Introduction

Mark Dean

Behavioral Economics G6943
Fall 2019

Aim for Today

- Nuts and bolts
  - See syllabus
- Utility and choice: A reminder
  - The importance of representation theorems
  - Some extensions
  - Testing Axioms
- Random Utility
- An Introduction to Bounded Rationality

A Representation Theorem for Utility Maximization

- The following should be familiar from your 1st year PhD class.
- First we defined a **data set**

Definition

For a finite set of alternatives X, a choice correspondence C is a mapping $C : 2^X/\emptyset \to 2^X/\emptyset$ such that $C(A) \subset A$ for all $A \in 2^X/\emptyset$.

- Next we defined a **model of behavior**

Definition

A utility function $u : X \to \mathbb{R}$ **rationalizes** a choice correspondence C if

$$C(A) = \arg\max_{x \in A} u(x)$$

If there exists a choice correspondence that rationalizes C then we say it has a **utility representation**

A Representation Theorem for Utility Maximization

- Then we defined some **conditions** (or **axioms**) on the data

Axiom α (AKA Independence of Irrelevant Alternatives) If
$$x \in B \subseteq A \text{ and } x \in C(A), \text{ then } x \in C(B)$$

Axiom β If $x, y \in C(A)$, $A \subseteq B$ and $y \in C(B)$ then $x \in C(B)$

- Before stating a **representation theorem** linking these conditions and the model

Theorem

A Choice Correspondence on a finite X has a utility representation **if and only if** *it satisfies axioms α and β*

A Representation Theorem for Utility Maximization

- And stating a uniqueness result

Theorem

Let $u : X \rightarrow \mathbb{R}$ be a utility representation for a Choice Correspondence C. Then $v : X \rightarrow \mathbb{R}$ will also represent C if and only if there is a strictly increasing function T such that

$$v(x) = T(u(x)) \; \forall \; x \in X$$

- If any of this is unfamiliar have a look at the detailed notes I'll put online

Representation Theorems: Why?

- Why was this a good idea?
- (For me) the most important reason is that the model of utility maximization has unobservable (or latent) variables
- Without a representation theorem it is hard to know what its observable implications are?
  - How could we test utility maximization in the lab if we don't observe utility
- Alternative: define an observable measure of utility
  - E.g. Bentham's felicific calculus
- But this is now a joint test of the hypothesis of utility maximization and the type of utility specified
- In contrast, a representation theorem gives a **precise** way to test the **entire class** of utility maximizing models
  - Necessary: if the data is consistent with utility maximization then it must satisfy those conditions
  - Sufficient: If it satisfies those conditions, then it is consistent with utility maximization

Representation Theorems: Why?

- Two added bonuses
  1. By making the observable implications clear, such theorems make it clear if and how different models make different predictions
  2. Uniqueness result tells us how seriously to take the unobservable elements of the model
     - e.g. how well identified utility is
- What has this got to do with behavioral economics?
- Throughout the course we are going to be adding constraints and motivations to our model of decision making
  - Attention costs, temptation, regret, beliefs etc
- Which may not be directly observable
- Without the use of representation theorem it is very hard to keep track of what behavior we are admitting by allowing these new psychological processes

- Will give an example of this in a minute
- First, a quick reminder about preferences

Definition

A **(complete) preference relation** is a (complete), transitive and reflexive binary relation

Definition

We say a complete preference relation $\succeq$ represents a choice correspondence C if

$$C(A) = \{x \in A | x \succeq y \ \forall \ y \in A\}$$

- You should also remember from your class last year two important theorems regarding preferences

Theorem

Let C be a choice correspondence on a finite set X. Then there exists a preference relation $\succeq$ which represents C - i.e.

$$C(A) = \{x \in A | x \succeq y \text{ for all } y \in A\}$$

if and only if C satisfies axioms α and β

Theorem

Let $\succeq$ be a binary relation on a finite set X. Then there exists a utility function $u : X \rightarrow \mathbb{R}$ which represents $\succeq$: i.e.

