04.12.2024 ***Adaptive data analysis***

Lecture 5: Azuma–Hoeffding and Composition for DP-Stability

Source: Lecture notes by Aaron Roth and Adam Smith

Lecturer: Uri Stemmer

The composition theorems we saw at the end of the previous class are decent but insufficient for achieving better results than those obtained via compression. Toward proving a stronger composition theorem for stability, we will need to study a generalization of the Chernoff-Hoeffding theorems discussed in the first lecture.

Reminders from the first lecture::

Theorem 0 (Hoeffding's Bound):
Let $a, \mu \in \mathbb{R}$ and let $X_1, X_2, \dots, X_k$ be independent random variables such that for all $i \in [k]$ we have $\Pr[|X_i| \leq a] = 1$ and $\mathbb{E}[X_i] \leq \mu$. Then, for any $z > 0$ it holds that

$$\Pr\left[\sum_{i=1}^{n} X_i \geq k\mu + z \cdot \sqrt{k} \cdot a\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

Toward the proof of a composition theorem for DP-stability, we will need a slightly more general version of the above theorem.

Theorem 1 (Azuma-Hoeffding Inequality):
Let $a, \mu \in \mathbb{R}$ and let $X_1, X_2, \dots, X_k$ be random variables such that for all $i \in [k]$ we have $\Pr[|X_i| \leq a] = 1$ and for any choice of $(x_1, \dots, x_{i-1}) \in \text{Support}(X_1, \dots, X_{i-1})$ it holds that

$$.\mathbb{E}[X_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}] \leq \mu$$

Then, for any $z > 0$ it holds that

$$\Pr\left[\sum_{i=1}^{k} X_i \geq k\mu + z \cdot \sqrt{k} \cdot a\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

Note that Theorem 0 is a special case of Theorem 1. Why? Because if $X_1, X_2, \dots, X_k$ are independent and $\mathbb{E}[X_i] \leq \mu$ then

$$\mathbb{E}[X_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}] = \mathbb{E}[X_i] \leq \mu$$

To prove this theorem, we will need to use the following fact:

Theorem 2 (Hoeffding's Lemma):
Let X be a real-valued random variable such that $\Pr[a \leq X \leq b] = 1$ and $\mathbb{E}[X] \leq \mu$. Then, for any $\lambda \in \mathbb{R}$ it holds that

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp\left(\lambda\mu + \frac{\lambda^2 (b-a)^2}{8}\right)$$

Proof idea of Theorem 2:

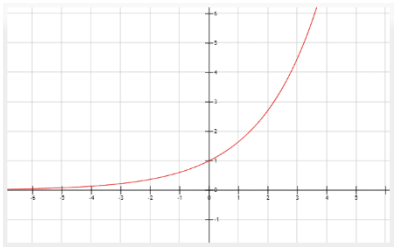
First, note that it suffices to prove the theorem for the case where $\mathbb{E}[X] = 0$.

Why? Assume we have proven the theorem for the case where $\mathbb{E}[X] = 0$. Now let X be a random variable with a nonzero expectation. Define a new random variable $\tilde{X} = X - \mathbb{E}[X]$ and note that $\mathbb{E}[\tilde{X}] = 0$. Therefore,

$$\mathbb{E}\left[e^{\lambda X}\right] = \mathbb{E}\left[e^{\lambda(\tilde{X}+\mathbb{E}[X])}\right] = e^{\lambda\cdot\mathbb{E}[X]} \cdot \mathbb{E}\left[e^{\lambda\tilde{X}}\right] \leq e^{\lambda\cdot\mu} \cdot \exp\left(\frac{\lambda^2(b-a)^2}{8}\right)$$

Thus, it remains to prove the theorem under the assumption that $\mathbb{E}[X] = 0$.

Consider the function $f(y) = e^{\lambda y}$ and note that this is a convex function.



Reminder: By definition, a function $f: \mathbb{R} \to \mathbb{R}$ is convex if for all $y_1, y_2 \in \mathbb{R}$ and for all $t \in [0,1]$ it holds that

$$f\big((1-t)\cdot y_1 + t\cdot y_2\big) \leq (1-t)\cdot f(y_1) + t\cdot f(y_2)$$

Therefore, in our case, for $y_1 = a,\ y_2 = b$ and $t = \frac{y-a}{b-a}$ where $y \in [a,b]$ it holds that

$$e^{\lambda y} = f(y) = f\big((1-t)a + tb\big) \leq (1-t)\cdot f(a) + t\cdot f(b) = \frac{b-y}{b-a}\cdot e^{\lambda a} + \frac{y-a}{b-a}\cdot e^{\lambda b}$$

This holds pointwise for every $y \in [a,b]$.
Now, consider our random variable $X \in [a,b]$. It follows that:

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \mathbb{E}\left[\frac{b-X}{b-a}\cdot e^{\lambda a} + \frac{X-a}{b-a}\cdot e^{\lambda b}\right] =$$

$$= \frac{b-\mathbb{E}[X]}{b-a}\cdot e^{\lambda a} + \frac{\mathbb{E}[X]-a}{b-a}\cdot e^{\lambda b} \underbrace{=}_{\substack{by\ assumption \\ \mathbb{E}[X]=0}} \frac{b}{b-a}\cdot e^{\lambda a} + \frac{-a}{b-a}\cdot e^{\lambda b}$$

Thus, we bounded $\mathbb{E}\left[e^{\lambda X}\right]$ with an expression that does not depend on X (only on a, b, λ)

The continuation of the proof follows from analyzing this expression. It can be shown that this expression is bounded by:

$$\exp\left(\frac{\lambda^2(b-a)^2}{8}\right)$$

q.e.d. (Theorem 2)

Additionally, we will need to recall the law of total expectation, which states that the expectation of the conditional expectation of a random variable equals the expectation of the random variable itself.

