Selling Experiments

Dirk Bergemann[1] Alessandro Bonatti[2] Alex Smolin[1]

[1]Yale University

[2]MIT Sloan

Sloan Marketing
December 2015

Introduction

Data buyer - a decision maker under uncertainty:

- has partial and private information
- can acquire additional information

Data seller offers additional information:

- how much information to provide and at what price?
- how to provide different information to different data buyers?

Interpretation: selling access to a database as in Acxiom, Bluekai, DoubleClick Ad Exchange

Example: Behavioral Retargeting

- firms tailor online advertising levels to individual users
- targeting requires information about characteristics of individual users.
- different firms have different “first-party” information on users $\Rightarrow$ heterogeneous valuation for additional information
- data seller has information "third party" on individual characteristics
- data seller can offer to reveal certain attributes

Information via Experiments

- data seller offers "information product":
- $=$ experiment (in the statistical sense of Blackwell)
- provide statistical information about payoff-relevant state
- value of experiment depends on
  decision maker's private information - his beliefs
- decision maker's private beliefs are his type
- data seller has (common) prior over types

Analysis

- optimal *versioning* of information product:
  design of information
- optimal *selling* of information product:
  price of information
- importantly: *only information product itself is contractible*
- by contrast, action of decision maker or realized state are not contractible

Results

- a menu of experiments is offered
- menu contains only "simple" items, experiments
- menu is coarser than diversity of data buyers (types)
- linearity (in probabilities) limits the use of versioning
- systematic distortions in information provided
- screening facilitated by "directional information"

Related Literature

Selling Information

Admati and Pfleiderer (1986, 1990), Eső and Szentes (2007), Babaioff (2012), "Selling Cookies" *AEJ Micro* (2015).

Information Impacts Prices

Johnson and Myatt (2006), Bergemann and Pesendorfer (2007), "Targeting in Advertising Markets" *RAND* (2011).

Persuasion

Rayo and Segal (2010), Kamenica and Gentzkow (2011),

Model

- single decision-maker (buyer of information)
- finite actions
$$a_1, ..., a_I \in A$$
- finite states
$$\omega_1, ..., \omega_J \in \Omega$$
- ex-post utility
$$u\left(a_i, \omega_j\right)$$
- leading example, matching action to state $|I| = |J|$:
$$u\left(a_i, \omega_j\right) = \mathbf{1}_{[i=j]}.$$

Common Prior and Private Information

- common prior probability over states

$$\mu \in \Delta\Omega$$

- decision maker privately observes an initial signal $r \in R$:

$$\lambda : \Omega \to \Delta R$$

- decision maker forms initial belief $\theta \in \Delta\Omega$ given signal r :

$$\theta_r(\omega) = \frac{\lambda(r\,|\omega)\,\mu(\omega)}{\sum_{\omega'}\lambda(r\,|\omega')\,\mu(\omega')}$$

- initial beliefs θ are private information of data buyer

- from data seller's point of view: λ induces distribution of initial beliefs $F(\theta)$.

Experiment

- data seller provides information as "experiment"
- an experiment (information structure) $I = \{S, \pi\}$ consists of signals $s \in S$ and likelihood function:

$$\pi : \Omega \rightarrow \triangle S$$

- type r and signal s independent – conditional on state ω:

$$\Pr\left((r,s)\left|\omega\right.\right) = \lambda\left(r\left|\omega\right.\right) \cdot \pi\left(s\left|\omega\right.\right), \quad \forall r, s, \omega$$

- costless provision of information (data is already stored);

Data Seller

- data seller can offer a menu of experiments

$$\mathcal{M} = \{\mathcal{I}, t\},$$

where each item on menu $\mathcal{I}$ is an experiment I:

$$\mathcal{I} = \{I\} \qquad t : \mathcal{I} \to \mathbb{R}^{+}$$

- each experiment I has a price t
- note: action a and state ω are not contractible
- thus: scoring rules and other belief elicitations schemes are not available
- price of information is determined before realization of ω

Timing of Information

1. type θ of decision maker is realized
2. seller offers menu of experiments $\mathcal{I}$
3. decision maker θ chooses among experiments $\mathcal{I}$
4. signal s of experiment is realized, action a is taken

Interpretation: Big Data

- a continuum of consumers: $i \in [0, 1]$,
- comsumer i spends $\omega \in \mathbb{R}_+$ per website (budget ω)
- distribution of budgets $\mu \in \Delta(\Omega)$ in population
- type θ of retailer is distribution of consumer budgets at its website
- distribution of consumers with budget ω over retailer θ: $\lambda(\cdot|\omega)$
- think $\theta =$ Walmart, JC Penney, Sears, Macy

Interpretation: Data Base and Demand for Data

- data seller (data base) has record of past digital purchases of i, thus knows of budget ω of i
- database can offer estimate, narrower or wider income brackets for every i and ω
- at random times consumer i with budget ω has change of taste
  i.e. new/renewal draw according to $\lambda\left(\cdot\left|\omega\right.\right)$
- when i appears for the first time at retailer θ website, retailer might wish to acquire more information about ω of i
- query or "machine" interpretation: for every i generate an estimate of ω
  - $\pi\left(s\left|\omega, i\right.\right)$ is independent of i conditionally on ω
  - $\pi\left(s\left|\omega\right.\right)$ is independent of r conditionally on ω

