Differential Privacy: A Primer for a Non-technical Audience*

Kobbi Nissim†[1], Thomas Steinke[2], Alexandra Wood[3], Micah Altman[5], Aaron Bembenek[6], Mark Bun[2], Marco Gaboardi[4], David R. O'Brien[3], and Salil Vadhan[2]

[1]Department of Computer Science, Georgetown University.
kobbi.nissim@georgetown.edu.
[2]Center for Research on Computation and Society, Harvard University.
{tsteinke|mbun|salil}@seas.harvard.edu.
[3]Berkman Klein Center for Internet & Society, Harvard University.
{awood|dobrien}@cyber.law.harvard.edu.
[4]State University of New York at Buffalo.
gaboardi@buffalo.edu.
[5]Program on Information Science. Massachusetts Institute of Technology.
escience@mit.edu.
[6]School of Engineering and Applied Sciences. Harvard University.
bembenek@g.harvard.edu.

February 14, 2018

Keywords: differential privacy, data privacy, social science research

*This document is the product of a working group of the *Privacy Tools for Sharing Research Data* project at Harvard University (http://privacytools.seas.harvard.edu). The working group discussions were led by Kobbi Nissim. Kobbi Nissim, Thomas Steinke, and Alexandra Wood were the lead authors of this document. Working group members Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, Kobbi Nissim, David R. O'Brien, Thomas Steinke, Salil Vadhan, and Alexandra Wood contributed to the conception of the document and to the writing. We thank Scott Bradner, Cynthia Dwork, Caper Gooden, James Honaker, Deborah Hurley, Rachel Kalmar, Georgios Kellaris, Daniel Muise, and Michel Reymond for their many valuable comments on earlier versions of this document. A preliminary version of this work was presented at the 9th Annual Privacy Law Scholars Conference (PLSC 2017), and the authors thank the participants for contributing thoughtful feedback. This material is based upon work supported by the National Science Foundation under Grant No. 1237235, as well as by the Alfred P. Sloan Foundation.

†Work towards this document was completed while the author was visiting the Center for Research on Computation and Society at Harvard University.

Executive Summary

Differential privacy is a strong, mathematical definition of privacy in the context of statistical and machine learning analysis. It is used to enable the collection, analysis, and sharing of a broad range of statistical estimates, such as averages, contingency tables, and synthetic data, based on personal data while protecting the privacy of the individuals in the data.

Differential privacy is not a single tool, but rather a criterion, which many tools for analyzing sensitive personal information have been devised to satisfy. It provides a mathematically provable guarantee of privacy protection against a wide range of *privacy attacks*, i.e., attempts to learn private information specific to individuals from a data release. Privacy attacks include re-identification, record linkage, and differencing attacks, but may also include other attacks currently unknown or unforeseen. These concerns are separate from *security attacks*, which are characterized by attempts to exploit vulnerabilities in order to gain unauthorized access to a system.

Computer scientists have developed a robust theory for differential privacy over the last fifteen years, and major commercial and government implementations have now started to emerge.

The differential privacy guarantee (§ 3). Differential privacy mathematically guarantees that anyone seeing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether or not that individual's private information is included in the input to the analysis.

The privacy loss parameter (§ 4.2). What can be learned about an individual as a result of her private information being included in a differentially private analysis is limited and quantified by a privacy loss parameter, usually denoted *epsilon* (ϵ). Privacy loss can grow as an individual's information is used in multiple analyses, but the increase is bounded as a function of ϵ and the number of analyses performed.

Interpreting the guarantee (§ 6.3). The differential privacy guarantee can be understood in reference to other privacy concepts:

- Differential privacy essentially protects an individual's information as if her information were not used in the analysis at all.
- Differential privacy essentially ensures that using an individual's data will not reveal any personally identifiable information that is specific to her. Here, *specific* refers to information that cannot be inferred unless the individual's information is used in the analysis.
- Differential privacy essentially masks the contribution of any single individual, making it impossible to infer any information specific to an individual, including whether the individual's information was used at all.

As these statements suggest, differential privacy is a new way of protecting privacy that is more quantifiable and comprehensive than the concepts of privacy that underlie many existing laws, policies, and practices around privacy and data protection. The differential privacy guarantee can be interpreted in reference to these other concepts, and can even accommodate variations in how they are defined across different laws. In many cases, data holders may use differential privacy to demonstrate that they have complied with legal and policy requirements for privacy protection.

Differentially private tools (§ 7). Differential privacy is currently in initial stages of implementation and use in various academic, industry, and government settings, and the number of practical tools providing this guarantee is continually growing. Multiple implementations of differential privacy have been deployed by corporations such as Google, Apple, and Uber, and federal agencies such as the U.S. Census Bureau. Additional differentially private tools are currently under development across industry and academia.

Some differentially private tools utilize an interactive mechanism, which enables users to submit queries about a dataset and receive corresponding differentially private results, such as custom-generated linear regressions. Others tools are non-interactive, enabling static data or data summaries, such as synthetic data or contingency tables, to be released and used.

In addition, some tools rely on a curator model, in which a database administrator has access to and uses private data to generate differentially private data summaries. Others rely on a local model, which does not require individuals to share their private data with a trusted third party. Instead, they answer questions about their data in a differentially private manner, and only these differentially private answers are ever shared.

Benefits of differential privacy (§ 8). Differential privacy is supported by a rich and rapidly advancing theory that enables one to reason with mathematical rigor about privacy risk. Adopting this formal approach to privacy yields a number of practical benefits for users:

- Systems that adhere to strong formal definitions like differential privacy provide protection that is robust to a wide range of potential *privacy attacks*, as defined above, including attacks that are unknown at the time of deployment. An analyst designing a differentially private data release need not anticipate particular types of privacy attacks, such as the likelihood that one could link particular fields with other data sources that may be available.

- Differential privacy provides provable privacy guarantees with respect to the cumulative risk from successive data releases and is the only existing approach to privacy that provides such a guarantee.

- Differentially private tools also have the benefit of transparency, as it is not necessary to maintain secrecy around a differentially private computation or its parameters. This feature distinguishes differentially private tools from traditional de-identification techniques which often require concealment of the extent to which the data have been transformed, thereby leaving data users with uncertainty regarding the accuracy of analyses on the data.

- Differentially private tools can be used to provide broad, public access to data or data summaries in a privacy-preserving way. They can enable wide access to data that cannot otherwise be shared due to privacy concerns, and do so with a guarantee of privacy protection that substantially increases the ability of the institution to protect the individuals in the data.

Differentially private tools can, therefore, help enable researchers, policymakers, and businesses to analyze and share sensitive data while providing strong guarantees of privacy to the individuals in the data.

Contents

1 Introduction

Businesses, government agencies, and research institutions often use and share data containing sensitive or confidential information about individuals. Improper disclosure of such data can have adverse consequences for a data subject's relationships, reputation, employability, insurability, or financial status, or even lead to civil liability, criminal penalties, or bodily harm. Due to these and related concerns, a large body of laws, regulations, ethical codes, institutional policies, contracts, and best practices has emerged to address potential privacy-related harms associated with the collection, use, and release of personal information. In the following discussion, we discuss aspects of the broader data privacy landscape that motivated the development of formal privacy models like differential privacy.

1.1 Introduction to legal and ethical frameworks for data privacy

The legal framework for privacy protection in the United States has evolved as a patchwork of highly sector- and context-specific federal and state laws. Federal information privacy laws, for instance, have been enacted to protect certain categories of personal information found in health, education, financial, and government records, among others. State data protection and breach notification laws prescribe specific data security and breach reporting requirements when managing certain types of personal information. In addition, federal regulations generally require researchers conducting research involving human subjects to secure approval from an institutional review board and fulfill ethical obligations to the participants, such as disclosing the risks of participation, obtaining their informed consent, and implementing specific measures to protect privacy. It is also common for universities and other research institutions to adopt policies that require their faculty, staff, and students to abide by certain ethical and professional responsibility standards and set forth enforcement procedures and penalties for mishandling data.

Further restrictions apply when privacy-sensitive data are shared under the terms of a data sharing agreement, which will often strictly limit how the data can be used or redisclosed by the recipient. Organizations may also require privacy measures set forth by technical standards, such as those specifying information security controls to protect personally identifiable information. In addition, laws such as the EU General Data Protection Regulation are in place to protect personal data about European citizens regardless of where the data are held. International privacy guidelines, such as the privacy principles developed by the Organisation for Economic Co-operation and Development, have also been adopted by governments across the world. Moreover, the right to privacy is also protected by various international treaties and national constitutions.

Taken together, the safeguards required by these legal and ethical frameworks are designed to protect the privacy of individuals and ensure they fully understand the scope of personal information to be collected and the associated privacy risks. They also help data holders avoid administrative, civil, and criminal penalties, as well as maintain the public's trust and confidence in commercial, government, and research activities involving personal data.

1.2 Traditional statistical disclosure limitation techniques

A number of technical measures for disclosing data while protecting the privacy of individuals have been produced within the context of these legal and ethical frameworks. In particular, a collection of *statistical disclosure limitation (SDL)* techniques has been widely adopted by statistical agencies,

data analysts, and social science researchers to analyze and share data containing privacy-sensitive data with the aim of making it more difficult (or impossible) to learn personal information pertaining to an individual. This category of techniques encompasses a wide range of methods for suppressing, aggregating, perturbing, and generalizing attributes of individuals in the data. [1] Such techniques are often applied with the explicit goal of *de-identification*, i.e., of making it difficult to link an identified person to a record in a data release by redacting or coarsening data.

Over time, changes in the way information is collected and analyzed, including advances in analytical capabilities, increases in computational power, and the expanding availability of personal data from a wide range of sources, are eroding the effectiveness of traditional SDL techniques. Since the 1990s, and with increasing frequency, privacy and security researchers have demonstrated that data that have been de-identified can often be successfully *re*-identified via record linkage. Re-identification via record linkage, or a *linkage attack*, refers to the re-identification of one or more records in a de-identified dataset by uniquely linking a record in a de-identified dataset with identified records in a publicly available dataset, such as a voter registration list. In the late 1990s, Latanya Sweeney famously applied such an attack on a dataset containing de-identified hospital records. Sweeney observed that records in the de-identified dataset contained the birth date, sex, and ZIP code of patients, that many of the patients had a unique combination of these three attributes, and that these three attributes were listed alongside individuals' names and addresses in publicly-available voting records. Sweeney was able to use this information to re-identify records in the de-identified dataset.[2] Subsequent attacks on protected data have demonstrated weaknesses in other traditional approaches to privacy protection, and understanding the limits of these traditional techniques is the subject of ongoing research.[3]

1.3 The emergence of formal privacy models

Re-identification attacks are becoming increasingly sophisticated over time, as are other types of attacks that seek to infer characteristics of individuals based on information about them in the data. Successful attacks on de-identified data have shown that traditional technical measures for privacy protection may be particularly vulnerable to attacks devised after a technique's deployment and use. Some de-identification techniques, for example, require the specification of attributes in the data as identifying (e.g., names, dates of birth, or addresses) or non-identifying (e.g., movie ratings or hospital admission dates). Data providers may later discover that attributes initially believed to be non-identifying can in fact be used to re-identify individuals. Similarly, de-identification procedures may require a careful analysis of present and future data sources that could potentially be linked with the de-identified data and enable re-identification of the data, but unanticipated sources of auxiliary information that can be used for re-identification may become available in the

[1] For an overview of traditional SDL techniques, see Federal Committee on Statistical Methodology, Report on Statistical Disclosure Limitation Methodology, Statistical Policy Working Paper 22 (2005), https://fcsm.sites.usa.gov/files/2014/04/spwp22.pdf.

[2] *See* Recommendations to Identify and Combat Privacy Problems in the Commonwealth: Hearing on H.R. 351 Before the House Select Committee on Information Security, 189th Sess. (Pa. 2005) (statement of Latanya Sweeney, Associate Professor, Carnegie Mellon University), http://dataprivacylab.org/dataprivacy/talks/Flick-05-10.html.

[3] *See, e.g.,* Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, Proceedings of the 2008 IEEE Symposium on Research in Security and Privacy 111 (2008); Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, 3 Nature Sci. Rep. 1376 (2013); Joseph A. Calandrino, *"You Might Also Like:" Privacy Risks of Collaborative Filtering*, Proceedings of the 2011 IEEE Symposium on Security and Privacy 231 (2011).

future.

Issues such as these have underscored the need for privacy technologies that are immune not only to linkage attacks, but to any potential attack, *including attacks that are currently unknown or unforeseen*. They have also demonstrated that privacy technologies must provide meaningful privacy protection in settings where extensive external information may be available to potential attackers, such as employers, insurance companies, relatives, and friends of an individual in the data. In addition, real-world attacks have illustrated that ex-post remedies, such as simply "taking the data back" when a vulnerability is discovered, are ineffective because many copies of a set of data typically exist; copies may even persist online indefinitely.[4]

In response to the accumulated evidence of weaknesses with respect to traditional approaches, a new privacy paradigm has emerged from the computer science literature: **differential privacy**. Differential privacy is primarily studied in the context of the collection, analysis, and release of aggregate statistics, from simple statistical estimations, such as averages, to machine learning. First presented in 2006,[5] differential privacy is the subject of ongoing research to develop privacy technologies that provide robust protection against a wide range of potential attacks, including types of attacks currently unforeseen. Importantly, differential privacy is not a single tool but a *definition* or *standard* for quantifying and managing privacy risks for which many technological tools have been devised. Analyses performed with differential privacy differ from standard statistical analyses, such as the calculation of averages, medians, and linear regression equations, in that random noise is added in the computation. Tools for differentially private analysis are now in early stages of implementation and use across a variety of academic, industry, and government settings.

