ROBUST RANDOM CHOICE

Evan Piermont
&
Roee Teper

21 May, 2018

Economic choice data is usually aggregated across

- many subjects, or,
- many different points in time, or both

For a choice problem: $D = \{x, y, z\}$

The analyst observes: $\rho_D(E)$ for $E \subseteq D$, representing the frequencies of choice.

Economic choice data is usually aggregated across

- many subjects, or,
- many different points in time, or both

For a choice problem: $D = \{x, y, z\}$

The analyst observes: $\rho_D(E)$ for $E \subseteq D$, representing the frequencies of choice.

Can we still ensure consistency with individual rationality?

Random Utility Models are a way of dealing with aggregated choice data:

Let $\mathcal{U}$ denote a set of utility functions. Then $\xi \in \Delta(\mathcal{U})$ is a random utility representing ρ if

$$\rho_D(E) = \xi\{u \in \mathcal{U} \mid \arg\max_D u \in E\}$$

Random utility models do not deal well with ties:

If with positive ξ-probability $u(x) = u(y)$, what is $\rho_D(x)$?

Random utility models do not deal well with ties:

If with positive ξ-probability $u(x) = u(y)$, what is $\rho_D(x)$?

Depends on how indifferences are broken.

- Gul and Pesendorfer (2006); Frick et al. (2017) assume ties occur with probability 0;
- Lu (2016) extends to allow ties with probability 0 or probability 1;
- Apesteguia et al. (2017) rule out ties by fiat, considering only linearly ordered preferences.
- Ahn and Sarver (2013) allow a tie breaking rule identified by preference over decision problems.

We present a model of random choice, ρ, such that

- Utility functions are linear (EU)
- Robust to how ties are broken:
    - We make **no** assumptions on how indifference is handled.
    - ρ is identified only up to choice by **strict** maximization.
- Necessary and sufficient for consistency with a EU-RUM, ξ.

Why do we care?

- Actual data reflects multiple tie breaking rules.
- Mathematically interesting.
- There is economic content to indifference.

If I learn an event did not occur, I do not care about the outcome within the event.

Using random choice to identify a rich model of info acquisition requires dealing with indifference.

Example

- State space $\{s_1, s_2\}$—ex ante equally likely.
- $x = [x^1, x^2] \in \mathbb{R}^2$ is a state contingent claim in utils.
- A 'utility function' is a belief $u \in \Delta(\{s_1, s_2\})$.
  - Utility of x is $u \cdot x$.
- Each agent i can learn the state at cost $c_i \in \mathbb{R}$.
  - Without info: $\xi([\frac{1}{2}, \frac{1}{2}]) = 1$.
  - With info: $\xi([1, 0]) = \xi([0, 1]) = \frac{1}{2}$.

Can we identify the distribution over c from choice frequencies?

Example

- $D_\lambda = \{x_\lambda, y_\lambda, z_\lambda\}$, $\lambda > 0$.
  - $x_\lambda = [\frac{2\lambda}{3}, \frac{2\lambda}{3}]$
  - $y_\lambda = [\lambda, 0]$
  - $z_\lambda = [0, \lambda]$
- Conditional choices are:
  - Without info: $\rho_{D_\lambda}(x_\lambda) = 1$.
  - With info: $\rho_{D_\lambda}(y_\lambda) = \rho_{D_\lambda}(z_\lambda) = \frac{1}{2}$.

Example

Then agent i obtains info whenever $\lambda - c_i \geq \frac{2\lambda}{3}$, i.e., when

$$\frac{1}{3}\lambda \geq c_i$$

So Prob$(c < \lambda)$ is probability info is obtained in $D_{3\lambda}$. I.e.,

$$\text{Prob}(c < \lambda) = 1 - \rho_{D_{3\lambda}}(x_{3\lambda}).$$

Example

But notice, when information is acquired

- $u \cdot [x^1, x^2] = u \cdot [x^1, y^2]$ with ξ prob $\frac{1}{2}$.
- Must allow non-trivial indifference.

This paper:

1. Outline a general model of random (linear) choice
2. Discuss how this model can identify endogenous information acquisition

Idea

ρ is identified only up to strict maximization:

- Different agent's break ties in different ways.
- We assume ρ is largest value consistent with *some* tie breaking rule.
- ρ is no longer a probability distribution.

Example

- As before, $x \in \mathbb{R}^2$.
- Let ξ be uniform over $u_1 = [1, 0]$, $u_2 = [\frac{1}{2}, \frac{1}{2}]$ and $u_3 = [0, 1]$.
- Let $D = \{[5, 0], [0, 5]\}$

What is $\rho_D([5, 0])$?