$$\begin{aligned} u(x) &\geq u(y) \text{ if and only if} \\ x &\succeq y \end{aligned}$$

if and only if $\succeq$ is a preference relation

The Importance of Representation Theorems: An Example

Gul and Pesendorfer

- As we will see in future lectures, choices may be affected by **reference points** as well as the set of available options
  - What you choose may depend on your point of reference
- One key question is where do reference points come from?
- In 2005 Koszegi and Rabin proposed a model of 'personal equilibrium'
  - People have 'rational expectations'
  - Reference point should be what you expect to happen
  - But what you expect to happen should be what you would choose given your reference point
  - An option is a personal equilibrium if **it is what you would choose if that is your reference point**

The Importance of Representation Theorems: An Example

Gul and Pesendorfer

- Let $U : X \times X \rightarrow \mathbb{R}$ be a reference dependent utility function
  - $U(x, z)$ is the utility of choosing alternative x when z is the status quo
- A choice correspondence satisfies the 'general' PE model if

$$C(A) = \{x \in A | U(x, x) \geq U(y, x) \; \forall \; y \in A\}$$

- A choice correspondence satisfies the 'specific' PE model if in addition it satisfies

1. U has the following functional form:

$$U(x, y) = \sum_{k \in K} u_k(x) + \sum_{j \in K} \mu(u_j(x) - u_j(y))$$

2. 'Status quo bias'

$$\begin{aligned} U(x, y) &\geq U(y, y) \\ &\Rightarrow U(x, x) > U(y, x) \end{aligned}$$

The Importance of Representation Theorems: An Example

Theorem

Let $C : 2^X / \varnothing \rightarrow 2^X / \varnothing$ be a choice function on a finite X The following statements are equivalent

1. *(General PE model): There exists a general PE utility function $U : X \times X \rightarrow \mathbb{R}$ such that*

$$C(A) = \{x \in A | U(x,x) \geq U(y,x) \ \forall \ y \in A\}$$

2. *There exists a complete, reflexive binary relation $\succeq$ such that*

$$C(A) = \{x \in A | x \succeq y \ \forall \ y \in A\}$$

3. *(Special PE model) There exists a special PE utility function $U : X \times X \rightarrow \mathbb{R}$ such that*

$$C(A) = \{x \in A | U(x,x) \geq U(y,x) \ \forall \ y \in A\}$$

- Recall the definition of the data set we have

Definition

For a finite set of alternatives X, a choice correspondence C is a mapping $C : 2^X/\emptyset \rightarrow 2^X/\emptyset$ such that $C(A) \subset A$ for all $A \in 2^X/\emptyset$.

- What are some problems with this data set?

1. X Finite
2. Observe choices from all choice sets
3. We allow for people to choose more than one option!
   - i.e. we allow for data of the form

$$C(\{x, y, z\}) = \{x, y\}$$

- Recall choices can be represented by preferences if α and β is satisfied regardless of the size of X

- For utility representation we usually require something else, typically continuity

Definition

A preference relation $\succeq$ on a metric space X is continuous if, for any $x, y \in X$ such that $x \succ y$, there exists an $\varepsilon > 0$ such that, for any $x' \in B(x, \varepsilon)$ and $y' \in B(y, \varepsilon)$, $x' \succ y'$

Theorem (Debreu)

Let X be a separable metric space, and $\succeq$ be a complete preference relation on X. If $\succeq$ is continuous, then it can be represented by a continuous utility function.

- Note: continuity cannon be violated in finite data sets.

Choices from all Choice Sets?

- Imagine running an experiment to try and test α and β
- The data that we need is the choice correspondence

$$C : 2^X / \emptyset \to 2^X / \emptyset$$

- How many choices would we have to observe?
- Lets say $|X| = 10$
  - Need to observe choices from every $A \in 2^X / \emptyset$
  - How big is the power set of X?
  - If $|X| = 10$ need to observe 1024 choices
  - If $|X| = 20$ need to observe 1048576 choices
- This is not going to work!

Choices from all Choice Sets?

