Differential Privacy Study Group **May 23, 2016**

Adaptive Data Analysis

Leader: KM Chung **Speaker: Mark Simkin; Notes: Chi-Ning Chou**

This week we are going to apply differential privacy results in adaptive data analysis.

1 Motivation

The purpose of data analysis is to generalize from the data/sample to the population distribution. Concretely, given data $\mathbf{x} = (x_1, \ldots, x_n)$ where each x_i is randomly sampled from unknown distribution $\mathcal{P}$ over alphabet set $\mathcal{X}$ and queries in the form $q : \mathcal{X} \to \mathcal{R}$. Our goal would be preventing from *false discovery*, *i.e.,* finding some queries such that the result of the analysis is significantly different from the population mean.

In the the *non-adaptive* setting, *i.e.,* the analyzing algorithm is decided before the analysts see the data, be using concentration bound, one can show that the probability of false discovery to happen is exponentially small w.r.t. the number of queries being used. However, in adaptive setting, false discovery might easily happen! Take a look at the following example.

Example 1 *Let $\mathcal{X} = [N]$ and $\mathcal{P} = Uniform(N)$. Given samples $\mathbf{x} = (x_1, \ldots, x_n) \sim \mathcal{P}^n$, where $n < \frac{N}{2}$. After using N adaptive queries asking how many samples are $1, 2, \ldots, N$, the analyst can find the set $S = \{x_i\}$ containing all the value in $\mathbf{x}$. As a result, when asking query $q_S(x) := \mathbf{1}_{x \in S}$, the sample average will be 1 while the population expectation is $\frac{|S|}{N} \leq \frac{1}{2}$, which is a false discovery.*

From Example 1, we can see that adaptive data analysis can lead to false discovery after adaptively asking a small number of queries. As a result, we would like to resolve this issue by proposing a way to prevent false discovery, *i.e.,* making sure that the answer to each query can generalize well to the population. On the other hand, we would also like to show some impossibility result that after using certain number of adaptive queries, false discovery might happen. Namely, there are two lines of study in this field:

- Proposing data analyzing mechanism to prevent false discovery.
- Showing lower bound in adaptive data analysis, *i.e.,* like in Example 1, show that after querying certain number of queries, one can overfit the data.

And this week we are focusing on the first one: proposing a data analyzing mechanism ot prevent false discovery.

2 Framework

The following discussion is based on [BSSU15].

2.1 Overview

There are three characters in our settings: population distribution $\mathcal{P}$, data analyst $\mathcal{A}$, and a mechanism $\mathcal{M}$. $\mathcal{P}$ is a distribution over finite set $\mathcal{X}$ and $\mathcal{M}$ receives n samples $\mathbf{x} = (x_1, \ldots, x_n)$ followed this population distribution. $\mathcal{A}$ asks queries $q_1, \ldots, q_k$ to $\mathcal{M}$ adaptively, and gets the answer $a_1^{\mathcal{M}}, \ldots, a_k^{\mathcal{M}}$. For simplicity, the query we focus here is simply a function from $\mathcal{X}$ to $[0, 1]$ and the error measure is the ℓ_1 norm.

The goal of $\mathcal{A}$ is simple: to approximate the underlying distribution $\mathcal{P}$, *i.e.,* to generalize from the sample data he/she can use. Namely, on query q_j, $\mathcal{A}$ wants the answers $a^{\mathcal{M}}(q_j)$ from $\mathcal{M}$ can be close enough to the population mean $a^{\mathcal{P}}(q_j) := \mathbb{E}_{X \sim \mathcal{P}}[q_j(X)]$. Formally, we say the mechanism is *accurate w.r.t. population*. If the population mean of q is conditioned on certain event E, then we denote it as $a^{\mathcal{P}}(q|E) := \mathbb{E}_{X \sim \mathcal{P}}[q(X)|E]$. Moreover, it is also natural to require the answers $a^{\mathcal{M}}(q_j)$ being not to far from the sample mean $a^{\mathcal{S}}(q_j) := \frac{1}{n} \sum_{i=1}^{n} q_j(x_i)$. Formally, we say the mechanism is *accurate w.r.t. sample*. Last but not least, it is also reasonable to expect the mechanism won't behave too different when given two sets of data that are close to each other, say only differ in one date point. Somehow, this is a artificial but realistic assumption. Surprisingly, it turns out that adding this requirement, the mechanism will have certain level of guarantee to behave accurately w.r.t. the population. We call this notion *stability*. See Figure 1 to understand the relationship between the three notions introduced in this paragraph.