Example

- As before, $x \in \mathbb{R}^2$.
- Let ξ be uniform over $u_1 = [1, 0]$, $u_2 = [\frac{1}{2}, \frac{1}{2}]$ and $u_3 = [0, 1]$.
- Let $D = \{[5, 0], [0, 5]\}$

What is $\rho_D([5, 0])$?

- If u_1 is realized, or,

Example

- As before, $x \in \mathbb{R}^2$.
- Let ξ be uniform over $u_1 = [1, 0]$, $u_2 = [\frac{1}{2}, \frac{1}{2}]$ and $u_3 = [0, 1]$.
- Let $D = \{[5, 0], [0, 5]\}$

What is $\rho_D([5, 0])$?

- If u_1 is realized, or,
- if u_2 is realized and indifference is broken for $[5, 0]$.

Example

- As before, $x \in \mathbb{R}^2$.
- Let ξ be uniform over $u_1 = [1, 0]$, $u_2 = [\frac{1}{2}, \frac{1}{2}]$ and $u_3 = [0, 1]$.
- Let $D = \{[5, 0], [0, 5]\}$

What is $\rho_D([5, 0])$?

- If u_1 is realized, or,
- if u_2 is realized and indifference is broken for $[5, 0]$.
- Each agent might choose $[5, 0]$ with prob between $\frac{1}{3}$ and $\frac{2}{3}$.

Example

- As before, $x \in \mathbb{R}^2$.
- Let ξ be uniform over $u_1 = [1, 0]$, $u_2 = [\frac{1}{2}, \frac{1}{2}]$ and $u_3 = [0, 1]$.
- Let $D = \{[5, 0], [0, 5]\}$

What is $\rho_D([5, 0])$?

- If u_1 is realized, or,
- if u_2 is realized and indifference is broken for $[5, 0]$.
- Each agent might choose $[5, 0]$ with prob between $\frac{1}{3}$ and $\frac{2}{3}$.
- $\rho_D([5, 0]) = \rho_D([0, 5]) = \frac{2}{3}$.

Set Up

$\mathcal{D}$ is the set of **decision problems**: all finite non-empty subsets of R^n.

Could be interpreted as:

- lotteries (wlog restrict attention to simplex).
- state contingent claims (in utils).

Primitive

Our primitive is a **random choice capacity** (RCC),

$$\rho = \{\rho_D : 2^D \to [0,1]\}_{D \in \mathcal{D}}.$$

ρ_D is

- **grounded**: $\rho_D(\emptyset) = 0$.
- **normalized**: $\rho_D(D) = 1$.
- **monotone**: $\rho_D(A \cup B) \geq \rho_D(A)$.
- not necessarily additive! (recall above example)

Utilities

For $u \in \mathbb{R}^n$, we can consider $u : \mathbb{R}^n \to \mathbb{R}$—a linear utility function—via the inner product:

$$u : x \mapsto u \cdot x = \sum_{i \leq n} u^i x^i$$

- If x is interpreted as a lottery:
  - u is interpreted as a vNM index
- If x is interpreted as state contingent claim:
  - u is interpreted as belief (prob over state space)

Utilities

For $A \subseteq D$, let

$$N(D, A) = \{u \in \mathbb{R}^n \mid A \cap (\arg\max_{y \in D} u \cdot y) \neq \emptyset\}.$$

- If something in A is chosen from D, the agent's utility must be in $N(D, A)$.
- $N(D, \{x\})$ is normal cone to D at x.
- $N(D, D) = \mathbb{R}^n$.
- Let Ω denote the smallest algebra containing all $N(D, A)$.

Random Linear Representations

Call a (finitely additive) probability measure, ξ over $(\mathbb{R}^n, \Omega)$, a **random linear representation** (RLR). Say that ρ **maximizes** ξ if

$$\rho_D(A) = \xi(N(D, A))$$

for all (D, A).

GP axioms

If ρ is additive then GP provide conditions for the existence of a RLR:

1. **Monotonicity**: $D \subseteq D' \implies \rho_D(x) \geq \rho_{D'}(x)$.
2. **Extremeness**: $\rho_D(\text{ext}(D)) = 1$
3. **Linearity**: $\rho_D(x) = \rho_{\lambda D+y}(\lambda x + y)$ for $\lambda > 0$.
4. **Mixture Cont**: $\rho_{\lambda D+\lambda' D'}$ is continuous in λ, λ' for $\lambda, \lambda' \geq 0$.