- So how about we forget about the requirement that we observe choices from all choice sets
- Are α and β still enough to guarantee a utility representation?

$$\begin{aligned} C(\{x,y\}) &= \{x\} \\ C(\{y,z\}) &= \{y\} \\ C(\{x,z\}) &= \{z\} \end{aligned}$$

- If this is our only data then there is no violation of α or β
- But no utility representation exists
- Note this is a problem for many behavioral models as well
  - see "Bounded Rationality and Limited Data Sets" de Clippel and Rozen [2018]

- We say that x is **directly revealed preferred to** y ($xR^D y$) if, for some choice set A

$$\begin{aligned} y &\in A \\ x &\in C(A) \end{aligned}$$

- We say that x is **revealed preferred to** y (xRy) if we can find a set of alternatives w_1, $w_2, \ldots . w_n$ such that
  - x is directly revealed preferred to w_1
  - w_1 is directly revealed preferred to w_2
  - ...
  - w_{n-1} is directly revealed preferred to w_n
  - w_n is directly revealed preferred to y
- I.e. R is the transitive closure of R^D

- We say x is **strictly revealed preferred to** y (xSy) if, for some choice set A

$$
\begin{aligned}
y &\in A \textbf{ but not } y \in C(A) \\
x &\in C(A)
\end{aligned}
$$

The Generalized Axiom of Revealed Preference

- Note that we can observe revealed preference and strict revealed preference from the data
- With these definitions we can write an axiom to replace α and β
- What behavior is ruled out by utility maximization?

Definition

A choice correspondence C satisfies the Generalized Axiom of Revealed Preference (GARP) if it is never the case that x is revealed preferred to y, and y is **strictly** revealed preferred to x

- i.e. xRy implies not ySx

Theorem

A choice correspondence C on an arbitrary subset of $2^X / \oslash$ satisfies GARP if and only if it has a preference representation

Corollary

A choice correspondence C on an arbitrary subset of $2^X / \oslash$ with X finite satisfies GARP if and only if it has a utility representation

Choice Correspondence?

- Another weird thing about our data is that we assumed we could observe a choice **correspondence**
  - Multiple alternatives can be chosen in each choice problem
- This is not an easy thing to do!
- What about if we only get to observe a choice function?
  - Only one option chosen in each choice problem
- How do we deal with indifference?
- One way is to figure out how to observe strict preferences

- The objects that the DM has to choose between are bundles of different commodities

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

- And they can choose any bundle which satisfies their budget constraint

$$\left\{ x \in \mathbb{R}^n_+ \middle| \sum_{i=1}^{n} p_i x_i \leq I \right\}$$

Definition

We say preferences $\succsim$ are **locally non-satiated** on a metric space X if, for every $x \in X$ and $\varepsilon > 0$, there exists

$$
\begin{aligned}
y &\in B(x, \varepsilon) \\
&\quad \text{such that} \\
y &\succ x
\end{aligned}
$$

Lemma

Let x^j and x^k be two commodity bundles such that $p^j x^k < p^j x^j$. If the DM's choices can be rationalized by a complete locally non-satiated preference relation, then it must be the case that $x^j \succ x^k$

- When dealing with choice from budget sets we say
  - x is **directly revealed preferred to** y if $p^x x \geq p^x y$
  - x is **revealed preferred to** y if we can find a set of alternatives w_1, $w_2, ...w_n$ such that
    - x is directly revealed preferred to w_1
    - w_1 is directly revealed preferred to w_2
    - ...
    - w_{n-1} is directly revealed preferred to w_n
    - w_n is directly revealed preferred to y
  - x is **strictly revealed preferred to** y if $p^x x > p^x y$

Theorem (Afriat)

Let $\left\{x^1, \ldots\ldots x^I\right\}$ *be a set of chosen commodity bundles at prices* $\left\{p^1, \ldots, p^I\right\}$*. The following statements are equivalent:*

1. *The data set can be rationalized by a locally non-satiated set of preferences* $\succeq$ *that can be represented by a utility function*
2. *The data set satisfies GARP (i.e. xRy implies not ySx)*
3. *There exists positive* $\left\{u^i, \lambda^i\right\}_{i=1}^{I}$ *such that*

$$u^i \leq u^j + \lambda^j p^j (x^i - x^j) \; \forall \; i, j$$

4. *There exists a continuous, concave, piecewise linear, strictly monotonic utility function u that rationalizes the data*

Testing Axioms in Practice

- So I have (hopefully) convinced you that representation theorems are a useful way of testing models with unobservable elements
- What do you think happens when we test these models in practice?
- They are (almost) always rejected!
- This is because axiomatic tests are 'all or nothing'
- One single mistake and an entire data set is declared irrational.

