Notes

Maxim Zhilyaev

March 8, 2020

1 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 $\boldsymbol{x} = \{x_1, \dots, x_n\}$ 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

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

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 $M(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \sum_{i=1}^n \boldsymbol{x}_i + \sum_{j=1}^m \boldsymbol{z}_j$. 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 $\boldsymbol{x}'$.

$$L(h) = \log\left(\frac{\Pr[M(\boldsymbol{x}) = h]}{\Pr[M(\boldsymbol{x}') = h]}\right) \tag{1.1}$$

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 $h_\ell \sim \mathrm{Bin}(m, 1/d)$, and the distribution of the fake bit counts across the bins takes the multinomial form $\hat{h} \sim \mathrm{Multinomial}(m, (1/d, \cdots, 1/d))$. 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,

$$Pr\left[\left|\bar{h}_k - \frac{m}{d}\right| > \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}\right] \leq \frac{\delta}{2} \tag{1.2}$$

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 $L(h)$

$$L(h) = \log\left(\frac{\hat{h}_\ell + 1}{\hat{h}_{\ell'}}\right) \leq \log\left(\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}\right) \leq \epsilon \tag{1.3}$$

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

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

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

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

$$\begin{aligned}
\implies &\frac{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}} \ge e^{-\epsilon} \\
\implies &\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1} \le e^{\epsilon} \\
\implies &\frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1} < \frac{m/d + \sqrt{\frac{3m}{d}\log\frac{4}{\delta}} + 1}{m/d - \sqrt{\frac{3m}{d}\log\frac{4}{\delta}}} \le e^{\epsilon}
\end{aligned}$$

2 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:

$$L(g) = \log \left(\frac{\Pr[M_r(\boldsymbol{x}) = g]}{\Pr[M_r(\boldsymbol{x'}) = g]} \right) \tag{2.1}$$

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:

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

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 (5.2) we have:

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

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.

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

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

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

Plugging the above formula into (2.1), 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 (3.3), 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 dimentions (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 both dimentions to stay between $e^{-\epsilon}$ and e^{ϵ} with probability $1 - \delta$.

$$P \left(e^{\epsilon} \geq \frac{Pr[g_1|h_1]}{Pr[g_1|h_1 - 1]} \cdot \frac{Pr[g_2|h_2]}{Pr[g_2|h_2 + 1]} \geq \frac{1}{e^{\epsilon}} \right) \leq 1 - \delta \tag{2.2}$$

We achieve condition of (3.40) by boiunding the ratio in each column separately with probability $1 - \delta/2$.

$$P \left(e^{\epsilon} \geq \frac{Pr[g_1|h_1]}{Pr[g_1|h_1 - 1]} \cdot \frac{Pr[g_2|h_2]}{Pr[g_2|h_2 + 1]} \geq \frac{1}{e^{\epsilon}} \right) \leq 1 - \delta \tag{2.3}$$

3 Appendix: Bit flipping in single dimension

There are $n+1$ single bit records being sent through a shuffler. Before sending, each bit is randomized with $\mathcal{R}$ - that is, flipped with probability q and kept unchanged with probability $p = 1-q$. the outcome S is the sum of randomized bits. Let D be a set of n bits and construct a neighboring pair of datasets by adding to D a set bit and a 0 zero bit. Then the privacy loss ratio at any given value of S is expressed as:

$$R(S|D) = \frac{P(S|D \cup 0)}{P(S|D \cup 1)} \tag{3.1}$$

Each probability allows conditioning on possible values the added bit could generate:

$$P(S|D \cup 0) = p \cdot P(S|D) + q \cdot P(S-1|D) \tag{3.2}$$

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

$$P(S|D \cup 1) = p \cdot P(S-1|D) + q \cdot P(S|D) \tag{3.3}$$

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

$$R(S|D) = \frac{p \cdot P(S|D) + q \cdot P(S-1|D)}{p \cdot P(S-1|D) + q \cdot P(S|D)} = \frac{p\frac{P(S|D)}{P(S-1|D)} + q}{p + q\frac{P(S|D)}{P(S-1|D)}} \tag{3.4}$$

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:

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

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

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

Therefore, properties of monotonicity and extrema established for $\rho(S)$ carry over to $R(S)$ as well.

