Utility Maximization

Mark Dean

Lecture Notes for Spring 2015 Behavioral Economics - Brown University

1 Lecture 1

1.1 Introduction

The first topic we are going to cover in the course isn't going to seem very 'behavioral': we are going to look at the question of whether or not people make choices as if they are maximizing a stable utility function. We are going to begin with this question for two reasons. First of all, it is pretty foundational: every standard economic model has at it's heart the assumption that people are utility maximizers. It is therefore worth understanding the behavioral implications of this model of behavior, and the circumstances in which it does and does not do a good job of predicting behavior. After all, if we are going to introduce behavioral alternatives to utility maximization, we should be doing so in cases where it does a poor job of explaining behavior.

The second reason to look at this question is that it introduces us to a problem that we are going to come to time and time again during this course, which is the 'latent variable' problem. Almost all models in both standard and behavioral economics are couched in terms of latent variables. These are variables that we cannot directly observe, but help us with our intuition about what is going on. In this case, the latent variable is 'utility': When we see people making choices between objects, we see the objects that they are choosing from, and the objects that they choose. What we don't see is the 'utility' of those objects, which makes the question "How do I tell if someone is a utility maximizer?" quite challenging to answer. Luckily, this is where the field of study that is "Decision Theory" is all about. You can therefore think of this section of the course as an introduction to some of the tools of decision theory.

Some of the materiel that we are going to cover here may seem familiar to you from ECON 1110. If so that's fine, you should consider this as revision - we will get onto new stuff pretty quickly.

1.2 How Can We Spot a Utility Maximizer?

In order to answer the question, (or indeed any question of this type) we need to identify two things:

1. **The data we wish to explain:** In this case, the data will consist of the choices that people make, and the set of available alternatives that they are choosing from. So, for example, if we want to model a decision maker's choices over snack foods, then the data might come in the following form:[1]

Available Snacks	**Chosen Snack**
Jaffa Cakes, Kit Kat	Jaffa Cakes
Kit Kat, Lays	Kit Kat
Lays, Jaffa Cakes	Jaffa Cakes
Kit Kat, Jaffa Cakes, Lays	Jaffa Cakes

   We can think of this data as resulting from a series of experiments in which a subject has been asked to choose from each of these sets of alternatives. Note that, for now, we only observe the decision maker choosing one item from each choice set. This is going to turn out to be important later on.

2. **The model we want to test**: The benchmark model in economics is that of utility maximization. We want to model an agent who acts as if they assign a fixed utility number $u(.)$ to each of the available alternatives, then chooses the one with the highest utility. Thus, for example, the decision maker might assign utilities

   $$\begin{aligned} u(jaffa\ cakes) &= 10 \\ u(kitkat) &= 5 \\ u(lays) &= 2 \end{aligned}$$

   Then, in any choice set, choose the snack with the highest utility. For the moment, let's assume that there are no ties.

[1] For the unaware, Jaffa Cakes are British biscuit-y type things that are one of the finest foodstuff known to humanity.

Notice that there is a feature of this model that makes it difficult to test: we do not directly observe utilities! If objects had utilities stamped on them, then testing the model would be easy: we could just look at the choices that a DM makes from different choice sets, and see if they always choose the object with highest utility. But in general objects do not have utilities stamped on them: if I look at a kit kat, I cannot directly observe what its utility is. How can we proceed?

One way to go would be to assume a particular utility function, and test that. For example, we could assume that people prefer more calories to less, and so utility should be equivalent to calories. However, there is a problem with this approach: We might pick the wrong utility function. Maybe the person we are observing is a dieter, and actually prefers less calories to more. Such a person would be maximizing a utility function, just not the one we assumed. Thus we could wrongly reject the model of utility maximization.

It would be better if we could ask the question whether there *exists any* utility function that explains peoples' choices. Put another way, can we identify patterns of choices that cannot be explained by the maximization of any utility function? Consider the following choices made by an experimental subject Ambrose

Ambrose's Choices

Available Snacks	**Chosen Snack**
Jaffa Cakes, Kit Kat	Jaffa Cakes
Kit Kat, Lays	Kit Kat
Lays, Jaffa Cakes	Lays
Kit Kat, Jaffa Cakes, Lays	Jaffa Cakes

Is it possible that Ambrose's choices can be explained by the maximization of some utility function? The answer is no. If Ambrose is making choices in order to maximize a utility function, then the choice of Jaffa Cakes over Kit Kat indicates that Jaffa Cakes must have a higher utility that Kit Kats[2], so $u(jaffa\ cakes) > u(kitkat)$. Similarly, the second choice tells us that $u(kitkat) > u(lays)$, while the third choice tells us that $u(lays) > u(jaffa\ cakes)$. Putting these three things together tells us that $u(jaffa\ cakes) > u(jaffa\ cakes)$, which is clearly impossible. Thus, there

[2]Note that we are ignoring the possibility that the subject is completely indifferent between the two snacks and chooses between them at random. There are ways round this problem, but they lie beyond the scope of this article. See any standard graduate microeconomics text for a discussion of this issue - for example Rubinstein [2007] chapter 3.

is no possible utility function that fits with Ambrose's choices.

