Notes

Maxim Zhilyaev

August 8, 2020

1 Abstract

We study a variety of the shuffling protocols for reporting one-hot vectors from multiple users with respect to privacy, sensitivity and practicality. From a practical standpoint, the cost of shuffling is not zero. Too many shuffled records may render a particular protocol impractical, even though its other metrics show good performance. We specifically consider protocols that minimize the number (but not necessarily the size) messages between a user device and the shuffler.

Assuming that the data comes from a universe $\mathcal{X} = [d]$ of d elements. Each individual $i \in [n]$ of n users has a data element $x_i \in \mathcal{X}$. We will write a data entry in bold $\boldsymbol{x}_i \in \{0,1\}^d$ to be the one-hot vector where x_i is zero in every position except position $\boldsymbol{x}_i \in \mathcal{X}$, where it is one. Furthermore, we will denote a dataset $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ to be a collection of all users' one-hot vectors. We consider multiple mix-net protocols for reporting 1-hot vectors. A simple one would require each user to donate his data $\boldsymbol{x}_i$ in clear, but, in addition, inject some fake reports $z_j \in \mathcal{X}$ for $j \in [m]$, and corresponding one-hot vector notation $\boldsymbol{z}_j$, where each data entry is chosen uniformly at random from $\mathcal{X}$. We then pass $\{\boldsymbol{x}_i : i \in [n]\}$ and $\{\boldsymbol{z}_j : j \in [m]\}$ to an anonymizer that shuffles the data and makes it impossible to determine whether a data record is real or fake. We call it the "clear-fake records" protocol and show that it provides adequate protection with the cost proportional to $[d]$. Hence if dimensions are not large then the "clear-fake records" protocol is preferred for its simplicity.

When $[d]$ is significant, the cost of sending and shuffling many fake records becomes prohibitive. Another protocol is developed, which parameters are independent of $[d]$. It's called a "fake and flip" protocol, whereby a user still generates true and fake one-hot report vectors that are both randomized by bit flipping before being sent to the shuffler. This enables adequate protection at reasonable cost. Depending on the data collection setting, various flavors of the "fake and flip" protocol are discussed.

In discussing mathematical properties of the protocols involving randomization we will rely upon results received for a single dimension bit reporting. Which results we provide in the first sections, along with some theoretical result claiming that if a randomization algorithm $\mathcal{R}$ is (ϵ, δ)-private on a dataset of n elements, it's also (ϵ, δ)-private on a dataset of $n+1$ elements, that is adding more elements to the shuffled set does not reduce privacy. These results are, then, used to develop

bounds for each protocol.

2 Differential Privacy Setup

A record is an element of some space $\mathcal{D}$, and a database $\mathbf{x}$ is a vector of n records: $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{D}^n$. A randomized algorithm $\mathcal{R}$ maps the database into another space: $\mathcal{R} : \mathcal{D}^n \to \mathcal{S}$. The result of applying an algorithm to a database is termed an **transcript**. The notion of differential privacy for an algorithm $\mathcal{R}$ is that the resulting transcripts does not change substantially when a record in the database is modified, i.e., transcripts are not sensitive to particular individual records in the database. Hence, releasing transcript of $\mathcal{R}$ publicly will not jeopardize privacy, since information regarding individual records cannot be gained by analyzing the outcome of $\mathcal{R}(\mathbf{x})$.

Differential privacy for a randomized algorithm $\mathcal{R}$ is formulated by comparing the transcripts generated by applying $\mathcal{R}$ to two very similar databases $\mathbf{x}, \mathbf{x}' \in \mathcal{D}^n$. We say the databases **differ in one row** if $\sum_{i=1}^{n} I(x_i \neq x_i') = 1$. Such datasets are commonly called **neighboring** database or **neighbors**.

Definition. A randomized algorithm $\mathcal{R}$ is (ϵ, δ)-**differentially private** if, for any two databases $\mathbf{x}, \mathbf{x}' \in \mathcal{D}^n$ differing in one row,

$$\mathsf{P}[\mathcal{R}(\mathbf{x}) \in S] \leq \exp(\epsilon) \cdot \mathsf{P}[\mathcal{R}(\mathbf{x}') \in S] + \delta \tag{2.1}$$

for all $S \subset \mathcal{S}$ (measurable).

In other words, the outcomes from the two databases databases differing in one row are close in distribution, may be with the exception of very unlikely outcomes whose probability is less than δ

Definition. A randomized algorithm $\mathcal{R}$ generates point-wise (ϵ, δ)-**indistinguishable** outcomes for two databases $\mathbf{x}, \mathbf{x}' \in \mathcal{D}^n$ when

$$\mathsf{P}\left(\frac{\mathsf{P}[\mathcal{R}(\mathbf{x}) = s]}{\mathsf{P}[\mathcal{R}(\mathbf{x}') = s]} \leq \exp(\epsilon)\right) \geq 1 - \delta \tag{2.2}$$

Proposition 2.1. *A randomized algorithm $\mathcal{R}$ is (ϵ, δ)-**differentially private** if for every pair of neighboring databases $\mathbf{x}, \mathbf{x}' \in \mathcal{D}^n$, $\mathcal{R}$ generates point-wise (ϵ, δ)-**indistinguishable** outcomes. Per reference [2]. (Ryan Rogers, etc..)*

To restate.

$$\mathsf{P}\left(\frac{\mathsf{P}[\mathcal{R}(\mathbf{x}) = s]}{\mathsf{P}[\mathcal{R}(\mathbf{x}') = s]} \leq \exp(\epsilon)\right) \geq 1 - \delta, \text{ for any two neighbors } \mathbf{x}, \mathbf{x}' \tag{2.3}$$

$$\implies \mathcal{R} \text{ is } (\epsilon, \delta) - \textbf{differentially private} \tag{2.4}$$

Proposition 2.2. *If $\mathcal{S}$ if finite, there exists a set $S_m \subset \mathcal{S}$, such that it maximizes the difference*

$$\mathsf{P}[\mathcal{R}(\mathbf{x}) \in S] - \exp(\epsilon) \cdot \mathsf{P}[\mathcal{R}(\mathbf{x}') \in S]$$

Suppose $\mathcal{S}$ if finite. Then the for any $S \subset \mathcal{S}$ respective set inclusion probabilities are

$$\mathsf{P}[\mathcal{R}(\mathbf{x}) \in S] = \sum_{s \in S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s]$$
$$\mathsf{P}[\mathcal{R}(\mathbf{x}) \in S] = \sum_{s \in S} \mathsf{P}[\mathcal{R}(\mathbf{x}') = s]$$

Consider a set $S_m \subset \mathcal{S}$ containing all and only $s \in S$, such that $\mathsf{P}[\mathcal{R}(\mathbf{x}) = s] > \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s]$, that is - all point-wise distinguishable values of s. The $\sum_{s \in S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s]$ reaches maximum in S_m. Indeed, any set S different from S_m will contain either point-wise indistinguishable values of s and they would reduce the sum or miss point-wise distinguishable values of s, which would also reduce the sum. More formally:

$$\begin{aligned}
&\sum_{s \in S_m} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] - \sum_{s \in S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] \\
= &\sum_{s \in S_m \setminus S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] - \sum_{s \in S \setminus S_m} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] \\
&\sum_{s \in S \setminus S_m} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] < 0, \ \textit{since } \forall s \in S \setminus S_m, \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] < \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] \\
&\sum_{s \in S_m \setminus S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] \geq 0, \ \textit{since } \forall s \in S_m \setminus S, \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] > \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] \\
\textit{hence } &\sum_{s \in S_m \setminus S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] - \sum_{s \in S \setminus S_m} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] > 0
\end{aligned}$$

Proposition 2.3. *For finite domains $\mathcal{S}$ the (ϵ, δ)-**differential privacy** and (ϵ, δ)-**indistinguishability** are equivalent - one implies the other.*

*Suppose $\mathcal{S}$ if finite. Then the (ϵ, δ)-**differentially private** condition holds if the following holds:*

$$\sum_{s \in S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] \leq \delta, \forall S \subset \mathcal{S} \tag{2.5}$$

And there exists the maximal S_m containing all and only point-wise distinguishable values of s. S_m maximizes the difference below.

$$\mathsf{P}[\mathcal{R}(\mathbf{x}) \in S] - \exp(\epsilon) \cdot \mathsf{P}[\mathcal{R}(\mathbf{x}') \in S]$$

*The probability of s being point-wise distinguishable is exactly the probability $P[s \in S_m]$ since it contains just the distinguishable s, and hence if $\mathcal{R}$ is (ϵ, δ)-**differentially private**, then it's also (ϵ, δ)-**indistinguishable**.*

*Now assume $\mathcal{R}$ is (ϵ, δ)-**indistinguishable**, then by definition*

$$\begin{aligned}
&\mathsf{P}(\mathsf{P}[\mathcal{R}(\mathbf{x}) = s] > \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s]) < \delta \\
\implies &P(\mathsf{P}[\mathcal{R}(\mathbf{x}) \in S_m] > \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') \in S_m]) < \delta \\
\implies &\sum_{s \in S_m} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] < \delta \\
\implies &\forall S \subset \mathcal{S}, \sum_{s \in S} \mathsf{P}[\mathcal{R}(\mathbf{x}) = s] - \exp(\epsilon)\,\mathsf{P}[\mathcal{R}(\mathbf{x}') = s] < \delta \qquad \textit{by maximality of } S_m
\end{aligned}$$

3 Single record shuffling protocol

There are n users, each holding a user value $x_i \in \mathcal{X}$. User values form a database of user records a dataset $\boldsymbol{x} = \{x_1, \ldots, x_n\}$. Each user applies a randomization procedure $\mathcal{R}(x) : \mathcal{R} : \mathcal{X} \to \mathcal{S}$, then submits $\mathcal{R}(x_i)$ to an anonymizer that shuffles the data and makes it impossible to determine which user submitted a record. We call this algorithm

$$M(x_1, \ldots, x_n) = \pi(\mathcal{R}(x_1), \ldots, \mathcal{R}(x_n)) \text{ where } \pi \text{ permutes its elements.}$$

If $\mathcal{S}$ a finite domain of dimension d, we can write the output of M as a histogram $h \in \mathbb{N}^d$ over the entire domain $\mathcal{S}$. $M(x_1, \ldots, x_n) = \{h_1, h_2, ..., h_d\}$. Where each histogram value h_i represents the number of users reported a particular value $s_i \in \mathcal{S}$. We then describe M as a mapping $M(\boldsymbol{x}) : M : \boldsymbol{x} \to \mathcal{H}_n = \mathbb{N}^d$, whereby an outcome of $M(\boldsymbol{x})$ is a particular histogram $h \in \mathcal{H}_n$, that is a histograms of d-bins and n-records.

