Utility Theory

CMPUT 654: Modelling Human Strategic Behaviour

S&LB §3.1

Recap: Course Essentials

Course webpage: jrwright.info/bgtcourse/

Contacting me:

- Discussion board: piazza.com/ualberta.ca/fall2019/cmput654/
  for **public** questions about assignments, lecture material, etc.
- Email: james.wright@ualberta.ca
  for **private** questions (health problems, inquiries about grades)
- Office hours: After every lecture, or by appointment

Utility, informally

A utility function is a real-valued function that indicates how much an agent **prefers** an outcome.

Rational agents act to maximize their expected utility.

Nontrivial claim:

1. Why should we believe that an agent's preferences can be adequately represented by a **single number**?

2. Why should agents maximize **expected value** rather than some other criterion?

Von-Neumann and Morgenstern's Theorem shows when these are true.

Outline

1. Informal statement
2. Theorem statement (von Neumann & Morgenstern)
3. Proof sketch
4. Fun game!
5. Representation theorem (Savage)

Formal Setting: Outcome

Definition: Let O be a set of **outcomes**:

$$O = Z \cup \Delta(O)$$

Not a typo!

where Z is some set of "actual outcomes", and

$\Delta(X)$ represents the set of **lotteries** over **finite** subsets of X:

$$[p_1 : x_1, \ldots, p_k : x_k]$$

with $\sum_{j=1}^{k} p_j = 1$ and $x_j \in X \;\; \forall 1 \leq j \leq k$

Formal Setting:
Preference Relation

A preference relation is a relationship between outcomes.

Definition

For a specific **preference relation** $\succeq$, write:

1. $o_1 \succeq o_2$ if the agent **weakly prefers** o_1 to o_2,

2. $o_1 \succ o_2$ if the agent **strictly prefers** o_1 to o_2,

3. $o_1 \sim o_2$ if the agent is **indifferent** between o_1 and o_2.

Formal Setting

Definition
A **utility function** is a function $u : O \to \mathbb{R}$. A utility function **represents** a preference relation $\succeq$ iff:

1. $o_1 \succeq o_2 \iff u(o_1) \geq u(o_2)$, and

2. $u([p_1 : o_1, \ldots, p_k : o_k]) = \sum_{j=1}^{k} p_j u(o_j)$.

Representation Theorem

Theorem: [von Neumann & Morgenstern, 1944]

Suppose that a preference relation $\succeq$ satisfies the axioms **Completeness**, **Transitivity**, **Monotonicity**, **Substitutability**, **Decomposability**, and **Continuity**.

Then there exists a function $u : O \to \mathbb{R}$ such that

1. $o_1 \succeq o_2 \iff u(o_1) \geq u(o_2)$, and

2. $u([p_1 : o_1, \ldots, p_k : o_k]) = \sum_{j=1}^{k} p_j u(o_j)$.

That is, there exists a utility function that **represents** $\succeq$.

Completeness and Transitivity

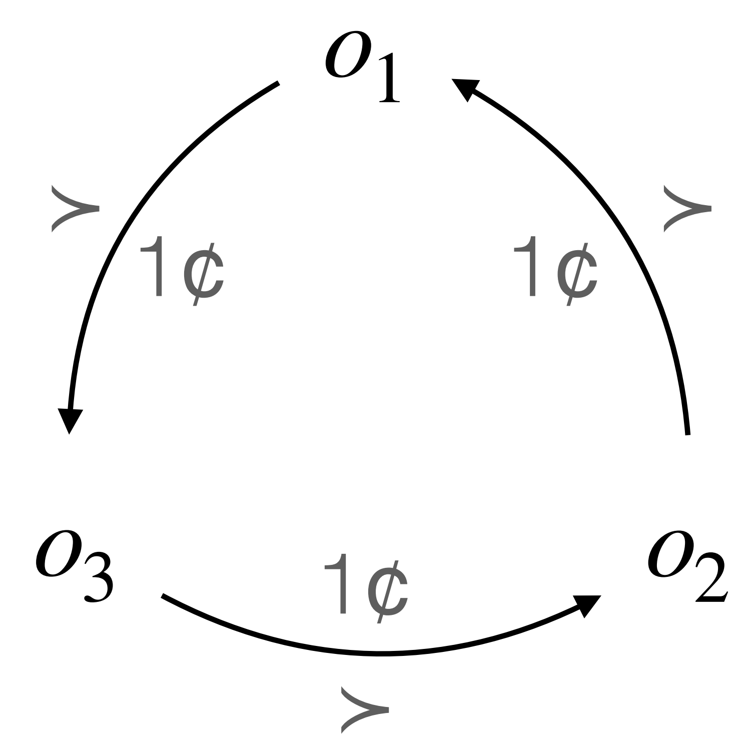
Definition (Completeness):

$$\forall o_1, o_2 : (o_1 \succ o_2) \vee (o_1 \prec o_2) \vee (o_1 \sim o_2)$$

Definition (Transitivity):

$$\forall o_1, o_2 : (o_1 \succeq o_2) \wedge (o_2 \succeq o_3) \implies o_1 \succeq o_3$$

Transitivity Justification: Money Pump



- Suppose that $(o_1 \succ o_2)$ and $(o_2 \succ o_3)$ and $(o_3 \succ o_1)$.
- Starting from o_3, you are willing to pay 1¢ (say) to switch to o_2
- But from o_2, you should be willing to pay 1¢ to switch to o_1
- But from o_1, you should be willing to pay 1¢ to switch back to o_3 again...