Now let's look at the choice of a second experimental subject, Bishop:

Bishop's Choices

Available Snacks	**Chosen Snack**
Jaffa Cakes, Kit Kat	Jaffa Cakes
Kit Kat, Lays	Kit Kat
Lays, Jaffa Cakes	Jaffa Cakes
Kit Kat, Jaffa Cakes, Lays	Kit Kat

Can Bishop's choices be explained as resulting from utility maximization? Again the answer is no. The first choice tells us that $u(jaffa\ cakes) > u(kitkat)$, while the fourth choice tells us that $u(kitkat) > u(jaffa\ cakes)$, again a contradiction.

Finally, let us think about the choice of Croft.

Croft's Choices

Available Snacks	**Chosen Snack**
Jaffa Cakes, Kit Kat	Jaffa Cakes
Kit Kat, Lays	Kit Kat
Lays, Jaffa Cakes	Jaffa Cakes
Kit Kat, Jaffa Cakes, Lays	Jaffa Cakes

Can Croft be thought of as a utility maximizer? The answer is yes! The final choice tells us that $u(jaffa\ cakes) > u(kitkat)$ and $u(jaffa\ cakes) > u(lays)$, while the second choice tells us that $u(kitkat) > u(lays)$. Furthermore, no other choices contradict this ordering. Thus, for example, we can model Croft as maximizing the utility function $u(jaffa\ cakes) = 10$, $u(kitkat) = 2$, $u(lays) = 1$

We have therefore identified some data sets in which choices can be explained by utility maximization, and some in which they cannot. The question is, can we come up with a general rule which differentiates one from the other?

1.3 Spotting Utility Maximization.

This question was first answered by Paul Samuelson writing in 1938 (though various authors have reformulated his original response. We are going to answer it in two parts

1. Provide conditions under which choices can be represented as the maximization of a complete, transitive *preference relation*
2. Show that a complete, transitive preference relation can be represented by a utility function

Putting these two results together will be enough to show that our conditions guarantee that choices can be represented by a utility function. In fact, our results are going to be stronger than that, in that they are going to be *if and only if* (or, put another way, will be both necessary and sufficient for utility maximization). This is also going to be very important, as we shall see.

In order to prove these results, we are going to have to introduce some formal definitions. First, the data that we are going to work with. The first thing we need is a define the set of things that we are going to observe our decision maker choose from. This is just going to be some set of objects X. At the moment, they are going to have no other defining characteristics. They are just going to be 'things'. In the above example X was the set {jaffa cakes, lays, kit kat}.

Next, we are going to need a description of the choices that our decision maker (DM) makes. This is going to come in the form of a *choice correspondence*, which is going to tell us, for every subset of X, what the decision maker would choose from that subset.

Definition 1 *A choice correspondence C is a mapping $C : 2^X/\emptyset \rightarrow 2^X/\emptyset$ such that $C(A) \subset A$ for all $A \in 2^X/\emptyset$.*

Don't let the notation confuse you. All we are doing is describing what our data looks like. Remember, in our previous example, the data told us what the decision maker chose from each subset of X - i.e. from {kit kat, lays}, {kit kat, jaffa cakes} and so on. This is what the mapping C is going to tell us. The notation says that C is a mapping (i.e. a function) that takes as its input a subset of X (that's the first $2^X/\emptyset$ - 2^X is the power set of X, i.e. all the subsets of X. $/\emptyset$ means we ignore the empty set) and the spits out as its output another subset of X (that is what we mean by the $\rightarrow 2^X/\emptyset$ bit). Thus, if we plug in a subset A of X - for example, the subset {kit kat, lays}, then C will spit out another subset of X, for example

$$C(\{kitkat, jaffacakes\}) = \{jaffacakes\}$$

with the interpretation that $C(A)$ is the thing that the decision maker would choose from set A. The condition $C(A) \subset A$ tells us that the thing the DM will choose from a set A must be something in A (the notation $\subset$ means 'is a subset of') which seems like a sensible restriction.

Note that here we are allowing the decision maker to choose more than one option from any given choice set. For example, we allow

$$C(\{kitkat, jaffacakes, lays\}) = \{jaffacakes, kitkat\}$$

This is a pretty weird thing to do: note that we don't mean that the decision maker gets both jaffa cakes AND kitkats from the choice set. Instead, we think of it as meaning something like "the decision maker would be happy with either jaffa cakes or lays from this choice set" or perhaps "the decision maker sometimes chooses jaffa cakes and sometimes chooses kitkats" from this choice set. Allowing this type of thing is a technically useful assumption (because it allows us to deal with indifference), but also pretty dubious. We will come back to it later on.

