Advanced Composition:

what if allow (ε,δ)-DP?

- Basic composition still holds: if A_i is (ε_i,δ_i)-DP, get total alg is $\left(\sum_{i=1}^{k}\varepsilon_i, \sum_{i=1}^{k}\delta_i\right)$-DP. (exercise). $(k\varepsilon, k\delta)$ if all are (ε,δ)-DP

- Can do better: basically $\sqrt{k}\cdot\varepsilon$!

$A_i : D \times R_1 \times \dots \times R_{i-1} \to R_i$

Thm: If each A_i is (ε,δ)-DP, then $\forall\, \delta' > 0$, the composed alg. A is $(\tilde{\varepsilon},\tilde{\delta})$-DP, where

$$\tilde{\varepsilon} = \varepsilon\sqrt{2k\ln\left(\tfrac{1}{\delta'}\right)} + k\varepsilon(e^{\varepsilon}-1), \quad \tilde{\delta} = k\delta + \delta'$$

get a feel for it before starting proof...

So, have an algorithm which runs k (ε,δ)-DP subalgorithms. want overall alg to be $(\tilde{\varepsilon},\tilde{\delta})$-DP. How should we set ε,δ as fn of $\tilde{\varepsilon},\tilde{\delta}$?

Basic composition: $\varepsilon = \frac{\tilde{\varepsilon}}{k}$, $\delta = \frac{\tilde{\delta}}{k}$

Advanced composition: $\delta' = \tilde{\delta}/2 \Rightarrow \delta = \frac{\tilde{\delta}}{2k}$

For ε: For $\varepsilon < 1$, $\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1} \approx \frac{\varepsilon}{2}$

$$\Rightarrow \tilde{\varepsilon} = \Theta\left(\varepsilon\sqrt{k\ln\left(\tfrac{2}{\delta}\right)} + k\varepsilon^2\right)$$

$\varepsilon^2 k \leq \tilde{\varepsilon} \Rightarrow \varepsilon \leq \sqrt{\frac{\tilde{\varepsilon}}{k}}$

$\varepsilon\sqrt{k\ln(1/\delta)} \leq \tilde{\varepsilon} \Rightarrow \varepsilon \leq \frac{\tilde{\varepsilon}}{\sqrt{k\ln 1/\delta}}$

$\sqrt{k}$ cost to composition!

Note: δ could equal $0 \Rightarrow$ set $\delta' = \delta$

still applies to pure DP mechanisms, but turns them into approximate DP!

Start proof. Pretty complex! Going to prove an easier version:

Thm: If each A_i is ε-DP, then $\forall \delta > 0$, the composed alg. A is $(\tilde{\varepsilon}, \delta)$-DP, where

$$\tilde{\varepsilon} = \varepsilon\sqrt{2k\ln\left(\tfrac{1}{\delta}\right)} + k\varepsilon(e^{\varepsilon}-1)$$

Idea: start with ε-DP, but by ending w/ approximate DP we can get better composition.

Q: How can we prove a mechanism is (ε,δ)-DP rather than DP? Can't just prove that $\forall D\sim D'$

$$\Pr[A(D)=r] \le e^{\varepsilon} \Pr[A(D')=r] + \delta \quad \forall r \in R$$ anymore!

Instead: Given $D\sim D'$ let $B^{\varepsilon}_{D,D'} = \{r \in R : \Pr[A(D)=r] > e^{\varepsilon} \Pr[A(D')=r]\}$

$$= \left\{r \in R : \ln\left(\frac{\Pr[A(D)=r]}{\Pr[A(D')=r]}\right) > \varepsilon\right\}$$

"bad" points.