Value of Experiment

- buyer's payoff under partial information

$$u(\theta) \triangleq \max_{a \in A} \mathbb{E}_\theta \left[u(a, \omega) \right].$$

- value of experiment (net value of augmented information)

$$V(I, \theta) \triangleq \mathbb{E}_{I,\theta}[\max_{a \in A} \mathbb{E}_{s,\theta} \left[u(a, \omega) \right]] - u(\theta).$$

Initial and Incremental Information

- interim probability

$$\theta_i = \Pr(\omega = \omega_i)$$

- likelihood function under experiment I ...:

$$\pi_{ij} = \Pr(s_j \mid \omega_i)$$

- ... and in matrix form π :

$$\begin{array}{cccc} & & s_j & \\ & \pi_{11} & \pi_{1j} & \cdots \\ & & & \\ \omega_i & \pi_{i1} & \pi_{ij} & \cdots \end{array}$$

Specific Experiments

- locally noise free (at s_j):

$$\begin{array}{ccccc} & & & s_j & \\ & \pi_{11} & & 0 & \cdots \\ & & & 0 & \\ \omega_i & \pi_{i1} & & \pi_{ij} & \cdots \\ & & & 0 & \end{array}$$

- locally non-dispersed (at ω_i)

$$\begin{array}{ccccc} & & & s_j & \\ & \pi_{11} & & \pi_{1_j} & \cdots \\ & & & & \\ \omega_i & 0 & 0 & 1 & 0 \end{array}$$

- perfectly informative

$$\pi_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

is noise free and non-dispersed, globally

Value of Experiment

- given matching action and state:

$$u(a_i, \omega_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

- value of experiment I for buyer θ :

$$V(I, \theta) = \sum_j \max_i \{\theta_i \pi_{ij}\} - \max_i \{\theta_i\}$$

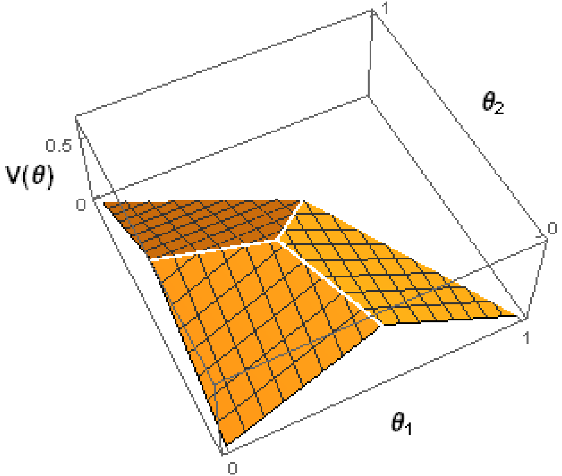
- posterior belief: interim belief θ_i and signal s_j:

$$\theta_i \pi_{ij}$$

- experiment provides a random allocation, $s_1, ..., s_J$ to an agent with unit demand $\max_i \{\theta_i \pi_{ij}\}$

Geometry of Value of Experiment

- three states $\omega_1, \omega_2, \omega_3$
- perfect information experiment
- interim belief $\theta = (\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$
- every edge represents a change in decision given interim belief



Seller's Problem

- seller offers a menu of experiments

$$\mathcal{M} = \{\mathcal{I}, t\}$$

with

$$\mathcal{I} = \{I\} \qquad t : \mathcal{I} \to \mathbb{R}^{+}.$$

- direct mechanism

$$\mathcal{M} = \{I(\theta), t(\theta)\}.$$

- seller's objective function is subject to incentive and participation constraints:

$$\max_{\{I(\theta), t(\theta)\}} \int t(\theta)\, \mathsf{d}F(\theta),$$

$$\text{s.t.} \quad V(I(\theta), \theta) - t(\theta) \geq V\left(I\left(\theta'\right), \theta\right) - t\left(\theta'\right) \quad \forall\, \theta, \theta',$$

$$V(I(\theta), \theta) - t(\theta) \geq 0 \quad \forall \theta.$$

First Steps

- possible continuum of experiments $I(\theta)$
- each experiment has a potentially complicated map:

$$\text{states} \rightarrow \text{signals} \rightarrow \text{actions}$$

- merge signals in $I(\theta)$ leading to the same action for type θ

Proposition (Maximal Cardinality of Signals)

In an optimal menu, the cardinality of the signal space of every experiment has at most the cardinality of the action space.

