CS 761: Randomized Algorithms Fall 2019

Lecture 12 — November 29, 2019

Prof. Gautam Kamath *By: Joakim Blikstad, Ethan Ward*
Edited by Vedat Levi Alev

Disclaimer: These notes have not been subject to the usual scrutiny reserved for formal publications.

1 Data Privacy

1.1 (Unsuccessful) Ad Hoc Methods

1. Netflix Challenge (2006–2009)

   Researchers were challenged to use machine learning techniques on Netflix's dataset to recommend new movies that users would enjoy.

   Netflix tried to anonymize their data by only storing:

   - Anonymised user id
   - What movies they rated
   - How high they rated it
   - When they rated it

   Paper: Narayanan-Shmatikov (2006) [6]: The authors found that they were able to identify individuals in the dataset. Their main idea was a cross-reference attack, where they looked at reviews on IMDB and cross-referenced the dates with when they watched the movie on Netflix. Since some IMDB user profiles had names and other identifying information, they could link people with watching potentially politically or socially sensitive movies. This stopped the 2nd Netflix Challenge from happening because of privacy concerns.

2. GWAS (*Genome Wide Association Study*)
   In these studies, researchers released supposedly anonymized genome data, with each data point having an individual's genetic markers and what diseases they got, in the hope of finding genetically-based diseases.

   In 2008, Homer et al found they were able to reidentify individuals, or tell if a specific person participated in a study. They were able to do this in a synthetic environment using χ^2 statistics [5]. Based on this finding, the American National Institutes of Health (NIH) stopped releasing certain classes of statistics to the public.

3. Massachusetts Group Insurance Commission (1997). The Massachusetts GIC had a database where each row contained

   name, SSN, zip, date of birth, sex, diagnosis, procedure

This was released to the public, but without the *name* and *SSN* fields, to hopefully anonymize the data.

To try to deanonymize the data, Latanya Sweeney paid $20 to get a list of

name, party affiliation, zip, date of birth, sex

for people in Cambridge, MA from a voter database. She matched this information to the above released data.

In order to match the voter information to the MGIC data, she found that, for example:
Narrowing to same date of birth $\rightarrow$ 6 people
Adding sex:male $\rightarrow$ 3 people
Adding same zip $\rightarrow$ 1 person

Using this, Sweeney found a governor's medical records and mailed them to him to prove the danger of the attack. In a further study, she found that someone's ZIP code, date of birth, and sex was enough to uniquely identify 87% of people in the US. [9]

1.2 k-anonymity (Samarati-Sweeney [8])

Definition 1. *k-anonymity $\iff$ each row in a database is identical to $\geq (k-1)$ other rows.*

The following is an example of making a database 2-anonymous (with respect to the "pseudo-identifiers," which are the columns besides Disease) using suppression (generalizing data in rows):

Original database:

First	Last	Age	Race	Disease
Harry	Stone	34	Af.Am	flu
John	Reyser	36	Cauc.	mumps
Beatrice	Stone	47	Af.Am.	mumps
John	Ramos	22	Hisp.	mumps

Suppressed database:

Age	Race	Disease
30-50	Af.Am	flu
20-40	*	mumps
30-50	Af.Am.	mumps
20-40	*	mumps

One major issue with this approach is that it takes polynomial time using suppression techniques to make a database k-anonymous for $k = 2$, but it is NP-hard for $k \geq 3$. Using only *s to suppress data, there exists $O(\log k)$-approximation algorithms in poly-time.

Another issue is that this approach still suffers from similar attacks as before. Say we know *John Ramos, 22* is in the above database, we can identify that he has mumps even in the suppressed version.

1.3 Dinur-Nissim attack [2]

Consider the following setting:

