Generalization through differential privacy

Jung, Ligett, Neel, Roth, Sharifi-Malvajerdi, Shenfeld '19

Geelon So
(agso@eng.ucsd.edu)

November 1, 2019

An elementary proof of the transfer theorem

A New Analysis of Differential Privacy's Generalization Guarantees

Christopher Jung[*] Katrina Ligett[†] Seth Neel[‡] Aaron Roth[§]
Saeed Sharifi-Malvajerdi Moshe Shenfeld

September 10, 2019

Abstract

We give a new proof of the "transfer theorem" underlying adaptive data analysis: that any mechanism for answering adaptively chosen statistical queries that is differentially private and sample-accurate is also accurate out-of-sample. Our new proof is elementary and gives structural insights that we expect will be useful elsewhere. We show: 1) that differential privacy ensures that the expectation of any query on the *posterior distribution* on datasets induced by the transcript of the interaction is close to its true value on the data distribution, and 2) sample accuracy on its own ensures that any query answer produced by the mechanism is close to its posterior expectation with high probability. This second claim follows from a thought experiment in which we imagine that the dataset is resampled from the posterior distribution after the mechanism has committed to its answers. The transfer theorem then follows by summing these two bounds, and in particular, avoids the "monitor argument" used to derive high probability bounds in prior work.

An upshot of our new proof technique is that the concrete bounds we obtain are substantially better than the best previously known bounds, even though the improvements are in the constants, rather than the asymptotics (which are known to be tight). As we show, our new bounds outperform the naive "sample-splitting" baseline at dramatically smaller dataset sizes compared to the previous state of the art, bringing techniques from this literature closer to practicality.

https://arxiv.org/abs/1909.03577.

Adaptive data analysis

Given a dataset S, we wish to perform a sequence of analyses,

$$q_1, q_2, \ldots, q_k$$

where the queries q_t adapt to the previous answers.

- **Problem:** reusing data can lead to overfitting

Adaptive data analysis

Given a dataset S, we wish to perform a sequence of analyses,

$$q_1, q_2, \ldots, q_k$$

where the queries q_t adapt to the previous answers.

- **Problem:** reusing data can lead to overfitting
- **Baseline solution:** partition data into k parts

Adaptive data analysis

Given a dataset S, we wish to perform a sequence of analyses,

$$q_1, q_2, \dots, q_k$$

where the queries q_t adapt to the previous answers.

- **Problem:** reusing data can lead to overfitting
- **Baseline solution:** partition data into k parts
  - amount of data grows linearly with number of queries

Connection to differential privacy

A **differentially private mechanism** is a randomized algorithm where the distributions of the outputs computed from similar datasets are also similar.

Connection to differential privacy

A **differentially private mechanism** is a randomized algorithm where the distributions of the outputs computed from similar datasets are also similar.

- Intuition: if an analysis is differentially private, then answers generalize since they don't depend closely on the particular dataset

Transfer theorem

Informal theorem. A differentially private analysis that has high *in-sample* accuracy must also have high *out-of-sample* accuracy.

Transfer theorem: prior works

Proofs

- Original connection to differential privacy: [DFHPRR15]
- Best analysis via the 'monitor argument': amount of data grows $\sqrt{k}$ with respect to number of queries k. [BNSSSU16]

Lower bounds

- The analysis in [BNSSSU16] is asymptotically tight, as seen in [HU14], [SU15]. This work [JLNRSS19] improves concrete bounds through new proof techniques.

Transfer theorem: this work

Adaptive data analysis consists of the following generating process:

1. collect a dataset S from an underlying distribution $\mathcal{P}^n$

Transfer theorem: this work

Adaptive data analysis consists of the following generating process:

1. collect a dataset S from an underlying distribution $\mathcal{P}^n$
2. interact with data to produce transcript Π of interaction

Transfer theorem: this work

The **Bayesian resampling lemma** states that this generating process is equivalent to the following:

1. sample an interaction Π

Transfer theorem: this work

The **Bayesian resampling lemma** states that this generating process is equivalent to the following:

1. sample an interaction Π
2. sample a dataset S' from the posterior distribution conditioned on the interaction Π

Transfer theorem: this work

Sketch of argument

1. The Bayesian resampling lemma implies that adaptive analysis is equivalent to non-adaptive analysis over posterior.

Transfer theorem: this work

Sketch of argument

1. The Bayesian resampling lemma implies that adaptive analysis is equivalent to non-adaptive analysis over posterior.
2. Relate analysis of dataset drawn from $\mathcal{P}^n$ to one drawn from the posterior through *posterior sensitivity*.

Transfer theorem: this work

Sketch of argument

1. The Bayesian resampling lemma implies that adaptive analysis is equivalent to non-adaptive analysis over posterior.
2. Relate analysis of dataset drawn from $\mathcal{P}^n$ to one drawn from the posterior through *posterior sensitivity*.
3. If an analysis has high in-sample accuracy and low posterior sensitivity, then it generalizes well.

