CS 860: Algorithms for Private Data Analysis Fall 2020

Lecture 3 — Intro to Differential Privacy

Prof. Gautam Kamath *Scribe: Gautam Kamath*

In this lecture, we will introduce differential privacy. We start with perhaps the first differentially private algorithm, by Warner from 1965 [War65].

Randomized Response

We work in a very simple setting. Suppose you are the instructor of a large class which has an important exam. You suspect that many students in the class cheated, but you aren't sure. How can you figure out what fraction of students cheated? Naturally, students would not be likely to honestly admit that they cheated.

Being a bit more precise: there are n people, and individual i has a sensitive bit $X_i \in \{0, 1\}$. They would like to ensure that no one else learns the value of X_i. Each person sends the analyst a message Y_i, which may depend on X_i and some random numbers which the individual can generate. Based on these Y_i's, the analyst would like to get an accurate estimate of $p = \frac{1}{n}\sum_{i=1}^n X_i$.

We can first start with the most obvious approach: individual i sends Y_i equal to the sensitive bit X_i. Foreshadowing, we write this in the following unconventional manner:

$$Y_i = \begin{cases} X_i & \text{with probability } 1 \\ 1 - X_i & \text{with probability } 0 \end{cases}$$

It is clear that the analyst can simply obtain $\tilde{p} = \frac{1}{n}\sum_{i=1}^n Y_i$, and that $\tilde{p} = p$ exactly. In other words, the result is perfectly accurate. However, the analyst sees Y_i, which is equal to X_i, and thus she learns the individual's private bit exactly: there is no privacy.

Consider an alternate strategy, as follows:

$$Y_i = \begin{cases} X_i & \text{with probability } 1/2 \\ 1 - X_i & \text{with probability } 1/2 \end{cases}$$

In this case, Y_i is perfectly private: in fact, it is a uniformly bit which does not depend on X_i at all, so the curator could not hope to infer anything about X_i from seeing it. But by the same token, this approach loses all sort of accuracy: $\tilde{Z} = \frac{1}{n}\sum_{i=1}^n Y_i$ is distributed as $\frac{1}{n}\text{Binomial}(n, 1/2)$, which is completely independent of the statistic Z.

At this point, we have two approaches: one which is perfectly accurate but not at all private, and one which is perfectly private but not at all accurate. The right approach will be to interpolate between these two extremes.

Consider the following strategy, which we will call *Randomized Response*, parameterized by some $\gamma \in [0, 1/2]$:

$$Y_i = \begin{cases} X_i & \text{with probability } 1/2 + \gamma \\ 1 - X_i & \text{with probability } 1/2 - \gamma \end{cases}$$

How private is this message Y_i, with respect to the true message X_i? We haven't built the tools to formally quantify this yet, so we'll be a bit informal for the time being. Note that $\gamma = 1/2$ corresponds to the first "honest" strategy, and $\gamma = 0$ is the second "uniformly random" strategy. What if we choose a γ in the middle, such as $\gamma = 1/4$? Then there will be a certain level of "plausible deniability" associated with the individual's disclosure: while $Y_i = X_i$ with probability $3/4$, it could be that their true bit was $1 - Y_1$, and this event happened with probability $1/4$. Informally speaking, how "deniable" their response is corresponds to the level of privacy they are afforded. In this way, they get a stronger privacy guarantee as γ approaches 0.

Let's put this aside for now, and focus on how accurate an estimate the analyst can obtain. Observe that

$$E[Y_i] = 2\gamma X_i + 1/2 - \gamma,$$

and thus

$$E\left[\frac{1}{2\gamma}\left(Y_i - 1/2 + \gamma\right)\right] = X_i.$$

This leads to the following natural estimator:

$$\tilde{p} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2\gamma}\left(Y_i - 1/2 + \gamma\right)\right].$$

The above calculation gives that $E[\tilde{p}] = p$. Next, we analyze the variance of $\tilde{p}$:

$$\mathbf{Var}[\tilde{p}] = \mathbf{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2\gamma}\left(Y_i - 1/2 + \gamma\right)\right]\right] = \frac{1}{4\gamma^2 n^2}\sum_{i=1}^{n}\mathbf{Var}\left[Y_i\right] \leq \frac{1}{16\gamma^2 n}.$$

The last inequality is due to the fact that the variance of a Bernoulli random variable is upper bounded by $1/4$. At this point, we can apply Chebyshev's inequality to obtain

$$|\tilde{p} - p| \leq O\left(\frac{1}{\gamma\sqrt{n}}\right).$$

This can also be obtained with high probability via a Chernoff bound.[1] As $n \to \infty$, this error goes to 0. An alternative way of wording this: if we wish to have additive error α, we require $n = O(1/\alpha^2\gamma^2)$ samples. Note that as γ gets closer to 0 (corresponding to stronger privacy), the error increases (or, with the second phrasing, the sample complexity). This is natural: the stronger the privacy guarantee we would like, the more data we require to achieve the same accuracy.