Theorem 3 ("Law of total expectation"):

Let X, Y be random variables such that $\mathbb{E}[|X|] < \infty$. Then, $\mathbb{E}[X] = \mathbb{E}\Big[\ \mathbb{E}[X|Y]\ \Big]$

Before recalling the proof of Theorem 3, let's first recall what $\mathbb{E}[\ \mathbb{E}[X|Y]\ ]$ means.

To this end, let's start by recalling what is "conditional expectation": For $y \in \text{Support}(Y)$ we have (in the discrete case):

$$\mathbb{E}[X|Y=y] = \sum_{x \in \text{Supp}(X)} x \cdot \Pr[X=x|Y=y] = \sum_{x \in \text{Supp}(X)} x \cdot \frac{\Pr[X=x,\ Y=y]}{\Pr[Y=y]}$$

Now, the conditional expectation of X w.r.t. Y is a function of the value of Y, defined as follows:

$$\mathbb{E}[X|Y](y) = \mathbb{E}[X|Y=y]$$

This means that $\mathbb{E}[X|Y]$ is a random variable (a function of Y)

So the theorem of total expectation says that the expectation of $\mathbb{E}[X|Y]$ equals the expectation of

A clearer way to phrase this theorem might be $\mathbb{E}[X] = \mathbb{E}_{y \leftarrow Y}\Big[\ \mathbb{E}[X|Y=y]\ \Big]$ which explicitly emphasizes that the outer expectation is over $y \leftarrow Y$. However, the formulation that appears in the theorem, $\mathbb{E}[X] = \mathbb{E}\Big[\ \mathbb{E}[X|Y]\ \Big]$ is the standard form used in probability theory. It's concise, but can sometimes feel less intuitive.

Proof of Theorem 3 (law of total probability) for the discrete case:

$$\mathbb{E}\Big[\ \mathbb{E}[X|Y]\ \Big] = \sum_y \mathbb{E}[X|Y=y] \cdot \Pr[Y=y] = \sum_y \left[\sum_x x \cdot \Pr[X=x|Y=y]\right] \cdot \Pr[Y=y]$$

$$= \sum_y \sum_x x \cdot \Pr[X=x,\ Y=y] = \sum_x \sum_y x \cdot \Pr[X=x,\ Y=y]$$

$$= \sum_x x \sum_y \Pr[X=x,\ Y=y] = \sum_x x \cdot \Pr[X=x] = \mathbb{E}[X]$$

q.e.d. (Theorem 3)

Class Exercise: Given a bag that initially contains R red balls and B blue balls: At each step, draw a ball randomly from the bag (uniformly and independently). Then, return the drawn ball to the bag along with one additional ball of the same color. Thus, after k steps, the bag contains R+B+k balls.

Define random variables X_i where $X_i = 1$ if a red ball is drawn in step i and $X_i = 0$ otherwise. Prove that

$$\mathbb{E}\left[\sum_{i=1}^{k} X_i\right] = k \cdot \frac{R}{R+B}$$

Solution:

We prove by induction that for every i it holds that $\mathbb{E}[X_i] = \frac{R}{R+B}$ and then the claim follows from linearity of expectation.

The base case for i=1 is trivial. Suppose it holds for every $i \leq \ell$. Denote $S_\ell = \sum_{i=1}^{\ell} X_i$. Then,

$$\mathbb{E}[X_{\ell+1}] = \mathbb{E}\left[\mathbb{E}[X_{\ell+1}|S_\ell]\right] = \mathbb{E}\left[\frac{R+S_\ell}{R+B+\ell}\right] = \frac{R+\mathbb{E}[S_\ell]}{R+B+\ell} \underbrace{=}_{induction} \frac{R+\ell\cdot\frac{R}{R+B}}{R+B+\ell} = \frac{R}{R+B}$$

Now we can prove the Azuma-Hoeffding Inequality. As we will see, the proof will resemble the proof of the Chernoff bound that we discussed in the first lecture.

Proof of Theorem 1 (Azuma-Hoeffding Inequality):

Let $a, \mu \in \mathbb{R}$ and let $X_1, X_2, \ldots, X_k$ be random variables such that for all $i \in [k]$ we have $\Pr[|X_i| \leq a] = 1$ and for any choice of $(x_1, \ldots, x_{i-1}) \in \text{Support}(X_1, \ldots, X_{i-1})$ it holds that

$$\mathbb{E}[X_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] \leq \mu.$$

Let $z > 0$. Denote $t = k\mu + z \cdot \sqrt{k} \cdot a$. We need to show that

$$\Pr\left[\sum_{i=1}^{k} X_i \geq t\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

Denote $c = \frac{z}{a\sqrt{k}}$. We calculate:

$$\Pr\left[\sum_{i=1}^{k} X_i \geq t\right] = \Pr\left[c \cdot \sum_{i=1}^{k} X_i \geq c \cdot t\right] = \Pr\left[e^{c\cdot\sum_{i=1}^{k} X_i} \geq e^{c\cdot t}\right] \underbrace{\leq}_{Markov} e^{-ct} \cdot \mathbb{E}\left[e^{c\cdot\sum_{i=1}^{k} X_i}\right] = ((1))$$