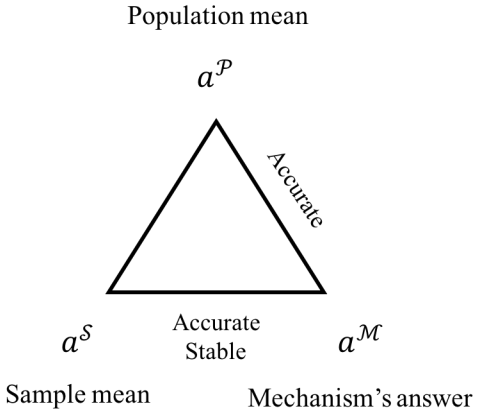


Figure 1: The relationship between three types of answers.

One can see that the in the above three notions: accurate w.r.t. population, accurate w.r.t. sample, and stability, there are two of them we can control in the mechanism, *i.e.,* accurate w.r.t. sample and stability. As to the accuracy w.r.t. population, we want the accuracy to be good, however, we can not directly touch it. The main contribution in [BSSU15] was showing that actually accuracy w.r.t. sample plus stability can imply accuracy w.r.t. population!

In the following, we will first formally define the three notions we are going to play with, then the main transfer theorem about generalization will be stated and proved.

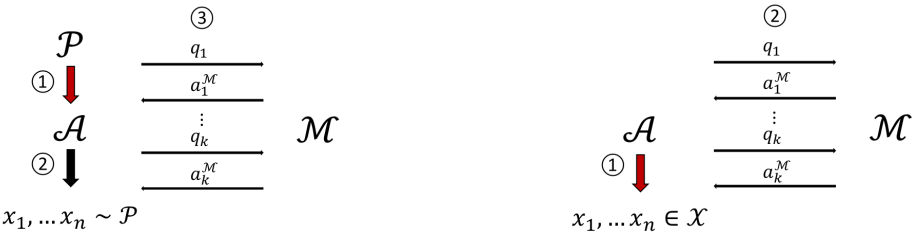
2.2 Definitions

First, let's formally define the type of queries we are going to focus on.

Definition 1 (Δ-sensitive query) *We say query $q : \mathcal{X}^n \to [0,1]$ is Δ-sensitive if for every pairs of $\mathbf{x}, \mathbf{x}' \in \mathcal{X}'$ differ in only one element, we have $|q(\mathbf{x}) - q(\mathbf{x}')| \leq \Delta$.*

Intuitively, the queries that are good for analyst should be *low-sensitive* in the sense that it is Δ-sensitive for some small Δ. A directly reason for this condition is that if the answer of a query varies a lot among datasets that are similar, the result of the analysis will be easily affected by the noise of the data. As a result, requiring a query to be low-sensitive is somehow guarantee the result of analysis being robust.

Next, let's define the two accuracy notions with the following accuracy game.



(a) The population accuracy game.

(b) The sample accuracy game.

Figure 2: The accuracy games.

Definition 2 ((α, β)-accurate w.r.t. population) *A mechanism $\mathcal{M}$ is (α, β)-accurate w.r.t. population for k adaptive queries if in any population accuracy game we have*

$$\mathbb{P}[\max_{j \in [k]} |a^{\mathcal{M}}(q_j) - a^{\mathcal{P}}(q_j)| \leq \alpha] \geq 1 - \beta$$

Definition 3 ((α, β)-accurate w.r.t. sample) *A mechanism $\mathcal{M}$ is (α, β)-accurate w.r.t. sample for k adaptive queries if in any sample accuracy game we have*

$$\mathbb{P}[\max_{j \in [k]} |a^{\mathcal{M}}(q_j) - a^{\mathcal{S}}(q_j)| \leq \alpha] \geq 1 - \beta$$

Next, we use a randomized meta algorithm $\mathcal{W}[\mathcal{M}, \mathcal{A}]$ to record the entire interaction between $\mathcal{A}$ and $\mathcal{M}$ and define stability on $\mathcal{W}$. Note that the input of $\mathcal{W}$ is the samples $x_1 \ldots, x_n$ and it will output all the query-answer pairs $(q_1, a_1^{\mathcal{M}}), \ldots (q_k, a_k^{\mathcal{M}})$. See Figure 3.