Proposition 3.1. *If M is (ϵ, δ)-**differentially private** for n records, then it's (ϵ, δ)-**differentially private** for $n+1$ records. Thus, protection for n records is enough for any number of records above n*

*Assume M is (ϵ, δ)-**differentially private** for n values of $\boldsymbol{x}$. Then for any neighboring datasets $\boldsymbol{x}$ and $\boldsymbol{x}'$ of size n, and any $H_n \subset \mathcal{H}_n$*

$$\mathsf{P}[M(\mathbf{x}) \in H_n] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}') \in H_n] \leq \delta \tag{3.1}$$

$$\sum_{h \in H_n} \big(\mathsf{P}[M(\mathbf{x}) = h] - \exp(\epsilon)\,\mathsf{P}[M(\mathbf{x}') = h]\big) \leq \delta \qquad \mathcal{H}_n \textit{ is finite} \tag{3.2}$$

Add another user record x, then respective probabilities of an outcome $h = \{h_1, h_2, ..., h_d\}$ produced by $M(\mathbf{x} \cup x)$ and $M(\mathbf{x}' \cup x)$ is given below:

$$\mathsf{P}[M(\mathbf{x} \cup x) = h] = \sum_{i}^{d} \mathsf{P}[\mathcal{R}(x) = s_i]\,\mathsf{P}(M(\mathbf{x}) = \{h_1, \cdots, h_i - 1, \cdots, h_d\} \tag{3.3}$$

$$\mathsf{P}[M(\mathbf{x}' \cup x) = h] = \sum_{i}^{d} \mathsf{P}[\mathcal{R}(x) = s_i]\,\mathsf{P}(M(\mathbf{x}') = \{h_1, \cdots, h_i - 1, \cdots, h_d\} \tag{3.4}$$

Since $\mathcal{H}_{n+1}$ if finite, then for every set $H_{n+1} \subset \mathcal{H}_{n+1}$:

$$\begin{aligned}
&\mathsf{P}[M(\mathbf{x} \cup x) \in H_{n+1}] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}' \cup x) \in H_{n+1}] \\
&= \sum_{h \in H_{n+1}} \left(\mathsf{P}[M(\mathbf{x} \cup x) = h] - \exp(\epsilon)\, \mathsf{P}[M(\mathbf{x}' \cup x) = h] \right) \\
&= \sum_{h \in H_{n+1}} \begin{pmatrix} \sum_i^d \mathsf{P}[\mathcal{R}(x) = s_i]\, \mathsf{P}(M(\mathbf{x}) = \{h_1, \cdots, h_i - 1, \cdots, h_d\} \\ -\exp(\epsilon) \sum_i^d \mathsf{P}[\mathcal{R}(x) = s_i]\, \mathsf{P}(M(\mathbf{x}') = \{h_1, \cdots, h_i - 1, \cdots, h_d\} \end{pmatrix} \qquad (3.3) and (3.4)
\end{aligned}$$

After rearranging the oder of summation and combining terms with same $\mathsf{P}[\mathcal{R}(x) = s_i]$, we have:

$$\begin{aligned}
&\mathsf{P}[M(\mathbf{x} \cup x) \in H_{n+1}] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}' \cup x) \in H_{n+1}] \\
&= \begin{pmatrix} & \mathsf{P}[\mathcal{R}(x) = s_1] \sum_{h \in H_{n+1}} (\mathsf{P}(M(\mathbf{x}) = \{h_1 - 1, h_2, \cdots, h_d\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1 - 1, h_2, \cdots, h_d\}) \\ + & \mathsf{P}[\mathcal{R}(x) = s_2] \sum_{h \in H_{n+1}} (\mathsf{P}(M(\mathbf{x}) = \{h_1, h_2 - 1, \cdots, h_d\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1, h_2 - 1, \cdots, h_d\}) \\ & \cdots\cdots \\ + & \mathsf{P}[\mathcal{R}(x) = s_d] \sum_{h \in H_{n+1}} (\mathsf{P}(M(\mathbf{x}) = \{h_1, h_2, \cdots, h_d - 1\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1, h_2, \cdots, h_d - 1\}) \end{pmatrix}
\end{aligned}$$

Note that each sum of the form

$$\sum_{h \in H_{n+1}} \left(\mathsf{P}(M(\mathbf{x}) = \{h_1, \cdots, h_i - 1, \cdots, h_d\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1 - 1, \cdots, h_i - 1, \cdots, h_d\} \right)$$

is done over histograms of size n, and represents a differential privacy difference, which by assumption (3.2) bounded by δ

$$\mathsf{P}[M(\mathbf{x}) \in H_n] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}') \in H_n] \leq \delta$$

From here:

$$\begin{aligned}
&\mathsf{P}[M(\mathbf{x} \cup x) \in H_{n+1}] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}' \cup x) \in H_{n+1}] \\
&= \begin{pmatrix} & \mathsf{P}[\mathcal{R}(x) = s_1] \sum_{h \in H_{n+1}} (\mathsf{P}(M(\mathbf{x}) = \{h_1 - 1, h_2, \cdots, h_d\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1 - 1, h_2, \cdots, h_d\}) \\ + & \mathsf{P}[\mathcal{R}(x) = s_2] \sum_{h \in H_{n+1}} (\mathsf{P}(M(\mathbf{x}) = \{h_1, h_2 - 1, \cdots, h_d\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1, h_2 - 1, \cdots, h_d\}) \\ & \cdots\cdots \\ + & \mathsf{P}[\mathcal{R}(x) = s_d] \sum_{h \in H_{n+1}} (\mathsf{P}(M(\mathbf{x}) = \{h_1, h_2, \cdots, h_d - 1\} - \exp(\epsilon)\, \mathsf{P}(M(\mathbf{x}') = \{h_1, h_2, \cdots, h_d - 1\}) \end{pmatrix} \\
&= \sum_i^d \mathsf{P}[\mathcal{R}(x) = s_i] \left(\mathsf{P}[M(\mathbf{x}) \in H_n^i] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}') \in H_n^i] \right) \\
&\leq \sum_i^d \mathsf{P}[\mathcal{R}(x) = s_i] \delta \\
&= \delta \sum_i^d \mathsf{P}[\mathcal{R}(x) = s_i]
\end{aligned}$$

Since $\sum_i^d \mathsf{P}[\mathcal{R}(x) = s_i] = 1$, we arrive to the desired proof

$$\mathsf{P}[M(\mathbf{x} \cup x) \in H_{n+1}] - \exp(\epsilon) \cdot \mathsf{P}[M(\mathbf{x}' \cup x) \in H_{n+1}] \leq \delta$$

4 Single bit shuffling protocol

We now consider an important scenario where users report bit values, which they flip and send randomized result to the shuffler. There are n users, each holding a value $x \in \{0,1\}$. User bits form a dataset $D \subset \mathcal{D} = \{0,1\}^n$. Each user applies a randomization procedure $\mathcal{R}(x) : \mathcal{R} : \{0,1\} \to \{0,1\}$, which flips the original bit value with probability q and keeps it unchanged with probability $p = 1-q$. A user, then submits a randomized bit $\mathcal{R}(x_i)$ to an anonymizer that shuffles the data and makes impossible to trace a reported bit to its sender. As before, we consider algorithm

$$M(D) = \pi(\mathcal{R}(x_1), \ldots, \mathcal{R}(x_n)) \text{ where } \pi \text{ permutes its elements.}$$

Since the the reported bits are shuffled, the measurer can only add them up, and the outcome $M(D)$ is uniquely determined by the sum of the reported randomized bits $s \in \mathcal{S} = \{0, 1, \cdots, n\}$. Let D be a set of n bits and construct a neighboring pair of datasets by adding to D a set bit 1 and a zero bit 0. Note that the domain $\mathcal{S}$ is finite. Hence, due to proposition 2.3, there are two equivalent expression for the (ϵ, δ)-**differential privacy** condition.

The algorithm M is differentially private if, for any subset $Z \subset \mathcal{S}$

$$P[M(D \cup 0) \in Z] \leq e^{\epsilon} P[M(D \cup 1) \in Z] + \delta \tag{4.1}$$
$$\& P[M(D \cup 1) \in Z] \leq e^{\epsilon} P[M(D \cup 0) \in Z] + \delta \tag{4.2}$$

Equivalently, the algorithm M is differentially private, if it is (ϵ, δ)-**indistinguishable**:

$$P\left(e^{-\epsilon} \leq \frac{P(M(D \cup 0) = s)}{P(M(D \cup 1) = s)} \leq e^{\epsilon}\right) \geq 1 - \delta, \forall s \in \{0, 1, \cdots, n+1\} \tag{4.3}$$

The quantity in parenthesis is referred as a **privacy loss ratio** R. For every instance of D, one can express the privacy loss ratio for a particular outcome s as:

$$R(s|D) = \frac{P(M(D \cup 0) = s)}{P(M(D \cup 1) = s)} = \frac{P(s|D \cup 0))}{P(s|D \cup 1)} \tag{4.4}$$

Note that 4.3 requires both R and its reciprocal $1/R$ be δ-bounded by e^{ϵ}. Whereby the first case corresponds to replacing a user bit 1 with 0 bit, and the second is reversed. We shall show the symmetry of both ratios later in the sequel.

Conditioning on possible values the added bit could generate:

$$P(s|D \cup 0) = p \cdot P(s|D) + q \cdot P(s-1|D) \tag{4.5}$$

Indeed, if 0 bit is randomized to itself (with probability p), then s must be generated by D alone, while if 0 bit was flipped (with probability q) then D must generate $s-1$ total bit sum. Similarly

$$P(s|D \cup 1) = p \cdot P(s-1|D) + q \cdot P(s|D) \tag{4.6}$$

Combining two conditioning expressions into the privacy loss ratio one arrives to:

$$R(s|D) = \frac{p \cdot P(s|D) + q \cdot P(s-1|D)}{p \cdot P(s-1|D) + q \cdot P(s|D)} = \frac{p\frac{P(s|D)}{P(s-1|D)} + q}{p + q\frac{P(s|D)}{P(s-1|D)}} \tag{4.7}$$

Let $\rho(s) = \frac{P(s|D)}{P(s-1|D)}$ be a **probability ratio** between adjacent values of s. It's related to $R(s)$ as in:

$$R(s|D) = \frac{p\frac{P(s|D)}{P(s-1|D)} + q}{p + q\frac{P(s|D)}{P(s-1|D)}} = \frac{q + p\rho(s)}{p + q\rho(s)} \tag{4.8}$$

Let $g(x) = \frac{q+px}{p+qx}$, the function g is increasing over $x > 0$, since

$$g'(x) = \frac{p-q}{(p+qx)^2} > 0.$$

Which gives as an important corollary

Corollary 4.1. *Properties of monotonicity and extrema established for $\rho(s)$ carry over to $R(s)$.*

Lemma 4.1. *The privacy loss ratio $R(s)$ decreases monotonically as s grows, reaching its maximum in $s = 0$ and minimum in $s = n$.*

Proof. Suppose D contains m set bits, then the distribution of s is a sum of two binomial distributions, and is a Poisson Binomial distribution.

$$s \sim Bin(m, p) + Bin(n - m, q) \tag{4.9}$$

As show by Wang, Y. H. (1993). "On the number of successes in independent trials", for any Poisson Binomial distribution, the probability of consecutive values are related as follows

$$\begin{aligned} & P(s)^2 > P(s-1) \cdot P(S+1) \\ \implies & \rho(s-1) > \rho(s) \\ \implies & R(s-1) > R(s) && \text{by 4.1} \end{aligned}$$

□

According to corollary 2.2 there exists a set S_m containing only point-wise distinguishable values of s. Then, by lemma 4.1 such set includes only values from 0 to k for which $R(s) > e^\epsilon$. This immediately gives us an expression for (ϵ, δ)-**differential privacy** in binary case.

