Lecture 11: The Valuation and Pricing of Information

I.Economic Foundations of Value of Information

Coin state ω

Prior belief $q(\omega)$

action		tail (1/2)	head (1/2)
	0	$2t$	0
	1	0	$2t$

$u(a,\omega)$

Without information, decision maker(DM) gets $\max_a \left[\mathbb{E}_{\omega\sim q} u(a,\omega)\right] = t$

With my (full) information, DM gets $\mathbb{E}_{\omega\sim q}[\max_a u(a,\omega)] = 2t$

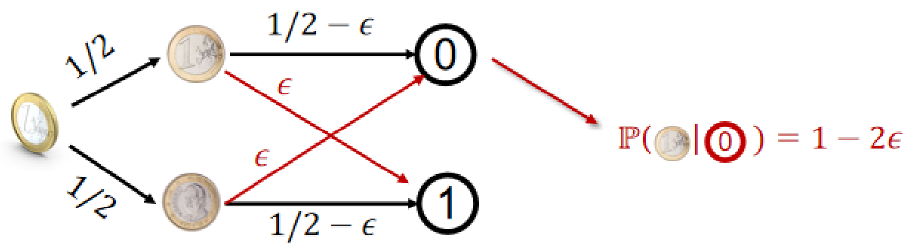
Value of (full) info $= \mathbb{E}_{\omega\sim q}[\max_a u(a,\omega)] - \max_a \left[\mathbb{E}_{\omega\sim q} u(a,\omega)\right]$

Posterior belief $p(\omega)$, given 0

Prior belief $q(\omega)$

action		tail ($1-2\epsilon$; 1/2)	head (2ϵ; 1/2)
	0	t	0
	1	0	$3t$

$u(a,\omega)$



Noisy information revelation

Realistically, can think of 0 as noisy prediction of state ω (e.g., stock trend, purchase prob)

Question: What is the value of this noisy signal 0?

- Without knowing this signal, DM takes action 1
- With this signal 0, DM takes action 0(assuming ε very small)
- However, true distribution is the posterior p regardless

$$\text{Value of knowing } 0 = \mathbb{E}_{\omega \sim p}[u(0,\omega)] - \mathbb{E}_{\omega \sim p} u(1,\omega)$$

Definition(FK'19):Consider an arbitrary desicion making problem $u(a,\omega)$, suppose a signal updates the DM's belief about state ω from $q \in \Delta(\Omega)$ to $p \in \Delta(\Omega)$, the value of this signal is defined as

$$D^u(p;q) = \mathbb{E}_{\omega \sim p}[u(a^*(p),\omega)] - \mathbb{E}_{\omega \sim p} u(a^*(q),\omega)$$

Example 1:
- $a \in A = \Delta(\Omega) \rightarrow$ action is to pick a distribution over states
- $u(a,\omega) = \log a_\omega$
- Which action $a \in \Delta(\Omega)$ maximizes expected utility $\mathbb{E}_{\omega \sim p}[u(a,\omega)]$?

$$a^*(p) = p$$

$$D^u(p;q) = \sum_\omega p_\omega \log \frac{p_\omega}{q_\omega} \qquad \text{KL-divergence}$$

Example 2:
- $a \in A = \Delta(A) \rightarrow$ action is to pick a distribution over states
- $u(a,\omega) = -\|a - e_\omega\|^2$

$$a^*(p) = p$$

$$D^u(p;q) = \|p-q\|^2 \qquad \text{Squared distance}$$

Some obvious properties
- Non-negativity: $D^u(p;q) \geq 0$
- Null information has no value: $D^u(q;q) = 0$
- Order-invariant: if DM receives signal σ_1, σ_2, the order of receiving them does not affect final expected total value

In this case, we say $D(p;q)$ is a valid measure for value of information

Theorem 1(FK'19): Consider any $D(p;q)$ function. There exists a decision problem $u(a,\omega)$ such that

$$D^u(p;q) = \mathbb{E}_{\omega \sim p}[u(a^*(p),\omega)] - \mathbb{E}_{\omega \sim p} u(a^*(q),\omega)$$

if and only if $D(p;q)$ satisfies
- Non-negativity: $D^u(p;q) \geq 0$
- Null information has no value: $D^u(q;q) = 0$
- Order-invariant: if DM receives signal σ_1, σ_2, the order of receiving them does not affect final expected total value

Theorem 2(FK'19):
1. For any concave function H, its Bregman divergence is a valid measure for value of information.
2. Conversely, for any valid measure $D(p;q)$ for value of information,

$$H(q) = \sum_\omega q^\omega D(e_\omega, q)$$

   is a concave function whosw Bregman divergence is $D(p;q)$

Why useful?
- Many functions - even natural ones like l_2 distance $\|p-q\|$- are not valid measures

- In fact, any metric is not valid, since metric cannot be a Bregman divergence
- There are efficient ways to tell whether a $D(p, q)$ is valid

