CS 860: Algorithms for Private Data Analysis Fall 2020

Lecture 6 — Advanced Composition

Prof. Gautam Kamath *Scribe: Gautam Kamath*

Today, we finally study the celebrated advanced composition theorem, originally due to Dwork, Rothblum, and Vadhan [DRV10]. Presentation today is based very heavily upon lecture notes of Adam Smith [Smi17] (edited mostly to add in my own little notes and comments), which is in turn based on the work of Kairouz, Oh, and Viswanath [KOV15], see also Section 3.5 of [DR14].

In particular, we prove the following theorem.

Theorem 1 (Advanced Composition)**.** *For all $\varepsilon, \delta, \delta' > 0$, let $M = (M_1, \ldots, M_k)$ be a sequence of (ε, δ)-differentially private algorithms, where the M_i's are potentially chosen sequentially and adaptively. Then M is $(\tilde{\varepsilon}, \tilde{\delta})$-differentially private, where $\tilde{\varepsilon} = \varepsilon\sqrt{2k\log(1/\delta')} + k\varepsilon\frac{e^\varepsilon - 1}{e^\varepsilon + 1}$ and $\tilde{\delta} = k\delta + \delta'$.*

There's a few points to discuss in this theorem. The second term in $\tilde{\varepsilon}$ might seem unusual, but if we are in the "high privacy" regime where ε is small, then $\frac{e^\varepsilon - 1}{e^\varepsilon + 1} \approx \varepsilon/2$, and thus the second term is $\approx k\varepsilon^2/2$, which may be considered as a lower order term for small ε. Also, we have a knob to tweak via the δ' – the more we increase it, the smaller the multiplicative factor in $\tilde{\varepsilon}$ becomes. Coarsely and informally speaking, one can think of this as guarantee as saying, if we wish to do k queries under ε-differential privacy, the cost is a factor of $\sqrt{k}$, rather than k.

As one suggested exercise: revisit the multivariate mean estimation example from the previous lecture. As we showed, the ℓ_2 error by using the Laplace mechanism and Gaussian mechanism was $d^{3/2}/\varepsilon n$ and $d/\varepsilon n$, respectively. Convince yourself that one could use advanced composition and the Laplace mechanism to acheive roughly the latter error $d/\varepsilon n$, albeit at the cost of relaxing to approximate differential privacy.

We will prove this theorem via sequence of two steps. First, we will reduce from general mechanisms to roughly binary ones (more precisely: algorithms which either output a 0, 1, or a blatant privacy violation). We then analyze the composition of mechanisms in this form.

Reduction to Binary-ish Mechanisms

What is the "simplest" pair of random variables that satisfy the style of "(ε, δ)-DP-esque" guarantees we would like? More precisely: what U, V have the property that $|L_{U||V}| \leq \varepsilon$ with probability $1 - \delta$ and $|L_{V||U}| \leq \varepsilon$ with probability $1 - \delta$? Recall that the privacy loss random variable between A and B is defined by drawing $t \sim A$ and outputting

$$L_{A||B} = \ln\left(\frac{\Pr[A = t]}{\Pr[B = t]}\right).$$

One candidate pair of random variables is the following, which has type "types" of outcomes: one to capture "catastrophic failure" with probability δ, and one for the "status quo" e^ε multiplicative guarantee. For the former, we will have outcomes that say "I am U" and "I am V," and for the latter we simply use bits 0 and 1. The random variables can be described in the following table:

t	$\Pr[U = t]$	$\Pr[V = t]$
0	$\frac{e^\varepsilon(1-\delta)}{1+e^\varepsilon}$	$\frac{(1-\delta)}{1+e^\varepsilon}$
1	$\frac{(1-\delta)}{1+e^\varepsilon}$	$\frac{e^\varepsilon(1-\delta)}{1+e^\varepsilon}$
I am U	δ	0
I am V	0	δ

Now, we claim that this "simple" pair of random variables is sufficient to express any pair of random variables with a bounded privacy loss.

