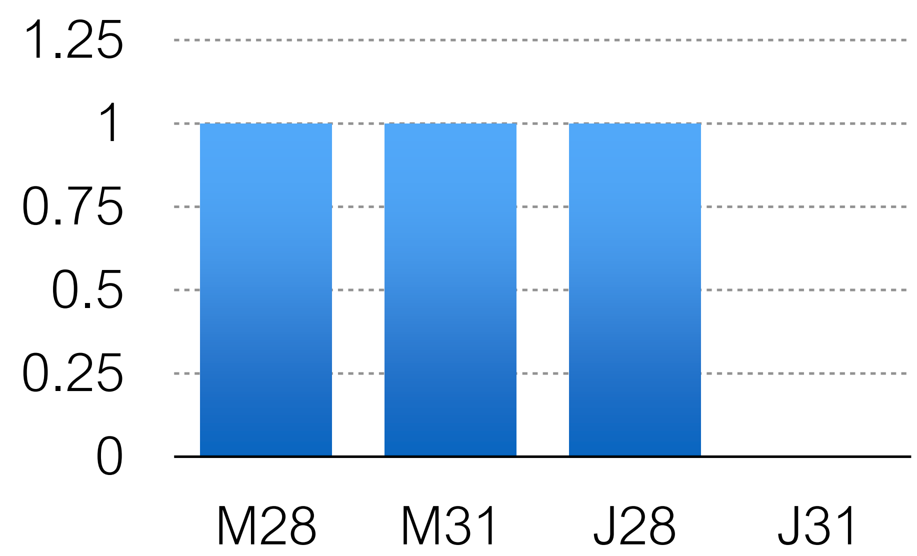
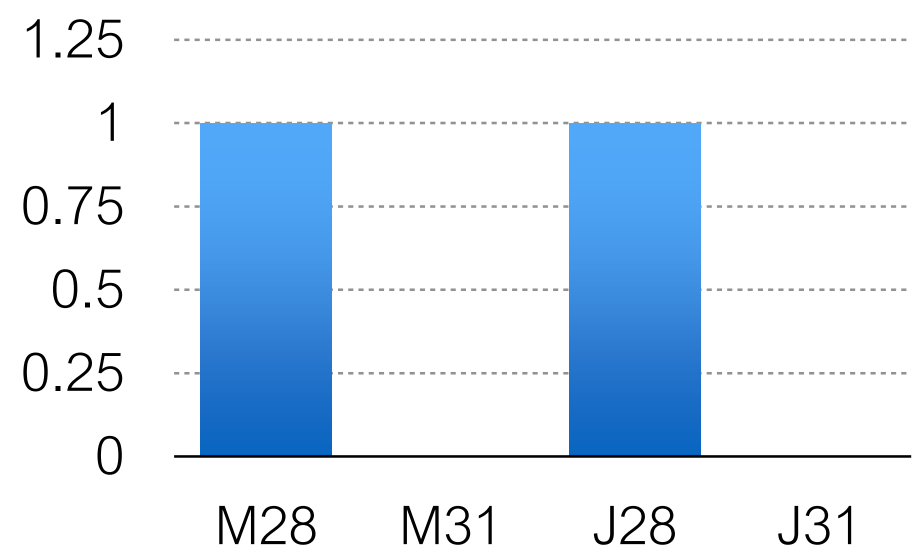
CONCENTRATION OF MEASURE FROM DIFFERENTIAL PRIVACY

Imagine a database representing a set of records. Can be represented as a frequency histogram instead. Neighboring databases differ in one record.

Name	Age
Marco	28
Julie	28
Marco	31



Name	Age
Marco	28
Julie	28



Definition 2.4 (Differential Privacy). A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ε, δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\varepsilon) \Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta,$$

IMPORTANT PROPERTY: Immunity to post-processing. Any algorithm that can be expressed as a randomized mapping run on top of a differentially private algorithm is differentially private.

Intuitive explanation: My participation in a survey should not compromise my privacy more than a ‘reasonable’ amount. Many ways of formalizing the exact nature of this guarantee- utility theoretically, cryptographically etc. But not focus of this talk.

Note that DP is also a stability notion- ‘small’ change in input should only produce ‘small’ change in the output.

<u>Why should you care?</u>

a) EVIL INSURANCE COMPANIES

b) Interesting Math Problems- for e.g. sample complexity of private PAC learning

c) New area- lots unsolved!

d) LOADS Of external applications

   i) Truthful Mechanisms
   ii) Generalization in Learning algorithms
   iii) Shadow tomography
   iv) Adversarial Robustness of Learning.
   many many more!