II.Optimal Pricing of Information

1.Plans

- Vignette 1: closed-form optimal mechanism for structed setups
- Vignette 2: algorithmic solution for general setups
- Vignette 3: from distilled data (i.e. information) to raw data

2.A Model of Information Pricing

- One seller, one buyer
- Buyers is a decision marker who faces a binary choice: an active action 1 and a passive action 0: Active action: come to talk, approve loan, invest stock X, etc.
- Payoff of passive action $\equiv 0$
- Payoff of active action$= v(\omega, t) = v_1(\omega)[t + \rho(\omega)]$
  - ω is a state of nature, t is buyer type
  - Assume $v(\omega, t)$ is linear and non-decreasing in $t \in [t_1, t_2]$
- Information structure:
  - Seller observes ω, and buyer knows t

Mechanism design question: How can seller optimally sell her information about ω to the buyer?

3.Design Space

Standard revelation principle implies optimal mechanism can w.l.o.g be a menu $\{\pi_t, p_t\}_{t \in T}$

- $\pi_t : \Omega \to S$ is an experiment (which generates signals) for type i
- $p_t \in \mathbb{R}$ is t's payment
- Each type is incentivized to report type truthfully

Concrete design question: design IC $\{\pi_t, p_t\}_{t \in T}$ to maximize seller's revenue

4.How Does It Differ from Selling Goods?

- Each experiment is like an item
  - In this sense, we are selling infinitely many goods
  - In fact, we are even "designing the goods"
- Participation constraint is different
  - Without any information, type t's utility is $\max\{\overline{v}(t), 0\}$

$$\overline{v}(t) = \int_{\omega \in \Omega} v(\omega, t) g(\omega) d\omega$$

Ex-ante expected utility of action 1

5.Threshold experiments turn out to suffice

Recall $v(\omega, t) = v_1(\omega)[t + \rho(\omega)]$

Def. π_t is a threshold experiment if π_t simply reveals $\rho(\omega) \geq \theta(t)$ or not some buyer-type-dependent threshold $\theta(t)$

Threshold is on $\rho(\omega)$

6.Virtual Value Functions

Recall virtual value function in [Myerson'81]: $\varphi(t)=t-\frac{1-F(t)}{f(t)}$

Def.

- Lower virtual value function: $\underline{\varphi}(t)=t-\frac{1-F(t)}{f(t)}$
- Upper virtual value function: $\overline{\varphi}(t)=t+\frac{F(t)}{f(t)}$
- Mixed virtual value function: $\varphi_c(t)=c\underline{\varphi}(t)+(1-c)\overline{\varphi}(t)$

Note:"upper" or "lower" is due to:

$$\underline{\varphi}(t)\le t\le\overline{\varphi}(t)$$

7.The Optimal Mechanism

Depend on two problem-related constants:

$$V_L=\max\{v(t_1),0\}+\int_{t_1}^{t_2}\int_{q:\rho(q)\ge-\underline{\varphi}^+(x)}g(q)v_1(q)dqdx$$

$$V_H=\max\{v(t_1),0\}+\int_{t_1}^{t_2}\int_{q:\rho(q)\ge-\overline{\varphi}^+(x)}g(q)v_1(q)dqdx$$

Note:

$$V_L<V_H$$

Theorem([LSX'21])

1. If $\overline{v}(t_2)\le V_L$, the mechanism with threshold experiments $\theta^*(t)=-\underline{\varphi}(t)$ and following payment function represents an optimal mechanism:

$$p^*(t)=\int_{\omega\in\Omega}\pi^*(\omega,t)g(\omega)v(\omega,t)d\omega-\int_{t_1}^{t}\int_{\omega\in\Omega}\pi^*(\omega,x)g(\omega)v_1(\omega)d\omega dx$$

2. If $\overline{v}(t_2)\ge V_H$, the mechanism wit threshold experiments $\theta^*(t)=-\overline{\varphi}(t)$ and following payment function represents an optimal mechanism:

$$p^*(t)=\int_{\omega\in\Omega}\pi^*(\omega,t)g(\omega)v(\omega,t)d\omega+\int_{t}^{t_2}\int_{\omega\in\Omega}\pi^*(\omega,x)g(\omega)v_1(\omega)d\omega dx-\overline{v}(t_2)$$

3. If $V_L\le\overline{v}(t_2)\le V_H$, the mechanism with threshold experiments $\theta^*(t)=-\varphi_c(t)$ and following payment function represents an optimal mechanism

$$p^*(t)=\int_{\omega\in\Omega}\pi^*(\omega,t)g(\omega)v(\omega,t)d\omega-\int_{t_1}^{t_2}\int_{\omega\in\Omega}\pi^*(\omega,x)g(\omega)v_1(\omega)d\omega dx$$

   where constant c is chosen such that

$$\int_{t_1}^{t_2}\int_{\omega:\rho(\omega)\ge\varphi_c^+(x)}g(\omega)v_1(\omega)d\omega dx=\overline{v}(t_2)$$