- $V(I(\theta), \theta)$ stays constant but $V(I(\theta), \theta')$ decreases $\forall \theta' \neq \theta$ as value of misreport is reduced

An Illustration: Binary States

- binary state, binary action:

$u(a,\omega)$	$a = a_H$	$a = a_L$
$\omega = \omega_H$	$\mathbf{1}$	$\mathbf{0}$
$\omega = \omega_L$	$\mathbf{0}$	$\mathbf{1}$

- let $\theta = \Pr(\omega = \omega_H)$
- by Proposition 1 restrict attention to experiments:

$$I = \begin{array}{c|cc} & s_H & s_L \\ \hline \omega_H & \alpha & 1-\alpha \\ \omega_L & 1-\beta & \beta \end{array}$$

- wlog convention that $\alpha + \beta \geq 1$ (equivalent to monotone likelihood ratio)

Value of Experiment with Binary Model

- value of experiment (α, β)

$$V(\alpha, \beta, \theta) = [\alpha\theta + \beta(1-\theta) - \max\{\theta, 1-\theta\}]^+ .$$

- *locally non-dispersed* at $\omega = \omega_L$, *locally noise free* at s_H:

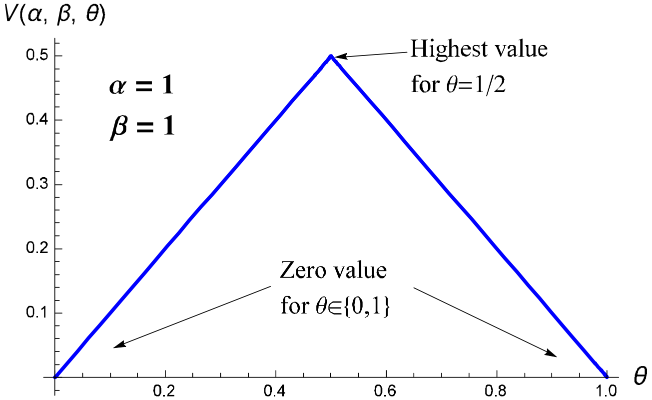
$I =$

	s_H	s_L
ω_H	α	$1-\alpha$
ω_L	0	1

- *directionally informative:* information valuable for some types, but not for others
  - valuable for DM who deems ω_L very likely
  - not valuable for DM who deems ω_H very likely
- directionally informative for null hypothesis of ω_L:
  - minimize false positive (type 1 error) to zero for ω_L,
  - maximize false negative (type 2 error) for ω_H

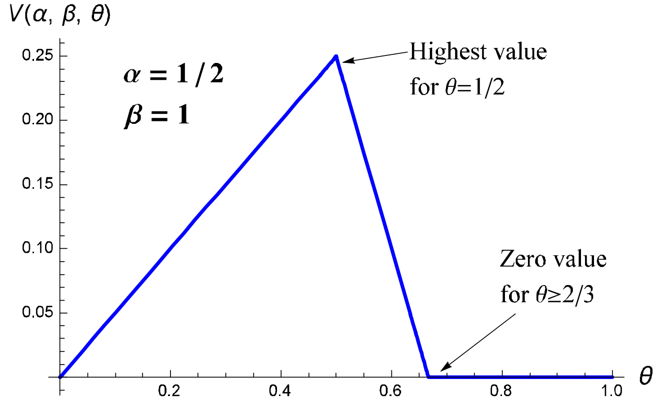
Value of a Perfectly Informative Experiment

- value of experiment $(\alpha, \beta) = (1, 1)$ for type θ.



- highest type is in the interior rather than on the boundary
- more than local incentive constraints, more than local participation constraints

Value of a Directionally Informative Experiment



- distance $|\theta - 1/2|$ not sufficient for value of experiment
- different slopes - differential gains of avoiding type 1 errors
- information has horizontal and vertical dimension of differentiation, information is always high-dimensional
- high degree of incompleteness in ranking of information structures

Preferences over Experiments

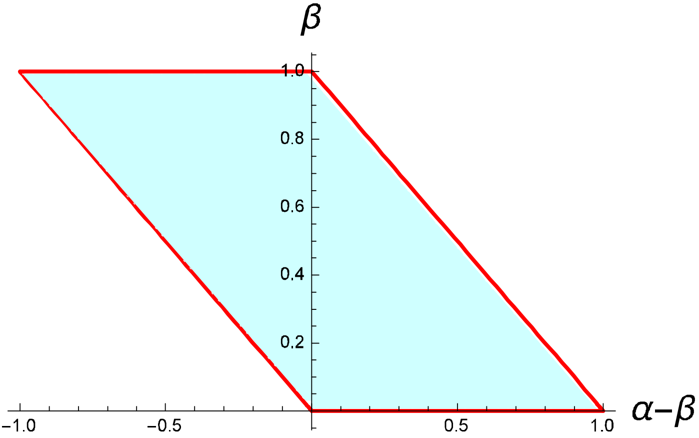
Value of experiment (α, β) for type θ

$$V(\alpha, \beta, \theta) = (\alpha - \beta)\,\theta + \beta - \max\{\theta, 1 - \theta\}.$$

- $\beta =$ *baseline* informativeness (from payoff normalization).

- $\alpha - \beta =$ *relative* informativeness.
- two "goods" that cannot be produced independently.