In the proof of the Chernoff bound, we used the assumption that the random variables are independent to split the expectation into a product of expectations, and then analyzed each individual expectation. Here, we cannot split into a product of expectations because we do not have independence. What can we use instead? The Law of Total Expectation:

$$((1)) = e^{-ct} \cdot \mathbb{E}\left[\mathbb{E}\left[e^{c\cdot\sum_{i=1}^{k} X_i} \middle| X_1, \ldots, X_{k-1}\right]\right] = ((2))$$

Here, the outer expectation is over the draws of $X_1, \dots, X_{k-1}$ and the inner expectation is over the draw of X_k (from the corresponding conditional space). Note that within the inner expectation $X_1, \dots, X_{k-1}$ are fixed (since we are conditioning on them), so they can be taken outside the inner expectation. Thus, we have:

$$\big((2)\big) = e^{-ct} \cdot \mathbb{E}\left[e^{c \cdot \sum_{i=1}^{k-1} X_i} \cdot \mathbb{E}[e^{c \cdot X_k} | X_1, \dots, X_{k-1}] \right] = \big((3)\big)$$

Now, recall that under our assumptions, for any fixed values of $X_1, \dots, X_{k-1}$ it holds that $\mathbb{E}[X_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}] \leq \mu$. Thus, we can apply Hoeffding's Lemma (Theorem 2) to the inner expectation and obtain:

$$\big((3)\big) \leq e^{-ct} \cdot \mathbb{E}\left[e^{c \cdot \sum_{i=1}^{k-1} X_i} \cdot \exp\left(c \cdot \mu + \frac{c^2 a^2}{2} \right) \right] = e^{-ct} \cdot e^{c\mu} \cdot e^{c^2 a^2/2} \cdot \mathbb{E}\left[e^{c \cdot \sum_{i=1}^{k-1} X_i} \right] = \big((4)\big)$$

So what did we get? We started from $e^{-ct} \cdot \mathbb{E}\left[e^{c \cdot \sum_{i=1}^{k} X_i} \right]$ and managed to eliminate one random variable (X_k) at the cost of accumulating a multiplicative factor of $e^{c\mu} \cdot e^{c^2 a^2/2}$.

By induction,

$$\big((4)\big) \leq e^{-ct} \cdot e^{kc\mu} \cdot e^{kc^2 a^2/2} = \big((5)\big)$$

Plugging our choice of $t = k\mu + z \cdot \sqrt{k} \cdot a$ we get

$$\big((5)\big) = e^{-ck\mu - cz\sqrt{k}a} \cdot e^{kc\mu} \cdot e^{kc^2 a^2/2} = e^{-cz\sqrt{k}a} \cdot e^{kc^2 a^2/2} = \big((6)\big)$$

Plugging our choice of $c = \frac{z}{a\sqrt{k}}$ we get

$$\big((6)\big) = e^{-z^2} \cdot e^{z^2/2} = e^{-z^2/2}$$

q.e.d. (Theorem 1)

<u>A Simple Example of Using the Azuma-Hoeffding Inequality (Non-Tight)</u>

Let's revisit the class exercise with the balls.

Simplifying Assumption: *The initial number of balls $N = R + B$ is much larger than the number of steps k (i.e., the number of times we draw one ball and return two). Additionally, assume $R = B = \frac{N}{2}$.*

Note that the random variables $X_1, \dots, X_k$ defined in the exercise are <u>not</u> independent.

Still, we can apply Azuma's inequality:

- *The random variables are bounded in the interval [0,1]*
- *For any fixed $X_1, \dots, X_i$ we have*

$$\mathbb{E}[X_{i+1} | X_1 = x_1, \dots, X_i = x_i] \leq \frac{R+i}{N+i} \leq \frac{R}{N} + \frac{k}{N} \approx \frac{1}{2}$$

Using Azuma's inequality, we get:

$$\Pr\left[\sum_{i=1}^{k} X_i \gtrsim \frac{k}{2} + z \cdot \sqrt{k}\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

In other words, with high probability $\sum_{i=1}^{k} X_i$ will not be much larger than its expectation, which is $\frac{k}{2}$ (based on the class exercise and the choice $R = B = \frac{N}{2}$).

***Remark:** This result is not tight for this example. A better result can be obtained using a slightly different version of Azuma's inequality (if you're interested, see **Pólya's urn** on Wikipedia).*

How could we avoid the assumption that $N \gg k$? (not presented in class)

We will use the following trick, which is useful much more broadly. Instead of applying Azuma's inequality directly to $\sum_{i=1}^{k} X_i$, we will apply it to an alternative sequence of RV's for which the conditional expectation is easier to control. Specifically, for $\ell \in [k]$ let us define the "partial sum"

$$S_\ell = \sum_{i=1}^{\ell} X_i$$

Our goal is to analyze S_k, but as we will see it will be useful to look also at intermediate S_ℓ's.

