Differential Privacy as a Stability Notion

Moshe Shenfeld

Hebrew University of Jerusalem

The Israel Statistical Association Annual Meeting
June 2019

Ask not what your data might do to your privacy,
ask what Privacy can do for your data analysis

What is Learning, and When Does it Fail?

- Learning / Science almost always happens on a sample

What is Learning, and When Does it Fail?

- Learning / Science almost always happens on a sample
- Most learning tasks can be described as a (series of) function assessment tasks (loss, gradient, etc.)

What is Learning, and When Does it Fail?

- Learning / Science almost always happens on a sample
- Most learning tasks can be described as a (series of) function assessment tasks (loss, gradient, etc.)
- Concentration bounds ensure that under certain conditions, for any given function, most samples are "representative"

What is Learning, and When Does it Fail?

- Learning / Science almost always happens on a sample
- Most learning tasks can be described as a (series of) function assessment tasks (loss, gradient, etc.)
- Concentration bounds ensure that under certain conditions, for any given function, most samples are "representative"
- Using the union bound, this guarantee can be extended to multiple functions (Bonferroni Correction)

What is Learning, and When Does it Fail?

- Learning / Science almost always happens on a sample
- Most learning tasks can be described as a (series of) function assessment tasks (loss, gradient, etc.)
- Concentration bounds ensure that under certain conditions, for any given function, most samples are "representative"
- Using the union bound, this guarantee can be extended to multiple functions (Bonferroni Correction)
- These guarantees do not hold for a function chosen based on the sample set, hence cannot be used for adaptive learning

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Holds a set, sampled from a distribution

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Does not have direct access to the sample set
- Holds a set, sampled from a distribution

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Does not have direct access to the sample set

- Holds a set, sampled from a distribution
- Does not know the underlying distribution

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Does not have direct access to the sample set
- Chooses functions to be evaluated by the Mechanism (queries)

- Holds a set, sampled from a distribution
- Does not know the underlying distribution

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Does not have direct access to the sample set
- Chooses functions to be evaluated by the Mechanism (queries)

- Holds a set, sampled from a distribution
- Does not know the underlying distribution
- Provides function estimations for the Adversary (responses)

Adaptive Data Analysis as a Game

In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Does not have direct access to the sample set
- Chooses functions to be evaluated by the Mechanism (queries)
- Tries to identify the sample set, via the Mechanism's responses to his chosen queries

- Holds a set, sampled from a distribution
- Does not know the underlying distribution
- Provides function estimations for the Adversary (responses)