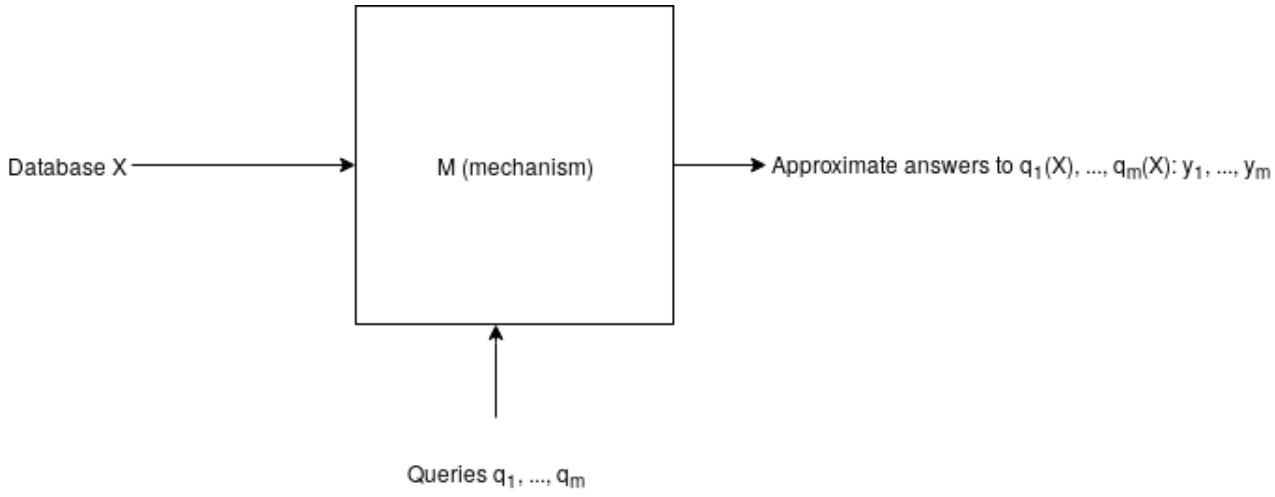
A trusted curator holds a database $X \in \mathcal{U}^n$ where the universe $\mathcal{U} = \{0,1\}^d$. Intuitively, n is the number of individuals, and the database stores d bits of data for each individual.

Queries $q_1, \ldots, q_m$, are questions about the dataset asked to the (trusted) curator. We want any query result the curator gives to be sufficiently private. We start by focusing on counting queries, or queries that ask how many people in the database satisfy some property.

Counting Queries:

$$q : \mathcal{U} \to \{0,1\}$$
$$q(X) = \frac{1}{n}\sum_{i=1}^{n} q(x_i)$$

We note that these can be too specific, as any answer that returns exactly $\frac{n-1}{n}$ means only one person does not satisfy a specific property. We then use the idea of a mechanism (or algorithm) M that transforms the answers to queries into some approximation according to the following diagram:



For the sake of making examples easier, we focus on $d = 1$ (one bit stored per person).

One example of this kind of query is a correlation query:

$$q_i(x) = \frac{1}{n}\sum_{j=1}^{n} q_{ij} X_j \quad q_i \in \{0,1\}^n, q_{ij} \in \{0,1\}$$

Definition 2. *Normalized Hamming distance between X and X' is $d_H(X, X') = \frac{1}{n}|\{i : x_i \neq x'_i\}|$.*

We can note that some kinds of mechanisms are not sufficiently private, and in fact we can compute an entire database that is relatively close to the original database, something we call blatantly non-private.

Theorem 3 (Dinur-Nissim [2])**.** *Let $\alpha > 0$. Suppose that for all $q \in \{0,1\}^n$, the algorithm outputs $y_q \in \mathbb{R}$ with $|y_q - q(x)| \leq \alpha$. Then we can compute a database $X' \in \{0,1\}^n$ such that $d_H(X, X') \leq 4\alpha$.*

Proof. Output any X' which is *consistent* (i.e. $|y_q - q(X')| \leq \alpha$) with all queries, after asking every query $q \in \{0,1\}^n$. At least one such X' exists, since X is *consistent* with every query. Consider the two queries $q_1 = X$ and $q_0 = 1 - X = \overline{X}$ (queries that should count everything in the database respectively nothing). We compute the *Normalized Hamming* distance:

$$\begin{aligned} d_H(X, X') &= \frac{1}{n} |\{i : x_i = 1, x'_i = 0\}| + \frac{1}{n} |\{i : x_i = 0, x'_i = 1\}| \\ &= (q_1(X) - q_1(X')) + (q_0(X') - q_0(X)) \\ &\leq (y_{q_1} + \alpha) - (y_{q_1} - \alpha) + (y_{q_0} + \alpha) - (y_{q_0} - \alpha) \\ &= 4\alpha \end{aligned}$$

□

We note that there are some issues with this approach. In the above proof, we use 2^n queries, which is both computationally inefficient and easy to detect by the trusted curator. However, we can show that only $O(n)$ queries are needed for a different approximation.

