An Introduction to Differential Privacy

2019 Joint Statistical Meetings

July 30, 2019

Instructors

- Robert Ashmead, PhD, Ohio Colleges of Medicine Government Resource Center
- Philip Leclerc, PhD, U.S. Census Bureau
- William Sexton, MS, U.S. Census Bureau

Email

- robert.ashmead@osumc.edu
- philip.leclerc@census.gov
- william.n.sexton@census.gov

Schedule

- 8:00 - 10:15 Introduction, Definitions and Basic Principles, Semantics
- 10:15 - 10:30 Mid-Morning Break
- 10:30 - 12:00 Analyzing Data with Differential Privacy, Extensions

Disclaimer
Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not those of the U.S. Census Bureau.

Learning Goals
At the conclusion of the class participants will be able to...

1. Describe why the release of de-identified data still creates a privacy risk.
2. Understand the difference between a trusted and untrusted curator model.
3. Understand the motivation and definition of differential privacy.
4. Explain common interpretations of the differential privacy gurantee.
5. Understand the definition and how to calculate query sensitivity.
6. Apply simple (Laplace, geometric) differentially private mechanisms to data examples (using R code).
7. Understand the rules for post-processing and composition of differentially private algorithms.
8. Understand how to adjust analysis to account for differentially private mechanisms.

1 Introduction

1.1 What is privacy and how do we measure it?

- Means different thing to different people
  - "Privacy is the ability of an individual or group to seclude themselves, or information about themselves, and thereby express themselves selectively." [1]
  - Concealment of information
  - Peace and quiet
  - Freedom
  - Learning something about you that you don't think someone should be able to learn
  - Others Suggestions?
- Most definitions are abstract and difficult to measure
- Try to think of privacy in a non-binary way

1.2 The Privacy-Accuracy Tradeoff

There is an inherent trade-off between privacy and data utility/accuracy discussed by many economists and statisticians [1, 2]. Broadly, the more information that is released (accurately) and with greater granularity, the better the data utility. However, at the same time this increases the risk of an attacker violating a person's privacy.

Think about releasing a dataset with individuals from some population. The more variables that are included in that dataset, the greater the ability to find relationships between variables and learn about the population. However, increasing the number of variables in the dataset also makes the individuals in the dataset more unique, and therefore easier to identify given external data.

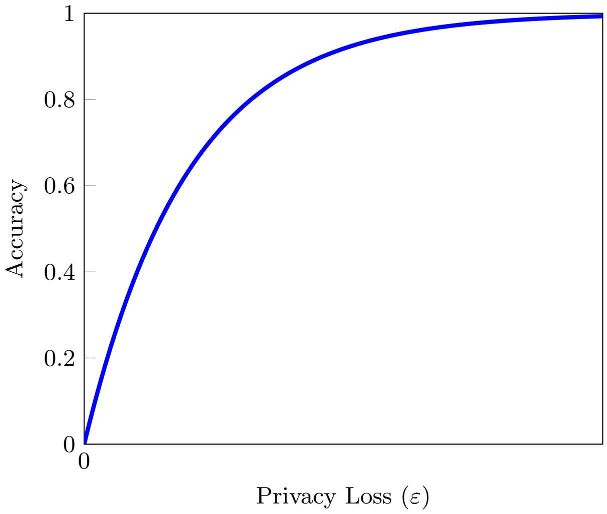

Figure 1: No data are used in this figure. An Example of the trade-off between privacy loss and accuracy in data publication[2]

[1] https://en.wikipedia.org/wiki/Privacy

[2] Source: Why the Economics Profession Must Actively Participate in the Privacy Protection Debate John M. Abowd, Ian M. Schmutte, William N. Sexton, and Lars Vilhuber AEA Papers and Proceedings. May 2019, Vol. 109, No. : Pages 397-402

1.3 Learning about the Population versus the Individual

An important goal of privacy methods in general should be to preserve the ability to learn about the population (general scientific knowledge) while limiting what can be learned about an individual (in violation of their privacy). For example, privacy methods should not deter the ability of a study to find a link between health behavior A and disease B. At the same time, an attacker should not be able to use the study publication to learn whether their neighbor participated in the study and their disease status.

There is also concept in formal privacy called *group privacy* which is concerned with protecting the privacy of all members of a group (e.g. family). So the distinction between the population and the individual can be blurred as the size of the group increases.

1.4 Concrete Attack Examples

There are several well-known examples of attacks on databases that are used to illustrate the vulnerability of de-identified databases. These include:

- The Netflix prize: Attackers were able to use user's public IMDb (Internet Movie Database) ratings to identify the user's in the Netflix database. [3]
- AOL: AOL released search data along with unique identifiers to link sets of searches together. Individuals could be identified from there searches. [4]
- Massachusetts Health Insurance: The governor of Massachusetts was identified in a de-identified health care dataset using voter list information. [5]
- 2010 Decennial Census: Internal attackers were able to reconstruct some records using commercial data combined with publicly released tables [6]

A general theme of most of these attacks is the ability to access outside information on specific persons and combine it with the database in order to learn additional information about those individuals.

1.5 Existing SDL Methods

Examples of popular statistical disclosure control methods are:

- Swapping: pairs of "similar" records are randomly paired and swapped. Probabilities, details of similarity calculation, etc are secret
- Top-coding: scalar observations (e.g. personal income) exceeding some threshold are perturbed or suppressed
- Cell suppression: in tabular summaries, a set of "primary suppression" cells is heuristically identified as sensitive and suppressed. A "secondary suppression" set of cells with known relationships to the primary suppressions are then further suppressed. The goal is to minimize severity of suppression while stopping an attacker who only has access to the published tables from precisely inferring values of the primary suppressions
- Synthetic data: a very broad class of techniques with few rules about the algorithms used, except that they all generate microdata. "Partially" synthetic methods may directly publish unmodified portions of initial input records, while "fully" synthetic approaches instead make statistical inferences to generate every value in every record
- k-anonymity: a relatively recent family of techniques primarily targeted at microdata releases. Their defining property is that the frequency of each type of published records must be at least $k-1$, so that the rarity or uniqueness of records is parametrically controlled
- ad hoc noise infusion: a heuristic label for a very broad class of techniques with no hard rules about their operation. Often ad hoc noise infusion involves additive or multiplicative infusion of noise drawn from a simple closed-form distribution into the attributes of

microdata records or aggregates/tabulations. The primary property distinguishing these techniques is negative: they do not come with any form of rigorous guarantee about the protection offered (even against narrow classes of attackers, with some exceptions)

- Controlled tabular adjustment: a close cousin of cell suppression, in which cells flagged as sensitive in published tables are protected by systematically (often determinstically) perturbing them, then adjusting the results in a manner controlled to ensure the table's interior cells still sum to its marginals. The primary distinction between controlled tabular adjustment and cell suppression is that tables published under controlled tabular adjustment are "complete" (no values missing due to suppression)

At the end of this section, we provide some general references for popular statistical disclosure control methods [7, 8]. An important difference between these traditional disclosure-avoidance methods and differential privacy is that these methods are generally aimed at protecting against specific classes of attackers, and often do not come with formal proofs of the protection they offer (with notable exceptions, such as cell suppression's formal guarantee against a narrow class of attackers), whereas differential privacy guards against very general classes of attackers, and this fact can be formally proven.

1.6 Data Curator Models

The *data curator* is the organization that collects and retains the data to be protected, often in a formal database; Google, the U.S. Census Bureau, Gallup, etc are all data curators. We consider two types of data curator models:

- The Trusted Curator Model (sometimes called the "central model")
  - Assumes an organization with the access to the collected data who is responsible for releasing for private query answers
  - Statistically efficient
  - Microdata difficult to generate
  - Interactive or non-interactive model
  - Examples: OnTheMap [9], Post Secondary Employment Outcomes [10], 2020 Decennial Census [11]

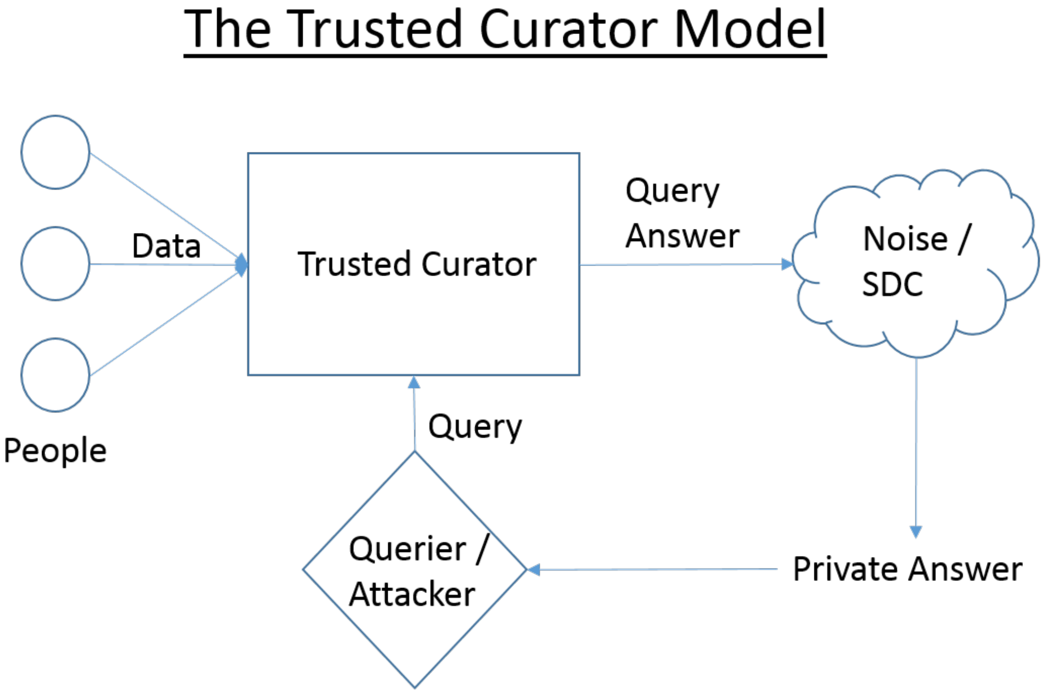

Figure 2: The Trusted Curator Model

- The Untrusted Curator Model (sometimes called the "local" model")
  - Guards against curator misuse
  - Simple and naturally generates microdata
  - Natural example is the randomized response technique
  - Examples: Google RAPPOR [12], Apple [13]
  - Noise scales with the number of records

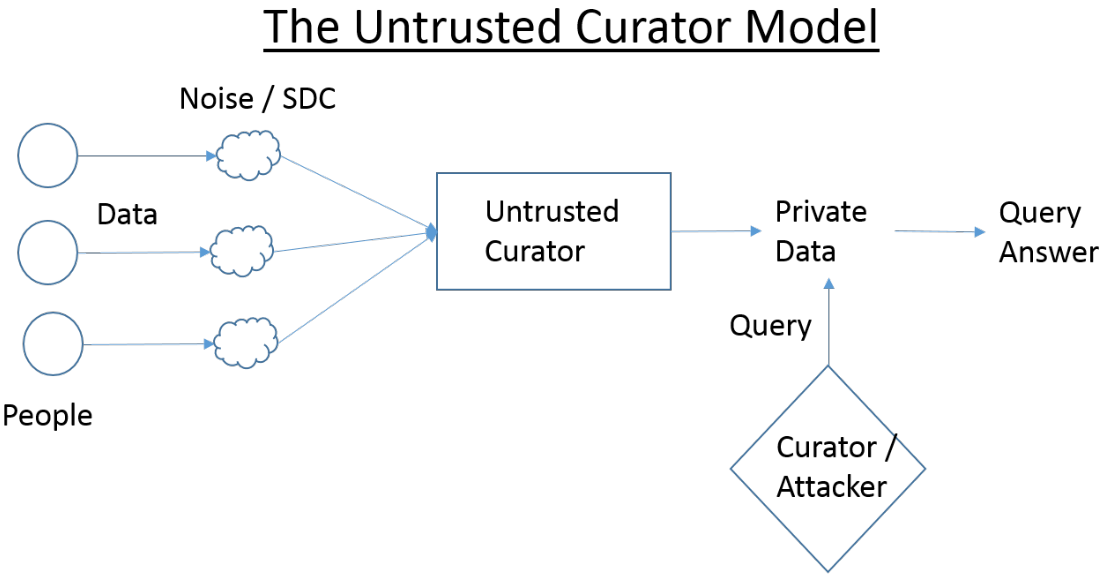

Figure 3: The Untrusted Curator Model

1.7 Differential Privacy History & Properties

- Differential privacy is not any single method, but a growing class of methods that satisfy a common definition.
- It is one of a (growing) body of formal privacy definitions. Here, we use "formal" to mean that these methods come equipped with a quantifiable measure of privacy-loss and provable guarantees against general classes of attackers.
- Differential Privay was originally defined and formalized in the 2006 paper of Dwork, McSherry, Nissim, and Smith [3].
- Dinur and Nissim published an important motivating paper [15] showing that if enough random queries are published from a database, then the whole database is revealed with high probability.