In the following sections, we provide a simplified and informal, but mathematically accurate, description of differential privacy. Using intuitive illustrations and limited mathematical formalism, we discuss the definition of differential privacy, how it addresses privacy risks, how differentially private analyses are constructed, and how such analyses can be used in practice. This discussion is intended to help non-technical audiences understand the guarantees provided by differential privacy. It can help guide practitioners as they make decisions regarding whether to use differential privacy and, if so, what types of promises they should make to data subjects about the guarantees differential privacy provides. In addition, these illustrations are intended to help legal scholars and policymakers consider how current and future legal frameworks and instruments should apply to tools based on formal privacy models such as differential privacy.

2 Privacy: A property of the analysis—not its outcome

This document seeks to explain how data containing personal information can be shared in a form that ensures the privacy of the individuals in the data will be protected. This question is motivated by real-world examples of data releases that were thought to be sufficiently protective of privacy but were later shown to carry significant privacy risks.

[4] As an example, in 2006 AOL published anonymized search history of 650,000 users over a period of three months. Shortly after the release, the New York Times identified a person in the release and AOL removed the data AOL from their site. However, in spite of its withdrawal by AOL, copies of the data are still accessible on the Internet today.

[5] Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, Proceedings of the Third Theory of Cryptography Conference 265 (2006), http://dx.doi.org/10.1007/11681878_14.

We begin this discussion with a cautionary tale about the re-identification of anonymized records released by the Massachusetts Group Insurance Commission.

> In the late 1990s, the Group Insurance Commission, an agency providing health insurance to Massachusetts state employees, allowed researchers to access anonymized records summarizing information about all hospital visits made by state employees. The agency anticipated that the analysis of these records would lead to recommendations for improving healthcare and controlling healthcare costs.
>
> Massachusetts Governor William Weld reassured the public that steps would be taken to protect the privacy of patients in the data. Before releasing the records to researchers, the agency removed names, addresses, Social Security numbers, and other pieces of information that could be used to identify individuals in the records.
>
> Viewing this as a challenge, Professor Latanya Sweeney, then a graduate student at MIT, set out to identify Gov. Weld's record in the dataset. She obtained demographic information about Gov. Weld, including his ZIP code and date of birth, by requesting a copy of voter registration records made available to the public for a small fee. Finding just one record in the anonymized medical claims dataset that matched Gov. Weld's gender, ZIP code, and date of birth enabled her to mail the Governor a copy of his personal medical records.

This case illustrates that, although a dataset may appear to be anonymous, it could nevertheless be used to learn sensitive information about individuals. Following Prof. Sweeney's famous demonstration, a long series of attacks have been carried out against different types of data releases anonymized using a wide range of techniques. These attacks have shown that risks remain even if additional pieces of information, such as those that were leveraged in Prof. Sweeney's attack (gender, date of birth, and ZIP code), are removed from a dataset prior to release. Risks also remain when using some frameworks for protecting privacy, such as k-anonymity, which is satisfied for a dataset in which the identifying attributes that appear for each person are identical to those of at least $k-1$ other individuals in the dataset.[6] Research has continually demonstrated that privacy measures that treat privacy as a property of the output, such as k-anonymity, will fail to protect privacy.

We offer a brief note on terminology before we proceed. Throughout this primer, we will use the terms *analysis* and *computation* interchangeably to refer to any transformation, usually performed by a computer program, of input data into some output.

As an example, consider an analysis on data containing personal information about individuals. The analysis may be as simple as determining the average age of the individuals in the data, or it may be more complex and utilize sophisticated modeling and inference techniques. In any case, the analysis involves performing a computation on input data and outputting the result. This notion of an analysis is illustrated in Figure 1.

[6] See, e.g., Ashwin Machanavajjhala et al., "ℓ-Diversity: Privacy Beyond k-Anonymity," *Proceedings of the 22nd International Conference on Data Engineering* (2006) ("In this paper we show with two simple attacks that a k-anonymized dataset has some subtle, but severe privacy problems.").

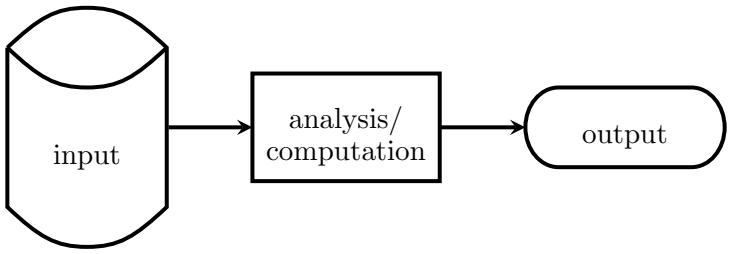

Figure 1: An analysis (or computation) transforms input data into some output.

This primer focuses, in particular, on analyses for transforming sensitive personal data into an output that can be released publicly. For example, an analysis may involve the application of techniques for aggregating or de-identifying a set of personal data in order to produce a sanitized version of the data that is safe to release. The data provider will want to ensure that publishing the output of this computation will not unintentionally leak information from the privacy-sensitive input data—but how?

As it turns out, one can be certain that the output of a computation is privacy-preserving if the computation itself is privacy-preserving. A key insight from the theoretical computer science literature is that *privacy is a property of the informational relationship between the input and output*, not a property of the output alone.[7] We illustrate why this is the case through a series of examples.

> Anne, a staff member at a high school, would like to include statistics about student performance in a presentation. She considers publishing the fact that a representative ninth-grade GPA is 3.5. Because the law protects certain student information held by educational institutions, she must ensure that the statistic will not inappropriately reveal student information, such as the GPA of any particular student.

One might naturally think that Anne could examine the statistic itself and determine that it is unlikely to reveal private information about an individual student. However, although the publication of this statistic might seem harmless, Anne needs to know *how* the statistic was computed to make that determination. For instance, if the representative ninth-grade GPA was calculated by taking the GPA of the alphabetically-first student in the school, then the statistic completely reveals the GPA of that student.[8]

[7] This insight follows from a series of papers demonstrating privacy breach enabled by leakages of information resulting from decisions made by the computation. *See, e.g.,* Krishnaram Kenthapadi, Nina Mishra, & Kobbi Nissim, *Denials Leak Information: Simulatable Auditing*, 79 Journal of Computer and System Sciences 1322 (2013), https://www.sciencedirect.com/science/article/pii/S002200001300113X. For a general discussion of the advantages of formal privacy models over ad-hoc privacy techniques, see Arvind Narayanan, Joanna Huey, & Edward W. Felten, *A Precautionary Approach to Big Data Privacy*, Working Paper (2015), http://randomwalker.info/publications/precautionary.pdf.

[8] One might object that the student's GPA is not traceable back to that student unless an observer knows how the statistic was produced. However, a basic principle of modern cryptography (known as Kerckhoffs' assumption) is that a system is not secure if its security depends on its inner workings being a secret. In our context, this means

Alternatively, Anne considers calculating a representative statistic based on average features of the ninth graders at the school. She takes the most common first name, the most common last name, the average age, and the average GPA for the ninth grade class. What she produces is "John Smith, a fourteen-year-old in the ninth grade, has a 3.1 GPA." Anne considers including this statistic and the method used to compute it in her presentation. In an unlikely turn of events, a new student named John Smith joins the class the following week.

This statistic does not reveal private information about John Smith because it is not based on his student records in any way. While Anne might decide not to use the statistic in her presentation because it is potentially confusing, using it would not reveal private information about John.

It may seem counter-intuitive that releasing a representative GPA violates privacy while releasing a GPA attached to a student's name would not. Yet these examples illustrate that the key to preserving privacy is the informational relationship between the private input and the public output, and not the output itself. Furthermore, not only is it necessary to examine the analysis itself to determine whether a statistic can be published, but it is also sufficient. In other words, we do not need to consider the output statistic at all, if we know whether the process used to generate that statistic preserves privacy.

3 What is the differential privacy guarantee?

The previous section illustrated why privacy should be thought of as a property of a computation—but how does one know whether a particular computation has this property?

Intuitively, a computation protects the privacy of individuals in the data if its output does not reveal any information that is specific to any individual data subject. Differential privacy formalizes this intuition as a *mathematical definition*. Just as we can show that an integer is even by demonstrating that it is divisible by two, we can show that a computation is differentially private by proving it meets the constraints of the definition of differential privacy. In turn, if a computation can be proven to be differentially private, we can rest assured that using the computation will not unduly reveal information specific to a data subject.

To see how differential privacy formalizes this intuitive privacy requirement as a definition, consider the following scenario.

Researchers have selected a sample of individuals across the U.S. to participate in a survey exploring the relationship between socioeconomic status and health outcomes. The participants were asked to complete a questionnaire covering topics such as where they live, their finances, and their medical history.

One of the participants, John, is aware that individuals have been re-identified in previous releases of de-identified data and is concerned that personal information he provides about himself, such as his medical history or annual income, could one day be revealed in de-identified data released from this study. If leaked, this

that we assume that the algorithm behind a statistical analysis is public (or could potentially be public).

information could lead to an increase in his life insurance premium or an adverse decision for a future mortgage application.

Differential privacy can be used to address John's concerns. If the researchers promise they will only share survey data after processing it with a differentially private computation, John is guaranteed that any data they release will not disclose anything that is *specific to him*, even though he participated in the study.

To understand what this means, consider a thought experiment, which we illustrate in Figure 2 and refer to as *John's opt-out scenario*. In John's opt-out scenario, an analysis is performed using data about the individuals in the study, except that information about John is omitted. His privacy is protected in the sense that the outcome of the analysis *does not depend on his specific information*—because it was not used in the analysis at all.

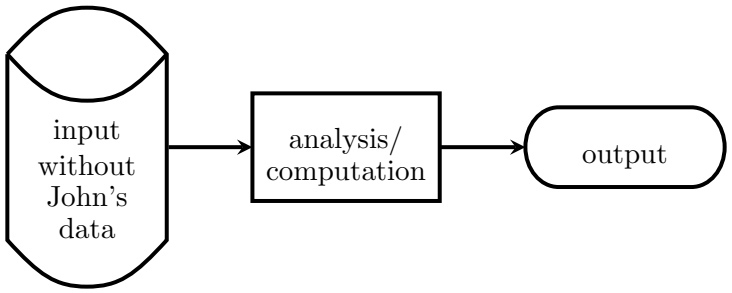

Figure 2: John's opt-out scenario.

John's opt-out scenario differs from the *real-world scenario* depicted in Figure 1, where the analysis is based on John's personal information along with the personal information of the other study participants. In contrast to his opt-out scenario, the real-world scenario involves some potential risk to John's privacy. Some of his personal information could be revealed by the outcome of the analysis because it was used as input to the computation.

3.1 What does differential privacy protect and what does it not protect?

Differential privacy aims to protect John's privacy in the real-world scenario in a way that mimics the privacy protection he is afforded in his opt-out scenario.[9] Accordingly, what can be learned about John from a differentially private computation is (essentially) limited to what could be learned about him from everyone else's data *without his own data being included in the computation*. Crucially, this same guarantee is made not only with respect to John, but also with respect to every other individual contributing his or her information to the analysis.

[9]It is important to note that the use of differentially private analyzes is *not* equivalent to the traditional use of opting out. On the privacy side, differential privacy does not require an explicit opt-out. In comparison, traditional use of opt-out requires an explicit choice that may cause privacy harms by calling attention to individuals that choose to opt out. On the utility side, there is no general expectation that using differential privacy would yield the same outcomes as adopting the policy of opt-out.

A precise description of the differential privacy guarantee requires the use of formal mathematical language, as well as technical concepts and reasoning that are beyond the scope of this document. In lieu of the mathematical definition, this document offers a few illustrative examples to discuss various aspects of differential privacy in a way we hope is intuitive and generally accessible.

Examples illustrating what differential privacy protects

The scenarios in this section illustrate the types of information disclosures that are addressed when using differential privacy.

> Alice and Bob are professors at Private University. They both have access to a database that contains personal information about students at the university, including information related to the financial aid each student receives. Because it contains personal information, access to the database is restricted. To gain access, Alice and Bob were required to demonstrate that they planned to follow the university's protocols for handling personal data, by undergoing confidentiality training and signing data use agreements proscribing their use and disclosure of personal information obtained from the database.
>
> In March, Alice publishes an article based on the information in this database and writes that "the current freshman class at Private University is made up of $3,005$ students, 202 of whom are from families earning over $\$350,000$ per year." Alice reasons that, because she published an aggregate statistic taken over $3,005$ people, no individual's personal information will be exposed. The following month, Bob publishes a separate article containing these statistics: "201 families in Private University's freshman class of $3,004$ have household incomes exceeding $\$350,000$ per year." Neither Alice nor Bob is aware that they have both published similar information.
>
> A clever student Eve reads both of these articles and makes an observation. From the published information, Eve concludes that between March and April one freshman withdrew from Private University and that the student's parents earn over $\$350,000$ per year. Eve asks around and is able to determine that a student named John dropped out around the end of March. Eve then informs her classmates that John's parents probably earn over $\$350,000$ per year.
>
> John hears about this and is upset that his former classmates learned about his parents' financial status. He complains to the university and Alice and Bob are asked to explain. In their defense, both Alice and Bob argue that they published only information that had been aggregated over a large population and does not identify any individuals.

