CIS 800/002 The Algorithmic Foundations of Data Privacy September 20, 2011

Lecture 4

Lecturer: Aaron Roth *Scribe: Aaron Roth*

Composition Theorems

We now have two basic ways of accessing a database privately:

1. The Laplace Mechanism: For answering non-adaptively chosen numeric queries.
2. The Exponential Mechanism: For answering non-numeric queries.

These two tools can be put together to construct more complicated private algorithms, but in order to do that, we need to understand how privacy parameters compose. That is, when we run multiple algorithms, each of which have privacy guarantees on their own, what is the privacy guarantee on the union of their outputs? How do the privacy parameters degrade?

Before we begin, we must discuss what exactly we mean by composition. We would like our definitions to cover the following two interesting scenarios:

1. Repeated use of differentially private algorithms on the same database. This allows both the repeated use of the same mechanism multiple times, as well as the modular construction of differentially private algorithms from basic private building blocks.
2. Repeated use of differentially private algorithms on *different* databases that may nevertheless contain information relating to the same individual. This allows us to reason about the cumulative privacy loss of a single individual whose data might be spread across multiple data sets, each of which may be used independently in a differentially private way.

Towards this end, we consider the following composition experiment, which takes as input an arbitrary adaptive algorithm $\mathcal{A}$ which we view as the "adversary" trying to break the privacy of some family of database access mechanisms $\mathcal{M}$ (e.g. the set of all ϵ-differentially private mechanisms), as well as a parameter b which can take value either 0 or 1.

Algorithm 1 k-Fold Adaptive Composition Experiment b for an adversary $\mathcal{A}$ and a family of database access mechanisms $\mathcal{M}$.

Compose$(\mathcal{A}, \mathcal{M}, k, b)$

for $i = 1$ to k **do**

$\mathcal{A}$ outputs two neighboring databases $D^{i,0}, D^{i,1} \in \mathbf{N}^{|\mathcal{X}|}$ and a mechanism $\mathcal{M}_i \in \mathcal{M}$.

$\mathcal{A}$ receives $y_i = \mathcal{M}_i(D^{i,b})$.

end for

With respect to a family of database access mechanisms $\mathcal{M}$ and an adversary $\mathcal{A}$, we say that the k-fold composition experiment b is: **Compose**$(\mathcal{A}, \mathcal{M}, k, b)$.

We say that the *view* of the adversary $\mathcal{A}$ of k-fold adaptive composition experiment b, denoted V^b, is the sequence $y_1, \ldots, y_k$ together with the outcome of any internal randomness of $\mathcal{A}$ (Note that this is enough to reproduce every step of the experiment):

$$V^b = (R^b, Y_1^b, \ldots, Y_k^b)$$

We consider V^b to be the output of the composition experiment, and require that the experiment be differentially private with respect to b. For intuition, imagine that $D^{i,1}$ is always a database that contains a certain individual Alice's data, and $D^{i,0}$ is the neighboring database identical to $D^{i,1}$ except in that Alice's data has been removed. Our requirement will be that intuitively, the adversary "can't tell" which experiment was run.

We start with a warm-up theorem.

Theorem 1 *The class of ϵ-differentially private mechanisms satisfies $k\epsilon$-differential privacy under k-fold adaptive composition.*

Proof A view of the adversary is a tuple $v = (r, y_1, \ldots, y_k)$. We have:

$$\begin{aligned}\frac{\Pr[V^0 = v]}{\Pr[V^1 = v]} &= \left(\frac{\Pr[R^0 = r]}{\Pr[R^1 = r]}\right) \cdot \prod_{i=1}^{k} \frac{\Pr[Y_i^0 = y_i | Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}]}{\Pr[Y_i^1 = y_i | Y_1^1 = y_1, \ldots, Y_{i-1}^1 = y_{i-1}]} \\ &\leq \prod_{i=1}^{k} \exp(\epsilon) \\ &= \exp(k\epsilon)\end{aligned}$$

■

i.e. "The ϵ's add up". The combination of k differentially private algorithms that are each ϵ_i-differentially private is $\sum_{i=1}^{k} \epsilon_i$-differentially private. An immediate consequence is that we can use the Laplace mechanism with adaptively chosen queries.

We'll now prove a more interesting theorem, showing how to trade a little bit of δ for a lot of ϵ. It will be helpful to build some notation first, and introduce a useful theorem from probability.

