Algorithms for Private Data Analysis Spring 2021
Lecture 8: Differential Privacy and Mechanism Design

Steven Wu

1 Differential Privacy and Mechanism Design

Differential privacy turns out to be a useful for mechanism design even when privacy is not the primary desideratum. For context, recall the following example of pricing digital goods.

Example 1.1 (Prices of a digital good)**.** Suppose you made an iPhone app. Now you want to sell it online. Suppose there are n buyers such that each buyer has their private valuation $x_i \in [0, 1]$—that is the maximum price they are willing to pay for a download of the app. Assuming that respondents answered truthfully, a reasonable estimate for the revenue you would get from selling the download at price p is

$$q(p; \mathbf{x}) = p \cdot \#\{i : x_i \geq p\} \ .$$

The goal is to find a price $\hat{p}$ such that $q(\hat{p}; \mathbf{x})$ is as large as possible.

A major assumption we make above is that each buyer reports their valuation truthfully. However, if they are strategic agents, they may have incentives to mis-report their private valuations. For example, suppose there are four buyers with private valuations: \$1, \$1, \$2.01, \$4. If all buyers report truthfully, the revenue maximizing price will be \$4.02. However, the buyer with the highest valuation has a strong incentive to lie about his valuation. If he happens to know the valuations of the other three buyers, he will likely report \$2.01 dollar as well. We will revisit this example. This is a typical problem studied in the field of *mechanism design*, which deals with problems that take inputs from strategic agents.

Differential privacy provides an appealing property for the price selection mechanism. If the mechanism is differentially private, then each agent has very little influence on the (random) price selection, and as a result they may have very little incentive to mis-report their private input. In particular, we can give a utility-theoretic definition of differential privacy which is equivalent to the standard definition we have been working with so far. Below we will write (x_{-i}, x_i') to denote the vector $(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)$, which is a common notation in game theory.

Proposition 1.2 (Utility-theoretic view of DP)**.** *An algorithm $A \colon \mathcal{X}^n \to \mathcal{Y}$ is ε-differentially private if and only if for every (utility) function $f : \mathcal{Y} \to \mathbb{R}_+$, and for every pair of neighboring data sets $x \in \mathcal{X}^n$ and $x_i' \in \mathcal{X}$:*

$$\mathop{\mathbb{E}}_{Y \sim A(x)}(f(Y)) \leq \exp(\varepsilon) \mathop{\mathbb{E}}_{Y \sim A(x_{-i}, x_i')}(f(Y))$$

The function f in the above definition can be interpreted as an agent's utility function for the outcomes selected by the mechanism A. In particular, when the algorithm is ε-differentially private, the agent cannot improve their utility by more than a multiplicative factor of $\exp(\varepsilon)$, no matter what their utility function may be. In the digital good example, this corresponds to the promise that even if any single agent has very little incentive to misreport their value for the app.

Proof. First, suppose that the algorithm A is ε-differentially private. Then

$$\begin{aligned}\mathop{\mathbb{E}}_{Y\sim A(x)}(f(Y)) &= \int_y f(y)\,\mathbb{P}\left(A(x)=y\right)dy \\ &\leq \int_y f(y)\,\exp(\varepsilon)\,\mathbb{P}\left(A(x')=y\right)dy \\ &= \exp(\varepsilon)\mathop{\mathbb{E}}_{Y\sim A(x_{-i},x_i')}(f(Y))\end{aligned}$$

In the reverse direction, consider any event $E \subseteq \mathcal{Y}$, and define f to be $f(y) = \mathbf{1}[y \in E]$. Then

$$\mathop{\mathbb{E}}_{Y\sim A(x)}(f(Y)) \leq \exp(\varepsilon)\mathop{\mathbb{E}}_{Y\sim A(x_{-i},x_i')}(f(Y))$$

implies

$$\mathop{\mathbb{P}}_{Y\sim A(x)}(Y \in E) \leq \exp(\varepsilon)\mathop{\mathbb{P}}_{Y\sim A(x_{-i},x_i')}(Y \in E)$$

which recovers the differential privacy definition. □

1.1 Mechanism Design Basics

We will consider n rational (utility maximizing) agents indexed by i who have privately known *types* $t_i \in \mathcal{T}$ (e.g., their private valuation on a good). A mechanism $M\colon \mathcal{T}^n \to \mathcal{Y}$ is a mapping between (reported) types of the n agents and some outcome space $\mathcal{Y}$. Each agent has some preference over outcomes, which are determined by their types: the utility that agent i gets why the outcome y is selected is defined to be:

$$u_i(y) \equiv u(t_i, y)$$

where $u\colon \mathcal{T}\times\mathcal{Y} \to [0,1]$ is some utility function.

In general, a mechanism designer wants to choose a desirable outcome (according to some objective function like revenue) and also wants to incentivize agents to report their true types.

Definition 1.3 (Dominant strategy truthfulness)**.** A mechanism $M\colon \mathcal{T}^n \to \mathcal{Y}$ is ε-approximate *dominant strategy truthful* if for all $t \in \mathcal{T}^n$ and for all i and $t_i' \in \mathcal{T}$:

$$u_i(M(t)) \geq u_i(M(t_{-i}, t_i')) - \varepsilon.$$

That is, no player can gain more than ε utility by mis-reporting their type. If M is randomized, then

$$u_i(M(t)) = \mathop{\mathbb{E}}_{Y\sim M(t)}(u_i(Y))$$

Differential privacy implies approximate truthfulness, due to the following claim from the result of [MT07].

