Costly Information Acquisition

Topics in Economic Theory

Duarte Gonçalves

University College London

2-16 March 2022

Motivation

"Information is power": Blackwell order

Motivation

"Information is power": Blackwell order

... but "There's no such thing as a free lunch": Information is costly

- Time
- Labor: need to pay people to gather the data (e.g. running surveys, parsing reviews), when not to elicit data in IC manner (e.g. econ experiments)
- Capital: labs, IT (data crunching), human capital (understanding what computer memory even is)

Today

Idea: learning about unknown state is costly

- Pay attention, process available information, gather information, etc.

Stylized model of costly info acquisition

- Can design/choose costly experiment
- Static problem, know available actions
- All bundled into cost of info
- Miss out on other obsevables: time, investment, limits to what is knowable
- Different models tailored to different problems

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Overview

1. Costly Information Acquisition
   - Model
   - Value and Cost of Information
   - State-Dependent Stochastic Choice Data
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Model

References: Caplin & Dean ('15 AER); Chambers, Liu, & Rehbeck ('20 JET)

Unknown state of the world $\theta \in \Theta$, finite

Prior prob distribution $\mu \in \Delta(\Theta)$

Decision problem: A

Action a mapping from states to prize space Z, $a : \Theta \to Z$

Finite decision problems: $\mathcal{A} := \{A \subset Z^{\Theta} \mid |A| < \infty\}$

vNM utility $u : Z \to \mathbb{R}$; write $u(a, \theta) \equiv u(a(\theta))$

Assumption: u and μ are known (more on this later)

Model

Information structure $\pi : \Theta \to \Delta(\Delta(\Theta))$ with finite support $S(\pi)$

Identify signals with posterior distributions: for $m \in S(\pi)$

$$m(\theta) = \frac{\mu(\theta)\pi(m \mid \theta)}{\sum_{\theta'} \mu(\theta')\pi(m \mid \theta')}$$

Π: set of all such info structures

Conditional posterior distrib: $\pi(m \mid \theta)$

Unconditional posterior distrib: $\pi(m) := \mathbb{E}_\mu[\pi(m|\theta)] = \sum_\theta \mu(\theta)\pi(m \mid \theta)$

$\mathbb{E}$ Posterior = Prior: $\mathbb{E}_\pi[m] = \mu$

$m \sim \pi$ MPS of μ

Implicit assumption here: two signals inducing the same posterior beliefs are treated as the same signal

Value of Information

EU max: $U^A(\mu) := \max_{a \in A} \mathbb{E}_\mu[u(a, \theta)]$

Continuous (convex) function on $\Delta(\Theta)$

Gross value of information π in decision problem A given prior μ:

$$G^A(\pi, \mu) := \mathbb{E}_\mu \left[\mathbb{E}_{\pi(\cdot|\theta)} \left[U^A(m) \right] \right] = \sum_\theta \mu(\theta) \sum_m \pi(m \mid \theta) \max_{a \in A} \mathbb{E}_m[u(a, \theta)]$$

Write $G^A(\pi, \mu) \equiv \pi \cdot U^A$

Will omit dependence on μ where unambiguous

Value of Information

EU max: $U^A(\mu) := \max_{a \in A} \mathbb{E}_\mu[u(a, \theta)]$

Continuous (convex) function on $\Delta(\Theta)$

Gross value of information π in decision problem A given prior μ:

$$G^A(\pi, \mu) := \mathbb{E}_\mu \left[\mathbb{E}_{\pi(\cdot|\theta)} \left[U^A(m) \right] \right] = \sum_\theta \mu(\theta) \sum_m \pi(m \mid \theta) \max_{a \in A} \mathbb{E}_m[u(a, \theta)]$$

Write $G^A(\pi, \mu) \equiv \pi \cdot U^A$

Will omit dependence on μ where unambiguous

Optimal Choices given Information π

Choice function $\sigma_A : S(\pi) \to \Delta(A)$ is optimal given π if, $\forall m \in S(\pi)$

$$\text{supp}(\sigma_A(m)) \subseteq \arg\max_{a \in A} \mathbb{E}_m[u(a, \theta)]$$

Note: $G^A(\pi) = \max_{\sigma_A} \mathbb{E}_\mu \left[E_{\pi(\cdot|\theta)} \left[E_{\sigma_A} [\mathbb{E}_m[u(a, \theta)]] \right] \right]$

Cost of Information

Net value of information π in decision problem A given prior μ

$$V(\pi \cdot U^A, \pi)$$

where $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup -\infty$ s.t. V strictly increasing in 1st arg (when finite)

Examples:

Additive costs: $V(\pi \cdot U^A, \pi; \mu) = G^A(\pi, \mu) - K(\pi; \mu)$

Multiplicative costs: $V(\pi \cdot U^A, \pi; \mu) = R(\pi; \mu)G^A(\pi, \mu)$

Combinations of the two

Additive prior-independent costs: $V(\pi \cdot U^A, \pi; \mu) = G^A(\pi, \mu) - K(\pi)$

Optimal Information Acquisition

$$\Pi^*(A) := \arg\max_{\pi \in \Pi} V(\pi \cdot U^A, \pi)$$

State-Dependent Stochastic Choice Data

Definition

A state-dependent stochastic choice dataset is

a finite collection of decision problems $\mathcal{D} \subset \mathcal{A}$ and

a set of state-dependent stochastic choice functions $P = \{P_A\}_{A \in \mathcal{D}}$ where $P_A : \Theta \to \Delta(A)$

$P_A(\theta)(a) \equiv P_A(a \mid \theta)$

Analyst can observe frequency with which each action is taken at each state:

{add egg,not}, {good,rotten} (Savage's problem)

Investment decisions, realized return

Pricing decisions, inflation

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
   - Revealed Information Structures
   - No Improving Action Switches (NIAS)
   - Generalized Axiom of Costly Information (GACI)
   - Representation Result
   - Unobservables
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Representation

Want "(A) *P* can be rationalized by a costly info acquisition model sat. [some properties]

iff (B) *P* sat. [some properties]"

(A) has two main requirements on *P*

Choices optimal given info structure

Info structure optimal

Definition

Given $\mu \in \Delta(\Theta)$ and $u : Z \to \mathbb{R}$, a state-dependent stochastic choice dataset $(\mathcal{D}, P)$ admits a **costly information acquisition representation** if there are

a) $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$

b) information structures $\{\pi_A\}_{A \in \mathcal{D}}$

c) choice functions $\{\sigma_A\}_{A \in \mathcal{D}}$

such that

1. **Monotonicity** of V: $\forall \pi \in \Pi, r' > r, V(r, \pi) > -\infty \implies V(r', \pi) > V(r, \pi)$ $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$
2. **Non-triviality**: $\forall r, \exists \pi_r$ such that $V(r, \pi_r) > -\infty$
3. **Choices σ_A are optimal**
4. **Information π_A is optimal**
5. **Data are matched**: $\forall A \in \mathcal{D}, \theta \in \Theta, a \in A$,

$$P_A(a|\theta) = \mathbb{E}_{\pi(\cdot|\theta)}[\sigma_A(a|m)] = \sum_{m \in S(\pi)} \pi(m \mid \theta)\sigma_A(a|m)$$

Revealed Information Structures

First step: base representation conditions on observables

Revealed Posterior: given $\mu, A \in \mathcal{D}, P_A \in P$, and $a \in \text{supp}(P_A)$, revealed posterior $\bar{m}^a \in \Delta(\Theta)$ is given by

$$\bar{m}^a(\theta) = \mathbb{P}(\theta \mid a \text{ is chosen from } A) = \frac{\mu(\theta) P_A(a \mid \theta)}{\sum_{\theta'} \mu(\theta') P_A(a \mid \theta')}$$

Revealed Information Structures

First step: base representation conditions on observables

Revealed Posterior: given $\mu, A \in \mathcal{D}, P_A \in P$, and $a \in \text{supp}(P_A)$, revealed posterior $\bar{m}^a \in \Delta(\Theta)$ is given by

$$\bar{m}^a(\theta) = \mathbb{P}(\theta \mid a \text{ is chosen from } A) = \frac{\mu(\theta) P_A(a \mid \theta)}{\sum_{\theta'} \mu(\theta') P_A(a \mid \theta')}$$

Revealed Information Structures: given $\mu, A \in \mathcal{D}, P_A \in P$, and $a \in \text{supp}(P_A)$, revealed information structure $\bar{\pi}_A$ is given by

$$\bar{\pi}_A(m \mid \theta) = \sum_{a \in \text{supp}(P_A) \mid \bar{m}^a = m} P_A(a \mid \theta)$$

inducing distrib on revealed posteriors $\bar{\pi}_A(\bar{m}^a) = \sum_\theta \mu(\theta) \bar{\pi}_A(\bar{m}^a \mid \theta)$

Revealed Information Structures

First step: base representation conditions on observables

Revealed Posterior: given $\mu, A \in \mathcal{D}, P_A \in P$, and $a \in \text{supp}(P_A)$, revealed posterior $\bar{m}^a \in \Delta(\Theta)$ is given by

$$\bar{m}^a(\theta) = \mathbb{P}(\theta \mid a \text{ is chosen from } A) = \frac{\mu(\theta) P_A(a \mid \theta)}{\sum_{\theta'} \mu(\theta') P_A(a \mid \theta')}$$

Revealed Information Structures: given $\mu, A \in \mathcal{D}, P_A \in P$, and $a \in \text{supp}(P_A)$, revealed information structure $\bar{\pi}_A$ is given by

$$\bar{\pi}_A(m \mid \theta) = \sum_{a \in \text{supp}(P_A) \mid \bar{m}^a = m} P_A(a \mid \theta)$$

inducing distrib on revealed posteriors $\bar{\pi}_A(\bar{m}^a) = \sum_\theta \mu(\theta) \bar{\pi}_A(\bar{m}^a \mid \theta)$

Idea: Suppose that $A = \{0, 1\}$, and $S(\pi) = \{m, m', m''\}$;

given m, m' chose 0; given m'' chose 1

Could do as well by bundling m, m' into a single signal m^0 at with $a = 0$ is optimal

Revealed Information Structures

Definition

π_A is **consistent** with P_A if $\exists \sigma_A : S(\pi) \to \Delta(A)$ such that

(i) Choices σ_A are optimal

$$\text{supp}(\sigma_A(m)) \subseteq \arg\max_{a \in A} \mathbb{E}_m[u(a,\theta)]\, P_A(a|\theta)$$

(ii) Data are matched

$$P_A(a|\theta) = \mathbb{E}_{\pi(\cdot|\theta)}[\sigma_A(m)] = \sum_{m \in S(\pi)} \pi(m \mid \theta)\sigma_A(m)$$

Lemma

If π_A is consistent with P_A, then it is sufficient for $\bar{\pi}_A$

Intuition: SDSC data cannot be + informative than info structure that induced it

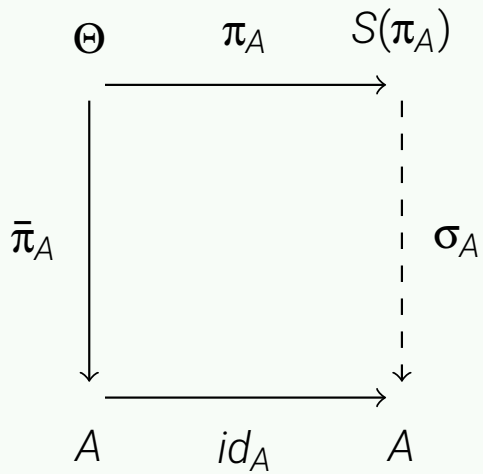
Revealed Information Structures

Lemma

If π_A is consistent with P_A, then it is sufficient for $\bar{\pi}_A$

Proof Sketch

Relabel $\bar{m}^a$ as a; Then $\sigma_A \circ \pi_A = P_A = id_A \circ \bar{\pi}_A = \bar{\pi}_A$ (recall Oliveira ('18 GEB))