1.8 Questions?

References

[1] John M Abowd and Ian M Schmutte. An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *American Economic Review*, 109(1):171–202, 2019.

[2] Stephen E Fienberg, Alessandro Rinaldo, and Xiaolin Yang. Differential Privacy and the Risk-Utility Tradeoff for Multi-Dimensional Contingency Tables. In *International Conference on Privacy in Statistical Databases*, pages 187–199. Springer, 2010.

[3] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large datasets (how to break anonymity of the netflix prize dataset). *University of Texas at Austin*, 2008. `https://arxiv.org/pdf/cs/0610105.pdf`.

[4] S Hansell. Aol Removes Search Data on Group of Web Users New York Times, 2006. `https://www.nytimes.com/2006/08/08/business/media/08aol.html`.

[5] D Barth-Jones. The debate over 'reidentification'of health information: What do we risk?, 2012. `https://www.healthaffairs.org/do/10.1377/hblog20120810.021952/full/`.

[6] S Borenstein. Potential privacy lapse found in americans' 2010 census data, 2019. `https://www.apnews.com/aba8e57c145047b5bab11b62baaa7f7a`.

[7] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Rainer Lenz, Jane Longhurst, E Schulte Nordholt, Giovanni Seri, and P Wolf. Handbook on statistical disclosure control. *ESSnet on Statistical Disclosure Control*, 2010. `https://ec.europa.eu/eurostat/cros/system/files/CENEX-SDC_handbook.pdf`.

[8] Laura McKenna et al. Disclosure avoidance techniques used for the 1970 through 2010 decennial censuses of population and housing. Technical report, 2018. `https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20Techniques%20for%20the%201970-2010%20Censuses.pdf`.

[9] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 277–286. IEEE Computer Society, 2008. `https://lehd.ces.census.gov/doc/help/ICDE08_conference_0768.pdf`.

[10] Andrew Foote, Ashwin Machanavajjhala, and Kevin McKinney. Releasing Earnings Distributions using Differential Privacy: Disclosure Avoidance System For Post Secondary Employment Outcomes (PSEO). Working Papers 19-13, Center for Economic Studies, U.S. Census Bureau, April 2019. `https://www2.census.gov/ces/wp/2019/CES-WP-19-13.pdf`.

[11] J Mervis. Can a set of equations keep u.s. census data private?, 2019. `https://www.sciencemag.org/news/2019/01/can-set-equations-keep-us-census-data-private`.

[12] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067. ACM, 2014. `https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42852.pdf`.

[13] Andy Greenberg. Apple's 'differential privacy' is about collecting your data – but not your data. *Wired Magazine*, 2016. `https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/`.

[14] 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.

[15] 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.

2 Definitions, basic principles

2.1 Basic definitions

A data record is a vector, an ordered list, of k attributes (e.g. sex, race, income, single-year age bins, etc.). The set of all possible data records is the data universe χ. We assume χ is finite and the vectors in χ are ordered by an arbitrary index. For example, suppose we have two attributes of interest: SEX={Male, Female} and MAR={Married, Single, Other}. Then a data record might be R = (Male, Married) and the data universe would be χ = {(Male, Married), (Male, Single), (Male, Other), (Female, Married), (Female, Single), (Female, Other)}. In other words, χ = SEX × MAR, the Cartesian product of the sets of levels for each variable. A database D is a collection of records from χ.

You can think of D like a spreadsheet of rows and columns where each column corresponds to an attribute and each row is a record. Our notation is similar to [4]. However, it is often useful to represent our database D as a vector $x \in \mathbb{N}^{|\chi|}$ where x_i is the number of data records for the ith element of χ. Note $\mathbb{N}$ refers to the set of nonnegative integers throughout.

Example 2.1.

SEX	*MAR*
Female	*Married*
Male	*Other*
Female	*Single*
Male	*Married*
Female	*Single*
Female	*Single*
Male	*Other*
Female	*Married*

INDEX	*SEX*	*MAR*	*COUNT*
1	*Male*	*Married*	*1*
2	*Male*	*Single*	*0*
3	*Male*	*Other*	*2*
4	*Female*	*Married*	*2*
5	*Female*	*Single*	*3*
6	*Female*	*Other*	*0*

$$x = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix}$$

```
mydata1 = c("Female", "Male", "Female", "Male", "Female", "Female", "Male", "Female")
mydata2 = c("Married", "Other", "Single", "Married", "Single", "Single", "Other",
    "Married")

example_data = data.frame(SEX = mydata1, MAR = mydata2)
example_data$SEX = factor(example_data$SEX, levels = c("Male", "Female"))
example_data$MAR = factor(example_data$MAR, levels = c("Married", "Single", "Other"))

table(example_data$MAR, example_data$SEX)

dataVect=as.vector(table(example_data$MAR, example_data$SEX))
dataVect
```

Definition 1. A *randomized algorithm* $\mathcal{A} : \mathbb{N}^{|\chi|} \rightarrow \mathcal{O}$ is a mapping where each input $x \in \mathbb{N}^{|\chi|}$ has an associated conditional probability distribution over the output space $\mathcal{O}$. We write this as $Pr[\mathcal{A}(x) \in S]$ for $x \in \mathbb{N}^{|\chi|}$ and $S \subset \mathcal{O}$.

In other words, $\mathcal{A}(x)$ is a random variable. For an input x, a randomized algorithm will draw a single sample point z from the output space according to $Pr[\mathcal{A}(x) = z]$. For example, if $\mathcal{O} = \{True, False\}$, $\mathcal{A}(x)$ might output $True$ with probability .3 and $False$ with probability .7 whereas $\mathcal{A}(x')$ might output $True$ with probability .6 and $False$ with probability .4.

An important concept in differential privacy is that of neighboring databases. The definition of *neighboring* varies in the literature but generally reflects some notion of closeness. We consider the two main variants of neighboring here. First denote the $L1$ norm as $||x||_1 = \sum_{i=1}^{|\chi|} |x_i|$ and define the distance between two vectors x, y as $||x - y||_1$. Note that $||x||_1$ equals the number of records in the database.

Definition 2. Data vectors x and y are *neighbors (add-delete)* if $||x - y||_1 = 1$.

In other words, D, D' are represented by neighboring data vectors if the collection of records in D differs from the collection of records in D' by the addition or removal of exactly one record. Differential privacy defined with this notion of neighbors is sometimes called "unbounded" differential privacy [2].

Example 2.2.

SEX	*MAR*
Female	*Married*
Male	*Other*
Female	*Single*
Male	*Married*
Female	*Single*
Female	*Single*
Male	*Other*
Female	*Married*

$$x = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix}$$

SEX	*MAR*
Female	*Married*
Male	*Other*
Female	*Single*
Male	*Married*
Female	*Single*
Female	*Single*
Male	*Other*
Female	*Married*
Male	*Single*

$$y = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix}$$

Definition 3. Data vectors x and y are *neighbors (change-a-record)* if $||x||_1 = ||y||_1$ and $||x - y||_1 = 2$.

In other words, D, D' are represented by neighboring data vectors if the collection of records in D differs from the collection of records in D' by changing exactly one record. Differential privacy defined with this notion of neighbors is sometimes called "bounded" differential privacy.

The reason for these two definitions of neighbors is somewhat due to mathematical convenience.Typically, the change-a-record definition corresponds to situations where the database size is considered public and not necessary to protect.

Example 2.3.

SEX	*MAR*
Female	*Married*
Male	*Other*
Female	*Single*
Male	*Married*
Female	*Single*
Female	*Single*
Male	*Other*
Female	*Married*

$$x = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix}$$

SEX	*MAR*
Female	*Married*
Male	*Other*
Female	*Single*
Male	*Married*
Female	*Single*
Female	*Single*
Male	*Other*
Male	*Single*

$$y = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 3 \\ 0 \end{pmatrix}$$

Definition 4. A randomized algorithm M is *ϵ-differentially private* if for all $S \subset \mathcal{O}$ and for all neighboring data vectors x, y:

$$Pr[M(x) \in S] \leq e^{\epsilon} Pr[M(y) \in S]$$

Intuitively, Definition 4 says that the probability distribution of the mechanism for neighboring databases should be "close". Note that the inequality can also be written as

$$log \left(\frac{Pr[M(x) \in S]}{Pr[M(y) \in S]} \right) \leq \epsilon.$$

In Definition 4, "close" is measured by the maximum of the log of the ratio of output probabilities for any possible output and any pair of neighboring databases, x, y. However, keep in mind that there are other measures of "closeness" used to define "formally private" variations on basic differential privacy.

To help motivate this definition, it helps to think about this behavior in terms of the output of the differentially private algorithm. For any output from the algorithm, the likelihood given x is close to the likelihood given y for any two neighboring data vectors x and y. The motivation being that it is difficult for an attacker to tell if x or y is the more likely dataset from the output of the differentially private algorithm alone. Taking this a step further, x and y differ by only a single person's record. Therefore an attacker has difficulty learning any single person's record from the output of the algorithm alone.

In Definition 4, ϵ is the parameter that specifies how close output distributions must be for databases that are neighbors, and is more commonly referred to as the *privacy-loss budget* of the algorithm. As we will see later, privacy loss here can be understood in the sense that ϵ bounds the distance between the distribution of the mechanism output using the actual data and the distribution of the mechanism's output if it were run with an arbitrary person's data deleted or replaced with a record at random (a situation which, presumably, is private for the reference person).

We now formalize the group privacy concept introduced in Section 1.3. The privacy loss of a group degrades linearly in the size of the group.

Theorem 2.1. *Any (unbounded or bounded) ϵ-differentially private mechanism M is $k\epsilon$-differentially private for groups of size k. That is, for all x, y that differ by k records (add/delete or change-a-record) and all $\mathcal{S} \in \mathcal{O}$:*

$$Pr[M(x) \in \mathcal{S}] \leq e^{k\epsilon} Pr[M(y) \in \mathcal{S}]$$

Proof. See [3]. □

Groups of any size are protected to some degree by differentially private mechanisms. However, the privacy guarantee becomes less meaningful as group size grows. This is a reasonable property as it allows for learning about a population while simultaneously prevents learning about an individual or small groups (e.g. a family).

2.2 Queries

A (vector-valued) linear query is a function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$ such that $f(x) = Bx$ for some $B \in [-1, 1]^{k \times |\mathcal{X}|}$. A (vector-valued) counting query is a linear query where $B \in \{0, 1\}^{k \times |\mathcal{X}|}$.