Transfer theorem: this work

Sketch of argument

1. The Bayesian resampling lemma implies that adaptive analysis is equivalent to non-adaptive analysis over posterior.
2. Relate analysis of dataset drawn from $\mathcal{P}^n$ to one drawn from the posterior through *posterior sensitivity*.
3. If an analysis has high in-sample accuracy and low posterior sensitivity, then it generalizes well.
4. Differential privacy implies low posterior sensitivity.

Preliminaries: setting

Ingredients

- $\mathcal{X}$ an abstract data domain

Preliminaries: setting

Ingredients

- $\mathcal{X}$ an abstract data domain
- $\mathcal{P}$ a distribution over $\mathcal{X}$

Preliminaries: setting

Ingredients

- $\mathcal{X}$ an abstract data domain
- $\mathcal{P}$ a distribution over $\mathcal{X}$
- $S = \{S_i\}_{i=1}^{n} \in \mathcal{X}^n$ a collection of n data records

Preliminaries: setting

Ingredients

- $\mathcal{X}$ an abstract data domain
- $\mathcal{P}$ a distribution over $\mathcal{X}$
- $S = \{S_i\}_{i=1}^n \in \mathcal{X}^n$ a collection of n data records
- Q is a family of queries q

Preliminaries: setting

Ingredients

- $\mathcal{X}$ an abstract data domain
- $\mathcal{P}$ a distribution over $\mathcal{X}$
- $S = \{S_i\}_{i=1}^{n} \in \mathcal{X}^n$ a collection of n data records
- Q is a family of queries q
- $q : \mathcal{X}^* \to [0, 1]$ a linear data query:

$$q(S) = \frac{1}{n} \sum_{i=1}^{n} q(S_i).$$

Preliminaries: setting

Additional notation

- Let S_i denote the random variable and x its realization

Preliminaries: setting

Additional notation

- Let S_i denote the random variable and x its realization
- If $\mathcal{D}$ is a distribution over datasets, $q(\mathcal{D})$ is the expectation:

$$q(\mathcal{D}) = \mathop{\mathbb{E}}_{S \sim \mathcal{D}}[q(S)].$$

Preliminaries: adaptive analysis

Interacting parties

- $\mathcal{A} : \mathbb{R}^* \to Q^*$ an analyst

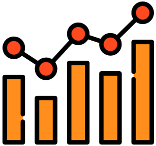
Preliminaries: adaptive analysis

Interacting parties

- $\mathcal{A} : \mathbb{R}^* \to Q^*$ an analyst
- $M : \mathcal{X}^n \times Q^* \to \mathbb{R}^*$ a (possibly stateful) statistical estimator

