Game theory

Katrina Ligett

differential privacy

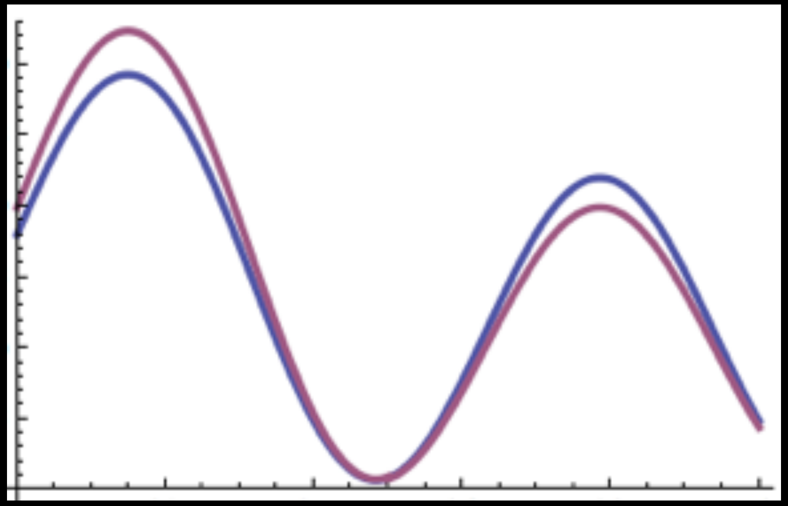
[DinurNissim03, DworkNissimMcSherrySmith06, Dwork06]

ε-Differential Privacy for algorithm M:

for any two neighboring data sets x_1, x_2, differing by the addition or removal of a single row

any $S \subseteq \text{range}(M)$,

$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$$



differential privacy

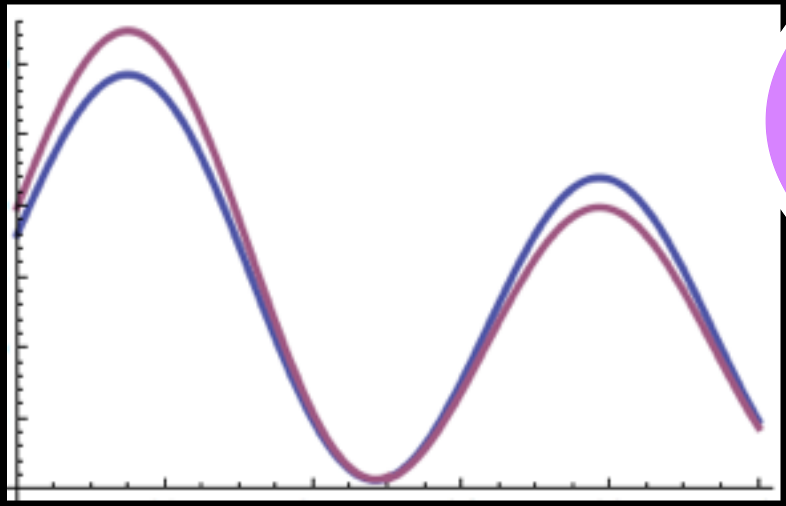
[DinurNissim03, DworkNissimMcSherrySmith06, Dwork06]

ε-Differential Privacy for algorithm M:

for any two neighboring data sets x_1, x_2, differing by the addition or removal of a single row

any $S \subseteq \text{range}(M)$,

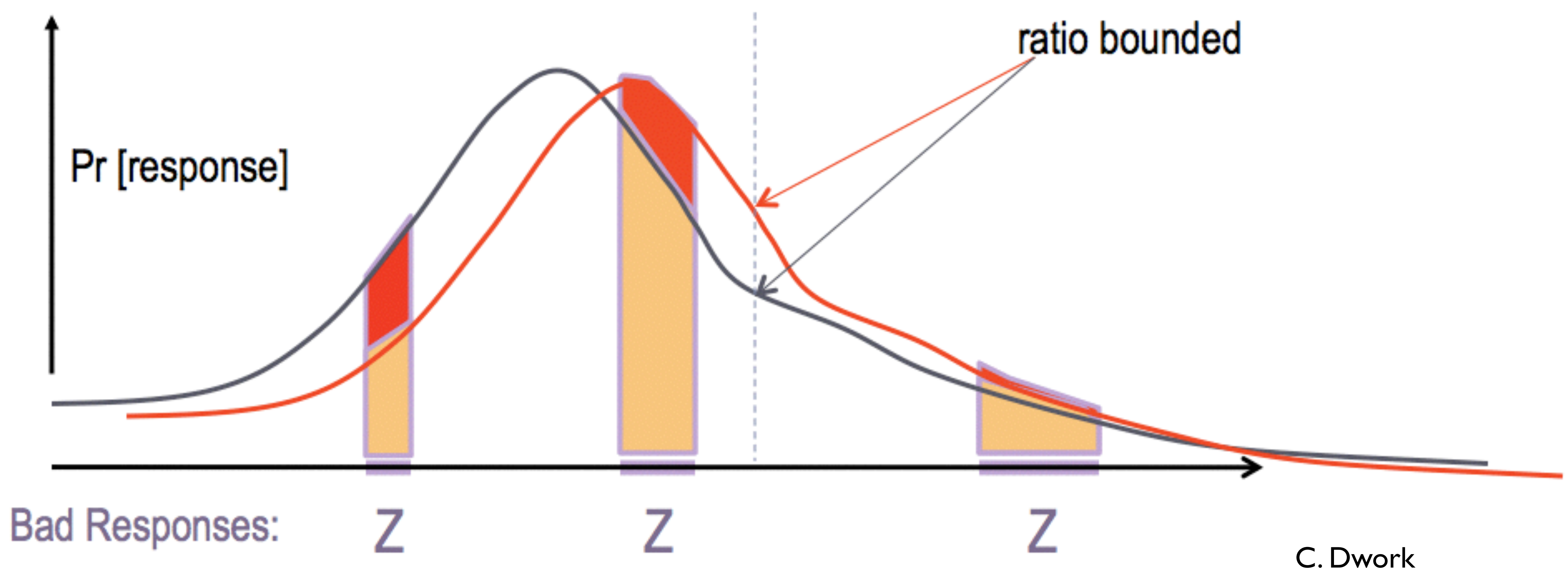
$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$