Lemma 4.2. *The algorithm M is (ϵ, δ)-**differentially** private if and only if*

$$(2pe^\epsilon - 1)P(k|D) - (e^\epsilon - 1)\sum_{i=0}^{k} P(i|D) \leq \delta, \forall k \in \{0, 1, \cdots, n\} \tag{4.10}$$

Proof. Consider probability $P[s \in S_m|D \cup 0]$, suppose $S_m = \{0, 1, \cdots, k\}$, then

$$\begin{aligned}
&P[s \in S_m|D \cup 0] \\
=&\sum_{i=0}^{k} [pP(i|D) + qP(i-1|D)] \\
=&pP(k|D) + (q+p)P(k-1|D) + (q+p)P(k-2|D) + \cdots + (q+p)P(0|D) \\
=&pP(k|D) + \sum_{i=0}^{k-1} P(i|D)
\end{aligned}$$

Similarly

$$P[s \in S_m|D \cup 0] = qP(k|D) + \sum_{i=0}^{k-1} P(i|D)$$

From here, the (ϵ, δ)-**differential privacy** condition fulfills when

$$P[s \in S_m|D \cup 0] \leq e^{\epsilon} P[s \in S_m|D \cup 1] + \delta \tag{4.11}$$

$$pP(k|D) + \sum_{i=0}^{k-1} P(i|D) - e^{\epsilon} \left(qP(k|D) + \sum_{i=0}^{k-1} P(i|D) \right) \leq \delta \tag{4.12}$$

$$(p - qe^{\epsilon})P(k|D) - (e^{\epsilon} - 1) \sum_{i=0}^{k-1} P(i|D) \leq \delta \tag{4.13}$$

$$(2pe^{\epsilon} - 1)P(k|D) - (e^{\epsilon} - 1) \sum_{i=0}^{k} P(i|D) \leq \delta \tag{4.14}$$

□

The formula 4.10 reveals the very nature of (ϵ, δ)-protection for the bit reporting. In essence, it's a difference in PDF and CDF of the underling distribution of s. The left tail probabilities grow as s increases, but so does the cumulative sum of them. At some point CDF becomes greater than the probability at given s, and then all consequent values of s are all pair-wise indistinguishable. As long as this difference stays under δ, the privacy is preserved.

Despite its simple form, the expression 4.10 does not immediately provide a simple way to express q from (ϵ, δ) and the size of the dataset. A theorem below makes possible to bound $R(s)$ of any distribution of the form 4.9 with the $R(s)$ of the dataset containing only 0 bits.

Proposition 4.1. *Denote a collections of n bits containing m set bits and $n - m$ zero bits as D_m. Further denote the corresponding quantities:*

- *privacy loss ratio at a particular value s as $R(s|D_m) = \frac{P(s|D_m \cup 0)}{P(s|D_m \cup 1)}$*
- *probability ratio at a particular value s as $\rho(s|D_m) = \frac{P(s|D_m)}{P(s-1|D_m)}$*

- *expected value of s as* $\mu_m = p \cdot m + q \cdot (n - m)$

Choose a distance l such that $l \geq npq$, then

$$\rho[\mu_m - l|D_r] \leq \rho[\mu_0 - (l + 2)|D_0]$$

That is, the probability ratio for any collection is bound by the probability ratio of the zero collection.

Proof. PROOF Is INVOLVED AND WILL BE GIVEN LATER IN APPENDIX. Max needs to fix Dave's notations, skipping for now. □

From lemma (4.1) and proposition 4.1 we immediately receive the corollary below that bounds the left tail of the distribution 4.9

Corollary 4.2. *For left deviations $l \geq npq$ from the mean the privacy loss ratio for the collection of n bits is bounded by the privacy loss ratio for the collection of n zero bits*

$$R[\mu_m - l|D_m] \leq R[\mu_0 - (l + 2)|D_0]$$

We present the properties of a zero valued dataset below along with derivation of the (ϵ, δ)-bound for such collection, and a formula to compute the the flipping frequency q. We then apply corollary 4.2 to bound an arbitrary set of bits.

4.1 properties of zero valued collection

Let D consists of n zero bits, the neighboring dataset D' is achieved by replacing a zero bit with set bit. The outcome of applying the algorithm M is a sum of randomized bits s. The following relationships hold.

$$\mu = E(s) = q \cdot n \tag{4.15}$$

$$P(s = i|D) = \binom{n}{i} q^i p^{n-i} \tag{4.16}$$

$$P(s = i|D') = \binom{n-1}{i} q^{i+1} p^{n-1-i} + \binom{n-1}{i-1} q^{i-1} p^{n-i+1} \tag{4.17}$$

$$R(i) = \frac{P(s = i|D)}{P(s = i|D')} = \frac{\binom{n}{i} q^i p^{n-i}}{\binom{n-1}{i} q^{i+1} p^{n-1-i} + \binom{n-1}{i-1} q^{i-1} p^{n-i+1}} \tag{4.18}$$

$$R(i) = \frac{1}{\frac{n-i}{n}\frac{q}{p} + \frac{i}{n}\frac{p}{q}} \tag{4.19}$$

By applying Chernoff bound to the distribution of s, we receive

$$P(|s - \mu| > t\mu) \leq 2e^{-\frac{t^2\mu}{3}} \tag{4.20}$$

Setting $t = \sqrt{\frac{3}{\mu} ln \frac{2}{\delta}}$ one arrives to

$$P\left(|s-\mu| > \sqrt{3nq \cdot ln\frac{2}{\delta}}\right) \leq \delta \tag{4.21}$$

Setting $l = \sqrt{3nq \cdot ln\frac{2}{\delta}}$, one is ensured that values of $P(s \in [\mu - l, \mu + l]) \geq 1 - \delta$. Conditioned on $s \in [\mu - l, \mu + l]$ we bound the privacy loss ratio $R(i)$ in this interval in the following way:

$$e^{\epsilon} \geq R(i) \geq e^{-\epsilon} \tag{4.22}$$

$$\implies \quad e^{-\epsilon} \leq \frac{1}{R(i)} \leq e^{\epsilon} \tag{4.23}$$

$$\implies \quad e^{-\epsilon} \leq \frac{n-i}{n}\frac{q}{p} + \frac{i}{n}\frac{p}{q} \leq e^{\epsilon} \tag{4.24}$$

We first bound the left side of the inequality, setting $i = \mu - l$

$$\frac{n-i}{n}\frac{q}{p} + \frac{i}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{4.25}$$

$$\frac{n-(\mu-l)}{n}\frac{q}{p} + \frac{\mu-l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{4.26}$$

$$\frac{n-(nq-l)}{n}\frac{q}{p} + \frac{nq-l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{4.27}$$

$$\frac{np+l}{n}\frac{q}{p} + \frac{nq-l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{4.28}$$

$$q + \frac{l}{n}\frac{q}{p} + p - \frac{l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{4.29}$$

$$1 - \frac{l}{n}\frac{p-q}{pq} \geq e^{-\epsilon}. \tag{4.30}$$

$$l \leq \left[1 - e^{-\epsilon}\right] \frac{npq}{p-q} \tag{4.31}$$

Plugging expression for l one arrives to the bound of q

$$\sqrt{3nq \cdot ln\frac{2}{\delta}} \leq \left[1 - e^{-\epsilon}\right] \frac{npq}{p-q} \tag{4.32}$$

$$\frac{(p-q)^2}{q \cdot p^2} \leq \frac{n\left[1 - e^{-\epsilon}\right]^2}{3ln\frac{2}{\delta}} \tag{4.33}$$

$$\text{since } \frac{(p-q)^2}{p^2} \leq 1 \text{ , then} \tag{4.34}$$

$$\frac{(p-q)^2}{q \cdot p^2} \leq \frac{1}{q} \leq \frac{n\left[1 - e^{-\epsilon}\right]^2}{3ln\frac{2}{\delta}} \tag{4.35}$$

$$q \geq \frac{3ln\frac{2}{\delta}}{n\left[1 - e^{-\epsilon}\right]^2} \tag{4.36}$$

In a similar fashion, one arrives to the right side bound

$$i = \mu + l \tag{4.37}$$

$$\frac{l}{n}\frac{p-q}{pq} \leq e^{\epsilon} - 1 \tag{4.38}$$

$$\sqrt{3nq \cdot ln\frac{2}{\delta}} \leq [e^{\epsilon} - 1]\,\frac{npq}{p-q} \tag{4.39}$$

$$q \geq \frac{3ln\frac{2}{\delta}}{n\,[e^{\epsilon} - 1]^2} \tag{4.40}$$

Since $e^{\epsilon} - 1 \geq 1 - e^{-\epsilon}$, if q bound (4.36) is met, then (4.40) is also met, which leads to the following lemma

Lemma 4.3. *The shuffling algorithm M on a collection of n zero bits is (ϵ, δ)-**differentially private** when the flipping frequency q of the randomization procedure $\mathcal{R}$ satisfy* (4.36)*.*

$$q \geq \frac{3 \cdot ln\frac{2}{\delta}}{n\,[1 - e^{-\epsilon}]^2}$$

4.2 Properties of an arbitrary distribution of the form 4.9

4.2.1 Symmetry

Consider a zero collection of D_0 of n zero bits, its distribution is binomial $s \sim Bin(n, q)$. The distribution for D_n - a collection of n set bits, is also binomial $s \sim Bin(n, p)$. These two distributions are mirror images of each other, which follows directly from the binomial probabilities for each collection, hence the corollary.

Corollary 4.3. *A binomial distribution with success probability q is a mirror image of a binomial distribution with success probability $p = 1 - q$*

Proof.

$$P(s = i|D_0) = \binom{n}{i} q^i p^{n-i}$$

$$P(s = n - i|D_n) = \binom{n}{n-i} p^{n-i} q^i$$

$$\binom{n}{n-i} = \binom{n}{i}$$

$$\implies P(s = i|D_0) = P(s = n - i|D_n)$$

□

This property extends to each D_m per the corollary below.

