X and X_{-j} are very large compared to the threshold τ^*, such that the chance of a noisy user count being less than τ^* goes to zero. Therefore the privacy loss random variables over the A_+ rows can behave arbitrarily close to the Gaussian mechanism by simply considering datasets with large enough counts on these rows. Hence there exists a pair of neighboring datasets X and X_{-j} such that the PLRV expressions from the Gaussian mechanism are arbitrarily close to those of applying the Gaussian sparse histogram mechanism on these A_+ rows. □

A.2.1. *Post-processing and PLRV's.* We modify the proof of Theorem 5 by Balle and Wang [2018] to prove the following claim.

Let M be a random function from O to R. Let f be a deterministic post-processing function from R to R'. Then for any datasets X and X', and any value of ϵ including $\epsilon < 0$, we have that

$$\mathbb{P}(L_{f\circ M,X,X'} \geq \epsilon) - e^{\epsilon}\,\mathbb{P}(L_{f\circ M,X',X} \leq -\epsilon) \leq \mathbb{P}(L_{M,X,X'} \geq \epsilon) - e^{\epsilon}\,\mathbb{P}(L_{M,X',X} \leq -\epsilon). \quad \text{(A.6)}$$

Proof. Let $T = \{r' \in R' : \log[\mathbb{P}(f(M(X)) = r')/\,\mathbb{P}(f(M(X')) = r')] \geq \epsilon\}$, and let $S = \{r \in R : f(r) \in T\}$. Let $E = \{r \in R : \log[\mathbb{P}(M(X) = r)/\,\mathbb{P}(M(X') = r)] \geq \epsilon\}$. Also $E_+ = S \cap E$ and $E_- = S \cap (R/E)$. Using these definitions we can write

$$\begin{aligned}
&\mathbb{P}(L_{f\circ M,X,X'} \geq \epsilon) - e^{\epsilon}\,\mathbb{P}(L_{f\circ M,X',X} \leq -\epsilon) \\
&= \int_T [\mathbb{P}(f(M(X)) = r') - e^{\epsilon}\,\mathbb{P}(f(M(X')) = r')]dr' \\
&= \int_S [\mathbb{P}(M(X) = s) - e^{\epsilon}\,\mathbb{P}(M(X') = s)]ds \\
&= (\int_{E_+} + \int_{E_-})[\mathbb{P}(M(X) = s) - e^{\epsilon}\,\mathbb{P}(M(X') = s)]ds \\
&\leq \int_{E_+} [\mathbb{P}(M(X) = s) - e^{\epsilon}\,\mathbb{P}(M(X') = s)]ds \\
&\leq \int_{E} [\mathbb{P}(M(X) = s) - e^{\epsilon}\,\mathbb{P}(M(X') = s)]ds \\
&= \mathbb{P}(L_{M,X,X'} \geq \epsilon) - e^{\epsilon}\,\mathbb{P}(L_{M,X',X} \leq -\epsilon), \qquad \text{(A.7)}
\end{aligned}$$

where we used that under the events in E_- the contributions are all non-positive for the first inequality, and then the contributions under any events in E are nonnegative and $E_+ \subseteq E$. □

Theorem 5.4. Recall our previous definitions that

$$\epsilon_2(a_=) = \epsilon - a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right),$$

$$\epsilon_3(a_=) = \epsilon + a_= \log \Phi\left(\frac{\tau^* - \tau}{\sigma}\right),$$

and

$$\mu(a_+) = \sqrt{\frac{a_+}{\sigma^2} + a_+\mu_o^2}.$$

Then Algorithm 1 with parameters τ^*, τ, σ, and Σ satisfies (ϵ, δ)-DP with $\epsilon \geq 0$ and $\delta \in [0, 1]$ if the following condition holds

$$\begin{aligned}
&\max\Bigg[1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u}, \\
&\max_{a_+ + a_= = C_u, a_+ > 0} 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} + \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=}\left[\Phi\left(\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right) - e^{\epsilon_2}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon_2}{\mu}\right)\right], \\
&\max_{a_+ + a_= = C_u, a_+ > 0} \Phi\left(\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right) - e^{\epsilon_3}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon_3}{\mu}\right)\Bigg] \leq \delta. \qquad \text{(5.8)}
\end{aligned}$$

Without further assumptions on A and the groups, this privacy accounting is exact.