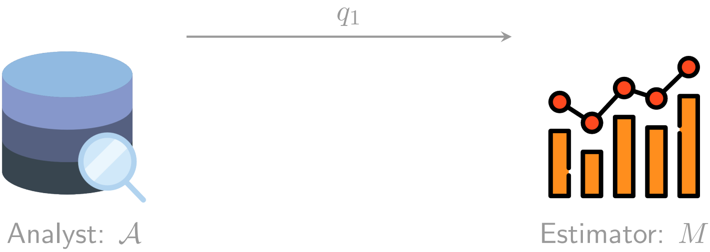
Preliminaries: adaptive analysis

Analyst: $\mathcal{A}$

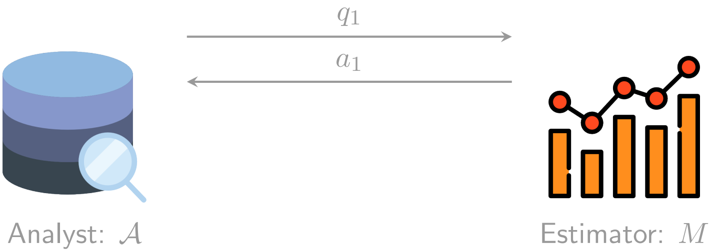


Estimator: M

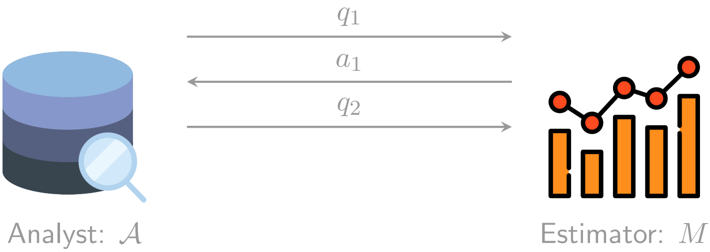
Preliminaries: adaptive analysis



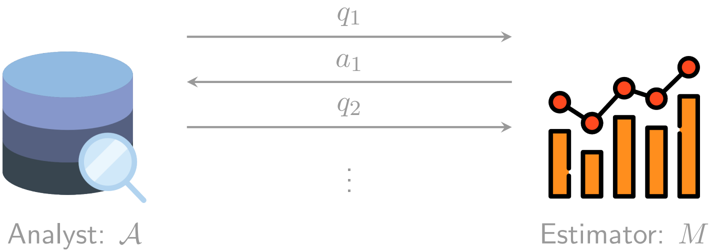
Preliminaries: adaptive analysis



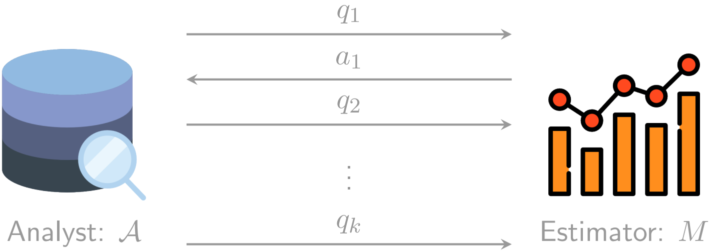
Preliminaries: adaptive analysis



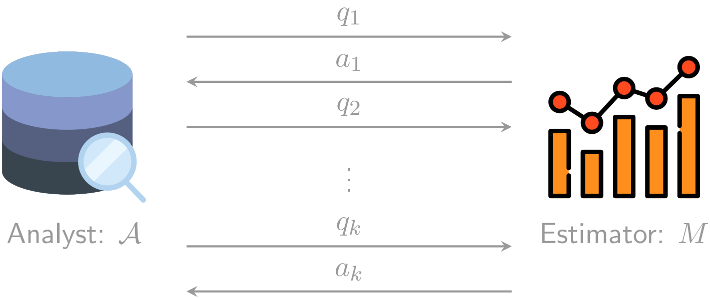
Preliminaries: adaptive analysis



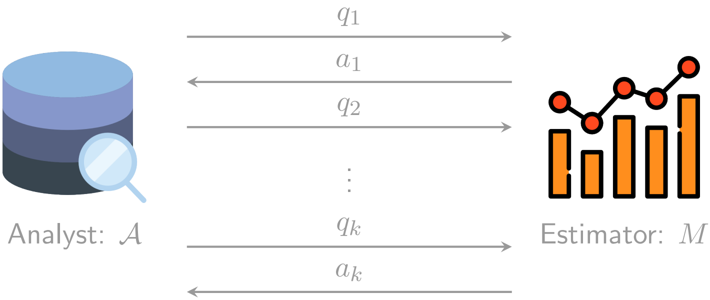
Preliminaries: adaptive analysis



Preliminaries: adaptive analysis



Preliminaries: adaptive analysis



Transcript: $\pi = \{(q_1, a_1), \ldots, (q_k, a_k)\}$.

Preliminaries: adaptive analysis

Notation

- $\pi \in \mathbf{\Pi} = (Q \times \mathbb{R})^*$ the transcript of interaction

Preliminaries: adaptive analysis

Notation

- $\pi \in \mathbf{\Pi} = (Q \times \mathbb{R})^*$ the transcript of interaction
- Π denotes the random variable and π its realizations

Preliminaries: adaptive analysis

Notation

- $\pi \in \mathbf{\Pi} = (Q \times \mathbb{R})^*$ the transcript of interaction
- Π denotes the random variable and π its realizations
- $\mathrm{Interact}(M, \mathcal{A}; S)$ is the transcript of the interaction

Preliminaries: adaptive analysis

Notation

- $\pi \in \mathbf{\Pi} = (Q \times \mathbb{R})^*$ the transcript of interaction
- Π denotes the random variable and π its realizations
- $\mathrm{Interact}(M, \mathcal{A}; S)$ is the transcript of the interaction
  - for brevity, abbreviate $\mathrm{Interact}(M, \mathcal{A}; S)$ by $I(S)$

Preliminaries: adaptive analysis

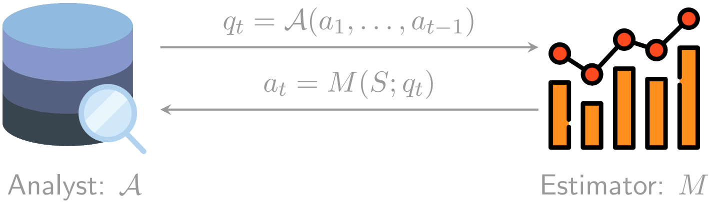


Figure 1: The interaction between $\mathcal{A}$ and M on dataset S generates a transcript $\text{Interact}(M, \mathcal{A}; S) \in \mathbf{\Pi}$.