Corollary 4.4. *Each distribution D_m, where $m \leq n/2$, has a symmetrical mirror distribution D_{n-m}. Where probabilities are related as below.*

$$P(s = i|D_m) = P(s = n - i|D_{n-m})$$

Proof. Split D_m into two sets r (which contains m set bits), and z (which contains $n-m$ zero bits). The probability $P(s = i|D_m)$ can be written as a sum of conditional probabilities of generating certain number of success from r snd z.

$$P(s = i|D_m) = \sum_{j=0}^{i} P(s = j|r)P(s = i - j|z)$$

The collection D_{n-m} again contains two sets - r' with m zero bits, and z' with $n-m$ set bits, which sets have mirror distributions of r snd z. Hence

$$\begin{aligned}
&P(s = n - i|D_{n-m}) = \sum_{j=0}^{i} P(s = m - j|r')P(s = n - m - (i - j)|z') \\
&P(s = j|r) = P(s = m - j|r') \\
&P(s = i - j|z) = P(s = n - m - (i - j)|z') \\
\implies &\sum_{j=0}^{i} P(s = j|r)P(s = i - j|z) = \sum_{j=0}^{i} P(s = m - j|r')P(s = n - m - (i - j)|z') \\
\implies &P(s = i|D_m) = P(s = n - i|D_{n-m})
\end{aligned}$$

□

4.2.2 Chernoff bounds of an arbitrary distribution

This section proves that the Chernoff bound for zero valued collection is also valid to an arbitrary collection of bits. That is, for any collection of n bits containing m set bits the following holds.

$$P(|s - \mu_m| > t\mu_m) \leq 2e^{-\frac{t^2\mu_0}{3}} = 2e^{-\frac{t^2nq}{3}}$$

Proposition 4.2. *Chernoff right tail bound for the distribution $P(s = i|D_0)$ holds for any distribution $P(s = i|D_m)$*

Proof. Recall that the right distribution tail is bounded by its moment generating function, hence for any D_m

$$P(s > (1 + \alpha)\mu_m|D_m) \leq \frac{\mathbb{E}(e^{ts}|D_m)}{e^{t(1+\alpha)\mu_m}} \tag{4.41}$$

$$\mathbb{E}(e^{ts}|D_m) = (q + pe^t)^m(p + qe^t)^{n-m} \tag{4.42}$$

$$\implies P(s > (1 + \alpha)\mu_m|D_m) \leq \frac{(q + pe^t)^m(p + qe^t)^{n-m}}{e^{t(1+\alpha)\mu_m}}. \tag{4.43}$$

Taking the ratio of the right side of the equality for D_0 and D_m

$$\frac{e^{-t(1+\alpha)\mu_0}(p+qe^t)^n}{e^{-t(1+\alpha)\mu_m}(q+pe^t)^m(p+qe^t)^{n-m}} = \tag{4.44}$$

$$e^{t(1+\alpha)(\mu_m-\mu_0)}\left(\frac{p+qe^t}{q+pe^t}\right)^m \tag{4.45}$$

Chernoff right bound is obtained by bounding 4.43 at $t = ln(1+\alpha)$, at which value the ratio above resolves to

$$(1+\alpha)e^{(1+\alpha)(\mu_m-\mu_0)}\left(\frac{p+q(1+\alpha)}{q+p(1+\alpha)}\right)^m =$$
$$(1+\alpha)e^{(1+\alpha)m(p-q)}\left(\frac{1+q\alpha}{1+p\alpha}\right)^m =$$
$$(1+\alpha)\left(e^{1+\alpha}e^{p-q}\frac{1+q\alpha}{1+p\alpha}\right)^m$$

Note that since $\alpha > 0$, $p > q$ and $p + q = 1$, the expression in parenthesis is always greater than e.

$$e^{p-q} > 1$$
$$\frac{1+q\alpha}{1+p\alpha} \geq \frac{1}{1+\alpha}$$
$$\implies e^{1+\alpha}e^{p-q}\frac{1+q\alpha}{1+p\alpha} > \frac{e^{1+\alpha}}{1+\alpha} > e$$
$$\implies (1+\alpha)\left(e^{1+\alpha}e^{p-q}\frac{1+q\alpha}{1+p\alpha}\right)^m > 1$$
$$\implies \frac{\mathbb{E}(e^{ts}|D_0)}{e^{t(1+\alpha)\mu_0}} > \frac{\mathbb{E}(e^{ts}|D_m)}{e^{t(1+\alpha)\mu_m}}$$

Hence, the Chernoff right tail bound for the distribution D_0 also bounds right tail distribution of any D_m. □

Proposition 4.3. *Chernoff left tail bound of D_n holds for any D_m*

Proof. The left tail of D_n distribution is the right tail of D_0 distribution. Should there exists D_m which left tail not bound by the Chernoff bound of D_n, then, by corollary 4.4, there exists a distribution D_{n-m} which right tail is not bound by the Chernoff bound of D_0, which contradicts proposition 4.2 □

Since distributions of D_0 and D_n are symmetrical the symmetrical bound of D_0 also applies to D_n. Then by two propositions above, both tails are bound by either D_0 or D_n bounds, which finally gives the desired theorem.

Proposition 4.4. *Any distribution D_m is bounded by the Chernoff bound of D_0*

$$P(|s-\mu_m| > t\mu_m) \leq 2e^{-\frac{t^2\mu_0}{3}} = 2e^{-\frac{t^2nq}{3}} \tag{4.46}$$

4.2.3 Bounding privacy loss for an arbitrary distribution

Due to the distributional symmetry (corollary 4.4), the reciprocal of the privacy loss ratio $\frac{1}{R(s)}$ behaves like a mirror-image of $R(s)$. It monotonically grows with s, reaches maximum at n, and the corollary 4.2 holds for collection D_n with respect to $\frac{1}{R(s)}$.

Corollary 4.5. *For the right deviations $l \geq npq$ from the mean the reciprocal of privacy loss ratio for the collection of n bits is bounded by the reciprocal of the privacy loss ratio for the collection of n zero bits*

$$\frac{1}{R[\mu_m + l|D_m]} \leq \frac{1}{R[\mu_n + (l+2)|D_n]}$$

proposition 4.1

Lemma 4.4. *Bit flipping randomization procedure $\mathcal{R}$ applied to a collection of n set bits is (ϵ, δ)-private if the bit flipping frequency q satisfies* (4.36)

$$q \geq \frac{3 \cdot ln\frac{2}{\delta}}{n\left[1 - e^{-\epsilon}\right]^2}$$

The next corollary bounds the right tail of distribution

Corollary 4.6. *For the right deviations $l \geq npq$ from the mean, the privacy loss ratio for the collection of n bits is bounded by the privacy loss ratio for the collection of n set bits*

$$R[\mu_r + l | D_r] \leq R[\mu_n + l + 2 | D_n]$$

Proof. TODO - proving by symmetry between D_0 and D_n distributions. □

Using the bound (4.36) and the fact that all-zero and all-set bit collections provide identical bounds to the privacy loss ratio for the left and the right side of distribution, we finally arrive to an important theorem.

Proposition 4.5. *randomization procedure $\mathcal{R}$ is (ϵ, δ)-private on a collection n bits, when the flipping frequency q obeys the bound below*

$$q \geq \frac{3 \cdot ln\frac{2}{\delta}}{n\left[1 - e^{-\epsilon} - 2\frac{p-q}{npq}\right]^2} \tag{4.47}$$

Note that the term $2\frac{p-q}{npq}$ appeared due to $l+2$ correction of both corollaries above. For sufficiently large n, this term diminishes to zero, which simplifies the bound to the form of lemma (4.3)

5 Clear Reports

Assume that the data comes from a universe $\mathcal{X} = [d]$ of d elements. Each individual $i \in [n]$ of n users has a data element $x_i \in \mathcal{X}$. We will write a data entry in bold $\boldsymbol{x}_i \in \{0,1\}^d$ to be the one-hot vector where x_i is zero in every position except position $\boldsymbol{x}_i \in \mathcal{X}$, where it is one. Furthermore, we will denote a dataset $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ to be a collection of all users' one-hot vectors. We will have each user donate his data $\boldsymbol{x}_i$. Further, we will inject some fake reports $z_j \in \mathcal{X}$ for $j \in [m]$, and corresponding one-hot vector notation $\boldsymbol{z}_j$, where each data entry is chosen uniformly at random from $\mathcal{X}$. We then pass $\{\boldsymbol{x}_i : i \in [n]\}$ and $\{\boldsymbol{z}_j : j \in [m]\}$ to an anonymizer that shuffles the data and makes it impossible to determine whether a data record is real or fake. We call this algorithm

$$M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \pi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_m) \text{ where } \pi \text{ permutes its elements.}$$

We then compute the privacy loss of such an algorithm M. Equivalently, we could write the output as a histogram over the entire database, as in $M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \boldsymbol{x}_i + \sum_{j=1}^{m} \boldsymbol{z}_j$. Note

that rather than inject random noise to these counts, as in central differential privacy, we want to consider *anonymized differential privacy*, where data records are transmitted through a mix net to break any identifiers with each data entry and the server sees the aggregated records in some random order. In this model, there is no trusted server that injects noise to ensure DP. Rather, the user needs to only trust the anonymizer to shuffle real and fake records.

We then consider the privacy loss for a general mechanism M. Consider an outcome $h \in \mathbb{N}^d$, which is a histogram over the full dataset domain and neighboring datasets $\boldsymbol{x}$ and $\boldsymbol{x}'$.

$$L(h) = \log\left(\frac{\Pr[M(\boldsymbol{x}) = h]}{\Pr[M(\boldsymbol{x}') = h]}\right) \tag{5.1}$$

If we can bound $L(h)$ by ϵ for any outcome h then we say that M is ϵ-DP. If we can bound $L(h)$ by ϵ with probability at least $1-\delta$ where the randomness is over $h \sim M(\boldsymbol{x})$, then we say that M is (ϵ, δ)-DP.

We now focus on M being the mechanism described above, which injects m fake reports. We can then write the privacy loss in the following way where we assume, without loss of generality, that $\boldsymbol{x}$ and $\boldsymbol{x}'$ only differ in the first record, i.e. $\boldsymbol{x}_i = \boldsymbol{x}'_i$ for all $i \neq 1$.

$$\begin{aligned}
L(h) &= \log\left(\frac{\Pr[x_1 + \sum_{i=2}^n \boldsymbol{x}_i + \sum_{j=1}^m \boldsymbol{z}_j = h]}{\Pr[x'_1 + \sum_{i=2}^n \boldsymbol{x}_i + \sum_{j=1}^m \boldsymbol{z}_j = h]}\right) \\
&= \log\left(\frac{\Pr[\sum_{j=1}^m \boldsymbol{z}_j = h - \boldsymbol{x}_1 - \sum_{i=2}^n \boldsymbol{x}_i]}{\Pr[\sum_{j=1}^m \boldsymbol{z}_j = h - \boldsymbol{x}'_1 - \sum_{i=2}^n x_i]}\right) \\
&= \log\left(\frac{\Pr[\sum_{j=1}^m \boldsymbol{z}_j = h - \sum_{i=1}^n \boldsymbol{x}_i]}{\Pr[\sum_{j=1}^m \boldsymbol{z}_j = h - \sum_{i=1}^n \boldsymbol{x}_i - (\boldsymbol{x}'_1 - \boldsymbol{x}_1)]}\right)
\end{aligned}$$

We denote $\hat{h}$ to be the histogram of the fake records only $\hat{h} = h - \sum_{i=1}^n \boldsymbol{x}_i$, with respective counts in each histogram bin $\hat{h} = \{\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_d\}$. Then the privacy loss ratio can be written as:

$$L(h) = \log\left(\frac{\Pr[\sum_{j=1}^m \boldsymbol{z}_j = \hat{h}]}{\Pr[\sum_{j=1}^m \boldsymbol{z}_j = \hat{h} + \boldsymbol{x}_1 - \boldsymbol{x}'_1)]}\right)$$

The one-hot vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}'_1$ may only differ in two positions, let these positions be ℓ and ℓ'. $\boldsymbol{x}_1$ and $\boldsymbol{x}'_1$ must have opposite bit-values in positions i and i' (otherwise these vectors are identical). Without loss of generality assume $x_{1,\ell} = 1, x_{1,\ell'} = 0$ and $x'_{1,\ell} = 0, x'_{1,\ell'} = 1$. Adding $\boldsymbol{x}_1$ adds 1 to h_i, while subtracting $\boldsymbol{x}'_1$ removes 1 from h'_ℓ. Hence, if $\hat{h} = \{\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_\ell, \ldots, \hat{h}_{\ell'}, \ldots, \hat{h}_d\}$, then $\hat{h} + \boldsymbol{x}_1 - \boldsymbol{x}'_1 = \{\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_\ell + 1, \ldots, \hat{h}_{\ell'} - 1, \ldots, \hat{h}_d\}$.