In order to proceed further in quantifying the level of privacy, we must (finally) introduce differential privacy. At its core, differential privacy is a broad formalization of this aforementioned notion of "plausible deniability."

Differential Privacy

In security and privacy, it is important to be precise about the precise setting in which we are working. We now define the setting for differential privacy, sometimes called *central differential*

[1] If you are not familiar with either the Chebyshev or Chernoff bound and the argument that we are applying here, it is important that you look it up and work out the details.

privacy or the *trusted curator* model. We imagine there are n individuals, X_1 through X_n, who each have their own datapoint. They send this point to a "trusted curator" – all individuals trust this curator with their raw datapoint, but no one else. Given their data, the curator runs an algorithm M, and publicly outputs the result of this computation. Differential privacy is a property of this algorithm M,[2] saying that no individual's data has a large impact on the output of the algorithm.

More formally, suppose we have an algorithm $M : \mathcal{X}^n \to \mathcal{Y}$. Consider any two datasets $X, X' \in \mathcal{X}^n$ which differ in exactly one entry. We call these *neighbouring datasets*, and sometimes denote this by $X \sim X'$. We say that M is *ε-(pure) differentially private* (ε-(pure) DP) if, for all neighbouring X, X', and all $T \subseteq \mathcal{Y}$, we have

$$\Pr[M(X) \in T] \leq e^{\varepsilon} \Pr[M(X') \in T],$$

where the randomness is over the choices made by M.

This definition was given by Dwork, McSherry, Nissim, and Smith in their seminal paper in 2006 [DMNS06]. It is now widely accepted as a strong and rigorous notion of data privacy. It has received acclaim in theory, winning the 2017 Gödel Prize, and the 2016 TCC Test-of-Time Award. At the same time, it has now seen adoption in practice at many organizations, including Apple [Dif17], Google [EPK14], Microsoft [DKY17], the US Census Bureau for the 2020 US Census [DLS+17], and much more.

Differential Privacy is an unusual sounding definition the first time you see it, so some discussion is in order.

- Differential privacy is quantitative in nature. A small ε corresponds to strong privacy, degrading as ε increases.

- ε should be thought of as a small-ish constant. Anything between (say) 0.1 and 5 might be a reasonable level privacy guarantee (smaller corresponds to stronger privacy), and you should be slightly skeptical of claims significantly outside this range.

- This is a worst-case guarantee, over all neighbouring datasets X and X'. Even if we expect our data to be randomly generated (and some realizations are incredibly unlikely), we still require privacy for all possible datasets nonetheless. While there do exist some notions of average-case privacy, these should be approached with caution – Steinke and Ullman write a series of posts which warn about the pitfalls of average-case notions of differential privacy [SU20a, SU20b].

- In words, the definition bounds the multiplicative increase (incurred by changing a single point in the dataset) in the probability of M's output satisfying any event.

- The use of a multiplicative e^{ε} in the probability might seem unnatural. For small ε, a Taylor expansion allows us to treat this as $\approx (1 + \varepsilon)$. The given definition is convenient because of the fact that $e^{\varepsilon_1} \cdot e^{\varepsilon_2} = e^{\varepsilon_1 + \varepsilon_2}$, which is useful when we examine the property of "group privacy" later.

- While the definition may look asymmetric, it is not: one can simply swap the role of X and X'.

[2] In differential privacy lingo, an algorithm is sometimes (confusingly) called a "mechanism."

- Convince yourself that any non-trivial (i.e., one that is not independent of the dataset) differentially private algorithm must be randomized.

- One might consider other notions of "closeness" of the distributions of $M(X)$ and $M(X')$. The given definition says the probability of any event is multiplicatively close. But at a glance, the statistical or total variation distance might also seem reasonable – essentially converting the multiplicative guarantee to an additive one. But this alternative notion would not give meaningful guarantees; we don't get into this here, but see Section 1.6 of [Vad17] for more discussion.

- Finally, we will generally use the notion "neighbouring datasets" where one point in X is changed arbitrarily to obtain X'. This is sometimes called "bounded" differential privacy, in contrast to "unbounded" differential privacy, where a point is either added or removed. In theory, these notions are equivalent up to a factor of 2, as an arbitrary change can be performed by removing one point and adding another. This can be formalized later, once we study the notion of group privacy. The former definition is usually more convenient mathematically.

That's it for technical comments on the definition.

As a brief interlude, let's discuss an alternative formulation of differential privacy in terms of hypothesis testing, due to Wasserman and Zhou [WZ10], and also explored by [KOV15, BBG+20].