Example 2.4 (scalar query). *Continuing with Example 2.1, the query "How many married people are there?" is defined by $f(x) = c^T x$ where $c^T = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$. And, of course, if x is the vector in Example 2.1, then*

$$c^T x = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix} = 3$$

```
#How many married people are there?
e = c(1,0,0,1,0,0)
query_answer=t(e) %*% dataVect
query_answer
```

Example 2.5 (vector-valued query)**.** *The query* $f(x) = Bx$ *where* $B = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$ *answers the following three questions: "How many married people are there?", "How many females are there?", and "How many married females are there?".*

Again using x *from Example 2.1,*

$$Bx = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}$$

```
#How many married people, females, married females are there?
B = matrix( c(1,0,0,1,0,0,
      0,0,0,1,1,1,
      0,0,0,1,0,0), byrow = TRUE,
      nrow = 3, ncol =6)

query_answer = B %*% dataVect
query_answer
```

Example 2.6. *Under the bounded differential privacy variant where the total* n *is known, we can ask proportions as linear queries. For example, suppose we know* $n = 8$*, and* x *is as in Example 2.1, then*

$$\begin{pmatrix} \frac{1}{8} & 0 & 0 & \frac{1}{8} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix} = 3/8$$

asks for the proportion of married individuals.

Example 2.7. *Under the unbounded differential privacy variant the total is not known and proportions cannot be asked* as linear queries*. We can however consider asking indirectly as follows:*

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \end{pmatrix}$$

The results can, of course, be combined to give the proportion of married individuals.

Definition 5. The *global* $L1$*-sensitivity* of a linear query f is

$$\Delta f = \max_{\substack{x,y \in \mathbb{N}^{|\mathcal{X}|} \\ x,y \text{ are neighbors}}} ||f(x) - f(y)||_1$$

Note that this is the max over *all possible pairs of neighboring databases*, not just neighbors of the actual or even likely databases. Global sensitivity is a worst-case measure of how much any one record can influence the value of a query and plays an important role in the scale of the random noise.

Definition 6. The identity query (sometimes called the histogram query) is defined as $h(x) = Ix = x$ where I is the $|\chi| \times |\chi|$ identity matrix.

1. The identity query for our SEX × MAR example represents asking the set of questions: "How many married males, single males, other males, married females, single females, and other females are there?":

$$Ix = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix}$$

What is the global L1-sensitivity of I?

Solution: Generally, the sensitivity depends on the definition of neighbors. For example, $||h(x) - h(y)||_1 = ||x - y||_1 = 1$ for all neighboring (add-delete) x, y and hence the global sensitivity would be 1. However, $||h(x) - h(y)||_1 = ||x - y||_1 = 2$ for all neighboring (change-a-record) x, y giving a global sensitivity of 2. Intuitively, adding/removing a record of type i changes the count of type i by 1, leaving all other counts fixed. On the other hand changing a record of type i to type j, increases the count of the latter by 1 while decreasing the count of the former by 1. All other counts are unaffected.

2. What is the global L1-sensitivity of an arbitrary vector-valued query?

Solution: For add/delete neighbors, if $f(x) = Ax$ is a vector-valued linear query, then Δf equals the maximum L1 norm taken over the columns of A. That is, $\Delta f = \max_i ||Ae_i||_1$ where e_i is the ith unit vector in $\mathbb{R}^{|\chi|}$. Intuitively, the L1 norm of the ith column represents the influence of a person of type i on the set of queries. In the case of (vector-valued) counting queries, the L1 norm of column i is literally the number of queries in which a person of type i is counted. Thus, finding the column with the maximum L1 norm is equivalent to finding which type of record, if added or removed, will shift the answer by the greatest distance. Notice this calculation is entirely independent of the actual data vector.

For change-a-record neighbors, $\Delta f = \max_{i,j} ||A(e_i - e_j)||_1$ where the max is over pairs of indices in $\mathbb{R}^{|\chi|}$. The intuition is similar and again the calculation is independent of the actual data vector.

Example 2.8. *Assume that our database includes a numeric variable and we are interested in calculating its mean over the database. (Note that in this case we might not want to represent the database as a vector). Assume that we are using the change-a-record definition of neighbors and the database has n records.*

3. What is the L1-sensitivity of the mean query?

Solution: This is actually a bit of a trick question because the answer depends importantly on details of the query's definition that the analyst can choose. If we allow the query to take on values $(-\infty, \infty)$ or even $(0, \infty)$, then the global sensitivity is infinite: adding/deleting or changing a single record could result in an infinite change in the mean, if that record's magnitude were large enough. When working with unbounded numeric variables, it is often necessary to constrain queries in order to achieve finite sensitivity, or to consider formally private frameworks that relax the notion of global sensitivity required for unbounded/bounded DP. It is common to "clip" or "truncate" a query's answer to a range $[L, U]$ (which must be chosen without looking at the true data; otherwise, we would need to treat the choice of L and U as sensitive queries too); in this case, the sensitivity of the mean query becomes $\frac{U-L}{n}$.

2.3 Standard Mechanisms

The most basic differential privacy mechanisms are designed to answer linear queries. They perturb the query output by adding noise drawn from a distribution centered at zero with scale calibrated to the sensitivity of the query and the privacy-loss parameter ϵ.

Definition 7. The Laplace distribution (centered at 0) with scale b has pdf:

$$Lap(b) = \frac{1}{2b} e^{(-\frac{|x|}{b})}$$

Note: The Laplace distribution is also called the double-exponential distribution. The Laplace distribution (centered at 0) has mean $\mu = 0$ and variance $\sigma^2 = 2b^2$.

Definition 8. Given a linear query f, the Laplace Mechanism is defined as $M(x; f, \epsilon) = f(x) + (Y)_{i=1}^{k}$ where each Y_i is an *i.i.d* draw from $Lap(\Delta f / \epsilon)$. We often write $M(x)$, suppressing the dependence on f and ϵ.

The Laplace Mechanism has mean $\mu = f(x)$ and variance $\sigma^2 = 2\left(\frac{\Delta f}{\epsilon}\right)^2$. In particular, the output distribution is centered on the true answer $f(x)$ and its variance increases with global sensitivity and decreases with ϵ. The noise introduced by the Laplace Mechanism is additive.

Example 2.9. *Let f and x be as in Example 2.4. We will use unbounded DP in this example. Hence, $\Delta f = 1$. We consider two values of epsilon below: $\epsilon = 1$ and $\epsilon = 10$.*

```
#if needed
#install.packages("rmutil")
library(rmutil)

laplace_mech = function(query_answer, epsilon=1, sens=1){
  n = length(query_answer)
  as.numeric(query_answer + rlaplace(n,0,sens/epsilon))
}
#note that sometimes the laplace distribution is paramaterized with a rate parameter
#rather than a scale parameter

#use the query from 4.3
query_answer = t(e) %*% dataVect
query_answer

laplace_mech(query_answer, epsilon=1, sens=1)

#Let's look at the distribution of the mechanism output
hist(replicate(10000, laplace_mech( query_answer, epsilon=1, sens=1)),prob=TRUE)
plot(density(replicate(10000, laplace_mech( query_answer, epsilon=1, sens=1))))

#What would happen if we change epsilon?
hist(replicate(10000, laplace_mech( query_answer, epsilon=10, sens=1)),prob=TRUE)
plot(density(replicate(10000, laplace_mech( query_answer, epsilon=10, sens=1))))
```

4. Let f and x be as in Example 2.5. What is the unbounded global L1 sensitivity? Will the variance of the Laplace Mechanism be larger with $\epsilon = 1$ or $\epsilon = 10$?

Solution: $\Delta f = \max_i ||Be_i||_1 = \max\{1, 0, 0, 3, 1, 1\} = 3$. The variance decreases in epsilon so $\epsilon = 1$ has smaller variance than $\epsilon = 10$.

```
#Variance for epsilon=1
2*(3/1)^2
#variance for epsilon=10
2*(3/10)^2
```

5. Draw the privacy-loss vs. accuracy tradeoff using ϵ and the variance as the measure of accuracy.

Solution:

```
#a short function to calculate the variance as a function of the sensitivity and
    epsilon
laplace_var = function(sens,epsilon){
     return(2*(sens/epsilon)^2)
     }

laplace_var(sens=3,epsilon=10)

#set our epsilon values starting at 0.01
epsilon_values = 0.01+ c(1:1000)/500

#apply the variance function to all our epsilon values
variance = sapply(epsilon_values, FUN = function(x) laplace_var(sens=3,epsilon=x) )

#plot using the SE
plot(epsilon_values, sqrt(variance), type = "l", xlab = "Epsilon", ylab = "Standard
    Error")

#plot relative to the worst variance
plot(epsilon_values,1 - variance/variance[1], type = "l", xlab = "Epsilon", ylab =
    "Accuracy")
```

Example 2.10. *Assume that we are using change-a-record sensitivity and that we have a database with $n = 25$ records. Say that we want to calculate a query that is the proportion of persons in the database that are Male (Assuming the database records are either Male or Female).*

6. If we were to use the Laplace mechanism with $\epsilon = 0.5$, what would be the standard error associated with the mechanism? What about for $n = 100$?

Solution: The sensitivity of the query is $1/n = 1/25$ because switching any record's sex in the database at most changes the proportion by $1/25$. The variance is $2\left(\frac{\Delta f}{\epsilon}\right)^2$. Therefore the standard error is

```
sqrt( 2 * (1/25 / 0.5 )^2 )
```

$n = 100$ should decrease the standard error

```
sqrt( 2 * (1/100 / 0.5 )^2 )
```

Theorem 2.2. *The Laplace Mechanism is ϵ-differentially private.*

Proof. The proof presented here follows [4]. Let M be defined as in Definition 8 for a given query f. Let $x, y \in \mathbb{N}^{|\mathcal{X}|}$ be neighbors. We must show $Pr[M(x) = z] \leq e^{\epsilon} Pr[M(y) = z]$ or alternatively that $\frac{Pr[M(x)=z]}{Pr[M(y)=z]} \leq e^{\epsilon}$.

$$\begin{aligned}
\frac{Pr[M(x) = z]}{Pr[M(y) = z]} &= \prod_{i=1}^{k} \left(\frac{exp\left(-\frac{\epsilon|f(x)_i - z_i|}{\Delta f}\right)}{exp\left(-\frac{\epsilon|f(y)_i - z_i|}{\Delta f}\right)} \right) \\
&= \prod_{i=1}^{k} exp\left(-\frac{\epsilon(|f(x)_i - z_i| - |f(y)_i - z_i|)}{\Delta f}\right) \\
&\leq \prod_{i=1}^{k} exp\left(-\frac{\epsilon(|f(x)_i - f(y)_i|)}{\Delta f}\right) \\
&= exp\left(-\frac{\epsilon(||f(x)_i - f(y)_i||_1)}{\Delta f}\right) \\
&\leq e^{\epsilon}
\end{aligned}$$

where the first line follows from the definition of the Laplace distribution (centered at $f(\cdot)$) with scale $\Delta f/\epsilon$, the first inequality follows from the triangle inequality, and the last inequality follows from the definition of global sensitivity since x and y are neighboring databases. The result holds for either bounded or unbounded DP. □

Definition 9. The two-sided geometric distribution with parameter $p \in (0, 1)$ is defined by:

$$Pr[X = k] = \frac{p}{2 - p}(1 - p)^{|k|}$$

where $k \in \mathbb{Z}$.

Note the geometric mechanism has mean $\mu = 0$ and variance $\sigma^2 = \frac{2(1-p)}{p^2}$. It can be computed as the difference of two geometric distributions.

Definition 10. Given a counting query f, the Geometric Mechanism is defined as $M(x; f, \epsilon) = f(x) + (Y)_{i=1}^{k}$ where each Y_i is a *i.i.d* draw from the two-sided geometric distribution with parameter $p = 1 - e^{-\epsilon/\Delta f}$

The Geometric Mechanism is a discrete variant of the Laplace Mechanism. It is useful when integer-valued output is desired. However, care must be taken when applying the Geometric Mechanism: without modification, we could not use it, for example, to directly estimate the answer to a question like, "What is $\frac{1}{3}$ the number of Married men in the database?" The trouble here is that neighboring databases can have disjoint supports, after the Geometric noise is infused.