Proof of Theorem 5.4. The three terms in the theorem immediately follow from Lemma 5.1, Lemma 2.1, and Lemma 5.3. Tightness follows from the tightness of these preceding Lemmas. The only remaining aspect is to prove that the inner maximization occurs when $a_+ + a_= = C_u$ instead of $a_+ + a_= \leq C_u$. To do so, we demonstrate that the PLRV difference given by the left-hand side of Eq. 2.3 is monotonically increasing with respect to μ for any ϵ.

Let

$$f(\mu, \epsilon) = \Phi\left(\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - e^{\epsilon}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right). \tag{A.8}$$

for arbitrary $\mu > 0$ and ϵ. Applying calculus, we have that

$$\frac{\partial f}{\partial \mu} = \phi\left(\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) \tag{A.9}$$

and

$$\frac{\partial f}{\partial \epsilon} = -e^{\epsilon}\Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right), \tag{A.10}$$

where ϕ is the PDF of the standard normal distribution. So the partial derivative of f with respect to μ is always positive, and the partial derivative with respect to ϵ is always negative.

This implies equality with C_u because a_+ only enters into these expressions via $\mu(a_+)$ and $\mu(a_+)$ in Equation 5.6 is monotonically increasing in a_+. □

Corollary 5.4.1. Let $\mu(C_u)$ be Equation 5.6 evaluated at C_u and define $m \geq 1$ generalizations of Eq. 4.1:

$$\begin{aligned}\delta_{\text{Gaussian}} &= \Phi\left(\frac{\mu(C_u)}{2} - \frac{\epsilon}{\mu(C_u)}\right) - e^{\epsilon}\Phi\left(-\frac{\mu(C_u)}{2} - \frac{\epsilon}{\mu(C_u)}\right)\\ \delta_{\text{infinite}} &= 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{C_u}.\end{aligned} \tag{5.9}$$

Algorithm 1 with parameters τ^*, τ, σ, and Σ has a minimal δ at a given $\epsilon \geq 0$, given by equality in Eq. 5.8, where

$$\max(\delta_{\text{infinite}}, \delta_{\text{Gaussian}}) \leq \delta < \delta_{\text{infinite}} + \delta_{\text{Gaussian}}. \tag{5.10}$$

For $C_u = 1$, the lower bound is an equality.

Proof of Corollary 5.4.1. Note for $\epsilon \geq 0$ as assumed here, $0 \leq \delta_{\text{Gaussian}} < 1$.

A.2.2. *Lower bound derivation.* The lower-bound on δ follows from the first term in the three-term maximization of Equation 5.8 and the third term in the three-term maximization evaluated at $a_+ = C_u$ and $a_= = 0$. The first term is identically δ_{infinite} and the third gives δ_{Gaussian}.

When $C_u = 1$, both the second and third terms are equal to δ_{Gaussian} because they are optimizations over $a_+ > 0$ so they can only be evaluated at $a_+ = 1 = C_u$ and $a_= = 0$. Hence the minimal $\delta = \max(\delta_{\text{infinite}}, \delta_{\text{Gaussian}})$ when $C_u = 1$.

A.2.3. *Upper bound derivation.* We start by recalling f from Equation A.8, that the partial derivative of f with respect to μ is always positive, and that the partial derivative with respect to ϵ is always negative.