Definition 2 (Max Divergence) *The Max Divergence between two random variables Y and Z taking values from the same domain is defined to be:*

$$D_\infty(Y||Z) = \max_{S \subset Supp(Y)} \left[\ln \frac{\Pr[Y \in S]}{\Pr[Z \in S]}\right]$$

The δ-Approximate Max Divergence between Y and Z is defined to be:

$$D_\infty^\delta(Y||Z) = \max_{S \subset Supp(Y): \Pr[Y \in S] \geq \delta} \left[\ln \frac{\Pr[Y \in S] - \delta}{\Pr[Z \in S]}\right]$$

Remark Note that a mechanism $\mathcal{M}$ is ϵ-differentially private if and only if on every two neighboring databases D and D', $D_\infty(\mathcal{M}(D)||\mathcal{M}(D')) \leq \epsilon$ and $D_\infty(\mathcal{M}(D')||\mathcal{M}(D)) \leq \epsilon$, and is (ϵ, δ)-differentially private if and only if on every two neighboring databases D, D': $D_\infty^\delta(\mathcal{M}(D)||\mathcal{M}(D')) \leq \epsilon$ and $D_\infty^\delta(\mathcal{M}(D')||\mathcal{M}(D)) \leq \epsilon$. We can also define an average case notion of divergence:

Definition 3 (KL-Divergence) *The KL-Divergence or Relative Entropy between two random variables Y and Z taking values from the same domain is defined to be:*

$$D(Y||Z) = \mathrm{E}_{y \sim Y}\left[\ln \frac{\Pr[Y = y]}{\Pr[Z = y]}\right]$$

Using these definitions, we can rephrase privacy under composition:

Definition 4 (Privacy Under Composition) *A class of mechanisms $\mathcal{M}$ satisfies ϵ-differential privacy under k-fold adaptive composition, if for every adversary $\mathcal{A}$, $D_\infty(V^0||V^1) \leq \epsilon$. It satisfies (ϵ, δ)-differential privacy under k-fold adaptive composition if for every adversary $\mathcal{A}$, $D_\infty^\delta(V^0||V^1) \leq \epsilon$*

The following probabilistic theorem is very useful whenever dealing with randomness:

Theorem 5 (Azuma's Inequality) *Let f be a function of m random variables $X_1, \ldots, X_m$, each X_i taking values from a set A_i such that $\mathrm{E}[f]$ is bounded. Let c_i denote the maximum effect of X_i on f – i.e. for all $a_i, a_i' \in A_i$:*

$$|\mathrm{E}[f|X_1, \ldots, X_{i-1}, X_i = a_i] - \mathrm{E}[f|X_1, \ldots, X_{i-1}, X_i = a_i']| \leq c_i$$

Then:

$$\Pr\left[f(X_1, \ldots, X_m) \geq \mathrm{E}[f] + t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^m c_i^2}\right)$$

We can now proof our more sophisticated composition theorem:

Theorem 6 *Let $\epsilon, \delta' \geq 0$. The class of ϵ-differentially private mechanisms satisfies (ϵ', δ')-differential privacy under k-fold adaptive composition for:*

$$\epsilon' = \sqrt{2k\ln(1/\delta')}\epsilon + k\epsilon(e^\epsilon - 1)$$

The next lemma which will be important in proving Theorem 6 says that if the *maximum* privacy loss is bounded by $\exp(\epsilon)$, then the *expected* privacy loss is actually quite a bit lower. Together with Azuma's inequality, this will allow us to prove a stronger composition theorem: except with small probability δ, the total privacy loss of a k-fold composition is not much more than the expected privacy loss of that composition, which scales more like $\sqrt{k}$ than like k.

Lemma 7 *Suppose that random variables Y and Z satisfy $D_\infty(Y||Z) \leq \epsilon$ and $D_\infty(Z||Y) \leq \epsilon$. Then $D(Y||Z) \leq \epsilon(e^\epsilon - 1)$.*

Proof We know that for any Y and Z it is the case that $D(Y||Z) \geq 0$ (Relative entropy is non-negative – look up the log-sum inequality!), and so it suffices to bound $D(Y||Z) + D(Z||Y)$. We get:

$$\begin{aligned}
D(Y||Z) &\leq D(Y||Z) + D(Z||Y) \\
&= \sum_y \Pr[Y = y] \cdot \left(\ln \frac{\Pr[Y = y]}{\Pr[Z = y]} + \ln \frac{\Pr[Z = y]}{\Pr[Y = y]}\right) \\
&\quad + (\Pr[Z = y] - \Pr[Y = y]) \cdot \left(\ln \frac{\Pr[Z = y]}{\Pr[Y = y]}\right) \\
&\leq \sum_y [0 + |\Pr[Z = y] - \Pr[Y = y]| \cdot \epsilon] \\
&= \epsilon \cdot \sum_y [\max\{\Pr[Y = y], \Pr[Z = y]\} - \min\{\Pr[Y = y], \Pr[Z = y]\}] \\
&\leq \epsilon \cdot \sum_y [(e^\epsilon - 1) \cdot \min\{\Pr[Y = y], \Pr[Z = y]\}] \\
&\leq \epsilon \cdot (e^\epsilon - 1).
\end{aligned}$$

■

We can now finish the proof of the theorem. The idea is that the expected privacy loss after k differentially private algorithms are run is bounded by the above lemma, and that with high probability, the total privacy loss is not much higher (By Azuma's inequality).

