Selling Privacy at Auction

隐私拍卖

Authors: Arpita Ghosh, Aaron Roth

Speaker： xxx

date： 2021-10-18

Background

Auctions

- Generalized First Price(GFP)
  - a non-truthful auction mechanism
  - the highest bidder pays the price bid by the highest bidder
- Generalized Second Price(GSP)
  - a non-truthful auction mechanism for multiple items
  - the highest bidder pays the price bid by the second-highest bidder
- Vickrey-Clarke-Groves(VCG)
  - Bidders submit bids that report their valuations for the items, without knowing the bids of the other bidders.
  - It gives bidders an incentive to bid their true valuations, by ensuring that the optimal strategy for each bidder is to bid their true valuations of the items.

Related Work

Differential Privacy as a Tool in Mechanism Design

- McSherry and Talwar proposed that differential privacy could itself be used as a solution concept in mechanism design

Auctions Which Preserve Privacy

- Feigenbaum, Jaggard, and Schapira study to what extent information must be leaked in second price auctions and in the millionaires problem.

Privacy in the Economics Literature

- primarily in the context of how preferences for privacy by agents may affect mechanisms,rather than in the context of markets for privacy

Relationship to the Privacy Literature

- Most Literature([Dwo08])almost exclusively focused on techniques for guaranteeing ε-differential privacy for various tasks,where has been taken as a given parameter.