EXPONENTIAL MECHANISM- [MT07]:

A common primitive used in DP. Will use in this talk. The idea is for some query on a database- I assume the existence of a utility function between database/output pairs.

$$u : N^{\mathbb{X}} \times R \rightarrow \mathbb{R}$$

Define sensitivity as:

$$max_{r \in R} max_{x,y \text{ neighbors}} |u(x,r) - u(y,r)|$$

Definition 3.4 (The Exponential Mechanism). The exponential mechanism $\mathcal{M}_E(x, u, \mathcal{R})$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp(\frac{\varepsilon u(x,r)}{2\Delta u})$.

$$\frac{\Pr[\mathcal{M}_E(x, u, \mathcal{R}) = r]}{\Pr[\mathcal{M}_E(y, u, \mathcal{R}) = r]} = \frac{\left(\frac{\exp(\frac{\varepsilon u(x,r)}{2\Delta u})}{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(x,r')}{2\Delta u})}\right)}{\left(\frac{\exp(\frac{\varepsilon u(y,r)}{2\Delta u})}{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(y,r')}{2\Delta u})}\right)}$$

$$= \left(\frac{\exp(\frac{\varepsilon u(x,r)}{2\Delta u})}{\exp(\frac{\varepsilon u(y,r)}{2\Delta u})}\right) \cdot \left(\frac{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(y,r')}{2\Delta u})}{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(x,r')}{2\Delta u})}\right)$$

$$= \exp\left(\frac{\varepsilon(u(x, r') - u(y, r'))}{2\Delta u}\right)$$

$$\cdot \left(\frac{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(y,r')}{2\Delta u})}{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(x,r')}{2\Delta u})}\right) \qquad \boxed{e^{\epsilon \frac{u(y,r')-u(x,r')}{2\Delta u}} \; e^{\epsilon \frac{u(x,r')}{2\Delta u}}}$$

$$\leq \exp\left(\frac{\varepsilon}{2}\right) \cdot \exp\left(\frac{\varepsilon}{2}\right) \cdot \left(\frac{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(x,r')}{2\Delta u})}{\sum_{r'\in\mathcal{R}} \exp(\frac{\varepsilon u(x,r')}{2\Delta u})}\right)$$

$$= \exp(\varepsilon).$$

Accuracy Lemma:

$$\mathbb{E}[u(x,r)] \geq max_j u(x,j) - 2\frac{ln|R|}{\epsilon}$$

PROOF:

By definition, P(output = r) = $\frac{e^{\epsilon u(x,r)/2\Delta u}}{K}$

$$u(x,r) = 2\frac{\Delta u}{\epsilon}(lnK + ln\mathrm{P}(\text{output} = \text{r}))$$

$$E[u(x,r)] = \sum_{i=1}^{r} \mathrm{P}(\text{output} = \text{r})\left(2\frac{\Delta u}{\epsilon}(lnK + ln\mathrm{P}(\text{output} = \text{r}))\right)$$

$$= 2\frac{\Delta u}{\epsilon}\left(lnK + \sum_{i=1}^{r} \mathrm{P}(\text{output} = \text{r})ln\mathrm{P}(\text{output} = \text{r})\right)$$

negative of Entropy Function H

Upper bound on entropy

$$\begin{aligned} H(X) &= \mathbf{E}[\log_2(1/p(X))] \\ &\leq \log_2 \mathbf{E}[1/p(X)] \qquad \text{(by applying Jensen with the r.v. } 1/p(X)) \\ &= \log_2 \sum_{i=1}^{n} p(a_i)\cdot(1/p(a_i)) \\ &= \log_2 \sum_{i=1}^{n} 1 \quad = \quad \log_2 n. \end{aligned}$$

$$
\begin{aligned}
&lnK = \\
&ln\sum_{i=1}^{r} e^{\frac{\epsilon u(x,r)}{2\Delta u}} \geq ln\ \mathrm{max} e^{\frac{\epsilon u(x,r)}{2\Delta u}} = \\
&\mathrm{max}\ lne^{\frac{\epsilon u(x,r)}{2\Delta u}} = \\
&\frac{\epsilon}{2\Delta u}\mathrm{max}\ u(x,r)
\end{aligned}
$$

Substituting back, we get the lemma.