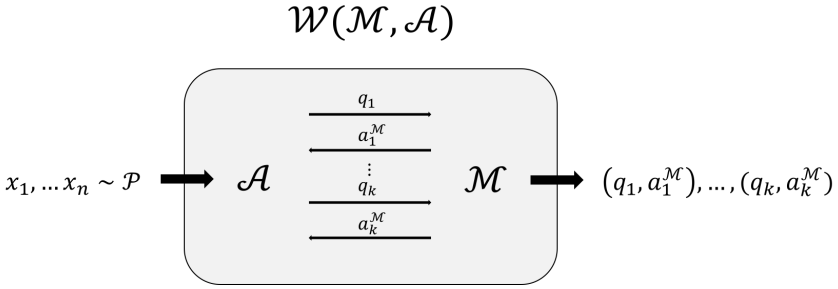


Figure 3: Meta algorithm $\mathcal{W}[\mathcal{M}, \mathcal{A}]$ for mechanism $\mathcal{M}$ and adversary/analyst $\mathcal{A}$.

Definition 4 (Max-stability of an algorithm) *We say an algorithm $\mathcal{W} : \mathcal{X}^n \to \mathcal{R}$ is (ϵ, δ)-max-stable if for every sample pairs $\mathbf{x}, \mathbf{x}'$ differ in exactly one data point and every $R \subseteq \mathcal{R}$,*

$$\mathbb{P}[\mathcal{W}(\mathbf{x}) \in R] \leq \mathbb{P}[\mathcal{W}(\mathbf{x}') \in R] + \delta$$

Definition 5 (Max-KL stability of a mechanism) *We say a mechanism $\mathcal{M}$ is (ϵ, δ)-max-stable if for any $\mathcal{A}$, $\mathcal{W}[\mathcal{M}, \mathcal{A}]$ is (ϵ, δ)-max-stable.*

An important lemma here is that the Max-KL stability will be preserved under post-processing. Namely, if we design an algorithm based on a stable algorithm, then it will also be stable with the same parameter setting.

Lemma 6 (post-processing preserve stability) *Let $\mathcal{W}$ be a (ϵ, δ)-Max-KL stable algorithm and f is an arbitrary function. Then $f(\mathcal{W}(\cdot))$ is also (ϵ, δ)-Max-KL stable.*

3 Transfer theorem

Let's quantitatively state the main transfer theorem: stable + accurate w.r.t. sample $\Rightarrow$ accurate w.r.t. population.

Theorem 7 (main transfer theorem) *Let Q be a family of Δ-sensitive queries on $\mathcal{X}$ and $\alpha, \beta \in (0, 0.1)$. If $\mathcal{M}$ is*

- *$(\epsilon = \frac{\alpha}{64\Delta n}, \delta = \frac{\alpha\beta}{32\Delta n})$-max-KL stable for k adaptively chosen queries from Q.*
- *$(\alpha' = \frac{\alpha}{8}, \beta' = \frac{\alpha\beta}{16\Delta n})$-accurate w.r.t. n samples for k adaptively chosen queries from Q.*

Then $\mathcal{M}$ is (α, β)-accurate w.r.t. population for k adaptively chosen queries from Q.

Here, we can think of the Δn term as the *uncertainty level* inherited from the date and the queries. That is, the larger the Δ is, the output might be easier to be affected by little changes in the date while in the same time, having more data means that there are more chances to go into a wrong way. As a result, as Δn becomes larger, one need to have a more stable and more accurate (w.r.t. sample) mechanism.

3.1 High level proof flow

Our goal is to show that stability plus accurate w.r.t. sample implies accurate w.r.t. population. The proof is shown in two steps as follow.

1. Using the main technical lemma to show that the among the queries asked by the analyst, the empirical average will not be too far from the population expectation.
2. Assume the mechanism does not generalize. Then, by the stability property, it implies that there are some queries such that the sample mean is far from population expectation, which contradicts to the main technical lemma.

Intuitively, we are playing between the three quantities in Figure 1, *i.e.,* the output from the mechanism $a^{\mathcal{M}}$, the empirical average $a^{\mathcal{S}}$, the population expectation $a^{\mathcal{P}}$. Our ultimate goal is to show that $a^{\mathcal{M}}$ is close to $a^{\mathcal{P}}$. To do so, we use the stability of $\mathcal{M}$ to show that $a^{\mathcal{S}}$ and $a^{\mathcal{P}}$ are close. Then, using the accuracy w.r.t. sample to show that $a^{\mathcal{S}}$ and $a^{\mathcal{M}}$ are close. As a result, $a^{\mathcal{M}}$ and $a^{\mathcal{P}}$ cannot be too far away, *i.e.,* our goal is fulfilled.