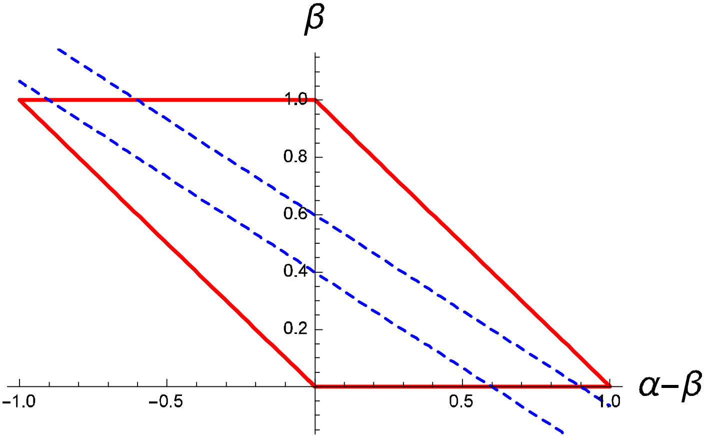
Feasible Set of Experiments



Indifference Curves for Given Type

- value of experiment is

$$V(\alpha, \beta, \theta) = (\alpha - \beta)\theta + \beta - \max\{\theta, 1 - \theta\}$$



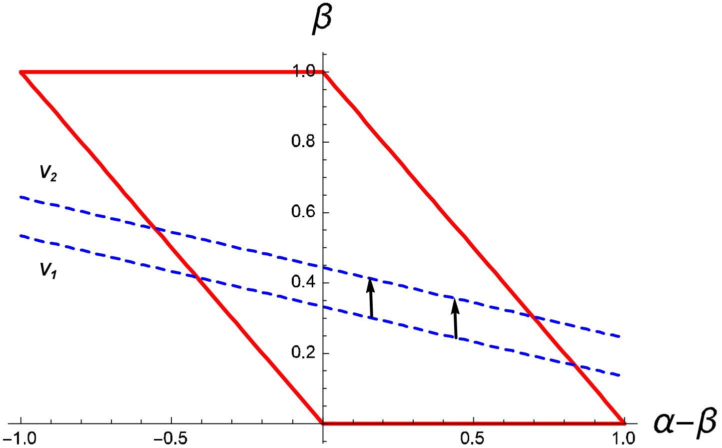
- higher θ have stronger preference for differential $\alpha - \beta$

Value of Baseline Information

- incremental change in the baseline information β
- while keeping the relative informativeness $\alpha - \beta$ constant

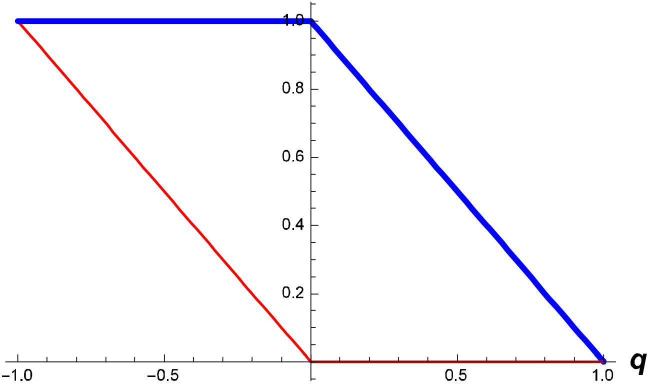
$$V(\alpha + \delta, \beta + \delta, \theta) - V(\alpha, \beta, \theta) = \delta, \quad \forall \theta$$

- uniform increase in value of experiment for all types



Set of Optimal Experiments

- maximal baseline informativeness for any given relative informativeness



- reduce choice of experiment to one-dimensional problem:

$$q \triangleq \alpha - \beta$$

Structure of Optimal Menu

- for finitely many states and actions, possibly continuum of types

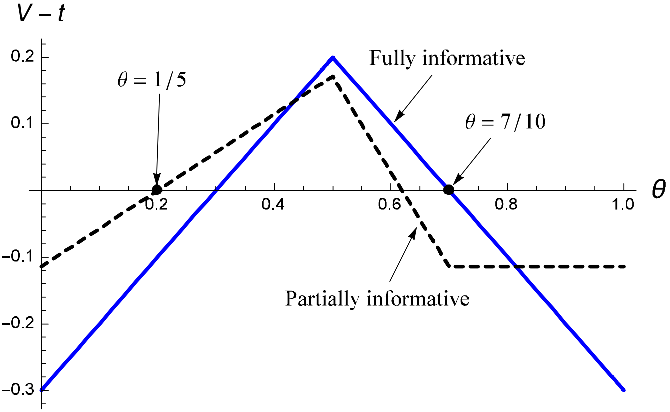
Proposition (Optimal Menu and Non-Dispersed Information)

1. *The fully informative experiment,* $\pi_{ii} = 1$ *for all* i*, is always part of the optimal menu.*
2. *Every experiment in an optimal menu is locally non-dispersed, i.e.,* $\pi_{ii} = 1$ *for some* i*.*

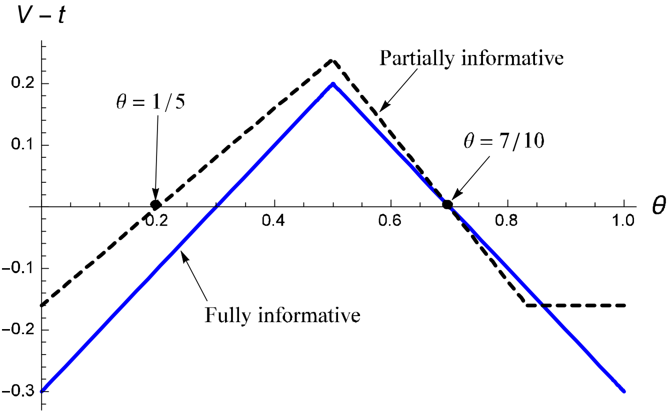
states/types	*binary*	*continuum*
binary	$\cdots$	$\cdots$
finite	$\cdots$	✓

