CS839: Differential Privacy and Learning February 5, 2026

Linear Queries and the Factorization Mechanism

Instructor: Gavin Brown *Scribe:* Anushka Singh

Disclaimer: This document is intended as an informal supplement to in-class note-taking. It has not been given the level of scrutiny expected in polished lecture notes, let alone that reserved for peer-reviewed publications.

Overview

In this lecture we study how to answer many *counting (linear) queries* under differential privacy. The key idea is to write the true answers as a matrix–vector product Fh_x, where h_x is the dataset histogram. Instead of adding noise to every query in F directly, we choose an alternative set of queries (a matrix M), answer those privately using Gaussian noise, and then reconstruct the answers to the original workload using a matrix R. This approach is called the *factorization mechanism* (because $F = RM$). We also reviewed the *exponential mechanism*, which is used for private *selection* (picking an output from a set).

1 Linear queries as linear functions

1.1 Histogram representation

Let $\mathcal{U} = \{u_1, \dots, u_m\}$ be a finite universe and let the dataset be $x = (x_1, \dots, x_n) \in \mathcal{U}^n$. Define the *normalized histogram* $h_x \in [0,1]^m$ by

$$(h_x)_j \;=\; \frac{1}{n}\bigl|\{i \in [n] : x_i = u_j\}\bigr| \qquad (j \in [m]).$$

We think of h_x as the empirical distribution of a uniformly random record from the dataset.

Lemma 1.1 (Adjacency in histogram form)**.** *If x and x' differ in exactly one record (i.e., are adjacent), then*

$$\|h_x - h_{x'}\|_1 \;\le\; \frac{2}{n}.$$

Proof sketch. Changing one record decreases one histogram coordinate by $1/n$ and increases another by $1/n$, so the ℓ_1 change is at most $2/n$. □

1.2 From predicates to vectors

For a predicate $\phi : \mathcal{U} \to \{0,1\}$ define its *truth-table vector* $v_\phi \in \{0,1\}^m$ by

$$(v_\phi)_j \;=\; \phi(u_j) \qquad (j \in [m]).$$

The corresponding linear query is

$$f_\phi(x) \;=\; \langle v_\phi, h_x\rangle \;=\; \sum_{j=1}^{m} \phi(u_j)\,(h_x)_j \;=\; \frac{1}{n}\sum_{i=1}^{n} \phi(x_i).$$

1.3 Workloads and the error metric

Let $f_1, \dots, f_k$ be k linear queries (predicates), and let $F \in \mathbb{R}^{k \times m}$ be the *query matrix* whose i-th row is $v_{\phi_i}^\top$. Then the workload answers can be written compactly as

$$f(x) \;=\; \begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix} \;=\; Fh_x.$$

We measure accuracy using *mean squared ℓ_2 error per query*: for a (randomized) mechanism $\mathcal{A}$ outputting $\mathbb{R}^k$,

$$\mathrm{err}(\mathcal{A}; F) \;=\; \frac{1}{k}\,\mathbb{E}\big[\|Fh_x - \mathcal{A}(h_x)\|_2^2\big]\,.$$

(Expectations are over the mechanism's randomness; the dataset x is treated as fixed.)

2 The factorization mechanism

2.1 Definition and algorithm

Fix a workload matrix $F \in \mathbb{R}^{k \times m}$. A *factorization* of F is a choice of matrices $R \in \mathbb{R}^{k \times r}$ and $M \in \mathbb{R}^{r \times m}$ such that

$$F = RM.$$

We refer to M as the *measurement* matrix and R as the *reconstruction* matrix.

Remark 2.1 (Measurement vs. reconstruction; what is "fixed"?)**.** For a given workload matrix F, we are free to choose a factorization $F = RM$. After we choose (R, M), they are treated as *fixed matrices* for the mechanism and its analysis. In particular, for a fixed factorization, R is constant (it does not depend on the dataset); only the noise is random. Different choices of factorization lead to different (fixed) R and M.

Algorithm 1 Factorization mechanism for workload F

```
Input: Dataset $x \in \mathcal{U}^n$ (via histogram $h_x$), matrices $R, M$ with $F = RM$, privacy parameters $(\varepsilon, \delta)$
Returns: A vector of noisy answers $\hat{f} \in \mathbb{R}^k$
1: Compute measurement $y \leftarrow Mh_x$
2: Draw noise $Z \sim \mathcal{N}(0, \sigma_{\varepsilon,\delta}^2 I_r)$ for an appropriate $\sigma_{\varepsilon,\delta}$ (chosen for $(\varepsilon, \delta)$-DP)
3: Output $\hat{f} \leftarrow R(y + Z)$
```

Expanding the output,

$$\hat{f} \;=\; R(Mh_x + Z) \;=\; Fh_x + RZ.$$

Thus, the mechanism answers the original workload with additive Gaussian noise whose covariance is shaped by R.