- We keep **Linearity** and **Mixture Continuity** exactly.
- Modify **Monotonicity** and **Extremeness**
- Add an additional restriction: **Convex Modularity**

Montonicity

Let $D \subset D'$, and let $A \subset D$. Then

$$\rho_D(A) \geq \rho_{D'}(A),$$

with equality whenever $\text{ext}(D) = \text{ext}(D')$.

- In GP, with ρ additive: $\text{ext}(D) = \text{ext}(D) = 1$ implies the final condition.

Let $D = \{x, y, \frac{1}{2}x + \frac{1}{2}y\}$. For all u:

$$u \cdot \left(\frac{1}{2}x + \frac{1}{2}y\right) = \max\{u \cdot x, u \cdot y\}$$

if and only if $u \cdot x = u \cdot y$.

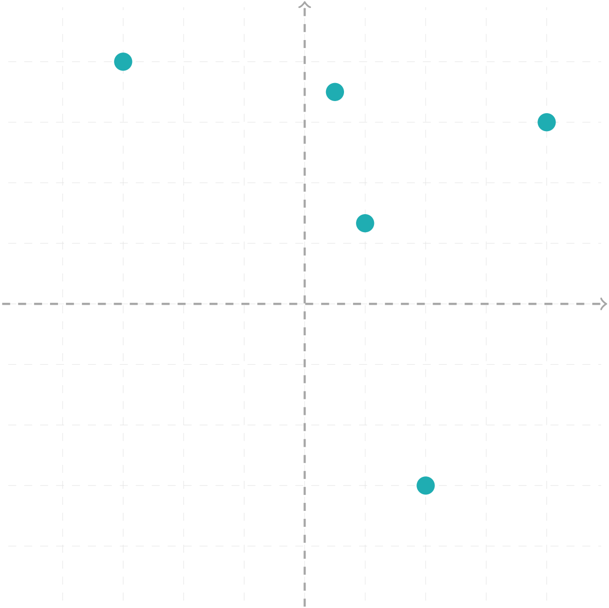
- Convex combinations are chosen only when all extreme points are chosen.
- GP assume this never happens, $\rho_D(\text{int}(D)) = 0$.

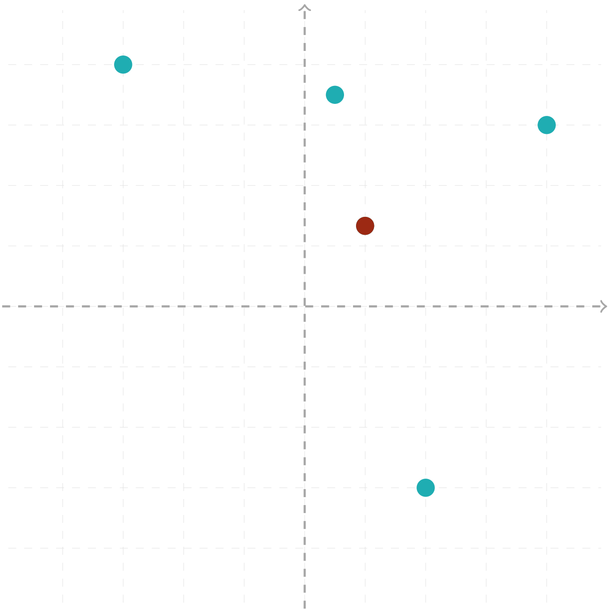
Let $\mathsf{pi}(D, A) =$

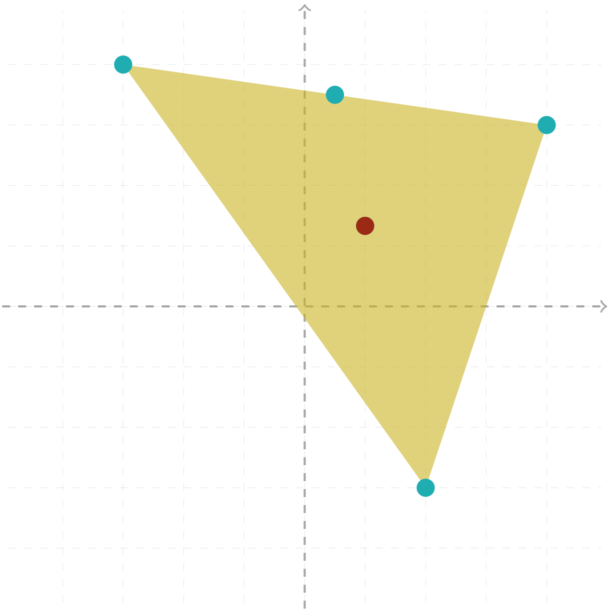
$$\{x \in \mathsf{conv}(D) \mid x = \alpha a + (1-\alpha)y, a \in A, y \in \mathsf{conv}(D), \alpha \in (0,1]\}$$