Further, note that the count the fake bits $\hat{h}_\ell = \sum_{j=1}^{m} \boldsymbol{z}_{j,\ell}$ is a binomial distribution $h_\ell \sim \text{Bin}(m, 1/d)$, and the distribution of the fake bit counts across the bins takes the multinomial form $\hat{h} \sim \text{Multinomial}(m, (1/d, \cdots, 1/d))$. We then aim to bound the following quantity.

$$\begin{aligned} L(h) &= \log\left(\frac{\Pr[\sum_{j=1}^{m} \boldsymbol{z}_j = \hat{h}]}{\Pr[\sum_{j=1}^{m} \boldsymbol{z}_j = \hat{h} + \boldsymbol{x}_1 - \boldsymbol{x}_1']}\right) \\ &= \log\left(\frac{\binom{m}{\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_\ell, \ldots, \hat{h}_{\ell'}, \ldots, \hat{h}_d}}{\binom{m}{hath_1, \hat{h}_2, \ldots, \hat{h}_\ell + 1, \ldots, \hat{h}_{\ell'} - 1, \ldots, \hat{h}_d}}\right) \\ &= \log\left(\frac{\hat{h}_\ell + 1}{\hat{h}_{\ell'}}\right) \end{aligned}$$

It must be stressed that for a given pair of $(\boldsymbol{x}_1, \boldsymbol{x}_1')$, the corresponding position pair (ℓ, ℓ') where their bits are different is fixed, and the privacy loss only surfaces while observing the counts in the corresponding histogram bins $(h_\ell, h_{\ell'})$. It's entirely possible to see high ratio between counts in some other histogram bins, but it wouldn't contribute to the privacy loss for a concrete pair $(\boldsymbol{x}_1, \boldsymbol{x}_1')$. This observation allows us to focus only on a single pair of the histogram bins, ignoring the rest of the histogram as immaterial.

By applying a Chernoff bound, we have a bound (symmetric for the upper and lower tail) for the sum of the fake bits in any bin $\hat{h}_k = \sum_{j=1}^{m} z_{j,k}, k \in [d]$

$$\Pr\left[\left|\hat{h}_k - \frac{m}{d}\right| > t\frac{m}{d}\right] \leq 2e^{-\frac{m}{d}\frac{t^2}{3}}, \qquad \text{for } 0 < t < 1.$$

Choose t to fit the expression below, hence $t = \sqrt{\frac{3d}{m}\log\frac{4}{\delta}}$. Using this expression for t turns our Chernoff bound into the following,

$$Pr\left[\left|\bar{h}_k - \frac{m}{d}\right| > \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}\right] \leq \frac{\delta}{2} \tag{5.2}$$

Given any pair of the histogram bins at positions (ℓ, ℓ'), the probability of observing large deviation from the mean in at least one bin obeys the unions bound.

$$\Pr\left[\max_{k \in (\ell, \ell')} \left|\hat{h}_k - \frac{m}{d}\right| > \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}\right] \leq \delta$$

We then condition on the event that both counts $\hat{h}_l$ or $\hat{h}_{l'}$ fall in the interval $m/d \pm \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}$, which event occurs with probability at least $1 - \delta$. Conditioned on there being the given number of fake records, we can upper bound the privacy ratio $L(h)$

$$L(h) = \log\left(\frac{\hat{h}_\ell + 1}{\hat{h}_{\ell'}}\right) \leq \log\left(\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}\right) \leq \epsilon \tag{5.3}$$

From the above, we then get a condition on the number of fake records, m, to ensure DP.

$$\begin{aligned}
&\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}} \leq e^\epsilon \\
\implies &\frac{m}{d}(e^\epsilon - 1) - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}(e^\epsilon + 1) - 1 \geq 0 \\
\implies &\frac{m}{d}(e^\epsilon - 1) - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}(e^\epsilon + 1) \geq 0 \\
\implies &\sqrt{\frac{m}{d}}\left(\sqrt{\frac{m}{d}}(e^\epsilon - 1) - \sqrt{3\log\frac{4}{\delta}}(e^\epsilon + 1)\right) \geq 0 \\
\implies &\sqrt{\frac{m}{d}}(e^\epsilon - 1) - \sqrt{3\log\frac{4}{\delta}}(e^\epsilon + 1) \geq 0 \\
\implies &\sqrt{\frac{m}{d}} \geq \frac{\sqrt{3\log\frac{4}{\delta}}(e^\epsilon + 1)}{e^\epsilon - 1} \\
\implies &\frac{m}{d} \geq 3\log\frac{4}{\delta}\left(\frac{e^\epsilon + 1}{e^\epsilon - 1}\right)^2
\end{aligned}$$

As for the lower bound of $L(h)$, it's met if the upper bound is met.

$$\begin{aligned}
&L(h) = \log\left(\frac{\hat{h}_\ell + 1}{\hat{h}_{\ell'}}\right) \geq \log\left(\frac{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}\right) \geq -\epsilon \\
\implies &\frac{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}} \geq e^{-\epsilon} \\
\implies &\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1} \leq e^\epsilon \\
\implies &\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1} < \frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}} \leq e^\epsilon
\end{aligned}$$

6 Fake records and bit flipping

We now apply the exact same protocol, whereby users produce n real and m fake reports, but require each 1-hot vector to bit bit-flipped with frequency q. A randomization procedure $\mathcal{R}(y)$ flips each bit of an arbitrary 1-hot-vector y with probability q and keeps it the same with probability $p = 1 - q$. The resulting mechanism M_r becomes a permutation of randomized true and fake records:

$$M_r(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \pi(\mathcal{R}(\boldsymbol{x}_1), \ldots, \mathcal{R}(\boldsymbol{x}_n), \mathcal{R}(\boldsymbol{z}_1), \ldots, \mathcal{R}(\boldsymbol{z}_m)) \text{ where } \pi \text{ permutes its elements.}$$

Without los of generality assume $\boldsymbol{x}_1$ is replaced with $\boldsymbol{x}'_1$ to receive a neighboring data set $\boldsymbol{x}'$. The outcome is a histogram $g \in \mathbb{N}^d$ containing sums of randomized bits in each dimension, and the privacy loss:

$$L(g) = \log\left(\frac{\Pr[M_r(\boldsymbol{x}) = g]}{\Pr[M_r(\boldsymbol{x}') = g]}\right) \tag{6.1}$$

The combined set $\boldsymbol{x} + \boldsymbol{z}$ gives raise to a histogram $h \in \mathbb{N}^d$ received by applying the before discussed mechanism M (clear true records plus fake records). Hence, the $Pr[M_r(x) = g]$ can be written as a sum of probabilities over the domain of h:

$$Pr[M_r(x) = g] = \sum_{h \in \mathbb{N}^d} Pr\left[M(\boldsymbol{x}) = h\right] \cdot Pr[g|h]$$
$$\implies L(g) = \log\left(\frac{\sum_{h \in \mathbb{N}^d} Pr\left[M(\boldsymbol{x}) = h\right] \cdot Pr[g|h]}{\sum_{h' \in \mathbb{N}^d} Pr\left[M(\boldsymbol{x}') = h'\right] \cdot Pr[g|h']}\right)$$

Noting that

$$Pr\left[M(\boldsymbol{x}) = h\right] = Pr\left[M(\boldsymbol{x}') = h - \boldsymbol{x}_1 + \boldsymbol{x}'_1\right]$$

And regrouping the privacy loss ratio to have summands with same $Pr\left[M(\boldsymbol{x}) = h\right]$ in identical positions in numerator and denominator, and applying (9.2) we have:

$$\log\left(\max_{h \in \mathbb{N}^d}\left(\frac{Pr\left[M(\boldsymbol{x}) = h\right] \cdot Pr[g|h]}{Pr\left[M(\boldsymbol{x}') = h - \boldsymbol{x}_1 + \boldsymbol{x}'_1\right] \cdot Pr[g|h - \boldsymbol{x}_1 + \boldsymbol{x}'_1]}\right)\right) \geq L(g) \text{ , and}$$
$$L(g) \geq log\left(\min_{h \in \mathbb{N}^d}\left(\frac{Pr\left[M(\boldsymbol{x}) = h\right] \cdot Pr[g|h]}{Pr\left[M(\boldsymbol{x}') = h - \boldsymbol{x}_1 + \boldsymbol{x}'_1\right] \cdot Pr[g|h - \boldsymbol{x}_1 + \boldsymbol{x}'_1]}\right)\right)$$

Probabilities $Pr\left[M(\boldsymbol{x}) = h\right]$ and $Pr\left[M(\boldsymbol{x}') = h - \boldsymbol{x}_1 + \boldsymbol{x}'_1\right]$ cancel each other out in each ratio, hence giving us the bounds of the privacy loss over domain of h .

$$\log\left(\max_{h\in\mathbb{N}^d}\left(\frac{Pr[g|h]}{Pr[g|h-\boldsymbol{x}_1+\boldsymbol{x}'_1]}\right)\right) \geq L(g) \geq log\left(\min_{h\in\mathbb{N}^d}\left(\frac{Pr[g|h]}{Pr[g|h-\boldsymbol{x}_1+\boldsymbol{x}'_1]}\right)\right)$$

Since bits are flipped independently, the probability of finding certain number of bits in a particular histogram bin g_l depends only on how many not-yet-randomized set bits there are in the dimension l, that is the value of h_l. Such independence allows to re-write $Pr[g|h]$ as a product of probabilities for each dimension.

$$Pr[M_r(x)=g] = Pr[\{g_1,g_2,,\ldots,g_d\}|\{h_1,h_2,,\ldots,h_d\}] = \prod_{i=1}^{d} Pr[g_i|h_i]$$