This story illustrates how, in combination, the results of multiple analyses using information about the same people may enable one to draw conclusions about individuals in the data. Alice and Bob each published information that, in isolation, seems innocuous. However, when combined, the information they published compromised John's privacy. This type of privacy breach is difficult for

Alice or Bob to prevent individually, as neither knows what information has already been revealed or will be revealed by others in future. This problem is referred to as the problem of *composition*.

Suppose, instead, that the institutional review board at Private University only allows researchers to access student records by submitting queries to a special data portal. This portal responds to every query with an answer produced by running a differentially private computation on the student records. As will be explained further below, differentially private computations introduce a carefully tuned amount of random noise to the statistics outputted. This means that the computation gives an approximate answer to every question asked through the data portal. We will see that the use of differential privacy prevents the privacy leak that occurred in the previous scenario.

> In March, Alice queries the data portal for the number of freshmen who come from families with a household income exceeding $350,000. The portal returns the noisy count of 204, leading Alice to write in her article that "the current freshman class at Private University is made up of 3,005 students, approximately 205 of whom are from families earning over $350,000 per year." In April, Bob asks the same question and gets the noisy count of 199 students. Bob publishes in his article that "approximately 200 families in Private University's freshman class of 3,004 have household incomes exceeding $350,000 per year." The publication of these noisy figures prevents Eve from concluding that one student, with a household income greater than $350,000, withdrew from the university in March. The risk that John's personal information could be uncovered based on these publications is thereby reduced.

This example hints at one of the most important properties of differential privacy: it is robust under composition. If multiple analyses are performed on data describing the same set of individuals, then, as long as each of the analyses satisfies differential privacy, it is guaranteed that all of the information released, when taken together, will still be differentially private. Notice how this scenario is markedly different from the previous hypothetical, in which Alice and Bob do not use differentially private analyses and inadvertently release two statistics that in combination lead to the full disclosure of John's personal information. The use of differential privacy rules out the possibility of such a complete breach of privacy. This is because differential privacy enables one to measure and bound the cumulative privacy risk from multiple analyses of information about the same individuals.

It is important to note, however, that *every* analysis, regardless whether it is differentially private or not, results in some leakage of information about the individuals whose information is being analyzed, and this leakage accumulates with each analysis. This is true for every release of data, including releases of aggregate statistics, as we describe in further detail in Sections 4.5 and 6.2 below. For this reason, there is a limit to how many analyses can be performed on a specific dataset while providing an acceptable guarantee of privacy. This is why it is critical to measure privacy loss and to understand quantitatively how risk accumulates across successive analyses.

Examples illustrating what differential privacy does not protect

Next, we provide examples that illustrate the types of information disclosures differential privacy does not aim to address.

> Suppose Ellen is a friend of John's and knows some of his habits, such as that he regularly consumes several glasses of red wine with dinner. Ellen later learns of a research study that found a positive correlation between drinking red wine and the likelihood of developing a certain type of cancer. She might therefore conclude, based on the results of this study and her prior knowledge of John's drinking habits, that he has a heightened risk of developing cancer.

It may seem at first that the publication of the results from the research study enabled a privacy breach by Ellen. After all, learning about the study's findings helped her infer new information about John that he himself may be unaware of, i.e., his elevated cancer risk. However, notice how Ellen would be able to infer this information about John even if John had not participated in the medical study—i.e., it is a risk that exists in both John's opt-out scenario and the real-world scenario. Risks of this nature apply to everyone, regardless of whether they shared personal data through the study or not.

Consider a second example:

> Ellen knows that her friend John is a public school teacher with five years of experience and that he is about to start a job in a new school district. She later comes across a local news article about a teachers union dispute, which includes salary figures for the public school teachers in John's new school district. Ellen is able to approximately determine John's salary at his new job, based on the district's average salary for a teacher with five years of experience.

Note that, as in the previous example, Ellen can determine information about John (i.e., his new salary) from the published information, even though the published information was not based on John's information. In both examples, John could be adversely affected by the discovery of the results of an analysis, even in his opt-out scenario. In both John's opt-out scenario and in a differentially private real-world scenario, it is therefore not guaranteed that *no* information about John can be revealed. The use of differential privacy only guarantees that (essentially) no information *specific to John* is revealed.

These examples suggest, more generally, that any useful analysis carries a risk of revealing some information about individuals. We argue, however, that such risks are largely unavoidable. In a world in which data about individuals are collected, analyzed, and published, John cannot expect better privacy protection than is offered by his opt-out scenario because he has no ability to prevent others from participating in a research study or a release of public records. Moreover, the types of information disclosures enabled in John's opt-out scenario often result in individual and societal benefits. For example, the discovery of a causal relationship between red wine consumption and elevated cancer risk can inform John about possible changes he could make in his habits that would likely have positive effects on his health. In addition, the publication of public school teacher salaries may be seen as playing a critical role in transparency and public policy, as it can help communities make informed decisions regarding appropriate salaries for their public employees.

4 How does differential privacy limit privacy loss?

In the previous section, we explained that the only things that can be learned about a data subject from a differentially private data release are essentially what could have been learned if the analysis had been performed without that individual's data.

How do differentially private analyses achieve this goal? And what do we mean by "essentially" when we say that the only things that can be learned about a data subject are *essentially* those things that could be learned without the data subject's information? We will see that the answers to these two questions are related. Differentially private analyses protect the privacy of individual data subjects by adding carefully-tuned random noise when producing statistics. Differentially private analyses are also allowed to leak *some* information specific to individual data subjects. A privacy parameter controls exactly how much information can be leaked and, relatedly, how much random noise is added during the differentially private computation.

4.1 Differential privacy and randomness

In the earlier example featuring Professors Alice and Bob at Private University, we saw that differentially private analyses add random noise to the statistics they produce. Intuitively, this noise masks the differences between the real-world computation and the opt-out scenario of each individual in the dataset. This means that the outcome of a differentially private analysis is not exact but an *approximation*. In addition, a differentially private analysis may, if performed twice on the same dataset, return different results. Because they intentionally add random noise, analyses performed with differential privacy differ from standard statistical analyses, such as the calculation of averages, medians, and linear regression equations.

> Consider a differentially private analysis that computes the number of students in a sample with a GPA of at least 3.0. Say that there are 10,000 students in the sample, and exactly 5,603 of them have a GPA of at least 3.0. An analysis that added no random noise would report that 5,603 students had a GPA of at least 3.0.
>
> However, a differentially private analysis adds random noise to protect the privacy of the data subjects. For instance, a differentially private analysis might report an answer of 5,521 when run on the student data; when run a second time on the same data, it might report an answer of 5,586.

Although a differentially private analysis might produce many different answers given the same dataset, it is often possible to calculate accuracy bounds for the analysis that tell us how much an output of the analysis is expected to differ from the noiseless answer. Section 6.1 discusses how the random noise introduced by a differentially private analysis affects statistical accuracy. Interested readers can find more information about the role randomness plays in the construction of differentially private analyses in Appendix A.1.

4.2 The privacy loss parameter

An essential component of a differentially private computation is the privacy loss parameter. This parameter determines how much noise is added to the computation. It can be thought of as a tuning knob for balancing privacy and accuracy. Each differentially private analysis can be tuned to provide more or less privacy (resulting in less or more accuracy, respectively) by changing the value of this parameter. The following discussion establishes an intuition for this parameter. It can be thought of as limiting how much a differentially private computation is allowed to deviate from the opt-out scenario of an individual in the data.

Consider the opt-out scenario for a certain computation, such as estimating the number of HIV-positive individuals in a surveyed population. Ideally, this estimate should remain exactly the same whether or not a single individual, such as John, is included in the survey. However, ensuring this property *exactly* would require the total exclusion of John's information from the analysis. It would also require the exclusion of Gertrude's and Peter's information, in order to provide privacy protection for them as well. If we continue with this line of argument, we come to the conclusion that the personal information of every single surveyed individual must be removed in order to satisfy that individual's opt-out scenario. Thus, the analysis cannot rely on any person's information, and is completely useless.

To avoid this dilemma, differential privacy requires only that the output of the analysis remain *approximately* the same, whether John participates in the survey or not. That is, differential privacy allows for a deviation between the output of the real-world analysis and that of each individual's opt-out scenario. A parameter quantifies and limits the extent of the deviation between the opt-out and real-world scenarios. As shown in Figure 3 below, this parameter is usually denoted by the Greek letter ϵ (epsilon) and referred to as the privacy parameter, or, more accurately, the privacy loss parameter. The parameter ϵ measures the effect of each individual's information on the output of the analysis. It can also be viewed as a measure of the additional privacy risk an individual could incur beyond the risk incurred in the opt-out scenario. Note that in Figure 3 we have replaced John with an arbitrary individual X to emphasize that the differential privacy guarantee is made simultaneously to *all* individuals in the sample, not just John.

Choosing a value for ϵ can be thought of as tuning the level of privacy protection required. This choice also affects the utility or accuracy that can be obtained from the analysis. A smaller value of ϵ results in a smaller deviation between the real-world analysis and each opt-out scenario, and is therefore associated with stronger privacy protection but less accuracy. For example, when ϵ is set to zero, the real-world differentially private analysis mimics the opt-out scenario of each individual perfectly. However, as we argued at the beginning of this section, an analysis that perfectly mimics the opt-out scenario of each individual would require ignoring all information from the input and accordingly could not provide any meaningful output. Yet when ϵ is set to a small number such as 0.1, the deviation between the real-world computation and each individual's opt-out scenario will be small, providing strong privacy protection while also enabling an analyst to derive useful statistics based on the data.

Although guidelines for choosing ϵ have not yet been developed, they are expected to emerge and evolve over time, as the expanded use of differential privacy in real-life applications will likely shed light on how to reach a reasonable compromise between privacy and accuracy. As a rule of thumb, however, ϵ should be thought of as a small number, between approximately 1/1000 and 1.

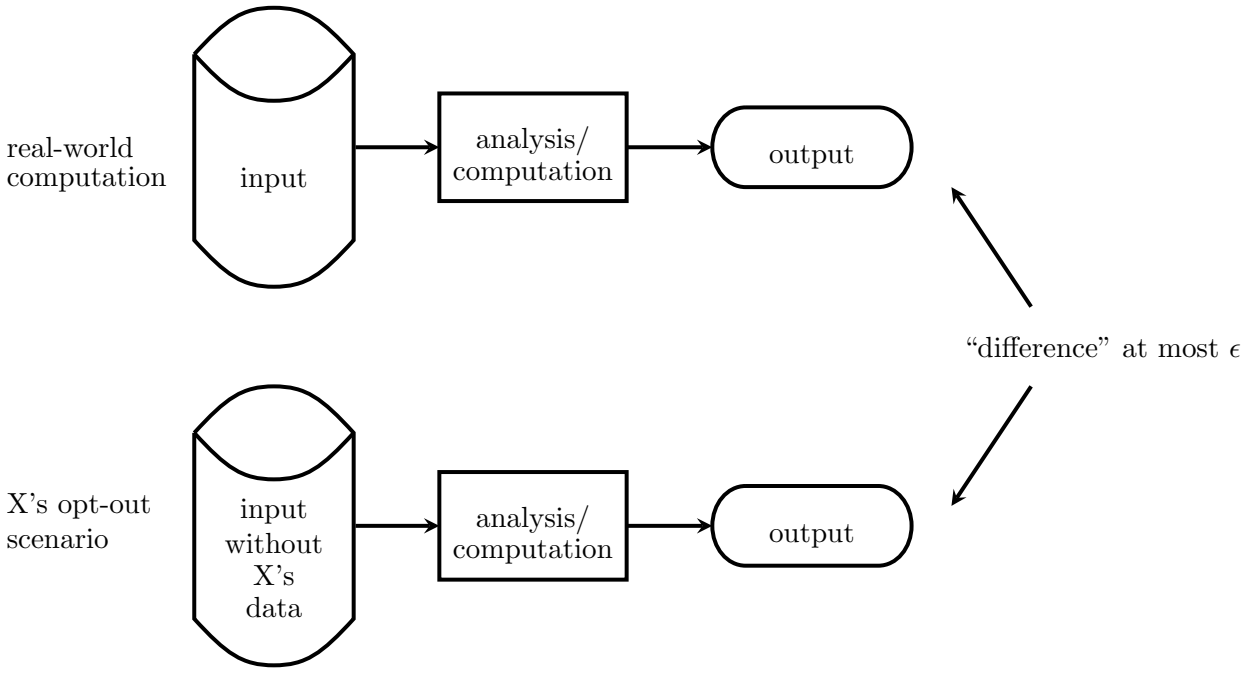

Figure 3: Differential privacy. The maximum deviation between the opt-out scenario and real-world computation should hold simultaneously for each individual X whose information is included in the input.

4.3 Bounding risk

We have discussed how the privacy loss parameter limits the deviation between the real-world computation and each data subject's opt-out scenario. However, it might not be clear how this abstract guarantee relates to privacy concerns in the real world. Therefore, in this section, we discuss a practical interpretation of the privacy loss parameter, as a bound on the financial risk incurred by participating in a study.

As explained above in Section 3.1, any useful analysis carries the risk that it will reveal information about individuals (which, in turn, might result in a financial cost). We will see in the following example that, while differential privacy necessarily cannot eliminate this risk, it can guarantee that the risk will be limited by quantitative bounds that depend on ϵ.