Testing Axioms in Practice

- This raises two related questions
  1. How close is a data set to satisfying a set of axioms?
  2. How much power does a particular data set have to identify violations of a set of axioms
- Techniques for answer these questions are very useful for behavioral economics
  - Most behavioral models include the standard model as a special case
  - Therefore they must (weakly) be able to explain more choice patterns than the standard model
  - How do we tell if the model is doing a good job?

- Which of these data sets do you think is closer to being rational?

Person A	Person B
$C_A(\{x, y\}) = \{x\}$	$C_B(\{x, y\}) = \{x\}$
$C_A(\{x, y, z\}) = \{z\}$	$C_B(\{x, y, z\}) = \{z\}$
$C_A(\{x, z\}) = \{z\}$	$C_B(\{x, z\}) = \{z\}$
$C_A(\{y, z\}) = \{y\}$	$C_B(\{y, z\}) = \{y\}$
$C_A(\{x, y, w\}) = \{w\}$	$C_B(\{x, y, w\}) = \{y\}$

- Arguably person A
- Because a **larger subset** of the data is consistent with rationality

- This is the basis of the HM index

Definition

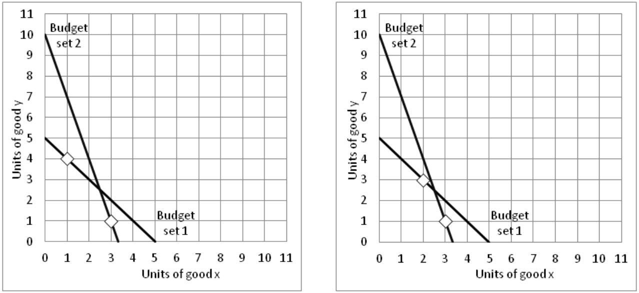
The HM index for a data set D is

$$\frac{|B|}{|D|}$$

where B is the largest subset of the data that satisfies the axiomatic system

- Advantages: Can be applied to any data set and axiomatic systems
- Disadvantages: Computationally complex, does not measure the size of the violation

- Which data set is closer to rationality?



- Arguably *b* as the budget set would have to be moved less in order to restore rationality
- This is the basis of the Afriat index

Definition

We say that x is revealed preferred to y at efficiency level e if $ep^x x > p^x y$.

- Note that $e = 1$ is standard revealed preference, and for $e = 0$ nothing is revealed preferred

Definition

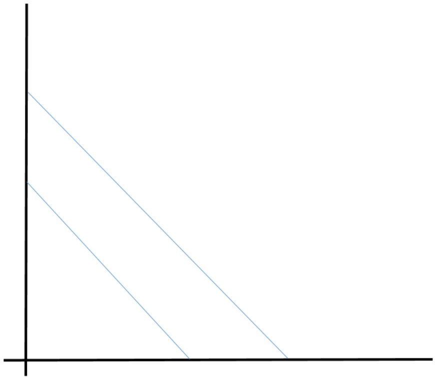
The Afriat index for a data set is the largest e such that the $e-$RP relation satisfies SARP

- Advantages: Computationally simple, takes into account the size of violations
- Disadvantages: Does not take into account number of violations, can only be applied to budget set data

Other Approaches

- There are a number of other approaches to this problem
- Possibly a sign that it has not been fully nailed.
    - Echenique, Federico, Sangmok Lee, and Matthew Shum. "The money pump as a measure of revealed preference violations." Journal of Political Economy 119.6 (2011): 1201-1223.
    - Dean, Mark, and Daniel Martin. "Measuring rationality with the minimum cost of revealed preference violations." Review of Economics and Statistics 98.3 (2016): 524-534.
    - Halevy, Yoram, Dotan Persitz, and Lanny Zrill. "Parametric recoverability of preferences." Journal of Political Economy 126.4 (2018): 1558-1593.
    - Aguiar, Victor, and Nail Kashaev. "Stochastic Revealed Preferences with Measurement Error: Testing for Exponential Discounting in Survey Data." (2017).
    - Maria Boccardi "Power of Revealed Preferences Tests and Predictive (Un)Certainty" (2018)

- Goodness of fit measures are important
- But they don't tell us everything we need to know



- How likely are we to observe a violation of GARP if we observe choices from these two choice sets?