Adaptive Data Analysis as a Game

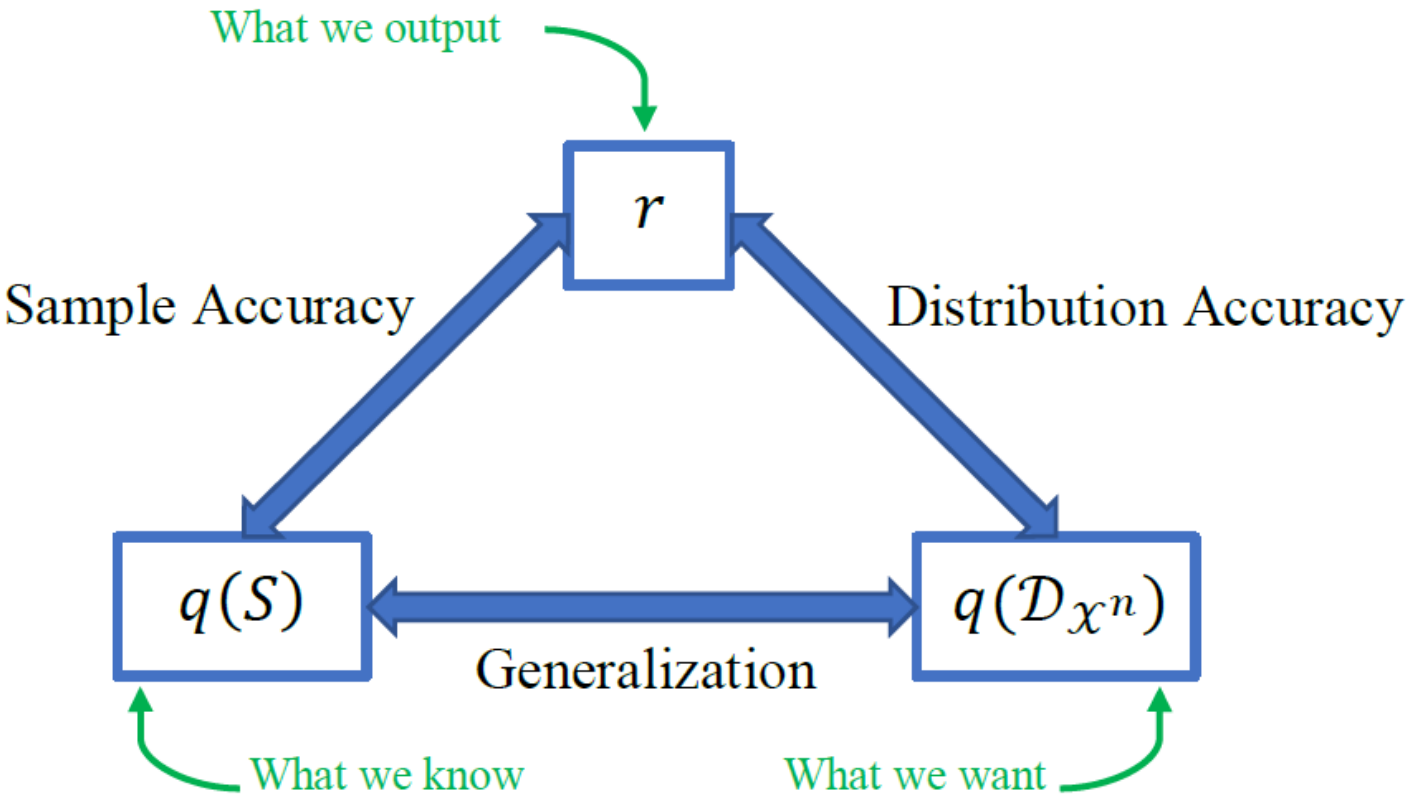
In the worst case, Adaptive Data Analysis can be thought of as a game between an Adversary and a Mechanism

- Might have any auxiliary information
- Does not have direct access to the sample set
- Chooses functions to be evaluated by the Mechanism (queries)
- Tries to identify the sample set, via the Mechanism's responses to his chosen queries

- Holds a set, sampled from a distribution
- Does not know the underlying distribution
- Provides function estimations for the Adversary (responses)
- Is expected to provide accurate responses while keeping the sample set secret from the Adversary

Informal Adaptive Data Analysis Setup

Domain - $\mathcal{X}$, Sample size - n, Sample set - $S \in \mathcal{X}^n$, Distribution over sample sets - $D_{\mathcal{X}^n}$, Query - q, Response - r,
$q(D_{\mathcal{X}^n}) := \mathop{\mathbb{E}}_{S}[q(S)]$



Generalization and Accuracy

Definition (Linear Queries)

A function $q : \mathcal{X}^n \to \mathbb{R}$ will be called a *linear query*, if it is defined by a function $q : \mathcal{X} \to \mathbb{R}$ such that,

$$q(S) := \frac{1}{n} \sum_{i=1}^{n} q(s_i)$$

If $q : \mathcal{X} \to [-\Delta, \Delta]$ it will be called a Δ*-bounded linear query*. The set of Δ-bounded linear queries will be denoted $\mathcal{Q}_\Delta$.