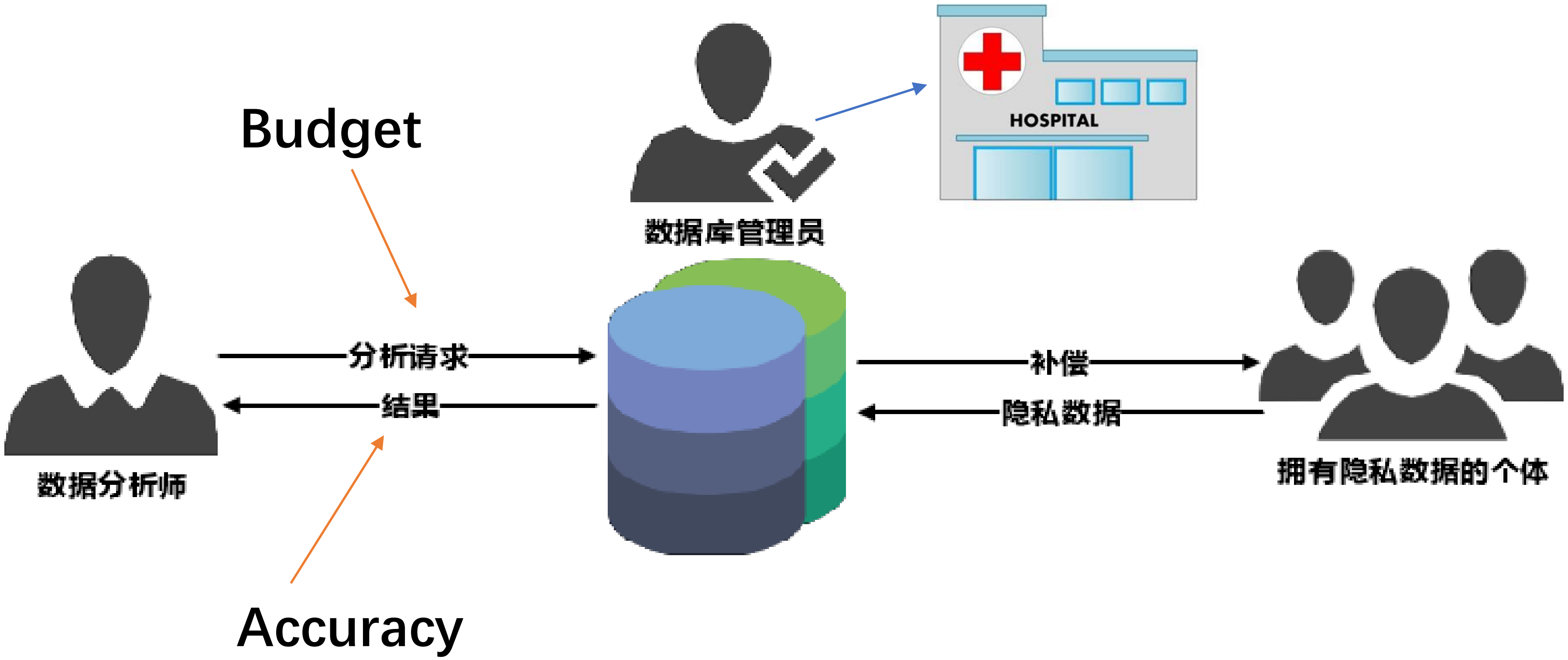
Contribution

1. This paper consider a setting in which a data analyst wishes to buy information from a population from which he can estimate some statistic.

2. any differentially private mechanism that guarantees a certain accuracy must purchase a certain minimum amount of privacy from a certain minimum number of agents

3. The main contribution of this paper is to formalize the notion of auctions for private data, and to show that the design space of such auctions can be taken to be the simple setting of multi-unit procurement auctions.

System Model



Two objectives for mechanism

- When the data analyst has a **fixed accuracy goal**, we show that an application of the classic Vickrey auction achieves the analyst's accuracy goal while **minimizing his total payment.**

- When the data analyst has **a fixed budget**, we give a mechanism which maximizes the accuracy of the resulting estimate while guaranteeing that the resulting sum payments **do not exceed the analyst's budget.**

Mechanism Design

Notation

- There are n individuals [n],Each individual i is associated with a private bit $b_i \in \{0,1\}$
- Each individual also has a certain cost function c_i: $\{\mathbb{R}_+ \to \mathbb{R}_+\}$
  - linear cost functions:$c_i = v_i \varepsilon$ for some unknown $v_i \in \mathbb{R}$
  - utility: $u_i = p_i(v) - v_i \epsilon_i(v)$
- $c_i(\varepsilon)$ determines what her cost is for her private bit $b_i$used in an ε-differentially private manner.
- Restrict our attention to values of $\varepsilon < 1$, So $\exp(\epsilon) \approx 1 + \epsilon$.
- The data analyst wishes to estimate the quantity $s = \sum_{i=1}^{n} b_i$
- The collection of all individuals' private bits is a database $D \in \{0,1\}^n$

Differential Privacy Algorithm

DEFINITION 2.1. *An algorithm $A : \{0,1\}^n \to \mathbb{R}$ satisfies ϵ_i-differential privacy with respect to individual i if for any pair of neighboring databases $D, D^{(i)} \in \{0,1\}^n$ differing only in their i'th bit, and for any $S \subset \mathbb{R}$:*

$$\frac{\Pr[A(D) \in S]}{\Pr[A(D^{(i)}) \in S]} \leq e^{\epsilon_i}$$

FACT 1. *Consider an algorithm $A : \{0,1\}^n \to \mathbb{R}$ that satisfies ϵ_i-differential privacy with respect to each individual i, and let $T \subset [n]$ denote a set of indices. Consider two databases $D, D^T \in \{0,1\}^n$ at Hamming distance $|T|$ that differ exactly on the indices in T. Then:*

$$\frac{\Pr[A(D) \in S]}{\Pr[A(D^T) \in S]} \leq e^{\sum_{i \in T} \epsilon_i}$$

Characterize the mechanism

A mechanism M : $\mathbb{R}_+^n \times \{0,1\}^n \to \mathbb{R} \times \mathbb{R}_+^n$

- Input: a vector of cost functions $v = (v_1, \ldots, v_n) \in \mathbb{R}_+^n$ and database $D \in \{0,1\}^n$
- Output: $\hat{s} = A(D) \in \mathbb{R}$ and a vector of payments $p(v) \in \mathbb{R}_+^n$