Monotonicity

Definition (Monotonicity):
If $o_1 \succ o_2$ and $p > q$, then

$$[p : o_1, (1-p) : o_2] \succ [q : o_1, (1-q) : o_2].$$

You should prefer a 90% chance of getting $1000 to
a 50% chance of getting $1000.

Substitutability

Definition (Substitutability):

If $o_1 \sim o_2$, then for all sequences $o_3, \ldots, o_k$ and $p, p_3, \ldots, p_k$ with $p + \sum_{j=3}^{k} p_j = 1$,

$$[p : o_1, p_3 : o_3, \ldots, p_k : o_k] \sim [p : o_2, p_3 : o_3, \ldots, p_k : o_k]$$

If I like apples and bananas equally, then I should be indifferent between a 30% chance of getting an apple and a 30% chance of getting a banana.

Decomposability aka "No Fun in Gambling"

Definition (Decomposability):
Let $P_\ell(o)$ denote the probability that lottery ℓ selects outcome o.

If $P_{\ell_1}(o_j) = P_{\ell_2}(o_j) \;\; \forall o_j \in O$, then $\ell_1 \sim \ell_2$.

Example:

Let $\ell_1 = [0.5 : [0.5 : o_1, 0.5 : o_2], 0.5 : o_3]$
Let $\ell_2 = [0.25 : o_1, 0.25 : o_2, 0.5 : o_3]$

Then $\ell_1 \sim \ell_2$, because

$$
\begin{aligned}
P_{\ell_1}(o_1) &= 0.5 \times 0.5 = 0.25 \qquad &= P_{\ell_2}(o_1) \\
P_{\ell_1}(o_2) &= 0.5 \times 0.5 = 0.25 \qquad &= P_{\ell_2}(o_2) \\
P_{\ell_1}(o_3) &= 0.5 \qquad &= P_{\ell_2}(o_3)
\end{aligned}
$$

Continuity

Definition (Continuity):

If $o_1 \succ o_2 \succ o_3$, then $\exists p \in [0,1]$ such that

$$o_2 \sim [p : o_1, (1-p) : o_3]$$

Proof Sketch: Construct the utility function

1. If $\succeq$ satisfies Completeness, Transitivity, Monotonicity, Decomposability, then for every $o_1 \succ o_2 \succ o_3$, there exists some p such that:

   (a) $o_2 \succ [q : o_1, (1-q) : o_3] \;\; \forall q < p$, and

   (b) $o_2 \prec [q : o_1, (1-q) : o_3] \;\; \forall q > p$.

2. If $\succeq$ additionally satisfies Continuity, then

$$\exists p : o_2 \sim [p : o_1, (1-p) : o_3].$$

3. Choose **maximal** $o^+ \in O$ and **minimal** $o^- \in O$.

4. Construct $u(o) = p$ such that $o \sim [p : o^+, (1-p) : o^-]$.

Question: Are o^+ and o^- guaranteed to exist?

Proof sketch:
Check the properties

1. $o_1 \succeq o_2 \iff u(o_1) \geq u(o_2)$

$u(o) = p$ such that $o \sim [p : o^+, (1 - p) : o^-]$.

Proof sketch:
Check the properties

2. $u([p_1 : o_1, \ldots, p_k : o_k]) = \sum_{j=1}^{k} p_j u(o_j)$

(i) Let $u^* = u([p_1 : o_1, \ldots, p_k : o_k])$

(ii) Replace o_j with $\ell_j = [u(o_j) : o^+, (1 - u(o_j)) : o^-]$, giving
$[p_1 : \ell_1, \ldots, p_k : \ell_k] = [p_1 : [u(o_1) : o^+, (1 - u(o_1)) : o^-], \ldots, p_k : [u(o_k) : o^+, (1 - u(o_k)) : o^-]]$

(iii) **Question:** What is $u([p_1 : \ell_1, \ldots, p_k : \ell_k])$?

$u([p_1 : \ell_1, \ldots, p_k : \ell_k]) = u^*$

(iv) **Question:** What is the probability of getting o^+ in $[p_1 : \ell_1, \ldots, p_k : \ell_k]$?