Lemma 4 (Paley-Zygmund Inequality[1])**.** *Suppose X is a non-negative random variable. Then* $\Pr[X \geq \lambda \mathrm{E}[X]] \geq (1-\lambda)^2 \frac{\mathrm{E}[X]^2}{\mathrm{E}[X^2]}$.

Proof.

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}\big[X \cdot \mathbf{1}_{\{x \leq \lambda \mathrm{E}[X]\}}\big] + \mathrm{E}\big[X \cdot \mathbf{1}_{\{x \geq \lambda \mathrm{E}[X]\}}\big] \\ &\leq \lambda \mathrm{E}[X] + \sqrt{\mathrm{E}[X^2] \cdot \mathrm{E}\big[\mathbf{1}_{\{x \geq \lambda \mathrm{E}[X]\}}\big]} && \text{(by Cauchy-Schwartz)} \\ &= \lambda \mathrm{E}[X] + \sqrt{\mathrm{E}[X^2] \cdot \Pr[X \geq \lambda \mathrm{E}[X]]} \\ \implies \quad \Pr[X \geq \lambda \mathrm{E}[X]] &\geq (1-\lambda)^2 \frac{\mathrm{E}[X]^2}{\mathrm{E}[X^2]} && \text{(by rearranging)} \end{aligned}$$

□

Theorem 5. *There exists a set of $m = O(n)$ queries $q_1, \ldots, q_m$ such that, given $y_1, \ldots, y_m$ with $|y_i - q_i(X)| \leq \frac{\alpha}{\sqrt{n}}$ for all i, we can compute a database X' such that $d_H(X, X') = O(\alpha^2)$. The algorithm works with probability $\geq \frac{1}{2}$.*

Proof. Choose queries $q_1, \ldots q_m$ uniformly at random in $\{0,1\}^n$. Similar to before, output any X'

[1]Compare to Markov's Inequality which shows $\Pr[X \geq \lambda \mathrm{E}[X]] \leq \frac{1}{\lambda}$.

such that $|y_i - q_i(X')| \leq \frac{\alpha}{\sqrt{n}}$. Consider any Y with $d_H(X,Y) = \sum_{j=1}^{n}(x_j - y_j)^2 \geq c\alpha^2$. We have:

$$\begin{aligned}
q_i(X) - q_i(Y) &= \frac{1}{n}\sum_{j=1}^{n} q_{ij}(x_j - y_j) \\
\mathrm{Var}[q_i(X) - q_i(Y)] &= \mathrm{Var}\left[\frac{1}{n}\sum_{j=1}^{n} q_{ij}(x_j - y_j)\right] \\
&= \frac{1}{4n^2}\sum_{j=1}^{n}(x_j - y_j)^2 && \text{(since } q_{ij} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\tfrac{1}{2})\text{)} \\
&\geq \frac{c\alpha^2}{4n^2} && \text{(by choice of } Y\text{)}
\end{aligned}$$

With substantial computations (omitted), Paley-Zygmund allows one to derive that

$$\begin{aligned}
&\Pr\left[\forall i,\ |q_i(X) - q_i(Y)| \leq \tfrac{2\alpha}{\sqrt{n}}\right] \leq 2^{-2n-1} \\
\implies \quad &\Pr\left[\exists X, Y, \forall i,\ |q_i(X) - q_i(Y)| \leq \tfrac{2\alpha}{\sqrt{n}}\right] \leq \frac{1}{2} && \text{(by union bound)}
\end{aligned}$$

Using the triangle inequality gives:

$$\begin{aligned}
&|y_i - q_i(X)| \leq \frac{\alpha}{\sqrt{n}} \quad \text{and} \quad |y_i - q_i(Y)| \leq \frac{\alpha}{\sqrt{n}} \\
\implies \quad &|q_i(X) - q_i(Y)| \leq \frac{2\alpha}{\sqrt{n}} \\
\implies \quad &\Pr\left[\exists X, Y \text{ both consistent with all queries} : d_H(X,Y) \geq c\alpha^2\right] \leq \frac{1}{2}
\end{aligned}$$

Hence the algorithm will with probability $\geq \frac{1}{2}$ return X' with the desired guarantee that $d_H(X, X') = O(\alpha^2)$. □

We note that there was a recent use of this form of attack in practice by Cohen and Nissim [1].