$e^{\varepsilon} \sim (1 + \varepsilon)$

differential privacy

$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$$



privacy, mechanisms, incentives, game theory

- Why would someone participate in a DP computation?
- Why would they give their true data?
- Would they need to be compensated? How much?
- How can the DP toolkit be used in game theory applications?

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
- incentives to participate and truth-tell in DP algorithms

game theory and mechanism design

- goal: solve some optimization problem
- catch: you don't have the inputs; they're held by self-interested agents
- common approach: design incentives and choice of solution ("mechanism") that incentivizes truth-telling

why truth-telling/strategy-proof?

- no need for participants to strategize
- simple to predict what will happen
- often, without loss of generality ("revelation principle"): if there is a non-truth-telling mechanism, replace it with a mechanism where the coordinator strategizes on behalf of the agents

LOTS of work in mechanism design on truthful mechanisms

- particular settings, constraints, goals, etc.

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
- incentives to participate and truth-tell in DP algorithms

the cheap answer (why participate, truth-tell?)

- Suppose agents $i \in [n]$ with types in X have utility functions $u_i : O \rightarrow [0, 1]$ over outcomes in O chosen by a mechanism M.

- We say $M : X^n \rightarrow \mathrm{O}$ is ε-approximately dominant strategy truthful if for every player i, for every $x_{-i} \in X^{n-1}$, and every $x'_i \in X$:

$$\mathbb{E}_{o \sim M(x)}[u_i(o)] \geq \mathbb{E}_{o \sim M(x'_i,\ x_{-i})}[u_i(o)] - \varepsilon$$

So, if a mechanism is ε-differentially private, it is also O(ε)-approximate dominant strategy truthful

the good news

- Composition very powerful! For example, if M_1 and M_2 are both ε-differentially private, their composition is O(ε)-approximately dominant strategy truthful.
- (Incentive properties of general strategy-proof mechanisms may not be preserved under composition.)

more good news

- If inputs x, y differ in the types of k players, we get

$$\mathbb{E}_{o \sim M(x)}[u(o)] \leq e^{\varepsilon k}\, \mathbb{E}_{o \sim M(y)}[u(o)]$$

- Changing up to k players' types changes the expected utility by at most ~$(1 + \varepsilon)$, when k << $1/\varepsilon$.
- DP mechanisms make truthful reporting a O($k\varepsilon$)-approximate dominant strategy, even for coalitions of k agents!
- In general dominant-strategy truthful mechanisms, robustness to collusion does not come for free.

more good news

- This is all without money!

the bad news

- Not only is truthfully reporting one's type an approximate dominant strategy, *any report* is an approximate dominant strategy.
- … perhaps we need to compensate (truthful) participation.

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
- incentives to participate and truth-tell in DP algorithms

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
  - DP gives asymptotic truthfulness
  - DP gives some new mechanism design results
  - DP and equilibrium selection
  - DP and exact truthfulness
- incentives to participate and truth-tell in DP algorithms

digital goods auctions

- unlimited supply of good with zero marginal cost of production
- n unit-demand buyers w/ valuations $v_i \in [0, 1]$
- $\mathrm{OPT} = \max_p \mathrm{Rev}(p, v) = \max_p\ p\,|\{\, i : v_i \geq p \}|$.

digital goods auctions

- unlimited supply of good with zero marginal cost of production
- n unit-demand buyers with valuations $v_i \in [0, 1]$
- $OPT = \max_p Rev(p, v) = \max_p\ p\ |\{ i : v_i \geq p\}|$.

digital goods auctions

- unlimited supply of good with zero marginal cost of production
- n unit-demand buyers with valuations $v_i \in [0, 1]$
- OPT = $\max_p \text{Rev}(p, v) = \max_p\ p\ |\{ i : v_i \geq p\}|$.
- [BBHM05] gives dominant strategy truthful mechanism with revenue ≥ OPT - O(sqrt(n))

digital goods auctions

- unlimited supply of good with zero marginal cost of production
- n unit-demand buyers with valuations $v_i \in [0, 1]$
- OPT = $\max_p \mathrm{Rev}(p, v) = \max_p\ p\,|\{ i : v_i \geq p\}|$.
- [BBHM05] gives dominant strategy truthful mechanism with revenue ≥ OPT - O(sqrt(n))
- [McSherryTalwar07] DP-based approach: discretize range, use exponential mechanism to select price. With high probability, gives price s.t. revenue is ≥ OPT - $O(\log n/\varepsilon)$. Approximately truthful if valuation reports binding. (Note: not the case that *every* report is an approximate dominant strategy.)

