1. Motivations

e.g. Robust De-anonymization of Large Sparse Datasets [Narayanan & Shmatikov '08]
   i.e. Breaking anonymity of Netflix Prize Dataset

e.g. Matching Known Patients to Health Records in Washington State Data [Sweeney '13]
   Breaking privacy with multiple overlapping datasets

e.g. Apple since '16, Google’s RAPPOR, TensorFlow Privacy, etc

Suppose STAT in Statistical Query model answers average salary about a company
What if I query average salary of a company, and do so again right after you leave the company?

Randomized response [Warner 1965]
Suppose you are taking a survey on a sensitive topic (e.g. have you taken drug illegally)
   Flip a fair coin, with prob 1/2, you answer Yes
   With prob 1/2, you answer honestly
If \( p \) fraction of population belongs to “Yes” group, in expectation \((1 + p)/2\) fraction will answer Yes
Survey researcher can deduce \( p \) from \((1 + p)/2\)
Even if you say Yes, you can plausibly deny

2. Definition

Dataset \( S = \{x_1, \ldots, x_m\} \subseteq X \) and another dataset \( S' \) differ in just one data point if
   \( S' \) is obtained from \( S \) by replacing \( x_i \) with \( x'_i \) for some \( 1 \leq i \leq m \)
A randomized algorithm \( A \) reads a dataset \( S \) and outputs \( y \in Y \)
   \( Y \) is called the range of \( A \)
   If \( A \) is a learning algorithm, then \( Y = \) hypothesis class \( \mathcal{H} \) of \( A \)
   But we also allow algorithms whose output isn’t a hypothesis, e.g. STAT(\( c, D \))

Definition 1. Randomized algorithm \( A \) satisfies \( \varepsilon \)-differential privacy if for any two datasets \( S, S' \) differing in just one data point, for any subset \( Y' \subseteq Y \) of outcomes of \( A \),
\[
P[A(S) \in Y'] e^{-\varepsilon} \leq P[A(S') \in Y'] \leq P[A(S) \in Y'] e^{\varepsilon}
\]
Since \( e^\varepsilon \approx 1 + \varepsilon \) and \( e^{-\varepsilon} \approx 1 - \varepsilon \)
Above definition requires \( P[A(S) \in Y'] / P[A(S') \in Y'] \) to be close to 1
If \( Y \) (range of \( A \)) is discrete, it’s equivalent to requiring that for any outcome \( y \in Y \) of \( A \),
\[
P[A(S) = y] e^{-\varepsilon} \leq P[A(S') = y] \leq P[A(S) = y] e^{\varepsilon}
\]
Original definition also covers the case where \( Y \) is continuous (e.g. \( Y = \mathbb{R} \))

3. Laplace mechanism

Suppose \( S \) consists of \( m \) points in \([0, b]\) and we want to estimate their average
Changing one data point in \( S \) changes the average by at most \( b/m \)
Laplace mechanism outputs the true average plus noise that is a Laplace random variable
Laplace distribution \( \text{Lap}(\mu, s) \) with mean \( \mu \) and scale \( s \) has density \( f(x \mid \mu, s) = \frac{1}{2s} \exp \left( -\frac{|x - \mu|}{s} \right) \)
Laplace mechanism
\[
\text{Output } v = \text{Lap}(a, b/\varepsilon m) \text{ where } a \text{ is the true average}
\]
In other words, \( v = a + x \) where \( x \) is the Laplace random variable \( \text{Lap}(0, b/\varepsilon m) \)
Smaller \( \varepsilon \) requires larger \( b/\varepsilon m \) i.e. more privacy requires larger noise
**Theorem 2.** Laplace mechanism satisfies $\varepsilon$-differential privacy

**Proof.** Fix two datasets $S$ and $S'$ differing in just one data point.

If $S$ has average $a$ and $S'$ has average $a'$, then $|a - a'| \leq b/m$.

Consider the ratio of densities $p_S(v)/p_{S'}(v)$ of outputting $v$ given $S$ (vs $S'$).

Ratio is smallest when $a' = a + b/m$ (the means are furthest apart) and $v \geq a'$

\[
\frac{p_S(v)}{p_{S'}(v)} \geq \frac{f(v \mid a, \frac{b}{m})}{\exp \left( -\frac{v-a}{b/\varepsilon m} \right)} = \exp \left( -\frac{v-a}{b/\varepsilon m} \right) \geq \exp(-\varepsilon)
\]

Last inequality follows from dropping the denominator (which is at most 1) and taking $v = a'$.

Likewise, ratio is largest when $a' = a + b/m$ and $v \leq a$

\[
\frac{p_S(v)}{p_{S'}(v)} \leq \frac{f(v \mid a, \frac{b}{m})}{\exp \left( -\frac{a-v}{b/\varepsilon m} \right)} = \exp \left( -\frac{a-v}{b/\varepsilon m} \right) \leq \exp(\varepsilon)
\]

Last inequality follows from dropping the numerator (which is at most 1) and taking $v = a$.

Required inequality for event $Y \subseteq [0, b]$ follows by integrating over all $v \in Y$.

□

**Proposition 3.** With prob $1 - \delta$, Laplace mechanism adds an error of magnitude at most $\frac{b}{\varepsilon m} \ln \frac{1}{\delta}$.

**Proof.** For $\tau \geq 0$

\[
P[x \geq \tau] = \varepsilon m \int_{\tau}^{\infty} e^{-x \varepsilon m/b} dx = \frac{1}{2} e^{-\tau m/b}
\]

So $P[x \geq \tau] = \delta/2$ when $\tau = \frac{b}{\varepsilon m} \ln \frac{1}{\delta}$

Identical analysis works for $P[x \leq -\tau] = \delta/2$.

□

**Generalization:** To compute some real-valued function (e.g. statistics) $g$ of dataset $S$.

Let $\Delta g = $ maximum change to $g$’s output when just one data point changes.

(General) Laplace mechanism outputs $v = \text{Lap}(g(S), \Delta g/\varepsilon)$

This mechanism satisfies $\varepsilon$-differential privacy, by the same proof.

**Composition:** Suppose independent mechanisms $A_1, \ldots, A_k$ answer $k$ queries.

Each satisfying $\varepsilon$-differential privacy.

Then the vector of $k$ responses $A = (A_1, \ldots, A_k)$ satisfies $k\varepsilon$-differential privacy, since

\[
P[A(S') = y] = P[A_1(S') = y_1] \cdots P[A_k(S') = y_k] \leq e^{\varepsilon} P[A_1(S) = y_1] \cdots e^{\varepsilon} P[A_k(S) = y_k] = e^{k\varepsilon} P[A(S) = y]
\]

The other inequality is analogous.