Without loss of generality assume that $\boldsymbol{x}_1$ and $\boldsymbol{x}'_1$ differ in the first and second positions, that is $\boldsymbol{x}_{1,1}=1, \boldsymbol{x}_{1,2}=0$ and $\boldsymbol{x}'_{1,1}=0, \boldsymbol{x}'_{1,2}=1$, then

$$\begin{aligned}
& h-\boldsymbol{x}_1+\boldsymbol{x}'_1 = \{h_1-1, h_2+1,,\ldots,h_d\} \\
\implies & \frac{Pr[g|h]}{Pr[g|h-\boldsymbol{x}_1+\boldsymbol{x}'_1]} = \frac{Pr[\{g_1,g_2,,\ldots,g_d\}|\{h_1,h_2,,\ldots,h_d\}]}{Pr[\{g_1,g_2,,\ldots,g_d\}|\{h_1-1,h_2+1,,\ldots,h_d\}]} \\
\implies & \frac{Pr[g|h]}{Pr[g|h-\boldsymbol{x}_1+\boldsymbol{x}'_1]} = \frac{Pr[g_1|h_1]Pr[g_2|h_2]\prod_{i=3}^{d} Pr[g_i|h_i]}{Pr[g_1|h_1-1]Pr[g_2|h_2+1]\prod_{i=3}^{d} Pr[g_i|h_i]} \\
\implies & \frac{Pr[g|h]}{Pr[g|h-\boldsymbol{x}_1+\boldsymbol{x}'_1]} = \frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]}\cdot\frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]}
\end{aligned}$$

Plugging the above formula into (6.1), the privacy loss bounds become:

$$\begin{aligned}
& \max_{h\in\mathbb{N}^d}\left(\log\left(\frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]}\cdot\frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]}\right)\right) \geq L(g) \geq \min_{h\in\mathbb{N}^d}\left(\log\left(\frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]}\cdot\frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]}\right)\right) \\
\implies & \max_{h\in\mathbb{N}^d}\left(\log\left(\frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]}\right)\right) + \max_{h\in\mathbb{N}^d}\left(\log\left(\frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]}\right)\right) \geq L(g) \text{ , and} \\
& L(g) \geq \min_{h\in\mathbb{N}^d}\left(\log\left(\frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]}\right)\right) + \min_{h\in\mathbb{N}^d}\left(\log\left(\frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]}\right)\right)
\end{aligned}$$

Basically, the privacy loss is bound by the sum of privacy losses in each of the affected dimensions. Which enables relatively simple path to the bound. We employ the results of lemma (**??**), which says that if $\mathcal{R}$ is (ϵ,δ)-private on a collection r bits, it's also (ϵ,δ)-private on collection of $r+1$ bits. Therefore, a privacy loss could be bounded for the m fake records only, and that will provide sufficient noise for the extra n real records.

Suppose that r fake records (out of m) happened to have zero bits in the affected dimensions (1 and 2). We will show later that $r \to m$, for large d. Then we are bounding the product of privacy loss ratios in the affected dimensions to stay between $e^{-\epsilon}$ and e^{ϵ} with probability $1-\delta$.

$$P\left(e^{\epsilon} \geq \frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]}\cdot\frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]} \geq \frac{1}{e^{\epsilon}}\right) \leq 1-\delta \tag{6.2}$$

We achieve condition of (6.2) by bounding the ratio in each dimension separately. Suppose that the following holds

$$P\left(e^{\frac{\epsilon}{2}} \geq \frac{Pr[g_1|h_1]}{Pr[g_1|h_1-1]} \geq \frac{1}{e^{\frac{\epsilon}{2}}}\right) \leq 1-\delta/2$$

and

$$P\left(e^{\frac{\epsilon}{2}} \geq \frac{Pr[g_2|h_2]}{Pr[g_2|h_2+1]} \geq \frac{1}{e^{\frac{\epsilon}{2}}}\right) \leq 1-\delta/2$$

Then, by the union bound, the combined probability of either ratio falling outside its bound is δ, and with probability $1-\delta$, both ratios stay between $e^{-\frac{\epsilon}{2}}$ and $e^{\frac{\epsilon}{2}}$, hence the product of the ratios is bounded in $[e^{-\epsilon}, e^{\epsilon}]$.

Proposition 6.1.

7 Appendix B - proof of zero collection maxiality

It is instructive to first consider the case where each record in the collection consists of a single bit, as the expressions simplify considerably.

When $L=1$, each original and synthetic record is either 1 or 0, and the transformation R flips each record with probability q. Partition the collection space $\mathcal{D}^n$ according to the number of records that are 1:

$$\mathcal{D}^n = \bigcup_{m=0}^{n} \mathcal{D}_m^n \qquad \text{where} \qquad \mathcal{D}_m^n := \left\{\mathbf{x} \in \mathcal{D}^n \,:\, \sum_{i=1}^{n} I(x_i=1) = m\right\}.$$

For $\mathbf{x} \in \mathcal{D}_m^n$, we have

$$A(\mathbf{x}) = \Phi \circ R(\mathbf{x}) = \big(A_n(m), n - A_n(m)\big),$$

where

$$\begin{aligned} A_n(m) &:= \sum_{i=1}^{n} I(R(x_i)=1) = \sum_{i:\, x_i=1} I(R(1)=1) + \sum_{i:\, x_i=0} I(R(0)=1) \\ &\sim Bin(m,p) + Bin(n-m,q), \end{aligned}$$

a sum of two independent Binomial random variables with support $\{0,\ldots,n\}$. Furthermore, if $\mathbf{x} \in \mathcal{D}_m^n$ and $\mathbf{x}, \mathbf{x}'$ differ in one row, then $\mathbf{x}' \in \mathcal{D}_{m-1}^n \cup \mathcal{D}_{m+1}^n$. Defining

$$\pi_n(s;m) := \frac{\mathsf{P}[A_n(m)=s]}{\mathsf{P}[A_n(m+1)=s]} \qquad \text{for} \quad s \in \{0,\ldots,n\} \text{ and } m \in \{0,\ldots,n-1\},$$

the privacy ratio becomes

$$\pi\big((s,n-s);\, \mathbf{x}, \mathbf{x}'\big) = \begin{cases} \pi_n(s;m-1) & x_1 = 1 \\ \pi_n(s;m)^{-1} & x_1 = 0 \end{cases}.$$

Hence, in the $L=1$ case, it suffices to study the behaviour of $\pi_n(s;m)$.

7.1 Recursive relationship over n and m

The conditioning argument (**??**) yields a recursive relationship that lets us express the distribution of A_n in terms of that of A_{n-1}.

Recall that $A_n(m)$ is the outcome of applying the bit transformation R to n original bits, m of which are 1 and $n-m$ are 0. For $m \geq 1$, we can condition on the outcome of one of the original 1s:

$$A_n(m) \sim Ber(p) + Bin(m-1,p) + Bin(n-m,q) \sim Ber(p) + A_{n-1}(m-1),$$

and so

$$\mathsf{P}[A_n(m) = s] = p\,\mathsf{P}[A_{n-1}(m-1) = s-1] + q\,\mathsf{P}[A_{n-1}(m-1) = s]. \tag{7.1}$$

If $s = 0$, the first term on the RHS is interpreted as 0, and if $s = n$, the last term is. Similarly, for $m \leq n-1$, conditioning on an original 0,

$$A_n(m) \sim Ber(q) + Bin(m,p) + Bin(n-m-1,q) \sim Ber(q) + A_{n-1}(m),$$

from which

$$\mathsf{P}[A_n(m) = s] = q\,\mathsf{P}[A_{n-1}(m) = s-1] + p\,\mathsf{P}[A_{n-1}(m) = s]. \tag{7.2}$$

The recursive formulas (7.1) and (7.2) give some insight into how the distribution of $A_n(m)$ changes as n and m vary:

- as n increases by 1, the probabilities shift slightly, with $\mathsf{P}[A_n(m) = 0] \leq \mathsf{P}[A_{n-1}(m) = 0]$ and $\mathsf{P}[A_n(m) = s]$ falling between $\mathsf{P}[A_{n-1}(m) = s-1]$ and $\mathsf{P}[A_{n-1}(m) = s]$ for each $s \geq 1$ (i.e., the hump of the pmf shifts to the right);
- the distribution of $A_n(m+1)$ is not so different to that of $A_n(m)$, since $\mathsf{P}[A_n(m) = s]$ and $\mathsf{P}[A_n(m+1) = s]$ both lie between consecutive pmf values of $A_{n-1}(m)$. In particular, this allows us to express the privacy ratio $\pi(s;m)$ in terms of $A_{n-1}(m)$.

Writing $P_{n,m}(s) := \mathsf{P}[A_n(m) = s]$, the formulas (7.1) and (7.2) can be expressed as

$$P_{n,m}(s) = pP_{n-1,m-1}(s-1) + qP_{n-1,m-1}(s) \quad \text{for} \;\; 0 \leq s \leq n, \;\; 1 \leq m \leq n$$

and

$$P_{n,m}(s) = qP_{n-1,m}(s-1) + pP_{n-1,m}(s) \quad \text{for} \;\; 0 \leq s \leq n, \;\; 0 \leq m \leq n-1.$$

7.2 The probability ratio

The probabilities in the privacy ratio represent the likelihood of observing the same synthetic collection outcome given two different original collections. In the expression $\pi_n(s;m) = P_{n,m}(s)/P_{n,m+1}(s)$, the probabilities correspond to the distributions of $A_n(m)$ and $A_n(m+1)$, respectively. However, using the decomposition (7.1) and (7.2), we can rewrite π_n in terms of probabilites from the same distribution, which is more convenient to work with.

Applying (7.2) to the numerator and (7.1) to the denominator, we obtain

$$\pi_n(s;m) = \frac{qP_{n-1,m}(s-1) + pP_{n-1,m}(s)}{pP_{n-1,m}(s-1) + qP_{n-1,m}(s)} = \frac{q + p\frac{P_{n-1,m}(s)}{P_{n-1,m}(s-1)}}{p + q\frac{P_{n-1,m}(s)}{P_{n-1,m}(s-1)}}$$

for $s \geq 1$, and $\pi_n(0;m) \equiv p/q$. Define the **probability ratio**

$$\rho_n(s;m) := \frac{P_{n,m}(s)}{P_{n,m}(s-1)} \qquad \text{for } \ 1 \leq s \leq n$$

a ratio of consecutive probabilities from the distribution of $A_n(m)$, and let $g(x) = \frac{q+px}{p+qx}$, so that $\pi_n = g \circ \rho_{n-1}$. The function g is increasing over $x > 0$, since

$$g'(x) = \frac{p-q}{(p+qx)^2} > 0.$$

Therefore, properties of monotonicity and extrema established for ρ_n (for all n) carry over to π_n as well.

The probability ratio can be expressed in a concise way using the following recursive property of the distribution of $A_n(m)$.

Lemma 7.1. *For $n \geq 1$,*

$$(s+1)P_{n,m}(s+1) = \left\{(m-s)\frac{p}{q} + (n-m-s)\frac{q}{p}\right\}P_{n,m}(s) + (n-s+1)P_{n,m}(s-1) \tag{7.3}$$

for $0 \leq m \leq n$ and $0 \leq s \leq n-1$ (with $P_{n,m}(-1) := 0$).