Claim: If $\Pr[A(D) \in B^{\varepsilon}_{D,D'}] \le \delta$ $\forall D\sim D'$, then A is (ε,δ)-DP

Pf: Fix $S \subseteq R$

$$\Pr[A(D)\in S] = \Pr[A(D) \in S \setminus B^{\varepsilon}_{D,D'}] + \Pr[A(D) \in S \cap B^{\varepsilon}_{D,D'}]$$

$$\le e^{\varepsilon} \Pr[A(D') \in S \setminus B^{\varepsilon}_{D,D'}] + \delta$$

$$\le e^{\varepsilon} \Pr[A(D') \in S] + \delta \quad ✓$$

So fix $D\sim D'$, let $B = B^{\tilde{\varepsilon}}_{D,D'}$

WTS: $\Pr[A(D) \in B] = \Pr_{y\sim A(D)}\left[\ln\left(\frac{\Pr[A(D)=y]}{\Pr[A(D')=y]}\right) > \tilde{\varepsilon}\right] \le \delta$

Let $y = (y_1, y_2, \ldots, y_k) \in R_1 \times R_2 \times \cdots \times R_k$ be point in range of A.

Random variables: $Y^D_1 = A_1(D)$, $Y^D_2 = A_2(D, Y_1)$

$\ldots$ $Y^D_k = A_k(D, Y_1, \ldots, Y_{k-1})$

Same for $Y^{D'}_1, Y^{D'}_2, \ldots, Y^{D'}_k$

For $\underline{\text{fixed}}$ y, get

$$\ln\left(\frac{\Pr[A(D)=y]}{\Pr[A(D')=y]}\right) = \ln\left(\frac{\Pr[Y_1^D = y_1]\cdot \Pr[Y_2^D = y_2 \mid Y_1^D = y_1]\cdot \ldots}{\Pr[Y_1^{D'} = y_1]\cdot \Pr[Y_2^{D'} = y_2 \mid Y_1^{D'} = y_1]\ldots}\right)$$

$$= \ln\left(\prod_{i=1}^{k} \frac{\Pr[Y_i^D = y_i \mid Y_1^D = y_1, \ldots, Y_{i-1}^D = y_{i-1}]}{\Pr[Y_i^{D'} = y_i \mid Y_1^{D'} = y_1, \ldots, Y_{i-1}^{D'} = y_{i-1}]}\right)$$

$$= \sum_{i=1}^{k} \underbrace{\ln\left(\frac{\Pr[Y_i^D = y_i \mid Y_1^D = y_1, \ldots, Y_{i-1}^D = y_{i-1}]}{\Pr[Y_i^{D'} = y_i \mid Y_1^{D'} = y_1, \ldots, Y_{i-1}^{D'} = y_{i-1}]}\right)}_{c_i(y_1, \ldots, y_i)}$$

$$= \sum_{i=1}^{k} c_i(y_1, \ldots, y_i) = \sum_{i=1}^{k} \ln\left(\frac{\Pr[A_i(D, y_1, \ldots, y_{i-1}) = y_i]}{\Pr[A_i(D', y_1, \ldots, y_{i-1}) = y_i]}\right)$$

So $\underline{\text{WTS}}$: $\Pr\limits_{y \sim A(D)}\left[\sum_{i=1}^{k} c_i(y_1, \ldots, y_i) > \tilde{\varepsilon}\right] \le \delta$

(Sum of rand. vars, each depends on previous). How to bound sum? Tail bound: expectation, range?

Rewrite: $c_i(y_1, \ldots, y_i) = \ln\left(\frac{\Pr[A(D, y_1, \ldots, y_{i-1}) = y_i]}{\Pr[A(D', y_1, \ldots, y_{i-1}) = y_i]}\right)$

Note: $\le \varepsilon \quad \forall y_1, \ldots, y_i$ by ε-DP of A_i!

So $\forall y$, $\sum_{i=1}^{k} c_i(y_1, \ldots, y_i) \le \varepsilon k$: basic composition!

By symmetry, $c_i(y_1, \ldots, y_i) \geq -\varepsilon \quad \forall y_1, \ldots, y_i$

$\Rightarrow |c_i(y_1, \ldots, y_i)| \leq \varepsilon \quad \forall y$

So each $c_i \in [-\varepsilon, \varepsilon] \quad \forall y$.