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
    - DP gives asymptotic truthfulness
    - DP gives some new mechanism design results
    - DP and equilibrium selection
    - DP and exact truthfulness
- incentives to participate and truth-tell in DP algorithms

game theory primer: equilibrium

- Nash equilibrium: an assignment of players to strategies so that no player would benefit by changing strategy, given how everyone else is playing

game theory primer: equilibrium

- Nash equilibrium: an assignment of players to strategies so that no player would benefit by changing strategy, given how everyone else is playing
- Correlated equilibrium: generalization, where players have access to correlating signal (traffic light; Waze)

equilibrium implementation with mediator[KearnsPaiRogersRothUllman14]

- setting: mechanism designer has limited power
  - cannot enforce that agents "use" the mediator
    - no ability to pay agents
    - can only recommend actions (not enforce them)
    - no prior over player types

equilibrium implementation with mediator[KearnsPaiRogersRothUllman14]

- Goal: agents report types; mechanism recommends equilibrium strategies to agents; agents incentivized to participate, report truthfully, and to follow equilibrium
- will want to use DP tools to make "robust" strategy recommendations
  - need game to be "large"
  - need to relax privacy notion

joint differential privacy

- my recommended strategy might reveal (too much about) my type
    - think: my suggested route from home to work tells you where my home and work are
- joint differential privacy: for each player, if she changes her input, the distribution over *everyone else*'s pieces of the output doesn't change too much

[KearnsPaiRogersRothUllman14]

in large games with private types, can implement a correlated equilibrium of the complete info game with a “strong” mediator (one who can verify your claim, if you do opt in, but can’t force you to take their recommendation)

[KearnsPaiRogersRothUllman14]

for more structured games (routing), can even achieve with “weak” mediator who can’t verify inputs

why this is surprising

not enough to compute equilibrium over those who opt-in, since may be an equilibrium of the wrong game—an agent could have a big effect on the equilibrium chosen, even if her actions within the game have limited impact on others' utilities

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
  - DP gives asymptotic truthfulness
  - DP gives some new mechanism design results
  - DP and equilibrium selection
  - DP and exact truthfulness
- incentives to participate and truth-tell in DP algorithms

obtaining exact truthfulness [NissimSmorodinskyTennenholtz12]

- one motivating question: facility location (each agent has a location and prefers to attend a school close to her; central designer must pick locations of schools to minimize overall travel time)
- might want to lie about your location in order to influence chosen locations

obtaining exact truthfulness [NissimSmorodinskyTennenholtz12]

- nonstandard environment
  - agents report types (locations)
  - mechanism picks outcome (locations of schools)
  - agents "react" (pick a school to attend)
  - reaction can be constrained based on reported type (you have to pick the school that's closest to your report)

obtaining exact truthfulness [NissimSmorodinskyTennenholtz12]

- Randomize between
  - a DP mechanism that gives approximate truthfulness
  - a punishing mechanism with bad guarantees on outcome utility, but that gives strict incentive to truth-tell

outline

- game theory primer
- DP gives approximate truthfulness
- DP as a tool in game theory
- incentives to participate and truth-tell in DP algorithms

differential privacy

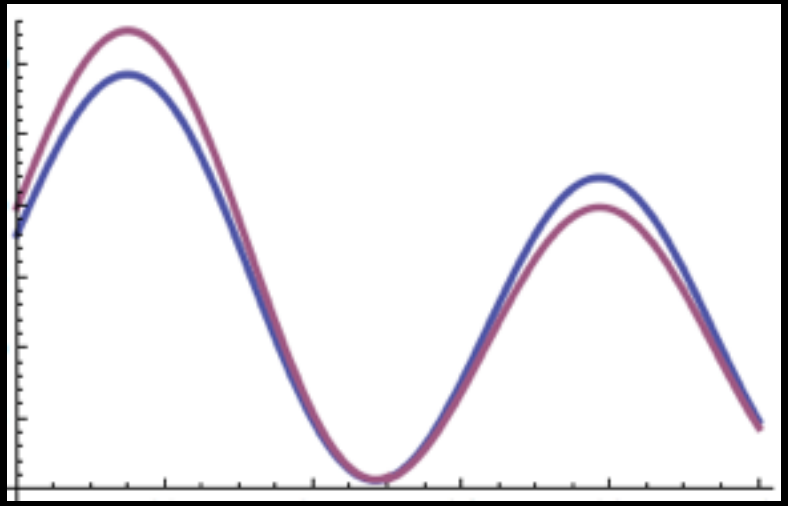
[DinurNissim03, DworkNissimMcSherrySmith06, Dwork06]

ε-Differential Privacy for algorithm M:

for any two neighboring data sets x_1, x_2, differing by the addition or removal of a single row

any $S \subseteq$ range(M),

$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$$



differential privacy

[DinurNissim03, DworkNissimMcSherrySmith06, Dwork06]

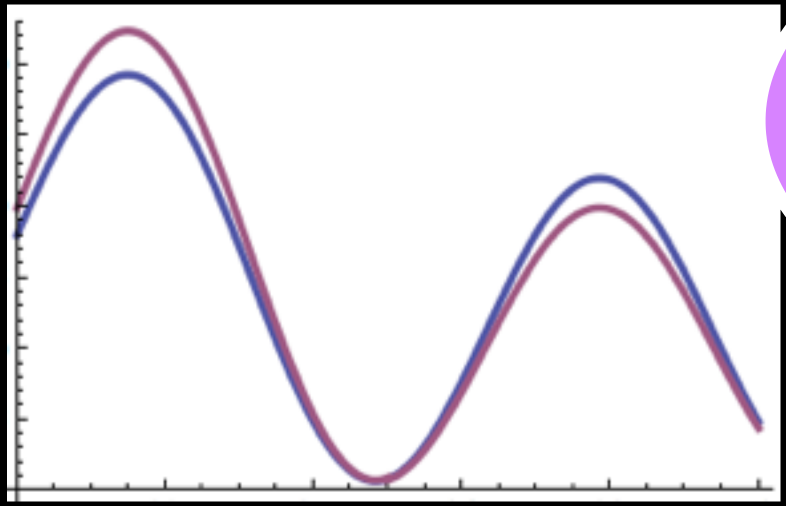
ε-Differential Privacy for algorithm M:

for any two neighboring data sets x_1, x_2, differing by the addition or removal of a single row

any $S \subseteq$ range(M),

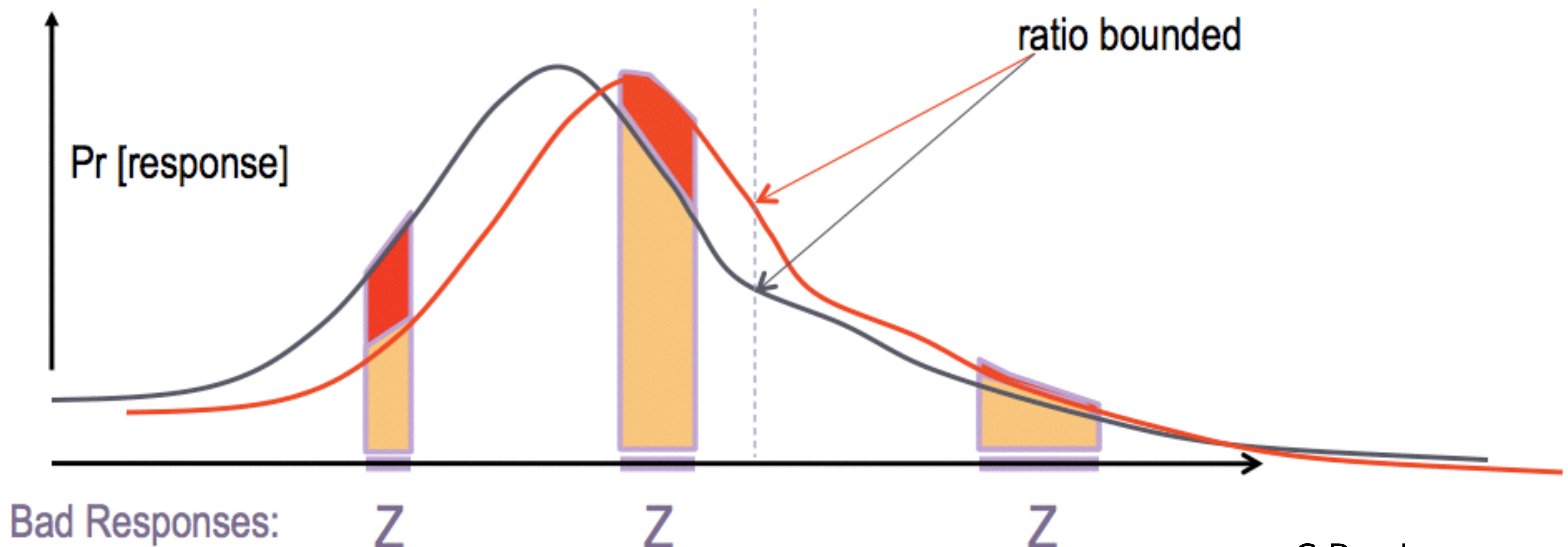
$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$$

$e^{\varepsilon} \sim (1 + \varepsilon)$



differential privacy

$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$$



C. Dwork

the issue of verification

the issue of verification

- challenging to strictly incentivize truth-telling in DP mechanisms, unless

the issue of verification

- challenging to strictly incentivize truth-telling in DP mechanisms, unless
  - agents care about the outcome

the issue of verification

- challenging to strictly incentivize truth-telling in DP mechanisms, unless
  - agents care about the outcome
  - responses are verifiable

the issue of verification

- challenging to strictly incentivize truth-telling in DP mechanisms, unless
  - agents care about the outcome
  - responses are verifiable
    - now or later

the issue of verification

- challenging to strictly incentivize truth-telling in DP mechanisms, unless
  - agents care about the outcome
  - responses are verifiable
    - now or later
    - with reasonable probability

Buying Private Data WITH Verification

Buying Private Data WITH Verification

- [GhoshRoth11] introduced problem of buying private data

Buying Private Data WITH Verification

- [GhoshRoth11] introduced problem of buying private data
- idea: want to buy sensitive information to estimate a population statistic, cheaply

[GhoshRoth11]: good news

[GhoshRoth11]: good news

- IF individuals don't care about privacy of their costs…

[GhoshRoth11]: good news

- IF individuals don't care about privacy of their costs…
- nearly optimal, truthful auctions

[GhoshRoth11]: good news

- IF individuals don’t care about privacy of their costs…
- nearly optimal, truthful auctions
  - fixed accuracy target, minimizing payments

[GhoshRoth11]: good news

- IF individuals don't care about privacy of their costs…
- nearly optimal, truthful auctions
  - fixed accuracy target, minimizing payments
  - fixed budget, maximizing accuracy

[GhoshRoth11, NissimVadhanXiao14]: bad news

[GhoshRoth11, NissimVadhanXiao14]: bad news

- Strong impossibility results for individually rational mechanisms when the costs themselves are private.

[GhoshRoth11, NissimVadhanXiao14]: bad news

- Strong impossibility results for individually rational mechanisms when the costs themselves are private.
- Wlog, assume true statistic is between 0 and n/2 with probability at least 1/2.