Generalization and Accuracy

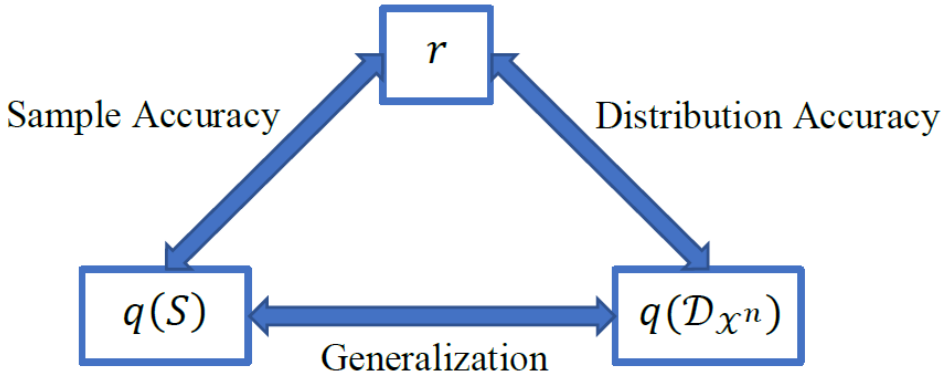
Definition (Sample / distribution Accuracy (single query))

A mechanism $M : \mathcal{X}^n \times \mathcal{Q} \to \mathbb{R}$ will be called:
(ϵ, δ)-*Sample-Accurate* Accurate, if

$$\mathbb{P}_{S,M}\left(|M(S,q) - q(S)| > \epsilon\right) \leq \delta.$$

(ϵ, δ)-*Distribution Accurate* if

$$\mathbb{P}_{S,M}\left(|M(S,q) - q(D_{\mathcal{X}^n})| > \epsilon\right) \leq \delta,$$



Generalization and Accuracy

Definition (Sample / distribution Accuracy (multiple queries))

Given a sequence of k queries, a mechanism M will be called: (ϵ, δ)-k-*Sample Accurate*, if

$$\mathbb{P}_{S,M}\left(\max_{i \in k} |M(S, q_i) - q_i(S)| > \epsilon\right) \leq \delta$$

(ϵ, δ)-k-*Distribution Accurate*, if

$$\mathbb{P}_{S,M}\left(\max_{i \in k} |M(S, q_i) - q_i(D_{\mathcal{X}^n})| > \epsilon\right) \leq \delta.$$

Generalization Guarantees

In order to guarantee (ϵ, δ)-k Distribution Accuracy, the sample size n should be:

Generalization Guarantees

In order to guarantee (ϵ, δ)-k Distribution Accuracy, the sample size n should be:

- Independent queries: $n = C\frac{\ln\left(\frac{k}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$

Generalization Guarantees

In order to guarantee (ϵ, δ)-k Distribution Accuracy, the sample size n should be:

- Independent queries: $n = C\frac{\ln\left(\frac{k}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$
- Re-sampling: $n = C\frac{k \cdot \ln\left(\frac{1}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$

Generalization Guarantees

In order to guarantee (ϵ, δ)-k Distribution Accuracy, the sample size n should be:

- Independent queries: $n = C\frac{\ln\left(\frac{k}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$
- Re-sampling: $n = C\frac{k \cdot \ln\left(\frac{1}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$
- Private efficient: $n = C\frac{\sqrt{k} \cdot \ln^{\frac{3}{2}}\left(\frac{k}{\epsilon\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^4$

Generalization Guarantees

In order to guarantee (ϵ, δ)-k Distribution Accuracy, the sample size n should be:

- Independent queries: $n = C\frac{\ln\left(\frac{k}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$
- Re-sampling: $n = C\frac{k \cdot \ln\left(\frac{1}{\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^0$
- Private efficient: $n = C\frac{\sqrt{k} \cdot \ln^{\frac{3}{2}}\left(\frac{k}{\epsilon\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^2}$, $C \approx 10^4$
- Private inefficient: $n = C\frac{\ln(|\mathcal{X}|)\ln(k)\ln^{\frac{3}{2}}\left(\frac{1}{\epsilon\delta}\right)}{\left(\frac{\epsilon}{\Delta}\right)^3}$, $C \approx 10^6$

What is Needed for Adaptive Generalization?