Next, we need to describe what we mean by a preference relation. A preference relation is a type of binary relation, and a binary relation B on a set X is a subset of $X \times X$., where if $\{x, y\}$ is in B, we interpret this as meaning that x is 'in relation to' y (which we will write xBy). We will use the symbol $\succeq$ to represent our binary relation, so $x \succeq y$ means that $\{x, y\} \in B$. We demand that our decision maker behaves according to a binary relation that has certain properties:

Definition 2 *A transitive binary relation is one in which $x \succeq y \succeq z$ implies $x \succeq z$. A reflexive binary relation is one such that $x \succeq x$ for all $x \in X$. A binary relation that is transitive and reflexive is called a preference relation (or a preorder). A complete binary relation is one for which, for any $x, y \in X$, either $x \succeq y$ or $y \succeq x$ or both.*

We will think of the complete preference relation as representing 'weak preferences' - i.e. $x \succeq y$ means that 'x is at least as good as y'. Our behavioral model is that people have preferences that are well behaved (in the sense of being complete, transitive and reflexive), and these preferences govern their choices. (It will become obvious why we think of such preferences as well behaved).

We can now formulate our question in the following way: under what circumstances can we find **some** complete preference relation such that choice is equal to the set of maximum objects in each set according to that preference ordering: the DM chooses the best objects according to that preference ordering. In other words, we want to find some $\succeq$ such that, for all $A \in 2^X/\varnothing$

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

Again, don't get panicked by the notation. What we are saying here is that we want to find a binary relation $\succeq$ such that, for every choice set A, the set of chosen objects ($C(A)$) is equal to a set of objects in A that satisfy some condition (this is what the notation $\{x \in A|....\}$ means). The condition that we want is that the chosen objects are the best in that set, according to the binary relation $\succeq$ - i.e that x is chosen if and only if x is as good as y ($x \succeq y$) for every y in A ($\forall \; y \in A$).

Note that we are allowing for the possibility that two objects are indifferent - that $x \succeq y$ and $y \succeq x$. In this case, and if both objects are preferred to all other objects in some choice set, then we want both objects to be chosen.

We can now define our problem more formally. The aim of our representation theorem is to find some conditions on the choice function C such that we can find some preference relation that *rationalizes* the DM's choices (i.e. the DM chooses the best objects according to those preferences). Remember again what it is that we can observe here. We assume that we **can** observe choices but **cannot** observe preferences. If we could observe preferences then it would be easy to test whether people are maximizing preferences - all we would have to do is look and see whether the item they chose in each set was the most preferred item. Instead, we are completely agnostic about what this preference relation is, we just want the DM to be behaving in a manner consistent with *some* preference relation.

So what are the relevant conditions? I suspect that, as you have done ECON1110, you have some inkling (you have probably come across the Weak Axiom of Revealed Preference at some point). In fact, we are going to write two axioms which between them are equivalent to the weak axiom of revealed preference. However, splitting them up in this way makes it more obvious what is going on.

Axiom 1 (Property α) *If $x \in B \subseteq A$ and $x \in C(A)$, then $x \in C(B)$*

Axiom 2 (Property β) *If $x, y \in C(A)$, $A \subseteq B$ and $y \in C(B)$ then $x \in C(B)$*

It is worth stopping and thinking for a minute about these two axioms. The first is sometimes called the independence of irrelevant alternatives, and is very intuitively appealing. It says that, if you choose an alternative x from a larger set (i.e. $x \in C(A)$), then take some objects out of that set that are not x (i.e. B is a subset of A such that $x \in B$), then you should still choose x from the smaller set. (i.e. $x \in C(B)$). This is clearly a property you would expect a 'rational' decision maker to obey - if they choose x from the larger set, they are telling you that they prefer x to all of the other objects in that set. They should therefore prefer x to any objects in a *subset* of that larger set.

What about the second property? The first thing to note is that this property only has matters in the case of choice *correspondences*$-$i.e. if the decision maker sometimes chooses two or more objects from the same set . If this is not the case, then C is single valued, and the conditions $x, y \in C(A)$ can never hold, so this axiom will be satisfied trivially. In the case of a choice correspondence, this condition says that, if x and y are chosen from a set, and y is chosen from a superset of that set, then x must also be chosen from that superset. Again, this makes sense if we think of a rational decision maker. If x and y are chosen together, then the DM must be indifferent between them. If y is chosen from some other set, then it must be at least as good as anything else in that set, and therefore so must x.

We are now in a position to state and prove our representation theorem.