Preliminaries: product and posterior distribution

We consider datasets S drawn from two distributions, $\mathcal{P}^n$ and $\mathcal{Q}_\pi$:

- $\mathcal{P}^n$ is the product distribution over datasets with n records

Preliminaries: product and posterior distribution

We consider datasets S drawn from two distributions, $\mathcal{P}^n$ and $\mathcal{Q}_\pi$:

- $\mathcal{P}^n$ is the product distribution over datasets with n records
- if $\pi \in \mathbf{\Pi}$ is a transcript, the *posterior* $\mathcal{Q}_\pi$ is the conditional distribution:

$$\mathcal{Q}_\pi = (\mathcal{P}^n) | \mathrm{Interact}(M, \mathcal{A}; S) = \pi.$$

Roadmap

Sketch

1. The Bayesian resampling lemma implies that adaptive analysis is equivalent to non-adaptive analysis over posterior.
2. Relate analysis of dataset drawn from $\mathcal{P}^n$ to one drawn from the posterior through *posterior sensitivity*.
3. If an analysis has high in-sample accuracy and low posterior sensitivity, then it generalizes well.
4. Differential privacy implies low posterior sensitivity.

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Relate analysis of dataset drawn from $\mathcal{P}^n$ to one drawn from the posterior through *posterior sensitivity*.
3. If an analysis has high in-sample accuracy and low posterior sensitivity, then it generalizes well.
4. Differential privacy implies low posterior sensitivity.

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. If an analysis has high in-sample accuracy and low posterior sensitivity, then it generalizes well.
4. Differential privacy implies low posterior sensitivity.

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. Prove general transfer theorem.
4. Differential privacy implies low posterior sensitivity.

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. Prove general transfer theorem.
4. Specialize transfer theorem to differential privacy

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. Prove general transfer theorem.
4. Specialize transfer theorem to differential privacy

Bayesian resampling lemma

Lemma

Let $E \subset \mathcal{X}^n \times \mathbf{\Pi}$ be any event. Then:

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[(S, \Pi) \in E \right] = \Pr_{S \sim \mathcal{P}, \Pi \sim I(S), S' \sim \mathcal{Q}_\Pi} \left[(S', \Pi) \in E \right].$$

Bayesian resampling lemma

Proof sketch.

- Expand out the probability.

$$\Pr_{S\sim\mathcal{P},\Pi\sim I(S),S'\sim\mathcal{Q}_\Pi}\left[(S',\Pi)\in E\right]$$

$$=\sum_{\pi}\sum_{x'}\Pr[\Pi=\pi]\Pr_{S'\sim\mathcal{Q}_\pi}[S'=x']\mathbf{1}[(x',\pi)\in E].$$

- Apply Bayes rule.
- Collapse terms.

□

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. Prove general transfer theorem.
4. Specialize transfer theorem to differential privacy

Definitions

1. (ϵ, δ)-posterior sensitivity
2. (α, β)-sample accuracy and (α, β)-distributional accuracy
3. (ϵ, δ)-differential privacy

Definitions

Definition

An interaction $\mathrm{Interact}(M, \mathcal{A}; \cdot)$ *is* (ϵ, δ)-**posterior sensitive** *if for every data distribution* $\mathcal{P}$*:*

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(\mathcal{P}^n) - q_j(\mathcal{Q}_\Pi)| \geq \epsilon \right] \leq \delta.$$

Definitions

Definition

*A statistical estimator M satisfies (α, β)-**sample accuracy** if for every data analyst $\mathcal{A}$ and every data distribution $\mathcal{P}$,*

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(S) - a_j| \geq \alpha \right] \leq \beta.$$

Definitions

Definition

*A statistical estimator M satisfies (α, β)-***distributional accuracy** *if for every data analyst $\mathcal{A}$ and every data distribution $\mathcal{P}$,*

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(\mathcal{P}^n) - a_j| \geq \alpha \right] \leq \beta.$$

Definitions

Definition

Two datasets $S, S' \in \mathcal{X}^n$ *are* **neighbors** *if they differ in at most one coordinate.*

Definition

An interaction $\mathrm{Interact}(M, \mathcal{A}; S)$ *satisfies* (ϵ, δ)-**differential privacy** *if for all data analysts* $\mathcal{A}$*, pairs of neighboring datasets* $S, S' \in \mathcal{X}^n$*, and for all events* $E \subset \mathbf{\Pi}$*,*

$$\Pr\left[I(S) \in E\right] \leq e^{\epsilon} \cdot \Pr\left[I(S') \in E\right] + \delta.$$

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. Prove general transfer theorem.
4. Specialize transfer theorem to differential privacy

General transfer theorem

Theorem (General transfer theorem)

Let $\mathrm{Interact}(M, \mathcal{A}; \cdot)$ *be* (α, β)*-sample accurate and* (ϵ, δ)*-posterior sensitive. Then, it is* (α', β')*-distributionally accurate, where* $\alpha' = \alpha + c + \epsilon$ *and* $\beta' = \frac{\beta}{c} + \delta$ *and* $c > 0$.