If D contains m set bits, then the distribution of S is a sum of two binomial distributions (a Poisson Binomial distribution):

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

Lemma 3.1. *When 0-bit is replaced with a set bit, the resulting privacy loss ratio $R(S)$ decreases monotonically as S grows, reaching its maximum in $S = 0$ and minimum in $S = N$.*

Proof. 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

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

□

Lemma 3.2. *When 0-bit is replaced with a set bit, the resulting privacy loss ratio $R(S)$ decreases monotonically as S grows, reaching its maximum in $S = 0$ and minimum in $S = N$.*

Proof. 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

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

□

Lemma 3.3. *If randomization procedure $\mathcal{R}$ is (ϵ, δ)-private on a collection n bits, it's also (ϵ, δ)-private on a collection of $n + 1$ bits. In other words, if n bit are protected with flipping frequency q, then all collections of greater size are also protected with same q.*

Proof. Let D_n be a collection of n bits and derive a neighboring collection D'_n by replacing a single bit. Since $\mathcal{R}$ is (ϵ, δ)-private for n bits, then for any set of outcomes W

$$P(\mathcal{R}(D'_n) \in W) \leq \exp(\epsilon)P(\mathcal{R}(D_n) \in W) + \delta \tag{3.5}$$

$$\implies \qquad P(\mathcal{R}(D'_n) \in W) - \exp(\epsilon)P(\mathcal{R}(D_n) \in W) \leq \delta \tag{3.6}$$

Now add a 0 bit to both D_n and D'_n. The probability of the extended collections D_{n+1} and D'_{n+1} generating a particular outcome S is given by (3.2)

$$P(\mathcal{R}(D_{n+1}) = S) = P(\mathcal{R}(D_n) = S)\cdot p + P(\mathcal{R}(D_n) = S-1)\cdot q \tag{3.7}$$

Assume that $\mathcal{R}(D_{n+1})$ is not (ϵ, δ)-private, then there must exists a set W' such that

$$P(\mathcal{R}(D'_{n+1}) \in W') > \exp(\epsilon)P(\mathcal{R}(D_{n+1}) \in W') + \delta \tag{3.8}$$

Suppose W' contains r distinct outcomes $W' = \{S_1, S_2, \cdots, S_r\}$, then

$$P(\mathcal{R}(D_{n+1}) \in W') = \sum_i^r P(\mathcal{R}(D_{n+1}) = S_i) \tag{3.9}$$

$$= \sum_i^r \left[P(\mathcal{R}(D_n) = S_i) + P(\mathcal{R}(D_n) = S_i - 1) \cdot q\right] \text{ by (3.7)} \tag{3.10}$$

$$= p \sum_i^r P(\mathcal{R}(D_n) = S_i) + q \sum_i^r P(\mathcal{R}(D_n) = S_i - 1) \tag{3.11}$$

Define a set $W'' = \{S_1 - 1, S_2 - 1, \cdots, S_r - 1\}$, then (3.2) can be rewritten in the form membership probabilities.

$$P(\mathcal{R}(D_{n+1}) \in W') = p \sum_i^r P(\mathcal{R}(D_n) = S_i) + q \sum_i^r P(\mathcal{R}(D_n) = S_i - 1) \tag{3.12}$$

$$= p \cdot P(\mathcal{R}(D_n) \in W') + q \cdot P(\mathcal{R}(D_n) \in W'') \tag{3.13}$$

In the same fashion we arrive to the expression of $P(\mathcal{R}(D'_{n+1}) \in W')$

$$P(\mathcal{R}(D'_{n+1}) \in W') = p \cdot P(\mathcal{R}(D'_n) \in W') + q \cdot P(\mathcal{R}(D'_n) \in W'') \tag{3.14}$$

Plugging the above probabilities into (3.11), we arrive to an inequality that must hold if $\mathcal{R}(D_{n+1})$ is not (ϵ, δ)-private.

$$P(\mathcal{R}(D'_{n+1}) \in W') > \exp(\epsilon) P(\mathcal{R}(D_{n+1}) \in W') + \delta \tag{3.15}$$

$$P(\mathcal{R}(D'_{n+1}) \in W') - \exp(\epsilon) P(\mathcal{R}(D_{n+1}) \in W') > \delta \tag{3.16}$$

$$p \cdot P(\mathcal{R}(D'_n) \in W') + q \cdot P(\mathcal{R}(D'_n) \in W'') - p \cdot P(\mathcal{R}(D_n) \in W') - q \cdot P(\mathcal{R}(D_n) \in W'') > \delta \tag{3.17}$$

$$p \left[P(\mathcal{R}(D'_n) \in W') - P(\mathcal{R}(D_n) \in W')\right] + q \left[P(\mathcal{R}(D'_n) \in W'') - P(\mathcal{R}(D_n) \in W'')\right] > \delta \tag{3.18}$$

