Utility Theory

CMPUT 654: Modelling Human Strategic Behaviour

S&LB §3.1

Recap: Course Essentials

Course webpage: jrwright.info/bgtcourse/

- This is the **main source** for information about the class
- Slides, readings, assignments, deadlines

Contacting me:

- Discussion board: piazza.com/ualberta.ca/winter2019/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 agents like 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 why (and when!) these are true.

- It is also a good example of some common elements in game theory (and economics):

  - Behaving “as-if”

  - Axiomatic characterization

Outline

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

Formal Setting

Definition
Let O be a set of possible **outcomes**. A **lottery** is a probability distribution over outcomes. Write $[p_1{:}o_1,\ p_2{:}o_2,\ \ldots,\ p_k{:}o_k]$ for the lottery that assigns probability p_j to outcome o_j.

Definition
For a specific **preference relation** $\geq$, write:

1. $o_1 \geq o_2$ if the agent **weakly prefers** o_1 *to* o_2,
2. $o_1 > o_2$ 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. $u([p_1 : o_1, \ldots, p_k : o_k]) = \sum_{i=1}^{k} p_i u(o_i)$.

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. $u([p_1 : o_1, \ldots, p_k : o_k]) = \sum_{i=1}^{k} p_i u(o_i)$.

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 ($o_1 > o_2$) and ($o_2 > o_3$) and ($o_3 > o_1$).
- 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

$$p + \sum_{i=3}^{k} p_i = 1,$$

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

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

Definition (Decomposability):

Let $P_\ell(o_i)$ denote the probability that lottery ℓ selects outcome o_i.

If $P_{\ell_1}(o_i) = P_{\ell_2}(o_i) \ \forall o_i \in O,$ 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

$$P_{\ell_1}(o_1) = P_{\ell_2}(o_1) = 0.25$$

$$P_{\ell_1}(o_2) = P_{\ell_2}(o_2) = 0.25$$

$$P_{\ell_1}(o_3) = P_{\ell_2}(o_3) = 0.5$$

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. For ≥ satisfying Completeness, Transitivity, Monotonicity, Decomposability, for every $o_1 > o_2 > o_3$, $\exists p$ such that:

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

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

2. For ≥ additionally satisfying Continuity,

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

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

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

Proof sketch:
Check the properties

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

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

Proof sketch: Check the properties

2. $u([p_1 : o_1, \ldots, p_k : o_k]) = \Sigma_{i=1}^{k} p_i u(o_i)$

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

   (ii) Replace o_i with $\ell_i = [u(o_i) : o^+, (1 - u(o_i)) : o^-]$, giving

   $$u^* = u([p_1 : [u(o_1) : o^+, (1 - u(o_1)) : o^-], \ldots, [p_k : [u(o_k) : o^+, (1 - u(o_k)) : o^-]]$$

   (iii) Question: What is the probability of getting ***o+***?
   Answer: $\Sigma_{i=1}^{k} p_i : u(o_i)$

   (iv) So $u^* = u\left(\left[\left(\Sigma_{i=1}^{k} p_i : u(o_i)\right) : o^+, \left(1 - \Sigma_{i=1}^{k} p_i : u(o_i)\right) : o^-\right]\right)$.

   (v) By definition of $u, u([p_1 : o_1, \ldots, p_k : o_k]) = \Sigma_{i=1}^{k} p_i u(o_i)$.

Caveats & Details

- Utility functions are **not uniquely defined**
  - Invariant to affine transformations (i.e., $m > 0$):
    $$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)] \iff X \succeq Y$$
    $$\iff \mathbb{E}[mu(X) + b] \geq \mathbb{E}[mu(Y) + b] \iff X \succeq Y$$
  - In particular, 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. Construct utility for some bounded range of outcomes
   $u : \{o_s, ..., o_e\} \rightarrow [0,1]$.
2. For outcomes outside that range, choose an overlapping range $\{o_{s'}, ..., o_{e'}\}$ with $s' < s < e' < e$
3. Construct $u' : \{o_{s'}, ..., o_{e'}\} \rightarrow [0,1]$ utility
4. Find $m > 0$, b such that $mu'(o_s) + b = u(o_s)$ and $mu'(o_{e'}) = u(o_{e'})$
5. Let $u(o) = mu'(o) + b$ for $o \in \{o_{s'}, ..., o_{e'}\}$

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.
  - If two people have different prices for step 1, what does that say about their utility functions for money?
- The second and third steps, not so much!
  - If two people have different prices for step 2, what does that say about their utility functions?
  - What if two people have the same prices for step 2 but different prices for step 3?

Another Formal Setting

- **States**: Set S of elements s, s', ... with subsets A, B, C, ...
- **Consequences**: Set F of elements f, g, h, ...
- **Acts**: Arbitrary functions $\boldsymbol{f} : S \rightarrow F$
- Preference relation ≥ **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 $\forall \mathbf{f}, \mathbf{g}, B : (\mathbf{f} \geq \mathbf{g}$ given $B) \vee (\mathbf{g} \geq \mathbf{f}$ given $B)$

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

P4 For every $A, B, (P[A] \leq P[B]) \vee (P[B] \leq P[A])$.

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

P6 (Sure-thing principle)

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.