Revealed Information Structures

Definition

π_A is **consistent** with P_A if $\exists \sigma_A : S(\pi) \to \Delta(A)$ such that

(i) Choices σ_A are optimal

$$\text{supp}(\sigma_A(m)) \subseteq \arg\max_{a \in A} \mathbb{E}_m[u(a,\theta)]\, P_A(a|\theta)$$

(ii) Data are matched

$$P_A(a|\theta) = \mathbb{E}_{\pi(\cdot|\theta)}[\sigma_A(m)] = \sum_{m \in S(\pi)} \pi(m \mid \theta)\sigma_A(m)$$

Lemma

If π_A is consistent with P_A, then it is sufficient for $\bar{\pi}_A$

Lemma

For all $A, B \in \mathcal{D}$, if π_A is consistent with P_A, then $\pi_A \cdot U^B \geq \bar{\pi}_A \cdot U^B$ and $\pi_A \cdot U^A = \bar{\pi}_A \cdot U^A$

$\geq$ immediately from Blackwell ordering; = follows from fact that data is matched

Optimality of Actions

Note that if $\boldsymbol{\pi}_A$ is consistent with P_A, then so is $\bar{\boldsymbol{\pi}}_A$

As $G^A(\boldsymbol{\pi}_A) = G^A(\bar{\boldsymbol{\pi}}_A)$, then $\boldsymbol{\sigma}_A(a' \mid \bar{m}^a) := \mathbf{1}_{a'=a}$ is optimal given $\bar{\boldsymbol{\pi}}_A$; i.e.

$$
\begin{aligned}
a \in \arg\max_{a' \in A} \mathbb{E}_{\bar{m}^a}[u(a', \boldsymbol{\theta})] &\iff \sum_{\boldsymbol{\theta}} \bar{m}^a(\boldsymbol{\theta})[u(a, \boldsymbol{\theta}) - u(b, \boldsymbol{\theta})] \geq \mathbf{0}, \ \ \forall b \in A \\
&\iff \sum_{\boldsymbol{\theta}} \frac{\boldsymbol{\mu}(\boldsymbol{\theta}) P_A(a \mid \boldsymbol{\theta})}{\sum_{\boldsymbol{\theta}'} \boldsymbol{\mu}(\boldsymbol{\theta}') P_A(a \mid \boldsymbol{\theta}')} [u(a, \boldsymbol{\theta}) - u(b, \boldsymbol{\theta})] \geq \mathbf{0}, \ \ \forall b \in A \\
&\iff \sum_{\boldsymbol{\theta}} \boldsymbol{\mu}(\boldsymbol{\theta}) P_A(a \mid \boldsymbol{\theta})[u(a, \boldsymbol{\theta}) - u(b, \boldsymbol{\theta})] \geq \mathbf{0}, \ \ \forall b \in A
\end{aligned}
$$

Optimality of Actions

Note that if π_A is consistent with P_A, then so is $\bar{\pi}_A$

As $G^A(\pi_A) = G^A(\bar{\pi}_A)$, then $\sigma_A(a' \mid \bar{m}^a) := \mathbf{1}_{a'=a}$ is optimal given $\bar{\pi}_A$; i.e.

$$
\begin{aligned}
a \in \arg\max_{a' \in A} \mathbb{E}_{\bar{m}^a}[u(a', \theta)] &\iff \sum_{\theta} \bar{m}^a(\theta)[u(a, \theta) - u(b, \theta)] \geq 0, \ \ \forall b \in A \\
&\iff \sum_{\theta} \frac{\mu(\theta) P_A(a \mid \theta)}{\sum_{\theta'} \mu(\theta') P_A(a \mid \theta')}[u(a, \theta) - u(b, \theta)] \geq 0, \ \ \forall b \in A \\
&\iff \sum_{\theta} \mu(\theta) P_A(a \mid \theta)[u(a, \theta) - u(b, \theta)] \geq 0, \ \ \forall b \in A
\end{aligned}
$$

Definition

Given $\mu, u, (\mathcal{D}, P)$ sat. **no improving action switches** (NIAS) if for every $A \in \mathcal{D}$, and every $a \in A$ such that $P_A(a) > 0$, $\sum_{\theta} \mu(\theta) P_A(a \mid \theta)[u(a, \theta) - u(b, \theta)] \geq 0, \ \ \forall b \in A$

Optimality of Actions

Taking stock

Enough to get one signal for each action chosen with positive prob.

NIAS: $\sum_{\theta} \mu(\theta) P_A(a \mid \theta)[u(a, \theta) - u(b, \theta)] \geq 0$, for a chosen with positive prob

Required for choices to be optimal given some info structure and data to be matched

(NIAS as an obedience constraint (BNE, BCE): player has no incentive to deviate from recommended course of action given signal realization)

Optimality of Actions

Taking stock

- Enough to get one signal for each action chosen with positive prob.
- NIAS: $\sum_{\theta} \mu(\theta) P_A(a \mid \theta)[u(a, \theta) - u(b, \theta)] \geq 0$, for a chosen with positive prob
  - Required for choices to be optimal given some info structure and data to be matched
- (NIAS as an obedience constraint (BNE, BCE): player has no incentive to deviate from recommended course of action given signal realization)

Next: implications of optimal information structure

Optimality of Information Structures

Definition

Given $\mu, u, (\mathcal{D}, P)$ sat. **generalized axiom of costly information** (GACI) if for all sequences $(\bar{\pi}_{A_1}, U^{A_1}), (\bar{\pi}_{A_2}, U^{A_2}), ..., (\bar{\pi}_{A_k}, U^{A_k})$ with $A_i \in \mathcal{D}$ for which $\bar{\pi}_{A_i} \cdot U^{A_i} \leq \bar{\pi}_{A_i} \cdot U^{A_{i+1}} \pmod{k}$, all inequalities hold with equality

Where does this come into play? Related to Afriat's theorem! (Chambers & Echenique '16)

Optimality of Information Structures

Definition

Given $\mu, u, (\mathcal{D}, P)$ sat. **generalized axiom of costly information** (GACI) if for all sequences $(\bar{\pi}_{A_1}, U^{A_1}), (\bar{\pi}_{A_2}, U^{A_2}), ..., (\bar{\pi}_{A_k}, U^{A_k})$ with $A_i \in \mathcal{D}$ for which $\bar{\pi}_{A_i} \cdot U^{A_i} \leq \bar{\pi}_{A_i} \cdot U^{A_{i+1}} \pmod{k}$, all inequalities hold with equality

Where does this come into play? Related to Afriat's theorem! (Chambers & Echenique '16)

Lemma (Afriat's Theorem)

Let $\mathcal{D}$ be finite and for each $A, B \in \mathcal{D}$ let $u_{A,B} \in \mathbb{R}$.

Suppose that for all $A \in \mathcal{D}$, $u_{A,A} = 0$ and that for any sequence $A_1, A_2, ..., A_k$ s.t. $u_{A_i,A_{i+1}} \leq 0$ for all $i \pmod{k}$, it holds that $u_{A_i,A_{i+1}} = 0$ for all $i \pmod{k}$.

Then there are numbers $U_A \in \mathbb{R}$ and $\lambda_A > 0$ s.t. $\forall A, B \in \mathcal{D}$, $U_A \leq U_B + \lambda_B u_{B,A}$

Restrictions on the Cost of Information

Would like to get implicit cost to sat. reasonable properties

$V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$ sat.

Weak monotonicity if π is sufficient for π' then for any $t \in \mathbb{R}$, $V(t, \pi) \leq V(t, \pi')$

Idea: More informative info structures are more costly

Quasiconcavity if $\forall \lambda \in [0, 1]$, $(t, \pi), (t', \pi') \in \mathbb{R} \times \Pi$, $V(t'', \pi'') \geq V(t', \pi') \wedge V(t, \pi)$, where $(t'', \pi'') = \lambda(t, \pi) + (1 - \lambda)(t', \pi')$

Idea: Implies qconcavity in π (setting $t = \pi \cdot U$ and $t' = \pi' \cdot U$)

Normalization if $V(0, \pi_0) = 0$, where π_0 is s.t. $\pi_0(\mu \mid \theta) = 1 \; \forall \theta$

Idea: utility is normalized to zero when ex-ante payoff is zero and info structure is fully uninformative (posteriors = prior)

Representation Result

Theorem (Chambers, Liu, & Rehbeck, '20 JET)

Fix μ and u and let $(\mathcal{D}, P)$ be a state-dependent stochastic choice dataset. The following are equivalent:

(i) $(\mathcal{D}, P)$ sat. NIAS and GACI

(ii) $(\mathcal{D}, P)$ admits a costly information acquisition representation

Restrictions on V come for free; i.e. are untestable

Representation Result

Theorem (Chambers, Liu, & Rehbeck, '20 JET)

Fix μ and u and let $(\mathcal{D}, P)$ be a state-dependent stochastic choice dataset. The following are equivalent:

(i) $(\mathcal{D}, P)$ sat. NIAS and GACI

(ii) $(\mathcal{D}, P)$ admits a costly information acquisition representation

(iii) $(\mathcal{D}, P)$ admits a costly information acquisition representation with V satisfying weak monotonicity, quasiconcavity, and normalization

Restrictions on V come for free; i.e. are untestable

Proof Sketch

- CIA rep w/ properties $\Longrightarrow$ NIAS immediate from fact that if $\boldsymbol{\pi}_A \in \boldsymbol{\Pi}^*(A)$ and choices are optimal, then $\boldsymbol{\pi}_A$ is consistent with P_A $\Longrightarrow$ (lemma) $\bar{\boldsymbol{\pi}}_A$ is consistent with P_A

Proof Sketch

- CIA rep w/ properties $\Longrightarrow$ NIAS immediate from fact that if $\boldsymbol{\pi}_A \in \boldsymbol{\Pi}^*(A)$ and choices are optimal, then $\boldsymbol{\pi}_A$ is consistent with P_A $\Longrightarrow$ (lemma) $\bar{\boldsymbol{\pi}}_A$ is consistent with P_A
- CIA rep w/ properties $\Longrightarrow$ GACI

$V(\boldsymbol{\pi}_A \cdot U^A, \boldsymbol{\pi}_A) > -\infty$: non-triviality $\forall r, \exists \boldsymbol{\pi}_r$ s.t. $-\infty < V(r, \boldsymbol{\pi}_r)$

Let $c := \inf_{m \in \Delta(\Theta)} U^A(m)$ and $\boldsymbol{\pi}_c : -\infty < V(c, \boldsymbol{\pi}_c)$

$\Longrightarrow -\infty < V(c, \boldsymbol{\pi}_c) \leq V(\boldsymbol{\pi}_c \cdot U^A, \boldsymbol{\pi}_c) \leq V(\boldsymbol{\pi}_A \cdot U^A, \boldsymbol{\pi}_A)$

Proof Sketch

- CIA rep w/ properties $\Longrightarrow$ NIAS immediate from fact that if $\boldsymbol{\pi}_A \in \boldsymbol{\Pi}^*(A)$ and choices are optimal, then $\boldsymbol{\pi}_A$ is consistent with P_A $\Longrightarrow$ (lemma) $\bar{\boldsymbol{\pi}}_A$ is consistent with P_A
- CIA rep w/ properties $\Longrightarrow$ GACI

$V(\boldsymbol{\pi}_A \cdot U^A, \boldsymbol{\pi}_A) > -\infty$: non-triviality $\forall r, \exists \boldsymbol{\pi}_r$ s.t. $-\infty < V(r, \boldsymbol{\pi}_r)$

Let $c := \inf_{m \in \Delta(\Theta)} U^A(m)$ and $\boldsymbol{\pi}_c : -\infty < V(c, \boldsymbol{\pi}_c)$