Binary Types and Binary States: First Example

- binary types: $\theta \in \{2/10, 7/10\}$ with equal probability
- type $\theta = 7/10$ is less informed
- a possible solution (with slack incentive constraints)



Binary Types: An Optimal Solution



- two experiments
  - no distortion at the top (θ closer to $1/2$);
  - no rent at the bottom;
  - corner solution – no rent at the top

Binary Types and Binary States

- two types are congruent if they choose the same action given their interim belief, otherwise non-congruent
- high type is less informed than low type:

$$\left|\theta^H - 1/2\right| \leq \left|\theta^L - 1/2\right|, \quad \theta^H \geq 1/2$$

- recall prior probability of high type

$$\gamma = \Pr\left(\theta = \theta^H\right)$$

- critical frequency of high vs low types:

$$\bar{\gamma} \triangleq \frac{1-\theta^L}{1-\theta^H}$$

Optimal Experiment

Proposition

1. *With congruent priors, the seller offers the perfectly informative experiment only; both types participate if and only if* $\gamma \leq \bar{\gamma} = \left(1 - \theta^L\right) / \left(1 - \theta^H\right)$.
2. *With non-congruent priors and* $\gamma \leq \bar{\gamma}$*, both types buy the fully informative experiment.*
3. *With non-congruent priors and* $\gamma > \bar{\gamma}$*, the high type buys the fully informative experiment and the low type buys a partially informative experiment:*

$$\alpha = \frac{2\theta^H - 1}{\theta^H - \theta^L} \qquad \text{and} \qquad \beta = 1;$$

and the seller extracts the entire surplus.

- *quality of information and comparative statics of* $1 - \alpha$ *in* $\gamma, \theta^L, \theta^H$

Many States and Many Actions

- we maintain binary states

$$\theta \in \{\theta_L, \theta_H\}$$

  and allow for many states (and many actions)
- order the states ω by their likelihood ratios:

$$\frac{\theta_1^L}{\theta_1^H} \leq \cdots \leq \frac{\theta_i^L}{\theta_i^H} \leq \cdots \leq \frac{\theta_N^L}{\theta_N^H}$$

- states ω_i with low indices are deemed more likely by θ^H

Optimal Experiment

- use disagreement across states to drive screening across types

Proposition

There exists i^* such that the optimal experiment $I(\theta^L)$ has $\pi_{ii} = 0$ for all $i < i^*$ and $\pi_{ii} = 1$ for all $i > i^*$.

- optimal experiment has lower-triangular shape

$$
\begin{matrix}
0 & \cdots & 0 & \pi_{1i} & \cdots & \cdots & \pi_{1n} \\
\vdots & & \vdots & \vdots & & & \vdots \\
\vdots & & \vdots & \pi_{ii} & \cdots & \cdots & \pi_{in} \\
\vdots & & \vdots & 0 & 1 & \cdots & 0 \\
\vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 1
\end{matrix}
$$

- distribution of $\pi_{i\cdot}$ is not uniquely determined

Continuum of Types

- return to binary states, allow continuum of types $\theta \in [0, 1]$
- recall the value of experiment $q \in [-1, 1]$ for type $\theta \in [0, 1]$:

$$V(q, \theta) = [\theta q - \max\{q, 0\} + \min\{\theta, 1 - \theta\}]^+ .$$

- single-crossing suggests q increasing in θ.
- types $\theta = 0$ and $\theta = 1$ receive zero rents.
- consider type $\theta = 1/2$, derive additional condition.

Incentive Compatibility

- rent of type $\theta = 1/2$

$$U(1/2) = U(0) + \int_0^{1/2} V_\theta(q,\theta)\mathrm{d}\theta = U(1) - \int_{1/2}^{1} V_\theta(q,\theta)\mathrm{d}\theta.$$

- define an allocation $q(\cdot)$ to be *responsive* if, for any θ

$$\theta q(\theta) - \max\{q(\theta), 0\} + \min\{\theta, 1-\theta\} \leq 0$$

$$\Rightarrow q(\theta) = \begin{cases} -1 & \text{if} \quad \theta < 1/2. \\ +1 & \text{if} \quad \theta \geq 1/2. \end{cases}$$

- if net utility of experiment is negative for θ, then assign zero information experiment

Incentive Compatibility

- rent of type $\theta = 1/2$

$$U(1/2) = \int_0^{1/2} (q(\theta) + 1)\mathrm{d}\theta = -\int_{1/2}^{1} (q(\theta) - 1)\mathrm{d}\theta.$$

Proposition (Necessary Conditions)

If the allocation $q(\theta)$ is implementable and responsive then

$$q(\theta) \in [-1, 1] \text{ is non-decreasing,}$$

and

$$\int_0^1 q(\theta)\, d\theta = 0.$$

- note: a different kind of constraint, a global constraint

Seller's Problem

$$\max_{q(\theta)} \int_0^1 \left[\left(\theta - \frac{1 - F(\theta)}{f(\theta)} \right) q(\theta) - \max\{q(\theta), 0\} \right] f(\theta)\, \mathsf{d}\theta,$$

$$\text{s.t. } q(\theta) \in [-1, 1] \text{ non-decreasing,}$$

$$\int_0^1 q(\theta)\, \mathsf{d}\theta = 0.$$

Piecewise linear (concave) problem with integral constraint.