Design Objective

- **Individually rational(Non-negative Utility)**

DEFINITION 2.4. *A mechanism $M : \mathbb{R}_+^n \times \{0,1\}^n \to \mathbb{R} \times \mathbb{R}_+^n$ is* individually rational *if for all $v \in \mathbb{R}_+^n$:*

$$\boxed{p_i(v) \geq v_i \epsilon_i(v)}$$

$$u_i = p_i(v) - v_i \epsilon_i(v) \geq 0$$

- **Truthfulness**

DEFINITION 2.5. *A mechanism $M : \mathbb{R}_+^n \times \{0,1\}^n \to \mathbb{R} \times \mathbb{R}_+^n$ is* dominant-strategy truthful *if for all $v \in \mathbb{R}_+^n$, for all $i \in [n]$, and for all $v_i' \in \mathbb{R}_+$:*

$$\boxed{p_i(v) - v_i \epsilon_i(v) \geq p_i(v_{-i}, v_i') - v_i \epsilon_i(v_{-i}, v_i'),}$$

that is, no player can ever increase his utility by misreporting his value for privacy.

True evaluation

Mis-report evaluation

- **Accuracy**

DEFINITION 2.6. *A mechanism M satisfies k-accuracy if for any $D \in \{0,1\}^n$, it outputs an estimate $\hat{s} = A(D)$ such that:*

$$\boxed{\Pr[|\hat{s} - s| \geq k] \leq \frac{1}{3}}$$

Charactering accurate mechanisms

- **we show necessary and sufficient conditions on the amount of privacy that a mechanism must purchase from each player in order to guarantee a fixed level of accuracy**

- a mechanism must purchase at least ε-privacy, from at least $|H|$ people, where the values of ε and $|H|$ depend on the desired accuracy.

- Necessary conditions

THEOREM 3.1. *Let* $0 < \alpha < 1$. *Any differentially private mechanism that is* $\alpha \cdot n/4$*-accurate must select a set of users* $H \subseteq [n]$ *such that:*

1. $\epsilon_i \geq \frac{1}{\alpha n}$ *for all* $i \in H$.
2. $|H| \geq (1-\alpha)n$.

- Sufficient conditions

THEOREM 3.3. *Let* $0 < \alpha < 1$. *There exists a differentially private mechanism that is* $(\frac{1}{2} + \ln 3)\alpha \cdot n$*-accurate and selects a set of individuals* $H \subseteq [n]$ *such that:*

1. $\epsilon_i = \begin{cases} \frac{1}{\alpha n}, & \text{for } i \in H; \\ 0, & \text{for } i \notin H. \end{cases}$
2. $|H| = (1-\alpha)n$.

$\Downarrow$

a multi-unit procurement auctions:

where we seek to purchase exactly $1/\alpha\, n$ units of some good from exactly $(1-\alpha)n$ individuals.

Deriving Truthful Mechanisms

Maximizing Accuracy Subject to a Budget Constraint

- **Problem**

  obtaining an estimate $\widehat{S}$ of maximum accuracy,subject to a hard budget constraint:$\sum_{i=1}^{n} p_i \leq B$

We give a truthful and individually rational mechanism for this problem, and show that it is instance-by-instance optimal among the class of envy-free mechanisms

- FairQuery algorithm

FairQuery(v, D, B) :

Sort v such that $v_1 \leq v_2 \leq \ldots \leq v_n$.

Let k be the largest integer such that $\frac{v_k}{n-k} \leq \frac{B}{k}$.

Output $\hat{s} = \sum_{i=1}^{k} b_i + \frac{n-k}{2} + \text{Lap}(n-k)$

Pay each $i > k$ $p_i = 0$ and each $i \leq k$ $p_i = \min(\frac{B}{k}, \frac{v_{k+1}}{n-k})$.