$\Longrightarrow -\infty < V(c, \boldsymbol{\pi}_c) \leq V(\boldsymbol{\pi}_c \cdot U^A, \boldsymbol{\pi}_c) \leq V(\boldsymbol{\pi}_A \cdot U^A, \boldsymbol{\pi}_A)$

Suppose $(\bar{\boldsymbol{\pi}}_{A_1}, U^{A_1}), \ldots, (\bar{\boldsymbol{\pi}}_{A_k}, U^{A_k})$ s.t. $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} \leq \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}$ for all i (mod k)

$$V(\boldsymbol{\pi}_{A_i} \cdot U^{A_i}, \boldsymbol{\pi}_{A_i}) = V(\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i}, \boldsymbol{\pi}_{A_i}) \leq V(\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_i})$$

$$\leq V(\boldsymbol{\pi}_{A_i} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_i}) \leq V(\boldsymbol{\pi}_{A_{i+1}} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_{i+1}}) = V(\bar{\boldsymbol{\pi}}_{A_{i+1}} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_{i+1}})$$

1st $\leq$ is strict if $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} < \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}$ (by strict monot in 1st arg of V)

If $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} < \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}$, we obtain a cycle and $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} < \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i}$, contradiction

Proof Sketch

- CIA rep w/ properties $\Longrightarrow$ NIAS immediate from fact that if $\boldsymbol{\pi}_A \in \boldsymbol{\Pi}^*(A)$ and choices are optimal, then $\boldsymbol{\pi}_A$ is consistent with P_A $\Longrightarrow$ (lemma) $\bar{\boldsymbol{\pi}}_A$ is consistent with P_A
- CIA rep w/ properties $\Longrightarrow$ GACI

$V(\boldsymbol{\pi}_A \cdot U^A, \boldsymbol{\pi}_A) > -\infty$: non-triviality $\forall r, \exists \boldsymbol{\pi}_r$ s.t. $-\infty < V(r, \boldsymbol{\pi}_r)$

Let $c := \inf_{m \in \Delta(\Theta)} U^A(m)$ and $\boldsymbol{\pi}_c : -\infty < V(c, \boldsymbol{\pi}_c)$

$\Longrightarrow -\infty < V(c, \boldsymbol{\pi}_c) \leq V(\boldsymbol{\pi}_c \cdot U^A, \boldsymbol{\pi}_c) \leq V(\boldsymbol{\pi}_A \cdot U^A, \boldsymbol{\pi}_A)$

Suppose $(\bar{\boldsymbol{\pi}}_{A_1}, U^{A_1}), ..., (\bar{\boldsymbol{\pi}}_{A_k}, U^{A_k})$ s.t. $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} \leq \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}$ for all i (mod k)

$$V(\boldsymbol{\pi}_{A_i} \cdot U^{A_i}, \boldsymbol{\pi}_{A_i}) = V(\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i}, \boldsymbol{\pi}_{A_i}) \leq V(\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_i})$$

$$\leq V(\boldsymbol{\pi}_{A_i} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_i}) \leq V(\boldsymbol{\pi}_{A_{i+1}} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_{i+1}}) = V(\bar{\boldsymbol{\pi}}_{A_{i+1}} \cdot U^{A_{i+1}}, \boldsymbol{\pi}_{A_{i+1}})$$

1st $\leq$ is strict if $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} < \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}$ (by strict monot in 1st arg of V)

If $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} < \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_{i+1}}$, we obtain a cycle and $\bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i} < \bar{\boldsymbol{\pi}}_{A_i} \cdot U^{A_i}$, contradiction

- CIA rep w/ properties $\Longleftarrow$ GACI + NIAS: application of Afriat's thm (see paper)

Unobservables

What if u/μ are unknown? Know that $u \in \mathcal{U}/\mu \in \mathcal{M}$: test NIAS and GACI for each u/μ

Easy-to-check conditions (linear conditions)

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
   - No Improving Attention Cycles (NIAC)
   - Optimal Information Structure with Additive Costs
   - Distinguishing Nonseparable and Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Additive Cost of Information

Cost of information structure $K : \Pi \to \mathbb{R} \cup \{\infty\}$ with $K(\pi) < \infty$ for some $\pi \in \Pi$

$\mathcal{K}$: set of all such cost functions

Additive cost: with loss

Infinite costs = hard constraints

e.g. bound on info acquired, restrictions on class of information structures

Optimal info structures

$G(A, \pi) - K(\pi) = \pi \cdot U^A - K(\pi)$

$$\Pi^*(K, A) := \arg\max_{\pi \in \Pi} \pi \cdot U^A - K(\pi)$$

Optimality of Information Structures

Idea:

$$\begin{cases} \pi_A \cdot U^A - K(\pi_A) \geq \pi_B \cdot U^A - K(\pi_B) \\ \pi_B \cdot U^B - K(\pi_B) \geq \pi_A \cdot U^B - K(\pi_A) \end{cases} \Longrightarrow \pi_A \cdot U^A + \pi_B \cdot U^B \geq \pi_B \cdot U^A + \pi_A \cdot U^B$$

$$\begin{aligned} \Longleftrightarrow \bar{\pi}_A \cdot U^A + \bar{\pi}_B \cdot U^B &= \pi_A \cdot U^A + \pi_B \cdot U^B \\ &\geq \pi_B \cdot U^A + \pi_A \cdot U^B \\ &\geq \bar{\pi}_B \cdot U^A + \bar{\pi}_A \cdot U^B \end{aligned}$$

where last $\geq$ due to $\pi_{A_i} \geq_B \bar{\pi}_{A_i}$

Interpretation: assignment problem

have $\{(\pi_A, U^A)\}_{A \in \mathcal{D}}$; $\pi_i \cdot U^j$: surplus generated by assigning i to j; $K(\pi_i)$: cost

Surplus maximized by assigning i to i

Optimality of Information Structures

Definition

Given μ, u, $(\mathcal{D}, P)$ sat. **no improving attention cycles** (NIAC) if for all sequences $(\bar{\pi}_{A_1}, U^{A_1}), (\bar{\pi}_{A_2}, U^{A_2}), ..., (\bar{\pi}_{A_k}, U^{A_k})$ with $A_i \in \mathcal{D}$, $\sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot U^{A_i} \geq \sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot U^{A_{i+1}} \pmod{k}$

Easy to check that $\neg$ GACI $\Longrightarrow \neg$ NIAC

Representation Result

Theorem (Caplin & Dean, '15 AER)

Fix μ and u and let $(\mathcal{D}, P)$ be a state-dependent stochastic choice dataset. The following are equivalent:

(i) $(\mathcal{D}, P)$ sat. NIAS and NIAC

(ii) $(\mathcal{D}, P)$ admits an additive costly information acquisition representation

(iii) $(\mathcal{D}, P)$ admits an additive costly information acquisition representation with $K \in \mathcal{K}$ nondecreasing in Blackwell order, convex

Restrictions on K come for free; i.e. are untestable

Can also normalize K s.t. zero cost for uninformative info structures

Representation Result

Proof Sketch

- Additive CIA rep w/ properties $\Longrightarrow$ NIAS, NIAC easier

focus instead on:

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

Representation Result

Proof Sketch

- Additive CIA rep w/ properties $\Longrightarrow$ NIAS, NIAC easier

focus instead on:

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

$$\boldsymbol{\pi} \cdot U^A - K(\boldsymbol{\pi}) \leq \bar{\boldsymbol{\pi}}_A \cdot U^A - K(\bar{\boldsymbol{\pi}}_A) \Longleftrightarrow K(\bar{\boldsymbol{\pi}}_A) - K(\boldsymbol{\pi}) \leq \bar{\boldsymbol{\pi}}_A \cdot U^A - \boldsymbol{\pi} \cdot U^A$$

Implies upper & lower bounds on cost difference of revealed info structures

$$\bar{\boldsymbol{\pi}}_A \cdot U^B - \bar{\boldsymbol{\pi}}_B \cdot U^B \leq K(\bar{\boldsymbol{\pi}}_A) - K(\bar{\boldsymbol{\pi}}_B) \leq \bar{\boldsymbol{\pi}}_A \cdot U^A - \bar{\boldsymbol{\pi}}_B \cdot U^A$$

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Implementation: WT find transfer $K : \cup_{A \in \mathcal{D}} \{\bar{\pi}_A\} \to \mathbb{R} \cup \{\infty\}$ s.t. makes info choice IC:

$$\bar{\pi}_A \cdot U^A - K(\bar{\pi}_A) \geq \bar{\pi}_B \cdot U^A - K(\bar{\pi}_B)$$

  $\forall A, B \in \mathcal{D}$; NIAC provides condition (Rockafellar '70)

▹ Skip

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Implementation: WT find transfer $K : \cup_{A \in \mathcal{D}} \{\bar{\pi}_A\} \to \mathbb{R} \cup \{\infty\}$ s.t. makes info choice IC:

$$\bar{\pi}_A \cdot U^A - K(\bar{\pi}_A) \geq \bar{\pi}_B \cdot U^A - K(\bar{\pi}_B)$$

  $\forall A, B \in \mathcal{D}$; NIAC provides condition (Rockafellar '70)

  Pick arbitrary A_0; let

  $T(A) := \sup_{\text{all chains } A_0 \text{ to } A=A_k} \sum_{n=0}^{k-1} \bar{\pi}_{A_i} \cdot U^{A_{i+1}} - \bar{\pi}_{A_i} \cdot U^{A_i}$

▹ Skip

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Pick arbitrary A_0; let

  $$T(A) := \sup_{\text{all chains } A_0 \text{ to } A=A_k} \sum_{n=0}^{k-1} \bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}$$

  NIAC implies (i) $T(A_0) = 0$; **(ii)** $T(A_0) \geq T(A_k) + \bar{\pi}_{A_k} \cdot U^{A_0} - \bar{\pi}_{A_k} \cdot U^{A_k}$

  $$\begin{aligned} T(A_k) + \bar{\pi}_{A_k} \cdot U^{A_0} - \bar{\pi}_{A_k} \cdot U^{A_k} &= (\bar{\pi}_{A_0} \cdot U^{A_1} - \bar{\pi}_{A_0} \cdot U^{A_0}) + (\bar{\pi}_{A_1} \cdot U^{A_2} - \bar{\pi}_{A_1} \cdot U^{A_1}) + \cdots \\ &\quad + (\bar{\pi}_{A_{k-1}} \cdot U^{A_k} - \bar{\pi}_{A_{k-1}} \cdot U^{A_{k-1}}) + (\bar{\pi}_{A_k} \cdot U^{A_0} - \bar{\pi}_{A_k} \cdot U^{A_k}) \\ &= -\sum_{n=0}^{k} \left(\bar{\pi}_{A_n} \cdot U^{A_n} - \bar{\pi}_{A_n} \cdot U^{A_{n+1 \,(\text{mod } k+1)}} \right) \leq 0 = T(A_0) \end{aligned}$$

  i.e. $T(A_i)$ is bounded

▹ Skip

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Pick arbitrary A_0; let

  $T(A) := \sup_{\text{all chains } A_0 \text{ to } A=A_k} \sum_{n=0}^{k-1} \bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}$

  NIAC implies (iii) $\forall A_i, A_j$, $T(A_i) \geq T(A_j) + \bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j}$

  Suppose for chain $(A_0, A_1, ..., A_k)$ with $A_k = A_j$, $T(A_j) := \sum_{n=0}^{k-1} \bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}$

  If $\nexists \ell : A_\ell = A_i$, then set $A_{k+1} = A_i$ and note

  $T(A_i) \geq \sum_{n=0}^{k-1} \left(\bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}\right) + \bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j} = \sum_{n=0}^{k} \left(\bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}\right)$

▹ Skip

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

Pick arbitrary A_0; let

$T(A) := \sup_{\text{all chains } A_0 \text{ to } A=A_k} \sum_{n=0}^{k-1} \bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}$

