Optimal Privacy-Constrained Mechanisms

Ran Eilat, Kfir Eliaz and Xiaosheng Mu*

Simons Institute, UC Berkeley

May 6, 2019

Motivation

Introduce and analyze a Bayesian measure of privacy loss.

- Most work on differential privacy (Dwork et al. '06) is "prior-free"
- From an outsider's perspective, the realized outcome of a DP mechanism does not reveal much about any individual participant's type

$$\mathbb{E}_{o \sim M(t)}[u(o)] \leq \exp(\epsilon) \cdot \mathbb{E}_{o \sim M(t')}[u(o)]$$

Motivation

Introduce and analyze a Bayesian measure of privacy loss.

- Most work on differential privacy (Dwork et al. '06) is "prior-free"
- From an outsider's perspective, the realized outcome of a DP mechanism does not reveal much about any individual participant's type

$$\mathbb{E}_{o \sim M(t)}[u(o)] \leq \exp(\epsilon) \cdot \mathbb{E}_{o \sim M(t')}[u(o)]$$

- But to implement this, types have to be reported
- We might worry about the designer knowing too much
- Our approach: mechanism design under a privacy constraint that *limits how much information the principal can collect from the agents*

Framework

We measure privacy loss by how much the principal learns about agent types through observing what they choose in the mechanism. Specifically,

Framework

We measure privacy loss by how much the principal learns about agent types through observing what they choose in the mechanism. Specifically,

1. Principal has prior belief F about agent types t

Framework

We measure privacy loss by how much the principal learns about agent types through observing what they choose in the mechanism. Specifically,

1. Principal has prior belief F about agent types t
2. He offers a general (potentially indirect) mechanism $\mathbb{M}$ specifying the message set M and how messages are mapped to outcomes
3. Agents play a Bayesian equilibrium

Framework

We measure privacy loss by how much the principal learns about agent types through observing what they choose in the mechanism. Specifically,

1. Principal has prior belief F about agent types t
2. He offers a general (potentially indirect) mechanism $\mathbb{M}$ specifying the message set M and how messages are mapped to outcomes
3. Agents play a Bayesian equilibrium
4. Given equilibrium message m, principal forms posterior belief $F(\cdot \mid m)$

Framework

We measure privacy loss by how much the principal learns about agent types through observing what they choose in the mechanism. Specifically,

1. Principal has prior belief F about agent types t
2. He offers a general (potentially indirect) mechanism $\mathbb{M}$ specifying the message set M and how messages are mapped to outcomes
3. Agents play a Bayesian equilibrium
4. Given equilibrium message m, principal forms posterior belief $F(\cdot \mid m)$
5. Privacy loss defined as expected KL-divergence between posterior and prior beliefs:

$$I(\mathbb{M}) = \mathbb{E}_m \left[D(F(\cdot \mid m) \,||\, F) \right]$$

Framework

We measure privacy loss by how much the principal learns about agent types through observing what they choose in the mechanism. Specifically,

1. Principal has prior belief F about agent types t
2. He offers a general (potentially indirect) mechanism $\mathbb{M}$ specifying the message set M and how messages are mapped to outcomes
3. Agents play a Bayesian equilibrium
4. Given equilibrium message m, principal forms posterior belief $F(\cdot \mid m)$
5. Privacy loss defined as expected KL-divergence between posterior and prior beliefs:

$$I(\mathbb{M}) = \mathbb{E}_m \left[D(F(\cdot \mid m) \mid\mid F) \right]$$

6. Principal constrained by $I(\mathbb{M}) \leq \kappa$ with κ exogenously given

Discussion

- Definition equivalent to MI between types and messages:
  - early version of Xiao ('13) considers MI as a cost to each agent
  - we take the paternalistic viewpoint of a regulator but do not directly model agent preferences for privacy
  - alternatively, each agent participates only if constraint is met
- Above measure of privacy loss takes average across different messages:
  - more stringent "ex-post" notion requires $D(F(\cdot \mid m) \,||\, F) \leq \kappa, \; \forall m$
  - results similar; focus on ex-ante case here
- Related issue of how to aggregate privacy loss across multiple agents:
  - paper studies an application with only one agent

Screening Environment

Focus on the monopolistic screening model of Mussa-Rosen ('78).

- A seller sells some quantity/quality $q \geq 0$ to a buyer for payment p
- Buyer type $\theta \in [\underline{\theta}, \overline{\theta}]$ distributed as F with positive density.

Screening Environment

Focus on the monopolistic screening model of Mussa-Rosen ('78).

- A seller sells some quantity/quality $q \geq 0$ to a buyer for payment p
- Buyer type $\theta \in [\underline{\theta}, \overline{\theta}]$ distributed as F with positive density.
- Profit net of production cost $p - \frac{q^2}{2}$. Utility $q \cdot \theta - p$
- Assume F has increasing and positive virtual values, so classic model predicts direct mechanism providing monotone quantities

Screening Environment

Focus on the monopolistic screening model of Mussa-Rosen ('78).