- Suppose we have already seen $X_1, X_2, \dots, X_\ell$ (so the partial sum S_ℓ is fixed). The total number of red balls in the bag at this time is $R' = R + S_\ell$ and the total number of blue balls in the bag at this time is $B' = B + \ell - S_\ell$.
- Then, the expected number of red balls seen in the future $(k - \ell)$ steps is

$$(k-\ell)\frac{R'}{R'+B'} = (k-\ell)\frac{R+S_\ell}{N+\ell}$$

Let us define another sequence of RV's: For all $\ell \in \{0,1,\dots,k\}$ define

$$F_\ell = S_\ell + (k-\ell)\frac{R+S_\ell}{N+\ell}$$

- Note that $F_k = S_k$ (which is the RV we want to analyze) and $F_0 = k\frac{R}{N}$ (which is the expectation of F_k).
- With these notations, our goal is to show that w.h.p. $|F_k - F_0|$ is small

To this end, let us define yet another sequence of RV's: For all $\ell \in \{1,2,\dots,k\}$ define

$$\Delta F_\ell = F_\ell - F_{\ell-1}$$

- We are going to apply Azuma's inequality to bound $\sum_{\ell=1}^{k} \Delta F_\ell = F_k - F_0$
- For this, we need to show that (1) the RV's ΔF_ℓ are bounded, and (2) we need to analyze their conditional expectation.

The RV's ΔF_ℓ are bounded:

$$|\Delta F_\ell| = |F_\ell - F_{\ell-1}| = \left|S_\ell + (k-\ell)\frac{R+S_\ell}{N+\ell} - S_{\ell-1} - (k-\ell+1)\frac{R+S_{\ell-1}}{N+\ell-1}\right|$$

$$\leq |S_\ell - S_{\ell-1}| + \left|\frac{R+S_{\ell-1}}{N+\ell-1}\right| + \left|(k-\ell)\frac{R+S_\ell}{N+\ell} - (k-\ell)\frac{R+S_{\ell-1}}{N+\ell-1}\right|$$

$$\leq 2 + (k-\ell)\left|\frac{R+S_\ell}{N+\ell} - \frac{R+S_{\ell-1}}{N+\ell-1}\right|$$

$$= 2 + (k-\ell)\left|\frac{R+S_{\ell-1}+X_\ell}{N+\ell} - \frac{R+S_{\ell-1}}{N+\ell-1}\right|$$

$$= 2 + (k-\ell)\left|\frac{(R+S_{\ell-1}+X_\ell)\cdot(N+\ell-1) - (R+S_{\ell-1})\cdot(N+\ell)}{(N+\ell)\cdot(N+\ell-1)}\right|$$

$$= 2 + (k-\ell)\left|\frac{-(R+S_{\ell-1}) + X_\ell\cdot(N+\ell-1)}{(N+\ell)\cdot(N+\ell-1)}\right|$$

$$\leq 2 + (k-\ell)\left|\frac{1}{(N+\ell)}\right| = 2 + \frac{k-\ell}{N+\ell} \leq 2 + \frac{k}{N}$$

Remark: Note that there is something wasteful here: while it's true that every $\boldsymbol{\Delta F_\ell}$ is bounded by $\left(2+\frac{k}{N}\right)$, most of them are bounded by something noticeably smaller (because our bound decreases with ℓ). A more fine-tuned version of Azuma would allow us to take advantage of this, thereby improving the resulting bound.

Conditional expectation:

$$\mathbb{E}[\Delta F_\ell | \Delta F_1 = f_1, \dots, \Delta F_{\ell-1} = f_{\ell-1}] = \mathbb{E}[F_\ell - F_{\ell-1} | (F_1 - F_0) = f_1, \dots, (F_{\ell-1} - F_{\ell-2}) = f_{\ell-1}] = \big((1)\big)$$

Note that, in particular, this conditioning fixes $F_{\ell-1}$ because

$$f_1 + f_2 + \cdots + f_{\ell-1} = (F_1 - F_0) + (F_2 - F_1) + \cdots + (F_{\ell-1} - F_{\ell-2}) = F_{\ell-1} - F_0 = F_{\ell-1} - k\frac{R}{N}$$

Thus, this conditioning also fixes $S_{\ell-1}$ via the equality

$$F_{\ell-1} = S_{\ell-1} + (k-\ell+1)\frac{R+S_{\ell-1}}{N+\ell-1}$$

So,

$$\big((1)\big) = \mathbb{E}[F_\ell - F_{\ell-1} | S_{\ell-1} = s_{\ell-1}]$$

$$= \mathbb{E}\left[\left(S_\ell + (k-\ell)\frac{R+S_\ell}{N+\ell}\right) - \left(S_{\ell-1} + (k-\ell+1)\frac{R+S_{\ell-1}}{N+\ell-1}\right) \middle| S_{\ell-1} = s_{\ell-1}\right]$$

$$= \mathbb{E}[S_\ell | s_{\ell-1}] + (k-\ell)\frac{R+\mathbb{E}[S_\ell | s_{\ell-1}]}{N+\ell} - s_{\ell-1} - (k-\ell+1)\frac{R+s_{\ell-1}}{N+\ell-1}$$

$$= \left(s_{\ell-1} + \frac{R + s_{\ell-1}}{N + \ell - 1}\right) + (k - \ell)\frac{R + \left(s_{\ell-1} + \frac{R + s_{\ell-1}}{N + \ell - 1}\right)}{N + \ell} - s_{\ell-1} - (k - \ell + 1)\frac{R + s_{\ell-1}}{N + \ell - 1}$$

$$= (k - \ell)\frac{R + s_{\ell-1} + \frac{R + s_{\ell-1}}{N + \ell - 1}}{N + \ell} - (k - \ell)\frac{R + s_{\ell-1}}{N + \ell - 1}$$

$$= (k - \ell) \cdot \frac{(N + \ell - 1)(R + s_{\ell-1}) + R + s_{\ell-1} - (N + \ell)(R + s_{\ell-1})}{N + \ell} = 0$$