To argue the above two propositions, we need to treat ourself as an outsider, *i.e.,* being out of the whole communication process and only observe but do not involve in the interaction among $\mathcal{M}$ and $\mathcal{A}$. Formally, we use a *monitoring algorithm* $\mathcal{W}$ to simulate this idea. See Figure 3. $\mathcal{M}$ and $\mathcal{A}$ are viewed as two randomized algorithm that are talking to each other. $\mathcal{W}$ can see the transcript, *i.e.,* all the queries and answers, among them. Thus, it can help us find the worst query in the whole process. To help $\mathcal{W}$ do so, we even let $\mathcal{W}$ know the underlying distribution $\mathcal{P}$ so that it can find the query having maximum error between sample mean and population expectation.

The original proof in [BSSU15] used *error amplification techniques* to optimize the parameters, however, making the proofs a little bit messy. Thus, in our following explanation, that part will be omitted.

3.2 First taste of main technical lemma

The main technical lemma uses the stability of $\mathcal{M}$ to show that $a^{\mathcal{S}}$ and $a^{\mathcal{P}}$ cannot be far away. To do so, we use the monitoring algorithm $\mathcal{W}$ to find the query q that maximizes the error among $a^{\mathcal{P}}(q)$ and $a^{\mathcal{S}}(q)$. And we have the following result.

Lemma 8 (main technical lemma for statistical queries) *Let $\mathcal{M}$ be an (ϵ, δ)-Max-KL stable mechanism, $\mathbf{x} \in \mathcal{X}^n$ be the n samples drawn from $\mathcal{P}$, and Q be a set of statistical queries. We have*

$$|a^{\mathcal{P}}(q|q = \mathcal{W}(\mathbf{x})) - \mathbb{E}[a^{\mathcal{S}}(q)|q = \mathcal{W}(\mathbf{x})]| \leq e^{\epsilon} - 1 + \delta$$

Proof: Before the proof, let's define some notations first. As we are going to replace the element in $\mathbf{x}$ with a fresh sample in order to relate sample mean with population mean, we use $\mathbf{x}_{i \to x'}$ to denote the database after replacing the i-th element in $\mathbf{x}$ with a fresh sample x'.

Now, let's rewrite the expectation of empirical average so that we can connect it to the popu-

lation expectation.

$$\begin{aligned}
\mathbb{E}[a^{\mathcal{S}}(q)|q = \mathcal{W}(\mathbf{x})] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[q(x_i)|q = \mathcal{W}(\mathbf{x})] \\
(\because \mathbb{E}[X] = \int \mathbb{P}[X > z]dz) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \mathbb{P}[q(x_i) > z|q = \mathcal{W}(\mathbf{x})]dz \\
(\because \text{stability of } \mathcal{M} + \text{Lemma 6}) &\leq \frac{1}{n}\sum_{i=1}^{n}\int_0^1 e^{\epsilon}\mathbb{P}[q(x_i)|q = \mathcal{W}(\mathbf{x}_{i\to x'})] + \delta dz \\
(\text{move } \delta \text{ out of the integral}[1]) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1 e^{\epsilon}\mathbb{P}[q(x_i)|q = \mathcal{W}(\mathbf{x}_{i\to x'})]dz + \delta \\
&= \frac{1}{n}\sum_{i=1}^{n} e^{\epsilon}\mathbb{E}[q(x_i)|q = \mathcal{W}(\mathbf{x}_{i\to x'})] + \delta \\
(\because (x_i, \mathbf{x}_{i\to x'}) \overset{d}{\approx} (x', \mathbf{x})[2]) &= \frac{1}{n}\sum_{i=1}^{n} e^{\epsilon}\mathbb{E}[q(x')|q = \mathcal{W}(\mathbf{x})] + \delta \\
&= e^{\epsilon}\mathbb{E}[q(x')|q = \mathcal{W}(\mathbf{x})] + \delta \\
(\because x' \text{ is independent to } \mathbf{x}) &= e^{\epsilon}a^{\mathcal{P}}(q|q = \mathcal{W}(\mathbf{x})) + \delta
\end{aligned}$$

With a simple plug-in, we get our desired result since the population mean is in [0,1].