Absent the integral constraint, corner solutions:

- $q^* \in \{-1, 0, 1\}$, i.e., all-or-nothing information, flat price.
- E.g., truncated support or symmetric distribution.

Optimal Menu

Proposition (Optimal Menu)

An optimal menu consists of at most two experiments.

1. *The first experiment is fully informative.*
2. *The second experiment is locally non-dispersed and locally noisefree.*

- coarse menu
- a continuum of types - yet only a binary choice is provided

Properties of the Optimal Menu

Optimal mechanism involves ≤ 2 bunching intervals.

Ideally, would sell $q = 0$ at two different prices (for $\theta \lesseqgtr 1/2$).

- Symmetric distribution or truncated support $\rightarrow$ flat pricing.
- Second-best menu may contain $q = 0$ only...
- ...or distort the "least profitable side."
- No further *versioning* is optimal.

Least informed types $\neq$ most valuable to the seller.

Type $\theta = 1/2$ need not get efficient $q = 0$.

Conclusions: Selling Information

- selling incremental information to privately informed buyers.
- costless acquisition & transmission, free degrading
- "uninterested seller" – packaging problem
- bayesian problem for buyers
- linear in probabilities: limited use of versioning
- screening *across agents* through directional information

Seller's Problem

$$\max_{q(\theta)} \int_0^1 \left[\left(\theta - \frac{1 - F(\theta)}{f(\theta)} \right) q(\theta) - \max\{q(\theta), 0\} \right] f(\theta)\, \mathsf{d}\theta,$$

$$\text{s.t. } q(\theta) \in [-1, 1] \text{ non-decreasing,}$$

$$\int_0^1 q(\theta)\, \mathsf{d}\theta = 0.$$

Seller's Problem

$$\max_{q(\theta)} \int_0^1 \left[\left(\theta f(\theta) + F(\theta) \right) q(\theta) - \max\left\{ q(\theta), 0 \right\} f(\theta) \right] \mathrm{d}\theta,$$

$$\text{s.t. } q(\theta) \in [-1, 1] \text{ non-decreasing},$$

$$\int_0^1 q(\theta)\, \mathrm{d}\theta = 0.$$

Consider "virtual values" for each experiment q separately:

$$\phi(\theta, q) \triangleq \begin{cases} \theta f(\theta) + F(\theta) & \text{for} \quad q < 0, \\ (\theta - 1) f(\theta) + F(\theta) & \text{for} \quad q > 0. \end{cases}$$

- $\phi =$ marginal value of going from $q(\theta) = -1$ to 0 to 1.
- Problem is not separable: virtual value ϕ depends on q.

General Case

- Let λ denote the multiplier on the integral constraint (shadow cost of providing higher "quantity").
- Let $\bar{\phi}(\theta, q)$ denote the *ironed* virtual value for experiment q.

Proposition (Optimal Allocation Rule)

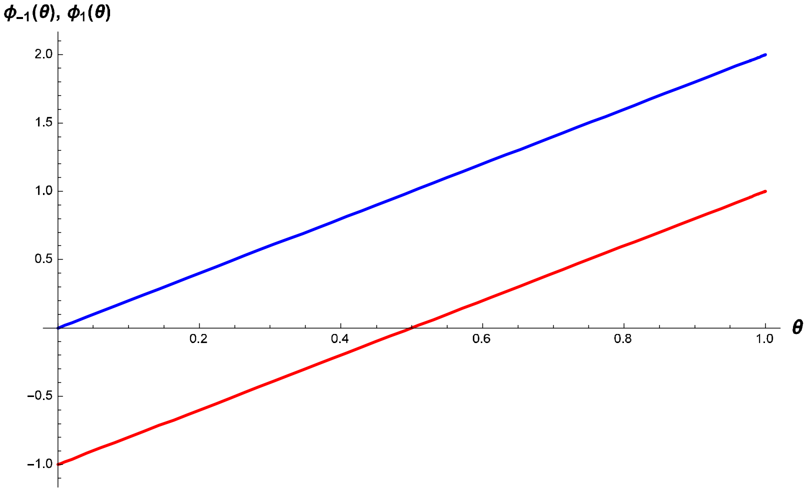
Allocation $q^*(\theta)$ *is optimal if and only if there exists* $\lambda^* \geq 0$ *s.t.*

$$q^*(\theta) \in \arg\max_{q \in [-1,1]} \left[\int_0^q \left(\bar{\phi}(\theta, x) - \lambda^* \right) dx \right] \text{ for all } \theta,$$

has the "pooling property," and satisfies the integral constraint.