NIAC implies (iii) $\forall A_i, A_j$, $T(A_i) \geq T(A_j) + \bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j}$

Suppose for chain $(A_0, A_1, ..., A_k)$ with $A_k = A_j$, $T(A_j) := \sum_{n=0}^{k-1} \bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}$

If $\exists \ell : A_\ell = A_i$, then $T(A_i) \geq \sum_{n=0}^{\ell-1} \left(\bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n} \right)$

Then

$$T(A_j) - T(A_i) \leq \sum_{n=\ell}^{k-1} \left(\bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n} \right)$$

$$\Longleftrightarrow T(A_j) - T(A_i) + (\bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j}) \leq \sum_{n=\ell}^{k-1} \left(\bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n} \right) + (\bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j})$$

$$= (\bar{\pi}_{A_i} \cdot U^{A_{\ell+1}} - \bar{\pi}_{A_i} \cdot U^{A_i}) + \cdots + (\bar{\pi}_{A_{k-1}} \cdot U^{A_j} - \bar{\pi}_{A_{k-1}} \cdot U^{A_{k-1}}) + (\bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j})$$

$$\leq 0 \qquad \text{(by NIAC)}$$

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Implementation: WT find transfer $K : \cup_{A \in \mathcal{D}} \{\bar{\pi}_A\} \to \mathbb{R} \cup \{\infty\}$ s.t. makes info choice IC:

$$\bar{\pi}_A \cdot U^A - K(\bar{\pi}_A) \geq \bar{\pi}_B \cdot U^A - K(\bar{\pi}_B)$$

  $\forall A, B \in \mathcal{D}$; NIAC provides condition (Rockafellar '70)

  Pick arbitrary A_0; let

  $T(A) := \sup_{\text{all chains } A_0 \text{ to } A=A_k} \sum_{n=0}^{k-1} \bar{\pi}_{A_n} \cdot U^{A_{n+1}} - \bar{\pi}_{A_n} \cdot U^{A_n}$

  NIAC implies $\forall A_i, A_j, \quad T(A_i) \geq T(A_j) + \bar{\pi}_{A_j} \cdot U^{A_i} - \bar{\pi}_{A_j} \cdot U^{A_j}$

  Set $K(\bar{\pi}_{A_j}) := \bar{\pi}_{A_j} \cdot U^{A_j} - T(A_j)$

▹ Skip

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Extend K to mixtures of revealed info structures

  Define $S := \cup_{A \in \mathcal{D}} S(\bar{\pi}_A)$ and consider $\bar{\pi}_A : \Theta \to \Delta(S)$

  Define $co(\{\bar{\pi}_A \mid A \in \mathcal{D}\}) := \{\pi \in \Pi \mid \pi = \sum_{A \in \mathcal{D}} \lambda(A)\bar{\pi}_A \text{ for some } \lambda \in \Delta(\mathcal{D})\}$

  Extend K to $co(\{\bar{\pi}_A \mid A \in \mathcal{D}\})$ as $K(\sum_A \lambda(A)\bar{\pi}_A) = \sum_A \lambda(A)K(\bar{\pi}_A)$

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Extend K to garblings of mixtures

  Let $\bar{\Pi}$ be the set of garblings of mixtures. For any $\pi' \in \mathcal{B}co(\{\bar{\pi}_A \mid A \in \mathcal{D}\})$, define

$$K(\pi') := \min\{K(\pi) \mid \pi \in co(\{\bar{\pi}_A \mid A \in \mathcal{D}\}) \text{ and } \pi \text{ is sufficient for } \pi'\}$$

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Extend K to garblings of mixtures

  Let $\bar{\Pi}$ be the set of garblings of mixtures. For any $\pi' \in \mathcal{B}co(\{\bar{\pi}_A \mid A \in \mathcal{D}\})$, define

  $$K(\pi') := \min\{K(\pi) \mid \pi \in co(\{\bar{\pi}_A \mid A \in \mathcal{D}\}) \text{ and } \pi \text{ is sufficient for } \pi'\}$$

  Extend K to Π: set $K(\pi) := \infty$ for all $\pi \in \Pi \setminus \bar{\Pi}$

  Easy to check that IC still satisfied: $\pi' \in \bar{\Pi}$ less informative than some $\pi = \sum_A \lambda(A)\bar{\pi}_A$

  and with $K(\pi') = K(\pi)$

  $\Longrightarrow \pi' \cdot U^B - K(\pi') = \pi' \cdot U^B - K(\pi) \leq \pi \cdot U^B - K(\pi) \leq \bar{\pi}_B U^B - K(\bar{\pi}_B)$ for any $B \in \mathcal{D}$

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

Extend K to garblings of mixtures

Let $\bar{\Pi}$ be the set of garblings of mixtures. For any $\pi' \in \mathcal{B}co(\{\bar{\pi}_A \mid A \in \mathcal{D}\})$, define

$$K(\pi') := \min\{K(\pi) \mid \pi \in co(\{\bar{\pi}_A \mid A \in \mathcal{D}\}) \text{ and } \pi \text{ is sufficient for } \pi'\}$$

Extend K to Π: set $K(\pi) := \infty$ for all $\pi \in \Pi \setminus \bar{\Pi}$

Easy to check that IC still satisfied: $\pi' \in \bar{\Pi}$ less informative than some $\pi = \sum_A \lambda(A)\bar{\pi}_A$

and with $K(\pi') = K(\pi)$

$\Longrightarrow \pi' \cdot U^B - K(\pi') = \pi' \cdot U^B - K(\pi) \leq \pi \cdot U^B - K(\pi) \leq \bar{\pi}_B U^B - K(\bar{\pi}_B)$ for any $B \in \mathcal{D}$

K convex and monotone in Blackwell order;

normalization $K(\pi) = 0$ for uninformative structures is straightforward

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Last step: construct $\boldsymbol{\sigma}_A : S(\boldsymbol{\pi}_A) \to \Delta(A)$ s.t. optimal choices

  $\forall m \in S(\boldsymbol{\pi}_A)$, let

$$\boldsymbol{\sigma}_A(a|m) := \begin{cases} \frac{P_A(a)}{\sum_{b \in A \,|\, \bar{m}^b = m} P_A(b)}, & \text{if } m = \bar{m}^a \\ 0, & \text{if otherwise} \end{cases}$$

Representation Result

Proof Sketch

- NIAC + NIAS $\Longrightarrow$ Additive CIA rep w/ properties

  Last step: construct $\boldsymbol{\sigma}_A : S(\boldsymbol{\pi}_A) \to \Delta(A)$ s.t. optimal choices

  $\forall m \in S(\boldsymbol{\pi}_A)$, let

$$\boldsymbol{\sigma}_A(a|m) := \begin{cases} \frac{P_A(a)}{\sum_{b \in A \,|\, \bar{m}^b = m} P_A(b)}, & \text{if } m = \bar{m}^a \\ 0, & \text{if otherwise} \end{cases}$$

  Note NIAS ensures $\boldsymbol{\sigma}_A$ is optimal

  $\sum_\theta \boldsymbol{\mu}(\theta) P_A(a \mid \theta)[u(a,\theta) - u(b,\theta)] \geq 0 \Longrightarrow \sum_\theta \bar{m}^a(\theta)[u(a,\theta) - u(b,\theta)] \geq 0$

  Can show that $\boldsymbol{\sigma}_A \circ \bar{\boldsymbol{\pi}}_A = P_A$ by construction: data are matched

  Done!

Optimal Information Structure with Additive Costs

DM chooses (π, σ)

(1) Distrib actions given beliefs: $\sigma : \Delta(\Theta) \to \Delta(A)$

$$\Sigma^*(A) := \left\{ \text{supp}(\sigma(m)) \in \arg\max_{a \in A} \mathbb{E}_m[u(a, \theta)],\ \forall m \in \Delta(\Theta) \right\}$$

(2) Distrib posterior beliefs given state: $\pi : \Theta \to \Delta(\Delta(\Theta))$ s.t. posterior distrib. MPS of prior: $\sum_\theta \mu(\theta) \sum_m \pi(m|\theta) m = \mathbb{E}_\mu[\mathbb{E}_{\pi(\cdot|\theta)}[m]] = \mu$

$$\Pi^*(A, \sigma) := \arg\max_{\pi \in \Pi} \mathbb{E}_\mu[\mathbb{E}_{\pi(\cdot|\theta)}[\mathbb{E}_m[u(\sigma(m), \theta)]]] - K(\pi; \mu)$$

Optimal Information Structure with Additive Costs

DM chooses (π, σ)

(1) Distrib actions given beliefs: $\sigma : \Delta(\Theta) \to \Delta(A)$

$$\Sigma^*(A) := \left\{ \text{supp}(\sigma(m)) \in \arg\max_{a \in A} \mathbb{E}_m[u(a,\theta)],\ \forall m \in \Delta(\Theta) \right\}$$

(2) Distrib posterior beliefs given state: $\pi : \Theta \to \Delta(\Delta(\Theta))$ s.t. posterior distrib. MPS of prior: $\sum_\theta \mu(\theta) \sum_m \pi(m|\theta) m = \mathbb{E}_\mu[\mathbb{E}_{\pi(\cdot|\theta)}[m]] = \mu$

$$\Pi^*(A, \sigma) := \arg\max_{\pi \in \Pi} \mathbb{E}_\mu[\mathbb{E}_{\pi(\cdot|\theta)}[\mathbb{E}_m[u(\sigma(m), \theta)]]] - K(\pi; \mu)$$

Lemma (Matejka & McKay '15 AER)

If K is strictly (weakly) increasing in $\succeq_B$ (it is wloo that) $\sigma \in \Sigma^*(A), \pi \in \Pi^*(A, \sigma)$ sat.:

$\nexists a, m, m'$ s.t. ((i) $\sigma(a|m) \in (0,1)$, and (ii)) $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

Intuition: at most one posterior belief per action chosen with positive probability; deterministic tie-breaking

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\boldsymbol{\sigma}(a|m)\boldsymbol{\pi}(m), \boldsymbol{\sigma}(a|m')\boldsymbol{\pi}(m') > \mathbf{0}$, and $m \neq m'$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

Note: $\tilde{m}(\theta) = \dfrac{\mu(\theta)\pi(\tilde{m}|\theta)}{\pi(\tilde{m})}$. Let $\delta := \dfrac{\pi(m)}{\pi(m) + \pi(m')}$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

Note: $\tilde{m}(\theta) = \dfrac{\mu(\theta)\pi(\tilde{m}|\theta)}{\pi(\tilde{m})}$. Let $\delta := \dfrac{\pi(m)}{\pi(m) + \pi(m')}$

Define $\pi' : \Theta \to \Delta(\Delta(\Theta))$:

$\pi'(\tilde{m}|\theta) := \pi(\tilde{m}|\theta) \quad \forall \tilde{m} \neq m, m'$, and $\pi'(m''|\theta) = \pi'(m|\theta) + \pi'(m'|\theta)$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

Note: $\tilde{m}(\theta) = \dfrac{\mu(\theta)\pi(\tilde{m}|\theta)}{\pi(\tilde{m})}$. Let $\delta := \dfrac{\pi(m)}{\pi(m) + \pi(m')}$

Define $\pi' : \Theta \to \Delta(\Delta(\Theta))$:

$\pi'(\tilde{m}|\theta) := \pi(\tilde{m}|\theta) \quad \forall \tilde{m} \neq m, m'$, and $\pi'(m''|\theta) = \pi'(m|\theta) + \pi'(m'|\theta)$

Then $m''(\theta) = \dfrac{\mu(\theta)(\pi(m|\theta) + \pi(m'|\theta))}{\pi(m) + \pi(m')} = \dfrac{\mu(\theta)\pi'(m''|\theta)}{\pi'(m'')}$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

Note: $\tilde{m}(\theta) = \dfrac{\mu(\theta)\pi(\tilde{m}|\theta)}{\pi(\tilde{m})}$. Let $\delta := \dfrac{\pi(m)}{\pi(m) + \pi(m')}$