For example, a database with 1 Married man could generate outputs of the form $\frac{1}{3} + k$ for any integer k, while a database with 2 married men could only generate outputs of the form $\frac{2}{3} + k$. Thus, if we saw a final publication value of $\frac{1}{3}$, we would be able to "rule out" the second database with certainty. This failure reflects that this use of two-sided Geometric noise is inappropriate, and, carefully analyzed, is not differentially private. This example recommends a useful "sanity check": a necessary condition for a mechanism to be DP is that all databases should have identical support, i.e., should agree on the set of *possible* outputs, though they can disagree on the probabilities of specific outputs.

Thus, in attempting to apply the Geometric mechanism, doing so with counting queries is appropriate, doing so with general linear queries is not. The distribution over outputs of the Geometric Mechanism has mean $\mu = f(x)$ and variance $\sigma^2 = \frac{2(1-p)}{p^2}$, with $p = 1 - e^{-\epsilon/\Delta f}$.

```
#the two-sided geometric distribution can be constructed as
#the difference between two independent random geometric variables
geom_mech = function(query_answer, epsilon=1, sens=1){
  n = length(query_answer)
  p = 1-exp(-epsilon / sens)
  x=rgeom(n, p)
  y=rgeom(n, p)
  return(as.numeric(query_answer + (x-y)))
}

#let's take a look at the geometric mechanism distribution
breaks = c(-12:12) + 0.5
hist(replicate(10000, geom_mech(0, epsilon=1, sens=1)), prob=TRUE, breaks = breaks)
```

Theorem 2.3. *The Geometric Mechanism is ϵ-differentially private.*

Proof. See [4]. □

2.3.1 Additional Reading

Other basic differential privacy mechanisms include NoisyMax, the exponential mechanism, and sparse vector. NoisyMax can be useful when the argmax of a set of queries is needed, but directly estimating all of the queries' values separately has high global sensitivity. The exponential mechanism is helpful when dealing with output spaces that are not subsets of $\mathbb{R}^n$ (e.g., some mechanisms may have outputs that are cat emojis, or 2-D images), while sparse vector is helpful when estimating a set of queries—most of which are strongly suspected *a priori* to have small true value—and which have large global sensitivity to estimate directly with the Laplace mechanism, similar to NoisyMax. You can read further about these mechanisms and their uses in [4]

2.4 Properties of Differential Privacy

2.4.1 Post-processing

One desirable privacy property that differential privacy possesses is robustness to post-processing. Colloquially, the post-processing property says you can't outsmart a differentially private algorithm. An attacker can't think of a clever way to use the output of a differentially private mechanism that somehow degrades its defined privacy guarantee. As a result of its immunity to post-processnig, differential privacy is future-proof in the sense that the guarantees hold regardless of what external information may become available ex post.

Theorem 2.4. *Let $M(x) : \mathbb{N}^{|\mathcal{X}|} \rightarrow R$ be an ϵ-differentially private algorithm. Let $\mathcal{A} : R \rightarrow R'$ be an arbitrary randomized algorithm. Then $\mathcal{A} \circ M$ is ϵ-differentially private.*

Proof. See [4]. □

Although post-processing does not compromise the privacy guarantee, post-processing can often be performed to improve statistical accuracy. For example, incorporating prior knowledge (that counts are non-negative) by truncating a Geometric Mechanism-derived estimate of a scalar counting query at 0 will reduce the estimator's variance:

Example 2.11. *Take Example 2.5*

$$Bx = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}$$

If we apply the Laplace mechanism, some of differentially private outputs may be negative. However since we know counts can't be negative, we can lower the variance of the post-processed outputs by truncating at zero.

We next present simplified versions of composition results which define how privacy loss degrades when outputs of two differentially private mechanisms are released.

2.4.2 Sequential Composition

Theorem 2.5. *Let M_1 be an unbounded ϵ_1-differentially private mechanism and unbounded M_2 be an ϵ_2-differentially private algorithm. Assume M_1 and M_2 have the same domain $\mathbb{N}^{|\mathcal{X}|}$. Define $M(x) = (M_1(x), M_2(x))$ for all $x \in \mathbb{N}^{|\mathcal{X}|}$. That is, the mechanism M runs M_1 and M_2 and returns the output from each. Then M is unbounded $(\epsilon_1 + \epsilon_2)$-differentially private. The result also holds replacing unbounded by bounded throughout.*

Proof. See [4]. □

In a general sense, sequential composition says that if a mechanism M_1 has privacy loss ϵ_1 and another mechanism M_2 has privacy loss ϵ_2, then the total privacy loss of the two mechanisms is no more than $\epsilon_1 + \epsilon_2$. There are a few useful consequences of this fact:

- In designing a differentially private mechanism, one can utilize multiple sub-mechanisms, and easily reason about their combined privacy loss.
- "Accounting" of the privacy-loss budget is easy. When there are multiple sub-mechanisms, the overall privacy-loss budget is sometimes called the *global privacy-loss budget.*
- Though it requires a slightly more general, "adaptive" form of the sequential composition theorem than the version we have stated here, it is also OK to use the output of one differentially private mechanism to choose which of several other differentially private mechanisms to invoke: the privacy loss still accumulates additively.

2.4.3 Parallel Composition

Parallel composition is a more efficient but restrictive alternative to sequential composition: when you can invoke parallel composition, you should do so, but you won't always have that option. We focus here on unbounded DP, but briefly describe parallel composition for bounded DP at this section's end.

An informal description of parallel composition is as follows: when asking separate queries about disjoint subpopulations the privacy loss does not accumulate beyond the maximum loss of the queries considered individually. For example, asking about the number of men with $\epsilon_1 = 2$ and asking about the number of women with $\epsilon_2 = 3$ results in a total loss of 3 rather than 5. Since we could have set $\epsilon_1 = 3$ as well without incurring any additional cost, it is generally considered wasteful to not use equal parameter settings on disjoint subpopulations.

Theorem 2.6. *Let χ_1, χ_2 partition χ. That is, χ_1 and χ_2 are disjoint and $\chi_1 \cup \chi_2 = \chi$. For convenience (also without loss of generality), we assume $\chi_1 = \{1, \dots, |\chi_1|\}$ and $\chi_2 = \{|\chi_1| + 1, \dots, |\chi|\}$ Then for any $x \in \mathbb{N}^{|\chi|}$ we can decompose x such that $x = x^1 + x^2$ where $x^1_i = 0$ if $i \in \chi_2$ and $x^2_i = 0$ if $i \in \chi_1$. Let M_1 be an unbounded ϵ_1-differentially private mechanism and M_2 be an unbounded ϵ_2-differentially private algorithm such that M_1 has domain $\mathbb{N}^{|\chi_1|} \times \{0\}^{|\chi_2|}$ and M_2 has domain $\{0\}^{|\chi_1|} \times \mathbb{N}^{|\chi_2|}$. Define $M(x) = (M_1(x^1), M_2(x^2))$ for all $x \in \mathbb{N}^{|\chi|}$. Then M is* $\max\{\epsilon_1, \epsilon_2\}$*-differentially private*

Proof.

$$\begin{aligned}
\frac{Pr[M(x) = (z_1, z_2)]}{Pr[M(y) = (z_1, z_2)]} &= \frac{Pr[M_1(x^1) = z_1]Pr[M_2(x^2) = z_2]}{Pr[M_1(y^1) = z_1]Pr[M_2(y^2) = z_2]} \\
&\leq \max\{\frac{Pr[M_1(x^1) = z_1]}{Pr[M_1(y^1) = z_1]}, \frac{Pr[M_2(x^2) = z_2]}{Pr[M_2(y^2) = z_2]}\} \\
&\leq e^{\max\{\epsilon_1, \epsilon_2\}}
\end{aligned}$$

where the first line follows by independence, the second line follows since adding/deleting a record can only change the probability of at most one partition so one of the ratio always equals 1, and the inequality comes from differential privacy guarantees of M_1 and M_2. □

7. Discuss how the composition of two scalar valued counting queries compares to asking a single equivalent vector-valued query. Assume unbounded differential privacy.

Solution: First consider asking the following two queries:

$$\begin{aligned}
f(x) &= \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} x \\
g(x) &= \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} x
\end{aligned}$$

compared with asking the single query:

$$h(x) = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Notice that $\Delta f = \Delta g = 1$ but $\Delta h = 2$. Hence the Laplace mechanism $M(x; h, \epsilon)$ that adds $Lap(2/\epsilon)$ noise to each component is ϵ-differentially private. The Laplace mechanism $M_1(x; f, \epsilon_1)$ that adds $Lap(1/\epsilon_1)$ to f alone is ϵ_1-differentially private and likewise $M_2(x; g, \epsilon_2)$ applied to g alone. However, releasing (M_1, M_2) together incurs a cost of $\epsilon_1 + \epsilon_2$ due to sequential composition. Note that by setting $\epsilon_i = \epsilon/2$ the composition approach matches the noise and privacy guarantee of M.

Second consider asking the following two queries:

$$\begin{aligned}
f(x) &= \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} x \\
g(x) &= \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} x
\end{aligned}$$

compared with asking the single query:

$$h(x) = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Suppose M_1 adds Laplace noise to f and M_2 add Laplace noise to g. Under unbounded differential privacy, $\Delta f = \Delta g = 1$. So M_1 and M_2 can each add $Lap(1/\epsilon)$ noise, and the parallel composition theorem says the combined privacy loss is only ϵ. Now notice that $\Delta h = 1$ so the (vector-valued) Laplace mechanism applied to h achieves ϵ-differential privacy by adding $Lap(1/\epsilon)$ noise to each component.

8. Identify which groups allow for parallel composition and which require sequential composition:

   1 Number of Germans, Number of Europeans

   2 Number of 1 person households, Number of multigenerational households

   3 Number of 65+ year olds, Number of expectant mothers.

   4 Number of girls, Number of Eagle Scouts.

   5 Number of Marvel fans, Number of DC fans

   6 Number of Married, Number of Singles

Solution:

1 Sequential Composition

2 Parallel Composition

3 May vary. Although uncommon (even highly improbable), is it impossible for a 65+ year old to be pregnant? If not, sequential composition. However, it may be common knowledge that elderly pregnant women were excluded from the study in which case parallel composition does apply.

4 May vary. In 2017, Boy Scouts allowed girls to join their ranks and so it is no longer impossible for girls to be Eagle Scouts. However if it is public knowledge the database only contains records from prior to 2017, then the subpopulations are disjoint so parallel composition works.

5 Sequential Composition

6 Parallel Composition

For bounded DP, parallel composition requires a little bit more thought: if M_1 and M_2 are bounded ϵ-DP, and M_1 asks about the count of Married Men while M_2 asks about the count of Women in a fixed database D, then it does not follow that (M_1, M_2) is bounded ϵ-DP, because we can modify a single Married Man record and convert it into a Woman record. If we do so, then both the true answer to M_1 and the true answer to M_2 would change by 1; this behavior is quite unlike the unbounded DP case, where only M_1 or M_2 could change when adding-deleting a record, but not both.

As this discussion suggests, we do recover a slightly worse variant of parallel composition for bounded DP, though: if each M_i in $(M_1, M_2, \ldots, M_k), k \geq 2$ is ϵ-bounded DP, the M_i each use independent random coins, and *a priori* we know that none of M_i can depend on overlapping sub-populations, then we may invoke parallel composition to conclude that $(M_1, M_2, \ldots, M_k)$ is no worse than bounded 2ϵ-DP. This version of parallel composition is of course only helpful when $k > 2$, since it is no better than sequential composition when $k \in \{1, 2\}$. The proof of this fact is identical to the proof of Theorem 2.6 , except that k factors appear initially, and only 2 survive reasoning about the impact of a change between neighboring datbases.