Accuracy over samples to accuracy over posterior

Lemma

Let M be (α, β)-sample accurate. Then for all $c > 0$,

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |a_j - q_j(\mathcal{Q}_\Pi)| > \alpha + c \right] \leq \frac{\beta}{c}.$$

Accuracy over samples to accuracy over posterior

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(S) - a_j| \geq \alpha \right] \leq \beta$$

$$\Downarrow$$

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |a_j - q_j(\mathcal{Q}_\Pi)| > \alpha + c \right] \leq \frac{\beta}{c}.$$

Accuracy over samples to accuracy over posterior

Proof sketch. First, we obtain a one-sided tail bound:

$$\Pr_{S\sim\mathcal{P}^n,\Pi\sim I(S)}\left[\max_j a_j - q_j(\mathcal{Q}_\Pi) > \alpha + c\right]$$

$$\leq$$

$$\frac{1}{c}\mathop{\mathbb{E}}_{S\sim\mathcal{P}^n,\Pi\sim I(S)}\left[\Pr_{S'\sim\mathcal{Q}_\Pi}\left[\max_j a_j - q_j(S') > \alpha\right]\right],$$

almost directly from an application of **Markov's inequality**.

Accuracy over samples to accuracy over posterior

Proof sketch (cont). But we can rewrite this upper bound as the following probability:

$$\frac{1}{c} \Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S), S' \sim \mathcal{Q}_\Pi} \left[\max_j a_j - q_j(S') > \alpha \right],$$

Accuracy over samples to accuracy over posterior

Proof sketch (cont). But we can rewrite this upper bound as the following probability:

$$\frac{1}{c} \Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S), S' \sim \mathcal{Q}_\Pi} \left[\max_j a_j - q_j(S') > \alpha \right],$$

which can be converted via the **Bayesian resampling lemma** into:

$$\frac{1}{c} \Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j a_j - q_j(S) > \alpha \right].$$

Accuracy over samples to accuracy over posterior

Proof sketch (cont). The same analysis holds for the other tail; combining them, we obtain the **two-sided tail bound**:

$$\Pr_{S\sim\mathcal{P}^n,\Pi\sim I(S)}\left[\left|\max_j a_j - q_j(\mathcal{Q}_\Pi)\right| > \alpha + c\right]$$

$$\leq$$

$$\frac{1}{c}\underbrace{\Pr_{S\sim\mathcal{P}^n,\Pi\sim I(S)}\left[\left|\max_j a_j - q_j(S)\right| > \alpha\right]}_{\leq\beta}.$$

□

Proof of general transfer theorem

Proof.

Lemma shows that if M is (α, β)-sample accurate, then

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |a_j - q_j(\mathcal{Q}_\Pi)| > \alpha + c \right] \leq \frac{\beta}{c}.$$

Proof of general transfer theorem

Proof.

Lemma shows that if M is (α, β)-sample accurate, then

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |a_j - q_j(\mathcal{Q}_\Pi)| > \alpha + c \right] \leq \frac{\beta}{c}.$$

If $\mathrm{Interact}(M, \mathcal{A}; \cdot)$ is (ϵ, δ)-posterior sensitive, then

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(\mathcal{P}^n) - q_j(\mathcal{Q}_\Pi)| \geq \epsilon \right] \leq \delta.$$

Proof of general transfer theorem

Proof.

Lemma shows that if M is (α, β)-sample accurate, then

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |a_j - q_j(\mathcal{Q}_\Pi)| > \alpha + c \right] \leq \frac{\beta}{c}.$$

If $\mathrm{Interact}(M, \mathcal{A}; \cdot)$ is (ϵ, δ)-posterior sensitive, then

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(\mathcal{P}^n) - q_j(\mathcal{Q}_\Pi)| \geq \epsilon \right] \leq \delta.$$

By the **triangle inequality**,

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\max_j |q_j(\mathcal{P}^n) - a_j)| > \alpha + c + \epsilon \right] < \frac{\beta}{c} + \delta.$$

□

Roadmap

Sketch

1. Bayesian resampling lemma.
2. Define posterior sensitivity.
3. Prove general transfer theorem.
4. Specialize transfer theorem to differential privacy

Differential privacy implies low posterior sensitivity

Let M be a statistical estimator for a family Q of linear queries.

Lemma

If M is (ϵ, δ)-differentially private, then for any data distribution $\mathcal{P}$ and any analyst $\mathcal{A}$, it is (ϵ', δ')-posterior sensitive, for all $c > 0$ and $\epsilon' = e^{\epsilon} - 1 + 2c$ and $\delta' = \delta/c$.

Differential privacy implies low posterior sensitivity

Notation. If $S \sim \mathcal{P}^n$, let S_i be uniformly random record from S.

Proof sketch. Proceed by contradiction. We will aim to define an event $E \subset \mathcal{X}^n \times \mathbf{\Pi}$ so that a high posterior sensitivity implies:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$

is large, but differential privacy implies that it is small.

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Notice that this difference can be split apart in two ways:[1]

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi)\in E]-\Pr_{S_i\otimes\Pi}[(S_i,\Pi)\in E]$$

[1] On *ad hoc* notation: let $E(\pi)$ be the set $\{x\in\mathcal{X}:(x,\pi)\in E\}$, and similarly for $E(\pi)$. Note that $E\subset\mathcal{X}\times\mathbf{\Pi}$.