Some Math Stuff

Expectation of a function of a discrete random variable is defined a

$$\sum_{i=1}^{n} f(x)p(x)$$

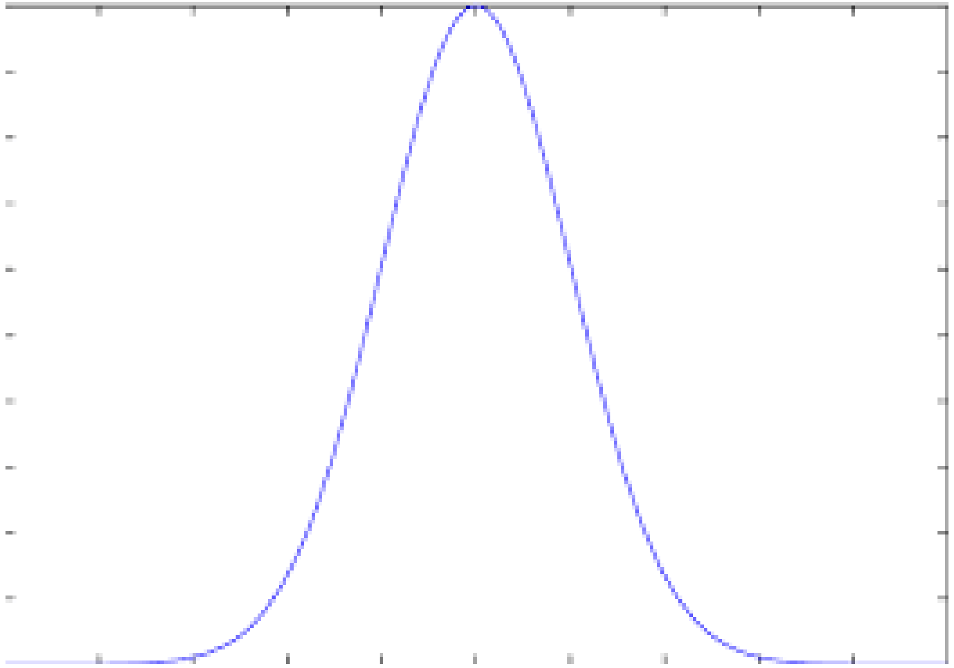
Will need Markov's inequality which states that for any positive ra

THIS TALK: CONCENTRATION OF MEASURE

How does a sum of independent and identically distributed random variables behave?

$$X = \frac{1}{n}(X_1 + X_2 + X_3 + ... + X_n)$$

How would we guess it behaves for finite n?- CLT intuition



Bounded in (a,b)- Hoeffding. 0-1 RVs- Chernoff. Examples:

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

One of the inequalities in Theorem 1 of Hoeffding (1963) states

$$\mathrm{P}\left(\overline{X} - \mathrm{E}\left[\overline{X}\right] \geq t\right) \leq e^{-2nt^2}$$

$$S_n = X_1 + \cdots + X_n$$

of the random variables:

$$\mathrm{P}(S_n - \mathrm{E}[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

$$\mathrm{P}(|S_n - \mathrm{E}[S_n]| \geq t) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

In this talk, going to show how to prove a statement very simila

Theorem 1.1 ([Ber24]). *If $X_1, \cdots, X_n$ are independent random variables supported on $[0,1]$ and $\mu_i = \mathbb{E}[X_i]$ for every i, then*

$$\forall \varepsilon \geq 0 \quad \mathbb{P}\left[\sum_{i=1}^{n} X_i - \mu_i \geq \varepsilon n\right] \leq e^{-\Omega(\varepsilon^2 n)}.$$

Approach: First, define Y to be

$$\sum_{i=1}^{n} X_i - \mu_i$$

I am going to consider many independent copies of Y-

Y1,Y2,…….. Ym. Going to reason about behavior of Y by real through the

oxy max(Y1,…….,Ym). Why is this a good proxy?

Lemma 1.2. *Let Y be a random variable and let $Y^1, Y^2, \cdots, Y^m$ be independent copies of Y. Then*

$$\mathbb{P}\left[Y \geq 2\mathbb{E}\left[\max\left\{0, Y^1, \cdots, Y^m\right\}\right]\right] \leq \frac{\ln(2)}{m}.$$

Proof. Let $y = 2\mathbb{E}\left[\max\{0, Y^1, \cdots, Y^m\}\right]$ and $\delta = \mathbb{P}[Y \geq y]$. By Markov's inequality,[2]

$$\mathbb{P}\left[\max\{0, Y^1, Y^2, \cdots, Y^m\} \geq y\right] \leq \frac{1}{2}.$$

However, if $\delta > \ln(2)/m$, then

$$\begin{aligned}\mathbb{P}\left[\max\{0, Y^1, Y^2, \cdots, Y^m\} \geq y\right] &= 1 - \mathbb{P}\left[\forall j \in [m] \quad Y^j < y\right] \\ &= 1 - \mathbb{P}[Y < y]^m = 1 - (1-\delta)^m \\ &> 1 - e^{-\delta m} > 1 - e^{-\ln(2)} = 1/2,\end{aligned}$$

which is a contradiction. Thus $\delta \leq \ln(2)/m$, as required. □

This puts us in business! What we need to do now is somehow bound E[max{0,Y1,Y2……,Ym}]. We'd hope that this is bounded by something that looks like log m (why?).