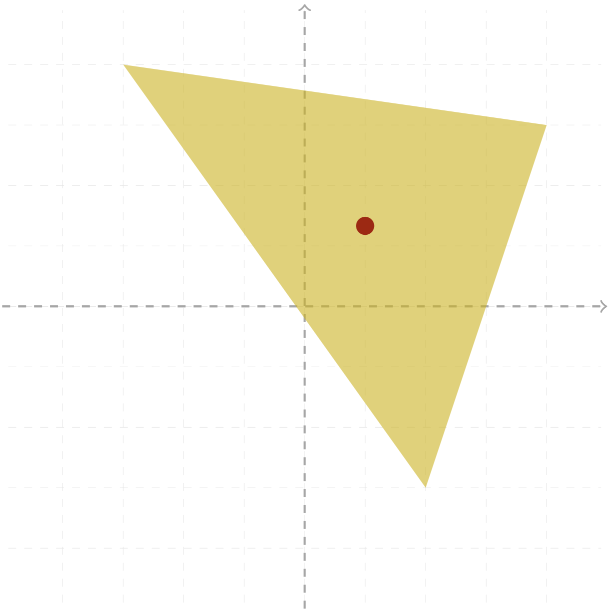
denote the projective interior of A in D.

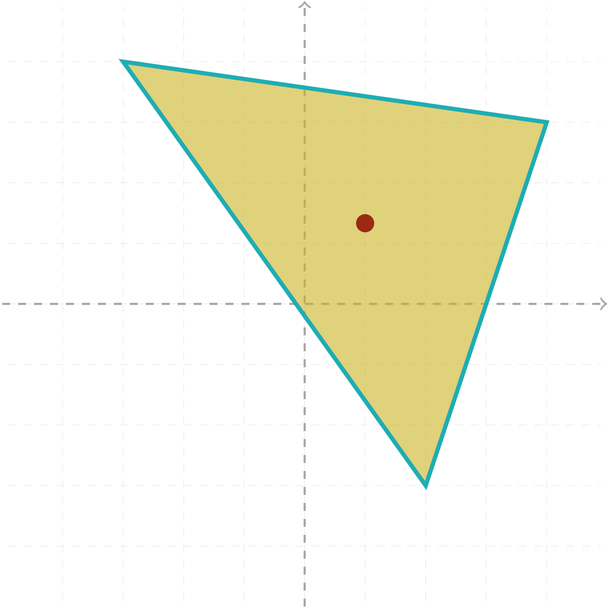
- $\mathsf{pi}(D, A)$ is the union of the relative interiors of all faces intersecting A.