8.Remarks

- Threshold mechanisms are common in real life:House/car inspections, stock recommendations: information seller only need to reveal it "passed" or "deserves a buy" or not

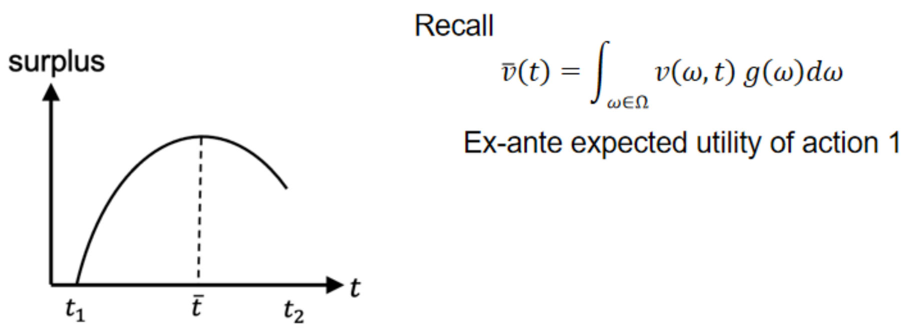
- Optimal mechanism has personalized thresholds and payments, tailored to accommodate different level of risk each buyer type can take:Different from optimal pricing of physical goods

What if seller is restricted to sell the same information to every buyer (e.g., due to regulation)? How will revenue change?

- This is the optimal price (Merson reserve) in previous example
- Revenue can be arbitrarily worse
- $\frac{1}{e}$ - approximation of optimal revenue if the value of full information as a function of t has monotone hazard rate

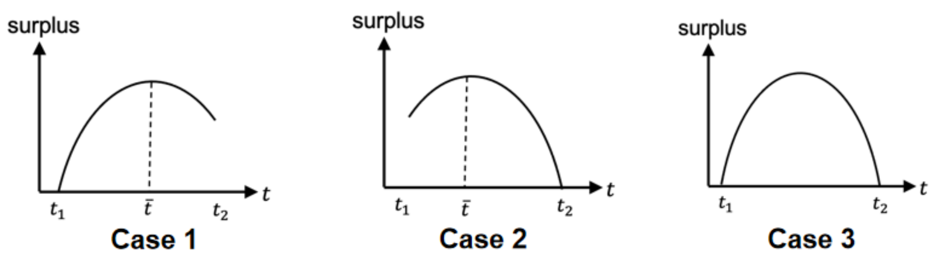
9.Additional Properties of Optimal Mechanism

Proposition 1([LSX'21]):Buyer surplus is increasing for $t \in [t_1, \bar{t}]$ and decreasing for $t \in [\bar{t}, t_2]$ where $\bar{t}$ satisfies $\bar{v}(\bar{t}) = 0$.



Prop.2([LSX'21]):Following properties hold in optimal mechanism.

1. In Case 1, surplus of t_1 is 0; In Case 2, surplus of t_2 is 0; In Case 3, surplus of both t_1 and t_2 is 0
2. Buyer payment is increasing in Case 1, decreasing in Case 2, and increase first then decrease in Case 3.



10.A Generalized Model of Selling Information

- Buyers takes one of n action $i \in [n] = \{1, ..., n\}$
- Buyers has an arbitrary utility function $u(i, \omega; t)$

Mechanism design question: How can seller optimally sell her information about ω to the buyer?

11.Existence of Simple "Direct Mechanisms"

Theorem (Revelation Principle, BBS'18, CXZ'20): Any information selling mechanism is "equivalent" to a direct and truthful mechanism:

1. Ask buyer to report type t
2. Charge buyer x_t and then directly make obedient action

Recommendation to buyer via a randomized scheme $\pi_t : Q \to [n]$. Moreover, the mechanism is incentive compatible(IC) — it is the buyer's best interest to truthfully report t.

12.This Optimal Mechanism is like Consulting

Consulting Mechanism w/ Bounded Payment [CXZ'20]

1. Elicit buyer type t
2. Charge buyer $x_t \leq B$ (bounded payment)
3. Observe realized state θ and recommend action i to the buyer with probability $\pi_t(\sigma_i, q)$

Theorem(CXZ'20): The optimal payment-limited consulting mechanism can be computed by a convex program.

III.Summary and Open Problems

1.Summary

- Raw and distilled data (i.e., information) both have economic values
- The pricing of information depends on its economic value
- Many recent progresses on pricing mechanisms for information/data
- But long way to go….

2. Open Direction I

- What if signals have error (e.g., predictions of ML algorithms)?
- What if the world is non-Bayesian? Difference between pricing signals vs pricing signal generation processes?
- How to be robust to uncertainty in information and ML models?

3. Open Direction II

Orchestrated information – acquiring information from multiple sources and selling aggregated information

4.Open Directions III

- From pricing information to pricing intelligence
- What is the most practical/efficient/feasible way to price intelligence? Directly sell raw data, or sell ML model, or sell inferences? Or personalized?