FairQuery is individually rational

FairQuery(v, D, B) :

Sort v such that $v_1 \leq v_2 \leq \ldots \leq v_n$.
Let k be the largest integer such that $\frac{v_k}{n-k} \leq \frac{B}{k}$.
Output $\hat{s} = \sum_{i=1}^{k} b_i + \frac{n-k}{2} + \mathrm{Lap}(n-k)$
Pay each $i > k$ $p_i = 0$ and each $i \leq k$ $p_i = \min(\frac{B}{k}, \frac{v_{k+1}}{n-k})$.

THEOREM 3.3. *Let $0 < \alpha < 1$. There exists a differentially private mechanism that is $(\frac{1}{2} + \ln 3)\alpha \cdot n$-accurate and selects a set of individuals $H \subseteq [n]$ such that:*

1. $\epsilon_i = \begin{cases} \frac{1}{\alpha n}, & \text{for } i \in H; \\ 0, & \text{for } i \notin H. \end{cases}$

☐ Proof of individually rational:

Proof: 根据定理3.3有

$$\varepsilon_i = \begin{cases} \frac{1}{n-k}, & i \leq k \\ 0, & i > k \end{cases}$$

因此，当 $i > k$ 时 $p_i = 0 = \varepsilon_i \cdot V_i = 0 \cdot V_i$ ---- ①

当 $i \leq k$ 时，$p_i = \underline{\min\left(\frac{B}{k}, \frac{V_{k+1}}{n-k}\right)}$

讨论一下 p_i：

当 $\frac{V_{k+1}}{n-k} \geq \frac{B}{k}$ 时，$p_i = \frac{B}{k} \geq \frac{V_i}{n-k}$ $(i \leq k)$

当 $\frac{V_{k+1}}{n-k} < \frac{B}{k}$ 时，$p_i = \frac{V_{k+1}}{n-k} \geq \frac{V_i}{n-k}$ $(i \leq k)$ $V_i \uparrow$

$\therefore p_i = \min\left(\frac{B}{k}, \frac{V_{k+1}}{n-k}\right) \geq \frac{V_i}{n-k}$ $(i \leq k)$

当 $i \leq k$ 时：

$\therefore p_i - \varepsilon_i V_i = p_i - \frac{V_i}{n-k} \geq 0$ ---- ②

因此：$p_i - \varepsilon_i V_i \geq 0$. 证毕.

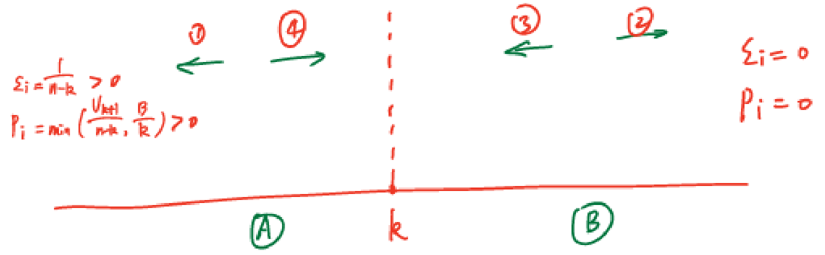
FairQuery is Truthful

☐ Proof of Truthfulness:

Fix any v, i, v_i' and consider $k = k(v)$, $k' = k(v_{-i}, v_i')$, $p_i = p_i(v)$, $p_i' = p_i'(v_{-i}, v_i')$, $\epsilon_i = \epsilon_i(v)$, and $\epsilon_i' = \epsilon_i'(v_{-i}, v_i')$. There are four cases:

1. Case 1: $v_i' < v_i$ and $p_i > 0$. In this case, v_i' moves earlier in the ordering and $\epsilon_i = \epsilon_i'$, and $p_i = p_i'$.
2. Case 2: $v_i' > v_i$ and $p_i = 0$. In this case, v_i' moves later in the ordering and $\epsilon_i = \epsilon_i' = p_i = p_i' = 0$.
3. Case 3: $v_i' < v_i$ and $p_i = 0$. In this case, v_i' moves earlier in the ordering, but if $p_i' > 0$ then by construction $p_i' = \min(\frac{B}{k'}, \frac{v_{k'+1}}{n-k'}) \leq v_i/(n-k')$. This follows because k' is such that $v_{k'+1} \leq v_i$ for all $i > k$ such that $p_i' > 0$.
4. Case 4: $v_i' > v_i$ and $p_i > 0$. In this case, v_i' moves later in the ordering, and either $p_i' = p_i$ and $\epsilon_i' = \epsilon_i$, or $p_i' = 0$ and $\epsilon_i = 0$. In the second case, by individual rationality, $p_i - v_i\epsilon_i \geq 0 = p_i' - v_i\epsilon_i'$.

Proof:



① $V_i' < V_i$ and $p_i > 0$ $i \in$ Ⓐ

排序移前. $\epsilon_i' = \epsilon_i$ and $p_i = p_i'$

② $V_i' > V_i$ and $p_i = 0$. $i \in$ Ⓑ

排序移后. $\epsilon_i' = \epsilon_i = 0 = p_i = p_i'$

③ $V_i' < V_i$ and $p_i = 0$ $i \in$ Ⓑ

排序移前. if $p_i' > 0$. $p_i' = \min(\frac{B}{k'}, \frac{V_{k'+1}}{n-k'})$

$p_i' = \min(\frac{B}{k'}, \frac{V_{k'+1}}{n-k'})$ 其中 $V_{k'+1} \le V_i$ $(i > k)$

其中 $\frac{V_{k'+1}}{n-k'} \le \frac{V_i}{n-k'} \le V_i$ $U_i = p_i' - V_i \le 0$

④ $V_i' > V_i$ and $p_i > 0$. $i \in$ Ⓐ

排序移后. ① $p_i' = p_i$ and $\epsilon_i' = \epsilon_i$ ⇒ 不变.

② $p_i' = 0$ and $\epsilon_i = 0$ ⇒ $U_i = 0$ (不变)

FairQuery is optimal envy-free mechanism

Envy-freeness:

OBSERVATION 4.3. *Any truthful envy-free mechanism which buys either no privacy or ϵ-privacy from each individual (i.e., if $\epsilon_i > 0, \epsilon_j > 0$ then $\epsilon_i = \epsilon_j$) must have the property that for all i, j with $\epsilon_i > \epsilon_j > 0$, $p_i = p_j$. Call such mechanisms* fixed purchase mechanisms. *That is, envy free fixed purchase mechanisms must pay each individual from whom privacy is purchased the same fixed price.*

Proof:

PROOF. First, observe the easy fact that FairQuery is indeed an envy free fixed purchase mechanism. We then merely observe that for any vector of valuations v, if FairQuery sets $\epsilon_i > 0$ for k individuals, then by the definition of k, it must be that $\frac{v_{k+1}}{(n-k-1)} > \frac{B}{k+1}$, and so any mechanism that set $\epsilon_i > 0$ for k' individuals for $k' > k$ must have $p_{k+1} > \frac{B}{(k+1)}$ by individual rationality. But by envy-freeness, it must have $\boxed{p_i = p_{k+1}} > \frac{B}{(k+1)}$ for all $i \leq k$. But in this case, we would have

$$\sum_{i=1}^{n} p_i \geq k' \cdot p_{k+1} > (k+1) \cdot \frac{B}{k+1} = B$$

which would violate the budget constraint. $\square$

• Minimizing Payment Subject to an Accuracy Constraint

By Theorem 3.3, Buy $\frac{1/2 + \ln 3}{an}$ units of privacy from $\left(1 - \frac{a}{\frac{1}{2} + \ln 3}\right) n$ people

The constraint on accuracy simply states that we must buy $\left(1 - \frac{a}{\frac{1}{2} + \ln 3}\right) n$ **units of the good.**

- MinCostAuction algorithm

MinCostAuction(v, D, α):

Let $\alpha' = \frac{\alpha}{1/2 + \ln 3}$ and $k = \lceil (1 - \alpha')n \rceil$.