Therefore, by Azuma's inequality, for any $z > 0$ it holds that

$$\Pr\left[F_k - F_0 \geq z \cdot \sqrt{k} \cdot \left(2 + \frac{k}{N}\right)\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

That is,

$$\Pr\left[S_k - k\frac{R}{N} \geq z\left(2\sqrt{k} + \frac{k^{1.5}}{N}\right)\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

So, as long as $k \leq N$, w.h.p. it holds that S_k does not exceed its expectation by mora than $\approx \sqrt{k}$. The bound can be improved to have roughly $\left(\sqrt{k} + \frac{k}{N}\right)$ instead of $\left(\sqrt{k} + \frac{k^{1.5}}{N}\right)$, see the green remark above.

COMPOSITION FOR DP-STABILITY

Suppose we have k mechanisms $M_1, M_2, \ldots, M_k$, each of which is (ε, δ)-DP-stable (each receives a database and a parameter). Consider the mechanism $\vec{M}$ that performs an **adaptive composition** of these mechanisms, meaning:

$$\vec{M}(S) = M_k\left(S, M_{k-1}\left(S, M_{k-2}\left(S, M_{k-3}(S, \ldots)\right)\right)\right)$$

Theorem 1: Let $0 < \varepsilon, \delta \leq 1$, and let $\vec{M}$ be a mechanism that performs k adaptive executions of mechanisms, each of which is (ε, δ)-DP-stable (without additional access to the database). Then, $\vec{M}$ is:

$$\left(2k \cdot \varepsilon^2 + \sqrt{2k \ln\left(\frac{1}{k\delta}\right)} \cdot \varepsilon \;\;,\;\; 2k\delta\right)\text{–DP-stable}$$

For simplicity, we will prove the following weaker version of the composition theorem:

Theorem 2: Let $0 < \varepsilon, \delta \leq 1$, and let $\vec{M}$ be a mechanism that performs k adaptive executions of mechanisms, each of which is $(\varepsilon, \mathbf{0})$-DP-stable (without additional access to the database). Then, $\vec{M}$ is $(\hat{\varepsilon}, \delta)$-DP-stable for

$$\hat{\varepsilon} = 2k\varepsilon^2 + \sqrt{2k \cdot \ln(1/\delta)}\,\varepsilon$$

Proof:

Fix neighboring datasets S, S'. We need to show that for any event B it holds that

$$\Pr\left[\vec{M}(S) \in B\right] \leq e^{\hat{\varepsilon}} \cdot \Pr\left[\vec{M}(S') \in B\right] + \hat{\delta}$$

Denote:

- $Y_1, \ldots, Y_k$ = *Random variables representing the outputs of* $M_1, \ldots, M_k$ *during the execution of* $\vec{M}(S)$
- $Y'_1, \ldots, Y'_k$ = *Random variables representing the outputs of* $M_1, \ldots, M_k$ *during the execution of* $\vec{M}(S')$
-

Also denote

$$V = (Y_1, \ldots, Y_k) \quad , \quad V' = (Y'_1, \ldots, Y'_k)$$

With these notations, it suffices to show that for any event B it holds that

$$\Pr[V \in B] \leq e^{\hat{\varepsilon}} \cdot \Pr[V' \in B] + \hat{\delta}$$

In other words, in order to show that $Y_k \approx_{(\varepsilon,\delta)} Y'_k$*, we are actually going to show something stronger: We will show that* $(Y_1, \ldots, Y_k) \approx_{(\varepsilon,\delta)} (Y'_1, \ldots, Y'_k)$

Auxiliary Claim 3: *It suffices to prove that:*

$$\Pr_{v \leftarrow V}\left[\ \ln\left(\frac{\Pr[V = v]}{\Pr[V' = v]}\right) > \hat{\varepsilon}\ \right] < \delta$$

Proof of Auxiliary Claim 3:

Define

$$W = \left\{ v \in \mathrm{Supp}(V) \ : \ \ln\left(\frac{\Pr[V = v]}{\Pr[V' = v]}\right) > \hat{\varepsilon} \right\}$$

By the conditions of the claim,

$$\Pr[V \in W] < \delta$$

Fix a set B*. Now,*

$$\Pr[V \in B] = \Pr[V \in B \cap W] + \Pr[V \in B \setminus W] \leq \delta + e^{\hat{\varepsilon}} \cdot \Pr[V' \in B \setminus W] \leq \delta + e^{\hat{\varepsilon}} \cdot \Pr[V' \in B]$$

q.e.d. (aux. claim 3)

Let's return to proving Theorem 2. According to Claim 3, it suffices to show that when sampling $v \leftarrow V$ then w.h.p. the resulting v is such that $\ln\left(\frac{\Pr[V=v]}{\Pr[V'=v]}\right)$ is small.