In order to ensure Adaptive Generalization, we must find a stability notions with the following properties:

What is Needed for Adaptive Generalization?

In order to ensure Adaptive Generalization, we must find a stability notions with the following properties:

- Post-Processing: Any subsequent analysis by the adversary over the mechanism's responses, remains stable

What is Needed for Adaptive Generalization?

In order to ensure Adaptive Generalization, we must find a stability notions with the following properties:

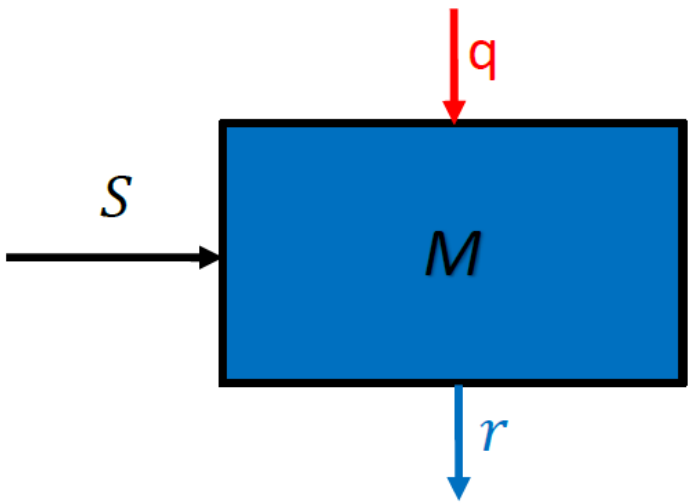
- Post-Processing: Any subsequent analysis by the adversary over the mechanism's responses, remains stable
- Composition: The stability guarantees of the mechanism for one query, extend to multiple adaptively chosen ones

What is Needed for Adaptive Generalization?

In order to ensure Adaptive Generalization, we must find a stability notions with the following properties:

- Post-Processing: Any subsequent analysis by the adversary over the mechanism's responses, remains stable
- Composition: The stability guarantees of the mechanism for one query, extend to multiple adaptively chosen ones
- Generalization: A stable mechanism implies generalization guarantees

We Have a Winner! - Differential Privacy



Definition (Differential Privacy)

A mechanism $M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{R}$ will be called (ϵ, δ) Differentially Private (or DP for short), if for any $S_1, S_2 \in \mathcal{X}^n$ which differ in one element, a query q, and a subset $R \subseteq \mathcal{R}$,