Other Approaches

- Some data sets have more power that others to detect violations of a particular axiom set
- How do we measure this?
- Bronars [1987] proposed comparing the pass rate observed in the data to the pass rate from **randomly generated** data using the same parameters
  - e.g. we run an experiment in which subjects are asked to make choices from 30 budget sets
  - Construct a data set consisting of random choices from the same budget sets
  - Compare the fraction of these random data sets that satisfy GARP to the fraction of subjects who do

Random Utility

- Until now, our model has been one of a decision maker who
  - Has a single, fixed utility function
  - Makes choices in order to maximize this utility function
- So if we observe the DM sometimes choose x and sometimes choose y we would declare them irrational
- But maybe this is harsh?
  - Preferences affected by some unobserved state
  - Aggregating across individuals
  - Imperfect perception leading to mistakes

- Maybe a better model is one that accounts for this
- Random utility: Allow for random fluctuations in the utility function
- In order to sensibly talk about this model we need to extend the data set

Definition

For a finite set X and collection of choice sets $\mathcal{D} \subset 2^X/\emptyset$ a random choice rule is a mapping $p : \mathcal{D} \to \triangle(X)$ such that $Supp(p(A)) \subset A$

- We will use $p(x, A)$ to represent the probability of choosing x from A
- Records the probability of choosing each option in each choice set
- Where does stochastic choice come from?
  - Observation from different individuals
  - Changes in choices by the same individual

Definition

A Random Utility Model (RUM) consists of a finite set of one-to-one utility functions $\mathcal{U}$ on X and a probability distribution π on $\mathcal{U}$

- Ruling out indifference (because its a pain)
- Finiteness of $\mathcal{U}$ is without loss of generality (why?)

Definition

A RUM represents a random choice rule ρ if, for every $A \in \mathcal{D}$

$$p(x, A) = \sum_{u \in \mathcal{U} | x = \arg\max u(A)} \pi(u)$$

- Probability of choosing x from A is equal to the probability of drawing a utility function such that x is the best thing in A

Rationalizing a Random Choice Rule

- Is any choice rule compatible with RUM?
- No! One necessary condition is monotonicity

Definition

A random choice rule satisfies monotonicity if for any $x \in B \subset A \subseteq X$

$$p(x, B) \geq p(x, A)$$

- Adding alternatives to a choice set cannot increase the probability of choosing an existing option

Fact

If a Random Choice Rule is rationalizable it must satisfy monotonicity

Proof.

Follows directly from the fact that

$$\begin{aligned} & \{u \in \mathcal{U} | x = \arg\max u(A)\} \\ \subseteq \quad & \{u \in \mathcal{U} | x = \arg\max u(B)\} \end{aligned}$$

□

Rationalizing a Random Choice Rule

- So is monotonicity also sufficient for a random choice rule to be consistent with RUM?
- Unfortunately not
- Consider the following example of a stochastic choice rule on $\{x, y, z\}$

$$
\begin{aligned}
p(x, \{x, y\}) &= \frac{3}{4} \\
p(y, \{y, z\}) &= \frac{3}{4} \\
p(z, \{x, z\}) &= \frac{3}{4}
\end{aligned}
$$

- Claim: this pattern of choice is not RUM rationalizable

Rationalizing a Random Choice Rule

- Why? Well consider preference ordering such that $z \succ x$
- We know the probability of utility functions consistent with these preferences is equal to $\frac{3}{4}$
- If $z \succ x$ there are three possible linear orders

$$\begin{aligned} z &\succ x \succ y \\ z &\succ y \succ x \\ y &\succ z \succ x \end{aligned}$$

- In each case, either $y \succ x$ or $z \succ y$ or both, meaning that

$$p(z, \{x, z\}) \leq p(y, \{x, y\}) + p(z, \{y, z\})$$

- Which is not true in this data

- Do we have necessary and sufficient conditions for RUM rationalizability?
- Yes, but they are pretty horrible

Theorem

A random choice rule is RUM rationalizable if and only it satisfies the Block Marschak inequalities: for all $A \in \mathcal{D}$ and $x \in A$

$$\sum_{B|A \subset B} (-1)^{|B/A|} p(x, B) \geq 0$$

- These can be tested, but only on complete data sets, and offer very little intuition.
- What can we do?

