# Differential Privacy

Berkeley Math Circle, Oct 16, 2019

## 1 Defining Privacy

In differential privacy, we measure privacy by quantifying information leakage.

**Definition 1 (Differential Privacy [Dwo06, DR$^+$14])** *A function $f$ is $\epsilon$-DP (**d**ifferentially-**p**rivate) if, for all neighboring databases $X$ and $X'$ (which are databases that differ in one row; we denote this by $X \sim X'$) and all subsets $S \subseteq \texttt{im}\ f$,*

$$\Pr(f(X) \in S) \leq e^{\epsilon} \Pr(f(X') \in S)$$

### 1.1 Properties: The Good and The Bad

**Good Property 1** Immunity to post-processing. If I take the output of a DP function, and do some additional processing, I won't learn more.

If a function $f$ is $\epsilon$-DP, and $g$ is any function, then $g \circ f$ is also $\epsilon$-DP.

**Good Property 2** Composition.

If functions $f$ and $g$ are $\epsilon$-DP, then $f \circ g$ is $2\epsilon$-DP.

**Bad Property 1** DP only measures a loose upper bound on information leakage.

If a function $f$ is $\epsilon$-DP, then $f$ is also $\epsilon'$-DP, for any $\epsilon' \geq \epsilon$.

**Bad Property 2** Any (non-trivial) DP function must be randomized.

A non-constant deterministic function $f$ is not $\epsilon$-DP for any $\epsilon$.

## 2 Mechanisms

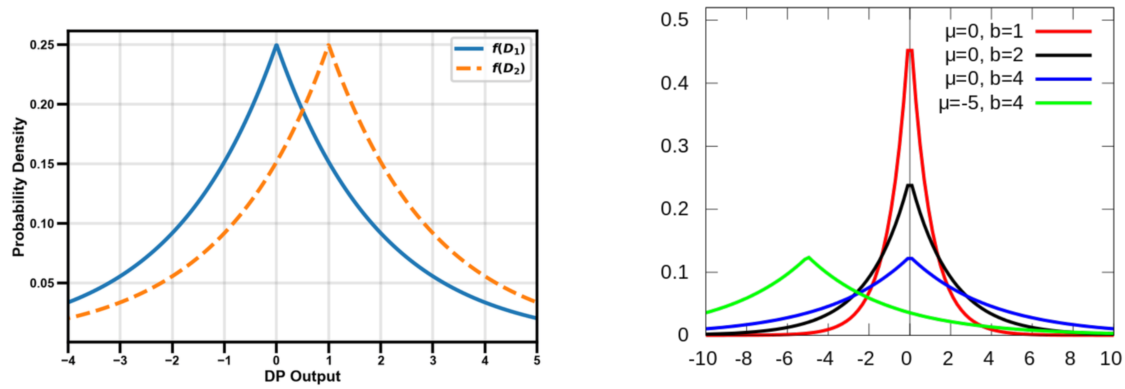
A "mechanism" is method of making database queries differentially private.

### 2.1 Laplace Mechanism

The Laplace mechanism makes queries private by adding random noise sampled from Laplace distribution. The Laplace distribution $\texttt{Lap}\ (\mu, b)$ $\mu$ = mean, $b$ = scale) has the probability density function

$$pdf(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

**Theorem 1 (Laplace Mechanism)** *Let $f$ be a (deterministic) function, and $\Delta f = \max_{X \sim X'} |f(X) - f(X')|$. Then, the function $M(X) = f(X) + \texttt{Lap}\ (0, \Delta f/\epsilon)$ is $\epsilon$-DP.*




(a) (b)

Figure 1: Left: Example of output distribution of Laplace mechanism on neighboring databases. Right: Laplace distribution for different scales $b$ (and mean $\mu$).

### 2.2 Other Mechanisms

Many other mechanisms have been developed over the years, e.g. Gaussian mechanism (use Gaussian distribution instead of Laplace), Exponential mechanism (considers *utility/accuracy* of the output), and Sparse Vector Technique (only output sums over a certain threshold).

## 3 Randomized Response and Local Differential Privacy

*"I have used, at least once, the resources of my institution for the benefit of a political party."*
— Survey to measure corruption in Bolivia, Brazil, and Chile, 2010 [BIZ15].

People often don't want to answer sensitive questions truthfully. However, it is difficult to quantify the bias in the answers.

Randomized response was first proposed by Warner in 1965 [War65] to address this problem. The idea is that, if people have *deniability* for their answers, then they are more likely to answer truthfully. Warner's method has been used in a survey to measure corruption, an ecological study, and a survey regarding people's sentiment about capital punishment [BIZ15].

### 3.1 Local Differential Privacy (LDP)

It turns out, the Warner's randomized response satisfies a strong notion called *local differential privacy (LDP)*. Informally, LDP is differential privacy against even the data curator. For example, when you send your data to Apple on your phone, you are protecting the data against Apple with a LDP mechanism.