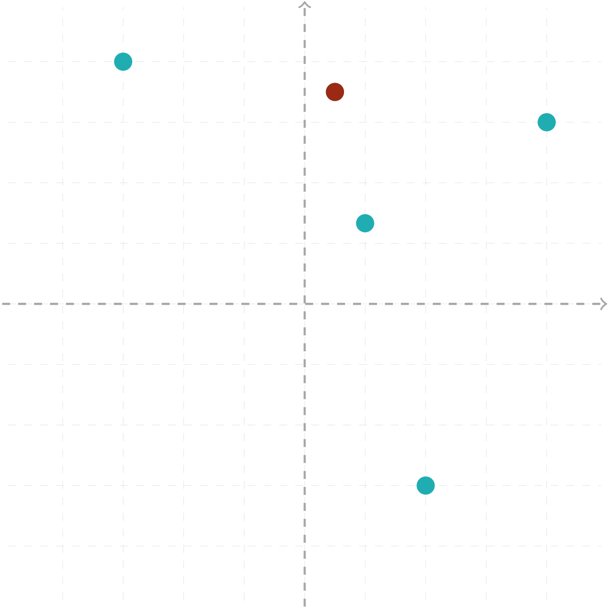


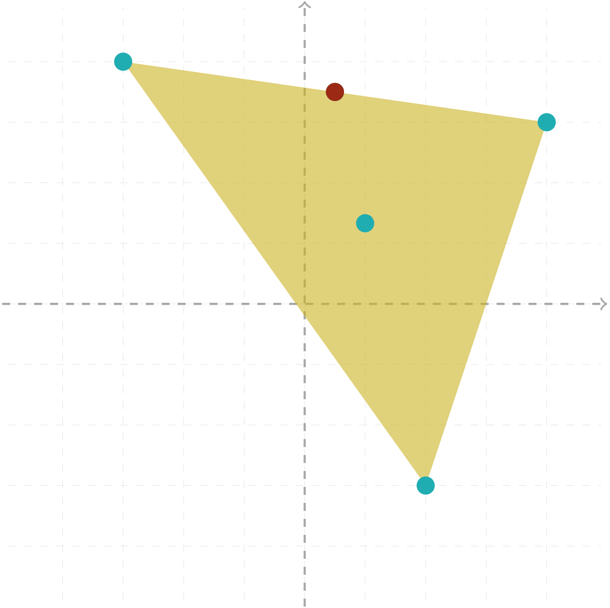


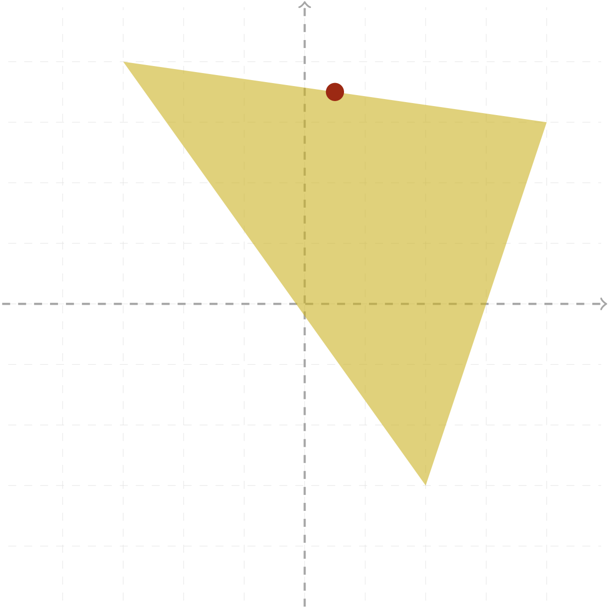


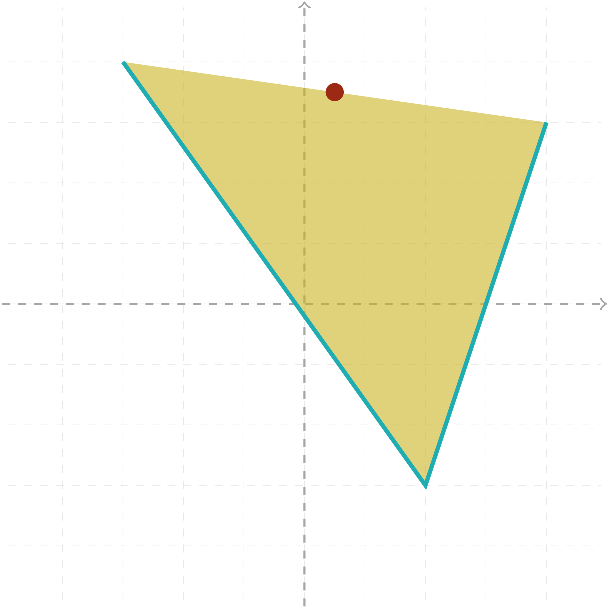


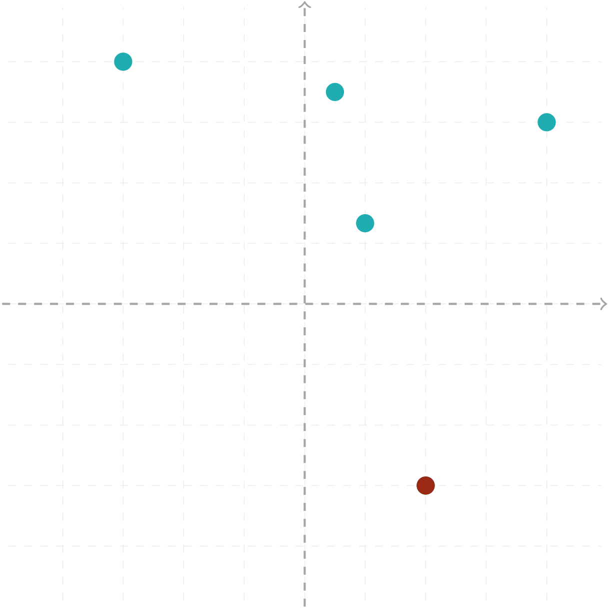


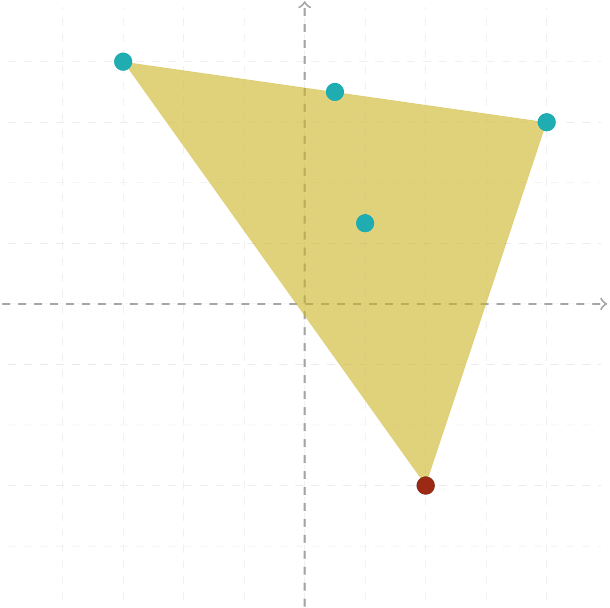


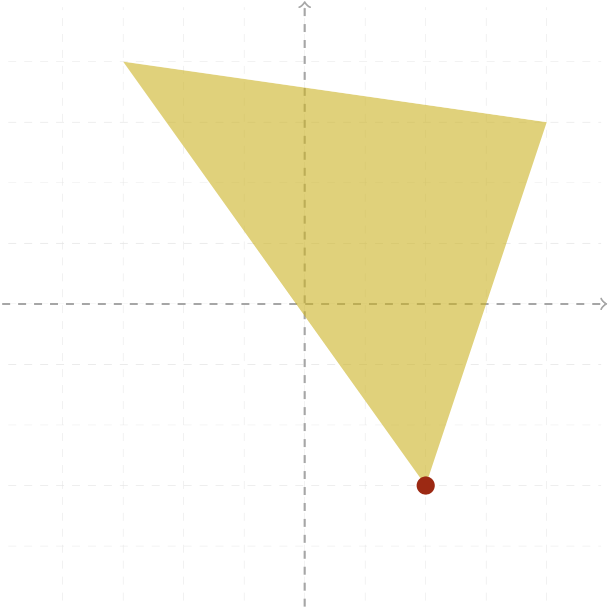


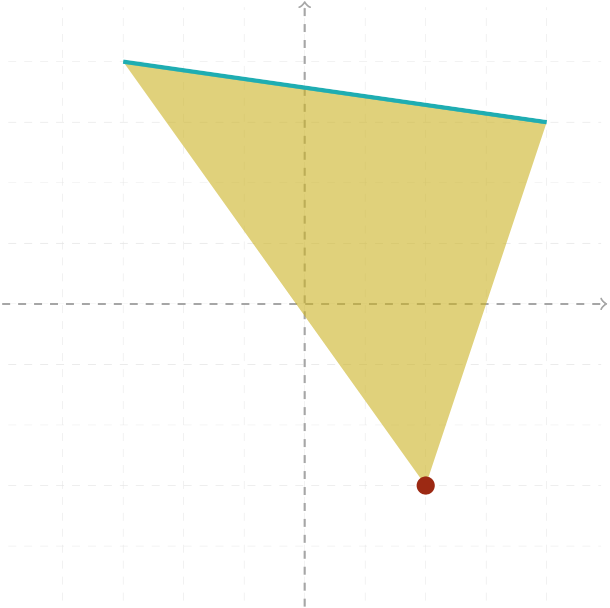












Extremeness

Let $A \subseteq D$. Then

$$\rho_D(\mathsf{pi}(D, A)) = \rho_D(A).$$

- Since $\mathsf{pi}(D, \mathsf{ext}(D)) = \mathsf{conv}(D)$. Extremeness implies $\rho_D(\mathsf{ext}(D)) = 1$.

Lastly, we need to restrict how non-additive ρ can be.

Let $D = \{x, y, \frac{1}{2}x + \frac{1}{2}y\}$. Notice:

$$\rho_D(\{x\}) = \xi\big(\{u \mid u(x) > u(y)\}\big) + \xi\big(\{u \mid u(x) = u(y)\}\big), \text{ and}$$
$$\rho_D(\{y\}) = \xi\big(\{u \mid u(y) > u(x)\}\big) + \xi\big(\{u \mid u(x) = u(y)\}\big).$$

So,

$$\rho_D(\{x, y\}) = \rho_D(\{x\}) + \rho_D(\{y\}) - \xi\big(\{u \mid u(x) = u(y)\}\big)$$