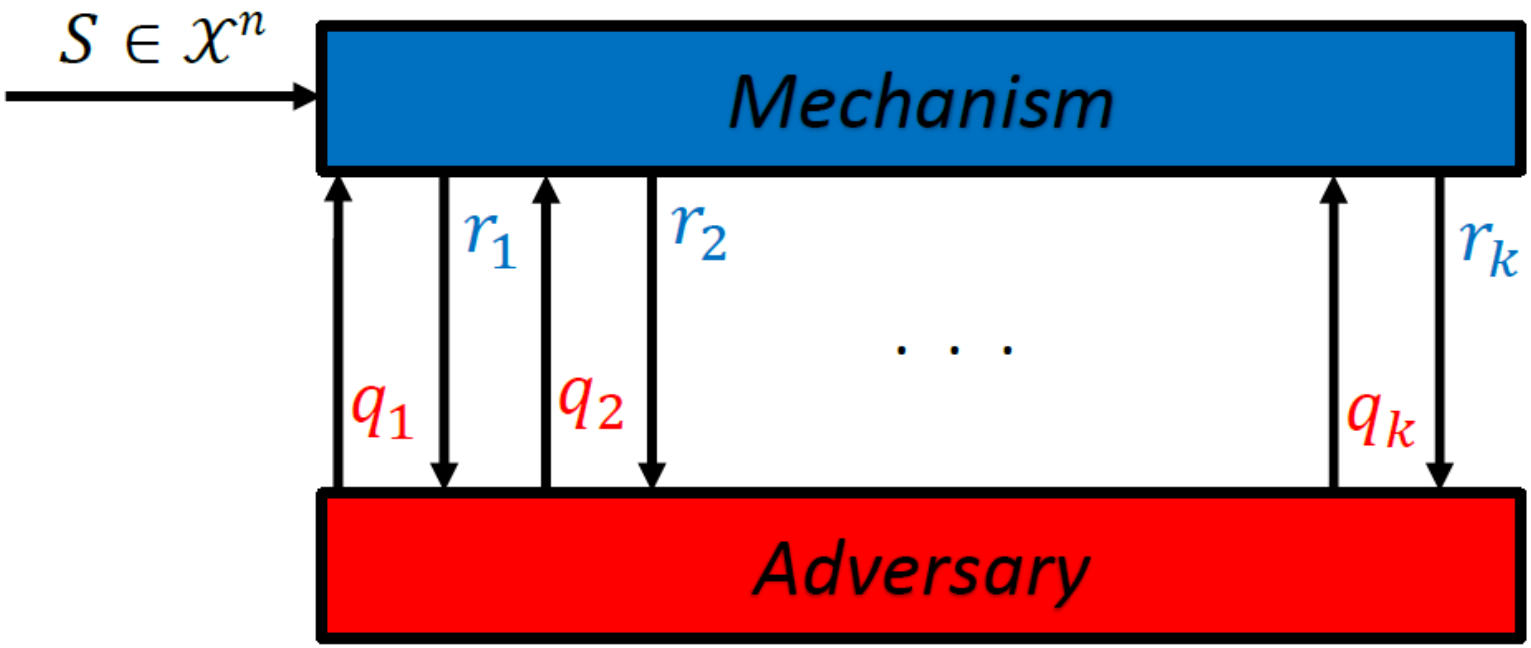
$$\mathbb{P}\left(M\left(S_1, q\right) \in R\right) \leq e^{\epsilon} \cdot \mathbb{P}\left(M\left(S_2, q\right) \in R\right) + \delta$$

Differential Privacy as a Stability Notion - Post-Processing

Theorem (DP holds under Post-Processing)

If a mechanism $M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{R}$ is (ϵ, δ)-DP, then for any arbitrary range $\mathcal{U}$ and (possibly non-deterministic) function $f : \mathcal{R} \to \mathcal{U}$, we have that $f \circ M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{U}$ is (ϵ, δ)-DP as well.

Differential Privacy as a Stability Notion - Adaptivity



Differential Privacy as a Stability Notion - Adaptivity

Adaptive Mechanism $\mathrm{Adp}_{M,k}$

Input: $S \in \mathcal{X}^n$, $A \in \mathcal{A}$

Output: $v_k \in \mathcal{V}_k$

$v_0 \leftarrow \emptyset$

for $i \in [k]$:

  $q_i \leftarrow A(v_{i-1})$

  $r_i \leftarrow M(S, q_i)$

  $v_i \leftarrow (v_{i-1}, r_i)$

return v_k

$$\mathcal{V}_k = \mathcal{R}^k,\ A : \mathcal{V}^* \to \mathcal{Q}$$

Differential Privacy as a Stability Notion - Adaptivity

Definition (k-fold Differential Privacy)

A mechanism $M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{R}$ will be called (ϵ, δ) k-fold Differentially Private (or k-DP for short), if for any $S_1, S_2 \in \mathcal{X}^n$ which differ in one element, an adversary A, and a subset $V \subseteq \mathcal{V}_k$,

$$\mathbb{P}\left(\mathrm{Adp}_{M,k}\left(S_1, A\right) \in V\right) \leq e^{\epsilon} \cdot \mathbb{P}\left(\mathrm{Adp}, \mathrm{k}_M\left(S_2, A\right) \in V\right) + \delta$$

Differential Privacy as a Stability Notion - Adaptivity

Definition (k-fold Differential Privacy)