[GhoshRoth11, NissimVadhanXiao14]: bad news

- Strong impossibility results for individually rational mechanisms when the costs themselves are private.
- Wlog, assume true statistic is between 0 and n/2 with probability at least 1/2.
- In order to be meaningfully accurate, when input database is all 1's, should return a value greater than n/2 w.p., say, at least 2/3.

[GhoshRoth11, NissimVadhanXiao14]: bad news

- Strong impossibility results for individually rational mechanisms when the costs themselves are private.
- Wlog, assume true statistic is between 0 and n/2 with probability at least 1/2.
- In order to be meaningfully accurate, when input database is all 1's, should return a value greater than n/2 w.p., say, at least 2/3.
- By DP, sum of the epsilons must be greater than ln 4/3.

[GhoshRoth11, NissimVadhanXiao14]: bad news

- Strong impossibility results for individually rational mechanisms when the costs themselves are private.
- Wlog, assume true statistic is between 0 and n/2 with probability at least 1/2.
- In order to be meaningfully accurate, when input database is all 1's, should return a value greater than n/2 w.p., say, at least 2/3.
- By DP, sum of the epsilons must be greater than ln 4/3.
- To get IR, total payment must exceed min v_i * sum of epsilons.

[GhoshRoth11, NissimVadhanXiao14]: bad news

- Strong impossibility results for individually rational mechanisms when the costs themselves are private.
- Wlog, assume true statistic is between 0 and n/2 with probability at least 1/2.
- In order to be meaningfully accurate, when input database is all 1's, should return a value greater than n/2 w.p., say, at least 2/3.
- By DP, sum of the epsilons must be greater than ln 4/3.
- To get IR, total payment must exceed min v_i * sum of epsilons.
- By DP, this must hold for *all* inputs, so cannot make finite payment.

[GhoshRoth11, NissimVadhanXiao14]: bad news

[GhoshRoth11, NissimVadhanXiao14]: bad news

- [NVX14] strengthen impossibility results of GR11, extending to much wider class of privacy valuations, including (ε, δ)-DP

responding to impossibility

- [FleischerLyu12]: c_i drawn from known prior given b_i; relies on knowing prior exactly
- [LigettRoth12]: take-it-or-leave-it offers (lose individual rationality); revised model of privacy costs
- [NissimVadhanXiao14]: monotonicity of correlation between bits and costs; known bound on how many players' costs exceed a given threshold

forms of report verification

- direct (check your driver's license, draw your blood)
- possibly randomized
- agents care about outcome (or can be scored based on future event) - prediction market
- correlations in population

Challenge: No observed outcome

Challenge: No observed outcome

Challenge: No observed outcome

- Should we acquire company X?

Challenge: No observed outcome

- Should we acquire company X?
- What is the prevalence of drug use?

Challenge: No observed outcome

- Should we acquire company X?
- What is the prevalence of drug use?
- Do our employees accept bribes?

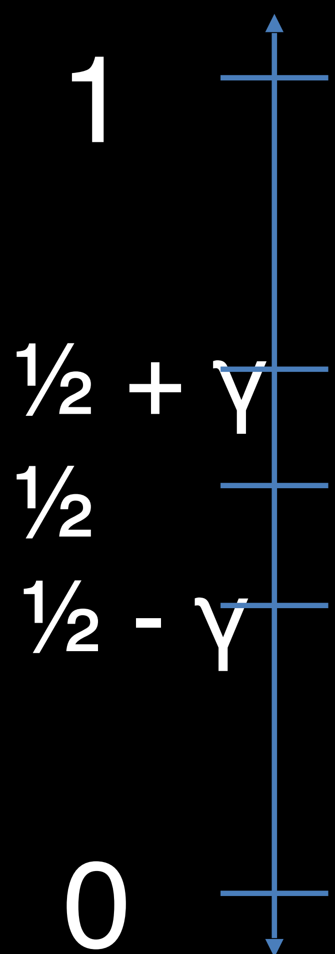
Challenge: No observed outcome

- Should we acquire company X?
- What is the prevalence of drug use?
- Do our employees accept bribes?
- Are students cheating in class?

Bayesian setting

- bit-cost pairs (b_i , c_i) drawn from known joint distribution
- agent's cost c_i does not give her additional information about other agents beyond what was conveyed by b_i

