Elicitability

Y. Azrieli[1] C. Chambers[2] P.J. Healy[1] N. Lambert[3]

[1]Ohio State University

[2]Georgetown University

[3]MIT

Northwestern/Kellogg MEDS Department Seminar
October 13, 2021

Research Question

A researcher is tasked with producing a statistical analysis.

He/She works from data that cannot be verified.

There is a conflict of interest and the researcher can manipulate the analysis.

Suppose we collect some data from an independent source and we can design payments using, as input, the researcher's report and the data.

What is the information we can elicit from the researcher?

Model

Model

Statistical Experiments

- There is an unknown parameter θ (belongs to parameter space Θ).
- The researcher performs Bayesian inference on the parameter.
- The elicitor collects data through statistical experiments.
  An *experiment* is a family of probability distributions indexed by the parameter.
  Formally a pair (Y, π):
  - Y = set of possible outcomes.
  - π = transition probability kernel from parameters to outcomes.
    An outcome from Y is drawn at random according to $\pi(\cdot|\theta)$.

Assumption: Parameter and outcome spaces are Polish spaces with their Borel σ-algebra.

An experiment is *categorical* if the outcome space is finite (e.g., probit, softmax, ...).
An experiment is *identified* when the parameter can be inferred from the outcome distribution.

Model

Elicitation Mechanisms

A (direct) elicitation *mechanism* for experiment (Y, π) is a mapping

$$\varphi : \Delta(\Theta) \times Y \to \mathbb{R}$$

- Inputs: reported parameter distribution and outcome randomly generated by the experiment.
- Output: A payoff.

The mechanism is *incentive compatible* when $\mathrm{E}_p[\varphi(p, y)] \geq \mathrm{E}_p[\varphi(q, y)]$ (for all p, q).

In the sequel, 'mechanism' means direct incentive-compatible elicitation mechanism.

Model

Elicitability Criterion

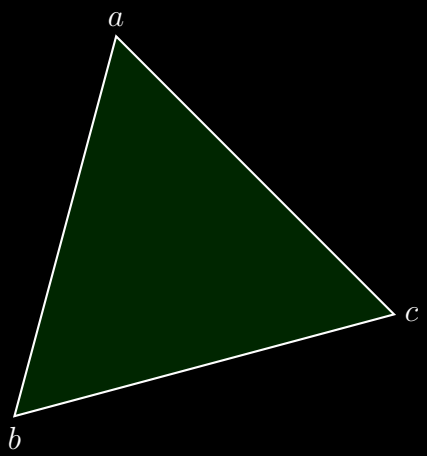
An *information partition* of distributions is a partition $\mathcal{P}$ of the space $\Delta(\Theta)$, defined by the equivalence relation $\sim_{\mathcal{P}}$, where $p \sim_{\mathcal{P}} q$ signifies that p and q belong to the same member.

Mechanism φ *elicits information* $\mathcal{P}$ when

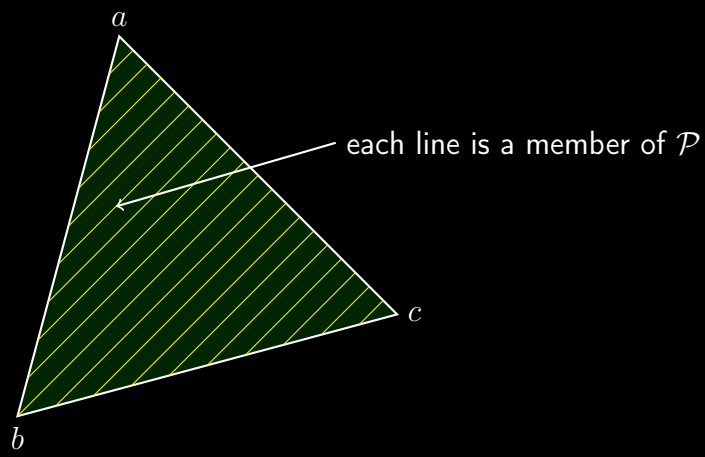
$$p \nsim_{\mathcal{P}} q \quad \Longrightarrow \quad \mathrm{E}_p[\varphi(p, y)] > \mathrm{E}_p[\varphi(q, y)].$$

$\mathcal{P}$ is *elicitable* with an experiment when there is a mechanism for this experiment that elicits $\mathcal{P}$.