Claim 1.4. *If M is ε-differentially private for $\varepsilon \leq 1$, then M is also ε-approximate dominant strategy truthful.*

Proof. First, observe that

$$\mathop{\mathbb{E}}_{Y\sim M(t)}(u_i(Y)) \geq \exp(-\varepsilon)\mathop{\mathbb{E}}_{Y\sim M(t_{-i},t_i')}(u_i(Y)) \geq (1-\varepsilon)\mathop{\mathbb{E}}_{Y\sim M(t_{-i},t_i')}(u_i(Y)) \geq \mathop{\mathbb{E}}_{Y\sim M(t_{-i},t_i')}(u_i(Y)) - \varepsilon$$

where the first inequality follows from the definition of differential privacy, the second follows from the fact that $\exp(-\varepsilon) \geq 1-\varepsilon$, and the last one follows from that the utility is in the range of $[0,1]$. □

It is worth noting a couple of things about this connection between privacy and approximate truthfulness.

- A reasonable objection is that this straightforward application of privacy makes not just truthful reporting an approximate dominant —it makes everything an approximate dominant strategy. Why should people tell the truth in such cases? As we will see, however, a typical application of this approach will give a more nuanced guarantee, in which truthful reporting remains an approximate dominant strategy, but not everything does
- The differential privacy definition composes across player deviations: if 4 players change their reports, then the probability of any event can change by at most $\exp(4\varepsilon)$. Hence, differential privacy automatically promises approximate group strategyproofness as well. No coalition of k players can improve their expected utility by more than a factor of $\exp(k\varepsilon)$ by deviating from truthtelling behavior.

Regarding the first point above, we make the following observation. Suppose the utility-relevant event for each player i in fact comes from some other outcome space $\mathcal{Y}_i'$ and players have utility functions $u_i : \mathcal{Y}_i' \to [0,1]$. Let us further suppose that there is some function $f : \mathcal{Y} \times \mathcal{T} \to \mathcal{Y}_i'$ mapping both outcome y selected by a mechanism M and agent i's reported type t_i' to an outcome $f(y, t_i') \in \mathcal{Y}_i'$. If we have that for every fixed y, $f(y, \cdot)$ makes truthful reporting a dominant strategy, then truthful reporting remains an ε-approximate dominant strategy to M, but non-truthful reports may no longer be approximate dominant strategies.

Claim 1.5. *If $M : \mathcal{T}^n \to \mathcal{Y}$ is ε-differentially private and $f(y, \cdot) : \mathcal{T} \to \mathcal{Y}_i'$ is dominant strategy truthful for every outcome $y \in \mathcal{Y}$, then truthful reporting is an ε-approximate dominant strategy for the mechanism $M' : \mathcal{T}^n \to \mathcal{Y}'$ that produces outcome $f(M(t), t_i)$ for agent i.*

Proof.

$$\begin{aligned}
\mathop{\mathbb{E}}_{Y \sim M(t)}(u_i(f(y, t_i))) &\geq \exp(-\varepsilon) \mathop{\mathbb{E}}_{Y \sim M(t_{-i}, t_i')}(u_i(f(Y, t_i))) \\
&\geq \exp(-\varepsilon) \mathop{\mathbb{E}}_{Y \sim M(t_{-i}, t_i')}\left(u_i(f(Y, t_i'))\right) \\
&\geq (1 - \varepsilon) \mathop{\mathbb{E}}_{Y \sim M(t_{-i}, t_i')}\left(u_i(f(Y, t_i'))\right) \\
&\geq \mathop{\mathbb{E}}_{Y \sim M(t_{-i}, t_i')}\left(u_i(f(Y, t_i'))\right) - \varepsilon
\end{aligned}$$

□

1.2 Revisiting Digital Goods

Now we have the formal language to think more formally about the problem of pricing digital goods. Each buyer has a private type, which is their private valuation $t_i \in [0,1]$ (after some re-scaling). The outcome space $\mathcal{Y} = 2^{[n]} \times [0,1]^n$ is the set of all subsets of bidders who might be selected to win an item, together with the set of all prices they might be charged: an outcome is a pair (S, p), where bidders $i \in S$ receive an item and pay p_i. Formally, the utility of a buyer i is defined as

$$u(S, p) = \mathbf{1}[i \in S]\,(t_i - p_i)$$

where x_i denotes the true value of the buyer. The mechanism proceeds as follows:

1. We can then use the exponential mechanism (or report noisy max) to select a price $\hat{p} \in [0, 1]$ from the exponential mechanism with revenue $q(p; t)$ as the quality score.

2. We will then sell to each bidder i at this price $\hat{p}$ if their reported value $x_i \geq p$. That is, for each buyer i: $i \in S$ and $p_i = \hat{p}$ if and only if $f(\hat{p}, x_i) \equiv \mathbf{1}[x_i \geq \hat{p}] = 1$.

3. Since for every fixed price $\hat{p}$, reporting one's true value is a dominant strategy, and $\hat{p}$ is chosen in an ε-differentially private way, the whole mechanism will be ε-approximately dominant strategy truthful.

Acknowledgement This lecture not is built on notes developed by Aaron Roth.

References

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *IEEE Symposium on Foundations of Computer Science*, FOCS '07, 2007.