Sort $v = c_i(\frac{1}{n-k})$ such that $v_1 \le v_2 \le \ldots \le v_n$.

Output $\hat{s} = \sum_{i=1}^{k} b_i + \frac{n-k}{2} + \mathrm{Lap}(\alpha' n)$

Pay each $\boxed{i > k\ p_i = 0 \text{ and each } i \le k\ p_i = v_{k+1}.}$

- Payments

$$\sum_{i=1}^{n} p_i = k \cdot v_{k+1}$$

Proof：

no other envy-free multi-unit procurement auction with the same accuracy guarantees (i.e. one that guarantees buying k units) makes smaller payments than MinCostAuction.

PROOF. For the sake of contradiction, suppose we have such a mechanism M. Fix some vector of valuations v that yields payments $p(v)$ such that $\sum_{i=1}^{n} p_i(v) < k \cdot v_{k+1}$ (again, note that v_i now denotes the total cost for purchasing data, not the per-unit privacy cost). First, if it is not already the case, we will construct a bid profile such that an item is purchased from some seller who is not among the k lowest sellers. It must be that there exists some i such that an item is purchased from i at a price of p_i, such that $v_i \leq p_i < v_{k+1}$ (otherwise $\sum_{i=1}^{n} p_i(v) \geq k \cdot v_{k+1}$). Let $v' = (v_{-i}, (p_i + v_{k+1})/2)$ be a bid profile in which bidder i raises his bid to be above p_i while remaining below v_{k+1}. Let $p' = p'(v)$ be the new payment vector. By individual rationality and truthfulness, it must be that in this new bid profile v', player i is no longer allocated an item: by individual rationality, he would have to be paid $p'_i > p_i$ if he were allocated an item, but if his true valuation were v_i, then this would be a beneficial deviation, contradicting truthfulness. Because the mechanism is constrained to always buy at least k items, it must be that in v', an item is now purchased from some seller j such that $j \geq k+1$. By individual rationality, $p'_j \geq v_j \geq v_{k+1}$. But by envy-freeness, it must be that for every seller i from whom an item was purchased, $p'_i = p'_j \geq v_{k+1}$. Because at least k items are purchased, we therefore have $\sum_{i=1}^{n} p'_i \geq k \cdot v_{k+1}$, which contradicts the purported payment guarantee of mechanism M. $\square$

Preserving the privacy of the bid

Can we design mechanisms that treat individuals valuations for privacy as private data as well, and compensate individuals for the privacy loss due to the use of their valuations v_i?

NO!

THEOREM 5.1. *If bidder valuations for privacy may be arbitrarily large (i.e., $v \in \mathbb{R}_+^n$) then no individually rational mechanism M can protect the privacy of the bidder valuations and promise k-accuracy for any $k < n/2$ (i.e., any nontrivial value).*

Solution!

restrict bidder valuations v_i to lie in a bounded range.

But re-introduces the very source of sampling bias that we wanted to solve by running an auction!

Future Directions

1. Studying Bayesian optimal mechanism design for these auctions would help identify and justify appropriate benchmarks.
2. It is unsatisfying to restrict individual valuations for privacy to lie in a bounded range. This requires the development of new models.
3. Is there any way to mediate the purchase of private data directly from individuals who have the power to lie about their private data?
4. How about a two sided market, in which there are multiple data analysts, competing for access to the private data from multiple populations.
5. In this paper we considered a one-shot mechanism. In reality, the administrator of a private database will face multiple requests for access to his data as time goes on.

Thanks for your attention!