To simplify, let function $\beta(a_=) = \Phi\left((\tau^* - \tau)/\sigma\right)^{a_=}$. Then the second and third terms of the three-term maximization in Theorem 5.4 written in terms of β and f are:

$$\max_{a_+ + a_= \leq C_u, a_+ > 0} 1 - \beta + \beta f(\mu(a_+), \epsilon - \ln \beta),$$
$$\max_{a_+ + a_= \leq C_u, a_+ > 0} f(\mu(a_+), \epsilon + \ln \beta) \tag{A.11}$$

Then $\mu(a_+)$ is maximized when $a_+ = C_u$. Let $\mu^* = \mu(C_u)$. Given that the partial derivative of f with respect to μ is always positive and the partial derivative with respect to ϵ is always negative (and $\ln \beta < 0$), we can write upper bounds for both terms as

$$\max_{a_= \leq C_u - 1} 1 - \beta + \beta f(\mu^*, \epsilon)$$

and

$$\max_{a_= \leq C_u - 1} f(\mu^*, \epsilon + \ln \beta). \tag{A.12}$$

Because $f(\mu^*, \epsilon) = \delta_{\text{Gaussian}} < 1$, the solution to the first optimization is $a_= = C_u - 1$, which evaluates to a quantity even larger when $a_= = C_u$. Evaluated at C_u, the first equation is $\delta_{\text{infinite}} + (1 - \delta_{\text{infinite}})\delta_{\text{Gaussian}} < \delta_{\text{infinite}} + \delta_{\text{Gaussian}}$, our desired upper-bound. So what remains to be shown is that the second equation is less than or equal to the first.

Define

$$r(\beta) = 1 - \beta + \beta f(\mu^*, \epsilon)$$

and

$$t(\beta) = f(\mu^*, \epsilon + \ln \beta). \tag{A.13}$$

We will show $r(\beta) \geq t(\beta)$ for continuous $\beta \in [0, 1]$, the relevant range of β for the above maximization over $a_=$. First, we note equality at the endpoints $r(0) = t(0) = 1$ and $r(1) = t(1) = \delta_{\text{infinite}}$. Then $dr/d\beta = -1 + \delta_{\text{infinite}}$ is a constant negative slope and

$$\frac{dt}{d\beta} = \frac{\partial f(\mu^*, \epsilon + \ln \beta)}{\partial \epsilon} \frac{1}{\beta} = -e^{\epsilon} \Phi\left(\frac{-\mu^*}{2} - \frac{\epsilon + \ln \beta}{\mu^*}\right).$$

Evaluated at $\beta = 0$, $dt/d\beta < dr/d\beta$ because $-e^{\epsilon} < -1 + \delta_{\text{infinite}}$. So slightly above $\beta = 0$, we have that $r > t$. Remembering that $e^y \phi(x/2 + y/x) = \phi(x/2 - y/x)$, where ϕ is the density function for the standard normal, we have that

$$\frac{d^2 t}{d\beta^2} = \frac{\phi\left(\frac{\mu^*}{2} - \frac{\epsilon + \ln \beta}{\mu^*}\right)}{\beta^2 \mu^*}. \tag{A.14}$$

This is always positive on the range of $\beta \in (0, 1]$.

Now we claim via the mean value theorem applied to the difference of the functions $r - t$, that because $d^2(r - t)/d\beta^2 < 0$ over $\beta \in (0, 1]$, and the two functions are equal at $\beta = 0$ and $\beta = 1$, there can be no other value of β such that $r = t$ over this range. Since $r - t > 0$ slightly above $\beta = 0$, therefore $r(\beta) \geq t(\beta)$ for $\beta \in [0, 1]$, and we have proven our upper bound. □

Corollary 5.4.2. Let $C_u = 1$ and suppose that $\delta \geq \delta_{\text{Gaussian}}$. Then the ratio of the minimal $\tau^* - \tau$ difference that satisfies (ϵ, δ)-DP for Algorithm 1 with other parameters σ and Σ under "add the deltas" (Google Anonymization Team [2020]) and exact accounting is given by

$$\frac{\Phi^{-1}(1 - \delta + \delta_{\text{Gaussian}})}{\Phi^{-1}(1 - \delta)}. \tag{5.11}$$