2.2 Privacy via the Gaussian mechanism

Let $\sigma_{\varepsilon,\delta} > 0$ be the Gaussian noise standard deviation chosen to ensure (ε, δ)-DP (see below).

Draw noise $Z \sim \mathcal{N}(0, \sigma_{\varepsilon,\delta}^2 I_r)$

To choose σ, we analyze the sensitivity of the *measurement map* $g(x) = Mh_x$ in ℓ_2. Define the ℓ_2 sensitivity

$$\Delta_2(g) \;=\; \max_{x \sim x'} \|Mh_x - Mh_{x'}\|_2.$$

A convenient quantity here is the norm $\|M\|_{1\to 2}$:

$$\|M\|_{1\to 2} \;=\; \max_{\substack{v \\ \|v\|_1 \le 1}} \|Mv\|_2.$$

Using the adjacency lemma, for adjacent $x \sim x'$,

$$\|Mh_x - Mh_{x'}\|_2 \;\le\; \|M\|_{1\to 2}\,\|h_x - h_{x'}\|_1 \;\le\; \frac{2}{n}\|M\|_{1\to 2},$$

so

$$\Delta_2(g) \;\le\; \frac{2}{n}\|M\|_{1\to 2}.$$

Claim 2.2 (A useful identity)**.** *For $M \in \mathbb{R}^{r\times m}$,*

$$\|M\|_{1\to 2} \;=\; \max_{j\in[m]} \|M_{*,j}\|_2,$$

i.e., the maximum ℓ_2 norm of a column of M.

Proof sketch. The maximum over $\|v\|_1 \le 1$ is achieved at an extreme point of the ℓ_1 ball, i.e., $v = \pm e_j$, so $\|Mv\|_2 = \|M_{*,j}\|_2$. □

Here σ denotes the *standard deviation* of the isotropic Gaussian noise added to each of the r measured queries.

Gaussian noise calibration. Using the standard Gaussian mechanism, choosing

$$\sigma \;\ge\; \frac{\Delta_2(g)\,\sqrt{2\ln(1.25/\delta)}}{\varepsilon}$$

is sufficient to make $Mh_x + Z$ (ε,δ)-DP. By post-processing, multiplying by R preserves privacy, hence the output $\hat{f} = R(Mh_x + Z)$ is also (ε,δ)-DP.

3 Error analysis

Recall $\hat{f} = Fh_x + RZ$, so the (per-query) mean squared error is

$$\mathrm{err}(\mathcal{A}; F) \;=\; \frac{1}{k}\mathbb{E}\left[\|RZ\|_2^2\right].$$

If $Z \sim \mathcal{N}(0, \sigma^2 I_r)$, then $RZ \sim \mathcal{N}(0, \sigma^2 RR^\top)$. For a mean-zero Gaussian with covariance Σ, we have $\mathbb{E}[\|Y\|_2^2] = \mathrm{tr}(\Sigma)$, hence

$$\mathbb{E}[\|RZ\|_2^2] \;=\; \mathrm{tr}(\sigma^2 RR^\top) \;=\; \sigma^2\|R\|_F^2.$$

Here we have use the *trace* and *Frobenius norm*

$$\mathrm{tr}(A) := \sum_i A_{ii} \quad \text{and} \quad \|A\|_F := \left(\sum_i \sum_j A_{ij}^2\right)^{1/2}.$$

The relationships between the norms of Gaussians, the trace, and the Frobenius norms are "standard" facts. If you don't have them memorized you can prove them all by writing out the vectors in terms of their components.

Therefore

$$\mathrm{err}(\mathcal{A}; F) \;=\; \frac{\sigma^2}{k}\|R\|_F^2.$$

Substituting $\sigma^2 = \frac{2\ln(1.25/\delta)}{\varepsilon^2}\Delta_2(g)^2$ and the bound $\Delta_2(g) \le \frac{2}{n}\|M\|_{1\to 2}$ yields the convenient upper bound

$$\mathrm{err}(\mathcal{A}; F) \;\le\; \frac{8\ln(1.25/\delta)}{k\,n^2\,\varepsilon^2}\;\|M\|_{1\to 2}^2\;\|R\|_F^2.$$

3.1 The (informal) factorization norm

The privacy/accuracy tradeoff above suggests optimizing over factorizations $F = RM$. The quantity that emerges is the *factorization norm*

$$\gamma_2(F) \;=\; \min_{F=RM}\;\|R\|_F\,\|M\|_{1\to 2}.$$

Up to constants and the $(\varepsilon, \delta, n, k)$ parameters, the factorization mechanism's error is controlled by $\gamma_2(F)$. In general, finding the optimal factorization can be reduced to convex optimization (e.g., an SDP); for structured workloads, one can sometimes do better.