**Definition 2 (Local Differential Privacy (LDP))** *A randomized function $f$ is $\epsilon$-LDP if, for all private data $x$ and $x'$ and all subsets $S \subseteq \texttt{im}\ f$,*

$$\Pr(f(x) \in S) \leq e^{\epsilon} \Pr(f(x') \in S)$$

## 3.2 Activity: Randomized Response

I want to know how popular ________________ is among teenagers, that is,
*How many people spend two hours or more on ________________, on an average day?*.

### 3.2.1 Instructions

1. Flip two coins. **Don't let anyone else know the result of the coin tosses!**
2. If *at least one* of the coins is *tails*, answer "True" or "False" to the following statement:

   - *A. On an average day, I spend two hours or more on ________________.*

3. If *both* coins are *heads*, answer "True" or "False" to the following statement:

   - *B. On an average day, I spend less than two hours on ________________.*

4. Let me know if you responded True or False, but not which statement you were answering.

### 3.2.2 Questions

Let $f(x)$ be the response to the survey, given that your answer to statement A is $x$.

1. If you didn't want to let people know about your secret obsession, how does this survey protect your privacy (even if you answered "True" in the survey)?
2. What is $\texttt{im}\ f$, the possible responses to the survey? List all subsets $S \subseteq \texttt{im}\ f$.
3. What is the probability $\Pr(f(\text{True}) = \text{True})$? How about $\Pr(f(\text{False}) = \text{True})$?
4. What is the smallest $\epsilon$ for which the survey is $\epsilon$-LDP?

## 3.3 LDP of Randomized Response

Let's say the survey tells us to answer statement A with probability $p = \Pr(f(x) = x)$, and statement B with probability $(1 - p)$, for some $p > 1/2$. What is the LDP of the survey?

**Claim 1** *The survey is* $\ln\left(\frac{p}{1-p}\right)$*-LDP.*

**Proof**
The set $\texttt{im}\ f = \{\text{True}, \text{False}\}$, so the possible $S \subseteq \texttt{im}\ f$ are: $\emptyset, \{\text{True}\}, \{\text{False}\}, \{\text{True}, \text{False}\}$.

- For $S = \emptyset$ and $S = \{\text{True}, \text{False}\}$, the inequality in LDP definition holds trivially (Why?)
- For $S = \{\text{True}\}$, the inequality becomes

  $$\Pr(f(x) = \text{True}) \leq e^{\epsilon} \Pr(f(x') = \text{True})$$

  Let's say $x = \text{True}$, $x' = \text{False}$. This inequality becomes

  $$\begin{aligned} \Pr(f(x) = x) &\leq e^{\epsilon} \Pr(f(x') = x) \\ p &\leq e^{\epsilon}(1 - p) \end{aligned}$$

Solving for $\epsilon$ gives us $\epsilon = \ln\left(\frac{p}{1-p}\right)$.

If instead $x = \text{False}$, $x' = \text{True}$, then the inequality becomes

$$\begin{aligned}\Pr(f(x) = x') &\leq e^{\epsilon}\Pr(f(x') = x') \\ (1-p) &\leq e^{\epsilon} p\end{aligned}$$

Solving for $\epsilon$ gives us $\epsilon = \ln\left(\frac{1-p}{p}\right)$. But since $p > 1/2$, $\ln\left(\frac{1-p}{p}\right) < \ln\left(\frac{p}{1-p}\right)$, so $f$ only satisfies $\epsilon = \ln\left(\frac{p}{1-p}\right)$-DP

- The proof for $S = \{\text{False}\}$ is very similar (Check!)

## 4 Further Reads

Data privacy horror stories
`https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/`
`http://techland.time.com/2012/02/17/how-target-knew-a-high-school-girl-was-pregnant-before-her-parents/`

Googles open source DP library, and implementation of private user data collection in Chrome
`https://github.com/google/differential-privacy`
`https://github.com/google/rappor`

Googles differentially private version of TensorFlow (training machine learning models)
`https://github.com/tensorflow/privacy`

Uber's open source real world differentially-private SQL queries
`https://medium.com/uber-security-privacy/differential-privacy-open-source-7892c82c42b6`
`https://github.com/uber/sql-differential-privacy`

Report of Apple's differential privacy settings
`https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf`

## References


[BIZ15] Graeme Blair, Kosuke Imai, and Yang-Yang Zhou. Design and analysis of the randomized response technique. *Journal of the American Statistical Association*, 110(511):1304–1319, 2015.

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[Dwo06] Cynthia Dwork. Differential privacy (invited paper). In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *ICALP 2006, Part II*, volume 4052 of *LNCS*, pages 1–12. Springer, July 2006.

[War65] Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.