Because, if we wanted Pr(Y> ϵn)

$$P(Y > \epsilon n) \leq P(Y > O(logm))$$

$$\leq P(Y > 2\mathbb{E}[max(Y_1, Y_2, ...)]) \leq log2/m$$

For the above string of inequalities to hold, we need that log m <= ϵn

This would allow us to set m as exponential in epsilon n, which would make the right hand side exponentially small, as we want.

Let X be a random n x m matrix with all values between 0 and 1. Let X_i^j represent the (i,j) position entry in the matrix. Clearly this can be used to represent our situation, with each row representing a fresh choice of Y - μ

The proof is dependent on the following Lemma:

Lemma 2.1 (Main Lemma). *If* **X** *is a random* $n \times m$ *matrix with entries supported on* $[0,1]$ *and independent rows, then*

$$\forall \eta > 0 \quad \mathbb{E}\left[\max_{j \in [m]} \sum_{i=1}^{n} X_i^j\right] \leq e^{\eta} \max_{j \in [m]} \mathbb{E}\left[\sum_{i=1}^{n} X_i^j\right] + \frac{2\ln(m)}{\eta}.$$

The main idea of the proof is that we can consider an algorithm to select the row with maximum sum. But we want the algorithm to be 'stable' in a privacy sense while also preserving accuracy. This can be modeled using the exponential mechanism! Choose row j according to the exponential mechanism with a utility function u(X,j) = $\sum_{i=1}^{n} X_i^j$

Very natural utility function if you want to select the row with the max sum! Let this algorithm be S.

We know that S is $(\epsilon, 0)$ DP

Also, from the lemma:

$$\mathbb{E}[u(x,r)] \geq max_j u(x,j) - 2\frac{ln|R|}{\epsilon}$$

In this case, substituting for the utility function and |R| = m

$$\mathbb{E}[\sum_{j=1}^{n} X_i^j] \geq max_j \sum_{j=1}^{n} X_i^j - 2\frac{lnm}{\epsilon}$$

This expectation is from randomness internal to the algorithm and is true for EVERY matrix. Taking expectation on both sides in terms of randomness of the matrix:

$$\mathbb{E}_{X,S}[\sum_{j=1}^{n} X_i^j] \geq \mathbb{E}_X[max_j \sum_{j=1}^{n} X_i^j - 2\frac{lnm}{\epsilon}]$$

We're close! Just one more lemma to prove the bigger lemma:

Claim 2.4. *If* $\mathbb{E}\left[\sum_{i=1}^{n} X_i^j\right] \leq \mu$ *for all* $j \in [m]$, *then*

$$\mathop{\mathbb{E}}_{\mathbf{X},\mathcal{S}_\eta}\left[\sum_{i=1}^{n} X_i^{\mathcal{S}_\eta(\mathbf{X})}\right] \leq e^\eta \mu.$$

How are we now done? Because mu can be set to max E[summation]!

But before that assume I had 2 independent random variables that had the same distribution i.e. that they took the same values with the same probabilities.

Consider the random variables $f(x,y)$ and $f(y,x)$. I claim that they have the same distribution. To see this, note that the output of the former is $f(n,m)$ when $x=n$ and $y=m$ which happens with probability $P(x=n)P(y=m)$. But the latter has output $f(n,m)$ when $y=n$ and $x=m$ which happens with probability $P(y=n)P(x=m)$. These are the same since x and y are identical!