1.4 Differential Privacy

1.4.1 Randomized Response (Warner 1965 [10])

Setting: We have n people, each with a sensitive bit $b_i \in \{0, 1\}$. We want to calculate $Z = \frac{1}{n}\sum_{i=1}^{n} b_i$ while still keeping the information private.

Take 1

Ask everyone to tell b_i truthfully, $Z = \frac{1}{n}\sum_{i=1}^{n} b_i$. This is perfectly accurate, but not private at all.

Take 2

Everyone reports $X_i = \begin{cases} b_i & \text{with probability } \frac{1}{2} \\ 1 - b_i & \text{with probability } \frac{1}{2} \end{cases}$. Compute $Z' = \frac{1}{n}\sum_{i=1}^{n} X_i$. This is perfectly private, but not accurate at all or even useful, as it is effectively just generating random bits for each person.

Take 3 (middle ground)

Everyone reports $X_i = \begin{cases} b_i & \text{with probability } \frac{1+\varepsilon}{2} \\ 1 - b_i & \text{with probability } \frac{1-\varepsilon}{2} \end{cases}$.

This idea is due to Warner [10]. We can note that $\varepsilon = 1$ corresponds to Take 1, and $\varepsilon = 0$ corresponds to Take 2. A good middle ground is something like $\varepsilon = \frac{1}{2}$, giving us a 3/4 chance of telling the truth and 1/4 change of lying. As a means of measuring privacy, we calculate the following:

$$\frac{\Pr[b_i = 1 \mid X_i = 1]}{\Pr[b_i = 1 \mid X_i = 0]} = \frac{(\frac{1+\varepsilon}{2})}{(\frac{1-\varepsilon}{2})} \approx 1 + 2\varepsilon$$

Note that when ε is small, this means it is very private (roughly the same chance of knowing someone's true bit and it being changed). We can then calculate the following for accuracy:

$$\begin{aligned} \mathrm{E}[X_i] &= b_i \tfrac{1+\varepsilon}{2} + (1 - b_i)\tfrac{1-\varepsilon}{2} = \varepsilon b_i + \tfrac{1-\varepsilon}{2} \\ Z' &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\varepsilon}(X_i - \tfrac{1-\varepsilon}{2}) \\ \mathrm{E}\big[Z'\big] &= \frac{1}{n}\sum_{i=1}^{n} b_i \\ \mathrm{Var}\big[Z'\big] &= \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] \leq O(\tfrac{1}{n\varepsilon^2}) \end{aligned}$$

So by Chebyshev's inequality we get

$$\text{Error } |Z - Z'| \approx O(\mathrm{SD}(Z')) = O(\tfrac{1}{\varepsilon\sqrt{n}})$$

This gives us accuracy α when $n \geq \Omega(\frac{1}{\alpha^2\varepsilon})$.

1.4.2 Central Differential Privacy (Dwork-McSherry-Nissim-Smith TCC 2006 [3])

Definition 6. *A randomized algorithm M is (ε, δ)-Differentially Private (DP) if, for all $S \subseteq \mathrm{Range}(M)$, and all x, y that differ in ≤ 1 entry, $\Pr[M(x) \in S] \leq e^{\varepsilon}\Pr[M(y) \in S] + \delta$.*

This guarantees good privacy if ε is a small constant (say $\varepsilon = 1$ is reasonable, $\varepsilon = 0.1$ is great), and δ "cryptographically small". When $\delta = 0$, we call it *pure* DP, and when $\delta > 0$ *approximate* DP.

Note that here x and y can be thought of as neighboring databases, or databases that only differ in one entry. Putting this into the context of previous attacks, it means that even if someone has sidechannel information about you, you can be safe adding your information to the database, as it will not change the answer to queries much. Note also that this kind of differential privacy requires a trusted curator. This makes it central differential privacy, as opposed to local differential privacy, which does not need a trusted participant [4] [7] [10].

1.4.3 Laplace Mechanism

Consider some random variable Y which is sampled from the *Laplace distribution* $\text{Lap}(b)$. Then Y has probability density function $P_Y(y) = \frac{1}{2b}\exp\Big(1 - \frac{|y|}{b}\Big)$. We also note that $\text{Var}[Y] = 2b^2$, and $\text{SD}(Y) = \sqrt{2} \cdot b$.

