Utility and Foundations

Modeling Human Strategic Behavior

Kevin Leyton-Brown

University of British Columbia
Canada CIFAR AI Chair, Amii

James R. Wright

University of Alberta
Canada CIFAR AI Chair, Amii

Student Introductions

Please introduce yourself by saying:

- what country you grew up in
- where you did your undergrad
- your current research interests
- something fun about you (your favourite band, book, flavour of ice cream, or anything else you'd like...)

Lecture Overview

Student Introductions

Informally

Theorem Statement

Utility Theory, Informally

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

Utility Theory, 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.

Utility Theory, 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.

This is a nontrivial claim!

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

Utility Theory, 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.

This is a 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?

Utility Theory, 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.

This is a 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.

Lecture Overview

Student Introductions

Informally

Theorem Statement

Formal Setting: Outcomes

Let O be the set of **outcomes**:

$$O = Z \cup \Delta(O)$$ (not a typo!)

where:

Formal Setting: Outcomes

Let O be the set of **outcomes**:

$$O = Z \cup \Delta(O)$$ (not a typo!)

where:

- Z is some set of "actual outcomes"

Formal Setting: Outcomes

Let O be the set of **outcomes**:

$$O = Z \cup \Delta(O)$$ (not a typo!)

where:

- Z is some set of "actual outcomes"
- $\Delta(X)$ represents the set of **lotteries** over **finite subsets** of X:

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

with $x_1, \ldots, x_k \in X$ and $\sum_{j=1}^{k} p_j = 1$.

Formal Setting: Preference Relation

A **preference relation** compares the relative desirability of outcomes.

For a given 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: Utility Function

A **utility function** is a function $u : O \to \mathbb{R}$.

Formal Setting: Utility Function

A **utility function** is a function $u : O \to \mathbb{R}$.

Definition

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

1. $\forall o_1, o_2 \in O : o_1 \succeq o_2 \iff u(o_1) \geq u(o_2)$, and
2. $\forall [p_1 : o_1, \; \ldots, p_k : o_k] \in O : u([p_1 : o_1, \; \ldots, p_k : o_k]) = \sum_{j=1}^{k} p_j u(o_j)$.

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. $\forall o_1, o_2 \in O : o_1 \succeq o_2 \iff u(o_1) \geq u(o_2)$, and
2. $\forall [p_1 : o_1, \ \ldots, p_k : o_k] \in O : u([p_1 : o_1, \ \ldots, p_k : o_k]) = \sum_{j=1}^{k} p_j u(o_j)$.

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

Completeness & Transitivity

Definition (Completeness)

A preference relation $\succeq$ satisfies **completeness** iff

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

Definition (Transitivity)

A preference relation $\succeq$ satisfies **transitivity** iff

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

Transitivity Justification: Money Pump

- Suppose that transitivity is violated: i.e., $(o_1 \succ o_2)$ and $(o_2 \succ o_3)$ and $(o_3 \succ o_1)$

Transitivity Justification: Money Pump

- Suppose that transitivity is violated: i.e., $(o_1 \succ o_2)$ 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

Transitivity Justification: Money Pump

- Suppose that transitivity is violated: i.e., $(o_1 \succ o_2)$ 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

Transitivity Justification: Money Pump

- Suppose that transitivity is violated: i.e., $(o_1 \succ o_2)$ 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...

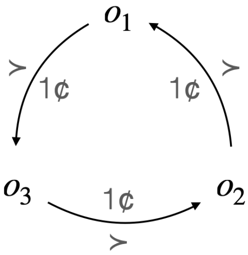
Transitivity Justification: Money Pump

- Suppose that transitivity is violated: i.e., $(o_1 \succ o_2)$ 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...

Transitivity Justification: Money Pump

- Suppose that transitivity is violated: i.e., $(o_1 \succ o_2)$ 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...

Agents with cyclic preferences are vulnerable to a money-pump!



Monotonicity

Definition (Monotonicity)

A preference relation $\succeq$ satisfies **monotonicity** iff for all $o_1, o_2 \in O$ and $p > q$,

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

You should prefer a 90% chance of getting \$1000 (or nothing) to a 50% chance of getting \$1000.

Substitutability

Definition (Substitutability)

A preference relation $\succeq$ satisfies **substitutability** iff for all $o_1, \ldots, o_k \in O$ and $p, p_3, \ldots, p_k$ satisfying $p + \sum_{j=3}^{k} p_j = 1$, if $o_1 \sim o_2$,

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

Substitutability

Definition (Substitutability)

A preference relation $\succeq$ satisfies **substitutability** iff for all $o_1, \ldots, o_k \in O$ and $p, p_3, \ldots, p_k$ satisfying $p + \sum_{j=3}^{k} p_j = 1$, if $o_1 \sim o_2$,

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

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)

A preference relation $\succeq$ satisfies **decomposability** iff for all lotteries ℓ_1, ℓ_2:

$$(\forall o \in O : P_{\ell_1}(o) = P_{\ell_2}(o)) \implies \ell_1 \sim \ell_2,$$

where $P_\ell(o)$ denotes the probability that outcome o is selected by lottery ℓ.

Decomposability (aka "No Fun in Gambling")

Definition (Decomposability)

A preference relation $\succeq$ satisfies **decomposability** iff for all lotteries ℓ_1, ℓ_2:

$$(\forall o \in O : P_{\ell_1}(o) = P_{\ell_2}(o)) \implies \ell_1 \sim \ell_2,$$

where $P_\ell(o)$ denotes the probability that outcome o is selected by lottery ℓ.

Example

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

Then $\ell_1 \sim \ell_2$ for any preference relation that satisfies decomposability, because

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

Continuity

Definition (Continuity)

A preference relation $\succeq$ satisfies **continuity** iff for all $o_1, o_2, o_3 \in O$,

$$o_1 \succ o_2 \succ o_3 \implies \exists p \in [0,1] : o_2 \sim [p : o_1,\ (1-p) : o_3].$$