Fix a sequence $v = (y_1, y_2, \dots, y_k)$ of outputs for $M_1, \dots, M_k$. We have that

$$\ln\left(\frac{\Pr[V = v]}{\Pr[V' = v]}\right) = \ln\left(\prod_{i=1}^{k} \frac{\Pr[Y_i = y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}]}{\Pr[Y_i' = y_i | Y_1' = y_1, Y_2' = y_2, \dots, Y_{i-1}' = y_{i-1}]}\right)$$

$$= \sum_{i=1}^{k} \ln\left(\frac{\Pr[Y_i = y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}]}{\Pr[Y_i' = y_i | Y_1' = y_1, Y_2' = y_2, \dots, Y_{i-1}' = y_{i-1}]}\right)$$

$$\triangleq \sum_{i=1}^{k} c_i(y_1, \dots, y_{i-1}, y_i)$$

Thus, we expressed $\ln\left(\frac{\Pr[V=v]}{\Pr[V'=v]}\right)$ *as the sum of* k *random variables* $C_1, C_2, \dots, C_k$ *where each* C_i *is a random variable taking the value* $c_i(y_1, \dots, y_{i-1}, y_i)$.

Now we want to apply the Azuma-Hoeffding inequality to these random variables.

Remark: *The probability is over sampling* $V = (Y_1, \dots, Y_k)$. *That is, we want to bound*

$$\Pr_{v \leftarrow V}\left[\ \ln\left(\frac{\Pr[V = v]}{\Pr[V' = v]}\right) > \hat{\varepsilon}\ \right] = \Pr_{v \leftarrow V}\left[\ \sum_{i=1}^{k} c_i(y_1, \dots, y_{i-1}, y_i) > \hat{\varepsilon}\ \right]$$

To apply Azuma's inequality, we need to bound $|C_i|$ *and analyze the expectation.*
Recall that, by definition

$$c_i(y_1, y_2, \dots, y_i) = \ln\left(\frac{\Pr[Y_i = y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}]}{\Pr[Y_i' = y_i | Y_1' = y_1, Y_2' = y_2, \dots, Y_{i-1}' = y_{i-1}]}\right) = ((1))$$

Where Y_j, Y_j' *are random variables representing the* j*th outcome during the executions on* S, S', *respectively.*

Recall that once the outcomes $y_1, y_2, \dots, y_{i-1}$ *are fixed, they determine the parameter* $p_i = p_i(y_1, y_2, \dots, y_{i-1})$ *that is fed into the execution of the mechanism* M_i.

Thus,

$$((1)) = \ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]}\right) = ((2))$$

Now recall that M_i *is* $(\varepsilon, 0)$*-DP-stable, and so, by definition of DP-stability,*

$$((2)) \in [-\varepsilon, \varepsilon]$$

In particular, for every $y_1, y_2, \dots, y_i$ we have that

$$|c_i(y_1, y_2, \dots, y_i)| \leq \varepsilon$$

And so $\Pr[|C_i| \leq \varepsilon] = 1$.
Now, let's compute the conditional expectation:

$$\mathbb{E}\left[C_i \,\middle|\, C_1 = c_1, \dots, C_{i-1} = c_{i-1} \right] = \mathbb{E}_{v \leftarrow V}\left[c_i(y_1, y_2, \dots, y_i) \,\middle|\, C_1 = c_1, \dots, C_{i-1} = c_{i-1} \right]$$

$$\underbrace{=}_{by\ ((1))} \mathbb{E}_{v \leftarrow V}\left[\ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]} \right) \middle| C_1 = c_1, \dots, C_{i-1} = c_{i-1} \right]$$

$$\underbrace{=}_{\substack{total \\ expectation}} \sum_{y_1, \dots, y_{i-1}} \mathbb{E}_{v \leftarrow V}\left[\ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]} \right) \middle| y_1, \dots, y_{i-1} \right] \cdot \Pr_{v \leftarrow V}\left[y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right]$$

$$= \sum_{y_1, \dots, y_{i-1}} \mathbb{E}_{y_i \leftarrow Y_i}\left[\ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]} \right) \middle| y_1, \dots, y_{i-1} \right] \cdot \Pr_{v \leftarrow V}\left[y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right] = ((3))$$

As we mentioned, once the outcomes $y_1, y_2, \dots, y_{i-1}$ are fixed, they determine the parameter $p_i = p_i(y_1, y_2, \dots, y_{i-1})$ and determines the distribution of Y_i to be the outcome of $M_i(S, p_i)$. Hence,

$$((3)) = \sum_{y_1, \dots, y_{i-1}} \mathbb{E}_{y_i \leftarrow M_i(S, p_i)}\left[\ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]} \right) \right] \cdot \Pr_{v \leftarrow V}\left[y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right] = ((4))$$

Auxiliary Claim 4 (to be proven shortly): *For any fixed p_i it holds that*

$$\mathbb{E}_{y_i \leftarrow M_i(S, p_i)}\left[\ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]} \right) \right] \leq 2 \cdot \varepsilon^2$$

Using this claim we get that

$$((4)) \leq \sum_{y_1, \dots, y_{i-1}} 2 \cdot \varepsilon^2 \cdot \Pr_{v \leftarrow V}\left[y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right] = 2 \cdot \varepsilon^2$$

Where are we in the proof of Theorem 2?

We want to prove that:

$$\Pr_{v\leftarrow V}\left[\sum_{i=1}^{k} c_i(y_1,\ldots,y_{i-1},y_i) > \hat{\varepsilon}\right] \le \hat{\delta}$$

So far we have shown that for every i it holds that $|C_i| \le \varepsilon$ and $\mathbb{E}\left[C_i \,\middle|\, C_1 = c_1,\ldots,C_{i-1} = c_{i-1}\right] \le 2\varepsilon^2$.