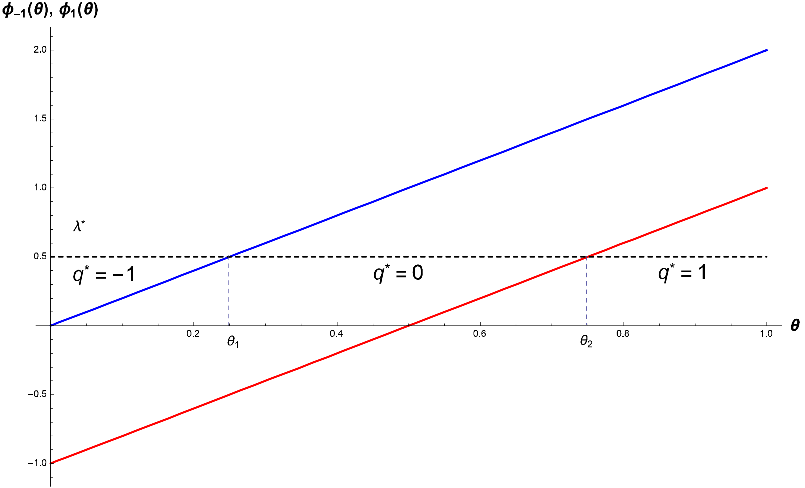
- Myerson (1981), Toikka (2011), Luenberger (1969).

Example 1: Uniform Distribution



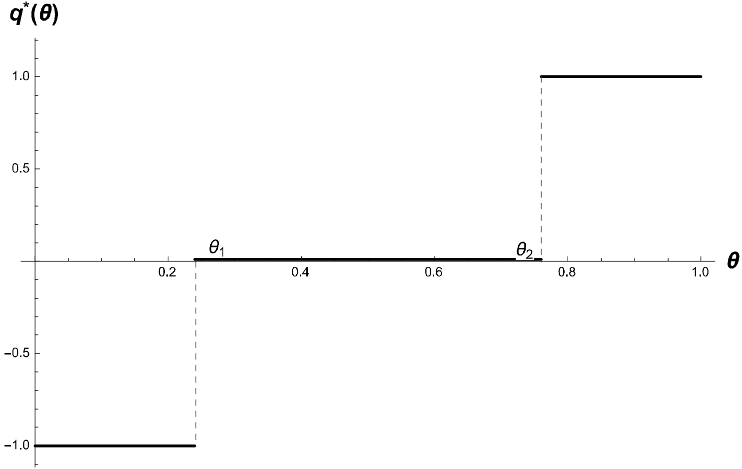
Virtual Values: $\phi(\theta, -1)$ in blue; $\phi(\theta, 1)$ in red.

Example 1: Uniform Distribution



Virtual Values: $\phi(\theta, -1)$ in blue; $\phi(\theta, 1)$ in red.

Example 1: Uniform Distribution



Optimal Menu: Single Experiment

Optimality of Flat Pricing

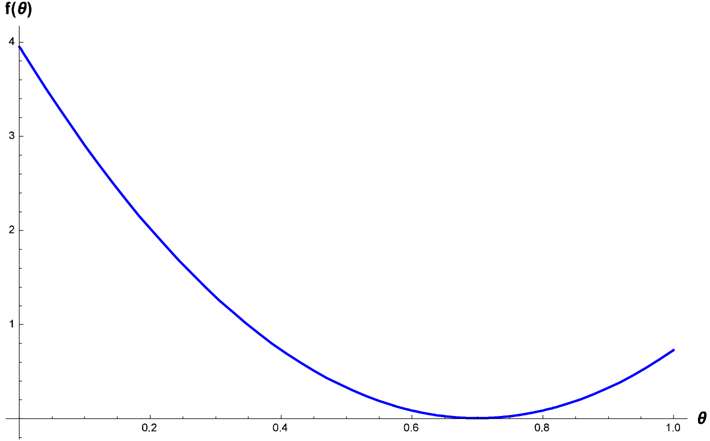
Proposition (Flat Pricing)

The optimal menu contains only the fully informative experiment when any of the following conditions hold:

1. *the density* $f(\theta) = 0$ *for all* $\theta > 1/2$ *or* $\theta < 1/2$*;*
2. *the density* $f(\theta)$ *is symmetric around* $\theta = 1/2$*.*
3. $F(\theta) + \theta f(\theta)$ **and** $F(\theta) + (\theta - 1) f(\theta)$ *are strictly increasing.*

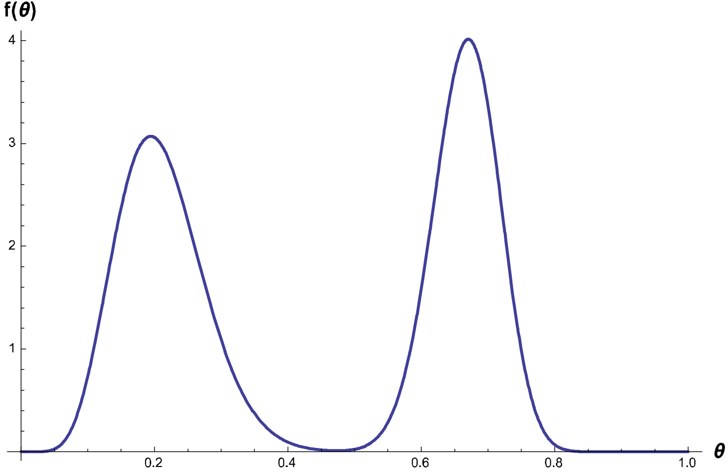
A second experiment is offered only if ironing is required.