Define $\pi' : \Theta \to \Delta(\Delta(\Theta))$:

$\pi'(\tilde{m}|\theta) := \pi(\tilde{m}|\theta) \quad \forall \tilde{m} \neq m, m'$, and $\pi'(m''|\theta) = \pi'(m|\theta) + \pi'(m'|\theta)$

Then $m''(\theta) = \dfrac{\mu(\theta)(\pi(m|\theta) + \pi(m'|\theta))}{\pi(m) + \pi(m')} = \dfrac{\mu(\theta)\pi'(m''|\theta)}{\pi'(m'')}$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m}) \; \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1-\delta)m'$

Note: $\tilde{m}(\theta) = \dfrac{\mu(\theta)\pi(\tilde{m}|\theta)}{\pi(\tilde{m})}$. Let $\delta := \dfrac{\pi(m)}{\pi(m)+\pi(m')}$

Define $\pi' : \Theta \to \Delta(\Delta(\Theta))$:

$\pi'(\tilde{m}|\theta) := \pi(\tilde{m}|\theta) \quad \forall \tilde{m} \neq m, m'$, and $\pi'(m''|\theta) = \pi'(m|\theta) + \pi'(m'|\theta)$

Then $m''(\theta) = \dfrac{\mu(\theta)(\pi(m|\theta)+\pi(m'|\theta))}{\pi(m)+\pi(m')} = \dfrac{\mu(\theta)\pi'(m''|\theta)}{\pi'(m'')}$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m}) \; \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

(3) Verifying $\sigma' \in \Sigma^*(A)$

For $\tilde{m} = m, m'$, $\sigma(a|\tilde{m}) > 0$, $\mathbb{E}_{\tilde{m}}[u(a,\theta)] \geq \mathbb{E}_{\tilde{m}}[u(a',\theta)] \; \forall a' \in A$

By linearity, same holds for $m'' = \delta m + (1-\delta)m'$ and $\sigma' \in \Sigma^*(A)$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m})\ \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

(3) Verifying $\sigma' \in \Sigma^*(A)$; (4) Verifying $\pi' \in \Pi^*(A, \sigma')$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m})\ \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

(3) Verifying $\sigma' \in \Sigma^*(A)$; (4) Verifying $\pi' \in \Pi^*(A, \sigma')$

$$\pi(m)\mathbb{E}_m[u(\sigma(m), \theta)] + \pi(m')\mathbb{E}_{m'}[u(\sigma(m'), \theta)] = \pi'(m'')\mathbb{E}_{m''}[u(\sigma'(m''), \theta)]$$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1-\delta)m'$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m})\ \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

(3) Verifying $\sigma' \in \Sigma^*(A)$; (4) Verifying $\pi' \in \Pi^*(A, \sigma')$

$$\pi(m)\mathbb{E}_m[u(\sigma(m),\theta)] + \pi(m')\mathbb{E}_{m'}[u(\sigma(m'),\theta)] = \pi'(m'')\mathbb{E}_{m''}[u(\sigma'(m''),\theta)]$$

$$G^A(\pi) = \sum_{\tilde{m}\neq m,m'} \pi(\tilde{m})\mathbb{E}_{\tilde{m}}[u(\sigma(\tilde{m}),\theta)] + \sum_{\tilde{m}=m,m'} \pi(\tilde{m})\mathbb{E}_{\tilde{m}}[u(\sigma(\tilde{m}),\theta)]$$

$$= \sum_{\tilde{m}\neq m''} \pi'(\tilde{m})\mathbb{E}_{\tilde{m}}[u(\sigma'(\tilde{m}),\theta)] + \pi'(m'')\mathbb{E}_{m''}[u(\sigma'(m''),\theta)] = G^A(\pi')$$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m})\ \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

(3) Verifying $\sigma' \in \Sigma^*(A)$; (4) Verifying $\pi' \in \Pi^*(A, \sigma')$

$$\pi(m)\mathbb{E}_m[u(\sigma(m), \theta)] + \pi(m')\mathbb{E}_{m'}[u(\sigma(m'), \theta)] = \pi'(m'')\mathbb{E}_{m''}[u(\sigma'(m''), \theta)]$$

$$G^A(\pi) = \sum_{\tilde{m} \neq m, m'} \pi(\tilde{m})\mathbb{E}_{\tilde{m}}[u(\sigma(\tilde{m}), \theta)] + \sum_{\tilde{m} = m, m'} \pi(\tilde{m})\mathbb{E}_{\tilde{m}}[u(\sigma(\tilde{m}), \theta)]$$

$$= \sum_{\tilde{m} \neq m''} \pi'(\tilde{m})\mathbb{E}_{\tilde{m}}[u(\sigma'(\tilde{m}), \theta)] + \pi'(m'')\mathbb{E}_{m''}[u(\sigma'(m''), \theta)] = G^A(\pi')$$

Optimal Information Structure with Additive Costs

Proof

Suppose not. Then, $\exists a, m, m'$ s.t. $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

(1) Defining garbled signal: $m'' := \delta m + (1 - \delta)m'$

(2) Defining strategy: $\sigma' : \Delta(\Theta) \to \Delta(A)$ s.t. $\sigma'(\tilde{m}) = \sigma(\tilde{m})\ \forall \tilde{m} \neq m''$, and $\sigma'(a|m'') = 1$

(3) Verifying $\sigma' \in \Sigma^*(A)$; (4) Verifying $\pi' \in \Pi^*(A, \sigma')$

$G^A(\pi) = G^A(\pi')$

As $\pi >_B \pi'$ and K is increasing in $\geq_B$, $G^A(\pi) - K(\pi) \leq G^A(\pi') - K(\pi')$

If K strictly increasing, contradicts optimality of π

Result extends to $\Theta \subseteq \mathbb{R}^n$

Distinguishing Multiplicative and Additive Costs

Menu (= decision problem) A yields **higher gross return from information** than menu B if

$\boldsymbol{\pi}, \boldsymbol{\pi}'$ s.t. $\boldsymbol{\pi} >_B \boldsymbol{\pi}'$, $\boldsymbol{\pi} \cdot U^A - \boldsymbol{\pi}' \cdot U^A > \boldsymbol{\pi} \cdot U^B - \boldsymbol{\pi}' \cdot U^B$

($\Longleftrightarrow U^A - U^B : \Delta(\Theta) \to \mathbb{R}$ is strictly convex)

Distinguishing Multiplicative and Additive Costs

Menu (= decision problem) A yields **higher gross return from information** than menu B if

$\boldsymbol{\pi}, \boldsymbol{\pi}'$ s.t. $\boldsymbol{\pi} >_B \boldsymbol{\pi}'$, $\boldsymbol{\pi} \cdot U^A - \boldsymbol{\pi}' \cdot U^A > \boldsymbol{\pi} \cdot U^B - \boldsymbol{\pi}' \cdot U^B$

($\iff U^A - U^B : \Delta(\Theta) \to \mathbb{R}$ is strictly convex)

Proposition (Chambers, Liu, & Rehbeck, '20 JET)

Fix $\boldsymbol{\mu}$ and $u \geq \mathbf{0}$ and let $(\mathcal{D}, P)$ be a state-dependent stochastic choice dataset with $\mathcal{D} = \{A, B\}$. Suppose A yields higher gross return from information than B.

(i) If $\bar{\boldsymbol{\pi}}_B >_B \bar{\boldsymbol{\pi}}_A$, then $(\mathcal{D}, P)$ cannot sat. NIAC.

(ii) If $U^A > U^B$, $\bar{\boldsymbol{\pi}}_B >_B \bar{\boldsymbol{\pi}}_A$, and $(\mathcal{D}, P)$ sat. NIAS, then it sat. GACI

Intuition (my reading)

(i) Additive costs: if higher return to info at A than B, then cannot choose + informative info structure at B than A

(ii) General nonseparable costs: people can do weird things

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
   - Homothetic Axiom of Costly Information (HACI)
   - Representation Result
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Multiplicative Cost of Information

Cost of information structure $R : \Pi \to \mathbb{R}_+$ with $R(\pi) > 0$ for some $\pi \in \Pi$

$\mathcal{R}$: set of all such cost functions

Multiplicative cost: with loss

Lower R, higher cost; zero R as if infinite cost; model hard constraints

e.g. bound on info acquired, restrictions on class of information structures

Optimal info structures

$R(\pi)G(A, \pi) = R(\pi)\pi \cdot U^A$

$$\Pi^*(K, A) := \arg\max_{\pi \in \Pi} R(\pi)\pi \cdot U^A$$

Note: restrict $u \geq 0$ to make sense

Optimality of Information Structures

Idea: Homothetic preferences!

Definition

Given μ and $u \geq 0$, let $\mathcal{D}_0 := \{A \in \mathcal{D} \mid \mathbb{E}_\mu[u(a,\theta) = 0, \forall a \in A]\}$. $(\mathcal{D}, P)$ sat. **homothetic axiom of costly information** (HACI) if for all sequences $(\bar{\pi}_{A_1}, U^{A_1}), (\bar{\pi}_{A_2}, U^{A_2}), ..., (\bar{\pi}_{A_k}, U^{A_k})$ with $A_i \in \mathcal{D} \setminus \mathcal{D}_0$, $\prod_{i=1}^{k} \frac{\bar{\pi}_{A_i} \cdot U^{A_{i+1}}}{\bar{\pi}_i \cdot U^{A_i}} \leq 1 \pmod{k}$

Representation Result

Theorem (Chambers, Liu, & Rehbeck, '20 JET)

Fix $\boldsymbol{\mu}$ and $u \geq \mathbf{0}$ and let $(\mathcal{D}, P)$ be a state-dependent stochastic choice dataset. The following are equivalent:

(i) $(\mathcal{D}, P)$ sat. NIAS and HACI

(ii) $(\mathcal{D}, P)$ admits a multiplicative costly information acquisition representation

(iii) $(\mathcal{D}, P)$ admits a multiplicative costly information acquisition representation with $R \in \mathcal{R}$ nonincreasing in Blackwell order and concave

Restrictions on R come for free; i.e. are untestable

Can also normalize R s.t. $R(\boldsymbol{\pi}) = \mathbf{1}$ for uninformative info structures

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
   - Binary Axiom of Costly Information (BACI)
   - Representation Result
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Constrained Information Acqusition

Feasible information structures $\Pi_C \subseteq \Pi$ with $\Pi_C \neq \emptyset$

Definition

Given μ and u, $(\mathcal{D}, P)$ sat. **binary axiom of costly information** (BACI) if for all $A, B \in \mathcal{D}$,

$\bar{\pi}_A \cdot U^A \geq \bar{\pi}_B \cdot U^A$

Note: BACI implies NIAC (same as $K(\pi) = 0$ if $\pi \in \Pi_C$ and $K(\pi) = \infty$ if ow) and HACI ($R(\pi) = 1_{\pi \in \Pi_C}$)

Representation Result

Theorem (Chambers, Liu, & Rehbeck, '20 JET)

Fix μ and u and let $(\mathcal{D}, P)$ be a state-dependent stochastic choice dataset. The following are equivalent:

(i) $(\mathcal{D}, P)$ sat. NIAS and BACI

(ii) $(\mathcal{D}, P)$ admits a constrained information acquisition representation

(iii) $(\mathcal{D}, P)$ admits a constrained information acquisition representation with Π_C convex

wlog to consider Π_C convex

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

From Information Acquisition to Stochastic Choice

Stochastic choice: data looks random

Random Utility

- Grand set of actions $\mathcal{A}$
- Utility randomly drawn $u : \mathcal{A} \to \mathbb{R}$ from some set $\mathcal{U}$; DM chooses action to max realized utility
  - Mistakes, trembles, etc
- Can think $\Theta = \mathcal{U}$, $u(a, \theta) = u(a)$ for realized utility
- Induces stochastic choice data $P_A : \Theta \to \Delta(A)$
- *State-dependent* Menu-monotonicity: $P_A(a \mid \theta) \geq P_{A \cup \{b\}}(a \mid \theta)$
  - (usually lit refers to state-independent menu-motonicity)