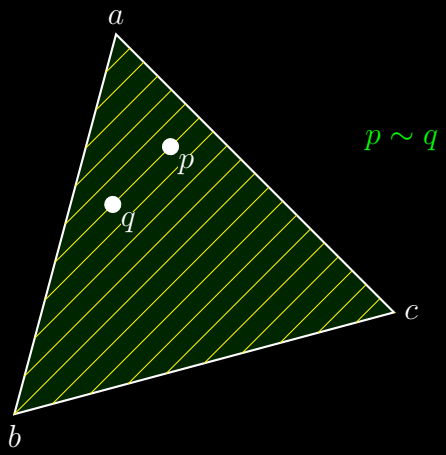
$$\Theta = \{a, b, c\}$$



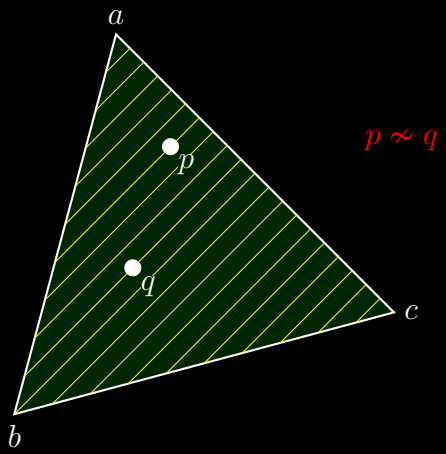
$$\Theta = \{a, b, c\}$$



$$\Theta = \{a, b, c\}$$



$$\Theta = \{a, b, c\}$$



$p \nsim q$

Example

Vaccine

We are interested in assessing the efficacy of a new vaccine.

A researcher is tasked with estimating the probability of getting sick after being vaccinated.

Statistical model is Bernoulli: $\Pr[y = 1] = p$ with $p \in [0, 1]$.

What's observable: sickness condition ($y \in \{0, 1\}$).

Example

Vaccine

We are interested in assessing the efficacy of a new vaccine.

A researcher is tasked with estimating the probability of getting sick after being vaccinated.

Statistical model is Bernoulli: $\Pr[y = 1] = p$ with $p \in [0, 1]$.

What's observable: sickness condition ($y \in \{0, 1\}$).

- $\rightsquigarrow$ We can elicit the mean of p with just one observation... but if we wanted to get the median, we would need infinite data!
- $\rightsquigarrow$ With two observations, we elicit whether researcher knows p, and if applicable, elicit p.
- $\rightsquigarrow$ With n observations we elicit an $O(1/n)$ mean-squared approximation of p's density.

Example

Cholesterol Medication

The researcher is tasked with evaluating the efficacy of a new cholesterol medication.

What's observable: the amount of cholesterol in the blood ($y \in \mathbb{R}$).

Let's assume cholesterol levels depend on age (x) through a linear model.

- Gaussian linear model: $y = \beta_0 + \beta_1 x + \sigma\varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 1)$.

Example

Cholesterol Medication

The researcher is tasked with evaluating the efficacy of a new cholesterol medication.

What's observable: the amount of cholesterol in the blood ($y \in \mathbb{R}$).

Let's assume cholesterol levels depend on age (x) through a linear model.

- Gaussian linear model: $y = \beta_0 + \beta_1 x + \sigma\varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 1)$.
  $\rightsquigarrow$ With a single observation we elicit the full parameter distribution.

Example

Cholesterol Medication

The researcher is tasked with evaluating the efficacy of a new cholesterol medication.

What's observable: the amount of cholesterol in the blood ($y \in \mathbb{R}$).

Let's assume cholesterol levels depend on age (x) through a linear model.

- Gaussian linear model: $y = \beta_0 + \beta_1 x + \sigma\varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 1)$.
  $\rightsquigarrow$ With a single observation we elicit the full parameter distribution.
- Non-parametric linear model: $\mathrm{E}[y|x] = \beta_0 + \beta_1 x$.
  $\rightsquigarrow$ With two observations we elicit the mean and variance of the parameters.

Example

Cholesterol Medication

The researcher is tasked with evaluating the efficacy of a new cholesterol medication.

What's observable: the amount of cholesterol in the blood ($y \in \mathbb{R}$).

Let's assume cholesterol levels depend on age (x) through a linear model.

- Gaussian linear model: $y = \beta_0 + \beta_1 x + \sigma\varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 1)$.
  $\rightsquigarrow$ With a single observation we elicit the full parameter distribution.
- Non-parametric linear model: $\mathrm{E}[y|x] = \beta_0 + \beta_1 x$.
  $\rightsquigarrow$ With two observations we elicit the mean and variance of the parameters.

Suppose instead $y \in \{0, 1\}$ and the model is probit: $y = \begin{cases} 1 & \text{if } \beta_0 + \beta_1 x + \varepsilon > 0 \\ 0 & \text{otherwise.} \end{cases}$

$\rightsquigarrow$ We need infinite data to elicit the means of β_0, β_1.

Outline and Contribution

(1) Describe what we can elicit for a given model and a given dataset.

(2) Compare statistical experiments based on their elicitation power.
How is elicitation different from estimation?
How does this difference impact the design of experiments?

(3) Investigate relation elicitation power $\iff$ incentive structures.

Related Literature(s)

Elicitation of preferences, probabilities, and statistical functionals:

- Preferences/Probabilities: Allais (1953), Savage (1971), ...

Elicitation of types/beliefs from multiple agents:

- Mechanism design: Crémer-McLean (1988), ...
- Peer prediction: Prelec (2004), Miller et al. (2005), ...

Statistical Theory:

- Testing forecasters: Olszewski-Sandroni (2008), Al-Najjar et al. (2010), ...
- Identification: Teicher (1961), ...
- Comparison of experiments: Blackwell (1953), ...

Information Elicited

Case of a Single Observation

Case of a Single Observation

Consider an experiment (Y, π) with random outcome y.
Every parameter θ induces a distribution over outcomes $\pi(\cdot|\theta)$.