Proof. We proceed by induction on n. Suppose first $n = 1$, $s = 0$. If $m = 1$, then $A_1(1) \sim Ber(p)$, and (7.3) holds since $(mp/q + (1-m)q/p) \cdot P_{1,1}(0) = p = P_{1,1}(1)$. The argument is similar when $m = 0$. Next assume (7.3) holds for $A_{n-1}(m)$, and suppose $m \leq n-1$ and $1 \leq s \leq n-2$. Observe

that

$$\begin{aligned}
&\left\{(m-s)\frac{p}{q}+(n-m-s)\frac{q}{p}\right\}P_{n,m}(s)+(n-s+1)P_{n,m}(s-1)\\
&\quad=\left\{(m-s)\frac{p}{q}+(n-1-m-s)\frac{q}{p}\right\}\big[qP_{n-1,m}(s-1)+pP_{n-1,m}(s)\big]\\
&\qquad\qquad+(n-1-s+1)\big[qP_{n-1,m}(s-2)+pP_{n-1,m}(s-1)\big]+\frac{q}{p}P_{n,m}(s)+P_{n,m}(s-1)\\
&\quad=p\left[\left\{(m-s)\frac{p}{q}+(n-1-m-s)\frac{q}{p}\right\}P_{n-1,m}(s)+(n-1-s+1)P_{n-1,m}(s-1)\right]\\
&\qquad+q\left[\left\{(m-(s-1))\frac{p}{q}+(n-1-m-(s-1))\frac{q}{p}\right\}P_{n-1,m}(s-1)\right.\\
&\qquad\qquad\left.+(n-1-(s-1)+1)P_{n-1,m}(s-2)\right]\\
&\qquad-\left(p+\frac{q^2}{p}\right)P_{n-1,m}(s-1)-qP_{n-1,m}(s-2)+\frac{q^2}{p}P_{n-1,m}(s-1)+qP_{n-1,m}(s)\\
&\qquad+qP_{n-1,m}(s-2)+pP_{n-1,m}(s-1)\\
&\quad=p(s+1)P_{n-1,m}(s+1)+qsP_{n-1,m}(s)+qP_{n-1,m}(s)\\
&\quad=(s+1)\big[qP_{n-1,m}(s)+pP_{n-1,m}(s+1)\big]=(s+1)P_{n,m}(s+1),
\end{aligned}$$

applying the induction hypothesis for s and for $s-1$ together with (7.2). If $s=0$, the argument is similar:

$$\begin{aligned}
\left\{m\frac{p}{q}+(n-m)\frac{q}{p}\right\}P_{n,m}(0)&=p\left\{m\frac{p}{q}+(n-1-m)\frac{q}{p}\right\}P_{n-1,m}(0)+qP_{n-1,m}(0)\\
&=pP_{n-1,m}(1)+qP_{n-1,m}(0)=P_{n,m}(1).
\end{aligned}$$

□

Given m, the probability ratio can be expressed using (7.3):

$$\begin{aligned}
\rho(s+1;m)&=\frac{m-s}{s+1}\frac{p}{q}+\frac{n-m-s}{s+1}\frac{q}{p}+\frac{n-s+1}{s+1}\frac{1}{\rho(s;m)}\\
\rho(1;m)&=m\frac{p}{q}+(n-m)\frac{q}{p}
\end{aligned}$$

Write

$$\eta(s):=\frac{n-s+1}{s+1}\qquad\text{and}\qquad\gamma_m(s):=\frac{1}{s+1}\left[(m-s)\frac{p}{q}+(n-m-s)\frac{q}{p}\right],$$

to get

$$\rho(s+1;m)=\eta(s)\rho(s;m)^{-1}+\gamma_m(s)\,;\quad\rho(1;m)=\gamma_m(0).\tag{7.4}$$

Note also that $\gamma_m(s)$ can be expressed in terms of $\mathsf{E}\,A_n(m)=\mu_m=nq+m(p-q)$:

$$(s+1)\gamma_m(s)=\frac{\mu_m-s}{pq}-n+2s.$$

The probability ratio has the following properties (TODO):

- decreasing in s for fixed m
- increasing in m for fixed s.

7.3 Bounding the probability ratio

For A to satisfy local differential privacy, the privacy ratio $\pi_n(s;m)$ must be bounded for all s except for a set of small probability with respect to the distribution $\mathsf{P}[A_n(m) = \cdot]$. Furthermore, this bound must hold regardless of the original collection described through m.

Fix $\delta > 0$. Given m, we show that the probability ratio for $s \in [\mu_m - \delta, n]$ is bounded by a value $\rho(s^*;0)$, where s^* is expressed in terms of $\mu_0 - \delta$. Together with the fact that $P_{n,m}(\mu_m - \delta) \leq P_{n,0}(\mu_0 - \delta)$ (TODO - is this necessary?), this implies that the bound for local differential privacy, required to hold for all m, can be computed in terms of $A_n(0)$ alone. Note that, since ρ is decreasing in s for fixed m, it is sufficient to consider the probability ratio at the smallest integer value belonging to the interval $[\mu_m - \delta, n]$.

TODO: how to handle the left endpoint. What is the min value of δ?

For $\delta > 0$ let $s_m(\delta) := \lceil \mu_m - \delta \rceil \vee 0$, and define $R_m(\delta) := \rho(s_m(\delta); m)$. Note that $R_m(\delta) \leq R_m(\delta')$ for $\delta \leq \delta'$, and $s_m(\delta + 1) = (s_m(\delta) - 1) \vee 0$.

Proposition 7.1.

$$R_m(\delta) \leq R_0(\delta + 2) \qquad \textit{for} \;\; m = 0, \ldots, n$$

provided $\delta > \sigma_0 + 1$, *where* $\sigma_0^2 = \mathsf{Var}\, A_n(0) = npq$.

Proof. Fix $\delta > \sigma_0 + 1$. (TODO) If $s_0(\delta) < 2$

Assume $s_0(\delta) \geq 2$, and suppose $R_m(\delta) > R_0(\delta + 2)$ for some m. Then, we have

$$R_0(\delta) \leq R_0(\delta + 1) \leq R_0(\delta + 2) < R_m(\delta) \leq R_m(\delta + 1),$$

implying that

$$R_0(\delta + 2)^{-1} = \frac{R_0(\delta + 1) - \gamma_0(s_0(\delta + 2))}{\eta(s_0(\delta + 2))} > \frac{R_m(\delta) - \gamma_m(s_m(\delta + 1))}{\eta(s_m(\delta + 1))} = R_m(\delta + 1)^{-1}$$

via (7.4). Write $s_m := s_m(\delta + 1)$, $s_0 := s_0(\delta + 2)$. Since $R_m(\delta) > R_0(\delta + 1)$ by assumption, we obtain:

$$\big\{\eta(s_m) - \eta(s_0)\big\} R_0(\delta + 1) + \big\{\eta(s_0)\gamma_m(s_m) - \eta(s_m)\gamma_0(s_0)\big\} > 0. \tag{7.5}$$

Furthermore,

$$\eta(s_m) - \eta(s_0) = \frac{n - s_m + 1}{s_m + 1} - \frac{n - s_0 + 1}{s_0 + 1} = -\frac{(n + 2)(s_m - s_0)}{(s_0 + 1)(s_m + 1)},$$

and

$$\begin{aligned}
&\eta(s_0)\gamma_m(s_m) - \eta(s_m)\gamma_0(s_0) \\
&\quad = \frac{n - s_0 + 1}{s_0 + 1} \cdot \frac{(\mu_m - s_m)/pq - n + 2s_m}{s_m + 1} - \frac{n - s_m + 1}{s_m + 1} \cdot \frac{(\mu_0 - s_0)/pq - n + 2s_0}{s_0 + 1} \\
&\quad = \frac{(n+2)(s_m - s_0) + (\mu_0 s_m - \mu_m s_0)/pq + (n+1)[\mu_m - \mu_0 - (s_m - s_0)]/pq}{(s_0 + 1)(s_m + 1)},
\end{aligned}$$

so (7.5) implies

$$\begin{aligned}
&-(n+2)(s_m - s_0)(R_0(\delta + 1) - 1) + \\
&\qquad \frac{\mu_0(s_m - s_0) - m(p-q)s_0}{pq} + \frac{(n+1)[m(p-q) - (s_m - s_0)]}{pq} > 0. \qquad (7.6)
\end{aligned}$$

Now, let $\delta_0 := \delta - \{\lceil \mu_0 - \delta \rceil - (\mu_0 - \delta)\} = \mu_0 - \lceil \mu_0 - \delta \rceil$, i.e., $\delta_0 = \inf\{\lambda : s_0(\lambda) = s_0(\delta)\}$. Then $s_0(\delta) = s_0(\delta_0) = \mu_0 - \delta_0$, an integer, and $s_m(\delta_0) - s_m(\delta) \in \{0, 1\}$, since $0 \le \delta - \delta_0 < 1$. Consequently, since

$$s_m(\delta_0) - s_0(\delta_0) = \lceil \mu_0 + m(p-q) - \delta_0 \rceil - (\mu_0 - \delta_0) = \lceil m(p-q) \rceil,$$

$$\begin{aligned}
s_m - s_0 &= (s_m(\delta) - 1) - (s_0(\delta) - 2) = s_m(\delta) - s_m(\delta_0) + \lceil m(p-q) \rceil + 1 \\
&\in \big\{ \lceil m(p-q) \rceil, \lceil m(p-q) \rceil + 1 \big\},
\end{aligned}$$

and

$$\begin{aligned}
\mu_0(s_m - s_0) - m(p-q)s_0 &= \mu_0(s_m - s_0) - m(p-q)(s_0(\delta_0) - 2) \\
&= \mu_0[(s_m - s_0) - m(p-q)] + m(p-q)(\delta_0 + 2).
\end{aligned}$$

Applying these identities in (7.6) gives

$$-(n+2)(s_m - s_0)(R_0(\delta+1) - 1) + \frac{m(p-q)(\delta_0 + 2)}{pq} + \frac{n + 1 - \mu_0}{pq}\big(m(p-q) - (s_m - s_0)\big) > 0.$$

Since $s_m - s_0 \ge m(p-q)$,

$$(n+2)(R_0(\delta+1) - 1) < \frac{\delta_0 + 2}{pq}\frac{m(p-q)}{s_m - s_0} < \frac{\delta_0 + 2}{pq}. \qquad (7.7)$$

Next, recall that $R_0(\delta+1) = P_{n,0}(s_0(\delta+1))/P_{n,0}(s_0(\delta+1) - 1)$. Since $P_{n,0}(\cdot) = \mathsf{P}[Bin(n, q) = \cdot\,]$,

$$R_0(\delta+1) - 1 = \frac{n - s_0(\delta+1) + 1}{s_0(\delta+1)} \cdot \frac{q}{p} - 1 = \frac{\mu_0 - (\mu_0 - \delta_0 - 1) + q}{p(\mu_0 - \delta_0 - 1)} = \frac{\delta_0 + q + 1}{p(\mu_0 - \delta_0 - 1)}.$$

Hence, substituting this expression in (7.7) yields

$$\begin{aligned}
&(\delta_0 + 2)(\mu_0 - \delta_0 - 1) > (n+2)q(\delta_0 + q + 1) > \mu_0(\delta_0 + q + 1) \\
\iff\ &-\delta_0^2 - 3\delta_0 + 2\mu_0 - 2 > (1+q)\mu_0 \\
\iff\ &-\delta_0^2 - 3\delta_0 + npq > 0,
\end{aligned}$$

which requires that δ_0 lie between the roots of the quadratic equation. In particular,

$$\delta_0 \leq -\frac{3}{2} + \frac{1}{2}\sqrt{9 + 4npq} \leq -\frac{3}{2} + \frac{3}{2} + \sqrt{npq} = \sqrt{npq}.$$

Finally, since $0 \leq \delta - \delta_0 < 1$, we conclude that

$$\delta = \delta_0 + \delta - \delta_0 < \sigma_0 + 1,$$

contradicting our initial choice of δ. □

8 Appendix: Ratios of sums: properties

Here we establish some results around bounding and comparing ratios of sums, which will be useful in working with the privacy ratio.