- A seller sells some quantity/quality $q \geq 0$ to a buyer for payment p
- Buyer type $\theta \in [\underline{\theta}, \overline{\theta}]$ distributed as F with positive density.
- Profit net of production cost $p - \frac{q^2}{2}$. Utility $q \cdot \theta - p$
- Assume F has increasing and positive virtual values, so classic model predicts direct mechanism providing monotone quantities
- In our model, seller maximizes profit subject to privacy. That is,

$$\max \ \mathbb{E}_m[p(m) - c(q(m))] \quad s.t. \quad \mathbb{E}_m\left[D(F(\cdot \mid m) \mid\mid F)\right] \leq \kappa$$

Coarse Revelation

Main Result

Given $0 < \kappa < \infty$. There exists an optimal privacy-constrained mechanism $\mathbb{M}$, where the set of types $[\underline{\theta}, \overline{\theta}]$ is partitioned into finitely many intervals, and in equilibrium each type truthfully reports its interval.

Coarse Revelation

Main Result

Given $0 < \kappa < \infty$. There exists an optimal privacy-constrained mechanism $\mathbb{M}$, where the set of types $[\underline{\theta}, \overline{\theta}]$ is partitioned into finitely many intervals, and in equilibrium each type truthfully reports its interval.

Further properties:

- privacy constraint binds in any optimal mechanism
- if κ small, exactly two intervals used

Why Intervals?

Several papers (e.g. Bergemann et al.) derived optimality of intervals by assuming upper bound on number of messages. For us,

1. First remove "redundant" messages: If two messages lead to same outcome, combine them into a single message
   $\Longrightarrow$ posterior belief is *averaged*, implying smaller privacy loss

2. Types that send different messages partition the type space

3. By single-crossing property, each partition is convex

4. Thus intervals – this part does not rely on specific form of KL; also extends to multiple agents with one-dimensional types

Finiteness

Where we use KL is to show finite intervals suffice.

- Technical difficulty as space of partitions is *not compact*
- We restore compactness by showing at most one short interval
- Otherwise, *merge two short intervals* and use saved privacy to *divide a long interval*. Profit would increase
- Intuition: "log" term in KL punishes heavily against getting precise information about even a small set of types

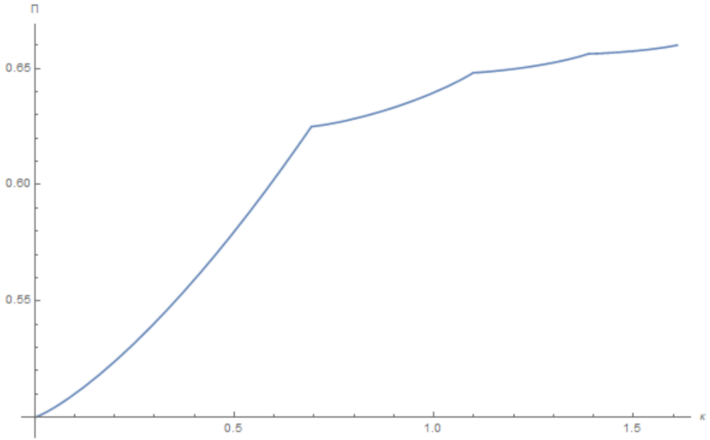
Uniform Case

Consider special case with uniform types. Can show "ordering" of intervals do not matter for profit and privacy measure.

Characterization

With uniform prior, for any κ, the optimal privacy-constrained mechanism partitions $[\underline{\theta}, \overline{\theta}]$ into $n-1$ *equally long* intervals and 1 *shorter* interval, such that the privacy constraint is exhausted.

Profit Frontier



plotted for θ~$U\left[1,2\right]$

Welfare Analysis

Comparative Statics w.r.t. κ

1. Profit from a κ-constrained optimal mechanism increases in κ

Welfare Analysis

Comparative Statics w.r.t. κ

1. Profit from a κ-constrained optimal mechanism increases in κ
2. Buyer surplus is maximized (resp. minimized) with full (resp. no) privacy

Welfare Analysis

Comparative Statics w.r.t. κ

1. Profit from a κ-constrained optimal mechanism increases in κ
2. Buyer surplus is maximized (resp. minimized) with full (resp. no) privacy
3. If prior density $f(\theta)$ decreases, no privacy maximizes total welfare

Recap

- Bayesian privacy measure: how much principal learns via mechanism
  $\Longrightarrow$ Coarse menu offered in the form of interval partition
- Implementation: where does the prior come from?
- Multiple agents: how to aggregate privacy?
- Dynamic mechanisms?

Thank You!