Let $\mathcal{P}^\star$ be the information that captures the mean outcome distribution:

$$p \sim_{\mathcal{P}^\star} q \qquad \Leftrightarrow \qquad \mathrm{E}_p[\pi(A|\theta)] = \mathrm{E}_q[\pi(A|\theta)] \quad \forall A \subset Y.$$

Theorem

$\mathcal{P}$ is elicitable with (Y, π) if and only if $\mathcal{P}$ is a (weak) coarsening of $\mathcal{P}^\star$. ▸ proof

$\mathcal{P}^\star$ is the *most refined information* that we can elicit with one observation from (Y, π).

Corollary

Let $g : \Theta \to \mathbb{R}$.
If there exists an unbiased estimator for g, then the mean of g is elicitable.

Proof: If Y is finite then

$$\begin{aligned} g(\theta) &= \mathrm{E}[w(y) \mid \theta] \\ &= \sum_y w(y)\pi(y|\theta) \\ \mathrm{E}_p[g(\theta)] &= \sum_y w(y)\ \mathrm{E}_p[\pi(y|\theta)] \end{aligned}$$

$\implies$ elicitation of mean probabilities implies elicitation of the mean of g.

The converse is true if the experiment is categorical.
Example: the mean parameters of the probit model cannot be elicited with finite data.

Example: German Tank Problem

There is a finite population with an unknown number of units numbered consecutively starting from 1.

- θ is the population size.
- Consider the experiment where one unit is drawn and its number observed: $Y = \{1, 2, \dots\}$ and

$$\pi(k|\theta) = \begin{cases} 1/\theta & \text{if } k \leq \theta, \\ 0 & \text{if } k > \theta. \end{cases}$$

With this experiment we elicit the full distribution over population sizes.

Proof: For m a positive integer, let

$$w_m(k) = \begin{cases} 1 & \text{if } k \leq m, \\ -m & \text{if } k = m+1, \\ 0 & \text{if } k > m+1. \end{cases}$$

Apply corollary with:

$$g_m(\theta) = \sum_{k=1}^{\infty} w_m(k)\pi(k|\theta)$$

$\Longrightarrow$ We elicit the mean of every g_m.

But $\mathrm{E}[g_m(\theta)] = \Pr[\theta \leq m]$, so we elicit the c.d.f. of θ.

■

Example: Bernoulli Model

The researcher is tasked with assessing the fraction of the population on which a new vaccine is effective. The elicitor gets data from independent clinical trials.

- θ is the fraction of the population on which the vaccine is effective.
- Consider the experiment corresponding to a single trial:

$$Y = \{0, 1\} \qquad \text{and} \qquad \begin{aligned} \pi(1|\theta) &= \theta, \\ \pi(0|\theta) &= 1 - \theta. \end{aligned}$$

This experiment elicits the mean fraction of the population on which the vaccine is effective, and no further information.

▸ more examples

Information Elicited

Adding Data Points

Case of Multiple Observations

Consider experiments $(Y_1, \pi_1), \ldots, (Y_n, \pi_n)$ that respectively generate outcomes $y_1, \ldots, y_n$ independently conditionally on the parameter.

The *product experiment* is the compound experiment that corresponds to observing $(y_1, \ldots, y_n)$.

Theorem

If (Y_i, π_i) elicits the mean of $g_i : \Theta \to \mathbb{R}$, then under regularity conditions,[1] the product experiment elicits the mean of $g_1 \times \cdots \times g_n$.

[1] If either

(1) Θ is compact, each (Y_i, π_i) is continuous, each g_i is continuous,

(2) or each (Y_i, π_i) is categorical.

Applications

- Consider an experiment and $g : \Theta \to \mathbb{R}$. If the mean of g is elicitable with n observations, then the variance is elicitable with $2n$ observations, the skewness is elicitable with $3n$ observations, etc.

Applications

- Consider an experiment and $g : \Theta \to \mathbb{R}$. If the mean of g is elicitable with n observations, then the variance is elicitable with $2n$ observations, the skewness is elicitable with $3n$ observations, etc.
- Consider any categorical experiment that is identified. If the parameter space is infinite, then the full parameter distribution cannot be elicited with finite data. But if the parameter space is finite of size n, then we can always elicit the full parameter distribution with $n - 1$ observations.

Applications

- Consider an experiment and $g : \Theta \to \mathbb{R}$. If the mean of g is elicitable with n observations, then the variance is elicitable with $2n$ observations, the skewness is elicitable with $3n$ observations, etc.
- Consider any categorical experiment that is identified. If the parameter space is infinite, then the full parameter distribution cannot be elicited with finite data. But if the parameter space is finite of size n, then we can always elicit the full parameter distribution with $n - 1$ observations.
- With two observations from an identified experiment, we can elicit two pieces of information: Whether the researcher knows the parameter, and, if applicable, the value of the parameter.

skip next application | details

Applications

On the Elicitation of Statistical Functionals

Let the parameter space be finite.

Fix an arbitrary experiment, and a finite dataset.

- If the median is elicitable, then the full parameter distribution is also elicitable.
- If the mode is elicitable, then the full parameter distribution is also elicitable.