Proof of Corollary 5.4.2. For $C_u = 1$ and $\delta \geq \delta_{\text{Gaussian}}$, "add the deltas" accounting would require that

$$\delta_{\text{infinite}} = \delta - \delta_{\text{Gaussian}} \geq 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right). \tag{A.15}$$

Exact accounting would require that

$$\delta \geq 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right). \tag{A.16}$$

Solving for the minimal $\tau^* - \tau$ under the two cases (equality in the two expressions) and dividing gives the corollary. □

A.3. **f-DP Calculations.** The f-DP framework [Dong et al., 2021, 2022] can also be used to analyze this thresholding algorithm. f-DP is a generalization of differential privacy which introduces a *tradeoff* function $f : [0, 1] \to [0, 1]$, which is any continuous convex function that is non-increasing and $f(y) \leq 1 - y$ for all $y \in [0, 1]$. If a mechanism M satisfies f-DP, then it guarantees the following. If two datasets X_1 and X_2 differ on one individual, then any (possibly randomized) classification rule that uses the output of M and tries to determine whether the input was X_1 or X_2 is going to have a tradeoff between its false positive rates and true positive and false negative rates. Specifically, if the false positive rate is at most α then the true positive rate is at most $1 - f(\alpha)$ and the false negative rate is at least $f(\alpha)$. Formally,

Definition A.1 (Dong et al. [2021, 2022])**.** Let $f : [0, 1] \to [0, 1]$ be a continuous convex function that is non-increasing and such that $f(y) \leq 1 - y$ for all $y \in [0, 1]$. A mechanism M satisfies f-DP is for all datasets X_1 and X_2 that differ on on individual's data and for all measurable sets S, then

$$P(M(X_2) \in S) \leq 1 - f(P(M(X_1) \in S)).$$

Next, we recall some facts from Dong et al. [2021, 2022].

Lemma A.1 ([Dong et al., 2022, Section 2.2])**.** Let mechanism M be the mechanism that adds $N(0, \sigma^2)$ noise to a query with L_2 sensitivity Δ. Then

- M satisfies μ-Gaussian DP with $\mu = \Delta/\sigma$; and
- The tradeoff function between Type I and Type II error is $f_\mu(\alpha) = \Phi(\Phi^{-1}(1 - \alpha) - \mu)$.

Lemma A.2 (Dong et al. [2021, 2022])**.** If M satisfies $(0, \delta)$-differential privacy then it satisfies f-DP with tradeoff function $f(y) = 1 - \delta - y$.

Given a sequence of mechanisms $M_1, \ldots, M_k$, f-DP allows one to compute their composition by evaluating the composed tradeoff function for each pair of neighbors X_1, X_2 and then taking the pointwise minimum of the resulting tradeoff functions. This is equivalent to fixing an ϵ, computing the δ approximate differential privacy parameter for each pair of neighbors separately and then taking the maximum δ.

For the case of the GSHM, let X_1 and X_2 be neighbors where, without loss of generality, X_1 is obtained from X_2 by removing one person, say person j. Thus $H_=$ be the groups such that person j participates in those groups in X_2 and they have counts equal to τ. Thus the number of such groups is $a_=$. Let H_+ (resp., H_-) be those groups such that person j participates in them in X_2 have counts $> \tau$ (resp., $< \tau$). Thus the number of groups in H_- is a_- and the number of groups in H_+ is a_+. When just considering the neighboring pair X_1 and X_2, the GSHM mechanism has the same effect as the following three mechanisms:

- M_- which just looks at the groups in H_- and always outputs $\emptyset$.
- $M_=$ which just looks at the groups in $H_=$ and produces one output per group. For each group, it adds $N(0, \sigma^2)$ noise to the group. If the true count is $< \tau$ or if the noisy count is $< \tau^*$ it outputs $\emptyset$. Otherwise it outputs the noisy count and noisy aggregate statistics.
- M_+ which just looks at the groups in $H_>$. It computes the noisy group counts and noisy group aggregates. For those groups whose noisy count is $\geq \tau^*$ it releases the noisy statistics, otherwise it suppresses them. This mechanism doesn't use a τ but it still coincides with the behavior of the GSHM algorithm on X_1 and X_2 because these groups are guaranteed to have at least τ people each.