□

3.3 Main technical lemma

Lemma 9 (main technical lemma for Δ-sensitive queries) *Let $\mathcal{M}$ be an (ϵ, δ)-Max-KL stable mechanism, $\mathbf{x} \in \mathcal{X}^n$ be the n samples drawn from $\mathcal{P}$, and Q be a set of Δ-sensitive queries. We have*

$$|\mathbb{E}[a^{\mathcal{P}}(q)|q = \mathcal{W}(\mathbf{x})] - \mathbb{E}[a^{\mathcal{S}}(q)|q = \mathcal{W}(\mathbf{x})]| \leq 2n\Delta \cdot (e^{\epsilon} - 1 + \delta)$$

Proof: In order to relate the sample mean to population mean, now we want to replace the whole database $\mathbf{x}$ to a fresh database $\mathbf{x}'$. Since the stability notion only allows neighboring database, here we would like to replace the elements in $\mathbf{x}$ with that in $\mathbf{x}'$ one by one. We denote $\mathbf{x}^l$ as the database after replacing the first l elements in $\mathbf{x}$. With a simple telescoping method, we have

$$\begin{aligned}
|a^{\mathcal{P}}(q|q = \mathcal{W}(\mathbf{x})) - \mathbb{E}[q(\mathbf{x})|q = \mathcal{W}(\mathbf{x})]| &= \mathbb{E}[q(\mathbf{x}') - q(\mathbf{x})|q = \mathcal{W}(\mathbf{x})] \\
(\because \text{telescoping}) &= |\sum_{l=1}^{n}\mathbb{E}[q(\mathbf{x}^l) - q(\mathbf{x}^{l-1})|q = \mathcal{W}(\mathbf{x})]| \\
(\because \text{triangle inequality}) &\leq \sum_{l=1}^{n}|\mathbb{E}[q(\mathbf{x}^l) - q(\mathbf{x}^{l-1})|q = \mathcal{W}(\mathbf{x})]| \qquad (1)
\end{aligned}$$

[1] Since here we deal with statistical queries, the range of the queries is nice, *i.e.,* [0,1], in which the additive term δ from the stability condition can be nicely handles. However, when it comes to more general queries such as Δ-sensitive queries, the range might be infinite, which is the main difficulty.

[2] As x' and every element in $\mathbf{x}$ are i.i.d. sampled from $\mathcal{P}$, clearly the two are distributed in the same way.

Observe that each term in the above summation is very small since they only differ in one element and the query is low-sensitive. However, since $\mathbf{x}^l$ and $\mathbf{x}^{l-1}$ differ in 2 elements, we cannot apply the sensitivity condition to yield an upper bound for the error. That is, if we want to directly approximate the difference, we need to integrate from $-\infty$ to ∞, which will blow up the additive term δ in the stability notion. As a result, we should try to transform this error term into another term that has bounded support.

The way [BSSU15] resolved this issue is based on a clever observation: the following two databases are identically distributed.

- Replace the l-th element of $\mathbf{x}^l$ with a fresh sample $x' \Rightarrow \mathbf{x}^l_{-l}$.
- Replace the l-th element of $\mathbf{x}^{l-1}$ with a fresh sample $x' \Rightarrow \mathbf{x}^{l-1}_{-l}$.

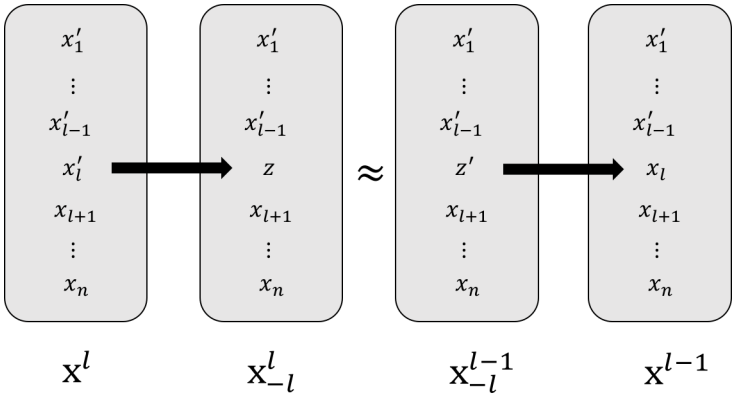


Figure 4: The neighboring databases. $\mathbf{x}^l$ and $\mathbf{x}^l_{-l}$ are neighboring, $\mathbf{x}^l_{-l}$ and $\mathbf{x}^{l-1}_{-l}$ are identically distributed, $\mathbf{x}^{l-1}_{-l}$ and $\mathbf{x}^{l-1}$ are neighboring.