Therefore, by the Azuma-Hoeffding Inequality, for every $z > 0$ we have

$$\Pr_{v\leftarrow V}\left[\sum_{i=1}^{k} c_i(y_1,\ldots,y_{i-1},y_i) > k2\varepsilon^2 + z\sqrt{k}\varepsilon\right] \le \exp\left(-\frac{z^2}{2}\right)$$

Setting $z = \sqrt{2\cdot\ln(1/\delta)}$ and denoting $\hat{\varepsilon} = 2k\varepsilon^2 + \sqrt{2k\cdot\ln\left(1/\hat{\delta}\right)}\,\varepsilon$ we get that

$$\Pr_{v\leftarrow V}\left[\sum_{i=1}^{k} c_i(y_1,\ldots,y_{i-1},y_i) > \hat{\varepsilon}\right] \le \delta$$

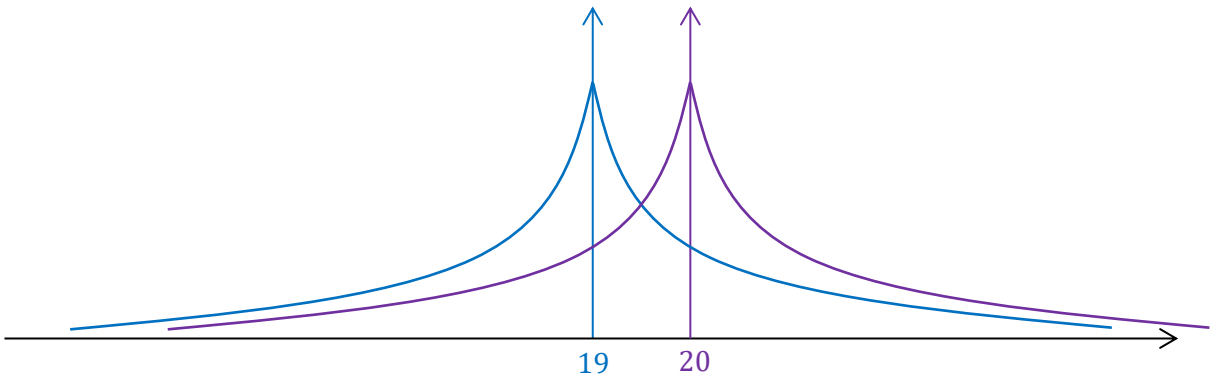
q.e.d. (Theorem 2)

It remains to prove Auxiliary Claim 4.

Intuition regarding composition:

Why should DP-stability degrade like $\sqrt{k}$ and not like k?

Suppose I runs the Laplace mechanism k times on an input S or on a neighboring input S', with the same query, having a value of 19 on S and a value of 20 on S'. In other words, in each round, I give you a sample from one of the two following distributions:



- *You want to try to guess whether the input is S or S'*
- *Output > 19.5 "makes you think" that S' is more likely.*
- *Output < 19.5 "makes you think" that S is more likely.*
- *But sometimes, you will get an output < 19.5 even on S' which only misleads you in your attempt to distinguish between S and S'*
- *Intuitively, this means that sometimes I "gain stability" instead of "losing stability".*

More intuition: Suppose we run the Laplace mechanism on S (i.e., we sample from the blue distribution in the diagram above). Clearly, the probability of getting an output less than 19 is exactly 1/2. Additionally, the probability of getting an output between 19 and 19.5 is approximately ε (this can be computed using an integral of the density function). Thus, the probability of getting an output in the region where the blue graph is higher than the purple one is approximately 1/2+ε, and with a probability of about 1/2−ε, we get an output that "misleads us", where the purple graph is higher.

How many times will we get an output that "does not mislead us"? This is a random variable with a binomial distribution and a success probability of 1/2+ε. The expected number of successes out of k samples is k/2+εk, meaning that, on average, there is indeed some advantage for "outputs that do not mislead us". However, the standard deviation of such a random variable is approximately $\frac{\sqrt{k}}{2}$. Therefore, if $\varepsilon k \ll \frac{\sqrt{k}}{2}$, then the advantage we have is "swallowed" by the standard deviation. In other words, as long as $\varepsilon \ll \frac{1}{2\sqrt{k}}$, it will be very difficult to distinguish between the purple and blue distributions.

Intuition regarding Claim 4 in the case of two Laplace mechanisms: We want to argue about the random variable defined as follows: sample a point y according to the blue distribution, and then compute the log ratio of the height of the blue line to the height of the purple line. We already know that this log ratio is always bounded by ±ε. However, it is quite clear that the expectation of this random variable must be much smaller than ε (because sometimes this value is even negative). As we now show, a simple calculation shows that this expectation is approximately ε^2.