Now draw $y \sim A(D)$, think of them as random vars, drawn sequentially.

$$c_1(Y_1) = \ln\left(\frac{\Pr[Y_1^D = y_1]}{\Pr[Y_1^{D'} = y_1]}\right) \quad (Y_1 \sim A_1(D))$$

$$c_2(y_1, Y_2) = \ln\left(\frac{\Pr[Y_2^D = y_2 \mid Y_1^D = y_1]}{\Pr[Y_2^{D'} = y_2 \mid Y_1^{D'} = y_1]}\right) \quad (Y_2 \sim A_2(D, y_1))$$

$\vdots$

$$c_i(y_1, \ldots, y_{i-1}, Y_i) \quad (Y_i \sim A_i(D, y_1, \ldots, y_{i-1}))$$

So have k random vars, <u>not independent</u> (dependent on earlier ones), but bounded no matter what.

<u>Q</u>: what is <u>expectation</u> of ith, i.e., what is

$$\mathop{E}_{Y_i \sim A_i(D, y_1, \ldots, y_{i-1})}[c_i(y_1, \ldots, y_{i-1}, Y_i)] \; ?$$

<u>Def</u>: For two random vars Y, Z with same domain,

$$D_\infty(Y \| Z) = \max_{S \subseteq \mathrm{Supp}(Y)}\left(\ln \frac{\Pr[Y \in S]}{\Pr[Z \in S]}\right) \quad \text{(max-divergence)}$$

$$D(Y \| Z) = \mathop{E}_{y \sim Y} \left[\ln \frac{\Pr[Y = y]}{\Pr[Z = y]} \right] \quad \text{(KL-divergence)}$$

Note: A is ε-DP iff $\forall D \sim D'$,

$$D_\infty(A(D) \| A(D')) \le \varepsilon \quad \text{and}$$
$$D_\infty(A(D') \| A(D)) \le \varepsilon$$

(Lemma 3.18)

Fact: If $D_\infty(Y \| Z) \le \varepsilon$ and $D_\infty(Z \| Y) \le \varepsilon$, then

$$D(Y \| Z) \le \varepsilon \cdot (e^\varepsilon - 1).$$

$$\Rightarrow \mathop{E}_{Y_i \sim A_i(D, y_1, \dots, y_{i-1})} [c_i(y_1, \dots, y_{i-1}, Y_i)] =$$

$$E\left[\ln \left(\frac{\Pr[Y_i^D = y_i \mid Y_1^D = y_1, \dots, Y_{i-1}^D = y_{i-1}]}{\Pr[Y_i^{D'} = y_i \mid Y_1^{D'} = y_1, \dots, Y_{i-1}^{D'} = y_{i-1}]} \right) \right]$$

$$= D(A_i(D, y_1, \dots, y_{i-1}) \| A_i(D', y_1, \dots, y_{i-1}))$$

$$\le \varepsilon \cdot (e^\varepsilon - 1) \quad \text{by fact.}$$

So want to bound sum of k random vars, each depends on what happened earlier, but no matter what happened earlier, expectation $\le \varepsilon \cdot (e^\varepsilon - 1)$, and in $[-\varepsilon, \varepsilon]$

Azuma's inequality: Let $X_1, \dots, X_k$ random vars

s.t. $\forall i \in [k]$, $\Pr[|X_i| \le \alpha] = 1$ and

$\forall$ $(x_1, \dots, x_{i-1}) \in \mathrm{Supp}(X_1, \dots, X_{i-1})$,

$E[X_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}] \le \beta$. Then $\forall z > 0$,

$$\Pr\left[\sum_{i=1}^{k} X_i > k\beta + z\sqrt{k} \cdot \alpha\right] \le e^{-z^2/2}$$

Apply Azuma's ineq. to c_i random vars:

$\alpha = \varepsilon$, $\beta = \varepsilon \cdot (e^{\varepsilon} - 1)$

$z = \sqrt{2\ln(1/\delta)}$

$$\Pr\left[A(D) \in B\right]$$

$$= \Pr_{y \sim A(D)}\left[\sum_{i=1}^{k} c_i(y_1, y_2, \dots y_{i-1}, Y_i) > \tilde{\varepsilon}\right]$$

$\tilde{\varepsilon} \ge k\varepsilon \cdot (e^{\varepsilon} - 1) + \sqrt{2\ln(1/\delta)}\sqrt{k} \cdot \varepsilon$

$$\le e^{-\ln\frac{1}{\delta}} = \delta$$

✓