Non-monotone Density



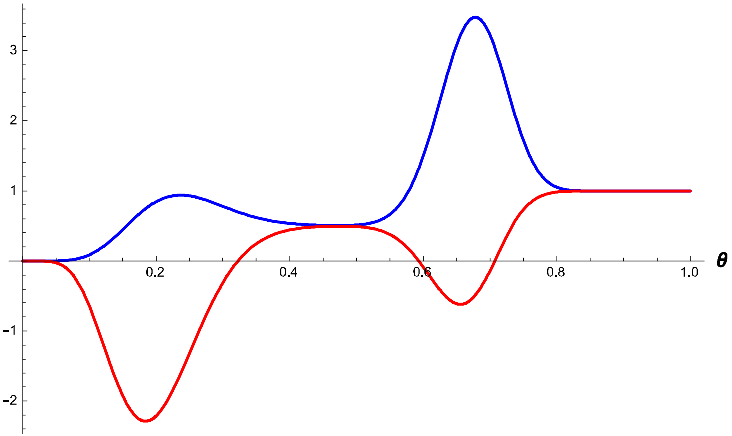
Probability Density Function: informed types are frequent

Example 2: Combination of Beta Distributions



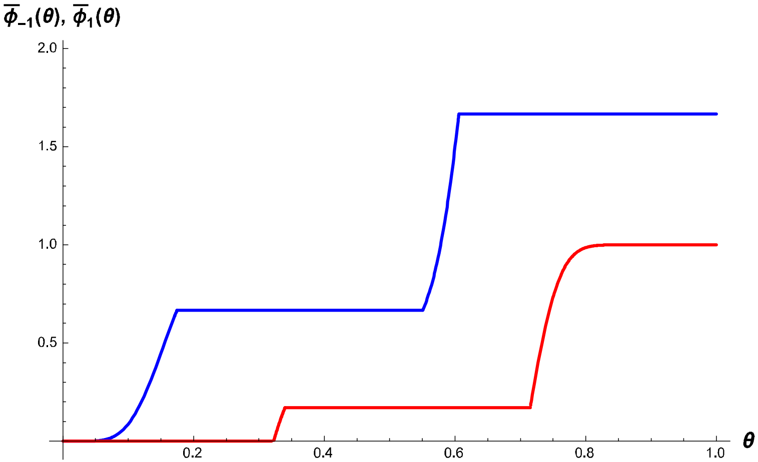
Probability Density Function

Example 2: Beta Distributions



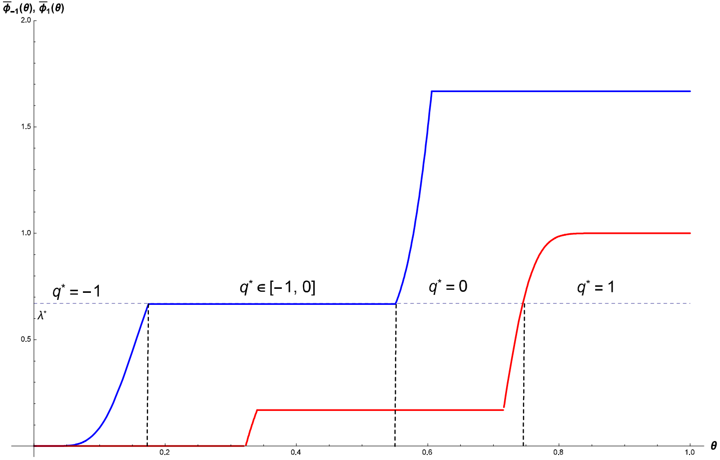
Virtual Values: $\phi(\theta, -1)$ in blue; $\phi(\theta, 1)$ in red.

Example 2: Beta Distributions



Ironed Virtual Values

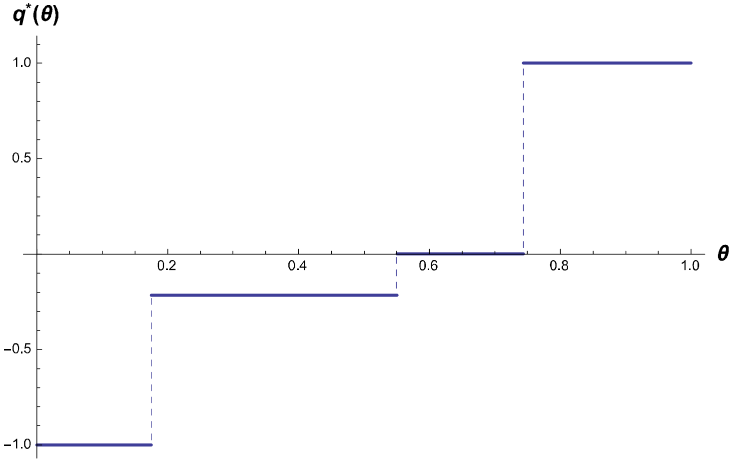
Example 2: Beta Distributions



Ironed Virtual Values

Example 2: Beta Distributions

One More Example



Optimal Menu: Two Experiments ($q \in \{-.21, 0\}$)

need not get efficient $q = 0$.

Implications for Observables

- how to damage an information good
- should not observe arbitrarily damaged goods
- directional information: only type-I or II errors
- should not observe multiple distortions of the same kind
- directional distortions $\sim$ disclosure of specific attributes (correlated with high- or low- value consumers).