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Notice that this difference can be split apart in two ways:[1]

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$= \sum_{\pi \in \mathbf{\Pi}} \Pr[\Pi = \pi] \sum_{x \in E(\pi)} \left(\Pr\left[S_i = x | \Pi = \pi\right] - \Pr[S_i = x]\right)$$

[1] On *ad hoc* notation: let $E(\pi)$ be the set $\{x \in \mathcal{X} : (x,\pi) \in E\}$, and similarly for $E(\pi)$. Note that $E \subset \mathcal{X} \times \mathbf{\Pi}$.

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Notice that this difference can be split apart in two ways:[1]

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$= \sum_{\pi \in \mathbf{\Pi}} \Pr[\Pi = \pi] \sum_{x \in E(\pi)} \left(\Pr[S_i = x | \Pi = \pi] - \Pr[S_i = x]\right)$$
$$= \sum_{x \in \mathcal{X}} \Pr[S_i = x]\big(\Pr[\Pi \in E(x) | S_i = x] - \Pr[\Pi \in E(x)]\big).$$

[1] On *ad hoc* notation: let $E(\pi)$ be the set $\{x \in \mathcal{X} : (x,\pi) \in E\}$, and similarly for $E(\pi)$. Note that $E \subset \mathcal{X} \times \mathbf{\Pi}$.

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Notice that this difference can be split apart in two ways:[1]

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$= \sum_{\pi \in \mathbf{\Pi}} \Pr[\Pi = \pi] \sum_{x \in E(\pi)} \left(\Pr\left[S_i = x | \Pi = \pi\right] - \Pr[S_i = x]\right)$$
$$= \sum_{x \in \mathcal{X}} \Pr[S_i = x] \big(\Pr\left[\Pi \in E(x) | S_i = x\right] - \Pr[\Pi \in E(x)] \big).$$

The first relates well to posterior sensitivity while the second to differential privacy.

[1] On *ad hoc* notation: let $E(\pi)$ be the set $\{x \in \mathcal{X} : (x,\pi) \in E\}$, and similarly for $E(\pi)$. Note that $E \subset \mathcal{X} \times \mathbf{\Pi}$.

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Suppose that M has high posterior sensitivity, where:

$$\Pr\left[\max_j |q_j(\mathcal{Q}_\Pi) - q_j(\mathcal{P}^n)| > \alpha\right] > \frac{\delta}{c}.$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Suppose that M has high posterior sensitivity, where:

$$\Pr\left[\max_j |q_j(\mathcal{Q}_\Pi) - q_j(\mathcal{P}^n)| > \alpha\right] > \frac{\delta}{c}.$$

One of the tails must have at least half of the probability mass. Without loss of generality, assume:

$$\Pr\left[\max_j q_j(\mathcal{Q}_\Pi) - q_j(\mathcal{P}^n) > \alpha\right] > \frac{\delta}{2c}.$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Define $\mathbf{\Pi}_\alpha$ to be the event:

$$\mathbf{\Pi}_\alpha = \left\{ \pi \in \mathbf{\Pi} : \max_j q_j(\mathcal{Q}_\pi) - q_j(\mathcal{P}^n) > \alpha \right\}.$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Define $\mathbf{\Pi}_\alpha$ to be the event:

$$\mathbf{\Pi}_\alpha = \left\{ \pi \in \mathbf{\Pi} : \max_j q_j(\mathcal{Q}_\pi) - q_j(\mathcal{P}^n) > \alpha \right\}.$$

These are the transcripts $\pi = \{(q_i, a_i)\}_{i=1}^k$ where one of the queries q_j distinguish between $\mathcal{Q}_\pi$ and $\mathcal{P}^n$ well. The previous slide just states that high posterior sensitivity implies:

$$\Pr\left[\Pi \in \mathbf{\Pi}_\alpha\right] > \frac{\delta}{2c}.$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Recall the first decomposition:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$= \sum_{\pi \in \mathbf{\Pi}} \Pr[\Pi = \pi] \sum_{x \in E(\pi)} \left(\Pr\left[S_i = x | \Pi = \pi\right] - \Pr[S_i = x]\right).$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Recall the first decomposition:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$= \sum_{\pi \in \mathbf{\Pi}_\alpha} \Pr[\Pi = \pi] \sum_{x \in E(\pi)} \left(\Pr\left[S_i = x | \Pi = \pi\right] - \Pr[S_i = x]\right).$$

If $E \subset \mathcal{X} \times \mathbf{\Pi}$ contains only transcripts from $\mathbf{\Pi}_\alpha$,

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Recall the first decomposition:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi)\in E] - \Pr_{S_i\otimes\Pi}[(S_i,\Pi)\in E]$$
$$= \sum_{\pi\in\boldsymbol{\Pi}_\alpha} \Pr[\Pi=\pi] \sum_{x\in E(\pi)} \left(\Pr\left[S_i=x|\Pi=\pi\right] - \Pr[S_i=x]\right).$$