Also, recall,

$$u(x) = u(y) \iff \frac{1}{2}x + \frac{1}{2}y \in \arg\max_{z \in D} u(z)$$

Hence

$$\rho_D(\{x, y\}) = \rho_D(\{x\}) + \rho_D(\{y\}) - \rho_D(\frac{1}{2}x + \frac{1}{2}y)$$

Convex-Modularity

Let $A, B \subseteq D$ be such that $\frac{1}{2}A + \frac{1}{2}B \subseteq D$. Then

$$\rho_D(A \cup B) = \rho_D(A) + \rho_D(B) - \rho_D(\frac{1}{2}A + \frac{1}{2}B)$$

Theorem

The following are equivalent:

1. ρ satisfies Monotonicity, Extremeness, Convex-Modularity, Linearity, and Mixture-Continuity.
2. ρ maximizes a finitely additive RLR, ξ.

Proof Sketch

Two preliminary facts:

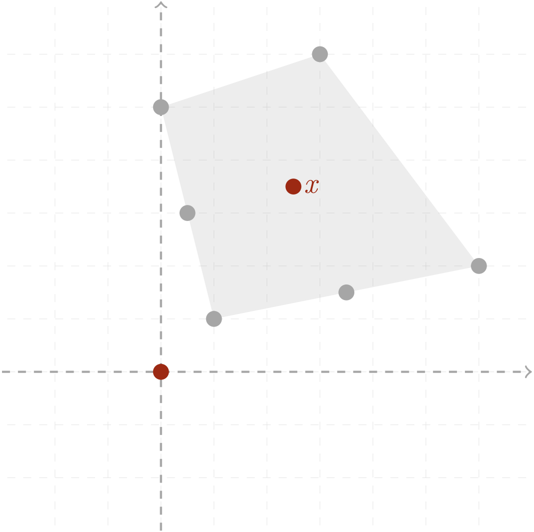
1. ρ is completely determined by its value over singletons.
   - Convex Modularity and Monotonicity

2. If $N(D, \{x\}) = N(D', \{x'\})$ then $\rho_D(x) = \rho_{D'}(x')$.
   - Linearity, mostly.

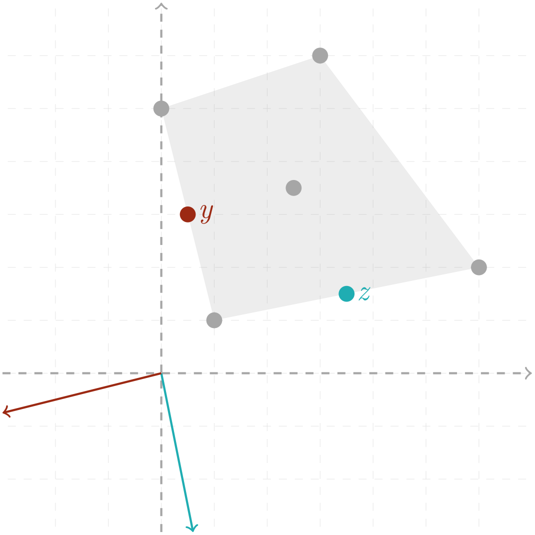
Proof Sketch

Construct ξ by setting value on ri$(N(D, \{x\})$ for each $(D, \{x\})$.

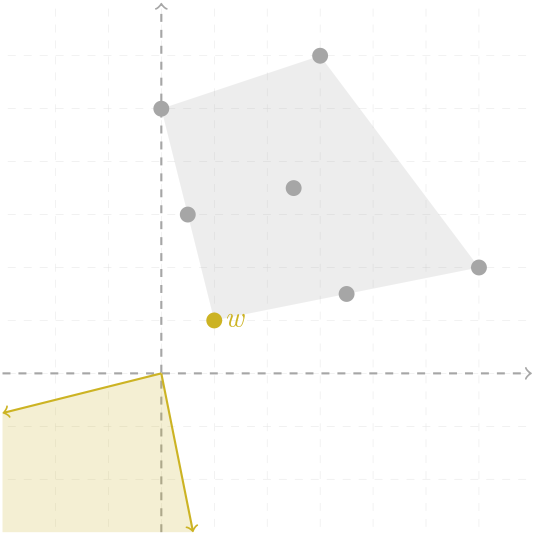
- Induction on the dimension of $N(D, \{x\})$.
- If $\dim(N(D, \{x\})) = 0$ then $N(D, \{x\}) = \mathbf{0}$.
  - $\xi(\mathbf{0}) = \rho_D(x)$