A mechanism $M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{R}$ will be called (ϵ, δ) k-fold Differentially Private (or k-DP for short), if for any $S_1, S_2 \in \mathcal{X}^n$ which differ in one element, an adversary A, and a subset $V \subseteq \mathcal{V}_k$,

$$\mathbb{P}\left(\mathrm{Adp}_{M,k}\left(S_1, A\right) \in V\right) \leq e^{\epsilon} \cdot \mathbb{P}\left(\mathrm{Adp, k}_M\left(S_2, A\right) \in V\right) + \delta$$

Theorem (DP Adaptively composes)

For any given $k \in \mathbb{N}$, if a mechanism $M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{R}$ is (ϵ, δ)-DP, then it is $(\epsilon', (k+1)\,\delta)$-$k$-DP, for $\epsilon' = \epsilon\sqrt{2k\ln\left(\frac{1}{\delta}\right)} + k\epsilon^2$.

Differential Privacy as a Stability Notion - Generalization [DFHPRR15b, BNSSSU16]

Theorem (DP implies generalization)

For any $0 \leq \epsilon, \delta < 0.1$*, if a mechanism* $M : \mathcal{X}^n \times \mathcal{Q}_\Delta \to \mathbb{R}$ *is both:*
$\left(\frac{\epsilon}{64\Delta}, \frac{\epsilon\delta}{32\Delta}\right)$*-k-DP and*
$\left(\frac{\epsilon}{8}, \frac{\epsilon\delta}{16\Delta}\right)$*-k-Sample Accurate,*
then it is (ϵ, δ)*-k-Distribution Accurate.*

Generalization of Expectation

Theorem (DP implies generalization of expectation)

For any $0 \leq \epsilon$, $0 \leq \delta < 0.1$, *if a mechanism* $M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{Q}_\Delta$ *is* (ϵ, δ)*-DP (we will denote the output queries as* r_q *for clarity), then*

$$\left| \mathop{\mathbb{E}}_{r_q = M(S,q)} \left[r_q\left(D_{\mathcal{X}^n}\right) - r_q\left(S\right) \right] \right| \leq 2\Delta\left(e^{\epsilon} - 1 + \delta\right)$$

The Monitor Mechanism

Monitor mechanism $\mathrm{Mon}_{M,k}$

Input: $\{S\} \in (\mathcal{X}^n)^T, A \in \mathcal{A}$

Output: $q \in \mathcal{Q}, r \in \mathbb{R}, t \in T$

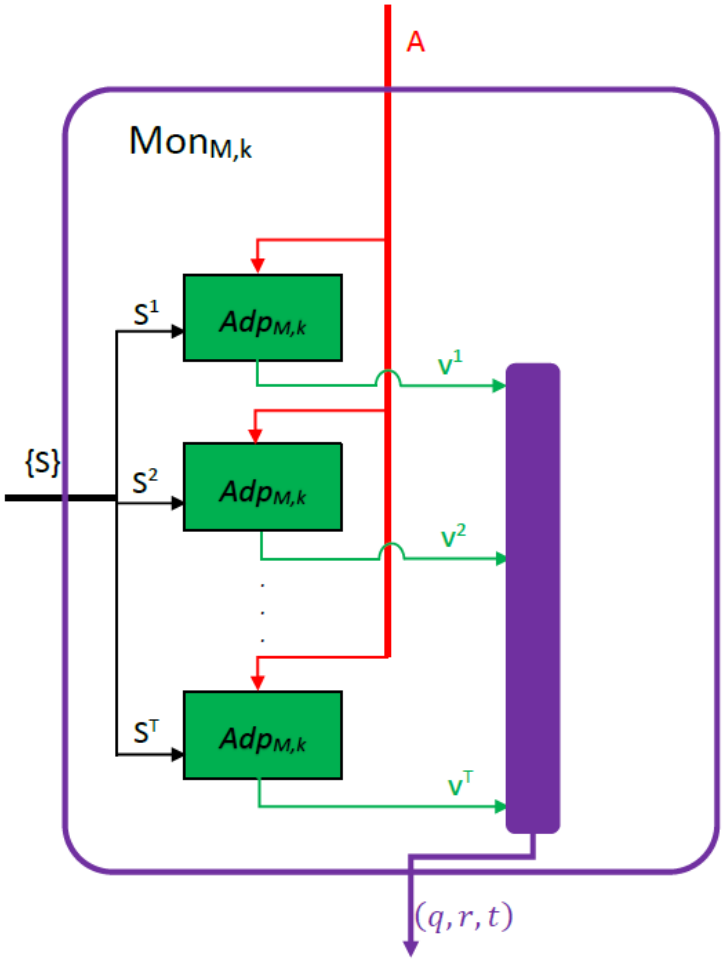
for $t = 1, \ldots, T$:

$v^t \leftarrow \mathrm{Adp}_{M,k}(S_t, A)$

$(q^t, r^t) \leftarrow \underset{(q,r) \in v^t}{\arg\max} |q(D_{\mathcal{X}^n}) - r|$

$t^* \leftarrow \underset{t \in [T]}{\arg\max} \left(q^t(D_{\mathcal{X}^n}) - r^t\right)$

return $\left(q^{t^*}, r^{t^*}, t^*\right)$



Properties of the Monitor Mechanism

Theorem (Mon generalization)

For any $0 \leq \epsilon$, $0 \leq \delta < 1$, *if a mechanism* M *is* (ϵ, δ)*-*k*-DP, then*

$$\left| \mathbb{E}\left[q\left(D_{\mathcal{X}^n}\right) - q\left(S_t\right)\right] \right| \leq 2\Delta\left(e^{\epsilon} - 1 + T\delta\right)$$

where $(q, r, t) = Mon_{M,k}\left(\{S\}, A\right)$*, and the expectation is taken over the sampling of* S *and the internal probability of* M

Properties of the Monitor Mechanism

Theorem (Mon is Sample Accuracy)

For any $0 \leq \epsilon$, $0 \leq \delta < 1$, *if a mechanism* M *is* (ϵ, δ)*-*k*-Sample Accurate, then*

$$\left| \mathbb{E}\left[r - q\left(S_t\right)\right] \right| \leq \epsilon + 2\Delta T \delta$$

where $(q, r, t) = Mon_{M,k}\left(\{S\}, A\right)$*, and the expectation is taken over the sampling of* S *and the internal probability of* M