Let X and X' be 2 independent random matrices as suggested. Let

Let (X_{-i}, X_i') represent the matrix X with the ith row replaced by the ith row of X'.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{X},\mathcal{S}_\eta}\left[\sum_{i=1}^{n} X_i^{\mathcal{S}_\eta(\mathbf{X})}\right] &= \mathop{\mathbb{E}}_{\mathbf{X}}\left[\sum_{j=1}^{m}\sum_{i=1}^{n} \mathop{\mathbb{P}}_{\mathcal{S}_\eta}\left[\mathcal{S}_\eta(\mathbf{X}) = j\right] X_i^j\right] \\
&\leq \mathop{\mathbb{E}}_{\mathbf{X},\tilde{\mathbf{X}}}\left[\sum_{j=1}^{m}\sum_{i=1}^{n} e^{\eta} \mathop{\mathbb{P}}_{\mathcal{S}_\eta}\left[\mathcal{S}_\eta(\mathbf{X}_{-i}, \tilde{\mathbf{X}}_i) = j\right] X_i^j\right] \\
&= \mathop{\mathbb{E}}_{\mathbf{X},\tilde{\mathbf{X}}}\left[\sum_{j=1}^{m}\sum_{i=1}^{n} e^{\eta} \mathop{\mathbb{P}}_{\mathcal{S}_\eta}\left[\mathcal{S}_\eta(\mathbf{X}) = j\right] \tilde{X}_i^j\right] \\
&\leq \mathop{\mathbb{E}}_{\mathbf{X},\tilde{\mathbf{X}}}\left[\sum_{j=1}^{m} e^{\eta} \mathop{\mathbb{P}}_{\mathcal{S}_\eta}\left[\mathcal{S}_\eta(\mathbf{X}) = j\right] \mu\right] \\
&= e^{\eta}\mu.
\end{aligned}$$

Proposition 2.5 (Proposition 1.3). *Let $X_1, \cdots, X_n$ be independent random variables supported on $[0,1]$ and $\mu_i = \mathbb{E}[X_i]$ for each i. Define $Y = \sum_{i=1}^{n} X_i - \mu_i$. Fix $m \in \mathbb{N}$ and let $Y^1, \cdots, Y^m$ be independent copies of Y. Then*

$$\mathbb{E}\left[\max\left\{0, Y^1, \cdots, Y^m\right\}\right] \leq 4\sqrt{n \cdot \ln(m+1)}.$$

Proof. Firstly, if $m \geq e^n - 1$, then the result holds trivially as $\max\left\{0, Y^1, \cdots, Y^m\right\} \leq n$ with certainty. So we may assume $m < e^n - 1$.

Let $\mu = \sum_{i=1}^n \mu_i$. For each $i \in [n]$, let $X_i^1, \cdots, X_i^m$ be independent copies of X_i, so that $Y^j = \sum_{i=1}^n X_i^j - \mu_i$ for all $j \in [m]$. Let $X_i^{m+1} = \mu_i$ be a constant "dummy random varaible" for each i.

Now we apply Lemma 2.1 to the random matrix $\mathbf{X} \in [0,1]^{n \times (m+1)}$:

$$\forall \eta > 0 \quad \mathbb{E}\left[\max_{j \in [m+1]} \sum_{i=1}^n X_i^j\right] \leq e^\eta \max_{j \in [m+1]} \mathbb{E}\left[\sum_{i=1}^n X_i^j\right] + \frac{2\ln(m+1)}{\eta}.$$

By construction, $\mathbb{E}\left[\sum_{i=1}^n X_i^j\right] = \sum_{i=1}^n \mu_i = \mu$ for all $j \in [m+1]$. Also $\sum_{i=1}^n X_i^j = Y^j + \mu$ for all $j \in [m]$ and $\sum_{i=1}^n X_i^{m+1} = 0 + \mu$. Substituting in these expressions yields

$$\forall \eta > 0 \quad \mathbb{E}\left[\max\left\{Y^1 + \mu, Y^2 + \mu, \cdots, Y^m + \mu, 0 + \mu\right\}\right] \leq e^\eta \mu + \frac{2\ln(m+1)}{\eta}.$$

$$\mathbb{P}[Y \geq \varepsilon n] \leq \mathbb{P}\left[Y \geq 8\sqrt{n \ln(m+1)}\right] \leq \mathbb{P}\left[Y \geq 2\mathbb{E}\left[\max\{0, Y^1, \cdots, Y^m\}\right]\right] \leq \frac{\ln(2)}{m} \leq \frac{\ln(2)}{e^{\varepsilon^2 n/64} - 2}.$$

Thus

$$\mathbb{P}[Y \geq \varepsilon n] \leq \min\left\{1, \frac{\ln(2)}{e^{\varepsilon^2 n/64} - 2}\right\} \leq (2 + \ln 2) \cdot e^{-\varepsilon^2 n/64} \leq e^{1-\varepsilon^2 n/64}.$$

$\square$

Yayyy! Works for loads of other concentration inequalities as well. In fact can get new concentration inequalities! [NS18]

Where can you learn more?

a) Take Rachel Cumming's class- ISYE/CS 8803- Foundations of Data Privacy! Offered in the Fall (not sure if this fall - ask her!)
b) Algorithmic Foundations of Data Privacy- Roth and Dwork