This phrasing is slightly more "operational" in nature, viewing things from the perspective of an adversary. Specifically, suppose the adversary is trying to decide between the following two scenarios, where X and X' are neighbouring datasets, and one of the two is guaranteed to hold:

$$H_0 : \text{the underlying dataset is } X$$

$$H_1 : \text{the underlying dataset is } X'$$

Using statistics terminology, these are called the null and the alternate hypothesis, respectively. Based on the output of some algorithm M which is run on the dataset, the adversary is trying to determine whether H_0 or H_1 is true. Intuitively, differential privacy says that the adversary shouldn't be to get significant advantage over randomly guessing. The actual guarantee is slightly more refined – for example, they could simply guess H_0 every time, and they would always be right when H_0 is true (compared to probability 1/2 by random guessing). Specifically, let p be the probability that the adversary predicts H_1 when H_0 is true (a "false positive") and q be the probability that the adversary predicts H_0 when H_1 is true (a "false negative"). ε-differential privacy implies that, simultaneously:

$$p + e^{\varepsilon} q \geq 1$$

$$e^{\varepsilon} p + q \geq 1$$

One can see that, when $\varepsilon = 0$, the adversary is essentially restricted to strategies that ignore the data and guess randomly (potentially in a biased way). As ε is increased, it allows the adversary some possibility of getting some advantage over blind guessing.

Why use this formulation of differential privacy? One reason is that it is more "operational" in nature, and gives one an alternative quantitative understanding of how well an adversary can detect the contribution of an individual. It is also used in understanding the privacy guarantees we get when we run multiple private algorithms on the same dataset [KOV15]. The recent notion of Gaussian differential privacy [DRS19] also embraces this interpretation, rephrasing the privacy guarantee in terms of hypothesis testing between two Gaussian distributions.

Let's take a step back: what does differential privacy *mean*? Simply repeating the definition: differential privacy says that, the probability of any event is comparable in the cases when an individual does or does not include their data in the dataset. This has a number of implications of what differential privacy does and does not ensure.

First, it prevents many of the types of attacks we have seen before. The linkage-style attacks that we have observed are essentially ruled out – if such an attack were effective with your data in the dataset, it would be almost as effective without. This holds true for existing auxiliary datasets, as well as any *future* data releases as well. It also prevents reconstruction attacks, in some sense "matching" the bounds shown in the Dinur-Nissim attacks [DN03], as we will quantify in a later lecture. In fact, it protects against *arbitrary* risks, which can be reasoned about by simply revisiting the fact that any outcome is comparably likely whether or not the individual's data was actually included.

Differential privacy does *not* prevent you from making inferences about individuals. Stated alternatively: differential privacy does not prevent statistics and machine learning. Consider the classic "Smoking Causes Cancer" example [DR14]. Suppose an individual who smokes cigarettes is weighing their options in choosing to participate in a medical study, which examines whether smoking causes cancer. They know that a positive result to this study would be detrimental to them, as it would cause their insurance premiums to rise. They also know that the study is being performed using differentially privately, so they choose to participate, and they know their privacy will be respected. Unfortunately for them, the study reveals that smoking does cause cancer! This is a privacy violation, right? No: differential privacy ensures that the outcome of the study would not be significantly impacted by their participation. In other words, whether they participated or not, the result was going to come out anyway. For more discussion of the compatibility of privacy and learning, see [McS16].

Differential privacy is also not suitable for the case where the goal is to identify a specific individual, and this is antithetical to the definition. As a timely example, despite the clamoring for privacy-preserving solutions for tracking the spread of COVID-19, it is not immediately clear how one could use differential privacy to facilitate *individual-level* contact tracing. This would seem to require information about where a specific individual has been, and which particular individuals they have interacted with. On the other hand, it might be possible to facilitate aggregate-level tracking, say if many people who tested positive all attended the same event. In this vein, there is some interesting work done by Google on DP analysis of location traces, to see which types of locations people spend more and less time at since COVID-19 struck [ABC$^+$20].

The definition of differential privacy is information theoretic in nature. That is, an adversary with unlimited amounts of computational power and auxiliary information is still unable to get an advantage. This is in contrast to cryptography, which typically focuses on computationally bounded adversaries. There has been some work on models of differential privacy where the adversary is computational bounded, see, e.g., [BNO08].

Randomized Response, Revisited

Design of differentially private algorithms is usually built around a few core primitives. One of these is randomized response, which we are now equipped to analyze the privacy of.