2.5 Complex Mechanisms

Highly sophisticated differential privacy mechanisms almost always use a core set of basic mechanisms as building blocks. Here are a few examples:

- The Matrix Mechanism and High-Dimensional Matrix Mechanism follow a "Select - Measure - Reconstruct" paradigm: given a set of target queries on which low accuracy is desired, these mechanisms Select an alternative set of queries, Measure estimated values of these alternative queries using the Laplace mechanism, and perform post-processing to optimally Reconstruct estimates to the original, target queries. As is common (though certainly not universal) with complex DP mechanisms, the privacy guarantee for MM/HDMM is simple to analyze, as it relies only on the Laplace mechanism and post-processing. (Accuracy requires more thought, as is also common.) See [5] and [6].
- The Multiplicative-Weights Exponential Mechanism uses a combination of two basic mechanisms: the exponential mechanism (see [4]) and the Laplace mechanism. These two basic mechanisms are used to iteratively improve on an initial guess at a histogram, focusing progress on queries for which answers appear to be worst. The privacy claim follows from a combination of composition and post-processing coupled with the guarantees of the basic mechanisms. See [7].
- The hierarchical count-of-counts mechanism derives its privacy guarantee from the geometric mechanism. See [8].

The reader may also find it interesting to explore HBTree [9] and the non-adaptive variant of the DP spatial-grid algorithm [10]. These mechanisms are sometimes called "data-independent", a label we have borrowed from Li et al. [11]; a mechanism is typically regarded as data-independent if the noise it introduces does not depend on the true data, and in particular if its error distribution can be described without dependence on the true data. Although restrictive, data-independent mechanisms can nevertheless be highly sophisticated, as there are many algorithm design decisions which can be optimized without looking at the true, sensitive data.

Mechanisms which are not data-independent may be (appropriately!) called data-dependent. Common examples are the Multiplicative-Weights Exponential Mechanism (MWEM) [7], PrivTree [12], iReduct [13], AHP [14], and the second, adaptive-grid algorithm [10]. Data-dependent mechanisms typically expend some privacy-loss budget on not just directly getting good estimates of the queries of interest, but in identifying important structural properties of the data, and only spending additional privacy-loss budget once major features of the data are understood. MWEm, for example, iteratively tries to identify the query on which performance is worst, then measures the true value of that query, uses the measurement to improve its output, repeats this process until the privacy-loss budget is exhausted. PrivTree recursively sub-divides a 2-dimensional space into quadrants, recursively continuing to sub-divide only those quadrants in which the count of records is detectably greater than 0.

2.6 Questions?

References

[1] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[2] Daniel Kifer and Ashwin Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 193–204, New York, NY, USA, 2011. ACM Digital Library.

[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] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 351–360, New York, NY, USA, 2009. ACM.

[5] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor. Optimizing linear counting queries under differential privacy. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 123–134. ACM, 2010.

[6] Ryan McKenna, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. Optimizing error of high-dimensional statistical queries under differential privacy. *Proceedings of the VLDB endowment*, 11(10), 2018.

[7] Moritz Hardt, Katrina Ligett, and Frank McSherry. A simple and practical algorithm for differentially private data release. *CoRR*, abs/1012.4763, 2010.

[8] Yu-Hsuan Kuo, Cho-Chun Chiu, Daniel Kifer, Michael Hay, and Ashwin Machanavajjhala. Differentially private hierarchical group size estimation. *CoRR*, abs/1804.00370, 2018.

[9] Understanding hierarchical methods for differentially private histograms. *Proceedings of the VLDB Endowment*, 6(14), 2013.

[10] W. Qardaji, Y. Weining, and L. Ninghui. Differentially private grids for geospatial data. *arXiv*, 2013.

[11] C. Li, M. Hay, G. Miklau, and Y. Wang. A data- and workload-aware algorithm for range queries under differential privacy. *Proceedings of the VLDB Endowment*, 7(5), 2014.

[12] J. Zhang, X. Xiao, and X. Xie. Privtree: A differentially private algorithm for hierarchical decompositions. *arXiv*, 2016.

[13] X. Xiao, G. Bender, M. Hay, and J. Gehrke. ireduct: Differential privacy with reduced relative errors. *SIGMOD '11*, 2011.

[14] X. Zhang, R. Chen, J. Xu, X. Meng, and Y. Xie. Towards accurate histogram publication under differential privacy. *SIAM*, 2014.

3 Semantics

The definition of differential privacy requires that the probabilities governing the output of a mechanism not depend sensitively on the data of any single database participant. Intuitively, we should suspect that this idea is related to some more concrete notion of privacy, but—without additional argument—the connection is vague. In the theory of formal privacy, *semantic frameworks*, or just "semantics," provide directly meaningful interpretations of privacy guarantees of DP mechanisms, and therefore clarify this connection.

Semantics compare the inferences an attacker can make given a data publication with the inferences they could have made in some "private baseline":

- *Counterfactual baselines*
  - Motivated by the idea that if *your* data wasn't used, then the data publication can't compromise *your* privacy
  - If an attacker's inference is not much better compared to what it would have been in the private baseline, then, intuitively, privacy loss must be minimal
  - *Relative* guarantee: measures improvement in inference relative to a world in which your data was not used. Contrast with *absolute* guarantees: that it be impossible for an attacker to learn some unknown secret fact about you
  - Equivalently, can be thought of as bounding the increase in risk to you of any good or bad inference being made specifically because you participated in data collection
  - All inferences guarded against: protection for inferences *about you*, whether helpful/harmful, for weird/boring inferences, inferences on other people/entities, etc
  - Are *worst-case*:
    - all possible mechanism outputs
    - attackers are assumed able to exactly Bayes' update with no computational restrictions
    - protection is bounded for all inferences & persons uniformly
    - protection is even offered against attackers with incorrect beliefs
  - Protection expressed *probabilistically*, not deterministically: guarantees are about increases in confidence, measured in ratios of probabilities or odds
- *Prior baselines*
  - Compares prior & posterior inferences of the attacker after viewing the publication output
  - Measures what additional information the attacker can learn about you from the publication output, beyond what they already know
  - Also a *relative* guarantee
  - Are also *worst-case*
  - Are also expressed probabilistically

In this tutorial, we'll focus primarily on counterfactual baselines, as they are somewhat more common in the formal privacy literature, and they encode one reasonable notion of it means for information to be private: in counterfactual approaches, *your information is defined as private if it could not be inferred by only looking at the information of other people.* Conversely, if some information *could* be inferred by an attacker combining their prior knowledge with a publication based on only other persons' data, then counterfactual semantics treat that information as non-private for you.

Although counterfactual baselines are uniquely important in the formal privacy literature, it is worth being aware that differential privacy does support proofs of semantic guarantees against general classes of attackers while using alternative baselines, such as prior baselines, although these other guarantees can be a bit more complicated to interpret in full generality. We briefly comment on some alternative approaches at the end of this section.

3.1 Notation

We'll need some notation in this section to represent the attacker's prior beliefs about the database and specific records in the database:

D : the database; a random variable in the view of the attacker

$\pi(D)$: the attacker's prior beliefs/knowledge about the database

$M()$: an ϵ-differentially private mechanism

D_i : record i from database D; a random variable, for the attacker

D^{-i} : the database with record i deleted

3.2 A Counterfactual Semantics Theorem

3.2.1 Simulation: A Counterfactual Baseline Semantics Framework

Semantic frameworks with a counterfactual baseline are often presented using a kind of "simulation" argument, with the goal of providing a guarantee like: for an arbitrary attacker, after viewing the output of a data publication mechanism, the attacker's posterior confidence will not be very different from what it would have been in a simulation of the data release where your data was not used (e.g., was never submitted to the data curator, or was deleted upon receipt).

Arguments of this kind appear in several places in the literature; Smith & Kasiviswanathan [1] provide one important, early contribution to this body of work. As a technical observation, note that Smith & Kasiviswanathan measure improvement in attacker confidence using the *statistical difference between related posterior distributions.* Here, we instead elect to use *posterior odds ratios* to emphasize counterfactual semantics' connection to an attacker comparing the likelihood of two specific, competing hypotheses:

Theorem 3.1. *Let $\pi(D)$ be an arbitrary attacker prior, M be a ϵ-differentially private mechanism, i be the reference person whose privacy guarantee we currently consider, and let S, S' be two competing hypotheses about record j: we may have either $D_j \in S$ or $D_j \in S'$, with $S \cap S' = \emptyset$. Then, for any mechanism output ω,*

$$e^{-2\epsilon} \leq \frac{\frac{\pi[D_j \in S | M(D)=\omega]}{\pi[D_j \in S' | M(D)=\omega]}}{\frac{\pi[D_j \in S | M(D^{-i})=\omega]}{\pi[D_j \in S' | M(D^{-i})=\omega]}} \leq e^{2\epsilon}$$

Intuitively, this counterfactual-baseline result says that, having seen the final data publication—regardless of what values were actually published—the attacker's posterior confidence that S is more likely than S' to be true of person i's record will not be very different (if ϵ is small) than it would have been had the publication mechanism instead deleted person i's data before running. Recalling some of our earlier discussion, note that j is unrestricted: we can set $j = i$ to investigate the way in which the reference person i's privacy guarantee protects against attacker inferences about properties of him or herself, but we can also set $j \neq i$ to explore bounds on how much we might learn about someone or something else because of i's participation in data collection.

Although counterfactual-baseline semantic guarantees are very general, applying to arbitrary attackers, there is important nuance to note in what they do *not* guarantee:

- Counterfactual-baseline results *don't* promise to bound the absolute improvement in attacker inference, only that their improvement in inferential confidence could have been about the same even without your data.
- Notice that, if the attacker is aware of strong statistical dependence between your and other persons' records, then even for small ϵ, an attacker might be able to dramatically improve their inferences about you from a data release, despite the counterfactual semantics bound.

- That this might occur does not contribute to the bound on the right-hand side of Theorem 3.1 because, when using a counterfactual baseline, *information that can be inferred about you using someone else's data is regarded as legitimate inference*, not as a privacy violation.

We will briefly revisit the issue of statistical dependence between records and its relation to semantic guarantees when we touch on alternatives to counterfactual semantics at the end of this section.

3.3 Some Simple Motivating Scenarios

In this section, we'll consider some more concrete examples of calculations involving Theorem 3.1. It will be helpful to consider 2 motivating scenarios:

- Alice has a rare disease and is a resident of an area widely publicized as suffering from high levels of air pollution. She was asked to partake in a medical study studying the link between this disease and hiking. Alice wants to help the research community, but is worried that fine-grained analyses released based on her participation could lead to her being identified in the sample, and to prospective medical insurance companies trying to charge her higher premiums. Could differential privacy be useful to alleviate Alice's concerns?
- Bob is considering providing his data to for use in a sample-based survey of demographic characteristics, but he lives in a sparsely populated area, and he has seen reports of racial discrimination in his local community against members of a race grouping that he self-identifies as a member of. Bob is worried that, if he honestly responds to the survey, then his race and location may be easily identified with the help of statistical information published as a result of the survey. If the organization collecting the survey data were to use a differentially private mechanism to publish their final statistical reports, should that alleviate Bob's concerns?

In both cases, if a differentially private publication mechanism is used, then we can provide clear, precise reasoning about the worst-case risk to Alice and Bob of their private information leaking in ways that they would not like, even if a sophisticated attacker uses the published information to make best guesses about Alice and Bob's characteristics.

3.3.1 Example: Bounding Alice's Privacy Risk

In assessing Alice's privacy risk from participating in the medical study, we need to first consider what could happen if she does not participate in the study. In general, we should consider many possibilities, focusing on those that, with suitable expert judgment, are most likely to occur and most harmful. However, to keep our illustration manageable, let us focus on just possible harm: if Alice does not participate in the medical study, it is possible that the study will nevertheless conclude that exposure to elevated levels of air pollution is causally linked to acquiring the rare disease from which Alice suffers.