However (3.5) implies that

$$\begin{aligned}
& & \left[P(\mathcal{R}(D'_n) \in W') - P(\mathcal{R}(D_n) \in W')\right] \le \delta \\
&\text{and} & \left[P(\mathcal{R}(D'_n) \in W'') - P(\mathcal{R}(D_n) \in W'')\right] \le \delta \\
&\implies & p \left[P(\mathcal{R}(D'_n) \in W') - P(\mathcal{R}(D_n) \in W')\right] + q \left[P(\mathcal{R}(D'_n) \in W'') - P(\mathcal{R}(D_n) \in W'')\right] \le p\delta + q\delta \\
&\implies & p \left[P(\mathcal{R}(D'_n) \in W') - P(\mathcal{R}(D_n) \in W')\right] + q \left[P(\mathcal{R}(D'_n) \in W'') - P(\mathcal{R}(D_n) \in W'')\right] \le \delta(p + q) \\
&\implies & p \left[P(\mathcal{R}(D'_n) \in W') - P(\mathcal{R}(D_n) \in W')\right] + q \left[P(\mathcal{R}(D'_n) \in W'') - P(\mathcal{R}(D_n) \in W'')\right] \le \delta
\end{aligned}$$

Which contradicts (3.18) and proves the lemma. □

3.1 properties of zero valued collection

A homogenous collection of n zero bits is an important spacial case, hence we present findings for it below. Let D consists of n zero bits, then the neighbor D' is achieved by replacing a zero bit with set bit. The outcome of applying procedure c is a sum of randomized bits S. The following relationships hold.

$$\mu = E(S) = q \cdot n \tag{3.19}$$

$$P(S = i|D) = \binom{n}{i} q^i p^{n-i} \tag{3.20}$$

$$P(S = i|D') = \binom{n-1}{i} q^{i+1} p^{n-1-i} + \binom{n-1}{i-1} q^{i-1} p^{n-i+1} \tag{3.21}$$

$$R(i) = \frac{P(s = i|D')}{P(s = i|D)} = \frac{\binom{n}{i} q^i p^{n-i}}{\binom{n-1}{i} q^{i+1} p^{n-1-i} + \binom{n-1}{i-1} q^{i-1} p^{n-i+1}} \tag{3.22}$$

$$\frac{1}{R(i)} = \frac{n-i}{n}\frac{q}{p} + \frac{i}{n}\frac{p}{q} \tag{3.23}$$

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

$$P(|S - \mu| > t\mu) \leq 2e^{-\frac{t^2\mu}{3}} \tag{3.24}$$

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

$$P\left(|S - \mu| > \sqrt{3nq \cdot ln\frac{2}{\delta}}\right) \leq \delta \tag{3.25}$$

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:

$$e^{\epsilon} \geq R(i) \geq e^{-\epsilon} \tag{3.26}$$

$$\implies \qquad e^{-\epsilon} \leq \frac{1}{R(i)} \leq e^{\epsilon} \tag{3.27}$$

$$\implies \qquad e^{-\epsilon} \leq \frac{n-i}{n}\frac{q}{p} + \frac{i}{n}\frac{p}{q} \leq e^{\epsilon} \tag{3.28}$$

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

$$\frac{n-i}{n}\frac{q}{p} + \frac{i}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{3.29}$$

$$\frac{n-(\mu-l)}{n}\frac{q}{p} + \frac{\mu-l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{3.30}$$

$$\frac{n-(nq-l)}{n}\frac{q}{p} + \frac{nq-l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{3.31}$$

$$\frac{np+l}{n}\frac{q}{p} + \frac{nq-l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{3.32}$$

$$q + \frac{l}{n}\frac{q}{p} + p - \frac{l}{n}\frac{p}{q} \geq e^{-\epsilon} \tag{3.33}$$