Lemma 8.1. *Suppose $a_1, \ldots, a_m, b_1, \ldots, b_m \in \mathbb{R}$ with $b_i > 0$ all i. Then*

$$\max\left(\frac{a_1}{b_1}, \ldots, \frac{a_m}{b_m}\right) \geq \frac{a_1 + \cdots + a_m}{b_1 + \cdots + b_m} \geq \min\left(\frac{a_1}{b_1}, \ldots, \frac{a_m}{b_m}\right).$$

Proof. Write

$$\frac{a_1 + \cdots + a_m}{b_1 + \cdots + b_m} = \frac{a_1}{b_1}\frac{b_1}{b_1 + \cdots + b_m} + \cdots + \frac{a_m}{b_m}\frac{b_m}{b_1 + \cdots + b_m} = \sum_{i=1}^{m} \frac{a_i}{b_i}\lambda_i$$

where $\lambda_1 + \cdots + \lambda_m = 1$. Then

$$\frac{a_1 + \cdots + a_m}{b_1 + \cdots + b_m} = \sum_{i=1}^{m} \frac{a_i}{b_i}\lambda_i \leq \sum_{i=1}^{m} \max\left(\frac{a_i}{b_i}\right)\lambda_i = \max\left(\frac{a_i}{b_i}\right)\sum_{i=1}^{m}\lambda_i = \max\left(\frac{a_i}{b_i}\right)$$

The low bound is derived in a similar fashion. □

References

[1] A Note on Differential Privacy: Defining Resistance to Arbitrary Side Information. Shiva Prasad Kasiviswanathan Adam Smith [2] Privacy Odometers and Filters: Pay-as-you-Go Composition. Ryan Rogers, Aaron Roth, Jonathan Ullman, Salil Vadhan

9 IGNORE BELOW THIS LINE

Suppose a new 1-bit is added to both collections, then $R_{n+1}(S)$ is derived by conditioning

$$R_{n+1}(S) = \frac{P(S|D \cup 1)}{P(S|D' \cup 1)} = \frac{P(S|D)q + P(S-1|D)p}{P(S|D')q + P(S-1|D')p}$$

By lemma (9.2) and lemma (4.1) we have

$$\frac{P(S-1|D)}{P(S-1|D'} \geq \frac{P(S|D)q + P(S-1|D)p}{P(S|D')q + P(S-1|D')p} \geq \frac{P(S|D)}{P(SD'} \tag{9.1}$$

$$\implies R_n(S-1) \geq R_{n+1}(S) \geq R_n(S) \tag{9.2}$$

$$\implies R_n(S) \geq R_{n+1}(S+1) \geq R_n(S+1) \tag{9.3}$$

Suppose R_n is (ϵ, δ)-private. Since R_n is monotonically decreasing with S (lemma (4.1)), there exist two values $\alpha + \beta \leq \delta$, such that R_n is upper bounded on the left at a particular limiting value S_α

$$R_n(S_\alpha) \leq e^\epsilon \text{ and } P_n(S \leq S_\alpha) \leq \alpha \tag{9.4}$$

And it's low bounded on the right at a particular limiting value S_β

$$R_n(S_\beta) \geq \frac{1}{e^\epsilon} \text{ and } P_n(S \geq S_\beta) \leq \beta \tag{9.5}$$

Consider the left (upper) bound first, and recall that according to (9.3)

$$R_n(S_\alpha) \geq R_{n+1}(S_\alpha + 1) \geq R_n(S_\alpha + 1)$$

$R_{n+1}(S_\alpha + 1)$ is bounded because $R_n(S_\alpha)$ is bounded per (9.4). Hence, R_{n+1} could only be over the bound at S_α, however the cumulative sum of probabilities up to S_α is always less for $n+1$ bits than for n bits.

$$P_{n+1}(S \leq S_\alpha) \leq P_n(S \leq S_\alpha) \tag{9.6}$$

We shall prove (9.6) in a moment. The important fact is that R_n upper bounds R_{n+1} at the left tail of distribution of S.

Similarly,

As show by Wang, Y. H. (1993). "On the number of successes in independent trials", for any Poisson Binomial distribution, the probability of consecutive values are related as follows

$$P(S)^2 > P(S-1) \cdot P(S+1)$$
$$\implies \rho(S-1) > \rho(S)$$
$$\implies R(S-1) > R(S)$$

Lemma 9.1. *The privacy loss reduces as S increases, reaching its maximum in $S = 0$ and minimum in $S = N$.*

the privacy loss reduces as S increases, reaching its maximum in $S = 0$ and minimum in $S = N$.

Note that $\frac{P(S|D)}{P(S-1|D)}$ as a **probability ratio** between adjacent values of S. It's easy to see that privacy loss ratio maximizes when **probability ratio** maximizes.

Recall that $p > q$ and consider two positive values A and B

$$\begin{aligned} \frac{p \cdot A + q}{p + q \cdot A} &\geq \frac{p \cdot B + q}{p + q \cdot B} \\ p^2 A + q^2 B &\geq p^2 B + q^2 A \\ A(p^2 - q^2) &\geq B(p^2 - q^2) \\ A &\geq B \end{aligned}$$

The above confirms that the distributions with largest **probability ratio** also exhibit larger privacy loss ratio. Hence we can focus on studying **probability ratio** instead of privacy loss ratio and choose those D that demonstrate sharpest decrease of probabilities in the left tail.

Consider a collection D of N bits, subjected to randomization procedure $\mathcal{R}$, whereby a bit is flipped with probability q and kept unchanged with probability $p = 1 - q$. The outcome of $\mathcal{R}$ is a a - sum of bits after randomization. The neigboring set D_m is recieved form Assuming that D contains m set bits, we consider a privacy loss ratio R_s computed for the outcome s:

$$R_s = \frac{P(s|D)}{P(s-1|D)}$$

consisting of m ones and $N - m$ zeros. Denote probability of number of successes for that collection as $P(S|D)$. The probability ratio at s is given by:

$$R_s = \frac{P(s|D)}{P(s-1|D)}$$

Denote expectation of s as μ:

$$\mu = mp + (N - m)q$$

For simplicity, denote probabilities at s for D as:

$$P(s|D) = P_s$$

It's known that for all $s < \mu$, the ratio R_s is greater than 1 and increasing:

Property 1.

$$R_{s-1} = \frac{P_{s-1}}{P_{s-2}} > R_s = \frac{P_s}{P_{s-1}} \tag{9.7}$$

$$P_{s-1}^2 > P_s P_{s-2} \tag{9.8}$$

Create two collections by adding to D one 1 and one 0. Call them D_1 and D_0 respectively. The probability of observing s from D_1 the is given by:

$$P(s|D_1) = pP_{s-1} + qP_s$$

Similarly for the second collection (with extra 0):

$$P(s|D_0) = qP_{s-1} + pP_s$$

Now consider the probability ratio for the collections D_1 and D_0 collections at some s:

$$R_s(D_1) = \frac{pP_{s-1} + qP_s}{pP_{s-2} + qP_{s-1}} \tag{9.9}$$

$$R_s(D_0) = \frac{qP_{s-1} + pP_s}{qP_{s-2} + pP_{s-1}} \tag{9.10}$$

N user bits are subjected to

Lemma 9.2. *Suppose $a_1, \ldots, a_m, b_1, \ldots, b_m \in \mathbb{R}$ with $b_i > 0$ all i. Then*

$$\max\left(\frac{a_1}{b_1}, \ldots, \frac{a_m}{b_m}\right) \geq \frac{a_1 + \cdots + a_m}{b_1 + \cdots + b_m} \geq \min\left(\frac{a_1}{b_1}, \ldots, \frac{a_m}{b_m}\right).$$

Proof. Write

$$\frac{a_1 + \cdots + a_m}{b_1 + \cdots + b_m} = \frac{a_1}{b_1}\frac{b_1}{b_1 + \cdots + b_m} + \cdots + \frac{a_m}{b_m}\frac{b_m}{b_1 + \cdots + b_m} = \sum_{i=1}^{m} \frac{a_i}{b_i}\lambda_i$$

where $\lambda_1 + \cdots + \lambda_m = 1$. Then

$$\frac{a_1 + \cdots + a_m}{b_1 + \cdots + b_m} = \sum_{i=1}^{m} \frac{a_i}{b_i}\lambda_i \leq \sum_{i=1}^{m} \max\left(\frac{a_i}{b_i}\right)\lambda_i = \max\left(\frac{a_i}{b_i}\right)\sum_{i=1}^{m}\lambda_i = \max\left(\frac{a_i}{b_i}\right)$$

The low bound is derived in a similar fashion. □

10 JUNK

Given the independence

$$Pr[M_r(x) = g] = \sum_{h \in \mathbb{N}^d} Pr\left[M(\boldsymbol{x}) = h\right] \cdot Pr[g|h] = \sum_{h \in \mathbb{N}^d} \left(Pr\left[M(\boldsymbol{x}) = h\right] \cdot \prod_{i=1}^{d} Pr[g_i|h_i] \right)$$

$$\implies L(g) = \log \left(\frac{\sum_{h \in \mathbb{N}^d} \left(Pr\left[M(\boldsymbol{x}) = h\right] \cdot \prod_{i=1}^{d} Pr[g_i|h_i] \right)}{\sum_{h' \in \mathbb{N}^d} \left(Pr\left[M(\boldsymbol{x}') = h'\right] \cdot \prod_{i=1}^{d} Pr[g_i|h'_i] \right)} \right)$$

More formally, the value of g_l is a sum of binomial distributions Where r is the number of set bits (both clear and fake) in the dimension l. This allows us to

The value of g_l distributed as a sum of two binomial variables.

$$g_l \sim Bin(h_l, p) + Bin(n + m - h_l, q)$$

Applying (9.2) gives bounds of $R(S)$

$$\frac{p \cdot P(S|D) + q \cdot P(S-1|D)}{p \cdot P(S-1|D) + q \cdot P(S|D)} R(S|D) = \frac{p \cdot P(S|D) + q \cdot P(S-1|D)}{p \cdot P(S-1|D) + q \cdot P(S|D)} \tag{10.1}$$