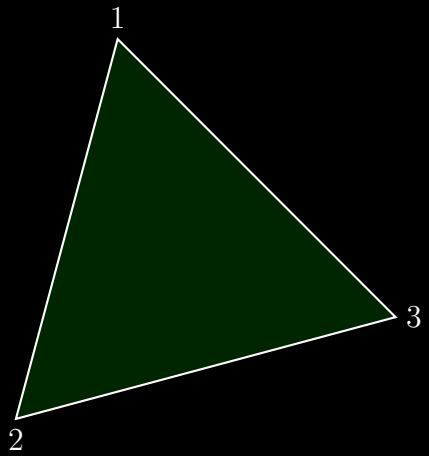
Assume $\Theta = \{\theta_1, \ldots, \theta_n\}$ and $Y = \{y_1, \ldots, y_m\}$.

Interpret $\pi(y|\theta)$ as an $n \times m$ Markov matrix, and $\Delta(\Theta) \subset \mathbb{R}^n$.

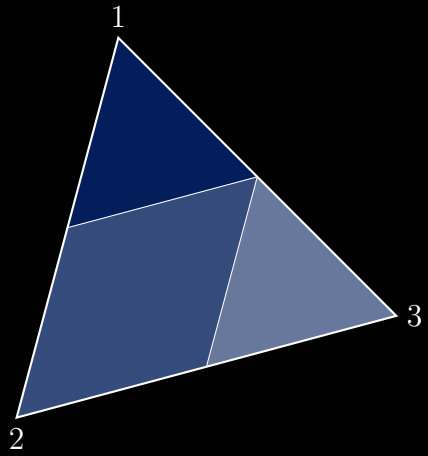
$$p \sim_{\mathcal{P}^\star} q \quad \Leftrightarrow \quad \mathrm{E}_p[\pi(y_i|\theta)] = \mathrm{E}_q[\pi(y_i|\theta)] \;\; (\forall i) \quad \Leftrightarrow \quad p\pi = q\pi.$$

$\Longrightarrow$ So $p \sim_{\mathcal{P}^\star} q$ iff $p = q + v$ where $v \in \ker \pi^\intercal$ (a linear space).

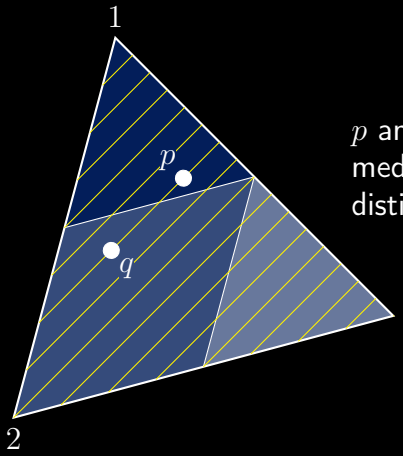
$$\Theta = \{1, 2, 3\}$$



Information partition for the median of θ



Information partition for the median of θ



p and q have different medians but cannot be distinguished

Applications

More Than Two Observations

Consider a categorical experiment (Y, π) with Y of size m.

If it is identified, we can identify Θ with the $(m-1)$-simplex of $\mathbb{R}^m$.

Suppose the researcher privately estimates a Lipschitz-continuous density function f over Θ.

Then with n observations, we can infer from the researcher's report a function $\widehat{f}$ that approximates f according to

$$\int_{\Theta} |f(\theta) - \widehat{f}(\theta)|^2 \, \mathrm{d}\theta = O(1/n).$$

Proof idea:

(1) Consider the Bernoulli model: $Y = \{0, 1\}$, $\Theta = [0, 1]$.

(2) Let $\mathcal{C}$ = collection of polynomials of degree $\leq n$.

(3) Let $Q_0, \ldots, Q_n$ be an orthonormal basis of $\mathcal{C}$: $\int_0^1 Q_i Q_j = \mathbb{1}\{i = j\}$.

(4) Since $\theta = \pi(y = 1|\theta)$, with n observations, we elicit the means of all $Q_k(\theta)$.

(5) Hence from the researcher's report we can infer the polynomial

$$\widehat{Q} = \sum_{k=0}^{n} \left(\int_\Theta Q_k(\theta) f(\theta) \, \mathrm{d}\theta \right) Q_k$$

$\widehat{Q}$ is the orthogonal projection of f on $\mathcal{C}$.

(6) Jackson's inequality: $\inf_{Q \in \mathcal{C}} \int_\Theta |f - Q|^2 \leq \gamma/n \quad \Longrightarrow \quad \int_\Theta |f - \widehat{Q}|^2 \leq \gamma/n = O(1/n)$

Applications

Optimal Sampling

Consider the Bernoulli model: $Y = \{0, 1\}$, $\Theta = [0, 1]$.

With n observations, the most refined elicitable information is $\mathcal{P}^\star$.

Suppose observations are costly. Can we elicit $\mathcal{P}^*$ by collecting fewer observations?

Yes: An 'optimal' sampling strategy consists in collecting observations until

- we observe outcome 1, or
- we already have collected n observations.

details

Applications

Optimal Sampling

Proof idea (sufficiency):

Any information we elicit with n observations is inferred from the mean of each $\theta^k(1-\theta)^{n-k}$.