Theorem 2. *Let A and B be random variables such that $|L_{A||B}| \leq \varepsilon$ with probability $1-\delta$ and $|L_{B||A}| \leq \varepsilon$ with probability $1-\delta$. Then there exists a randomized mapping F such that $F(U) \sim A$ and $F(V) \sim B$.*

Essentially, this says that we can first sample from one of the simple distributions (either U or V) and map this into a sample from one of the more complicated distributions. We do not prove this here, but claim that it follows by appropriately mapping the four outcomes to the corresponding subsets of the domain (in some non-uniform way that depends on A and B). Roughly, the outcomes "I am U," 0, 1, and "I am V" are mapped to points x where $L(x) = \frac{\Pr[A=x]}{\Pr[B=x]}$ is in the following ranges, respectively: $L(x) > e^\varepsilon$, $0 \leq L(x) \leq e^\varepsilon$, $-e^\varepsilon \leq L(x) \leq 0$, and $L(x) < e^{-\varepsilon}$. But with this reduction in place, it essentially suffices to examine the behaviour of the privacy loss between U and V, and the same will be implied for A and B via post-processing.

Now that we can reduce a pair of random variables to studying this simple case, let's examine how to reduce from a sequence of differentially private algorithms to a sequence of simple pairs of random variables. The algorithms M_1 through M_k each have domain $\mathcal{X}^n$. Since they can be chosen adaptively, they implicitly also take in a "transcript" of results so far: that is, M_j takes as input the dataset $\in \mathcal{X}^n$ as well as a_1^{j-1}, which is a vector representing the results of computations M_1 through M_{j-1}. Each M_j is (ε, δ)-DP, for any possibility of the transcript a_1^{j-1}. We also fix two neighbouring datasets X, X'.

The idea is that, at each step of the process, we apply the mapping implied by the theorem above. Suppose we have a sequence $z_1, \ldots, z_k$ which is generated i.i.d. from either U or V. For each j in $\{1, \ldots, k\}$, we apply Theorem 2, where the inputs are $M_j(X)$ and $M_j(X')$ (with the transcript a_1^{j-1} in both cases as well). This will give us a function F_j, such that $M_j(X) \sim F_j(U)$ and $M_j(X') \sim F_j(V)$. Applying this sequentially gives us the following theorem, where F^* is the sequence of functions $F_1, \ldots, F_k$.

Theorem 3. *There exists a randomized mapping F^* such that the algorithm M defined by the composition of $M_1, \ldots, M_k$ has the following two properties:*

- *$M(X) \sim F^*(U_1, \ldots, U_k)$, where $U_1, \ldots, U_k$ are drawn i.i.d. from U*
- *$M(X') \sim F^*(V_1, \ldots, V_k)$, where $V_1, \ldots, V_k$ are drawn i.i.d. from V*

At this point, if we can prove that the privacy loss random variable between $(U_1, \ldots, U_k)$ and $(V_1, \ldots, V_k)$ is appropriately bounded, then the post-processing property of differential privacy (i.e., when we apply F^*) will imply the same guarantee for $M(X)$ and $M(X')$.

Composition of Binary-ish Mechanisms

At this point, we are left to prove that the absolute value of the privacy loss random variable between the sequence $(U_1, \ldots, U_k)$ and $(V_1, \ldots, V_k)$ is bounded by $\tilde{\varepsilon}$ with probability $1 - \tilde{\delta}$. Let z_j be the realization of the jth random variable in this sequence – that is, we can imagine that each $z_j \sim U$ independently. We will define two events as we go, which will help us partition the outcomes into different "cases". Since we're trying to bound the privacy loss random variable, let's first discount the cases where it is "obviously" unbounded. More precisely, if any of the z_j is "I am U," then this is an "obvious" privacy violation, which we want to ignore as we proceed. One can think of these as contributing the $k\delta$ term to the expression of $\tilde{\delta}$.

$$E_1 = \{z : \text{at least one } z_j \text{ is "I am } U\text{"}\}.$$

We can see that $\Pr[E_1] = 1 - (1-\delta)^k \leq \delta k$.

For essentially the remainder of the proof, we're going to condition on $\bar{E}_1$, to avoid the privacy loss random variable taking an infinite value. Suppose we have some string of outcomes $z \in \{0,1\}^k$, which leads to the following realization of the privacy loss random variable:

$$\ln\left(\frac{\Pr[(U_1,\ldots,U_k)=z]}{\Pr[(V_1,\ldots,V_k)=z]}\right) = \sum_{j=1}^{k} \ln\left(\frac{\Pr[U_j=z_j]}{\Pr[V_j=z_j]}\right) = \sum_{j=1}^{k} \ln\left(\frac{(1-\delta)e^{\varepsilon(1-z_j)}/(e^\varepsilon+1)}{(1-\delta)e^{\varepsilon z_j}/(e^\varepsilon+1)}\right) = \sum_{j=1}^{k} \varepsilon(1-2z_j).$$

The first equality holds because the z_j's are independent. Note that the privacy loss random variable is a sum of independent random variables, which take the value either ε or $-\varepsilon$. This is very convenient for use of the Chernoff bound.