example Bayesian setting



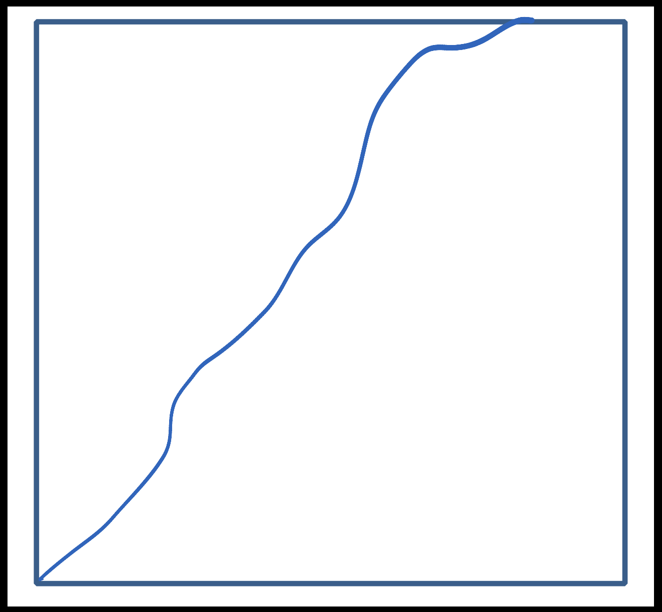
$\Pr[b_j = b \mid b_i = b]$

$\Pr[b_j = b]$

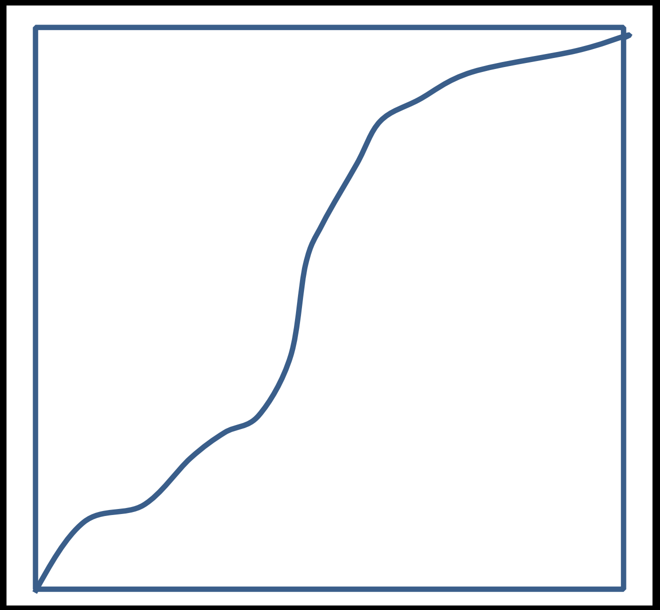
$\Pr[b_j = b \mid b_i = \neg b]$

example Bayesian setting

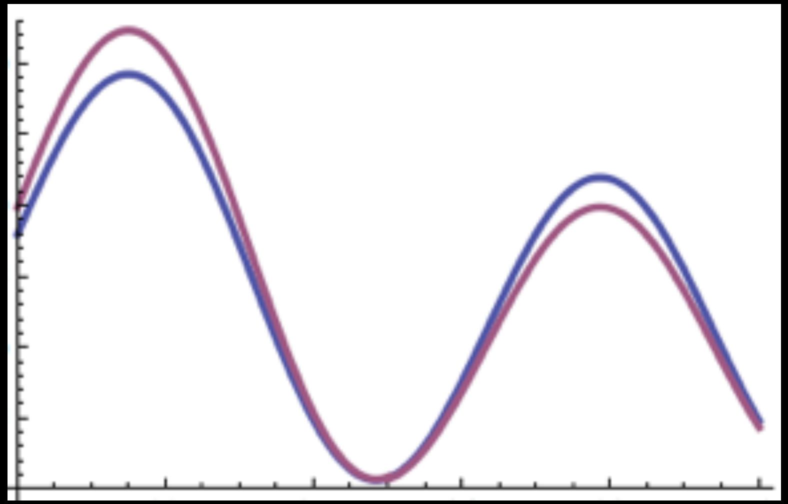
$Pr[c_i > c|b_i=0]$



$Pr[c_i > c| b_i=1]$

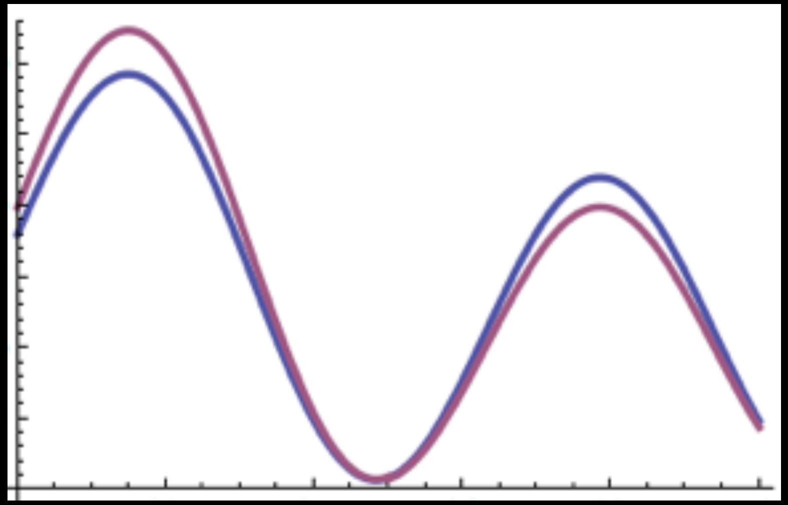


modeling privacy costs



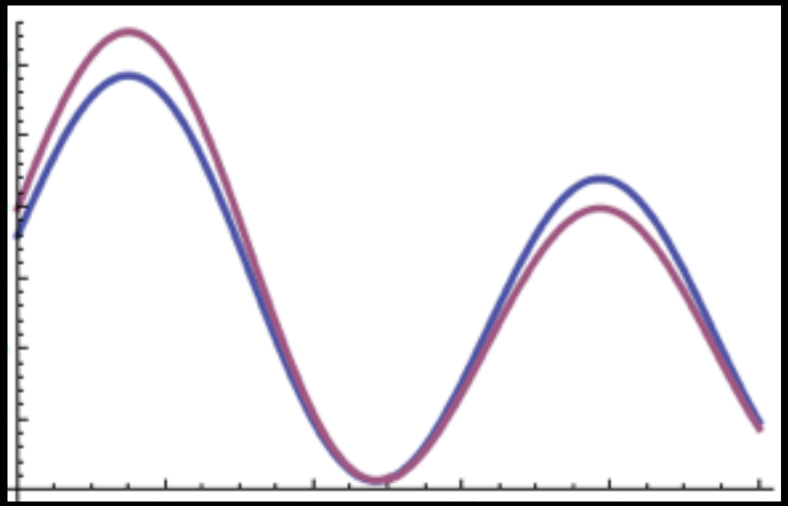
- for most results, adopt model of [NissimOrlandiSmorodinsky12]: privacy costs can be arbitrary, but upper-bounded by linear cost c_i ε