Under the suggested sampling strategy:

- From the 1st observation, we elicit the mean of θ and $1-\theta$.
- From the 2nd observation, we elicit the mean of θ^2 and $\theta(1-\theta)$.
- ...
- From the k$^{\text{th}}$ observation, we elicit the mean of θ^k and $\theta^{k-1}(1-\theta)$.

The linear span of these polynomials is the set of polynomials of degree $\leq n$, so we can infer the mean of each $\theta^k(1-\theta)^{n-k} \implies$ we elicit as much information as with all the n observations.

Information Elicited

Adding Covariates

details

Comparison of Experiments

In the sequel, all experiments are categorical unless mentioned otherwise.

Blackwell Dominance

An experiment (Y, π_Y) *dominates* an experiment (Z, π_Z) *in the sense of Blackwell* if

$$z = h(y, \varepsilon)$$

with ε an independent random noise (equality is in distribution).

This is equivalent to the existence of an $Y \times Z$ Markov matrix M s.t.

$$\pi_Z(z|\theta) = \sum_{y \in Y} M(z|y)\pi_Y(y|\theta) \qquad (\pi_Z = \pi_Y M \text{ in matrix notation}).$$

Elicitation Dominance

An experiment (Y, π_Y) *dominates* an experiment (Z, π_Z) *in the sense of elicitation* if for every $\mathcal{P}$ that can be elicited by (Z, π_Z), $\mathcal{P}$ can be elicited by (Y, π_Y).

Lemma

The mean of g is elicitable with (Y, π_Y) if and only if, for some $w : Y \to \mathbb{R}$,

$$g(\theta) = \sum_y w(y)\pi(y|\theta).$$

▸ proof

So, domination is equivalent to the existence of an $Y \times Z$ matrix M s.t.

$$\pi_Z(z|\theta) = \sum_{y \in Y} M(z|y)\pi_Y(y|\theta) \qquad (\pi_Z = \pi_Y M \text{ in matrix notation}).$$

$\implies$ Blackwell dominance implies Elicitation dominance, but not conversely.

Noisy Transforms

Call (Y, π_Y) a noisy transform of (Z, π_Z) if

- $Y = Z$.
- With positive probability, $y = z$, and with the complementary probability, $y = \varepsilon$ where ε is an independent random variable.

Noisy transformations preserve the information that can be elicited, because $\pi_Y = \pi_Z M$ with M invertible, so $\pi_Z = \pi_Y M^{-1}$.

Blackwell and Elicitation Orders

The elicitation order is the transitive closure of the union of two orders:

- The Blackwell order.
- The order induced by noisy transformations.

Proposition

(Y, π_Y) dominates (Z, π_Z) in the sense of elicitation if and only if (Y, π_Y) dominates a noisy transform of (Z, π_Z) in the sense of Blackwell.

Proof idea:

Suppose (Y, π_Y) dominates (Z, π_Z) in the sense of elicitation.
Then $\pi_Z = \pi_Y M$ for some matrix M.
Choose M so that each row sums to one (always possible).

Fix $\lambda \in (0, 1)$.

- (Z', π'_Z) = noisy tranform of (Z, π_Z) where:
  $z' = z$ with probability λ and z' is uniformly drawn with probability $1 - \lambda$.
- U = Markov matrix that transforms each y into a uniformly distributed outcome of Z.

$$\begin{aligned}
\pi'_Z &= \lambda \pi_Z + (1 - \lambda)\pi_U \\
&= \lambda \pi_Y M + (1 - \lambda)\pi_Y U \\
&= \pi_Y(\lambda M + (1 - \lambda)U) \\
&= \pi_Y N \qquad \text{with } \; N = \lambda M + (1 - \lambda)U
\end{aligned}$$

Each row of N sums to one, and N has positive entries if λ is small enough.
$\implies$ N is a Markov matrix, so (Y, π_Y) Blackwell-dominates (Z', π'_Z).

Incentives

Elicitation Power and Incentives Structure

- If an experiment (Y, π_Y) dominates (Z, π_Z) in the sense of elicitation, then for any mechanism for (Z, π_Z), there exists a payoff-equivalent mechanism for (Y, π_Y).

$\Longrightarrow$ If two experiments elicit the same information, the incentives we can design are the same with both, *even if one is a strict garbling of the other.*

Elicitation Power and Incentives Structure

- If an experiment (Y, π_Y) dominates (Z, π_Z) in the sense of elicitation, then for any mechanism for (Z, π_Z), there exists a payoff-equivalent mechanism for (Y, π_Y).

$\Longrightarrow$ If two experiments elicit the same information, the incentives we can design are the same with both, *even if one is a strict garbling of the other*.

- If we can elicit some information with a given number of observations from an experiment, collecting more data does not help to strengthen incentives to elicit this information (provided we fix the payoffs of truthful reports).

Elicitation Power and Incentives Structure

- If an experiment (Y, π_Y) dominates (Z, π_Z) in the sense of elicitation, then for any mechanism for (Z, π_Z), there exists a payoff-equivalent mechanism for (Y, π_Y).

$\Longrightarrow$ If two experiments elicit the same information, the incentives we can design are the same with both, *even if one is a strict garbling of the other*.