CIA allows for violations of state-dependent menu-monotonicity, for all states

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
   - Introduction to Entropy
   - Characterizing Choice with Entropy Costs
   - Entropy and Posterior Separable Costs
   - Entropy and MLRP-ordered Signals
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
   - Introduction to Entropy
   - Characterizing Choice with Entropy Costs
   - Entropy and Posterior Separable Costs
   - Entropy and MLRP-ordered Signals
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Entropy

Particular functional form for information cost based on (Shanon) entropy

Popular in applied literature and applications

Originated in information theory (Shannon ‘48)

(good reference: Cover & Thomas, ‘06)

Entropy

Entropy: $H(X) := -\mathbb{E}_{P_X}[\ln P_X(X)]; \quad p \in \Delta(X)$

measure amount of information in X; convention $0 \ln 0 = 0$

Entropy

Entropy: $H(X) := -\mathbb{E}_{P_X}[\ln P_X(X)]; \quad p \in \Delta(X)$

measure amount of information in X; convention $0 \ln 0 = 0$

Joint Entropy: $H(X, Y) = -\mathbb{E}_{P_{X,Y}}[\ln P_{X,Y}(X, Y)]$

Entropy

Entropy: $H(X) := -\mathbb{E}_{P_X}[\ln P_X(X)]; \quad p \in \Delta(X)$

measure amount of information in X; convention $0 \ln 0 = 0$

Joint Entropy: $H(X, Y) = -\mathbb{E}_{P_{X,Y}}[\ln P_{X,Y}(X, Y)]$

$\mathbb{E}$ Conditional Entropy $H(Y|X) := \sum_x P_X(x) H(Y|x)$

Note: if $X \perp\!\!\!\perp Y$, then $H(Y|X) = H(Y)$

Entropy

Entropy: $H(X) := -\mathbb{E}_{P_X}[\ln P_X(X)]; \quad p \in \Delta(X)$

measure amount of information in X; convention $0 \ln 0 = 0$

Joint Entropy: $H(X, Y) = -\mathbb{E}_{P_{X,Y}}[\ln P_{X,Y}(X, Y)]$

$\mathbb{E}$ Conditional Entropy $H(Y|X) := \sum_x P_X(x) H(Y|x)$

Note: if $X \perp\!\!\!\perp Y$, then $H(Y|X) = H(Y)$

Entropy Chain Rule: $H(X, Y) = H(X) + H(Y|X); \quad H((X, Y)|Z) = H(X|Z) + H(Y|X, Z)$

$H(X_1, ..., X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i \mid X_1, ..., X_{i-1})$

Entropy

Entropy: $H(X) := -\mathbb{E}_{P_X}[\ln P_X(X)]; \quad p \in \Delta(X)$

measure amount of information in X; convention $0 \ln 0 = 0$

Joint Entropy: $H(X,Y) = -\mathbb{E}_{P_{X,Y}}[\ln P_{X,Y}(X,Y)]$

$\mathbb{E}$ Conditional Entropy $H(Y|X) := \sum_x P_X(x) H(Y|x)$

Note: if $X \perp\!\!\!\perp Y$, then $H(Y|X) = H(Y)$

Entropy Chain Rule: $H(X,Y) = H(X) + H(Y|X); \quad H((X,Y)|Z) = H(X|Z) + H(Y|X,Z)$

$H(X_1,...,X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i \mid X_1,...,X_{i-1})$

Theorem (Khinchin, '57)

H is unique function defined on arbitrary finite-dimensional simplices that

(i) depends only on prob distr (not values): $H(X) = H(p)$

(ii) max at uniform distr: $\max_{p \in \Delta^n} H(p) = H((1/n, 1/n, ..., 1/n))$

(iii) is unaffected by adding zero prob state: $H((p_1, p_2, ..., p_n)) = H((p_1, p_2, ..., p_n, 0))$

(iv) is additive: $(X,Y) \sim P_{XY}, H(X,Y) = H(X) + H(Y|X)$

KL and Mutual Information

Kullback–Leibler divergence (KL) ≡ Relative Entropy: $D(p||q) = \mathbb{E}_p[\ln(p(X)/q(X))]$,

with $\mathbf{0} \cdot \ln(\mathbf{0}/\mathbf{0}) = \mathbf{0} \ln \mathbf{0} = \mathbf{0}$, and $p \cdot \ln(p/\mathbf{0}) = \infty$ for $p > \mathbf{0}$

KL is basis of Bayesian inference!

$D(p||q) \geq \mathbf{0}$; w/ = iff $p = q$

Not symmetric: $D(p||q)$ in general $\neq D(q||p)$

KL and Mutual Information

Kullback–Leibler divergence (KL) ≡ Relative Entropy: $D(p||q) = \mathbb{E}_p[\ln(p(X)/q(X))]$,

with $\mathbf{0} \cdot \ln(\mathbf{0}/\mathbf{0}) = \mathbf{0} \ln \mathbf{0} = \mathbf{0}$, and $p \cdot \ln(p/\mathbf{0}) = \infty$ for $p > \mathbf{0}$

KL is basis of Bayesian inference!

$D(p||q) \geq \mathbf{0}$; w/ = iff $p = q$

Not symmetric: $D(p||q)$ in general $\neq D(q||p)$

$\mathbb{E}$ Conditional KL $D(p_{Y|X}||q_{Y|X}) := \mathbb{E}_{x\sim p_X}[D(p_{Y|X}||q_{Y|X})]$

KL Chain Rule: $D(p_{X,Y}||q_{X,Y}) = D(p_X||q_X) + D(p_{Y|X}||q_{Y|X})$

Entropy and Mutual Information

Mutual Information (MI): $I(X;Y) = D(p_{X,Y}||p_X p_Y) = \sum_{x,y} p_{X,Y}(x,y)\left[\ln\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right)\right]$

how much info rv contains about another

reduction in uncertainty on rv due to knowledge of the other

Relation to Entropy

$I(X;X) = H(X)$

$I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$

Properties

Conditional MI: $I((X;Y)|Z) = H(X|Z) - H(X|Y,Z)$

MI Chain Rule: $I(X_1,...,X_n;Y) = I(X_1;Y) + \sum_{i=2}^{n} I((X_i;Y)|X_1,...,X_{i-1})$

$I(X;Y) \geq 0$; w/ = iff $X \perp\!\!\!\perp Y$

reduction in uncertainty on rv due to knowledge of the other

Entropy

Convexity/Concavity

Entropy $H(p)$ concave in p

KL $D(p||q)$ convex in (p, q)

MI $I(X;Y)$ concave in p_X given $p_{Y|X}$

$$I(X;Y) = H(Y) - H(Y|X) = H(\sum_x p_X(x) p_{Y|X}(\cdot|x)) - \sum_x p_X(x) H(p_{Y|X}(\cdot|x))$$

MI $I(X;Y)$ convex in $p_{Y|X}$ given p_X

$$I(X;Y) = D(p_{X,Y}||p_X p_Y) = D(p_{Y|X} p_X || \mathbb{E}_{p_X}[p_{Y|X}] p_X)$$

Entropy

More Properties

Conditioning reduces entropy: $H(X) \geq H(X|Y) \iff I(X;Y) \geq 0$

But conditioning *can increase* MI: possible $I(X;Y) < I((X;Y)|Z)$, e.g. $X \perp\!\!\!\perp Y$ and $Z = X + Y$

Transformation reduces entropy and MI: $H(f(X)) \leq H(X); \quad I(X;Y) \geq I(X;f(Y))$

No uncertainty, zero entropy: $H(f(X)|X) = 0$ ($f(X)$ is known given X)

Entropy of a sum: $X \perp\!\!\!\perp Y \implies H(X+Y) \geq \max\{H(X), H(Y)\}$ $H(X+Y|X) = H(Y|X)$

Independence bound on entropy: $H(X_1, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$ (via chain rule for H)

Entropy

Probability Bounds

if $X \perp\!\!\!\perp Y, X \sim p, Y \sim q$, then $\mathbb{P}(X = Y) \geq \max\{\exp(-H(p) - D(p||q)), \exp(-H(q) - D(q||p))\}$

if X, Y iid, then $\mathbb{P}(X = Y) \geq \max\{\exp(-H(p)\}$; w/ = iff unif distr

Derived Measures

Metric: $\delta(X;Y) := H(X,Y) - I(X;Y)$

Correlation measure: $I(X;Y)/H(X)$

Statistics

$\mu \in \Delta(\Theta)$; $p_\theta \in \Delta(X)$; T statistic of data X

Note: $I(\theta; T(X)) \leq I(\theta; X)$

Sufficient statistic: T s.t. $I(\theta; T(X)) = I(\theta; X)$

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
   - Introduction to Entropy
   - Characterizing Choice with Entropy Costs
   - Entropy and Posterior Separable Costs
   - Entropy and MLRP-ordered Signals
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Entropy Costs

Recall: CIA with additive costs

Given $\boldsymbol{\mu} \in \Delta(\boldsymbol{\Theta})$ and $u : A \times \boldsymbol{\Theta} \to \mathbb{R}$

$$\max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \boldsymbol{\pi} \cdot U^A - K(\boldsymbol{\pi})$$

Posterior beliefs: $m \sim \boldsymbol{\pi}(\cdot \mid \boldsymbol{\theta})$

Entropy Costs

Recall: CIA with additive costs

Given $\mu \in \Delta(\Theta)$ and $u : A \times \Theta \to \mathbb{R}$

$$\max_{\pi \in \Pi} \pi \cdot U^A - K(\pi)$$

Posterior beliefs: $m \sim \pi(\cdot \mid \theta)$

Entropy (MI) costs: $K(\pi; \mu) := \lambda I(\theta; m) = \lambda(H(\theta) - H(\theta|m))$

$\lambda > 0$: scalar that parameterizes cost

Cost measures reduction in prior entropy due to information acquired

Fully uninformative π: $m = \mu$; $\theta|m = \theta$ and thus $H(\theta|m) = H(\theta)$, $I(\theta; m) = 0$

Fully informative π: m is suff statistic for θ; $H(\theta|m) = 0$, $I(\theta; m) = H(\theta)$

For any π: $H(\theta) \geq I(\theta; m) \geq 0$

Characterizing Choice with Entropy Costs

Lemma (Matejka & McKay '15 AER)

If K is strictly (weakly) increasing in $\geq_B$ (it is wloo that) $\boldsymbol{\sigma} \in \Sigma^*(A), \boldsymbol{\pi} \in \Pi^*(A, \boldsymbol{\sigma})$ sat.:

$\nexists a, m, m'$ s.t. ((i) $\boldsymbol{\sigma}(a|m) \in (0, 1)$, and (ii)) $\boldsymbol{\sigma}(a|m)\boldsymbol{\pi}(m), \boldsymbol{\sigma}(a|m')\boldsymbol{\pi}(m') > 0$, and $m \neq m'$

Given $\boldsymbol{\pi}$, associate each $m \in S(\boldsymbol{\pi})$ with action a optimal at $m \in \Delta(\Theta)$; denoted m^a

Characterizing Choice with Entropy Costs

Lemma (Matejka & McKay '15 AER)

If K is strictly (weakly) increasing in $\succeq_B$ (it is wloo that) $\sigma \in \Sigma^*(A), \pi \in \Pi^*(A, \sigma)$ sat.:

$\nexists a, m, m'$ s.t. ((i) $\sigma(a|m) \in (0, 1)$, and (ii)) $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