The mechanism M_- is uninformative and has no privacy impact. The mechanism $M_=$ satisfies $(0, \delta)$-differential privacy, with $\delta = 1 - \Phi(\tau^* - \tau/\sigma)^{a_=}$, because if all the outputs are $\emptyset$ it is uninformative, but if some output is not $\emptyset$, then one can immediately tell that the input was X_2. The mechanism M_+ just behaves like a Gaussian mechanism followed by postprocessing and satisfies Gaussian DP with privacy parameter $\mu(a+)$ (defined in Equation 5.6).

Composition of $(0, \delta)$-DP and Gaussian mechanisms, and obtaining the (ϵ, δ) parameters of the composed mechanisms works as follows.

Lemma A.3 ([Dong et al., 2022, Section 3.3])**.** If M has tradeoff function f, then the composition of M and an $(0, \delta)$-differentially private mechanism has tradeoff function g lower bounded by

$$g(\alpha) = \begin{cases} (1-\delta) f(\frac{\alpha}{1-\delta}) & \text{if } \alpha \in [0, 1-\delta] \\ 0 & \text{otherwise.} \end{cases} \tag{A.17}$$

The convex conjugate g^* of a function g defined over $[0, 1]$ is:

$$g^*(y) = \sup_{\alpha \in [0,1]} y\alpha - g(\alpha)$$

Lemma A.4 ([Dong et al., 2021, proof of corollary 2.13])**.** The convex conjugate f_μ^* of the tradeoff function f_μ (of the Gaussian mechanism) is

$$f_\mu^*(y) = y\Phi\left(-\frac{\mu}{2} - \frac{1}{\mu}\log(-y)\right) - \Phi\left(-\frac{\mu}{2} + \frac{1}{\mu}\log(-y)\right)$$

Lemma A.5 ([Dong et al., 2022, Section 3.3])**.** A mechanism M with tradeoff function lower bounded by f satisfies (ϵ, δ)-differential privacy for $\delta = 1 + f^*(-e^\epsilon)$.

Lemma A.6. Let f be a tradeoff function and define g as in Equation A.17. The the convex conjugate of g is $\max\{(1-\delta)f^*(y),\ y\}$.

Proof. We calculate as follows:

$$\begin{aligned} g^*(y) &= \max\left\{ \sup_{\alpha\in[0,1-\delta]} y\alpha - g(\alpha), \quad \sup_{\alpha\in(1-\delta,1]} y\alpha - g(\alpha) \right\} \\ &= \max\left\{ \sup_{\alpha\in[0,1-\delta]} y\alpha - (1-\delta)f(\frac{\alpha}{1-\delta}), \quad \sup_{\alpha\in(1-\delta,1]} y\alpha \right\} \\ &= \max\left\{ \sup_{\alpha\in[0,1-\delta]} y\alpha - (1-\delta)f(\frac{\alpha}{1-\delta}), \quad y \right\} \end{aligned}$$

(substituting $z = \alpha/(1-\delta)$)

$$\begin{aligned} &= \max\left\{ \sup_{z\in[0,1]} (1-\delta)yz - (1-\delta)f(z), \quad y \right\} \\ &= \max\left\{ (1-\delta) \sup_{z\in[0,1]} yz - f(z), \quad y \right\} \\ &= \max\{(1-\delta)f^*(y),\ y\}. \end{aligned}$$

□

Lemma A.7. Let M_1 be a mechanism that satisfies μ-Gaussian DP and let M_2 be a mechanism that satisfies $(0, \delta)$-DP. Then the composition of M_1 and M_2 satisfies (ϵ, δ^*)-DP for $\epsilon > 0$ with

$$\delta^* \geq\ \delta + (1-\delta)\left(\Phi\left(-\frac{\epsilon}{\mu} + \frac{\mu}{2}\right) - e^\epsilon \Phi\left(-\frac{\epsilon}{\mu} - \frac{\mu}{2}\right).\right).$$

In particular, this means that if the Gaussian mechanism satisfies (ϵ_2, δ_2)-DP, then the composed mechanism satisfies $(\epsilon_2, \delta + (1-\delta)\delta_2)$-DP.