Let's start by computing the expectation of this random variable, recalling that we are conditioning on the event $\bar{E}_1$ to ensure $z \in \{0,1\}^k$.

$$\mathbf{E}_{z\sim(U_1,\ldots,U_k)}\left[\ln\left(\frac{\Pr[(U_1,\ldots,U_k)=z]}{\Pr[(V_1,\ldots,V_k)=z]}\right)|\bar{E}_1\right] = k\varepsilon \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}.$$

To see this, observe that if we condition on $\bar{E}_1$, then $(1-2z_j)$ is 1 with probability $\frac{e^\varepsilon}{1+e^\varepsilon}$ and -1 with probability $\frac{1}{1+e^\varepsilon}$ (easily derivable using, say, Bayes rule). We note that this term is one of the terms in $\tilde{\varepsilon}$ – the other term will express "how far past the mean" the privacy loss random variable goes.

In fact, we let E_2 denote the event that the privacy loss random variable goes "too far". Let $t > 0$ be some parameter we will specify later:

$$E_2 = \left\{z \in \{0,1\}^k : \ln\left(\frac{\Pr[(U_1,\ldots,U_k)=z]}{\Pr[(V_1,\ldots,V_k)=z]}\right) > k\varepsilon\frac{e^\varepsilon-1}{e^\varepsilon+1} + t\varepsilon\sqrt{k}\right\}.$$

Observe that the right-hand side of this inequality is $\tilde{\varepsilon}$, up to the specification of t.

We can use the following Chernoff bound for independent random variables $Z_1, \ldots, Z_k$ in the range $[\ell, u]$:

$$\Pr\left[\sum_{j=1}^{k} Z_j \geq \sum \mathbf{E}[Z_j] + \tau\right] \leq \exp\left(-\frac{2\tau^2}{k(u-\ell)^2}\right).$$

In our context, we set $[\ell, u] = [-\varepsilon, \varepsilon]$ and $\tau = t\varepsilon\sqrt{k}$, giving

$$\Pr\left[E_2|\bar{E}_1\right] \leq \exp\left(-\frac{t^2}{2}\right).$$

At this point, one can observe we "have all the pieces" – we see that if neither E_2 nor E_1 holds, then we have a bounded privacy loss. Furthermore, E_1 and E_2 are both low probability events. Let us assemble these components in a more rigorous manner.

We first note that $\bar{E}_1$ and $\bar{E}_2$ both hold, then the privacy loss random variable is bounded, giving the following for any set of outcomes z:

$$\Pr[U = z \cap \bar{E}_1 \cap \bar{E}_2] \leq e^{\tilde{\varepsilon}} \Pr[V = z \cap \bar{E}_1 \cap \bar{E}_2] \leq e^{\tilde{\varepsilon}} \Pr[V = z].$$

The latter inequality is because we require fewer events to occur.

With this in hand,

$$\begin{aligned}
\Pr[U = z] &= \Pr[U = z \cap \bar{E}_1 \cap \bar{E}_2] + \Pr[U = z \cap E_1] + \Pr[U = z \cap \bar{E}_1 \cap E_2] \\
&\leq \Pr[U = z \cap \bar{E}_1 \cap \bar{E}_2] + \Pr[E_1] + \Pr[E_2|\bar{E}_1] \Pr[\bar{E}_1] \\
&\leq e^{\tilde{\varepsilon}} \Pr[V = z] + k\delta + \exp(-t^2/2) \cdot 1.
\end{aligned}$$

The first equality is by the law of total probability (you might find it helpful to draw the Venn diagram of outcomes). The first inequality again requires fewer events to occur. Setting $t = \sqrt{2\ln(1/\delta')}$ completes the proof.

References

[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.

[DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil Vadhan. Boosting and differential privacy. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, FOCS '10, pages 51–60, Washington, DC, USA, 2010. IEEE Computer Society.

[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.

[Smi17] Adam Smith. The algorithmic foundations of adaptive data analysis - lecture 13: Strong composition. https://adaptivedataanalysis.files.wordpress.com/2017/11/lect13.pdf, 2017.