Thus, $\mathbb{E}[q(\mathbf{x}^l_{-l})|q = \mathcal{W}(\mathbf{x})] = \mathbb{E}[q(\mathbf{x}^{l-1}_{-l})|q = \mathcal{W}(\mathbf{x})]$. Moreover, now $\mathbf{x}$ and $\mathbf{x}^l_{-l}$ are neighboring! (so does $\mathbf{x}^{l-1}$ and $\mathbf{x}^{l-1}_{-l}$) Namely, by the sensitivity property of the query q, we know that $|q(\mathbf{x}^l) - q(\mathbf{x}^l_{-l})| \leq \Delta$. Finally, define $B^l(\mathbf{x}, \mathbf{z}) := q(\mathbf{z}) - q(\mathbf{z}_{-l}) + \Delta$, where $q = \mathcal{W}(\mathbf{x})$. Now, we can rewrite the above telescoping sum.

$$\begin{aligned}(1) &= \sum_{l=1}^{n} |\mathbb{E}[(q(\mathbf{x}^l) - q(\mathbf{x}^l_{-l}) + \Delta) - (q(\mathbf{x}^{l-1}) - q(\mathbf{x}^{l-1}{-}l) + \Delta)|q = \mathcal{W}(\mathbf{x})]| \\ &= \sum_{l=1}^{n} \mathbb{E}[B^l(\mathbf{x}, \mathbf{x}^l) - B^l(\mathbf{x}, \mathbf{x}^{l-1})|q = \mathcal{W}(\mathbf{x})] \end{aligned} \tag{2}$$

Now, consider the relation between $B^l(\mathbf{x}, \mathbf{x}^l)$ and $B^l(\mathbf{x}, \mathbf{x}^{l-1})$ in (2). From Figure 4, we can see that $(\mathbf{x}, \mathbf{x}^l) \overset{d}{\sim} (\mathbf{x}^l, \mathbf{x})$ and $(\mathbf{x}^{l-1}, \mathbf{x}) \overset{d}{\sim} (\mathbf{x}, \mathbf{x}^{l-1})$, *i.e.,* $B^l(\mathbf{x}, \mathbf{x}^l) \overset{d}{\sim} B^l(\mathbf{x}^l, \mathbf{x})$ and $B^l(\mathbf{x}^{l-1}, \mathbf{x}) \overset{d}{\sim} B^l(\mathbf{x}, \mathbf{x}^{l-1})$. Moreover, as $\mathbf{x}^l$ and $\mathbf{x}^{l-1}$ only differs in one element, when we replace the database fed into $\mathcal{W}$ from $\mathbf{x}^l$ to $\mathbf{x}^{l-1}$, the stability condition will guarantee the closeness of $B(\mathbf{x}^l, \mathbf{x})$ and $B(\mathbf{x}^{l-1}, \mathbf{x})$. Concretely, $\forall z \in [0, 2\Delta]$, we have $\mathbb{P}[B(\mathbf{x}^l, \mathbf{x}) = z] \leq e^\epsilon \mathbb{P}[B(\mathbf{x}^{l-1}, \mathbf{x}) = z] + \delta$.

$$B(\mathbf{x}, \mathbf{x}^l) \overset{d}{\sim} B(\mathbf{x}^l, \mathbf{x}) \overset{(\epsilon,\delta)}{\sim} B(\mathbf{x}^{l-1}, \mathbf{x}) \overset{d}{\sim} B(\mathbf{x}, \mathbf{x}^{l-1}) \tag{3}$$

Finally, we can replace the $\mathbf{x}^{l-1}$ in $B(\mathbf{x}, \mathbf{x}^{l-1})$ with $\mathbf{x}^l$ as follow.

$$\begin{aligned}\mathbb{E}[B(\mathbf{x}^{l-1}, \mathbf{x})|q = \mathcal{W}(\mathbf{x}^{l-1})] &= \int_0^{2\Delta} \mathbb{P}[B(\mathbf{x}^{l-1}, \mathbf{x}) = z|q = \mathcal{W}(\mathbf{x}^{l-1})]dz \\ (\because \mathcal{W} \text{ is } (\epsilon, \delta)\text{-Max stable}) &\leq \int_0^{2\Delta} e^{\epsilon}\mathbb{P}[B(\mathbf{x}^{l}, \mathbf{x}) = z|q = \mathcal{W}(\mathbf{x}^{l})] + \delta dz \\ &= e^{\epsilon}\mathbb{E}[B(\mathbf{x}^{l}, \mathbf{x})|q = \mathcal{W}(\mathbf{x}^{l})] + 2\Delta \cdot \delta\end{aligned}$$