> Gertrude, a 65-year-old woman, is considering whether to participate in a medical research study. While she can envision many potential personal and societal benefits resulting in part from her participation in the study, she is concerned that the personal information she discloses over the course of the study could lead to an increase in her life insurance premium in the future.
>
> For example, Gertrude is concerned that the tests she would undergo as part of

the research study would reveal that she is predisposed to suffer a stroke and is significantly more likely to die in the coming year than the average person of her age and gender. If such information related to Gertrude's increased risk of morbidity and mortality is discovered by her life insurance company, it will likely increase her premium substantially.

Before she opts to participate in the study, Gertrude wishes to be assured that privacy measures are in place to ensure that her participation will have, at most, a limited effect on her life insurance premium.

4.3.1 A baseline: Gertrude's opt-out scenario

It is important to note that Gertrude's life insurance company may raise her premium based on something it learns from the medical research study, even if Gertrude does not herself participate in the study. The following example is provided to illustrate such a scenario.[10]

Gertrude holds a $\$100,000$ life insurance policy. Her life insurance company has set her annual premium at $\$1,000$, i.e., 1% of $\$100,000$, based on actuarial tables that show that someone of Gertrude's age and gender has a 1% chance of dying in the next year.

Suppose Gertrude opts out of participating in the medical research study. Regardless, the study reveals that coffee drinkers are more likely to suffer a stroke than non-coffee drinkers. Gertrude's life insurance company may update its assessment and conclude that, as a 65-year-old woman who drinks coffee, Gertrude has a 2% chance of dying in the next year. The company decides to increase Gertrude's annual premium from $\$1,000$ to $\$2,000$ based on the findings of the study.

In this example, the results of the study led to an increase in Gertrude's life insurance premium, even though she did not contribute any personal information to the study. A potential increase of this nature is unavoidable to Gertrude because she cannot prevent other people from participating in the study. Using the terminology of Section 3 above, this type of effect is taken into account by Gertrude's insurance premium in her *opt-out scenario*.

4.3.2 Reasoning about Gertrude's risk

Next, we consider the increase in risk that is due to Gertrude's participation in the study.

Suppose Gertrude decides to participate in the research study. Based on the results of medical tests performed on Gertrude over the course of the study, the researchers conclude that Gertrude has a 50% chance of dying from a stroke in the next year. If the data from the study were to be made available to Gertrude's insurance

[10] Figures in this example are based on data from Social Security Administration, Actuarial Life Table: Period Life Table, 2011, `http://www.ssa.gov/oact/STATS/table4c6.html`.

company, it might decide to increase her insurance premium from $\$2,000$ to more than $\$50,000$ in light of this discovery.

Fortunately for Gertrude, this does not happen. Rather than releasing the full dataset from the study, the researchers release only a differentially private summary of the data they collected. Differential privacy guarantees that, if the researchers use a value of $\epsilon = 0.01$, then the insurance company's estimate of the probability that Gertrude will die in the next year can increase from 2% to at most

$$2\% \cdot (1 + 0.01) = 2.02\%.$$

Thus Gertrude's insurance premium can increase from $\$2,000$ to, at most, $\$2,020$. Gertrude's first-year cost of participating in the research study, in terms of a potential increase in her insurance premium, is at most $20.

Note that this analysis above *does not* imply that the insurance company's estimate of the probability that Gertrude will die in the next year must increase as a result of her participation in the study, nor that if the estimate increases it must increase to 2.02%. What the analysis shows is that if the estimate were to increase it would not exceed 2.02%.

Consequently, this analysis *does not* imply that Gertrude would incur an increase in her insurance premium, or that if she were to see such an increase it would cost her $20. What is guaranteed is that, if Gertrude should see an increase in her premium, this increase would not exceed $20.

Gertrude may decide that the potential cost of participating in the research study, $20, is too high and she cannot afford to participate with this value of ϵ and this level of risk. Alternatively, she may decide that it is worthwhile. Perhaps she is paid more than $20 to participate in the study or the information she learns from the study is worth more than $20 to her. The key point is that differential privacy allows Gertrude to make a more informed decision based on the worst-case cost of her participation in the study.

It is worth noting that, should Gertrude decide to participate in the study, her risk might increase even if her insurance company is not aware of her participation. For instance, Gertrude might actually have a higher chance of dying in the next year, and that could affect the study results. In turn, her insurance company might decide to raise her premium because she fits the profile of the studied population, even if it does not believe her data were included in the study. On the other hand, differential privacy guarantees that, even if the insurance company knows that Gertrude *did* participate in the study, it can essentially only make inferences about her that it could have made if she had not participated in the study.

4.4 A general framework for reasoning about privacy risk

We can generalize from Gertrude's scenario and view differential privacy as a framework for reasoning about the increased risk that is incurred when an individual's information is included in a data analysis. Differential privacy guarantees that an individual will be exposed to essentially the same privacy risk whether or not her data are included in a differentially private analysis. In this

context, we think of the privacy risk associated with a data release as the potential harm that an individual might experience due to a belief that an observer forms based on that data release.

In particular, when ϵ is set to a small value, the probability that an observer will make some inference that is harmful to a data subject based on a differentially private data release is at most $1 + \epsilon$ times the probability that the observer would have made that inference without the data subject's inclusion in the data set.[11] For example, if ϵ is set to 0.01, then the privacy risk to an individual resulting from participation in a differentially private computation grows by at most a multiplicative factor of 1.01.

As shown in the Gertrude scenario, there is the risk to Gertrude that the insurance company will see the study results, update its beliefs about the mortality of Gertrude, and charge her a higher premium. If the insurance company infers from the study results that Gertrude has probablility p of dying in the next year, and her insurance policy is valued at $100,000$, this translates into a risk (in financial terms) of a higher premium of $p \times \$100,000$. This risk exists even if Gertrude does not participate in the study. Recall how in the first hypothetical, the insurance company's belief that Gertrude will die in the next year doubles from 1% to 2%, increasing her premium from $1,000 to $2,000, based on general information learned from the individuals who did participate. We also saw that, if Gertrude does decide to participate in the study (as in the second hypothetical), differential privacy limits the change in this risk relative to her opt-out scenario. In financial terms, her risk increases by at most $20, since the insurance company's beliefs about her probability of death change from 2% to at most $2\% \cdot (1 + \epsilon) = 2.02\%$, where $\epsilon = 0.01$.

Note that the above calculation requires certain information that may be difficult to determine in the real world. In particular, the 2% baseline in Gertrude's opt-out scenario (i.e., Gertrude's insurer's belief about her chance of dying in the next year) is dependent on the results from the medical research study, which Gertrude does not know at the time she makes her decision whether to participate. Fortunately, differential privacy provides guarantees relative to every baseline risk.

> Say that, without her participation, the study results would lead the insurance company to believe that Gertrude has a 3% chance of dying in the next year (instead of the 2% chance we hypothesized earlier). This means that Gertrude's insurance premium would increase to $3,000. Differential privacy guarantees that, if Gertrude had instead decided to participate in the study, the insurer's estimate for Gertrude's mortality would have been at most $3\% \cdot (1 + \epsilon) = 3.03\%$ (assuming an ϵ of 0.01), which means that her premium would not increase beyond $3,030.

Calculations like those used in the analysis of Gertrude's privacy risk can be performed by referring to Table 1. For example, the value of ϵ used in the research study Gertrude considered participating in was 0.01, and the baseline privacy risk in her opt-out scenario was 2%. As shown in Table 1, these values correspond to a worst-case privacy risk of 2.02% in her real-world scenario. Notice also how the calculation of risk would change with different values. For example, if the privacy risk in Gertrude's opt-out scenario were 5% rather than 2% and the value of epsilon remained the same, then the worst-case privacy risk in her real-world scenario would be 5.05%.

The fact that the differential privacy guarantee applies to *every* privacy risk means that Gertrude can know for certain how participating in the study might increase her risks relative to opting out, even if she does not know *a priori* all the privacy risks posed by the data release. This enables

[11] In general, the guarantee made by differential privacy is that the probabilities differ by at most a factor of $e^{\pm\epsilon}$, which is approximately $1 \pm \epsilon$ when ϵ is small.

posterior belief given $A(x')$ in %	value of ϵ					
	0.01	0.05	0.1	0.2	0.5	1
0	0	0	0	0	0	0
1	1.01	1.05	1.1	1.22	1.64	2.67
2	2.02	2.1	2.21	2.43	3.26	5.26
5	5.05	5.24	5.5	6.04	7.98	12.52
10	10.09	10.46	10.94	11.95	15.48	23.2
25	25.19	25.95	26.92	28.93	35.47	47.54
50	50.25	51.25	52.5	54.98	62.25	73.11
75	75.19	75.93	76.83	78.56	83.18	89.08
90	90.09	90.44	90.86	91.66	93.69	96.07
95	95.05	95.23	95.45	95.87	96.91	98.1
98	98.02	98.1	98.19	98.36	98.78	99.25
99	99.01	99.05	99.09	99.18	99.39	99.63
100	100	100	100	100	100	100
	maximum posterior belief given $A(x)$ in %					

Table 1: Maximal change between posterior beliefs in Gertrude's opt-out and real-world scenarios. The notation $A(x')$ refers to the application of the analysis A on the dataset x', which does not include Gertrude's information. As this table shows, the use of differential privacy provides a quantitative bound on how much one can learn about an individual from a computation.

Gertrude to make a more informed decision about whether to take part in the study. For instance, she can calculate exactly how much additional risk she incurs by participating in the study over a range of possible baseline risk values, and decide whether she is comfortable with taking on the risks entailed by these different scenarios.

4.5 Composition

Privacy risk accumulates with multiple analyses on an individual's data, and this is true whether or not any privacy-preserving technique is applied.[12] One of the most powerful features of differential privacy is its robustness under composition. We are able to reason about—and bound—the privacy risk that accumulates when multiple differentially private computations are performed on an individual's data.

The parameter ϵ quantifies how privacy risk accumulates across multiple differentially private analyses. In particular, say that two differentially private computations are performed on datasets about the same individuals. If the first computation uses a parameter of ϵ_1 and the second uses a parameter of ϵ_2, then the cumulative privacy risk resulting from these computations is no greater than the risk associated with an aggregate parameter of $\epsilon_1 + \epsilon_2$. In other words, the privacy risk from running the two analyses is bounded by the privacy risk from running a single differentially

[12] We emphasize that this observation is true for *any* use of information, and, hence, for any approach to preserving privacy. It is not unique to differentially private analyses. However, the fact that the cumulative privacy risk from multiple analyses can be bounded is a distinguishing property of differential privacy.

private analysis with a parameter of $\epsilon_1 + \epsilon_2$.

This means that, while it cannot get around the fundamental law that privacy risk increases when multiple analyses are performed on the same individual's data, differential privacy guarantees that privacy risk accumulates in a bounded and graceful way. For instance, two differentially private analyses cannot be combined in a way that leads to a privacy breach that is disproportionate to the privacy risk associated with each analysis in isolation. Differential privacy is currently the only framework with quantifiable guarantees on how risk accumulates across multiple analyses.

Suppose that Gertrude decides to opt into the medical study because it is about heart disease, an area of research she considers critically important. The study leads to a published research paper, which includes results from the study produced by a differentially private analysis with a parameter of $\epsilon_1 = 0.01$. A few months later, the researchers decide that they want to use the same study data for another paper. This second paper would explore a hypothesis about acid reflux disease, and would require calculating new statistics based on the original study data. Like the analysis results in the first paper, these statistics would be computed using differential privacy, but this time with a parameter of $\epsilon_2 = 0.02$.

Because she only consented for her data to be used in research about heart disease, the researchers must obtain Gertrude's permission to reuse her data for the paper on acid reflux disease. Gertrude is concerned that her insurance company could compare the results from both papers and learn something negative about Gertrude's life expectancy and drastically raise her insurance premium. She is not particularly interested in participating in a research study about acid reflux disease and is concerned the risks of participation might outweigh the benefits to her.

Because the statistics from each study are produced using differentially private analyses, Gertrude can precisely bound the privacy risk that would result from contributing her data to the second study. The combined analyses can be thought of a single analysis with a privacy loss parameter of

$$\epsilon_1 + \epsilon_2 = 0.01 + 0.02 = 0.03.$$

Say that, without her participation in either study, the insurance company would believe that Gertrude has a 3% chance of dying in the next year, leading to a premium of \$3,000. If Getrude participates in both studies, the insurance company's estimate of Gertrude's mortality would increase to at most

$$3\% \cdot (1 + 0.03) = 3.09\%.$$

This corresponds to a premium increase of \$90 over the premium that Gertrude would pay if she had not participated in either study.

5 What types of analyses are performed with differential privacy?

A large number of analyses can be performed with differential privacy guarantees. The following is a non-exhaustive list of analyses for which differentially private algorithms are known to exist:

- **Count queries:** The most basic statistical tool, a count query returns an estimate of the number of individual records in the data satisfying a specific predicate. For example, a count query could be used to return the number of records corresponding to HIV-positive individuals in a sample. Differentially private answers to count queries can be obtained through the addition of random noise, as demonstrated in the detailed example found above in Section A.1.