- $\xi(\mathrm{ri}(N(D,x))) = \xi(\mathbf{0}) = \rho_D(x)$



- $\xi(\mathsf{ri}(N(D,y))) = \xi(N(D,y) \setminus \mathbf{0}) = \rho_D(y) - \rho_D(x)$
- $\xi(\mathsf{ri}(N(D,z))) = \xi(N(D,z) \setminus \mathbf{0}) = \rho_D(z) - \rho_D(x)$



- $\xi(\mathsf{ri}(N(D,w))) = \rho_D(w) - \rho_D(z) - \rho_D(y) + \rho_D(x)$

- We motivated the non-additivity of ρ as coming from tie breaking rules; but ρ is an arbitrary capacity.
- The structure imposed on ρ_D, implies that it is a *coherent upper probability*.
  - Exists a $M \subseteq \Delta(D)$ such that

$$\rho_D(A) = \sup_{m \in M} m(A),$$

Define

$$M(\xi, D) = \Big\{ \int_{R^n} \tau_u(A)\xi(du) \mid \tau_u \in \Delta(\mathbb{R}^n),\ \mathsf{supp}(\tau_u) = \arg\max_{y \in D} u(y) \Big\}.$$

Theorem

Let ρ maximize ξ. Then $\rho_D = \sup_{m \in M(\xi,D)} m(A)$ for all D.

Info Acquisition

- Take interpretation that $\mathcal{D}$ is sets of AA-acts on n-dimensional state space.
- A RLR ξ is over n-dimensional simplex.
  - Utility u is a belief.
  - Requires state monotonicity and non-triviality axioms, as in Lu (2016).

All agents have the same prior, but

- have different costs to acquire information.
    - I.e., run an experiment
    - Experiments are partitions of the state space.
- observe different realizations for the same exp,
    - according to the prior (exp are i.i.d.)

The Model

1. Agent i faces $D \in \mathcal{D}$, with prior μ.
2. Chooses partition, Γ to maximize

$$\sum_{s_i, i \leq n} \mu(s_i) \max_{x \in D} \mu(\cdot|\Gamma(s_i)) \cdot x - c_i(\Gamma)$$

3. Observes $\Gamma(s_i)$ with probability $\mu(\Gamma(s_i))$
4. Chooses $x \in D$ to maximize $\mu(\cdot|\Gamma(s_i)) \cdot x$

The observer only sees the final choice frequencies.

The resulting choice rule will satisfy all axioms but linearity and monotonicity.
Weakened to translation invariance.

Why not linearity?

- $D = \{[0,1],[1,0]\}$.
- As $\lambda \to 0$, choose no info when faced with λD.
- As $\lambda \to \infty$, choose maximal info when faced with λD.

As $\lambda \to 0$, choice becomes degenerate a dictated by prior, μ.

We can exploit this lack of linearity to identify the distribution of costs.

- (Assuming cost of $c_i(\Gamma)$ and $c_i(\Gamma')$ are independent.)
- If $\Gamma \subset \Gamma'$, but $c_i(\Gamma) \geq c_i(\Gamma')$ then agent never chooses Γ.

We proceed by induction on the number of cells of Γ.

If Γ has two cells E_1, E_2, we can find c_i as in the example. (Assume $\mu(E_1) < \mu(E_2)$ for simplicity).

- $D_\lambda(\Gamma)$ is acts
  1. Constant act (at least) $\lambda\gamma$ with $1 > \gamma \geq (1 - \min_{i \leq n} \mu(a_i))$.
  2. Act $x_\lambda^{\Gamma_i}$: λ on E_i and 0 on E_j, $j \neq i$.
- For a given λ, choose info iff $\lambda - c \geq \lambda\gamma$.
- Iff info is acquired will agent (maybe, depending on the info) choose $x_\lambda^{\Gamma_1}$.

If Γ' has three cells F_1, F_2, F_3, and refines
$\Gamma = \{E_1 = \{F_1\}, E_2 = \{F_2, F_3\}\}$.
(Assume $\mu(F_1) < \mu(F_2) < \mu(F_3)$ for simplicity).

- $D_{\lambda,\lambda'} = D_\lambda(\Gamma) \cup D_{\lambda'}(\Gamma')$
- For small $\lambda' = \lambda$: agents choose out of $D_\lambda(\Gamma)$.
- Only agents with low enough cost $c(\Gamma)$ will choose $x_\lambda^{\Gamma_1}$, independent of λ'.

- Fixing λ as λ' increases, the rate of switching form $x_{\lambda}^{\Gamma_1}$ to $x_{\lambda}^{\Gamma_1'}$ identifies the cost.
- Only if $c(\Gamma') > c(\Gamma)$, observe switching rate
- Can vary λ to observe the entire distribution.