Given π, associate each $m \in S(\pi)$ with action a optimal at $m \in \Delta(\Theta)$; denoted m^a

$$\sum_{\theta} \sum_{m} U^A(m)\pi(m|\theta)\mu(\theta) = \sum_{\theta} \sum_{a} U^A(m^a)\pi(m^a|\theta)\mu(\theta)$$

Characterizing Choice with Entropy Costs

Lemma (Matejka & McKay '15 AER)

If K is strictly (weakly) increasing in $\succeq_B$ (it is wloo that) $\sigma \in \Sigma^*(A), \pi \in \Pi^*(A, \sigma)$ sat.:

$\nexists a, m, m'$ s.t. ((i) $\sigma(a|m) \in (0,1)$, and (ii)) $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

Given π, associate each $m \in S(\pi)$ with action a optimal at $m \in \Delta(\Theta)$; denoted m^a

$$\sum_{\theta}\sum_{m} U^A(m)\pi(m|\theta)\mu(\theta) = \sum_{\theta}\sum_{a} U^A(m^a)\pi(m^a|\theta)\mu(\theta)$$

$$= \sum_{a} U^A(m^a)\sum_{\theta}\pi(m^a|\theta)\mu(\theta) = \sum_{a}\sum_{\theta} u(a,\theta)m^a(\theta)P_A(a)$$

Characterizing Choice with Entropy Costs

Lemma (Matejka & McKay '15 AER)

If K is strictly (weakly) increasing in $\succeq_B$ (it is wloo that) $\sigma \in \Sigma^*(A), \pi \in \Pi^*(A, \sigma)$ sat.:

$\nexists a, m, m'$ s.t. ((i) $\sigma(a|m) \in (0,1)$, and (ii)) $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

Given π, associate each $m \in S(\pi)$ with action a optimal at $m \in \Delta(\Theta)$; denoted m^a

$$\sum_\theta \sum_m U^A(m)\pi(m|\theta)\mu(\theta) = \sum_\theta \sum_a U^A(m^a)\pi(m^a|\theta)\mu(\theta)$$

$$= \sum_a U^A(m^a) \sum_\theta \pi(m^a|\theta)\mu(\theta) = \sum_a \sum_\theta u(a,\theta) m^a(\theta) P_A(a)$$

$$= \sum_a \sum_\theta u(a,\theta) P_A(\theta|a) P_A(a) = \sum_\theta \sum_a u(a,\theta) P_A(a|\theta)\mu(\theta)$$

Characterizing Choice with Entropy Costs

Lemma (Matejka & McKay '15 AER)

If K is strictly (weakly) increasing in $\succeq_B$ (it is wloo that) $\sigma \in \Sigma^*(A), \pi \in \Pi^*(A, \sigma)$ sat.:

$\nexists a, m, m'$ s.t. ((i) $\sigma(a|m) \in (0,1)$, and (ii)) $\sigma(a|m)\pi(m), \sigma(a|m')\pi(m') > 0$, and $m \neq m'$

Given π, associate each $m \in S(\pi)$ with action a optimal at $m \in \Delta(\Theta)$; denoted m^a

$$\sum_\theta \sum_m U^A(m)\pi(m|\theta)\mu(\theta) = \sum_\theta \sum_a U^A(m^a)\pi(m^a|\theta)\mu(\theta)$$

$$= \sum_a U^A(m^a) \sum_\theta \pi(m^a|\theta)\mu(\theta) = \sum_a \sum_\theta u(a,\theta) m^a(\theta) P_A(a)$$

$$= \sum_a \sum_\theta u(a,\theta) P_A(\theta|a) P_A(a) = \sum_\theta \sum_a u(a,\theta) P_A(a|\theta)\mu(\theta)$$

Can rewrite cost function:

$$K(\pi;\mu) = \lambda I(\theta; m^a) = \lambda(H(\theta) - H(\theta|m^a)) = \lambda(H(m) - H(m|\theta))$$

$$= \lambda(H(P_A(a)) - H(P_A(a|\theta))) = \lambda\left(\sum_\theta \mu(\theta) \sum_a P_A(a|\theta) \ln P_A(a|\theta) - \sum_a P_A(a) \ln P_A(a)\right)$$

Characterizing Choice with Entropy Costs

$$\max_{\pi \in \Pi} \pi \cdot U^A - K(\pi; \mu) = \max_{P_A : \Theta \to \Delta(A)} \mathbb{E}_\mu[\mathbb{E}_{P_A(\cdot|\theta)}[u(a, \theta)]] - \lambda(H(P_A(a)) - \mathbb{E}_\mu[H(P_A(a|\theta))])$$

$$\text{s.t. } P_A(a) = \mathbb{E}_\mu[P_A(a|\theta)]$$

Characterizing Choice with Entropy Costs

$$\max_{\pi \in \Pi} \pi \cdot U^A - K(\pi; \mu) = \max_{P_A : \Theta \to \Delta(A)} \mathbb{E}_\mu[\mathbb{E}_{P_A(\cdot|\theta)}[u(a, \theta)]] - \lambda(H(P_A(a)) - \mathbb{E}_\mu[H(P_A(a|\theta))])$$

$$\text{s.t. } P_A(a) = \mathbb{E}_\mu[P_A(a|\theta)]$$

Theorem 1 (Matejka & McKay '15 AER)

P_A is optimal only if $P_A(a|\theta) = \dfrac{P_A(a)\exp(u(a,\theta)/\lambda)}{\sum_b P_A(b)\exp(u(b,\theta)/\lambda)}$ a.s.

Characterizing Choice with Entropy Costs

$$\max_{\pi \in \Pi} \pi \cdot U^A - K(\pi; \mu) = \max_{P_A : \Theta \to \Delta(A)} \mathbb{E}_\mu[\mathbb{E}_{P_A(\cdot|\theta)}[u(a,\theta)]] - \lambda(H(P_A(a)) - \mathbb{E}_\mu[H(P_A(a|\theta))])$$

$$\text{s.t. } P_A(a) = \mathbb{E}_\mu[P_A(a|\theta)]$$

Theorem 1 (Matejka & McKay '15 AER)

P_A is optimal only if $P_A(a|\theta) = \dfrac{P_A(a)\exp(u(a,\theta)/\lambda)}{\sum_b P_A(b)\exp(u(b,\theta)/\lambda)}$ a.s.

Proposition 1 (Caplin, Dean & Leahy '19 REStud)

P_A is optimal iff sat. (1) MM condition, and (2) $\forall a \in A$

$$\sum_\theta \mu(\theta) \frac{\exp(u(a,\theta)/\lambda)}{\sum_b P_A(b)\exp(u(b,\theta)/\lambda)} \leq 1$$

with equality is $P_A(a) > 0$

Proof: constrained optimization

Computing a Solution

Blahut-Arimoto Algorithm (Cover & Thomas '12)

Ensures convergence

$$P_A^{n+1}(a|\theta) = \frac{P_A^n(a)\exp(u(a,\theta)/\lambda)}{\sum_{b\in A} P_A^n(b)\exp(u(b,\theta)/\lambda)}$$

$$P_A^{n+1}(a) = \left(\sum_{\theta} \mu(\theta)\frac{\exp(u(a,\theta)/\lambda)}{\sum_{b\in A} P_A^n(b)\exp(u(b,\theta)/\lambda)}\right) P_A^n(a)$$

Uniqueness of a Solution with Entropy Costs

Vectors $\{x_i\}_i$, $x_i \in \mathbb{R}^n$ are affinely independent if

$\forall \boldsymbol{\lambda}_i \in \mathbb{R}$ s.t. $\sum_i \boldsymbol{\lambda}_i = \mathbf{0}$ and $\sum_i \boldsymbol{\lambda}_i x_i = \mathbf{0}$, then $\boldsymbol{\lambda}_i = \mathbf{0}$ $\forall i$

Implied by linear independence

Uniqueness of a Solution with Entropy Costs

Vectors $\{x_i\}_i$, $x_i \in \mathbb{R}^n$ are affinely independent if

$\forall \boldsymbol{\lambda}_i \in \mathbb{R}$ s.t. $\sum_i \boldsymbol{\lambda}_i = \mathbf{0}$ and $\sum_i \boldsymbol{\lambda}_i x_i = \mathbf{0}$, then $\boldsymbol{\lambda}_i = \mathbf{0}$ $\forall i$

Implied by linear independence

Theorem 2 (Caplin & Dean '13 WP)

If $\{\{\exp(u(a, \boldsymbol{\theta})/\boldsymbol{\lambda})\}_{\boldsymbol{\theta}}, a \in A\}$ are affinely independent, there is a unique optimal $\boldsymbol{\pi}$

Characterizing Beliefs with Entropy Costs

Can also characterize optimal posterior beliefs

$\Longrightarrow$ Rewrite conditions in terms of m^a

Characterizing Beliefs with Entropy Costs

Can also characterize optimal posterior beliefs

$\Longrightarrow$ Rewrite conditions in terms of m^a

Proposition 2 (Caplin, Dean, & Leahy '19 REStud)

π is optimal iff

(1) $\sum_a \pi(m^a) m^a = \mu$

(2) [Invariant LR for Chosen Actions] $\forall a, b : P_A(a), P_A(b) > 0$

$$\frac{m^a(\theta)}{\exp(u(a,\theta)/\lambda)} = \frac{m^b(\theta)}{\exp(u(b,\theta)/\lambda)}$$

(3) [LR for Unchosen Actions] $\forall a, b : P_A(a) > 0 = P_A(b)$

$$\sum_\theta \frac{m^a(\theta)}{\exp(u(a,\theta)/\lambda)} \exp(u(b,\theta)/\lambda) \leq 1$$

Endogenous Consideration Sets

Caplin, Dean, & Leahy ('19 REStud) study endogenous consideration sets $\{a \mid P_A(a) > 0\}$

- Needle in haystack (3.1): only 1 action good, all others bad
- Independent values (3.2): $u(a, \cdot) \perp\!\!\!\perp u(b, \cdot)$; finite utility levels

See also Jung, Kim, Matejka, & Sims ('19 REStud)

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
   - Introduction to Entropy
   - Characterizing Choice with Entropy Costs
   - Entropy and Posterior Separable Costs
   - Entropy and MLRP-ordered Signals
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Posterior-Based Approach

Definition

$K(\pi;\mu)$ is a **posterior separable** cost function if for each $\mu \in \Delta(\Theta)$, $\exists k_\mu : \Delta(\Theta) \to \mathbb{R} \cup \{\infty\}$

s.t. $K(\pi;\mu) := \mathbb{E}_\pi[k_\mu(m)] - k_\mu(\mu)$

where k_μ is strictly convex and finite for $m \in \text{int}\Delta(\Theta)$, and $\pi(m) = \mathbb{E}_\mu[\pi(m|\theta)]$

$K(\pi;\mu)$ is a **uniformly posterior separable** cost function if $\exists k : \Delta(\Theta) \to \mathbb{R} \cup \{\infty\}$ s.t.