- **Histograms:** A histogram contains the counts of data points as they are classified into disjoint categories. For example, in the case of numerical data, a histogram shows how data are classified within a series of consecutive non-overlapping intervals. A **contingency table (or cross tabulation)** is a special form of a histogram representing the interrelation between two or more variables. The categories of a contingency table are defined as conjunctions of attribute variables. Differentially private histograms and contingency tables provide noisy counts for the data classified in each category.

- **Cumulative distribution function (CDF):** For data over an ordered domain, such as age (where the domain is integers, say, in the range $0 - 100$), or annual income (where the domain is real numbers, say, in the range $\$0.00 - \$1,000,000.00$), a cumulative distribution function depicts for every domain value x an estimate of the number of data points with a value up to x. A CDF can be used for computing the median of the data points (the value x for which half the data points have value up to x) and the interquartile range, among other statistics. A differentially private estimate of the CDF introduces noise that needs to be taken into account when the median or interquartile range is computed from the estimated CDF.[13]

- **Linear regression:** Social scientists are often interested in modeling how some dependent variable varies as a function of one or more explanatory variables. For instance, a researcher may seek to understand how a person's health depends on her education and income. In linear regression, an underlying linear model is assumed, and the goal of the computation is to fit a linear model to the data that minimizes a measure of "risk" (or "cost"), usually the sum of squared errors. Using linear regression, social scientists can learn to what extent a linear model explains their data, and which of the explanatory variables correlates best with the dependent variable. Differentially private implementations of linear regression introduce noise in its computation. Note that, while this noise may in some cases hide existing correlations in the data, researchers are engaged in ongoing work towards the development of algorithms where this undesirable effect of noise addition can be controlled and minimized.

- **Clustering:** Clustering is a data analysis technique that involves grouping data points into clusters, so that points in the same cluster are more similar to each other than to points in other clusters. Data scientists often use clustering as an exploratory tool to gain insight

[13] For a more in depth discussion of differential privacy and CDFs, see Daniel Muise and Kobbi Nissim, "Differential Privacy in CDFs," Slide Deck (2016), http://privacytools.seas.harvard.edu/files/dpcdf_user_manual_aug_2016.pdf.

into their data and identify the data's important sub-classes. Researchers are developing a variety of differentially private clustering algorithms, and such tools are likely to be included in future privacy-preserving tool kits for social scientists.

- **Classification:** In machine learning and statistics, classification is the problem of identifying which of a set of categories a data point belongs in, based on a training set of examples for which category membership is known. Data scientists often utilize data samples that are pre-classified (e.g., by experts) to train a classifier which can later be used for labeling newly-acquired data samples. Theoretical work has shown that it is possible to construct differentially private classification algorithms for a large collection of classification tasks, and, furthermore, that, at least in principle, the performance of these classification algorithms is comparable with the performance of similar non-privacy-preserving algorithms.

- **Synthetic data:** Synthetic data are "fake" data produced from a statistical model based on the original data. Synthetic data resemble the original sensitive data in format, and, for a large class of analyses, results are similar whether performed on the synthetic or original data. Theoretical work has shown that differentially private synthetic data can be generated for a large variety of tasks. A significant benefit is that, once a differentially private synthetic data set is generated, it can be analyzed any number of times, without any further implications for the privacy budget. As a result, synthetic data can be shared freely or even made public in many cases.

6 Practical considerations when using differential privacy

In this section, we discuss some of the practical challenges to using differential privacy, including challenges related to the accuracy of differentially private statistics, and challenges due to the degradation of privacy that results from multiple analyses. It is important to note that the challenges of producing accurate statistics while protecting privacy and addressing composition are not unique to differential privacy. It is a fundamental law of information that privacy risk grows with the use of data, and hence this risk applies to any disclosure control technique. Traditional statistical disclosure limitation techniques, such as suppression, aggregation, and generalization, often reduce accuracy and are vulnerable to loss in privacy due to composition, and the impression that these techniques do not suffer accumulated degradation in privacy is merely due to the fact that these techniques have not been analyzed with the higher level of rigor that differential privacy has been.[14] A rigorous analysis of the effect of composition is important for establishing a robust and realistic understanding of how multiple statistical computations affect privacy.

6.1 Accuracy

In this Section, we discuss the relationship between differential privacy and accuracy. The accuracy of an analysis is a measure of how its outcome can deviate from the true quantity or model it attempts to estimate. There is no single measure of accuracy, as measures of deviations differs

[14] For a discussion of privacy and utility with respect to traditional statistical disclosure limitation techniques, see Bee-Chung Chen, Daniel Kifer, Kristen LeFevre, and Ashwin Machanavajjhala, Privacy-Preserving Data Publishing, Foundations and Trends in Databases 2.1-2 (2009): 1-167.

across applications.[15] Multiple factors have an effect on the accuracy of an estimate, including, e.g., measurement and sampling errors. The random noise introduced in differentially private computations to hide the effect of single individuals on the computed estimate similarly affects accuracy.

Put differently, differential privacy increases the minimal sample size required to produce accurate statistics. In practice, the amount of noise that is added to differentially private analyses makes it difficult to obtain much utility from small- to moderately-sized datasets. As a rule of thumb, almost no utility is expected from datasets containing $1/\epsilon$ or fewer records.[16] Much of the ongoing research on differential privacy is focused on understanding and improving the tradeoff between privacy and utility, i.e., obtaining the maximum possible utility from data while preserving differential privacy.

Procedures for estimating the accuracy of certain types of analyses have been developed. These procedures take as input the number of records, a value for ϵ, and the ranges of numerical and categorical fields, among other parameters, and produce guaranteed accuracy bounds. Alternatively, a desired accuracy may be given as input instead of ϵ, and the computation results in a value for ϵ that would provide this level of accuracy. Figure 4 illustrates an example of a cumulative distribution function and the results of its (noisy) approximation with different settings of the privacy parameter ϵ.[17] Procedures for estimating the accuracy of an analysis are being developed for practical implementations of differential privacy, including the tools that are being developed for Harvard's Dataverse project, as discussed below.

We note that statistical analyses are in general estimates and not exact, and that in some cases the noise introduced by differential privacy is insignificant compared to other sources of error, such as sampling error. Interested readers can find a more detailed discussion on this topic in Appendix A.2.

Another concept related to accuracy is *truthfulness*. This term has appeared regularly, if infrequently, in the statistical disclosure control literature since the mid 1970s, though it does not have a well-recognized formal definition.[18] Roughly speaking, a privacy-protecting method is considered truthful if one can determine unambiguously which types of statements, when true of the protected data, are also necessarily true of the original data.

This concept may have an intuitive appeal. For data protected via suppressing some of the cells

[15] For example, a researcher interested in estimating the average income of a given population may care about the absolute error of this estimate, i.e., the difference between the real average and the estimate, whereas a researcher interested in the median income may care about the difference between the number of respondents whose income is below the estimate and the number of respondents whose income is above the estimate.

[16] An exception is when the amplification technique known as "secrecy of the sample" is used. See Section A.4 for a discussion on this topic.

[17] This figure first appeared in Daniel Muise and Kobbi Nissim, "Differential Privacy in CDFs," Slide Deck (2016), http://privacytools.seas.harvard.edu/files/dpcdf_user_manual_aug_2016.pdf.

[18] *See, e.g.,* Ferrer, Sanchez & Soria Comez, Database Anonymization 3.1 (Morgan & Claypool, eds., 2017) (distinguishing between "perturbative masking (which distorts the original data and leads to the publication of non-truthful data)" and "non-perturbative masking (which reduces the amount of information, either by suppressing some of the data or by reducing the level of detail, but preserves truthfulness)"); Fung, Chang, Wen & Yu, *Privacy Preserving Data Publication*, ACM Computing Surveys 42(14) (2010) (describing, without defining, truthfulness at the record level by explaining that "[i]n some data publishing scenarios, it is important that each published record corresponds to an existing individual in real life. . . . Randomized and synthetic data do not meet this requirement. Although an encrypted record corresponds to a real life patient, the encryption hides the semantics required for acting on the patient represented."

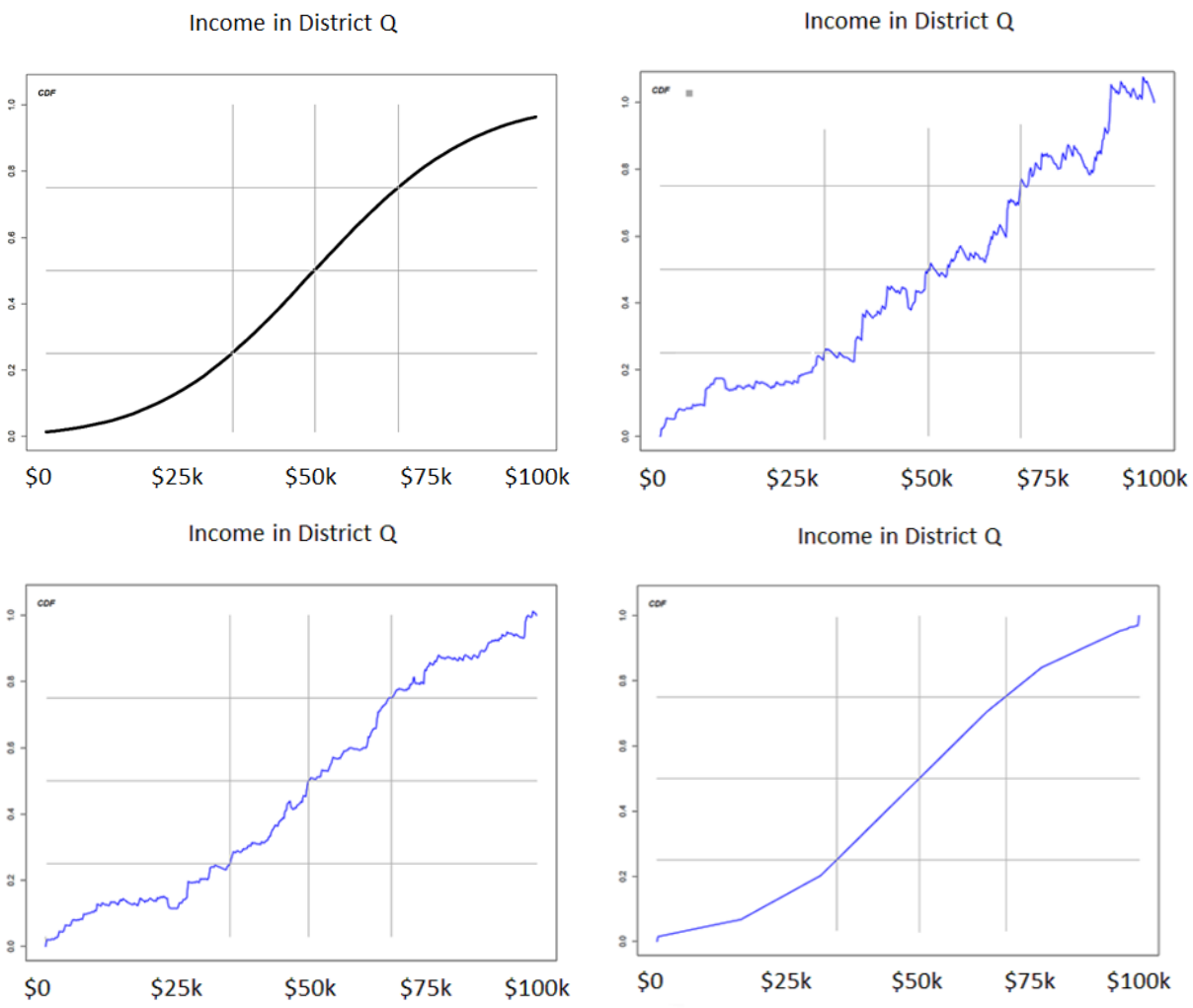

Figure 4: An example of the outcome of a differentially private computation of the cumulative distribution function (CDF) of income in District Q. The top left graph presents the original CDF (without noise) and the subsequent graphs show the result of applying differentially private computations of the CDF with ϵ values of 0.005 (top right), 0.01 (bottom left), and 0.1 (bottom right). Notice that, as smaller values of ϵ imply better privacy protection, they also imply less accuracy due to noise addition compared to larger values of ϵ.

in the database, statements of the form "there is (are) record(s) with characteristics X and Y" are true in the original data if they are true in the protected data. For example, one might definitively state, using only the protected data, that "some plumbers earn over $50,000." Arguably, one cannot make this same statement definitively and truthfully for data that has been synthetically generated.

One must be careful, however, to identify and communicate the types of true statements a protection method supports. For example, neither suppression nor synthetic data support truthful

non-existence claims at the microdata level. For example a statement such as "there are no plumbers in Wisconsin who earn over $50,000" cannot be made definitively by examining the protected data alone, if income or occupation values have been suppressed or synthetically generated. Moreover, protection methods may, in general, preserve truth at the individual record level, but not at the aggregate level (or vice versa). For example, synthetic data supports reliable truthful statements about many aggregate computations (e.g. "the median income of a plumber is $\$45,000 \pm \$2,000$"), but data subject to suppression generally cannot support truthful aggregate statements and protect privacy simultaneously.[19]

As mentioned above, there are many individual sources of error that contribute to the total uncertainty in a calculation. These are traditionally grouped by statisticians into into the categories of sampling and non-sampling errors. Generally, differentially private methods introduce more uncertainty. However, it is a property of differential privacy that the method itself does not need to be kept secret. This means the amount of noise added to the computation can be taken into account in the measure of accuracy, and therefore lead to truthful statements about the data. This is in contrast to many traditional statistical disclosure control methods which only report sampling error, and keep the information needed to estimate the "privacy error" secret.