Proof. By Lemma A.3, the tradeoff function of the composition of M_1 and M_2 is lower bounded by the function g in Equation A.17 in which f_μ is used in place of f. Combining Lemmas A.4 and A.6, the convex conjugate of g is

$$g^*(y) = \max\left\{ y,\ (1-\delta)y\Phi\left(-\frac{\mu}{2} - \frac{1}{\mu}\log(-y)\right) - (1-\delta)\Phi\left(-\frac{\mu}{2} + \frac{1}{\mu}\log(-y)\right) \right\}.$$

Substituting $y = -e^\epsilon$,

$$g^*(-e^\epsilon) = \max\left\{ -e^\epsilon,\ -(1-\delta)e^\epsilon\Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - (1-\delta)\Phi\left(-\frac{\mu}{2} + \frac{\epsilon}{\mu}\right) \right\}$$

and so by Lemma A.5, the composed mechanism satisfies (ϵ, δ)-DP for

$$\begin{aligned}
\delta &= 1 + g^*(-e^\epsilon) \\
&= 1 + \max\left\{-e^\epsilon,\ -(1-\delta)e^\epsilon \Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - (1-\delta)\Phi\left(-\frac{\mu}{2} + \frac{\epsilon}{\mu}\right)\right\} \\
&= \max\left\{1 - e^\epsilon,\ \delta + (1-\delta) - (1-\delta)e^\epsilon \Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - (1-\delta)\Phi\left(-\frac{\mu}{2} + \frac{\epsilon}{\mu}\right)\right\} \\
&= \max\left\{1 - e^\epsilon,\ \delta + (1-\delta)\left(1 - e^\epsilon \Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) - \Phi\left(-\frac{\mu}{2} + \frac{\epsilon}{\mu}\right)\right)\right\} \\
&= \max\left\{1 - e^\epsilon,\ \delta + (1-\delta)\left(-e^\epsilon \Phi\left(-\frac{\mu}{2} - \frac{\epsilon}{\mu}\right) + \Phi\left(\frac{\mu}{2} - \frac{\epsilon}{\mu}\right)\right)\right\} \\
&= \max\left\{1 - e^\epsilon,\ \delta + (1-\delta)\left(\Phi\left(-\frac{\epsilon}{\mu} + \frac{\mu}{2}\right) - e^\epsilon \Phi\left(-\frac{\epsilon}{\mu} - \frac{\mu}{2}\right)\right)\right\}.
\end{aligned}$$

Finally, note that for $\epsilon > 0$, the first term is negative, so the second term is the maximum. □

Using Lemma A.7, we can work out the (ϵ, δ^*) parameters that result from the composition of the tradeoff functions of $M_=$ and M_+ for the pair of neighboring databases X_1 and X_2. Specifically, we have $\mu = \mu(a_+)$ and $\delta = 1 - \Phi((\tau^* - \tau)/\sigma)^{a_=}$. Taking the maximum δ^* over all pairs of neighbors is the same as taking the maximum over all a_+ and $a_=$ that sum to C_u. The result is:

$$\begin{aligned}
\max_{a_= + a_+ = C_u} & 1 - \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \\
& + \Phi\left(\frac{\tau^* - \tau}{\sigma}\right)^{a_=} \left[\Phi\left(-\frac{\epsilon}{\mu(a_+)} + \frac{\mu(a_+)}{2}\right) - e^\epsilon \Phi\left(-\frac{\epsilon}{\mu(a_+)} - \frac{\mu(a_+)}{2}\right)\right].
\end{aligned}$$