Namely,

$$\begin{aligned}|\mathbb{E}[B^l(\mathbf{x}, \mathbf{x}^l) - B^l(\mathbf{x}, \mathbf{x}^{l-1})|q = \mathcal{W}(\mathbf{x})]| &\leq (e^{\epsilon} - 1) \cdot \mathbb{E}[B(\mathbf{x}^l, \mathbf{x})|q = \mathcal{W}(\mathbf{x})] + 2\Delta \cdot \delta \\ (\because B(\cdot,\cdot)leq2\Delta) &\leq (e^{\epsilon} - 1) \cdot 2\Delta + 2\Delta \cdot \delta \\ &= 2\Delta(e^{\epsilon} - 1 + \delta) \end{aligned} \tag{4}$$

Plug (4) into (1) and (2), we have

$$|a^{\mathcal{P}}(q|q = \mathcal{W}(\mathbf{x})) - \mathbb{E}[q(\mathbf{x})|q = \mathcal{W}(\mathbf{x})]| \leq 2n\Delta \cdot (e^{\epsilon} - 1 + \delta)$$

□

We have gone through the simple version proof for the main technical lemma. To achieve optimal parameter setting, we need to use an error amplification argument. The idea is to feed $\mathcal{M}$ with T i.i.d. size n databases $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ and use the monitoring algorithm $\mathcal{W}$ to find the query with largest generalization error. Namely $\mathcal{W}(\mathbf{X}) \arg\max_{q,t} |a_t^{\mathcal{P}}(q) - a_t^{\mathcal{S}}(q)|$.

Lemma 10 (main technical lemma for Δ-sensitive queries with error amplification) *Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ be a collection of i.i.d. databases in which $\mathbf{x}_i \in \mathcal{X}^n$ follows distribution $\mathcal{P}$, $\mathcal{M}$ be an (ϵ, δ)-Max-KL stable mechanism, and Q be a set of Δ-sensitive queries. Suppose $\mathcal{W}$ is the monitoring algorithm $\mathbf{X}$, we have*

$$|\mathbb{E}[a^{\mathcal{P}}(q)|(q, t) = \mathcal{W}(\mathbf{X})] - \mathbb{E}[a^{\mathcal{S}}(q)|(q, t) = \mathcal{W}(\mathbf{X})]| \leq 2n\Delta \cdot (e^{\epsilon} - 1 + T\delta)$$

Note that in this error amplification version, the additive term has been blown up from δ to $T\delta$. Moreover, with the parameter setting in Theorem 7, we have the following corollary.

Corollary 11 *If $\mathcal{M}$ is $(\frac{\alpha}{64\Delta n}, \frac{\alpha\beta}{32\Delta n})$-Max-KL stable, for any monitoring algorithm $\mathcal{W}$, we have*

$$|\mathbb{E}[a^{\mathcal{P}}(q) - a_t^{\mathcal{S}}(q)|(q, t) = \mathcal{W}(\mathbf{X})]| \leq 2\Delta n(e^{\frac{\alpha}{64\Delta n}} - 1 + t(\frac{\alpha\beta}{32\Delta n}))$$

When we take $T = \lfloor 1/\beta \rfloor$, we have

$$|\mathbb{E}[a^{\mathcal{P}}(q) - a_t^{\mathcal{S}}(q)|(q, t) = \mathcal{W}(\mathbf{X})]| \leq \frac{\alpha}{8}$$

3.4 Final step

To show that $\mathcal{M}$ is (α, β)-accurate w.r.t. population, we instead assume it is not and try to derive contradiction. Assume $\mathcal{M}$ is not (α, β)-accurate w.r.t. population, then by Definition 2, $\forall t \in [T]$, $\mathbb{P}[\max_{j\in[k]} |a^{\mathcal{P}}(q_{t,j}) - a_j^{\mathcal{M}}(q_{t,j})| > \alpha] > \beta$. Our goal is to show that $|\mathbb{E}[a^{\mathcal{P}}(q) - a_t^{\mathcal{S}}(q)|(q,t) = \mathcal{W}(\mathbf{X})]| \geq \frac{\alpha}{4}$, which contradicts to the corollary of main technical lemma in Corollary 11.

Let's factorize the expected error between population and sample as follow.

$$\begin{aligned}
|\mathbb{E}[a^{\mathcal{P}}(q) - a_t^{\mathcal{S}}(q)|(q,t) = \mathcal{W}(\mathbf{X})]| &= |\mathbb{E}[a^{\mathcal{P}}(q) + a^{\mathcal{M}}(q) - (a^{\mathcal{M}}(q) - a_t^{\mathcal{S}}(q))|(q,t) = \mathcal{W}(\mathbf{X})]| \\
(\because \text{triangle inequality}) &\geq |\mathbb{E}[a^{\mathcal{P}}(q) - a^{\mathcal{M}}(q)|(q,t) = \mathcal{W}(\mathbf{X})]| \\
&\quad - |\mathbb{E}[a^{\mathcal{M}}(q) - a_t^{\mathcal{S}}(q)|(q,t) = \mathcal{W}(\mathbf{X})]| \qquad (5)
\end{aligned}$$