To summarize, any privacy-preserving method, if misused or misinterpreted, can produce untruthful statements. In addition, the truthfulness of some methods, such as suppression and aggregation, is inherently limited to particular levels of computations, e.g., to existence statements on microdata, or statements about selected aggregate statistical properties, respectively. Differential privacy may be used truthfully for a broader set of computations, so long as the uncertainty of each calculation is estimated and reported.

6.2 The "privacy budget"

One can think of the parameter ϵ as determining the overall privacy protection provided by a differentially private analysis. Intuitively, ϵ determines "how much" of an individual's privacy an analysis may utilize, or, alternatively, by how much the risk to an individual's privacy can increase. A smaller value for ϵ implies better protection, i.e., less risk to privacy. Conversely, a larger value for ϵ implies worse protection, i.e., higher potential risk to privacy. In particular, $\epsilon = 0$ implies perfect privacy, i.e., the analysis does not increase any individual's privacy risk at all. Unfortunately, analyses that satisfy differential privacy with $\epsilon = 0$ must completely ignore their input data and therefore are useless.

We can also think of ϵ as a "privacy budget" to be spent by analyses of individuals' data. If a single analysis is expected to be performed on a given set of data, then one might allow this analysis to exhaust the entire privacy budget ϵ. However, a more typical scenario is that several analyses are expected to be run on a dataset, and therefore one needs to calculate the total utilization of the privacy budget by these analyses.

Fortunately, a number of composition theorems have been developed for differential privacy, as mentioned above in Section 4.5. In particular, these theorems state that the composition of two differentially private analyses results in a privacy loss that is bounded by the sum of the privacy losses of each of the analyses.

[19] Correctly calculating and truthfully reporting the uncertainty induced by suppression would require revealing the full details of the suppression algorithm and its parameterization. Revealing these details allows information to be inferred about individuals. Traditional statistical disclosure control methods require that the mechanism itself be kept secret in order to protect against this type of attack.

To understand how overall privacy loss is accounted for in this framework, consider the following example.

> Suppose a data analyst using a differentially private analysis tool is required to do so while maintaining differential privacy with an overall privacy loss parameter $\epsilon = 0.1$. This requirement for the overall privacy loss parameter may be guided by an interpretation of a regulatory standard, institutional policy, or best practice, among other possibilities. It means that all of the analyst's analyses, taken together, must have a value of ϵ that is at most 0.1.
>
> Consider how this requirement would play out within the following scenarios:
>
> **One-query scenario:** The data analyst performs a differentially private analysis with a privacy loss parameter $\epsilon_1 = 0.1$. In this case, the analyst would not be able to perform a second analysis over the data without risking a breach of the policy limiting the overall privacy loss to $\epsilon = 0.1$.
>
> **Multiple-query scenario:** The data analyst first performs a differentially private analysis with $\epsilon_1 = 0.01$, which falls below the limit of $\epsilon = 0.1$. This means that the analyst can also apply a second differentially private analysis, say with $\epsilon_2 = 0.02$. After the second analysis, the overall privacy loss amounts to
>
> $$\epsilon_1 + \epsilon_2 = 0.01 + 0.02 = 0.03,$$
>
> which is still less than $\epsilon = 0.1$, and hence allows the analyst to perform additional analyses before exhausting the budget.

The multiple-query scenario can be thought of as if the data analyst has a *privacy budget* of $\epsilon = 0.1$ that is consumed incrementally as she performs differentially private analyses, until the budget has been exhausted. Performing additional analyses after the overall budget has been exhausted may result in a privacy parameter that is larger (i.e., worse) than ϵ. Any further use would result in a privacy risk that is too significant.

Note that, in the sample calculation for the multiple-query example, we bounded the accumulated privacy risk simply by adding the privacy parameters of each analysis. It is in fact possible to obtain better bounds on the accumulation of the privacy loss parameter than suggested by this example. Various tools for calculating the bounds on the accumulated privacy risks in real-world settings using more sophisticated approaches are currently under development.

6.3 Complying with legal requirements for privacy protection

Statistical agencies, companies, researchers, and others who collect, process, analyze, store, or share data about individuals must take steps to protect the privacy of the data subjects in accordance with various laws, institutional policies, contracts, ethical codes, and best practices. In some settings, tools that satisfy differential privacy can be used to analyze and share data, while both complying with such legal obligations and providing strong mathematical guarantees of privacy protection for the individuals in the data.

Privacy regulations and related guidance do not directly answer the question of whether the use of differentially private tools is sufficient to satisfy existing regulatory requirements for protecting privacy when sharing statistics based on personal data. This issue is complex because privacy laws are often context-dependent and there are significant gaps between differential privacy and the concepts underlying regulatory approaches to privacy protection. Different regulatory requirements are applicable depending on the sector, jurisdiction, actors, and types of information involved. As a result, datasets held by an organization may be subject to different requirements. In some cases, similar or even identical datasets may be subject to different requirements when held by different organizations. In addition, many legal standards for privacy protection are, to a large extent, open to interpretation and therefore require a case-specific legal analysis by an attorney.

Other challenges arise as a result of differences between the concepts appearing in privacy regulations and those underlying differential privacy. For instance, many laws focus on the presence of "personally identifiable information" or the ability to "identify" an individual's personal information in a release of records. Such concepts do not have precise definitions, and their meaning in the context of differential privacy applications is unclear. In addition, many privacy regulations emphasize particular requirements for protecting privacy when disclosing individual-level data, such as removing personally identifiable information, which are arguably difficult to interpret and apply when releasing aggregate statistics. While in some cases it may be clear whether a regulatory standard has been met by the use of differential privacy, in other cases—particularly along the boundaries of a standard—there may be considerable uncertainty.

Regulatory requirements relevant to issues of privacy in computation rely on an understanding of a range of different concepts, such as personally identifiable information, de-identification, linkage, inference, risk, consent, opt out, and purpose and access restrictions. In the following discussion, we show how the definition of differential privacy can be interpreted to address each of these concepts while accommodating differences in how these concepts are defined across various legal and institutional contexts.

Personally identifiable information (PII) and de-identification are central concepts in information privacy law. Regulatory protections typically extend only to personally identifiable information; information not considered personally identifiable is not protected. Although definitions of personally identifiable information vary, they are generally understood to refer to the presence of pieces of information that are linkable to the identity of an individual or to an individual's personal attributes.[20] PII is also related to the concept of *de-identification*, which refers to a collection of techniques devised for transforming identifiable information into non-identifiable information while also preserving some utility of the data. In principle, it is intended that de-identification, *if performed successfully*, can be used as a tool for removing PII, or transforming PII into non-PII.

When differential privacy is used, it can be understood as (essentially) ensuring that using an

[20] For a general definition of personally identifiable information, see, e.g., Government Accountability Office, Alternatives Exist for Enhancing Protection of Personally Identifiable Information (2008). ("For purposes of this report, the terms *personal information* and *personally identifiable information* are used interchangeably to refer to any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individuals identity, such as name, Social Security number, date and place of birth, mothers maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information."). For a survey of various definitions of *personally identifiable information*, see Paul M. Schwartz & Daniel J. Solove, The PII Problem: Privacy and a New Concept of Personally Identifiable Information, 86 N.Y.U. L. Rev. 1814 (2011).

individual's data will not reveal personally identifiable information specific to him or her.[21] Here, the use of the term "specific" refers to information that is unique to the individual and cannot be inferred unless the individual's information is used in the analysis.

Linkage is a mode of privacy loss recognized, implicitly or explicitly, by a number of privacy regulations.[22] Linkage typically refers to the matching of information in a database to a specific individual, often by leveraging information from external sources. Linkage is also closely related to the concept of *identifying* an individual in a data release, as identifying an individual is often accomplished via a successful linkage. Linkage has a concrete meaning when data are published as a collection of individual-level records, often referred to as microdata. However, what is considered a successful linkage when a publication is made in other formats (including, e.g., statistical models and synthetic data) is open to interpretation.

Despite this ambiguity, it can be argued that differential privacy addresses record linkage in the following sense. Differentially private statistics provably hide the influence of every individual, and even groups of individuals, providing protection not only against releasing exact records but also approximate statistics that leak individual-level information that could be used in a linkage attack. Furthermore, differential privacy provides a robust guarantee of privacy protection that is independent of the auxiliary information available to an attacker. When differential privacy is used, an attacker utilizing auxiliary information cannot learn much more about an individual in a database than she could if that individual's information were not in the database at all.

Inference is another mode of privacy loss that is implicitly or explicitly referenced by some privacy regulations. For example, some laws protect information that enables the identity of an individual to be "reasonably inferred,"[23] and others protect information that enables one to determine an attribute about an individual with "reasonable certainty."[24] When discussing inference as a mode of privacy loss, it is important to distinguish between two types: inferences about individuals and inferences about large groups of individuals. Although privacy regulations generally do not draw a clear distinction between these two types of inference, the distinction is key to understanding which privacy safeguards would be appropriate in a given setting.

Differential privacy (essentially) protects an individual from inferences about attributes that are specific to him or her (i.e., information that is unique to the individual and cannot be inferred unless the individual's information is used in the analysis). Interventions other than differential privacy may be necessary in contexts in which inferences about large groups of individuals, such as uses of data that result in discriminatory outcomes by race or sex, are a concern.

Risk is another concept that appears in various ways throughout regulatory standards for privacy

[21] Note that the term "use" in this statement refers to the inclusion of an individual's data in an analysis.

[22] For example, by defining personally identifiable information in terms of information "linked or linkable to a specific student," FERPA appears to emphasize the risk of a successful record linkage attack. 34 C.F.R. § 99.3. The Department of Health & Human Services in guidance on de-identifying data in accordance with the HIPAA Privacy Rule includes an extended discussion of examples of record linkage attacks and de-identification strategies for mitigating them. *See* Office for Civil Rights, Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule* (2012).

[23] *See, e.g.,* Confidential Information Protection and Statistical Efficiency Act, Pub. L. 107-347 § 502(4) (protecting "any representation of information that permits the identity of the respondent to whom the information applies to be reasonably inferred by either direct or indirect means").

[24] *See, e.g.,* FERPA, 34 C.F.R. § 99.3 (defining "personally identifiable information," in part, in terms of information that would allow one to identify a student "with reasonable certainty").

protection and related guidance. For example, some regulatory standards include a threshold level of risk that an individual's information may be identified in a data release.[25] Similarly, some regulations also acknowledge, implicitly or explicitly, that any disclosure of information carries privacy risks, and therefore the goal is to minimize rather than eliminate such risks.[26]

Differential privacy can readily be understood in terms of risk. Specifically, differential privacy enables a formal quantification of risk. It (essentially) guarantees that the risk to an individual is the same with or without her participation in the dataset, and this holds true for any notion of risk adopted by a regulatory standard or institutional policy. In this sense, differential privacy can be interpreted as (essentially) guaranteeing that the risk to an individual is minimal or very small. Moreover, the privacy loss parameter epsilon can be tuned according to different requirements for minimizing risk.

Consent and opt out are concepts underlying common provisions set forth in information privacy laws. Consent and opt out provisions enable individuals to choose to allow, or not to allow, their information to be used by or redisclosed to a third party.[27] Such provisions are premised on the assumption that providing individuals with an opportunity to opt in or out gives them control over the use of their personal information and effectively protects their privacy. However, this assumption warrants a closer look. Providing consent or opt-out mechanisms as a means of providing individuals with greater control over their information is an incomplete solution as long as individuals are not fully informed about the consequences of uses or disclosures of their information. In addition, allowing individuals the choice to opt in or out can create new privacy concerns. An individual's decision to opt out may (often unintentionally) be reflected in a data release or analysis and invite scrutiny into whether the choice to opt out was motivated by the need to hide compromising information.[28]

The differential privacy guarantee can be interpreted as providing stronger privacy protection than a consent or opt-out mechanism. This is because differential privacy can be understood as automatically providing all individuals in the data with the protection that opting out is intended to provide. When differential privacy is used, the consequences for an individual's privacy are essentially the same whether or not an individual's information is included in an analysis. Moreover, differential privacy provides all individuals with this privacy guarantee, thereby avoiding the possibility that individuals who choose to opt out would, by doing so, inadvertently reveal a sensitive attribute about themselves or attract attention as individuals who are potentially hiding sensitive

[25] For example, the HIPAA Privacy Rule requires covered entities to use de-identification techniques prior to releasing data in order to create a dataset with only a "very small" risk of identification. 45 C.F.R. § 164.514(b)(1); 65 Fed. Reg. 82,461, 82,543 (Dec. 28, 2000).

[26] Guidance on complying with CIPSEA requires agencies to "collect and handle confidential information to minimize risk of disclosure." *See* Office of Management and Budget, Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), 72 Fed. Reg. 33,361 (June 15, 2007). Guidance from the Department of Health & Human Services recognizes that de-identification methods "even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds." Office for Civil Rights, Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule* (2012).

[27] For example, FERPA includes a provision requiring educational agencies and institutions to offer students an opportunity to opt out of the disclosure of their personal information in school directories. 34 C.F.R. § 99.37.

[28] For a real-world example, consider recent reports that the National Security Agency's surveillance efforts specially target users of privacy services. *See* Kim Zetter, *The NSA Is Targeting Users of Privacy Services, Leaked Code Shows*, Wired, July 3, 2014.

facts about themselves.