Properties of the Monitor Mechanism

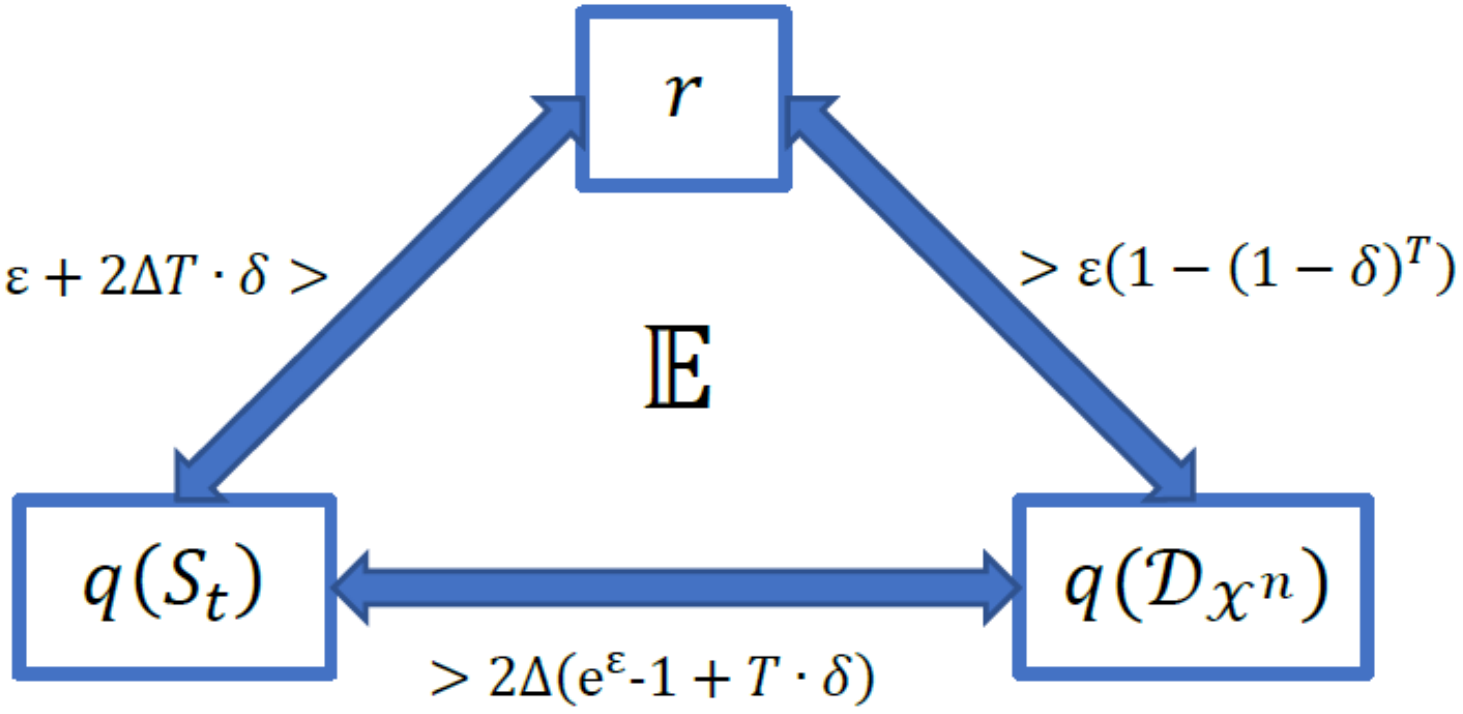
Theorem (Mon Distribution Accuracy)

For any $0 \leq \epsilon$, $0 \leq \delta < 1$, *if a mechanism* M *is not* (ϵ, δ)*-*k*-Accurate Accurate, then*

$$\left| \mathbb{E}\left[q\left(D_{\mathcal{X}^n}\right) - r\right] \right| \geq \epsilon \left(1 - (1-\delta)^T\right)$$

where $(q, r, t) = Mon_{M,k}\left(\{S\}, A\right)$*, and the expectation is taken over the sampling of* S *and the internal probability of* M

Generalization in Probability Proof



Summary

- DP implies generalization
- DP holds under post processing
- DP composes sub linearly

Combining these three properties, we get a generalization guarantee that holds for any adaptive learning process

Beyond Differential Privacy

Beyond Differential Privacy

- DP is a worst case type of guarantee, in particular with regards to the sample sets

Beyond Differential Privacy

- DP is a worst case type of guarantee, in particular with regards to the sample sets
- Learning guarantees are based on sampling

Beyond Differential Privacy

- DP is a worst case type of guarantee, in particular with regards to the sample sets
- Learning guarantees are based on sampling
- Other - "weaker" - stability notion were proposed, such as Max Information

Beyond Differential Privacy

- DP is a worst case type of guarantee, in particular with regards to the sample sets
- Learning guarantees are based on sampling
- Other - "weaker" - stability notion were proposed, such as Max Information
- A natural question that arises is what is the "weakest" stability notion which still ensures generalization

Beyond Differential Privacy

- DP is a worst case type of guarantee, in particular with regards to the sample sets
- Learning guarantees are based on sampling
- Other - "weaker" - stability notion were proposed, such as Max Information
- A natural question that arises is what is the "weakest" stability notion which still ensures generalization
- Lately Ligett and Shenfeld proposed a new stability notion - Local Statistical Stability, and proved it is both necessary and sufficient for generalization

Questions?

Email: moshe.shenfeld@mail.huji.ac.il