Recall that our goal is to lower (5) with $\frac{\alpha}{4}$. Consider the first term in (5). By the definition of $\mathcal{W}$ we know that $a^{\mathcal{P}}(q) \geq a^{\mathcal{M}}(q)$. Moreover, as we assume $\mathcal{M}$ is not(α, β)-accurate w.r.t. population, we have

$$\mathbb{P}[a^{\mathcal{P}}(q_t) - a^{\mathcal{M}}(q_t) > \alpha | q_t = \mathcal{W}(\mathbf{x}_t)] > \beta, \ \forall t \in [T] \qquad (6)$$

Furthermore,

$$\begin{aligned}
\mathbb{P}[a^{\mathcal{P}}(q) - a^{\mathcal{M}}(q) > \alpha | (q,t) = \mathcal{W}(\mathbf{X})] &= 1 - \mathbb{P}[a^{\mathcal{P}}(q) - a^{\mathcal{M}}(q) \leq \alpha | (q,t) = \mathcal{W}(\mathbf{X})] \\
&= 1 - \mathbb{P}[\cup_{t\in[T]} \{a^{\mathcal{P}}(q_t) - a^{\mathcal{M}}(q_t) \leq \alpha | q_t = \mathcal{W}(\mathbf{x}_t)\}] \\
(\because (6)) &\geq 1 - (1-\beta)^T \\
(\because T = \left\lfloor \frac{1}{\beta} \right\rfloor) &\geq \frac{\alpha}{2} \qquad (7)
\end{aligned}$$

Now, consider the second error term between sample mean and output of $\mathcal{M}$. As long as $\mathcal{M}$ is $(\frac{\alpha}{8}, \frac{\alpha\beta}{16\Delta n})$-stable, we know that this error term should not be too large. Concretely, for the output of each database $\mathbf{x}_t$, most of the time the error will be at most $\frac{\alpha}{8}$ while at most with probability $\frac{\alpha\beta}{16\Delta n}$ the error will exceed $\frac{\alpha}{8}$. However, even in this rare situation, we can still bound the error with $2\Delta n$ by the low-sensitivity property of the queries[3]. As a result, we have

$$\begin{aligned}
|\mathbb{E}[a^{\mathcal{M}}(q) - a_t^{\mathcal{S}}(q)|(q,t) = \mathcal{W}(\mathbf{X})]| &\leq \frac{\alpha}{8} \cdot \mathbb{P}[\forall t \in [T], j \in [k], |a^{\mathcal{M}}(q_{t,j}) - a_t^{\mathcal{S}}(q_{t,j})| \leq \frac{\alpha}{8}] \\
&\quad + 2\Delta n \cdot \mathbb{P}[\exists t \in [T], j \in [k], |a^{\mathcal{M}}(q_{t,j}) - a_t^{\mathcal{S}}(q_{t,j})| > \frac{\alpha}{8}] \\
&\leq \frac{\alpha}{8} + 2\Delta n \cdot \sum_{t\in[T]} \mathbb{P}[\exists j \in [k], |a^{\mathcal{M}}(q_{t,j}) - a_t^{\mathcal{S}}(q_{t,j})| > \frac{\alpha}{8}] \\
&\leq \frac{\alpha}{8} + 2\Delta n \cdot [T \cdot (\frac{\alpha\beta}{16\Delta n})] \\
(\because T = \left\lfloor \frac{1}{\beta} \right\rfloor) &\leq \frac{\alpha}{8} + \frac{\alpha}{8} = \frac{\alpha}{4} \qquad (8)
\end{aligned}$$

Finally, combine (7) and (8) and plug in (5), we achieve our desired goal: $|\mathbb{E}[a^{\mathcal{P}}(q) - a_t^{\mathcal{S}}(q)|(q,t) = \mathcal{W}(\mathbf{X})]| \geq \frac{\alpha}{4}$, which contradicts to Corollary 11 and hence $\mathcal{M}$ is (α, β)-accurate w.r.t. population.

[3]This loose upper bound can be achieved by replacing all the element in the database one by one with random sample.

References

[BSSU15] Raef Bassily, Adam Smith, Thomas Steinke, and Jonathan Ullman. More general queries and less generalization error in adaptive data analysis. *arXiv preprint arXiv:1503.04843*, 2015.