modeling privacy costs



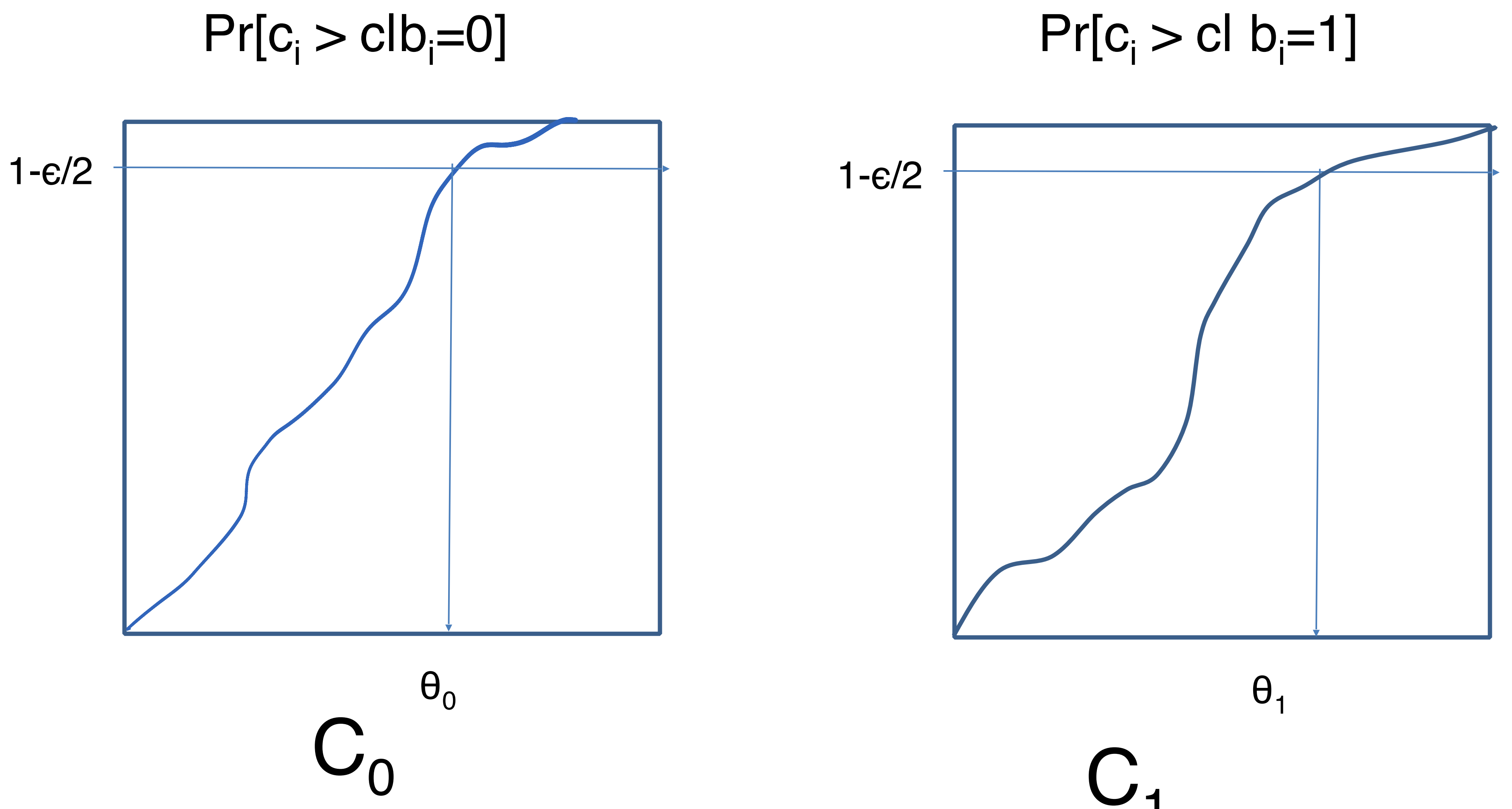
- for most results, adopt model of [NissimOrlandiSmorodinsky12]: privacy costs can be arbitrary, but upper-bounded by linear cost $c_i\ \varepsilon$
- utility model: bounded by $c_i\ \varepsilon$ - p_i

modeling privacy costs



- for most results, adopt model of [NissimOrlandiSmorodinsky12]: privacy costs can be arbitrary, but upper-bounded by linear cost $c_i\ \varepsilon$
- utility model: bounded by $c_i\ \varepsilon$ - p_i
- could also incorporate explicit preferences to manipulate outcome

participation threshold



Let $\theta = \max\{\theta_0, \theta_1\}$

if *verification* weren't an issue…

1. Collect $\widehat{b_i} \in \{0, 1, \perp\}$
2. Release $\frac{\left|\{i : \widehat{b_i} = 1\}\right| + \lambda(\frac{\epsilon n}{2})}{n}$
3. Pay $\frac{2\theta}{\epsilon n}$