- If we can elicit some information with a given number of observations from an experiment, collecting more data does not help to strengthen incentives to elicit this information (provided we fix the payoffs of truthful reports).
- For any two experiments (Y, π_Y), (Z, π_Z), if for any mechanism for (Z, π_Z) with non-negative payoffs there exists a payoff-equivalent mechanism for (Y, π_Y) with non-negative payoffs, then (Y, π_Y) dominates (Z, π_Z) in the sense of Blackwell.

▸ details

Summary

Summary

The problem: eliciting information on a researcher's belief about the parameter of a statistical model, when we collect independent outcome realizations and incentive provision is done via transfers that depend on reported information and outcome realizations.

(1) We characterize the information that can be elicited with a given experiment/dataset.
Small differences in information can make elicitation as hard as estimation.

(2) We propose an ordering of experiments based on their elicitation power.
It is connected to, but different from, the canonical Blackwell ordering.

(3) We relate elicitation power with flexibility in incentives design.
More elicitation power $\implies$ more flexibility (except under limited liability).
More data does not help to strengthen incentives (except under limited liability).

Appendix

Proof of Theorem 1

Proof idea:

(1) Any parameter distribution p induces outcome distribution μ_p with

$$\mu_p(A) = \mathrm{E}_p[\pi(A|\theta)].$$

This mean outcome distribution is the most we can elicit.

(2) Hahn–Kolmogorov extension theorem implies existence of $A_i \subset Y$, $i \in \mathbb{N}$, such that outcome distributions are identified on the $A_i's$.

(3) From report $p \in \Delta(\Theta)$ and realized outcome y, the following protocol elicits the mean outcome distribution:
   (a) Draw $i \in \mathbb{N}$ at random according to a full support distribution.
   (b) Pay $1 - (\mu_p(A_i) - \mathbb{1}\{y \in A_i\})^2$.

back

Poisson Model

Example: Poisson Model

An investor lends money to a population of borrowers.
A risk officer is tasked with estimating the default rate.

- θ captures the default rate over the period of interest.
- Consider the experiment that reveals the number of loan defaults over a given period: $Y = \{0, 1, 2, \dots\}$ and

$$\pi(k|\theta) = \frac{\theta^k}{k!} e^{-\theta}.$$

With this experiment we elicit the full distribution over default rates.

Example: Poisson Model

An investor lends money to a population of borrowers.
A risk officer is tasked with estimating the default rate.

- θ captures the default rate over the period of interest.
- Consider the experiment that reveals the number of loan defaults over a given period: $Y = \{0, 1, 2, \dots\}$ and

$$\pi(k|\theta) = \frac{\theta^k}{k!} e^{-\theta}.$$

With this experiment we elicit the full distribution over default rates.

Proof: For $t \in \mathbb{R}$, let $w_t(k) = (1 + \mathbf{i}t)^k$ with $\mathbf{i}$=imaginary unit. Apply theorem with

$$g_t(\theta) = \sum_{k=0}^{\infty} w_t(k)\pi(k|\theta) = e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta + \mathbf{i}t\theta)^k}{k!} = e^{\mathbf{i}t\theta}.$$

We elicit expected value of every $e^{\mathbf{i}t\theta}$ = the characteristic function of θ, and so we elicit its distribution.

back

Details on Applications of Product Experiments

Applications

Finite vs. Infinite Parameter Space

- If the parameter space is infinite, then the full parameter distribution cannot be elicited with finite data.
- If the parameter space is finite of size n, then we can always elicit the full parameter distribution with $n - 1$ observations.

Proof: Let (Y, π) be the product experiment that corresponds to a fixed number of observations.

If the parameter space is finite, π corresponds to a Markov matrix.

If $|\Theta| > |Y|$ the matrix has more rows than columns.

$\implies$ two distinct parameter distributions induce the same outcome distribution.

■

Applications

Finite vs. Infinite Parameter Space

- If the parameter space is infinite, then the full parameter distribution cannot be elicited with finite data.
- If the parameter space is finite of size n, then we can always elicit the full parameter distribution with $n-1$ observations.

Proof: Enumerate $\Theta = \{1, \ldots, n\}$. There exists $g : \Theta \to \mathbb{R}$ that is one-to-one whose mean is elicitable with one observation.

With $n-1$ observations we elicit the mean of $(1, g, g^2, \ldots, g^{n-1})$. The mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ defined as

$$(p_1, \ldots, p_n) \mapsto \sum_{i=1}^{n} p_i \begin{bmatrix} 1 \\ g(i) \\ \vdots \\ g^{n-1}(i) \end{bmatrix}$$

is invertible $\implies$ the mean outcome distribution identifies the parameter distribution.

■

Applications

Case of Two Observations

With two observations from an identified experiment, we can elicit two pieces of information:

(1) Whether the researcher knows the parameter...

(2) ...and, if applicable, the value of the parameter.

Applications

Case of Two Observations

With two observations from an identified experiment, we can elicit two pieces of information:

(1) Whether the researcher knows the parameter...

(2) ...and, if applicable, the value of the parameter.

Proof idea:

There exists $g : \Theta \to \mathbb{R}$ that is one to one and whose mean is elicitable with one observation.

$\implies$ with two observations we elicit the variance of g.