Definition 7. *Given $f : X \to \mathbb{R}$, the sensitivity Δ is defined as*

$$\Delta = \max_{\substack{x,y \in X \\ x,y \text{ are} \\ \text{“neighbors”}}} |f(x) - f(y)|$$

Note that in the above definition, the term “neighbors” is intentionally vague, as what we want differs substantially from setting to setting.

Theorem 8 (Laplace Mechanism)**.** *$f(x) + \text{Lap}(\frac{\Delta}{\varepsilon})$ is ε-Differentially Private, where Δ is the sensitivity of f.*

Proof. Suppose $f : X \to \mathbb{R}$, and $x, y \in X$ are neighbors (i.e. they differ in at most one entry). Define P_x to be the probability density function of $f(x) + \text{Lap}(\frac{\Delta}{\varepsilon})$ and P_y similarly. Now evaluate at any point $z \in \mathbb{R}$:

$$\begin{aligned}
\frac{P_x(z)}{P_y(z)} &= \frac{\exp\Big(-\frac{\varepsilon|f(x)-z|}{\Delta}\Big)}{\exp\Big(-\frac{\varepsilon|f(y)-z|}{\Delta}\Big)} \\
&= \exp\Big(\frac{\varepsilon}{\Delta}(|f(y)-z| - |f(x)-z|)\Big) \\
&\leq \exp\Big(\frac{\varepsilon}{\Delta}|f(y)-f(x)|\Big) \\
&\leq \exp(\varepsilon)
\end{aligned}$$

This shows that the mechanism $M(x) = f(x) + \text{Lap}(\frac{\Delta}{x})$ is ε-DP. We also note that this is a pure-DP algorithm, since $\delta = 0$. □

1.4.4 Average of bits with Laplace Mechanism

We now revisit the average of bits problem, but use the Laplace Mechanism to construct a centrally differentially private algorithm.

Consider $f(x) = \frac{1}{n}\sum_{i=1}^{n} b_i$, and that the sensitivity $\Delta \leq \frac{1}{n}$. Then

$$Z'' = \frac{1}{n}\sum_{i=1}^{n} b_i + \mathrm{Lap}\left(\frac{1}{n\varepsilon}\right) \qquad \mathrm{E}\big[Z''\big] = \frac{1}{n}\sum_{i=1}^{n} b_i$$

$$\mathrm{Var}\big[Z''\big] = \frac{2}{\varepsilon^2 n^2} \qquad \mathrm{SD}[Z''] = \frac{\sqrt{2}}{\varepsilon n}$$

This means error $\approx O(\frac{1}{\varepsilon n})$ and accuracy α requires $n \geq \Omega(\frac{1}{\varepsilon\alpha})$. This is quadratically better than the randomized response algorithm, but requires a trusted curator.

Nice properties of Differential Privacy

- Post processing.
  If M is (ε, δ)-DP, and f a function, then $f \circ M$ is (ε, δ)-DP too.
- Group privacy.
  If M is (ε, δ)-DP, then for all x, x' which differ in at most k entries, $\Pr[M(x) \in S] \leq e^{k\varepsilon}\Pr[M(x') \in S] + ke^{k\varepsilon}\delta$. This can be thought of as the generalization of the theorem above, where we are interested in the difference when multiple people change their entry in a database, do not include it in a database, etc.
- Basic composition.
  If $M(x) = (M_1(x), \ldots, M_k(x))$ where each M_i is(ε, δ)-DP, then M is $(k\varepsilon, k\delta)$-DP.
- Advanced composition.
  In the same setting as above, for all $\delta' > 0$, M is $(\sqrt{k\log(1/\delta')}\cdot\varepsilon, k\delta + \delta')$-DP.

References

[1] Aloni Cohen and Kobbi Nissim. Linear program reconstruction in practice. *CoRR*, abs/1810.05692, 2018.

[2] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210. ACM, 2003.

[3] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[4] Alexandre Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, pages 211–222, New York, NY, USA, 2003. ACM.

[5] Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V. Pearson, Dietrich A. Stephan, Stanley F. Nelson, David W. Craig, and et al. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS Genetics*, 4(8), 2008.

[6] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the netflix prize dataset. *CoRR*, abs/cs/0610105, 2006.

[7] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *arXiv preprint arXiv:0803.0924*, 2008.

[8] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI International, 1998.

[9] Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, 671:1–34, 2000.

[10] Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. PMID: 12261830.