- $\frac{2}{\epsilon n}$-Differentially Private
- Expected Error: $\frac{\epsilon}{2}$ from noise, $\leq \frac{\epsilon}{2}$ from non-participation
- Cost: $\frac{2\theta}{\epsilon}$

if *privacy* weren't an issue…

- peer-prediction literature [MillerResnickZeckhauser05]

if *privacy* weren't an issue…

- peer-prediction literature [MillerResnickZeckhauser05]
- key idea: reward participants for reports that are predictive of *others*' reports

if *privacy* weren't an issue…

- peer-prediction literature [MillerResnickZeckhauser05]
- key idea: reward participants for reports that are predictive of *others'* reports
- uses proper scoring rule, which incentivizes participants to truthfully report beliefs (e.g., log of probability mass you placed on event that actually occurred)

peer-prediction algorithm

- randomly pair players i and j
- pay player i $\text{properScoringRule}(r_j, p_{ri})$
  - r_j is player j's reported bit
  - p_{ri} is the posterior based on player i's reported bit

challenges of privacy

challenges of privacy

- my payment reveals too much about me

challenges of privacy

- my payment reveals too much about me
- being paid based on a single other player's bit too revealing

challenges of privacy

- my payment reveals too much about me
- being paid based on a single other player’s bit too revealing
- can’t get full participation at any fixed cost

challenges of privacy

- my payment reveals too much about me
- being paid based on a single other player’s bit too revealing
- can’t get full participation at any fixed cost
- incentive to truth-tell must be robust to noise in aggregation and to error due to lack of full participation

challenges of privacy

- my payment reveals too much about me
- being paid based on a single other player's bit too revealing
- can't get full participation at any fixed cost
- incentive to truth-tell must be robust to noise in aggregation and to error due to lack of full participation
- more noise: directly harms accuracy, but encourages participation (which helps accuracy)

joint differential privacy

- the amount you are paid is too revealing
- give a guarantee under “joint differential privacy,” wherein the closeness differential privacy requires is on the computation’s outcome and everyone else’s payments

private peer-prediction [GhoshLigettRothSchoenebeck15]

private peer-prediction [GhoshLigettRothSchoenebeck15]

private peer-prediction [GhoshLigettRothSchoenebeck15]

1. Collect $\hat{b}_i \in \{0, 1, \perp\}$
2. Compute $\bar{b} = \left|\{i : \hat{b}_i = 1\}\right| + \lambda\left(\frac{\epsilon n}{2}\right)$
3. Compute $\bar{a} = \frac{\bar{b}}{n}$, $\overline{a_{-i}} = \frac{\bar{b} - b_i}{n-1}$
4. Release $\bar{a}$
5. Payment $p_i = \frac{2\theta}{\epsilon n(2\gamma - \epsilon)} \left(1 - \overline{a_{-i}}\right)$ if $\hat{b}_i = 0$;

   $p_i = \frac{2\theta}{\epsilon n(2\gamma - \epsilon)} \overline{a_{-i}}$ if $\hat{b}_i = 1$;

   $p_i = 0$ if $\hat{b}_i = \perp$

- $\frac{2}{\epsilon n}$-JointDP
- Equilibrium for agents with costs $< \theta$ to truth-tell
- Expected Error: $\frac{\epsilon}{2}$ from noise, $\leq \frac{\epsilon}{2}$ from non-participation
- Cost: $\frac{2\theta}{\epsilon(2\gamma - \epsilon)}$

private peer-prediction: sketch of accuracy proof

- accuracy comes from truthfulness of enough players

private peer-prediction: sketch of accuracy proof

- accuracy comes from truthfulness of enough players
- show existence of threshold strategy equilibrium, where all agents with cost below threshold are incentivized to truth-tell

private peer-prediction: sketch of accuracy proof

- accuracy comes from truthfulness of enough players
- show existence of threshold strategy equilibrium, where all agents with cost below threshold are incentivized to truth-tell
- find threshold such that a large fraction of players have costs below it, and for all players, conditioning on having either bit, posterior says large fraction of others have costs below it

slightly more specific measure of privacy costs [ChenChongKashMoranVadhan13]

- adversary
  - cannot see agents' participation
  - updates belief about agent based on outcome

slightly more specific measure of privacy costs [ChenChongKashMoranVadhan13]

- adversary
  - cannot see agents' participation
  - updates belief about agent based on outcome
- if mechanism is ε-DP then agent only affects outcome with probability ε

slightly more specific measure of privacy costs [ChenChongKashMoranVadhan13]

- adversary
  - cannot see agents' participation
  - updates belief about agent based on outcome
- if mechanism is ε-DP then agent only affects outcome with probability ε
- with probability ε, adversary changes view by ε, so cost of participation is $c_i\varepsilon^2$

slightly more specific measure of privacy costs [ChenChongKashMoranVadhan13]

- adversary
  - cannot see agents' participation
  - updates belief about agent based on outcome
- if mechanism is ε-DP then agent only affects outcome with probability ε
- with probability ε, adversary changes view by ε, so cost of participation is $c_i\varepsilon^2$
- can achieve 0 cost in limit of n

privacy + game theory

- DP gives asymptotic truthfulness, some new mechanism design and equilibrium selection results
- the asymptotic truthfulness toolkit is sometimes useful for getting exact truthfulness
- interesting challenge of modeling costs for privacy
- interesting challenges in elicitation/payment for private data

if privacy is for humans…

- do we need to understand…
    - how people currently value it?
        - how people behave with respect to it? (revealed preferences)
    - how people “should” value it (if they were rational, understood risks, etc.)?
    - how the technologies we enable and implement change people’s value for and expectations of privacy?
- what are the right promises to give?