Selling Data to Agent with Endogenous Information
Yingkai Li (Northwestern University → Yale University (Postdoc))

Selling Information
Seller maximizes the revenue by selling information to a buyer. Buyer is a decision maker who uses acquired information to improve the expected utility from better decisions.

- buyer has private value of information;
- buyer can endogenously acquire additional information.

Timeline:
1. Seller offers a menu for selling information.
2. Buyer chooses an option for buying information from seller.
3. Buyer receives a signal from the seller, and then purchase additional information endogenously from other sources.
4. Buyer makes a decision and final utility is realized.

Comparison to exogenous information:
Buyer is endowed with additional information regardless of the selling mechanism.

Sequential-learning Proof
Buyer has costs for acquiring additional information.

- buyer can sequentially acquire any feasible information;
- expected cost of any sequential experiment is weakly higher than any single experiment that is equally informative.

Example: Gaussian Learning

- prior is Gaussian;
- acquire a signal with Gaussian noise with fixed cost $c > 0$.

Optimal Mechanisms

**Theorem.** In the optimal mechanisms, the buyer has no incentive to acquire additional costly information in eqm.

Comparative Statics:

- optimal revenue increases when the set of feasible information the buyer can purchase shrinks;
- optimal revenue increases when buyer’s cost of acquiring additional information increases.

Linear Valuation
Buyer’s value for information is separable between his private type and the improvement from refining posterior belief.

- expected value $V(\mu, \theta) = \theta \cdot v(\mu)$
  
  $\Rightarrow$ value of info $V(\sigma, \mu, \theta) = \theta \cdot (E_{\hat{\mu} \sim D}[v(\hat{\mu})] - v(\mu))$;

Example: matching utilities

- buyer with private type $\theta$ tries to predict unknown outcome;
- buyer receives value $\theta$ if the prediction is correct and 0 o.w.

**Theorem.** Assuming linear valuation, in the optimal mechanisms, there is a threshold on the type such that

- higher types receives full information;
- low types receives partial information which is what they would purchase from other sources if the seller offers no information.

Pricing for Full Information

**Theorem (Exogenous Information [Bergemann, Cai, Velegkas, Zhao ’21])** Assuming linear valuation and regular distribution, for any $\epsilon > 0$, there exists an instance such that pricing for full information is at most $\epsilon$ fraction of the optimal revenue.

**Theorem (Endogenous Information)** Assuming linear valuation and regular distribution, pricing for full information achieves at least half of the optimal revenue.

![Figure: Reduction on identifying the type distribution that maximizes the approximation ratio between the optimal revenue and optimal pricing.]

Interpretation:

- price discrimination is crucial if buyer’s additional information is exogenous, but not crucial if it is endogenous.
- extreme information structure such that pricing is a bad approximation will not arise endogenously.

Working paper at arXiv:2103.05788