Remark that:

- The researcher knows the parameter iff the variance of g is non-zero.
- If the variance of g is zero, then the mean of g identifies the parameter.

Covariates

Case of Covariates

Suppose each observation is associated with a dependent variable $y \in Y$ (e.g. cholesterol level) and also a covariate $x \in X$ (e.g. age and weight). Assume the distribution of covariates is known.

The experiment for one random observation can be decomposed into two parts:

- The randomization on x.
- Given the covariate, the randomization on y. Assume it is captured by (Y, π_x).

Suppose information $\mathcal{P}_x$ is elicitable with (Y, π_x).

Theorem

Let $x_1, x_2, \ldots,$ be covariates drawn independently according to λ, and $\mathcal{P}$ be an information partition on parameter distributions.

If, with positive probability, the join of $\mathcal{P}_{x_1}, \mathcal{P}_{x_2}, \ldots,$ is a (weak) refinement of $\mathcal{P}$, then $\mathcal{P}$ is elicitable.

back

Gaussian Linear Models

Suppose each observation is associated with a dependent variable $y \in \mathbb{R}$ (e.g., cholesterol level) *and also a vector of covariates* $x = (x_1, \ldots, x_K) \in \mathbb{R}^K$ (e.g., age and weight).

A Gaussian linear model postulates a linear relationship

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \sigma\varepsilon$$

with ε a standard normal error.

The parameter is $\theta = (\beta_0, \ldots, \beta_K, \sigma) \in \mathbb{R}^{n+1} \times \mathbb{R}_+$.

The associated experiment can be decomposed in two parts.

- The randomization on the covariates $(x_1, \ldots, x_K)$ (arbitrary).
- Given the covariates, the randomization on the outcome y

$$y \sim \mathcal{N}(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K, \sigma^2).$$

back

Applications

Gaussian Linear Models

Suppose the vector of covariates $x = (x_1, \ldots, x_K)$ is distributed according to λ (known).

If:

(1) the support of λ has non-empty interior, and

(2) $(\beta_0, \ldots, \beta_K, \sigma)$ takes value in compact set,

then with a single observation, we can elicit the full distribution of the parameters.

Applications

Gaussian Linear Models

Suppose the vector of covariates $x = (x_1, \ldots, x_K)$ is distributed according to λ (known).

If:

(1) the support of λ has non-empty interior, and

(2) $(\beta_0, \ldots, \beta_K, \sigma)$ takes value in compact set,

then with a single observation, we can elicit the full distribution of the parameters.

Proof idea:

- For a fixed x, y is Gaussian, and mixtures of Gaussians whose mean and variance belong to a compact set are identified.
- The distribution over $(\beta_0, \ldots, \beta_K, \sigma)$ is determined by the distribution of its one-dimensional projections.

back

Applications

Non-Parametric Linear Models

The parameter specifies the probabilities of $y \in \mathbb{R}$ given a vector of covariates $x = (x_1, \ldots, x_K) \in \mathbb{R}^K$.

The shape of the probability distribution is free, but the joint probability of (x, y) satisfies:

$$\begin{aligned} \mathrm{E}[y|x] &= \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K, \\ \mathrm{var}[y|x] &= \sigma^2. \end{aligned}$$

Under regularity conditions:

- With one observation, we can elicit the mean of each β_i.
- With two observations, we can elicit the mean and variance of each β_i, and the mean of σ^2.
- With four observations, we can elicit the mean and variance of each β_i, and of σ^2.

back

Elicitation with Categorical Experiments

Proof idea (finite parameter space):

(1) Assume $\Theta = \{\theta_1, \ldots, \theta_n\}$ and $Y = \{y_1, \ldots, y_m\}$.

(2) Interpret $\pi(y|\theta)$ as an $n \times m$ Markov matrix, and $\Delta(\Theta) \subset \mathbb{R}^n$.

(3) $p \sim_{\mathcal{P}^\star} q \quad \Leftrightarrow \quad \mathrm{E}_p[\pi(y_i|\theta)] = \mathrm{E}_q[\pi(y_i|\theta)] \;\; (\forall i) \quad \Leftrightarrow \quad p\pi = q\pi.$

$\Longrightarrow$ So $p \sim_{\mathcal{P}^\star} q$ iff $p = q + v$ where $v \in \ker \pi^\intercal$ (a linear space).

(4) Take $g : \Theta \to \mathbb{R}$. Interpret g as a vector of $\mathbb{R}^n$.

(5) If we elicit the mean of g, then $p \sim_{\mathcal{P}^\star} q$ implies $\underbrace{\mathrm{E}_p[g(\theta)]}_{p \cdot g} = \underbrace{\mathrm{E}_q[g(\theta)]}_{q \cdot g}$.

(6) Hence $g \perp \ker \pi^\intercal$.

(7) Standard linear algebra: $(\operatorname{img} \pi)^\perp = \ker \pi^\intercal$.

$\Longrightarrow$ Hence $g \in \operatorname{img} \pi$.

back

Sufficient Statistics for Elicitation

Sufficient Statistics

Sufficiency in the Classical Sense

Consider an experiment (Y, π) that generates outcome y.

A *statistic* (for this experiment) is a function of the observation $T(y)$.