If $E \subset \mathcal{X}\times\boldsymbol{\Pi}$ contains only transcripts from $\boldsymbol{\Pi}_\alpha$, and the **inner sum** is $o(1)$,

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Recall the first decomposition:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi)\in E] - \Pr_{S_i\otimes\Pi}[(S_i,\Pi)\in E] = \sum_{\pi\in\mathbf{\Pi}_\alpha}\Pr[\Pi=\pi]\sum_{x\in E(\pi)}\left(\Pr\left[S_i=x|\Pi=\pi\right]-\Pr[S_i=x]\right).$$

If $E\subset\mathcal{X}\times\mathbf{\Pi}$ contains only transcripts from $\mathbf{\Pi}_\alpha$, and the **inner sum** is $o(1)$, then posterior sensitivity gives a lower bound:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi)\in E] - \Pr_{S_i\otimes\Pi}[(S_i,\Pi)\in E] > o(1)\cdot\Pr\left[\Pi\in\mathbf{\Pi}_\alpha\right].$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). Construct E so that:

- The projection of E to $\mathbf{\Pi}$ is $\mathbf{\Pi}_\alpha$
- The terms in the inner sum is nonnegative

$$E(\pi) = \{x \in \mathcal{X} : \Pr\left[S_i = x | \Pi = \pi\right] - \Pr[S_i = x] > 0\}$$

Explicitly, define E to be:

$$E = \bigcup_{\pi \in \mathbf{\Pi}_\alpha} E(\pi) \times \{\pi\}.$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). If E defined this way, then:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi)\in E]-\Pr_{S_i\otimes\Pi}[(S_i,\Pi)\in E]$$
$$>o(1)\cdot\Pr[\Pi\in\mathbf{\Pi}_\alpha]$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). If E defined this way, then:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E] > \alpha \cdot \Pr[\Pi \in \mathbf{\Pi}_\alpha]$$

we obtain a good lower bound.

Differential privacy implies low posterior sensitivity

Proof sketch (cont). On the other hand, if M is differentially private, then the second decomposition:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$= \sum_{x \in \mathcal{X}} \Pr[S_i = x]\big(\Pr[\Pi \in E(x) | S_i = x] - \Pr[\Pi \in E(x)]\big).$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). On the other hand, if M is differentially private, then the second decomposition:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi)\in E] - \Pr_{S_i\otimes\Pi}[(S_i,\Pi)\in E] \leq \sum_{x\in\mathcal{X}} \Pr[S_i = x]\left[(e^{\epsilon}-1)\Pr\left[\Pi\in E(x)\right]+\delta\right]$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). This provides an upper bound:

$$\Pr_{(S_i,\Pi)}[(S_i,\Pi) \in E] - \Pr_{S_i \otimes \Pi}[(S_i,\Pi) \in E]$$
$$< ((e^{\epsilon} - 1) + 2c) \cdot \Pr[\Pi \in \mathbf{\Pi}_{\alpha}].$$

Differential privacy implies low posterior sensitivity

Proof sketch (cont). This provides an upper bound:

$$\Pr_{(S_i,\Pi)}\left[(S_i,\Pi)\in E\right] - \Pr_{S_i\otimes\Pi}\left[(S_i,\Pi)\in E\right] < \left((e^{\epsilon}-1)+2c\right)\cdot \Pr\left[\Pi\in\mathbf{\Pi}_{\alpha}\right].$$

We obtain a contradiction if $\alpha \geq (e^{\epsilon}-1)+2c$. Thus, a differentially private mechanism must also have low posterior sensitivity. □

Transfer theorem for differential privacy

Theorem

Suppose that M is (ϵ, δ)-differentially private and (α, β)-sample accurate for linear queries. Then for every analyst $\mathcal{A}$ and $c, d > 0$, it is also (α', β')-distributionally accurate.[2]

[2] $\alpha' = \alpha + (e^{\epsilon} - 1) + c + 2d$ and $\beta' = \frac{\beta}{c} + \frac{\delta}{d}$.

Application to the Gaussian mechanism

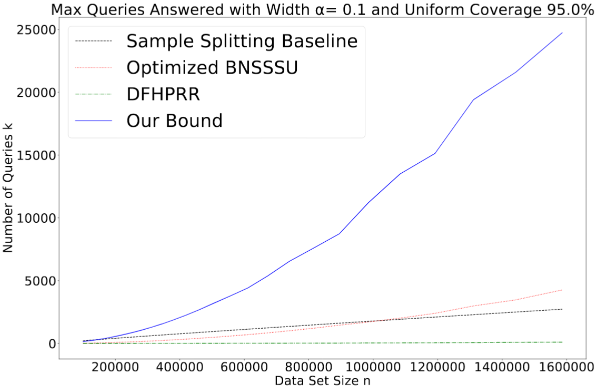


Figure 2: Comparison of lower bounds on the number of adaptive linear queries that can be answered using the Gaussian mechanism.