$\forall \mu \in \Delta(\Theta)$, $K(\pi;\mu) := \mathbb{E}_\pi[k(m)] - k(\mu)$

where k is strictly convex and finite for $m \in \text{int}\Delta(\Theta)$, and $\pi(m) = \mathbb{E}_\mu[\pi(m|\theta)]$

Caplin, Dean, & Leahy ('18 WP) provide representation results, building on Caplin & Dean ('15 AER)

Posterior-Based Approach

Definition

$K(\pi;\mu)$ **posterior separable**: $K(\pi;\mu) := \mathbb{E}_\pi[k_\mu(m)] - k_\mu(\mu)$

PS Costs = NIAS, NIAC (for additivity), technical restrictions (non-linearity, convexity, continuity, and a richness condition)

+ Separability Condition: makes cost of particular posterior independent from distrib of posteriors

Fix A and μ; P_A state-dep stochastic choice

$S(P_A)$:= actions taken with positive probability

(Revealed) posteriors for chosen actions m^a: $M(\mu, A)$

Can find problem $\tilde{A} = (A \setminus \{b\}) \cup \{c\}$ s.t. posteriors m^a the same for $a \neq b, c$

Posterior-Based Approach

Definition

$K(\pi;\mu)$ **uniformly posterior separable**: $K(\pi;\mu) := \mathbb{E}_\pi[k(m)] - k(\mu)$

UPS Costs = PS Costs + Locally Invariant Posteriors

Fix A and μ; $P_A : \Theta \to \Delta(A)$; $S(P_A)$:= actions taken with positive probability

(Revealed) posteriors for chosen actions m^a: $M(\mu, A)$

If revealed posteriors under μ still feasible under $\tilde{\mu}$ — i.e. $\tilde{\mu} \in \text{co}(M^A(\mu))$

then $M(\tilde{\mu}, A) = M(\mu, A)$

If different problem $\tilde{A} \subseteq S(P_A)$, but $\{m^a\}_{a\in\tilde{A}}$ still feasible given $\tilde{\mu}$,

then $M(\tilde{\mu}, A) = \{m^a\}_{a\in\tilde{A}}$

Only choice probabilities adjust so that $\tilde{\mu} = \sum_{a\in\tilde{A}} \tilde{P}_{\tilde{A}}(a) m^a$

Useful for problem-solving!

Posterior-Based Approach

Entropy Costs = UPS + Invariance Under Compression

States θ, θ' payoff equivalent if $u(a, \theta) = u(a, \theta') \; \forall a$

Invariance under compression: DM doesn't waste resources in distinguishing between payoff equivalent states; i.e. treats θ and θ' the same

Posterior-Based Approach

Entropy Costs = UPS + Invariance Under Compression

States θ, θ' payoff equivalent if $u(a, \theta) = u(a, \theta') \; \forall a$

Invariance under compression: DM doesn't waste resources in distinguishing between payoff equivalent states; i.e. treats θ and θ' the same

Reasonable? Appeals to maximizer in us, but

seems at odds with situation in which

(1) learning about state is informative about likelihood of nearby states,

(2) payoffs at states nearby θ may be very different from those near θ'

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
   - Introduction to Entropy
   - Characterizing Choice with Entropy Costs
   - Entropy and Posterior Separable Costs
   - Entropy and MLRP-ordered Signals
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Entropy and MLRP-ordered Signals

A, Θ ordered sets; $|\Theta| \geq 3$

$u : A \times \Theta$ sat. **(strict) increasing differences** if $\forall a' > a, \theta' > \theta$,

$u(a', \theta') - u(a, \theta')(>) \geq u(a', \theta) - u(a, \theta)$

π sat. **MLRP** if $\forall m, m' \in S(\pi)$, either

$$\frac{m'(\theta')}{m'(\theta)} \geq \frac{m(\theta')}{m(\theta)}$$

whenever $\theta' > \theta$, or vice-versa

K sat. uniform posterior separability, smoothness, invariance to relabeling states

Entropy and MLRP-ordered Signals

A, Θ ordered sets; $|\Theta| \geq 3$

$u : A \times \Theta$ sat. **(strict) increasing differences** if $\forall a' > a, \theta' > \theta$,

$u(a', \theta') - u(a, \theta')(>) \geq u(a', \theta) - u(a, \theta)$

π sat. **MLRP** if $\forall m, m' \in S(\pi)$, either

$$\frac{m'(\theta')}{m'(\theta)} \geq \frac{m(\theta')}{m(\theta)}$$

whenever $\theta' > \theta$, or vice-versa

K sat. uniform posterior separability, smoothness, invariance to relabeling states

Theorem (Mensch '19 WP)

Take assumptions as given. π sat. MLRP for all u sat. incresing differences, and all priors $\mu \in \Delta(\Theta)$ if and only if information costs are entropic

Entropy and MLRP-ordered Signals

Theorem (Mensch '19 WP)

Take assumptions as given. π sat. MLRP for all u sat. incresing differences, and all priors $\mu \in \Delta(\Theta)$ if and only if information costs are entropic

Intuition for sufficiency:

At optimal, MM15 imply: $u(a', \theta) - u(a, \theta) = \lambda(\ln m^{a'}(\theta) - \ln m^{a}(\theta))$

$\forall \theta$ and a', a chosen with positive probability

Increasing differences: $(\ln m^{a'}(\theta') - \ln m^{a}(\theta')) - (\ln m^{a'}(\theta) - \ln m^{a}(\theta)) \geq 0$

$\implies \frac{m^{a'}(\theta')}{m^{a'}(\theta)} \geq \frac{m^{a}(\theta')}{m^{a}(\theta)}$

Entropy and MLRP-ordered Signals

Theorem (Mensch '19 WP)

Take assumptions as given. π sat. MLRP for all u sat. incresing differences, and all priors $\mu \in \Delta(\Theta)$ if and only if information costs are entropic

Provide foundation for typically assumed use of MLRP signals

Result extends to A, Θ compact subsets of $\mathbb{R}$, with μ absolutely continuous, u continuous

Applications to games (see sections 4, 5)

- e.g. uniqueness of symmetric eqm of IPV 1PA and 2PA with 1st stage of info acq wrt own valuation, and eqm strategies sat. MLRP; revenue equiv holds

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Limitations of Entropy Costs

Example

WT invest if return on risky asset r greater than return on risk-free asset r_0

with entropy, prob choosing best action depends only on whether $r \lessgtr r_0$, not on distance $r - r_0$

Limitations of Entropy Costs

Example

WT invest if return on risky asset r greater than return on risk-free asset r_0

with entropy, prob choosing best action depends only on whether $r \lessgtr r_0$, not on distance $r - r_0$

WT detect mistakes on data entry on volume of sales

with entropy, prob detecting mistake depends on whether or not there was a mistake, not on whether the mistake was GPB 1 or GPB 1 bn

Limitations of Entropy Costs

Example

WT invest if return on risky asset r greater than return on risk-free asset r_0

with entropy, prob choosing best action depends only on whether $r \lessgtr r_0$, not on distance $r - r_0$

WT detect mistakes on data entry on volume of sales

with entropy, prob detecting mistake depends on whether or not there was a mistake, not on whether the mistake was GPB 1 or GPB 1 bn

No notion of similarity of states

"Entropy does not depend on a metric, i.e., the distance between states does not matter. With entropy, it is as difficult to distinguish the temperature of 10C from 20C, as 1C from 2C." (Mackowiak, Matejka, and Wiederholt, 2018)

Same holds for Tsallis entropy (Caplin, Dean, & Leahy, '18 WP; 8.3)

Limitations of Entropy Costs

Metric-dependent Additive Cost Functions

Pomatto, Strack, & Tamuz ('19 WP)

$$C(\pi) = \sum_{\theta,\tilde{\theta}\in\Theta} \beta_{\theta,\tilde{\theta}} D(\pi(\cdot \mid \theta)||\pi(\cdot \mid \tilde{\theta})), \quad \text{where } \beta_{\theta,\tilde{\theta}} \propto \frac{1}{(\theta-\tilde{\theta})^2}$$

Hebert & Woodford ('21 AER): Fisher information costs (limit result)

$$\mathbb{E}_{\mu}\left[\sum_{a|P_A(a)>0} \frac{\left(\frac{\partial}{\partial\theta}P_A(\theta|a)\right)^2}{P_A(\theta|a)}\right]$$

Other Formulations

Capacity constraints

Fixed information structures

As with utility functions (e.g. pervasiveness of CES and quasi-linear utility), whether assumptions on costs are reasonable or not, depends on application

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Implications of CIA for Choice

Strictly dominated actions never chosen

(true in general for CIA)

Adding duplicates to action *a* doesn't affect prob. choosing other actions

(true in general for CIA when info is about θ; not necessarily true for search models)

Increasing payoffs to action *a* increases *unconditional* $P_A(a)$ under entropy costs

(sequential CIA with fixed info structures can drop unconditional: Gonçalves '20 WP)

MM15 fn 18: claim that with exogenous info structure, making action better can make it chosen less often in expectation (wrt μ)

However, in example info structure changes as well

If info structure is truly fixed, this never happens

Implications of CIA for Choice

CIA allows for violations of state-dependent menu-monotonicity, for all states

State-dependent menu-monotonicity: $P_A(a \mid \boldsymbol{\theta}) \geq P_{A \cup \{b\}}(a \mid \boldsymbol{\theta})$

MM15, Prop. 6: Example of $P_A(a \mid \boldsymbol{\theta}) < P_{A \cup \{b\}}(a \mid \boldsymbol{\theta})$ $\forall \boldsymbol{\theta}$ with entropy costs

Contrastingly: sequential CIA with fixed info structures, state-dependent menu-monotoniticity holds! (Gonçalves '20 WP)

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Experimental Evidence

Choice in perception problems:

Dewan & Neligh ‘20 JET: subjects mostly satisfy NIAS and NIAC, and mostly responsive to incentives in terms of performance

Dean & Neligh ‘20 WP: support for CIA, but violations of RUM

In general entropy poorly fitted for perceptual problems

Caplin, Csaba, Leahy, & Nov ‘21 QJE: cool method to recover info costs experimentally

Experimental Evidence

Choice in perception problems:

Dewan & Neligh ‘20 JET: subjects mostly satisfy NIAS and NIAC, and mostly responsive to incentives in terms of performance

Dean & Neligh ‘20 WP: support for CIA, but violations of RUM

In general entropy poorly fitted for perceptual problems

Caplin, Csaba, Leahy, & Nov ‘21 QJE: cool method to recover info costs experimentally

Choosing information structures:

Ambuehl & Li ‘15 GEB: undervalue info; beliefs underreact to info; biases due to nonstandard belief updating

Charness, Oprea, & Yuksel ‘21 JEEA: rely on heuristics esp. confirmation-seeking

Novak, Matveenko, & Ravaioli ‘21 WP: mostly in line with CIA, prob choosing best π depends on value difference, tend to prefer simpler π

Overview

1. Costly Information Acquisition
2. Non-Separable Costs
3. Additive Costs
4. Multiplicative Costs
5. Constrained Information Acquisition
6. Relation to Stochastic Choice
7. Entropy Costs
8. Limitations of and Alternatives to Entropy Costs
9. Implications of CIA for Choice
10. Experimental Evidence
11. Applications

Applications

- Business cycle dynamics (Mackowiak & Wiederhold ‘15 REStud; Benhabib, Liu, & Wang ‘16 JET)
- Labor market discrimination (Bartos, Bauer, Chytilova, & Matejka ‘16 AER)
- Hiring (Acharya & Wee ‘20 AEJMacro)
- Choosing health insurance (Heiss, McFadden, & Winter ‘21 AER; Brown & Jeon ‘21 WP)
- Product pricing (Matejka ‘16 REStud; Martin ‘17 GEB; Stevens ‘20 REStud)
- Electoral competition (Matejka & Tabellini ‘21 JEEA; Feddersen & Sandroni ‘06 QJPS)
- Persuasion (Le Treust & Tomala ‘19 JET, Bloedel & Segal ‘18 WP; Wei ‘21 JMathE)
- Trade (Dasgupta & Mondria ‘18 JInternationalE)
- Principal-Agent problems (Lipnowski, Mathevet, & Wei ‘20 AERI)
- News media (Nimark & Pitschner ‘19 JET)
- Financial trading (Van Nieuwerburgh & Veldkamp ‘10 REStud; Kacperczyk, Van Nieuwerburgh & Veldkamp ‘16 Ecta; Vives & Yang ‘18 WP)

Applications

- Structural IO (Aguirregabiria & Jeon ‘20 RIO; Matveenko ‘17 WP; Fosgerau, Monardo, & De Palma ‘20 WP)
- Phillips curve (Afrouzi & Yang ‘21 WP)
- Bargaining (Ravid ‘20 AER)
- Monetary policy (Angeletos & La’O ‘20 JPE)
- Auctions (Persico ‘00 Ecta; Shi ‘12 GEB)