Purpose and access provisions often appear in privacy regulations as restrictions on the use or disclosure of personal information to specific parties or for specific purposes. Legal requirements reflecting purpose and access restrictions can be divided into two categories. The first category includes restrictions, such as those governing confidentiality for statistical agencies,[29] prohibiting the use of identifiable information except for statistical purposes. The second category broadly encompasses other types of purpose and access provisions, such as those permitting the use of identifiable information for legitimate educational purposes.[33]

Restrictions limiting use to statistical purposes, including statistical purposes involving population-level rather than individual-level analyses or statistical computations, are consistent with the use of differential privacy. Tools that satisfy differential privacy can be understood to restrict uses to only those that are for statistical purposes. However, other use and access restrictions, such as provisions limiting use to legitimate educational purposes, are orthogonal to differential privacy and demand alternative privacy safeguards.

The foregoing interpretations of the differential privacy guarantee can be used to demonstrate that in many cases a differentially private mechanism would prevent the types of disclosures of personal information that privacy regulations have been designed to address. Moreover, in many cases, differentially private tools provide privacy protection that is more robust than that provided by techniques commonly used to satisfy regulatory requirements for privacy protection. However, further research to develop methods for proving that differential privacy satisfies legal requirements and set the privacy loss parameter epsilon based on such requirements is needed.[34] In practice, data providers should consult with legal counsel when considering whether differential privacy tools, potentially in combination with other tools for protecting privacy and security, are appropriate within their specific institutional settings.

7 Tools for differentially private analysis

At the time of this writing, differential privacy is transitioning from a purely theoretical mathematical concept to one that underlies software tools for practical use by analysts of privacy-sensitive data.The first real-world implementations of differential privacy have been deployed by companies such as Google, Apple, and Uber, and government agencies such as the U.S. Census Bureau.

[29]Title 13 restricts the use of confidential information from respondents, prohibiting uses "for any purpose other than the statistical purposes for which it is supplied,"[30] and restricting access to agency employees and approved researchers with Special Sworn Status. CIPSEA prohibits the use of protected information "for any use other than an exclusively statistical purpose,"[31] where *statistical purpose* "means the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups."[32]

[33]FERPA generally prohibits the disclosure of personally identifiable information from education records, with limited exceptions such as disclosures to school officials with a legitimate educational interest in the information, 34 C.F.R. §§ 99.31(a)(1), 99.7(a)(3)(iii), and organizations conducting studies for, or on behalf of, schools, school districts, or postsecondary institutions, 34 C.F.R. § 99.31(a)(6).

[34]For an extended discussion of the gaps between legal and computer science definitions of privacy and a demonstration that differential privacy can be used to satisfy an institution's obligations under the Family Educational Rights and Privacy Act, see Kobbi Nissim, Aaron Bembenek, Alexandra Wood, Mark Bun, Marco Gaboardi, Urs Gasser, David R. O'Brien, Thomas Steinke, & Salil Vadhan, *Bridging the Gap between Computer Science and Legal Approaches to Privacy*, Harvard Journal of Law & Technology (forthcoming 2018).

Researchers in industry and academia are currently building and testing additional tools for differentially private statistical analysis. This section briefly reviews some of these newly-emerging tools, with a particular focus on the tools that inspired the drafting of this primer.

7.1 Government and commercial applications of differential privacy

Since 2005, the U.S. Census Bureau has published an online interface enabling the exploration of the commuting patterns of workers across the United States, based on confidential data collected by the Bureau through the Longitudinal Employer-Household Dynamics program.[35] Through this interface, members of the public can interact with synthetic datasets generated from confidential survey records. Beginning in 2008, the computations used to synthesize the data accessed through the interface provide formal privacy guarantees and satisfy a variant of differential privacy.[36] In 2017, the Census Bureau announced that it was prototyping a system that would protect the full set of publication products from the 2020 decennial Census using differential privacy.

Google, Apple, and Uber have also experimented with differentially private implementations. For instance, Google developed the RAPPOR system, which applies differentially private computations in order to gather aggregate statistics from consumers who use the Chrome web browser.[37] This tool allows analysts at Google to monitor the wide-scale effects of malicious software on the browser settings of their users, while providing strong privacy guarantees for individuals.[38]

7.2 Differential privacy in Harvard's Privacy Tools project

The Harvard Privacy Tools project[39] develops tools to help social scientists and other researchers collect, analyze, and share data while providing privacy protection for individual research subjects. To this end, the project seeks to incorporate definitions and algorithmic tools from differential privacy into Dataverse, an open-source software application developed at Harvard. Dataverse provides a preservation and archival infrastructure that enables institutions to host data repositories through which researchers can upload their data or access data made available by other researchers for the purposes of replication or secondary research.

New privacy tools being developed for integration with the Dataverse infrastructure include a private data sharing interface, *PSI*,[40] which facilitates data exploration and analysis using differ-

[35] *See* U.S. Census Bureau, OnTheMap Application for the Longitudinal Employer-Household Dynamics program, http://onthemap.ces.census.gov (last visited Apr. 30, 2016).

[36] *See* Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, & Lars Vilhuber, *Privacy: Theory Meets Practice on the Map*, Proceedings of the IEEE 24th International Conference on Data Engineering 277 (2008).

[37] *See* Úlfar Erlingsson, *Learning Statistics with Privacy, aided by the Flip of a Coin*, Google Research Blog (Oct. 30, 2014), http://googleresearch.blogspot.com/2014/10/learning-statistics-with-privacy-aided.html; Úlfar Erlingsson, Vasyl Pihur & Aleksandra Korolova, *RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response*, Proceedings of the 21st ACM Conference on Computer and Communications Security (2014).

[38] Other examples for using differential privacy (for which, to the best of our knowledge, no technical reports have been published) include Google's use of differential privacy in analyzing urban mobility and Apple's use of differential privacy in iOS 10. *See* Andrew Eland, *Tackling Urban Mobility with Technology*, Google Europe Blog (Nov. 18, 2015), http://googlepolicyeurope.blogspot.com/2015/11/tackling-urban-mobility-with-technology.html, and Andy Greenberg, *Apples 'Differential Privacy' Is About Collecting Your Data–But Not Your Data*, Wired (Jun. 13, 2016), http://www.wired.com/2016/06/apples-differential-privacy-collecting-data/.

[39] Harvard Privacy Tools Project, `http://privacytools.seas.harvard.edu`.

[40] *See* Marco Gaboardi, James Honaker, Gary King, Jack Murtagh, Kobbi Nissim, Jonathan Ullman, & Salil P. Vadhan, *PSI (ψ): a Private data Sharing Interface*, Working paper (2016), `http://arxiv.org/abs/1609.04340`.

ential privacy. The PSI interface provides researchers depositing datasets into Dataverse, who are not necessarily data privacy experts, guidance on how to partition a limited privacy budget among the many statistics to be produced or analyses to be run. It also provides researchers seeking to explore a dataset available on Dataverse with guidance on how to interpret the noisy results produced by a differentially private algorithm. PSI offers a basic collection of tools for producing differentially private statistics whose results can be visualized using *TwoRavens*,[41] a browser-based interface for exploring and analyzing data. Through the differentially private access enabled by PSI, researchers will be able to perform rough preliminary analyses of privacy-sensitive datasets that currently cannot be safely shared. Such access will help researchers determine whether it is worth the effort to apply for full access to the raw data.

PSI is also being designed to integrate with other tools available through Dataverse, such as *DataTags*,[42] which are simple, iconic labels that categorically describe certain requirements for handling privacy-sensitive data. Each DataTag maps to a different set of transfer, storage, and access requirements, from completely open data (a "blue" tag) to maximally-protected data (a "crimson" tag).[43] The Privacy Tools project seeks to develop tools using the DataTags framework to denote handling policies for different versions of a dataset or for statistics derived from a dataset. For example, while a raw version of a privacy-sensitive dataset might be assigned a more restrictive DataTag (e.g., "red" or "crimson") that enables access only by approved researchers, differentially private statistics derived from the data might be assigned a less restrictive DataTag (e.g., "green") that enables access by any user registered with the Dataverse repository. In addition, members of the Privacy Tools project are assessing the privacy protection guaranteed by different settings of the differential privacy parameters (ϵ and δ), so that they can make recommendations regarding the values of these parameters that are appropriate for producing statistics from a dataset that has been assigned a given DataTag.

7.3 Other experimental implementations of differential privacy

Several other experimental systems enable data analysts to construct privacy-preserving analyses without requiring an understanding of the subtle technicalities of differential privacy. Systems such as Privacy Integrated Queries (PINQ),[44] Airavat,[45] and GUPT[46] aim to make it easier for users to write programs that are guaranteed to be differentially private, either by composition of differentially private building blocks,[47] or through the use of general frameworks such as "partition-

[41] TwoRavens, `http://datascience.iq.harvard.edu/about-tworavens`.

[42] DataTags, `http://datatags.org`.

[43] *See* Latanya Sweeney, Merce Crosas, & Michael Bar-Sinai, *Sharing Sensitive Data with Confidence: The Datatags System*, Technology Science (2015), `http://techscience.org/a/2015101601`.

[44] *See* Frank McSherry, *Privacy Integrated Queries: An Extensible Platform for Privacy-preserving Data Analysis*, Proceedings of the 2009 Int'l Conference on Management of Data 19 (2009), http://doi.acm.org/10.1145/1559845.1559850.

[45] *See* Indrajit Roy, Srinath T. V. Setty, Ann Kilzer, Vitaly Shmatikov, & Emmett Witchel, *Airavat: Security and privacy for MapReduce*, Proceedings of the 7th USENIX Symposium on Networked Systems Design and Implementation 297 (2010), `http://www.usenix.org/events/nsdi10/tech/full_papers/roy.pdf`.

[46] *See* Prashanth Mohan, Abhradeep Thakurta, Elaine Shi, Dawn Song, & David E. Culler, *GUPT: Privacy Preserving Data Analysis Made Easy*, Proceedings of the ACM SIGMOD International Conference on Management of Data 349 (2012), `http://doi.acm.org/10.1145/2213836.2213876`.

[47] *See* Frank McSherry, *Privacy Integrated Queries: An Extensible Platform for Privacy-preserving Data Analysis*, Proceedings of the 2009 Int'l Conference on Management of Data 19 (2009), http://doi.acm.org/10.1145/1559845.1559850; Andreas Haeberlen, Benjamin C. Pierce, & Arjun Narayan, *Differen-*

and-aggregate" or "subsample-and-aggregate"[48] to convert non-private programs into differentially private ones.[49] These systems rely on a common approach: they keep the data safely stored and allow users to access them only via a programming interface which guarantees differential privacy. They also afford generality, enabling one to design many types of differentially private programs that are suitable for a wide range of purposes. However, note that most of these systems do not provide much guidance for a lay user who has limited expertise in programming. Moreover, they do not provide much guidance on deciding how to partition a limited privacy budget among many statistics or analyses, or how to interpret the noisy results given by a differentially private algorithm.

7.4 Tools for specific data releases or specific algorithms

There have been a number of successful applications of differential privacy with respect to specific, structured sources of data, including commuter patterns,[50] mobility data,[51] client-side software data,[52] genome-wide association studies,[53] location history data,[54] and usage patterns for phone technology.[55] In these settings, differential privacy experts carefully optimize the choice of differentially private algorithms and the partitioning of the privacy budget to maximize utility for the particular data source. These tools are specific to the type of data they are designed to handle, and they cannot be applied in contexts in which the collection of data sources and the structure of the datasets are too heterogenous to allow for such optimizations.

Beyond these examples, there is a vast literature on the design of differentially private algorithms for specific data analysis tasks, including substantial experimental work on comparing and optimizing such algorithms across a wide range of datasets. For example, the recent work on DPBench,[56] a framework for standardized evaluation of the accuracy of privacy algorithms, provides

tial Privacy Under Fire, Proceedings of the 20th USENIX Security Symposium (2011), http://www.cis.upenn.edu/ ahae/papers/fuzz-sec2011.pdf.

[48] *See* Kobbi Nissim, Sofya Raskhodnikova, & Adam Smith, *Smooth Sensitivity and Sampling in Private Data Analysis*, Proceedings of the 39th ACM Symposium on Theory of Computing (2007), http://doi.acm.org/10.1145/1250790.1250803.

[49] *See* Indrajit Roy, Srinath T. V. Setty, Ann Kilzer, Vitaly Shmatikov, & Emmett Witchel, *Airavat: Security and privacy for MapReduce*, Proceedings of the 7th USENIX Symposium on Networked Systems Design and Implementation 297 (2010), http://www.usenix.org/events/nsdi10/tech/full_papers/roy.pdf; Prashanth Mohan, Abhradeep Thakurta, Elaine Shi, Dawn Song, & David E. Culler, *GUPT: Privacy Preserving Data Analysis Made Easy*, Proceedings of the ACM SIGMOD International Conference on Management of Data 349 (2012), http://doi.acm.org/10.1145/2213836.2213876.

[50] *See* Ashwin Machanavajjhala, Daniel Kifer, John M. Abowd, Johannes Gehrke, & Lars Vilhuber, *Privacy: Theory Meets Practice on the Map*, Proceedings of the 24th International Conference on Data Engineering 277 (2008), http://dx.doi.org/10.1109/ICDE.2008.4497436.