A prospective insurer might then use Alice's address and the results of the study to infer that—owing to the pollution near her residence—she is 10% more likely to have the rare disease than not to have it. We further suppose that the prospective insurer uses a very simple system to calculate premiums: the insurer is assumed to charge Alice a premium equal to $\$500 \cdot p$, where p is the insurer's percentage-wise belief about how much more likely Alice is to have the rare disease than not to have it. In this scenario, $\$500 \cdot 0.1 = \$50$ is Alice's *baseline*: whether or not she participates in the study, in this scenario she can expect a \$50 increase in premium offer from a prospective insurer.

We can now use the counterfactual semantics codified in Theorem 3.1 to bound Alice's increase in risk from participating in the study, which is—in the counterfactual semantics—identical to Alice's privacy risk. Letting $S = \{\text{Alice has the rare disease}\}, S' = \{\text{Alice does not have the rare disease}\}$, we posited in our baseline scenario that $\frac{Pr_\theta[D_i \in S | \mathbb{M}(D^*_{-i})=\omega]}{Pr_\theta[D_i \in S' | \mathbb{M}(D^*_{-i})=\omega]} = 1.1$, and we know from Theorem 3.1 that

$$\frac{\frac{\pi[D_i \in S | M(D)=\omega]}{\pi[D_i \in S' | M(D)=\omega]}}{\frac{\pi[D_i \in S | M(D^{-i})=\omega]}{\pi[D_i \in S' | M(D^{-i})=\omega]}} = \frac{\frac{\pi[D_i \in S | M(D)=\omega]}{\pi[D_i \in S' | M(D)=\omega]}}{1.1} \leq e^{2\epsilon}$$

Hence,

$$\frac{\pi[D_i \in S | M(D) = \omega]}{\pi[D_i \in S' | M(D) = \omega]} \leq 1.1 e^{2\epsilon} \tag{1}$$

When $\epsilon = 0.01$, for example, inequality 1 says that, even if Alice participates in the medical study, then the prospective insurer cannot think Alice is more than 12% more likely to have the rare disease than not to have it (given that the insurer would have believed Alice was only 10% more likely to have the rare disease than not to have it, in a world where the medical study's results were published without using Alice's data). Given our assumptions about the prospective insurer's pricing model, we conclude that Alice cannot face more than a \$500 · 0.12 = \$60 increase in her insurance premium because of her participation in the study, if she chooses to participate.

Of course, the counterfactual guarantee degrades swiftly as ϵ increases, owing to the exponential in the bound of Theorem 1, and so an algorithm designer, data curator concerned about privacy risks faced by participants in data collection—or a prospective study participant—should carefully consider the choice of ϵ, weighing the quality of the worst-case counterfactual semantics guarantees against the greater accuracy and associated social benefit that can be achieved by infusing less noise (i.e., larger ϵ). Figure 3.3.1 illustrates how the counterfactual semantics worst-case bound varies as the counterfactual belief $p = \frac{Pr_\theta[D_i \in S | \mathbb{M}(D^*_{-i})=\omega]}{Pr_\theta[D_i \in S' | \mathbb{M}(D^*_{-i})=\omega]}$ and ϵ are varied:

Figure 3.3.1

Privacy Risk as a Function of ϵ

p / ϵ	0.01	0.1	1	2	4	8	16
10e−10	1e−9	1e−9	7e−9	5e−8	3e−6	0.009	80000
1.1	1.12	1.34	8.12	60	3300	10e7	10e14
2	2.04	2.44	15	109	6000	10e8	10e15
4	4	5	29	218	12000	10e8	10e15
8	8.16	10	60	436	23847	10e8	10e15
16	16.32	19.5	118	873	47695	10e9	10e16

In closing our discussion of this example, we make a few comments on Figure 3.3.1:

- It is common in explanations of this issue in the literature to use linear approximations to the exponential function in the bound of theorems like Theorem 1, which are appropriate for small ϵ. We have chosen not to use such an approximation, to emphasize how rapidly the semantic bound degrades as ϵ increases.
- As this table shows, the specific setting of ϵ is fundamental to understanding the guarantee provided by an ϵ-DP mechanism (bounded or unbounded). In the literature, it is common to assume $\epsilon \approx 0$, but so far practical applications have tended to use ϵ between 1 and 40.
- The bounds in the bottom-right of Figure 3.3.1 are very weak, but the reader should not necessarily infer that these bounds provide no privacy. Guarantees of this type must be understood in context: if $p \approx 0$ (as in the second row, where $p =$ 10e−10), even a very poor privacy guarantee will not allow an attacker to infer that S is true of Alice with confidence, and so it may still be valuable to have this guarantee carefully, deductively proven.

3.3.2 Exercise: Bounding Bob's Privacy Risk

Bob is concerned about racial discrimination in his area. For our stylized example, suppose that Bob is comfortable responding to the survey if he is confident that the probability of an attacker identifying his race cannot increase, as a result of his participation in data collection, by more than 25%. Bob believes the probability of this happening before survey data is collected to be about $\pi[D_i \in S] = 0.02$, where D_i refers to Bob's record and $S = \{\text{Bob is of race R}\}$. If an unbounded ϵ-differentially private mechanism is used to publish the set of final statistics based on the collected data, at what values of ϵ would Bob be comfortable participating?

Before we try to solve this problem, we should first notice that Bob's risk tolerance is not stated in the same form as the left-hand-side that is bounded in Theorem 3.1. Directly applying the theorem is therefore not very easy, but let us try to do so and see where it takes us.

Let $S' = S^c$, and M be the unbounded ϵ-DP mechanism assumed in use to publish the survey statistics. The attacker's confidence about Bob's race may improve even if Bob does not participate, but we can reasonably interpret $\frac{\pi[D_i \in S|M(D)=\omega]}{\pi[D_i \in S|M(D^{-i})=\omega]}$ as the multiplicative increase in this probability due to Bob's participation, in the event that the published output statistics are ω. We therefore understand Bob's concern as requiring that, for any attacker and possible published statistics ω,

$$\frac{\pi[D_i \in S|M(D) = \omega]}{\pi[D_i \in S|M(D^{-i}) = \omega]} \leq 1.25$$

From Theorem 3.1, we know

$$\frac{\frac{\pi[D_i \in S|M(D)=\omega]}{\pi[D_i \in S'|M(D)=\omega]}}{\frac{\pi[D_i \in S|M(D^{-i})=\omega]}{\pi[D_i \in S'|M(D^{-i})=\omega]}} \leq e^{2\epsilon}$$

and, using $S' = S^c$, we can derive

$$\frac{\pi[D_i \in S|M(D) = \omega]}{\pi[D_i \in S|M(D^{-i}) = \omega]} \leq e^{2\epsilon} \frac{1 - \pi[D_i \in S|M(D^{-i}) = \omega]}{1 - \pi[D_i \in S|M(D) = \omega]}$$

With a bit of work, we can transform this to show that:

$$\frac{\pi[D_i \in S|M(D) = \omega]}{\pi[D_i \in S|M(D^{-i}) = \omega]} \leq e^{2\epsilon} + \frac{1}{1 - \pi[D_i \in S|M(D^{-i}) = \omega]}$$

Unfortunately, to use this bound requires us to be able to bound $\pi[D_i \in S|M(D^{-i}) = \omega]$ away from 1, and doing *that* requires us to reason about probabilities over the data set D, the attacker's prior π, and the algorithm used. Can we somehow do better than this solution, and derive a guarantee that does not depend on these additional details?

In this case, it can be fruitful to use the definition of DP directly, and to try to bound the quantity of interest. To see how this might work, we adopt the same notation as in the previous, partial solution, and seek to establish an ϵ at which we can meet Bob's desired bound:

$$\frac{\pi[D_i \in S|M(D) = \omega]}{\pi[D_i \in S|M(D^{-i}) = \omega]} \leq 1.25$$

Notice we can use Bayes' theorem to re-express Bob's desired bound as:

$$\frac{\frac{\pi[M(D)=\omega|D_i \in S]\pi[D_i \in S]}{\pi[M(D)=\omega]}}{\frac{\pi[M(D^{-i})=\omega|D_i \in S]\pi[D_i \in S]}{\pi[M(D^{-i})=\omega]}} = \frac{\frac{\pi[M(D)=\omega|D_i \in S]}{\pi[M(D)=\omega]}}{\frac{\pi[M(D^{-i})=\omega|D_i \in S]}{\pi[M(D^{-i})=\omega]}} \leq 1.25$$

We can now bound the two ratios in this expression separately. First, we argue that:

$$\frac{\pi[M(D) = \omega|D_i \in S]}{\pi[M(D^{-i}) = \omega|D_i \in S]} = \frac{\pi[M(D) = \omega, D_i \in S]}{\pi[M(D^{-i}) = \omega, D_i \in S]} = \frac{\sum_{d:d_i \in S} \pi[M(d) = \omega]\pi[D = d]}{\sum_{d:d_i \in S} \pi[M(d^{-i}) = \omega]\pi[D = d]}$$

We then notice that, for any d with $d_i \in S$, there is exactly one term in numerator and denominator, and d, d^{-i} are neighboring databases in the unbounded sense. Invoking the definition of unbounded ϵ-DP for each of these terms separately, we can see that

$$\frac{\sum_{d:d_i \in S} \pi[M(d) = \omega]\pi[D = d]}{\sum_{d:d_i \in S} \pi[M(d^{-i}) = \omega]\pi[D = d]} \leq \frac{\sum_{d:d_i \in S} e^{\epsilon}\pi[M(d^{-i}) = \omega]\pi[D = d]}{\sum_{d:d_i \in S} \pi[M(d^{-i}) = \omega]\pi[D = d]} = e^{\epsilon}$$

This argument gets us half-way to bounding Bob's privacy risk. You should try now to complete this argument! If you're stuck, see the solution box below, or raise your hand and let the instructors know you're having a bit of trouble. After a few minutes, we'll walk through the solution together to make sure we all understand it.

Solution: For the remaining ratio, we can get an analogous lower bound, arguing term-by-term:

$$\frac{\pi[M(D) = \omega]}{\pi[M(D^{-i}) = \omega]} = \frac{\sum_d \pi[M(d) = \omega]\pi[D = d]}{\sum_d \pi[M(d^{-i}) = \omega]\pi[D = d]} \geq e^{-\epsilon}$$

Combining the two bounds, we now see that

$$\frac{\frac{\pi[M(D)=\omega|D_i \in S]\pi[D_i \in S]}{\pi[M(D)=\omega]}}{\frac{\pi[M(D^{-i})=\omega|D_i \in S]\pi[D_i \in S]}{\pi[M(D^{-i})=\omega]}} = \frac{\frac{\pi[M(D)=\omega|D_i \in S]}{\pi[M(D)=\omega]}}{\frac{\pi[M(D^{-i})=\omega|D_i \in S]}{\pi[M(D^{-i})=\omega]}} \leq e^{2\epsilon}$$

We can therefore solve:

$$e^{2\epsilon} = 1.25 \Rightarrow \epsilon = 0.5ln(1.25) \approx 0.11$$

And conclude that, for $\epsilon < 0.5ln(1.25)$, the increase in Bob's risk of his individual race being identified by an attacker cannot be more than 25% higher than it would have been had Bob not participated in the survey at all, no matter what final statistics are published, regardless of what data happens to be collected for other persons in the survey, and without needing to know the details of the unbounded ϵ-DP publication algorithm, M. Given Bob's assumed tolerance for this risk, for these values of ϵ he would therefore be OK with participating in data collection.