Now that we have the definition in hand, let's analyze the differential privacy guarantee when our algorithm M is randomized response. In fact, we will actually show that the bit-string $M(X_1, \dots, X_n) = (Y_1, \dots, Y_n)$ is differentially private – privacy of our estimate $\tilde{p}$ will follow by the post-processing property of differential privacy (essentially saying that a function of a differentially private object is also private), which we will discuss next lecture. We consider any particular realization $a \in \{0,1\}^n$ of $(Y_1, \dots, Y_n)$. We have that $\Pr[M(X) = a] = \prod_{i=1}^n \Pr[Y_i = a_i]$. Suppose that X and X' differ only in coordinate j. Then we have that

$$\frac{\Pr[M(X) = a]}{\Pr[M(X') = a]} = \frac{\prod_{i=1}^n \Pr[Y_i = a_i]}{\prod_{i=1}^n \Pr[Y_i' = a_i]} = \frac{\Pr[Y_j = a_j]}{\Pr[Y_j' = a_j]} \leq \frac{1/2 + \gamma}{1/2 - \gamma} \leq e^{O(\gamma)},$$

where the last inequality holds for γ (say) smaller than 1/4. Therefore, we have that ε-randomized response is $O(\varepsilon)$-differentially private, and achieves accuracy $O\left(\frac{1}{\varepsilon\sqrt{n}}\right)$. Actually, randomized response provides a stronger privacy guarantee than (central) differential privacy, it provides *local* differential privacy, in which individuals trust no one but themselves. This will be the topic of later lectures.

We'll end here, but next time we will start with the Laplace Mechanism. This is a very flexible algorithm which applies in more general settings, also achieves ε-differential privacy, and much better accuracy for this task than randomized response.

References

[ABC+20] Ahmet Aktay, Shailesh Bavadekar, Gwen Cossoul, John Davis, Damien Desfontaines, Alex Fabrikant, Evgeniy Gabrilovich, Krishna Gadepalli, Bryant Gipson, Miguel Guevara, Chaitanya Kamath, Mansi Kansal, Ali Lange, Chinmoy Mandayam, Andrew Oplinger, Christopher Pluntke, Thomas Roessler, Arran Schlosberg, Tomer Shekel, Swapnil Vispute, Mia Vu, Gregory Wellenius, Brian Williams, and Royce J. Wilson. Google covid-19 community mobility reports: Anonymization process description (version 1.0). *arXiv preprint arXiv:2004.04145*, 2020.

[BBG+20] Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. Hypothesis testing interpretations and rényi differential privacy. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, AISTATS '20, pages 2496–2506. JMLR, Inc., 2020.

[BNO08] Amos Beimel, Kobbi Nissim, and Eran Omri. Distributed private data analysis: Simultaneously solving how and what. In *Proceedings of the 28th Annual International Cryptology Conference*, CRYPTO '08, pages 451–468, Berlin, Heidelberg, 2008. Springer.

[Dif17] Differential Privacy Team, Apple. Learning with privacy at scale. https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf, December 2017.

[DKY17] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *Advances in Neural Information Processing Systems 30*, NIPS '17, pages 3571–3580. Curran Associates, Inc., 2017.

[DLS$^+$17] Aref N. Dajani, Amy D. Lauger, Phyllis E. Singer, Daniel Kifer, Jerome P. Reiter, Ashwin Machanavajjhala, Simson L. Garfinkel, Scot A. Dahl, Matthew Graham, Vishesh Karwa, Hang Kim, Philip Lelerc, Ian M. Schmutte, William N. Sexton, Lars Vilhuber, and John M. Abowd. The modernization of statistical disclosure limitation at the U.S. census bureau, 2017. Presented at the September 2017 meeting of the Census Scientific Advisory Committee.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284, Berlin, Heidelberg, 2006. Springer.

[DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, pages 202–210, New York, NY, USA, 2003. ACM.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[DRS19] Jinshuo Dong, Aaron Roth, and Weijie J. Su. Gaussian differential privacy. *arXiv preprint arXiv:1905.02383*, 2019.

[EPK14] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, 2014. ACM.

[KOV15] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *Proceedings of the 32nd International Conference on Machine Learning*, ICML '15, pages 1376–1385. JMLR, Inc., 2015.

[McS16] Frank McSherry. Statistical inference considered harmful. https://github.com/frankmcsherry/blog/blob/master/posts/2016-06-14.md, June 2016.

[SU20a] Thomas Steinke and Jonathan Ullman. The pitfalls of average-case differential privacy. https://differentialprivacy.org/average-case-dp/, July 2020.

[SU20b] Thomas Steinke and Jonathan Ullman. Why privacy needs composition. https://differentialprivacy.org/privacy-composition/, August 2020.

[Vad17] Salil Vadhan. The complexity of differential privacy. In Yehuda Lindell, editor, *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, chapter 7, pages 347–450. Springer International Publishing AG, Cham, Switzerland, 2017.

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

[WZ10] Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.