$\sum_{j=1}^{k} \left(p_j \times u(o_j)\right)$

(v) Construct $\ell^* = \left[\sum_{j=1}^{k} \left(p_j \times u(o_j)\right) : o^+, \left(1 - \sum_{j=1}^{k} \left(p_i \times u(o_j)\right)\right) : o^- \right]$

$u(\ell^*) = \sum_{j=1}^{k} \left(p_j \times u(o_j)\right)$

(vi) Observe that $[p_1 : \ell_1, \ldots, p_k : \ell_k] \sim \ell^*$ (**why?**)

$u([p_1 : \ell_1, \ldots, p_k : \ell_k]) = u^* = u(\ell^*) = \sum_{j=1}^{k} \left(p_j \times u(o_j)\right)$ ∎

Caveats & Details

Utility functions are **not uniquely defined.** (**Why?**)

- Invariant to affine transformations (i.e., $m > 0$):

$$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)] \iff X \succeq Y$$
$$\iff \mathbb{E}[mu(X) + b] \geq \mathbb{E}[mu(Y) + b] \iff X \succeq Y$$

This means we're not stuck with a range of [0,1]!

Caveats & Details

The proof depended on **minimal** and **maximal** elements of O, but that is not critical.

Construction for **unbounded** outcomes/preferences:

1. Pick two outcomes $o_s \prec o_e$. Construct utility for all outcomes $o_s \preceq o \preceq o_e$:

$$u : \{o \in O \mid o_s \preceq o \preceq o_e\} \to [0,1]$$

2. For outcomes o' outside that range, choose $o_{s'} \prec o' \prec o_s \prec o_e \prec o_{e'}$.
3. Construct utility $u' : \{o \in O \mid o_{s'} \preceq o \preceq o_{e'}\} \to [0,1]$.
4. Find $m > 0$ and $b \in \mathbb{R}$ such that $mu'(o_s) + b = u(o_s)$ and $mu'(o_e) + b = u(o_e)$.
5. Let $u(o) = mu'(o) + b$ for all $o \in \{o' \in O \mid o_{s'} \preceq o' \preceq o_{e'}\}$.

Fun game: Buying lottery tickets

Write down the following numbers:

1. How much would you pay for the lottery
   [0.3 : $5, 0.3 : $7, 0.4 : $9]?

2. How much would you pay for the lottery
   [p : $5, q : $7, ($1 - p - q$) : $9]?

3. How much would you pay for the lottery
   [p : $5, q : $7, ($1 - p - q$) : $9]
   if you knew the last seven draws had been 5,5,7,5,9,9,5?

Beyond von Neumann & Morgenstern

- The first step of the fun game was a good match to the utility theory we just learned.
  - **Question:** If two agents have different prices for [0.3 : \$5, 0.3 : \$7, 0.4 : \$9], what does that say about their utility functions for money?
- The second and third steps, not so much!
  - **Question:** If two agents have different prices for [p : \$5, q : \$7, $(1 - p - q)$: \$9], what does that say about their **utility functions**?
  - What if two people have the same prices for step 2 but different prices once they hear what the last few draws were?

Another Formal Setting

- **States**: Set S of elements $s, s', \dots$ with subsets $A, B, C, \dots$
- **Consequences**: Set F of elements $f, g, h, \dots$
- **Acts**: Arbitrary functions $\mathbf{f} : S \to F$
- Preference relation $\geq$ **between acts**
- $(\mathbf{f} \geq \mathbf{g}$ given $B) \iff$

  $\mathbf{f}' \geq \mathbf{g}'$ for every $\mathbf{f}', \mathbf{g}'$ that agree with $\mathbf{f}, \mathbf{g}$ respectively on B and each other on $\overline{B}$

Another Representation Theorem

Theorem: [Savage, 1954]
Suppose that a preference relation ≥ satisfies postulates P1-P6.
Then there exists a utility function U and a probability measure P such that

$$\mathbf{f} \geq \mathbf{g} \iff \sum_i P[B_i]U[f_i] \geq \sum_i P[B_i]U[g_i].$$

Postulates

P1 $\succeq$ is a simple order

P2 $\forall \mathbf{f}, \mathbf{g}, B : (\mathbf{f} \succeq \mathbf{g} \text{ given } B) \vee (\mathbf{g} \succeq \mathbf{f} \text{ given } B)$

P3 $(\mathbf{f}(s) = g \wedge \mathbf{f}'(s) = g' \ \ \forall s \in B) \implies (\mathbf{f} \succeq \mathbf{f}' \text{ given } B \iff g \succeq g')$

P4 For every A, B, either $A \leq B$ or $B \leq A$ (see D4)

P5 It is false that for every f, f', $f \succeq f'$.

P6 For all $\mathbf{g} \succ \mathbf{h}$ and consequence f, there exists a partition of S such that the consequence of either $\mathbf{g}$ or $\mathbf{h}$ can be replaced by f without changing the ordering of the two acts.

Summary

- Using very simple axioms about **preferences over lotteries**, utility theory proves that rational agents ought to act **as if** they were maximizing the **expected value** of a real-valued function.
    - **Rational** agents are those whose behaviour satisfies a certain set of **axioms**
    - *If you don't buy the axioms, then you shouldn't buy that this theorem is about rational behaviour*
- Can extend beyond this to "subjective" probabilities, using axioms about **preferences over uncertain "acts"** that do not describe how agents manipulate probabilities.