It is worth noticing an unasked-for benefit of this solution as well: although in our example, Bob committed to a particular prior probability with which he suspected an attacker might believe him to be of race R, $\pi[D_i \in S] = 0.02$, but the privacy guarantee we derived does not depend on this assumption. Even if Bob is wrong and the prior probability of an attacker inferring that he is of race R is much larger or much smaller than 0.02, we can still guarantee Bob that the attacker's improvement in probability of believing that Bob is of race R won't be more than 25%, compared to a world where Bob's data was not collected.

Alice and Bob have qualitatively different risks that concern them, and even qualitatively different forms in which they have chosen to express their tolerance for risk from participating in surveys (which is identical to their privacy risk, in the counterfactual semantics framework). Nevertheless, in both cases, we saw that, if a differentially private algorithm was used to perform data publication, then we could explicitly calculate an ϵ at which Alice and Bob should be willing to participate in data collection, given each of their risk tolerances.

The ability to directly model arbitrary types of privacy risks faced by respondents from contributing their data to a survey or census is a testament to differential privacy's mathematical fecundity, and allows for providing respondents with transparent privacy guarantees that do not depend on secret details of the disclosure-avoidance/data publication mechanism in use. As it turns out, even this generality is only scratching the surface: to close out this section, we'll consider several alternative ways in which semantic guarantees can be derived and stated for DP mechanisms.

3.4 Semantics: Caveats & Extensions

As we saw in working with Theorem 3.1, and in deriving mathematical bounds to satisfy the concrete risks concerning Alice and Bob, differentially private release mechanisms can provide very general, precise guarantees. However, we should also be careful to remember the kinds of guarantees we have *not* provided, and to emphasize some of the key assumptions we have made:

- We have provided bounds relative to a counterfactual world in which the respondent's data was not used
- We have provided bounds considering only the possibility of a single person's data being deleted at a time, and have identified this as synonymous with "privacy risk." We have not provided privacy bounds for groups of people
- We have *not* provided bounds on how much an attacker can learn relative to what they already knew, only relative to what they could learn when the reference person's data is not used
- We have assumed the use of pure unbounded ϵ DP, which in particular means that every output statistic either must be noisy or must be classified as "public knowledge" and outside of the data collector's legal or ethical obligations to guard against
- We have used pure unbounded ϵ DP, which treats all information about respondents as equally important to protect, and all possible attackers as of equal importance to protect against

All 5 of these restrictions can be relaxed, or alternatives to them presented, though some are more easily loosened than others:

- Although the counterfactual approach is popular, it is also possible to present the privacy guarantees in a semantic framework that directly bounds how much an attacker can learn about an individual relative to what the attacker knew prior to looking at the published data.
  - The Pufferfish framework [2] provides an important example of this approach. Unlike the counterfactual approach, for this kind of result it is necessary to assume a bound on the strength of statistical dependence between records in the attacker's prior
  - The Pufferfish paper contains, for example, a theorem showing an equivalence between unbounded ϵ-DP and protecting against all inferences, but only for attackers who treat all records as statistically independent
- Guarantees for *groups of persons* can be derived in at least two ways
  - The database representation may actually contain "individuals" that correspond to groups of persons (e.g., a database row may represent a company or family).
  - The DP "group privacy" guarantee Theorem 2.1 can be invoked. A group variant of Theorem 3.1 may be derived, with bound degrading like $\epsilon^{2k\epsilon}$ for a size-k group.
- Relaxations of DP/DP semantics that allow for "invariants" (statistics with no noise infused, but which are not regarded as public/common knowledge) are hard, & in their infancy
  - Blowfish privacy [3] provides one important approach to dealing with this issue
  - Invariants restrict the class of secrets that can receive any worst-case inferential protection, but many choices must be made in quantifying their impact on the inferences for which worst-case protection is still possible
- Relaxations of pure DP exist that still provide general guarantees but weaken the set of information that is considered "secret" for respondents
  - Pufferfish provides one way of formalizing specific sets of secrets. Modeling a relaxed secret set is fundamental for data where single individuals can be very important
  - For example, annual revenue data on the software sector cannot be very good if the presence or absence of Microsoft, Google, or Apple is hard to infer. See [4] for an example of a concrete application where secrets were carefully relaxed in this sense

3.5 Related Reading

In addition to the references cited above, the interested reader is encouraged to consult the "non-technical primer" of Wood et al. [5]. The structure of this section of our tutorial was heavily influenced by Wood et al's presentation, although our examples and choice of emphasis often differ from theirs.

3.6 Questions?

References

[1] Shiva P Kasiviswanathan and Adam Smith. On the 'semantics' of differential privacy: A bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1), 2014.

[2] D. Kifer and A. Machanavajjhala. A rigorous and customizable framework for privacy. *PODS '12 Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems*, 2012.

[3] Machanavajjhala A. He, X. and B. Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. *SIGMOD '14*, 2014.

[4] Machanavajjhala A. Abowd J. Graham M. Kutzbach M. Haney, S and L. Vilhuber. Utility cost of formal privacy for releasing national employer-employee statistics. *SIGMOD '17*, 2017.

[5] Altman M. Bembenek A. Bun M. Gaboardi M. Honaker J. Nissim K. O'Brien D. Stienke T. Wood, A. and S. Vadhan. Differential privacy: A primer for a non-technical audience. *Vanderbilt Journal of Entertainment & Technology Law*, 21(1), 2018.

4 How Do We Analyze Data Once Differentially Private Mechanisms Have Been Applied

One of the most attractive properties of DP disclosure-avoidance mechanisms for data users is that, unlike traditional disclosure-avoidance techniques, the privacy guarantees associated with DP mechanisms do not depend on secrecy. Our earlier discussion of Theorem 3.1 illustrates why this is the case: DP's worst-case semantic guarantees are generally developed by assuming that an attacker can exactly Bayes update after seeing the mechanism's output, using full knowledge of the DP mechanism's implementation.

As a result of this approach, both the mathematical description and code underlying a DP disclosure-avoidance mechanism can be shared with the user community. Only random seeds and the true values of the sensitive data must be kept secret. Use of DP methods therefore make available to user communities an option that could only be handled with guesswork previously: the noise distributions used in DP mechanisms can be exactly, transparently studied and discussed. In particular, the noise introduced by DP distributions can be adjusted for in statistical inference and estimation, just as traditional statistical analysis adjusts for other sources of random error, such as sampling procedures.

4.1 MSE & Point Estimates for Basic Mechanisms

The Laplace and Geometric mechanisms are often used as building blocks in more complex mechanisms, and are therefore important to understand when performing statistical inference. They are also probabilistically straightforward: both mechanisms yield simple unbiased estimators, they both have simple, closed-form descriptions for their primary moments, and for both there are a number of theorems describing their error properties under a variety of norms.

Despite their many virtues, the basic Laplace and Geometric mechanisms can often be improved upon when performing statistical estimation or inference, both when they are used in isolation and when they are used as part of a more complex mechanism. We'll first consider the basic Laplace and Geometric mechanisms. Two situations are especially common:

- The output is inconsistent with allowable query values (e.g. negative output for a counting query)
- The vector-output has length > 1, and its components can be combined to form a more accurate answer

Example 4.1. *Let $f(x) = Ax$ be a query with $A = \begin{bmatrix} I \\ B \end{bmatrix}$ and $B = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$, where our database is as in Example 2.1 (Sex by marital status).*

9. What is the (add-delete) sensitivity of the query?

Solution: The sensitivity is 3. The maximum of the column sums of A is 3 (4th column).

10. Assume that that we used the geometric mechanism with $\epsilon = 1$ and output the following differntially private vector $\begin{pmatrix} 3 & 0 & -1 & 4 & 2 & 4 & 2 & 7 \end{pmatrix}^T$. How might we use all the information to estimate "How many married people are there" (the first row of B)?

Solution: The variance of the noise infused by the Geometric mechanism into the estimate corresponding to the first row of B is $\frac{1-p}{p^2}, p = 1 - exp(-1/3)$. However, we can also get an estimate of this same query's by combining the outputs corresponding to the first and fourth rows of the identity matrix, and so we have several distinct estimators for the the number of married persons in the database. The Geometric mechanism infuses independent noise into each of the 8 components of the output vector, so we can minimize mean-squared error by taking an average of the several estimates for the query, weighting according to each estimate's inverse-variance.

```
#variance of single component
p = 1- exp(-1/3)
var_1 = (1-p)/p^2
var_1
inv_var1 = 1/var_1

#The variance of the sum of the first and fourth row of the identity matrix
var_2 = (1-p)/p^2 + (1-p)/p^2
var_2
inv_var2 = 1/var_2

#combined estimate
inv_var1/ (inv_var1+ inv_var2) * 2 + inv_var2/ (inv_var1+ inv_var2) * (3+4)

#The variance of the combined estimate is
var_new = (inv_var1/ (inv_var1+ inv_var2))^2 * var_1 + (inv_var2/ (inv_var1+
    inv_var2))^2*var_2
var_new
```

4.2 Uncertainty Due to the Differentially Private Mechanism

In this section we will think about working with the output of the differentially private mechanism. To begin with, we'll only focus on the uncertainty due to the randomness in the differentially private mechanism. Let's start off with the Laplace mechanism:

Theorem 4.1. *For any $\beta \in (0,1]$ and $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$, let $y = M(x)$ be the output of the ϵ differentially private Laplace mechanism. Then*

$$Pr\left[||f(x) - z||_\infty \geq \frac{\Delta f}{\epsilon} \ln\left(\frac{k}{\beta}\right)\right] \leq \beta$$

Proof. See [4]. □

For example, with $\beta = .05$, Theorem 4.1 says that a maximum component-wise absolute difference between $f(x)$ and y greater than $\frac{\Delta f}{\epsilon} \ln\left(\frac{k}{.05}\right)$ can only occur with probability less than .05. Recall that k is the size of the output vector. One can be 95% confident that the maximum componentwise absolute difference is less than $\frac{\Delta f}{\epsilon} \ln\left(\frac{k}{.05}\right)$. We can use this result to make simultaneous confidence intervals or a confidence region.

Corollary 4.1.1. *For any $\beta \in (0,1]$ and $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$, let $y = M(x)$ be the output of the ϵ differentially private Laplace mechanism.*

$$Pr\left[|f(x) - z| \geq \frac{\Delta f}{\epsilon} \ln\left(\frac{1}{\beta}\right)\right] \leq \beta.$$

11. Consider the query f from Example 2.4 (not necessarily the same x). Suppose the Laplace Mechanism with $\epsilon = 1$ outputs $z = 4$. Use Corollary 4.1.1 to define a 90% confidence interval.

Solution: Let $\beta = .1$. Recall $\Delta f = 1$. Then $Pr[|f(x) - 4| \leq \ln(10)] \geq .9$. Hence, the 90% CI is $(4 - \ln(10), 4 + \ln(10))$ or $(1.7, 6.3)$.

12. What would the 90% CI be if $\epsilon = 10$ instead?

Solution: We have $Pr[|f(x) - 4| \leq \frac{1}{10}\ln(10)] \geq .9$ so the 90% CI is $(3.77, 4.23)$.

13. Consider the query f from Example 2.5 (not necessarily the same x). Suppose the Laplace Mechanism with $\epsilon = 1$ outputs $z = \begin{pmatrix} 5 \\ -3 \\ 1 \end{pmatrix}$. Apply Theorem 4.1. Does the accuracy guarantee differ using unbounded vs bounded differential privacy?