The experiment induced by the statistic is written $(T(Y), T(\pi))$.

The statistic is *sufficient in the classical sense* if

$$\Pr[y|\theta, T(y)] = \Pr[y|\theta', T(y)] \qquad \forall \theta \neq \theta'.$$

Sufficient Statistics

Sufficiency in the Classical Sense

Consider an experiment (Y, π) that generates outcome y.

A *statistic* (for this experiment) is a function of the observation $T(y)$.

The experiment induced by the statistic is written $(T(Y), T(\pi))$.

The statistic is *sufficient in the classical sense* if

$$\Pr[y|\theta, T(y)] = \Pr[y|\theta', T(y)] \qquad \forall \theta \neq \theta'.$$

If T is a sufficient statistic in the classical sense, (Y, π) is Blackwell equivalent to $(T(Y), T(\pi))$.

$\implies$ If T is a sufficient statistic, we elicit the same information with (Y, π) as with $(T(Y), T(\pi))$.

back

Sufficient Statistics

Sufficiency in Elicitation

Suppose we get n observations $y_1, \ldots, y_n$.

Statistic (for the product experiment) is now $T(y_1, \ldots, y_n)$.

Cost of T = the minimum number of observations needed to compute T.

$\rightsquigarrow$ Formally cost $C(y_1, \ldots, y_n)$ = minimum index j such that $T(y_1, \ldots, y_j, y'_{j+1}, \ldots, y'_n)$ does not depend on $y'_{j+1}, \ldots, y'_n$.

T is a *sufficient statistic in the elicitation sense* if, when we can elicit $\mathcal{P}$ with (Y, π), we can also elicit $\mathcal{P}$ with $(T(Y), T(\pi))$.

Sufficiency in the classical sense $\implies$ sufficiency in the elicitation sense, but not conversely.

Sufficient statistic T is *optimal* when there does not exist a sufficient statistic T' such that $C' \leq C$ and for some $y_1, \ldots, y_n$, $C'(y_1, \ldots, y_n) < C(y_1, \ldots, y_n)$.

back

Optimal Sampling Strategy

Enumerate $Y = \{1, \ldots, m\}$.

Below, y captures the current observation.

Starting with the 1st observation, apply the following algorithm:

- If $y = 1$ then keep observing until $y \neq 1$ (or we have n observations).
  If we have collected n observations we stop, else we continue below.
- If $y = 2$ then keep observing until $y \neq 2$ (or we have n observations).
  If we have collected n observations we stop, else we continue below.
- ...
- If $y = m - 1$ then keep observing until $y \neq m - 1$ (or we have n observations).

back

Incentives

Elicitation Power and Incentive Structures

Here we consider general mechanisms $\varphi : \mathcal{M} \times Y \to \mathbb{R}$.

If an experiment (Y, π_Y) dominates an experiment (Z, π_Z) in the sense of elicitation, then for every general mechanism φ_Z for (Z, π_Z), there exists a general mechanism φ_Y for (Y, π_Y) such that

$$\mathrm{E}_p[\varphi_Y(\mathfrak{m}, y)] = \mathrm{E}_p[\varphi_Z(\mathfrak{m}, z)] \qquad \text{(for all } \mathfrak{m}, p\text{)}.$$

$\implies$ If (Y, π_Y) is equivalent to (Z, π_Z) in the sense of elicitation, then the incentives we can design are the same with both—even if one is a strict garbling of the other.

back

Elicitation Power and Incentive Structures

Fix an experiment. Suppose we can elicit information $\mathcal{P}$ with n observations.

Q: Does collecting more data help strengthen incentives to elicit $\mathcal{P}$?

A: No (fixing the payoffs of the truthful reports).

back

Elicitation Power and Incentive Structures

Assume that, for every general mechanism φ_Z for (Z, π_Z) with $\varphi_Z \geq 0$, there exists a general mechanism φ_Y for (Y, π_Y) with $\varphi_Y \geq 0$,

$$\mathrm{E}_p[\varphi_Y(\mathfrak{m}, y)] = \mathrm{E}_p[\varphi_Z(\mathfrak{m}, z)] \qquad \text{(for all } \mathfrak{m}, p\text{)}.$$

Then, (Y, π_Y) dominates (Z, π_Z) in the sense of Blackwell.

The converse is true.

back

BDM Representation

Consider an experiment (Y, π) with $Y = \{1, \ldots, m\}$ and the $(m-1)$-simplex as parameter space that captures the outcome distribution (e.g., Bernoulli model).

Let φ be a (direct, IC) mechanism. Assume:

For parameter θ^*: $\mathrm{E}_p[\theta] = \lambda\theta^* + (1-\lambda)\,\mathrm{E}_q[\theta] \quad \Longrightarrow \quad \mathrm{E}_{\theta^*}[\varphi(p,y)] \geq \mathrm{E}_{\theta^*}[\varphi(q,y)]$.

Under smootness conditions, we can write

$$\varphi(p,y) = s_0(y) + \sum_{i=1}^{m} \int_0^{\mathrm{E}_p[\theta_i]} (\mathbb{1}\{y=i\} - t) w_i(t)\,\mathrm{d}t$$

where $s_0(y)$ and $w_i(t)$ are arbitrary.

back