Extensions

- Recall in the proof that **accuracy over samples implies accuracy over posterior**, we made use of Markov's inequality to upper bound the error of analysis performed over the posterior. In the setting of $(\epsilon, 0)$-differential privacy, they obtain even better bounds by directly bounding the tail using a Chernoff-like concentration.
- In addition to linear queries, they also provide transfer theorems for **low sensitive** and **minimization** queries.

References

[HU14] Hardt, M., Ullman, J. "Preventing false discovery in interactive data analysis is hard." *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 2014.

[DFHPRR15] Dwork, C., et al. "Preserving statistical validity in adaptive data analysis." *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. ACM, 2015.

[SU15] Steinke, T., and Ullman, J. "Interactive fingerprinting codes and the hardness of preventing false discovery." *Conference on Learning Theory*. 2015.

[BNSSSU16] Bassily, R., et al. "Algorithmic stability for adaptive data analysis." *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. ACM, 2016.

[JLNRSS19] Jung, C., et al. "A New Analysis of Differential Privacy's Generalization Guarantees." *arXiv preprint arXiv:1909.03577* (2019).

Figures

1. Database icon made by Smashicons from www.flaticon.com.
2. Graph icon made by Freepik from www.flaticon.com.
3. Interaction LaTeXcode from https://tex.stackexchange.com/questions/211779/security-protocols-in-latex.

Appendix: Bayesian resampling lemma

Proof.

Expanding out the probability:

$$\Pr_{S\sim\mathcal{P},\Pi\sim I(S),S'\sim\mathcal{Q}_\Pi}\left[(S',\Pi)\in E\right]$$

$$= \sum_{\pi}\sum_{x'}\Pr[\Pi=\pi]\Pr_{S'\sim\mathcal{Q}_\pi}[S'=x']\mathbf{1}[(x',\pi)\in E]$$

$$= \sum_{\pi}\sum_{x'}\Pr[\Pi=\pi]\Pr[S'=x'|\Pi=\pi]\mathbf{1}[(x',\pi)\in E]$$

$$= \sum_{\pi}\sum_{x'}\Pr[\Pi=\pi]\frac{\Pr[\Pi=\pi|S'=x']\cdot\Pr[S=x']}{\Pr[\Pi=\pi]}\mathbf{1}[(x',\pi)\in E]$$

$$= \Pr_{S\sim\mathcal{P}^n,\Pi\sim I(S)}[(S,\Pi)\in E],$$

where we made use of Bayes' rule. □

Appendix: accuracy over samples and posterior

Proof. Denote by $j^*(\pi) = \arg\max_j |a_j - q_j(\mathcal{Q}_\pi)|$.
Consider the one-sided tail probability:

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[a_{j^*(\Pi)} - q_{j^*(\Pi)}(\mathcal{Q}_\Pi) > \alpha + c \right].$$

Expanding out the expectation $q_{j^*(\Pi)}(\mathcal{Q}_\Pi)$, we get equality with:

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\mathop{\mathbb{E}}_{S' \sim \mathcal{Q}_\Pi} \left[a_{j^*(\Pi)} - q_{j^*(\Pi)}(S') - \alpha \right] > c \right],$$

which is bounded above by:

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\mathop{\mathbb{E}}_{S' \sim \mathcal{Q}_\Pi} \left[\left(a_{j^*(\Pi)} - q_{j^*(\Pi)}(S') - \alpha \right)_+ \right] > c \right],$$

where $(x)_+ = \max\{0, x\}$.

Appendix: accuracy over samples and posterior

Proof (cont). Markov's inequality yields the upper bound:

$$\frac{1}{c} \mathop{\mathbb{E}}_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\mathop{\mathbb{E}}_{S' \sim \mathcal{Q}_\Pi} \left[\left(a_{j^*(\Pi)} - q_{j^*(\Pi)}(S') - \alpha \right)_+ \right] \right].$$

Since $(a_{j^*(\Pi)} - q_{j^*(\Pi)}(S') - \alpha)_+ \leq 1$, a further upper bound:

$$\frac{1}{c} \mathop{\mathbb{E}}_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[\Pr_{S' \sim \mathcal{Q}_\Pi} \left[a_{j^*(\Pi)} - q_{j^*(\Pi)}(S') - \alpha > 0 \right] \right],$$

which can be collapsed to:

$$\frac{1}{c} \Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S), S' \sim \mathcal{Q}_\Pi} \left[a_{j^*(\Pi)} - q_{j^*(\Pi)}(S') > \alpha \right].$$

Appendix: accuracy over samples and posterior

Proof (cont). Applying the Bayesian resampling lemma, we see that the original one-sided tail probability is upper bounded:

$$\Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[a_{j^*(\Pi)} - q_{j^*(\Pi)}(\mathcal{Q}_\Pi) > \alpha + c \right]$$

$$\leq$$

$$\frac{1}{c} \Pr_{S \sim \mathcal{P}^n, \Pi \sim I(S)} \left[a_{j^*(\Pi)} - q_{j^*(\Pi)}(S) > \alpha \right].$$

This fills in the gaps in the above proof sketch. □