Extension 1: Kitamura Stoye

- In a recent paper Kitamura Stoye [ECMA 2018] offered an approach that has two advantages over the Block Marschak inequalities
    1. Applies to incomplete data
    2. Has an associated statistical test which takes into account the fact that we only observe estimates of $\hat{p}$
- Will describe the former (see paper for latter)

- Consider a data set consisting of choices from $\{a_1, a_2\}$, $\{a_1, a_2, a_3\}$ and $\{a_1, a_2, a_3, a_4\}$
- Construct vectors each entry of which relates to a given choice from each choice set

$$
\begin{gathered}
a_1 | \{a_1, a_2\} \\
a_2 | \{a_1, a_2\} \\
a_1 | \{a_1, a_2, a_3\} \\
a_2 | \{a_1, a_2, a_3\} \\
a_3 | \{a_1, a_2, a_3\} \\
a_1 | \{a_1, a_2, a_3, a_4\} \\
a_2 | \{a_1, a_2, a_3, a_4\} \\
a_3 | \{a_1, a_2, a_3, a_4\} \\
a_4 | \{a_1, a_2, a_3, a_4\}
\end{gathered}
$$

- Construct a matrix of all possible rationalizable choice vectors

$$\begin{array}{c} a_1 | \{a_1, a_2\} \\ a_2 | \{a_1, a_2\} \\ a_1 | \{a_1, a_2, a_3\} \\ a_2 | \{a_1, a_2, a_3\} \\ a_3 | \{a_1, a_2, a_3\} \\ a_1 | \{a_1, a_2, a_3, a_4\} \\ a_2 | \{a_1, a_2, a_3, a_4\} \\ a_3 | \{a_1, a_2, a_3, a_4\} \\ a_4 | \{a_1, a_2, a_3, a_4\} \end{array} \quad \left\{ \begin{array}{cccc} 1 & 1 & 0 & \\ 0 & 0 & 1 & \\ 1 & 1 & 0 & \\ 0 & 0 & 0 & \\ 0 & 0 & 1 & .... \\ 1 & 0 & 0 & \\ 0 & 0 & 0 & \\ 0 & 1 & 1 & \\ 0 & 0 & 0 & \end{array} \right\} = A$$

- Let P be the observed choice probabilities associated with each row of the matrix A

Theorem

P is rationalizable by RUM if and only if their exists a probability vector v such that

$$Av = P$$

- Computationally the tricky bit is computing A
  But KS have techniques for this

- A second approach we could take is to restrict ourselves to a specific class of random utility models: e.g. Luce

Definition

A Random Choice rule on a finite set X has a Luce representation if there exists a utility function $u : X \rightarrow \mathbb{R}_{++}$ such that for every $A \in \mathcal{D}$ and $x \in A$

$$p(x, A) = \frac{u(x)}{\sum_{y \in A} u(y)}$$

- Advantages:
  - Captures the intuitive notion that 'better things are chosen more often'
  - Equivalent to the Logit form where

    $$u(x) = v(x) + \varepsilon$$

    and ε has an extreme value type 1 distribution

- The Luce model also has a very clean axiomatization

Definition

A random choice rule p on a set X satisfies stochastic independence of irrelevant alternatives if and only if, for any $x, y \in X$ and $A, B \in \mathcal{D}$ such that $x, y \in A \cap B$

$$\frac{p(x, A)}{p(y, A)} = \frac{p(x, B)}{p(y, B)}$$

Theorem

A random choice rule is rationalizable by the Luce model if and only if it satisfies Stochastic IIA

- Problem: Stochastic IIA sometimes not very appealing:
    - Consider {red bus, car} vs {red bus, blue bus, car}

Extension 3: Change the Domain

- It is beyond the scope of this course, but (perhaps surprisingly) characterizing RUM becomes easier if we put more structure on the choice objects
  - Lotteries: Gul, Faruk, and Wolfgang Pesendorfer. "Random expected utility." Econometrica 74.1 (2006): 121-146.
  - Time dated rewards: Lu, Jay, and Kota Saito. "Random intertemporal choice." Journal of Economic Theory 177 (2018): 780-815.
- See also Lu, Jay. "Random choice and private information." Econometrica 84.6 (2016): 1983-2027.