When sampling a value y according to the blue distribution (with value 19) we have that

- $\ln\left(\frac{\Pr[M_i(S,p)=y]}{\Pr[M_i(S',p)=y]}\right) = \varepsilon \;\; if \;\; y \leq 19$
- $\ln\left(\frac{\Pr[M_i(S,p)=y]}{\Pr[M_i(S',p)=y]}\right) = -\varepsilon \;\; if \;\; y \geq 20$
- *In between, it is somewhere in between these values. Let us pessimistically assume that* $\ln\left(\frac{\Pr[M_i(S,p)=y]}{\Pr[M_i(S',p)=y]}\right) = \varepsilon$ *holds also for* $19 \leq y \leq 20$.
- *Now note that the probability of being in the middle is small:*

$$\Pr_{y_i \leftarrow M_i(S,p_i)}[19 \leq y_i \leq 20] = \int_0^1 \frac{\varepsilon}{2} \cdot \exp(-\varepsilon \cdot y)\, dy = \left[\frac{\varepsilon}{2} \cdot \frac{\exp(-\varepsilon \cdot y)}{-\varepsilon}\right]_0^1$$
$$= \frac{1}{2}(1 - e^{-\varepsilon}) \approx \frac{1}{2}\big(1 - (1-\varepsilon)\big) = \frac{\varepsilon}{2}$$

Therefore,

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] \leq \Pr[y_i \leq 19] \cdot \varepsilon + \Pr[19 < y_i < 20] \cdot \varepsilon + \Pr[y_i > 20] \cdot (-\varepsilon)$$
$$\approx \frac{1}{2}\varepsilon + \frac{\varepsilon}{2}\varepsilon - \left(\frac{1}{2} - \frac{\varepsilon}{2}\right)\varepsilon = \varepsilon^2$$

A formal proof for Claim 4:

Observe that

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right] = \sum_{y_i} \Pr[M_i(S,p_i)=y_i] \cdot \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \geq 0$$

Explanation: This follows from the log-sum inequality, which states that if $a_1, \ldots, a_n, b_1, \ldots, b_n$ are non-negative numbers, then

$$\sum_i a_i \cdot \ln\left(\frac{a_i}{b_i}\right) \geq \left(\sum_i a_i\right) \cdot \ln\left(\frac{\sum_i a_i}{\sum_i b_i}\right)$$

Similarly, it holds that

$$\mathbb{E}_{y_i \leftarrow M_i(S',p_i)}\left[\ln\left(\frac{\Pr[M_i(S',p_i)=y_i]}{\Pr[M_i(S,p_i)=y_i]}\right) \right] \geq 0$$

Now,

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right] \leq$$

$$\leq \mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right] + \mathbb{E}_{y_i \leftarrow M_i(S',p_i)}\left[\ln\left(\frac{\Pr[M_i(S',p_i)=y_i]}{\Pr[M_i(S,p_i)=y_i]}\right) \right]$$

$$= \mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right] - \mathbb{E}_{y_i \leftarrow M_i(S',p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right]$$

$$= \sum_{y_i} \Pr_{M_i(S,p_i)}[y_i] \cdot \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) - \sum_{y_i} \Pr_{M_i(S',p_i)}[y_i] \cdot \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)$$

$$= \sum_{y_i} \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \cdot \left(\Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right)$$

$$\leq \sum_{y_i} \left| \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right| \cdot \left| \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right|$$

$$\leq \max_{y_i \in \mathrm{Supp}(Y_i)} \left| \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right| \cdot \sum_{y_i} \left| \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right|$$

$$\leq \varepsilon \cdot \sum_{y_i} \left| \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right| \underbrace{=}_{\text{see below}} \varepsilon \cdot 2 \cdot \mathrm{SD}\left(M_i(S,p_i) \ , \ M_i(S',p_i) \right)$$

$$= 2\varepsilon \cdot \max_{T \subseteq \mathrm{Supp}(Y_i)} \left| \Pr_{M_i(S,p_i)}[y_i \in T] - \Pr_{M_i(S',p_i)}[y_i \in T] \right| \underbrace{\leq}_{\substack{M_i \text{ is stable} \\ \text{(see below)}}} 2\varepsilon \cdot (1 - e^{-\varepsilon}) \ \leq \ 2\varepsilon^2$$

Reminder: The statistical distance $\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2)$ is a measure of distance between two distributions $\mathcal{D}_1, \mathcal{D}_2$ and it has two equivalent definitions:

$$\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \sum_{y} \left| \Pr_{x \sim \mathcal{D}_1}[x = y] - \Pr_{x \sim \mathcal{D}_2}[x = y] \right| = \max_{T} \left| \Pr_{x \sim \mathcal{D}_1}[x \in T] - \Pr_{x \sim \mathcal{D}_2}[x \in T] \right|$$

Additional; explanation for the transition in the last line: Consider the event T that maximizes the expression, and assume without loss of generality that

$$\Pr_{M_i(S,p_i)}[y_i \in T] \geq \Pr_{M_i(S',p_i)}[y_i \in T]$$

Then, as M_i is $(\varepsilon, 0)$-DP-stable, we have

$$\left| \Pr_{M_i(S,p_i)}[y_i \in T] - \Pr_{M_i(S',p_i)}[y_i \in T] \right| = \Pr_{M_i(S,p_i)}[y_i \in T] - \Pr_{M_i(S',p_i)}[y_i \in T] \leq$$

$$\leq \Pr_{M_i(S,p_i)}[y_i \in T] - e^{-\varepsilon} \cdot \Pr_{M_i(S,p_i)}[y_i \in T] = (1 - e^{-\varepsilon}) \cdot \Pr_{M_i(S,p_i)}[y_i \in T] \leq (1 - e^{-\varepsilon})$$

q.e.d. (claim 4). This concludes the proof of the composition theorem (Theorem 2).