$$1 - \frac{l}{n}\frac{p-q}{pq} \geq e^{-\epsilon}. \tag{3.34}$$

$$l \leq \left[1-e^{-\epsilon}\right]\frac{npq}{p-q} \tag{3.35}$$

Plugging expression for l one arrives to the bound of q

$$\sqrt{3nq \cdot ln\frac{2}{\delta}} \leq \left[1-e^{-\epsilon}\right]\frac{npq}{p-q} \tag{3.36}$$

$$\frac{(p-q)^2}{q \cdot p^2} \leq \frac{n\left[1-e^{-\epsilon}\right]^2}{3ln\frac{2}{\delta}} \tag{3.37}$$

$$\text{since } \frac{(p-q)^2}{p^2} \leq 1 \text{ , then} \tag{3.38}$$

$$\frac{(p-q)^2}{q \cdot p^2} \leq \frac{1}{q} \leq \frac{n\left[1-e^{-\epsilon}\right]^2}{3ln\frac{2}{\delta}} \tag{3.39}$$

$$q \geq \frac{3ln\frac{2}{\delta}}{n\left[1-e^{-\epsilon}\right]^2} \tag{3.40}$$

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

$$i = \mu + l \tag{3.41}$$

$$\frac{l}{n}\frac{p-q}{pq} \leq e^{\epsilon} - 1 \tag{3.42}$$

$$\sqrt{3nq \cdot ln\frac{2}{\delta}} \leq \left[e^{\epsilon}-1\right]\frac{npq}{p-q} \tag{3.43}$$

$$q \geq \frac{3ln\frac{2}{\delta}}{n\left[e^{\epsilon}-1\right]^2} \tag{3.44}$$

Note that since $x + 1/x \geq 2$, then

$$e^{\epsilon} - 1 \geq 1 - e^{-\epsilon}$$

Which implies that if q bound (3.40) is met, then (3.44) is also met, which leads to the following lemma

Lemma 3.4. *Bit flipping randomization procedure $\mathcal{R}$ applied to a collection of n zero bits is (ϵ, δ)-private if the bit flipping frequency q satisfies* (3.40)

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

4 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 4.1. *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. □

5 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 (5.2) and lemma (3.2) we have

$$\frac{P(S-1|D)}{P(S-1|D'} \geq \frac{P(S|D)q + P(S-1|D)p}{P(S|D')q + P(S-1|D')p} \geq \frac{P(S|D)}{P(SD'} \tag{5.1}$$

$$\implies R_n(S-1) \geq R_{n+1}(S) \geq R_n(S) \tag{5.2}$$

$$\implies R_n(S) \geq R_{n+1}(S+1) \geq R_n(S+1) \tag{5.3}$$

Suppose R_n is (ϵ, δ)-private. Since R_n is monotonically decreasing with S (lemma (3.2)), 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

$$R_n(S_\alpha) \leq e^\epsilon \text{ and } P_n(S \leq S_\alpha) \leq \alpha \tag{5.4}$$

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

$$R_n(S_\beta) \geq \frac{1}{e^\epsilon} \text{ and } P_n(S \geq S_\beta) \leq \beta \tag{5.5}$$

Consider the left (upper) bound first, and recall that according to (5.3)

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

$R_{n+1}(S_\alpha + 1)$ is bounded because $R_n(S_\alpha)$ is bounded per (5.4). 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.

$$P_{n+1}(S \leq S_\alpha) \leq P_n(S \leq S_\alpha) \tag{5.6}$$

We shall prove (5.6) 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

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

Lemma 5.1. *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.

$$R_{s-1} = \frac{P_{s-1}}{P_{s-2}} > R_s = \frac{P_s}{P_{s-1}} \tag{5.7}$$

$$P_{s-1}^2 > P_s P_{s-2} \tag{5.8}$$

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:

$$R_s(D_1) = \frac{pP_{s-1} + qP_s}{pP_{s-2} + qP_{s-1}} \tag{5.9}$$

$$R_s(D_0) = \frac{qP_{s-1} + pP_s}{qP_{s-2} + pP_{s-1}} \tag{5.10}$$

N user bits are subjected to

Lemma 5.2. *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. □

6 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 (5.2) gives bounds of $R(S)$

$$\frac{p \cdot P(S|D) + q \cdot P(S-1|D)}{p \cdot P(S-1|D) + q \cdot P(S|D)} R(S|D) = \frac{p \cdot P(S|D) + q \cdot P(S-1|D)}{p \cdot P(S-1|D) + q \cdot P(S|D)} \tag{6.1}$$