[51] *See* Darakhshan J. Mir, Sibren Isaacman, Ramón Cáceres, Margaret Martonosi, & Rebecca N. Wright, *DP-WHERE: Differentially Private Modeling of Human Mobility*, Proceedings of the 2013 IEEE International Conference on Big Data 580 (2013), http://dx.doi.org/10.1109/BigData.2013.6691626.

[52] *See* Úlfar Erlingsson, Vasyl Pihur & Aleksandra Korolova, *RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response*, Proceedings of the 21st ACM Conference on Computer and Communications Security (2014).

[53] *See* Xiaoqian Jiang, Yongan Zhao, Xiaofeng Wang, Bradley Malin, Shuang Wang, Lucila OhnoMachado, & Haixu Tang, *A community assessment of privacy preserving techniques for human genomes*, BMC Medical Informatics and Decision Making 14(Suppl 1) (2014), https://www.ncbi.nlm.nih.gov/pubmed/25521230.

[54] *See* Andrew Eland, *Tackling Urban Mobility with Technology*, Google Europe Blog (Nov. 18, 2015), http://googlepolicyeurope.blogspot.com/2015/11/tackling-urban-mobility-with-technology.html.

[55] *See* Apple Press Info, *Apple previews iOS 10, the biggest iOS release ever* (2016), https://www.apple.com/pr/library/2016/06/13Apple-Previews-iOS-10-The-Biggest-iOS-Release-Ever.html.

[56] *See* Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, & Dan Zhang, *Principled evaluation*

a thorough comparison of different algorithms and different ways of optimizing them.[57]

8 Summary

Differential privacy provides a formal, quantifiable measure of privacy. It is established by a rich and rapidly evolving theory that enables one to reason with mathematical rigor about privacy risk. Quantification of privacy is achieved by the privacy loss parameter ϵ, which controls, simultaneously for every individual contributing to the analysis, the deviation between one's opt-out scenario and the actual execution of the differentially private analysis.

This deviation can grow as an individual participates in additional analyses, but the overall deviation can be bounded as a function of ϵ and the number of analyses performed. This amenability to *composition*, or the ability to provable privacy guarantees with respect to the cumulative risk from successive data releases, is a unique feature of differential privacy. While it is not the only framework that quantifies a notion of risk for a single analysis, it is currently the only framework with quantifiable guarantees on the risk resulting from a composition of several analyses.

The parameter ϵ can be interpreted as bounding the excess risk to an individual resulting from her data being used in analysis (compared to her risk when her data are not being used). Indirectly, the parameter ϵ also controls the accuracy to which a differentially private computation can be performed. For example, researchers making privacy-sensitive data available through a differentially private tool may, through the interface of the tool, choose to produce a variety of differentially private summary statistics while maintaining a desired level of privacy (quantified by an accumulated privacy loss parameter), and then compute summary statistics with formal privacy guarantees.

Systems that adhere to strong formal definitions like differential privacy provide protection that is robust to a wide range of potential privacy attacks, including attacks that are unknown at the time of deployment.[58] An analyst designing a differentially private data release need not anticipate particular types of privacy attacks, such as the likelihood that one could link particular fields with other data sources that may be available. Differential privacy *automatically* provides a robust guarantee of privacy protection that is independent of the methods and resources used by a potential attacker.

Differentially private tools also have the benefit of transparency, as it is not necessary to maintain secrecy around a differentially private computation or its parameters. This feature distinguishes differentially private tools from traditional de-identification techniques which often require concealment of the extent to which the data have been transformed, thereby leaving data users with uncertainty regarding the accuracy of analyses on the data.

Differentially private tools can be used to provide broad, public access to data or data summaries in a privacy-preserving way. Differential privacy can help enable researchers, policymakers, and businesses to analyze and share sensitive data that cannot otherwise be shared due to privacy concerns. Further, it ensures that they can do so with a guarantee of privacy protection that substantially increases their ability to protect the individuals in the data. This, in turn, can further the progress of scientific discovery and innovation.

---

of differentially private algorithms using DPBench, Proceedings of the 2016 International Conference on Management of Data (2016), http://dl.acm.org/citation.cfm?id=2882931.

[57] See also DPComp, https://www.dpcomp.org.

[58] We define privacy attacks as attempts to learn private information specific to individuals from a data release.

Further reading

Differential privacy was introduced in 2006 by Dwork, McSherry, Nissim and Smith.[59] This primer's presentation of the opt-out scenario vs. real-world computation is influenced by Dwork (2006),[60] and its risk analysis is influenced by Kasiviswanathan and Smith (2008).[61] For other presentations of differential privacy, see Dwork (2011) and Heffetz and Ligett (2014).[62] For a thorough technical introduction to differential privacy, see Dwork and Roth (2014).[63]

[59] Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, Proceedings of the Third Theory of Cryptography Conference 265 (2006), http://dx.doi.org/10.1007/11681878_14.

[60] Cynthia Dwork, *Differential privacy*, Proceedings of the 33rd International Colloquium on Automata, Languages and Programming 1 (2006), http://dx.doi.org/10.1007/11787006_1.

[61] Shiva Prasad Kasiviswanathan & Adam Smith, *A note on differential privacy: Defining resistance to arbitrary side information* (2008), http://arxiv.org/abs/0803.3946.

[62] Cynthia Dwork, *A firm foundation for private data analysis*, 54 Communications of the ACM 86 (2011), http://doi.acm.org/10.1145/1866739.1866758; Ori Heffetz & Katrina Ligett, *Privacy and data-based research*, 28 Journal of Economic Perspectives 75 (2014), http://www.aeaweb.org/articles.php?doi=10.1257/jep.28.

[63] Cynthia Dwork & Aaron Roth, *The algorithmic foundations of differential privacy*, 9 Foundations and Trends in Theoretical Computer Science 211 (2014), http://dx.doi.org/10.1561/0400000042.

A Advanced topics

We conclude with some advanced topics for readers interested in exploring differential privacy further. This section explores how differentially private analyses are constructed and how the noise introduced by differential privacy compares to statistical sampling error, discusses the protection differential privacy can provide for small groups of individuals, and introduces the concept of the secrecy of the sample.

A.1 How are differentially private analyses constructed?

As indicated above, the construction of differentially private analyses relies on the careful introduction of uncertainty in the form of random noise. This section provides a simple example illustrating how a carefully-calibrated amount of random noise can be added to the outcome of an analysis in order to provide privacy protection.

> Consider computing an estimate of the number of HIV-positive individuals in a sample, where the sample contains $n = 10,000$ individuals of whom $m = 38$ are HIV-positive. In a differentially private version of the computation, random noise Y is introduced into the count so as to hide the contribution of a single individual. That is, the result of the computation would be $m' = m + Y = 38 + Y$ instead of $m = 38$.

The magnitude of the random noise Y affects both the level of privacy protection provided and the accuracy of the count. For instance, a larger amount of noise would result in better privacy protection and worse accuracy—and vice versa. The magnitude of Y depends on the privacy loss parameter ϵ, where a smaller value of ϵ is associated with a larger noise magnitude.[64]

When choosing the noise distribution, one possibility is to sample the random noise Y from a normal distribution with zero mean and standard deviation $1/\epsilon$.[65] Because the choice of the value of ϵ is inversely related to the magnitude of the noise introduced by the analysis, the mechanism is designed to provide a quantifiable tradeoff between privacy and utility. Consider the following example.

> A researcher uses the estimate m', as defined in the previous example, to approximate the fraction p of HIV-positive people in the population. The computation would result in the estimate
>
> $$p' = \frac{m'}{n} = \frac{38 + Y}{10,000}.$$

[64] In some implementations of differential privacy, a second parameter denoted by the Greek letter δ (delta) is also used. The parameter δ controls the probability that a privacy breach event would happen, and hence should be kept very small (e.g., one in a billion). To simplify the presentation here, we will assume that δ is set to zero.

[65] More accurately, the noise Y is sampled from the Laplace distribution with zero mean and standard deviation $\sqrt{2}/\epsilon$. The exact shape of the noise distribution is important for proving that outputting $m + Y$ preserves differential privacy, but can be ignored for the current discussion.

For instance, suppose the sampled noise is $Y = 4.2$. Then, the estimate would be

$$p' = \frac{38 + Y}{10,000} = \frac{38 + 4.2}{10,000} = \frac{42.2}{10,000} = 0.42\%,$$

whereas without added noise, the estimate would have been $p = 0.38\%$.

A.2 Two sources of error: sampling error and added noise

We continue with the example from the previous section. Note that there are two sources of error in estimating p: sampling error and added noise. The first source, sampling error, would cause m to differ from the expected $p \cdot n$ by an amount of roughly

$$|m - p \cdot n| \approx \sqrt{p \cdot n}.$$

For instance, consider how the researcher from the example above would calculate the sampling error associated with her estimate.

The researcher reasons that m' is expected to differ from $p \cdot 10,000$ by roughly

$$\sqrt{p \cdot 10,000} \approx \sqrt{38} \approx 6.$$

Hence, the estimate 0.38% is expected to differ from the true p by approximately

$$\frac{6}{10,000} = 0.06\%,$$

even prior to the addition of the noise Y by the differentially private mechanism.

The second source of error is the addition of random noise Y in order to achieve differential privacy. This noise would cause m' and m to differ by an amount of roughly

$$|m' - m| \approx 1/\epsilon.$$

The researcher in the example would calculate this error as follows.

The researcher reasons that, with a choice of $\epsilon = 0.1$, she should expect $|m' - m| \approx 1/0.1 = 10$, which can shift p' from the true p by an additional $\frac{10}{10,000} = 0.1\%$.

Taking both sources of noise intro account, the researcher calculates that the difference between noisy estimate p' and the true p is roughly

$$0.06\% + 0.1\% = 0.16\%.$$

Because the two sources of noise are statistically independent, the researcher can use the fact that their variances add to produce a slightly better bound:

$$|p' - p| \approx \sqrt{0.06^2 + 0.1^2} = 0.12\%.$$

Generalizing from this example, we find that the standard deviation of the estimate p' (hence the expected difference between p' and p) is of magnitude roughly

$$|p' - p| \approx \sqrt{p/n} + 1/n\epsilon,$$

which means that for a large enough sample size n the sampling error would far exceed the noise added for the purposes of privacy protection.

Note also that the literature on differentially private algorithms has identified many other noise introduction techniques that result in better accuracy guarantees than the simple technique used in the examples above. Such techniques are especially important for more complex analyses, for which the simple noise addition technique discussed in this section is often sub-optimal in terms of accuracy.

A.3 Group privacy

By holding individuals' opt-out scenarios as the relevant baseline, the definition of differential privacy directly addresses disclosures of information localized to a single individual. However, in many cases, information may be shared between multiple individuals. For example, relatives may share an address or certain genetic attributes.

How does differential privacy protect information of this nature? Consider the opt-out scenario for a group of k individuals. This is the scenario in which the personal information of all k individuals is omitted from the input to the analysis. For instance, John and Gertrude's opt-out scenario is the scenario in which both John's and Gertrude's information is omitted from the input to the analysis.

Recall that the parameter ϵ controls by how much the real-world scenario can differ from any individual's opt-out scenario. It can be shown that the difference between the real-world and opt-out scenarios of a group of k individuals grows to at most

$$k \cdot \epsilon.$$

This means that the privacy guarantee degrades moderately as the size of the group increases. Effectively, a meaningful privacy guarantee can be provided to groups of individuals of a size of up to about

$$k \approx 1/\epsilon$$

individuals. However, almost no protection is guaranteed to groups of

$$k \approx 10/\epsilon$$

individuals or greater. This is the result of a design choice to not *a priori* prevent analysts using differentially private mechanisms from discovering trends across moderately-sized groups.

A.4 Amplifying privacy: Secrecy of the sample

As discussed in Section 6, differential privacy limits accuracy, and the extent of the inaccuracy depends inversely on the privacy parameter ϵ. Sometimes, the dataset used as input to a differentially private mechanism is a random sample from a large population, as in the following example.

Alice, a researcher at State University, collected personal information from individuals in a study exploring the relationship between coffee consumption, economic status, and health status. The personal information she collected in this study is based on a uniformly random and secret sample of $3,000$ individuals living in the city of Boston.

Because Alice's study uses a uniformly random sample,[66] and, furthermore, the identities of the participating individuals are kept confidential, Alice can apply a theorem in differential privacy known as "secrecy of the sample." This theorem effectively allows for a savings in the privacy parameter ϵ that corresponds to the ratio of sizes between the dataset and the larger population. For instance, for a population the size of the city of Boston, approximately $600,000$, the savings in ϵ can be $3,000/600,000 = 0.05$. This means that greater accuracy, corresponding to a 0.05 decrease in epsilon, can be provided for the differentially private analyses performed on the data from Alice's study.

This topic comes with two notes of caution. First, sampling from the sampling frame is usually not uniform in practice. Alice should therefore be conservative in her estimate of the underlying population. For example, if Alice draws her survey sample from a Boston phonebook, then she should take the underlying population size to be no larger than the number of Boston residents who are listed in the phonebook. Second, the effective decrease in ϵ is conditioned on the identities of the sampled individuals being kept secret. This may be a hard condition to maintain in practice. For example, if Alice sends surveyors to respondents' homes, then their neighbors may learn that they participated in Alice's research study. A more subtle consideration is that secrecy of the sample also requires the identities of individuals who have *not* been sampled to be kept secret.

[66] By *uniformly random* we mean that each individual in the sampling frame is selected to be in the sample with equal probability and independently of the other individuals in the sampling frame.