Solution: Let $\beta = .05$, $k = 3$. In this case, $\Delta f = 3$ regardless of which neighboring definition is used. However, in general, unbounded and bounded differentially privacy will give different results. We have that $\frac{\Delta f}{\epsilon}\ln\left(\frac{k}{\beta}\right) = \frac{3}{1}\ln(\frac{3}{.05}) \approx 12.28$ so $Pr[||f(x) - z||_\infty \leq 12.28] \geq .95$. Thus, we can be 95% confident that

$$\begin{pmatrix} -7.28 \\ -15.28 \\ -11.28 \end{pmatrix} << f(x) << \begin{pmatrix} 17.28 \\ 9.28 \\ 13.28 \end{pmatrix}$$

4.3 Expending Additional Privacy-Loss Budget for Estimates of Error

In the previous sections, we have considered one natural strategy adjusting to respect the uncertainty introduced by DP mechanisms: directly studying the error distribution of the DP mechanism. However, sometimes an alternative option is available. If the data curator is willing to expend privacy-loss budget, they can directly use a simple DP mechanism to estimate, for example, the L1 error of a mechanism with respect to collected, sensitive sample data. This can be attractive when the variance or MSE of the mechanism does not have a closed form (due to post-processing or data-dependence), but a measure of uncertainty of the is needed.

Example 4.2. *We collect data on the number of Men and number of Women in a small town in Texas, and our sampled data set consists of these two counts: $x = (3, 5)$. To protect the privacy of our sample's respondents, we publish estimates of these counts using the Laplace mechanism M with $\epsilon = 1$. Assume the mechanism output is $(-3, 7)$. Because we know that we can't have a negative count, we use a post-processing technique to round negative numbers up to zero making our output vector $(0, 7)$ instead.*

14. Assume unbounded differential privacy. What is the variance of the mechanism's output including post-processing?

Solution: Without post-processing, the variance of each of the components is:

```
#The sensitivity is 1
laplace_var(sens=1, epsilon=1)
#2(1/0.1)^2
```

However we don't have a closed-form solution for the variance including the post-processing. That is because the variance is dependent on the true value of the true query.

```
#write a function that combines the Laplace mechanism
#with rounding up to zero
round_to_zero = function(query_answer=0, epsilon = 1, sens=1){
  x = laplace_mech(query_answer=query_answer, epsilon = epsilon, sens=sens)
  if(x <0) x =0
  return(x)}

#estimated variance with query answer 0
var( replicate(10000,round_to_zero(query_answer=0, epsilon = 1, sens=1)) )
#estimated variance with query answer 7
var( replicate(10000,round_to_zero(query_answer=7, epsilon = 1, sens=1)) )
#estimated variance with query answer 1000
var( replicate(10000,round_to_zero(query_answer=1000, epsilon = 1, sens=1)) )
```

Since we needed the true value of the query to estimate the variance, this would violate the rules of differential privacy. Instead, another option is to spend additional privacy-loss budget on error measures to enable us to get a better measure of uncertainty about our estimate. We'll again use the Laplace mechanism, but this time we'll estimate the signed L1 error given that our output vector is $(0, 7)$. For two scalar counts being compared, c, d, the signed L1 error is $c - d$.

15. Compute a DP estimate of the error for the the estimated count of men using additional privacy-loss budget $\epsilon = 1$. What is the sensitivity of the error measure?

Solution: The sensitivity is just 1 because in this case we treat our differentially private output 0 as fixed. We therefore can estimate it using the Laplace mechanism in our example by computing:

$$(3 - 0) + \text{Lap}\left(\frac{1}{1}\right)$$

```
set.seed(42)
error_est = laplace_mech(3, epsilon = 1, sens=1)
error_est
```

16. We now have an error estimate in hand. Use the error estimate to make a 90% confidence interval around for the estimated count of men. What is our total PLB expenditure now? Have our privacy guarantees degraded?

Solution: Applying Theorem 4.1 a 90% confidence interval around the error is

$$\widehat{error} \pm ln(10)$$

```
c(error_est - log(10), error_est + log(10))
```

Since the error itself is the difference between the true value and zero, this is actually a confidence interval for the estimated count of men as well.

Our privacy-loss budget expenditure is now 2 (using sequential composition). As a result, our privacy guarantees have degraded.

4.4 Hypothesis Testing

Hypothesis tests are an essential part of classical statistical inference, but if we ignore the fact that for example Laplace noise was added to sample counts, then the hypothesis test may not achieve its intended probability of a Type 1 error. Because we have the information about the differentially private mechanism parameters, we should be able to incorporate that information into our hypothesis tests, and ensure that it does not unexpectedly alter our error rate. The literature focused on modified hypothesis tests is young, but we'll look at an approach using the paper by Gaboardi et. al. [2] as motivation. They develop a method for performing differentailly private goodness-of-fit and independence tests using a chi-squared statistic.

In the context of differential privacy, we want our adjusted hypothesis tests to have significance level α, where α is the probability of a Type 1 error. Given this constraint, the added noise from the differentially private mechanism will cause the power of the test to decrease compared with utilizing the non-DP variant of this test.

Example 4.3. *Say that a random sample of 100 persons was taken from a particular population and each person's disease status for a particular disease is collected. Assume that the Laplace mechanism with* $\epsilon = 0.5$ *was applied to the count of persons with the disease in the sample and that the differentially private output was* $M(D) = 34.21$. *From previous studies the disease prevalence was found to be* 25.5% *in the general population. Test the hypothesis that* $H_o : p = .255$ *vs.* $H_A : p \neq .255$.

How might we go about incorporating the noise from the Laplace mechanism into the hypothesis test? In the Gaboardi et. al. paper [2], the authors utilize two main approaches:

- Derive the (asymptotic) distribution of the test statistic, including the impact of the differentially private mechanism
- Estimate the distribution of the test statistic under the null hypothesis using a Monte Carlo technique

Let's consider the second approach. The typical test statistic for this type of hypothesis would be the normal approximation to the binomial test with test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Without the Laplace noise, under the null hypothesis, the test statistic is approximately normally distributed. Let's use a Monte Carlo approach to estimate the distribution of the test statistic under the null hypothesis including the Laplace noise.

```
#We assume the data have a binomial distribution with p = 0.255

#mc draws
reps = 10000
#sample size
n=100
#null hypothesis
p0 = 0.255

#sample from binomial distribution with prob p0
x = rbinom(n=reps,size=n,prob=p0)
#apply the laplace mechanism to each count
x_dp = sapply(x, FUN= function(y) laplace_mech(y, epsilon=0.5, sens=1))
#as a proportion
phat = x_dp / n

#calculate the distribution of the z staistic
z_dist = (phat - p0)/(sqrt(p0*(1-p0)/n))

#what does the distribution look like?
hist(z_dist)

#what is the observed z statistic?
z_obs = (0.3421 - p0)/(sqrt(p0*(1-p0)/n))

#this will be right tailed
mean(z_obs >= z_dist)

#p-value
1- mean(z_obs >= abs(z_dist))

#how would this have compared if we had ignored the mechanism?
#what would have been the p-value?
2*(1-pnorm(z_obs))
```

```
#Let's take a look at the type 1 error of our method
test_error1 = function(n=100){

  #mc draws
  reps = 10000
  #null hypothesis
  p0 = 0.255

  phat_obs = laplace_mech(rbinom(n=1,size=n,prob=p0), epsilon=0.5, sens=1) / n

  #sample from binomial distribution with prob p0
  x = rbinom(n=reps,size=n,prob=p0)
  #apply the Laplace mechanism to each count
  x_dp = sapply(x, FUN= function(y) laplace_mech(y, epsilon=0.5, sens=1))
  #calculate the proportion
  phat = x_dp / n

  #calculate the distribution of the z statistic
  z_dist = (phat - p0)/(sqrt(p0*(1-p0)/n))

  #what is the observed z statistic?
  z_obs = (phat_obs - p0)/(sqrt(p0*(1-p0)/n))

  #p-value
  p_value=1- mean(abs(z_obs) >= abs(z_dist))
  return(p_value)
}

#Simulate it 1000 times
error_test = replicate(1000, test_error1(n=100))
#Using a significance level of 0.05 should imply we reject the null 5% of the time
mean(error_test <=0.05)
mean(error_test <=0.10)
```

This general strategy could be applied to a number of test statistics, but additional parameters might need to be estimated as well if they are not specified as part of the null hypothesis. See the following references for additional discussion of hypothesis tests [3, 4].

4.5 Questions

References

[1] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[2] Marco Gaboardi, Hyun-Woo Lim, Ryan M Rogers, and Salil P Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*. JMLR, 2016.

[3] Yue Wang, Jaewoo Lee, and Daniel Kifer. Differentially private hypothesis testing, revisited. *arXiv preprint arXiv:1511.03376*, 1, 2015.

[4] Ryan Rogers and Daniel Kifer. A new class of private chi-square hypothesis tests. In *Artificial Intelligence and Statistics*, pages 991–1000, 2017.

5 Extensions

5.1 Combined Sample-DP Uncertainty

The development of adjusted measures of uncertainty and modified hypothesis tests that incorporate both noise due to sampling and noise due to use of a DP statistical disclosure control mechanism is in its infancy. Of the works we covered in the last section, most do not do this (with the notable exception of the hypothesis test we examined), and instead opt to just measure the error introduced by the DP mechanism relative to the realized sample. However, some preliminary works have examined this problem, both asymptotically or in terms of results for finite sample sizes.

Karwa and Vadhan [1] present an important example of this developing literature: they study the derivation of differentially private conservative confidence intervals for a normally distributed mean (reflecting, e.g., sample variation). They examine both the known and unknown variance cases. In both, they rely fundamentally on “clipping” at least some unknown parameters to an interval specified *a priori* (i.e., without looking at the data). Karwa & Vadhan provide algorithms for generating CIs in both the known and unknown variance cases (Algorithms 4 and 5 in [1]).

5.2 Secrecy of the Sample

Suppose a differentially private mechanism is run on a sample of a database rather than the full database. Is the privacy guarantee improved? Intuitively, taking a sample increases privacy since a person’s data is only used with some probability. This can be interpreted in the context of a probability sample survey as well; that is, even if the sample is just performed in the usual sense from some population, where the population’s data may not be formally curated in any database, then “Secrecy of the Sample” can still be invoked, so long as the list of persons sampled from the population is kept secret.

Smith [2] and Li et al. [3] both consider this problem and show that sampling amplifies the privacy guarantee. Li et al. show that under Poisson sampling with sampling rate β, applying an ϵ-differentially private mechanism with privacy-loss budget ϵ_1 to the resulting sample leads to a differentially privacy-loss guarantee with

$$\epsilon_2 = ln(1 + \beta(e^{\epsilon} - 1)).$$

For example, with $\beta = 0.05$ and $\epsilon_1 = 1$, the effective privacy-loss guarantee has $\epsilon_2 = ln(1 + 0.05(e^1 - 1)) = 0.0824$. As a result, with sampling we need to add less noise to get the same privacy guarantee. Additional work is needed to consider more advanced sampling techniques.

5.3 Variants of Differential Privacy

There are several variants of differential privacy in the literature. Each of the following relaxes the privacy guarantee in favor of improved accuracy. Each redefines how the closeness of output distributions is measured.

- Approximate differential privacy or (ϵ, δ)-differential privacy differs from pure differential privacy by allowing for a δ chance that the differential privacy bound fails . See [4]
- Another pair of alternatives are concentrated differential privacy and zero-concentrated differential privacy both of which require the privacy loss to have a small mean and to besubguassian. See [5] and [6].
- Rényi Differential privacy utilizes the Rényi Divergence to measure the closeness of output distributions. See [7].

References

[1] V. Karwa and S. Vadhan. Finite sample differentially private confidence intervals. *arXiv*, 2017.

[2] Adam Smith. Differential privacy and the secrecy of the sample, Sep 2009.

[3] Ninghui Li, Wahbeh Qardaji, and Dong Su. On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, pages 32–33. ACM, 2012.

[4] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[5] Cynthia Dwork and Guy N. Rothblum. Concentrated differential privacy. *CoRR*, abs/1603.01887, 2016.

[6] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985*, pages 635–658, New York, NY, USA, 2016. Springer-Verlag New York, Inc.

[7] Ilya Mironov. Rényi differential privacy. *CoRR*, abs/1702.07476, 2017.