Proof [Proof of Theorem 6] A view of the adversary A consists of a tuple of the form $v = (r, y_1, \ldots, y_k)$, where r is the internal randomness of A and $y_1, \ldots, y_k$ are the outputs of the mechanisms $\mathcal{M}_1, \ldots, \mathcal{M}_k$. Let

$$B = \{v : \Pr[V^0 = v] > e^{\epsilon'} \cdot \Pr[V^1 = v]\}.$$

We will show that $\Pr[V^0 \in B] \leq \delta$, and hence for every set S, we have

$$\Pr[V^0 \in S] \leq \Pr[V^0 \in B] + \Pr[V^0 \in (S \setminus B)] \leq \delta + e^{\epsilon'} \cdot \Pr[V^1 \in S].$$

This is equivalent to saying that $D_\infty^\delta(V^0||V^1) \leq \epsilon'$.

It remains to show $\Pr[V^0 \in B] \leq \delta$. Let random variable $V^0 = (R^0, Y_1^0, \ldots, Y_k^0)$ denote the view of A in Experiment 0 and $V^1 = (R^1, Y_1^1, \ldots, Y_k^1)$ the view of A in Experiment 1. Then for a fixed view $v = (r, y_1, \ldots, y_k)$, we have

$$\begin{aligned}
\ln\left(\frac{\Pr[V^0 = v]}{\Pr[V^1 = v]}\right) &= \ln\left(\frac{\Pr[R^0 = r]}{\Pr[R^1 = r]} \cdot \prod_{i=1}^{k} \frac{\Pr[Y_i^0 = y_i | R^0 = r, Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}]}{\Pr[Y_i^1 = y_i | R^1 = r, Y_1^1 = y_1, \ldots, Y_{i-1}^1 = y_{i-1}]}\right) \\
&= \sum_{i=1}^{k} \ln\left(\frac{\Pr[Y_i^0 = y_i | R^0 = r, Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}]}{\Pr[Y_i^1 = y_i | R^1 = r, Y_1^1 = y_1, \ldots, Y_{i-1}^1 = y_{i-1}]}\right) \\
&\stackrel{\text{def}}{=} \sum_{i=1}^{k} c_i(r, y_1, \ldots, y_i).
\end{aligned}$$

Now for every prefix $(r, y_1, \ldots, y_{i-1})$ we condition on $R^0 = r, Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}$, and analyze the expectation and maximum possible value of the random variable $c_i(R^0, Y_1^0, \ldots, Y_i^0) = c_i(r, y_1, \ldots, y_{i-1}, Y_i^0)$, . Once the prefix is fixed, the next pair of databases $D^{i,0}$ and $D^{i,1}$ are also determined (in both Experiment 0 and 1). Thus Y_i^0 is distributed according to $\mathcal{M}_i(D^{i,0})$. Moreover for any value y_i, we have

$$c_i(r, y_1, \ldots, y_{i-1}, y_i) = \ln\left(\frac{\Pr[\mathcal{M}_i(D^{i,0}) = y_i]}{\Pr[\mathcal{M}_i(D^{i,1}) = y_i]}\right).$$

By ϵ-differential privacy this is bounded by ϵ. We can also reason as follows:

$$|c_i(r, y_1, \ldots, y_{i-1}, y_i)| \leq \max\left\{D_\infty(\mathcal{M}_i(D^{i,0})||\mathcal{M}_i(D^{i,1})), D_\infty(\mathcal{M}_i(D^{i,1})||\mathcal{M}_i(D^{i,0}))\right\} = \epsilon.$$

By Lemma 7, we have:

$$\mathrm{E}\left[c_i(R^0, Y_1^0, \ldots, Y_i^0) | R^0 = r, Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}\right] = D(\mathcal{M}_i(D^{i,0})||\mathcal{M}_i(D^{i,1})) \leq \epsilon \cdot (e^\epsilon - 1).$$

Thus we can apply Azuma's Inequality (Theorem 5) to the random variables $C_i = c_i(R^0, Y_1^0, \ldots, Y_i^0)$, letting $f(C_1, \ldots, C_k) = \sum_{i=1}^k C_i$, and noting that that $\mathrm{E}[f] = k\epsilon \cdot (e^\epsilon - 1)$. Taking $t = \sqrt{2k\log(1/\delta)}\epsilon$, we find:

$$\Pr[V^0 \in B] = \Pr[f(C_1, \ldots, C_k) \geq \mathrm{E}[f] + \sqrt{2k\log(1/\delta)}\epsilon] \leq \delta$$

as desired. ■

Bibliographic Information The model of adaptive composition, and the composition theorem presented here are from "Boosting and Differential Privacy" by Dwork, Rothblum, and Vadhan, 2010.