4 Exponential mechanism (core DP primitive)

In the last few minutes of class, we will introduce the *exponential mechanism*. It represents one of the four fundamental mechanisms we will see in this class (along with Laplace noise, Gaussian noise, and randomized response).

4.1 Definition

The exponential mechanism is used to make a *choice* from a set of candidates $\mathcal{Y}$.

Let $u : \mathcal{U}^n \times \mathcal{Y} \to \mathbb{R}$ be a *utility* function and define its sensitivity

$$\Delta u \;=\; \max_{x\sim x'}\max_{y\in\mathcal{Y}} |u(x,y) - u(x',y)|.$$

Given privacy parameter $\varepsilon > 0$, the exponential mechanism outputs $y \in \mathcal{Y}$ with probability

$$\Pr[\mathcal{M}(x) = y] \;\propto\; \exp\Big(\frac{\varepsilon}{2\Delta u}\,u(x,y)\Big)\,.$$

4.2 Privacy guarantee

Theorem 4.1. *If u has sensitivity Δu, then the exponential mechanism is ε-differentially private.*

Proof sketch. Fix adjacent $x \sim x'$ and any $y \in \mathcal{Y}$. Let $Z(x) = \sum_{y'} \exp(\frac{\varepsilon}{2\Delta u}u(x,y'))$. Then

$$\frac{\Pr[\mathcal{M}(x) = y]}{\Pr[\mathcal{M}(x') = y]} \;=\; \exp\Big(\frac{\varepsilon}{2\Delta u}(u(x,y) - u(x',y))\Big)\cdot\frac{Z(x')}{Z(x)}.$$

Because u has sensitivity Δu, for every $y' \in \mathcal{Y}$ we have $u(x', y') \le u(x, y') + \Delta u$. Therefore,

$$\exp\Big(\frac{\varepsilon}{2\Delta u} u(x', y')\Big) \le \exp\Big(\frac{\varepsilon}{2\Delta u}(u(x, y') + \Delta u)\Big) = e^{\varepsilon/2} \exp\Big(\frac{\varepsilon}{2\Delta u} u(x, y')\Big) .$$

Summing over y' gives $Z(x') \le e^{\varepsilon/2} Z(x)$, i.e., $\frac{Z(x')}{Z(x)} \le e^{\varepsilon/2}$. Similarly, $u(x, y) - u(x', y) \le \Delta u$ implies the numerator ratio is also $\le e^{\varepsilon/2}$, so the total ratio is $\le e^{\varepsilon}$.

The first factor is at most $e^{\varepsilon/2}$ by sensitivity. For the second, each term in $Z(x')$ differs from the corresponding term in $Z(x)$ by at most a factor $e^{\varepsilon/2}$, so $Z(x')/Z(x) \le e^{\varepsilon/2}$. Multiplying gives the ε bound. □

4.3 Example: private winner selection

Suppose $y \in \mathcal{Y}$ are candidates in an election and the dataset consists of votes. Let $u(x, y)$ be the number of votes for candidate y. Changing one vote changes $u(x, y)$ by at most 1, so $\Delta u = 1$. The exponential mechanism then samples candidates with probability proportional to $\exp(\varepsilon u(x, y)/2)$, biasing toward the winner while preserving ε-DP.

Exercise 4.2 (Laplace as an exponential mechanism)**.** Let $f : \mathcal{U}^n \to \mathbb{R}$ be a real-valued query with ℓ_1 sensitivity $\Delta_1 = \max_{x \sim x'} |f(x) - f(x')|$. Consider the exponential mechanism over output space $\mathcal{Y} = \mathbb{R}$ with utility

$$u(x, y) := -|y - f(x)|.$$

1. Show that $\Delta u \le \Delta_1$.
2. Show that the resulting output distribution has density proportional to

   $$\exp\Big(-\frac{\varepsilon}{2\Delta_1} |y - f(x)|\Big) ,$$

   which is exactly a Laplace distribution centered at $f(x)$ (up to the parameterization convention).

Conclude that the Laplace mechanism can be viewed as a special case of the exponential mechanism.