Definition 3 *A choice correspondence $C : 2^X/\varnothing \rightarrow 2^X/\varnothing$ for some $S \subset 2^X/\varnothing$ is **rationalized** by a preference relation $\succeq$ if, for every $A \in 2^X/\varnothing$, it is the case that*

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

Theorem 1 *For any finite set X and complete choice correspondence $C : 2^X/\varnothing \rightarrow 2^X/\varnothing$, there exists a complete preference relation $\succeq$ that rationalizes that choice correspondence if and only if C satisfies property α and β.*

Proof. *The first thing to do is note that this proof must come in two parts, as we are making two claims: this comes from the fact that the statement is "if and only if', so we have to show (i) that*

α and β imply that we can find a rationalizing preference relation and (ii) any rationalizable choice function satisfies α and β. We will start with the former, as this is the more tricky bit (in fact, we have already argued informally for the latter.) ■

Proof (axioms imply representation). *We will break the proof down into the following steps*

1. ***Generate a candidate binary relation.*** *Our claim is that, if the choice correspondence satisfies α and β, then it is rationalizable by some complete preference relation. The first stage of the proof is to describe such a relation, which we will then shows does the necessary job. We will define the relationship using choices from two objects we will say that $x \trianglerighteq y$ if and only if $x \in C(\{x, y\})$, so x is 'weakly preferred' to y (according to our candidate preference relation) if it is chosen from the set containing x and y only. We will stretch this definition somewhat by saying that $x \trianglerighteq x$, as x is definitionally chosen from the set $\{x\}$.*

2. ***Show that $\trianglerighteq$ is a complete, transitive preference relation.*** *So we have defined a binary relation. Great. However, our theorem demands that choices be rationalized by a complete preference relation - i.e. a complete, transitive, reflexive binary relation. We next need to show that $\trianglerighteq$ has these properties. Reflexivity is easy - in fact we defined $\trianglerighteq$ explicitly so that it is reflexive. Completeness is also relatively straightforward. By definition, $C(\{x, y\})$ is either $\{x\}$, $\{y\}$ or $\{x, y\}$. Thus, by the construction of $\trianglerighteq$ either $x \trianglerighteq y$, $y \trianglerighteq x$ or both. Finally, we need to show transitivity, which we will do by contradiction. Imagine there exists $x, y, z \in X$ such that $x \trianglerighteq y \trianglerighteq z$ but not $x \trianglerighteq z$. This implies*

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

This in turn implies that $z \in C(\{x, z\})$. We can now show that we must have a violation of either property α or property β. Consider the set $\{x, y, z\}$. If $x \in C(\{x, y, z\})$, then the fact that $x \notin C(\{x, z\})$ is a direct violation of property α. If $y \in C(\{x, y, z\})$, then by property α, $y \in C(\{x, y\}) = \{x, y\}$. Property β then implies that $x \in C(\{x, y, z\})$, which we have already shown leads to a violation of α. If $z \in C(\{x, y, z\})$, then by α $z \in C(\{y, z\}) = \{y, z\}$, and so by β $y \in C(\{x, y, z\})$. Again, we have already shown that this leads to a violation. However, as $C(\{x, y, z\})$ is nonempty, one of these cases must occur, and so a failure of transitivity *implies a failure of either α or β.*

3. ***Show that* $\trianglerighteq$ *rationalizes* C.** *We now need to show that, for all sets, our DM chooses as if they are maximizing $\trianglerighteq$. In other words, for some arbitrary $A \in 2^X/\varnothing$ we need to show that $C(A) = \{x \in A | x \trianglerighteq y \ \forall \ y \in A\}$. As we are proving the equality of two sets, this in itself takes two stages*

   (a) *$C(A) \subseteq \{x \in A | x \trianglerighteq y \ \forall \ y \in A\}$. Say $x \in C(A)$. Take any $y \in A$. We need to show that $x \trianglerighteq y$ - in other words that $x \in C(\{x, y\})$. However, this follows directly from property α. Thus, anything that is chosen from A must be 'preferred' to everything else in A*

   (b) *$C(A) \supseteq \{x \in A | x \trianglerighteq y \ \forall \ y \in A\}$. Say $x \trianglerighteq y \ \forall \ y \in A$. Then, $x \in C(\{x, y\})$ for all $y \in A$. Now $C(A)$ must be non-empty, so either $x \in C(A)$ (in which case we are done), or $y \in C(A)$ for $y \neq x$. By property α, this implies that $\{x, y\} = C(\{x, y\})$, and so by property β, $x \in C(\{x, y\})$*

This shows that properties α and β are sufficient for rationalizability ■

Proof (representation implies axioms). *Homework* ■

This completes the proof of our first representation theorem.

One final thing to note. The preference relation that rationalizes a complete set of choice data is unique. For completeness we will prove this claim as well:

Theorem 2 *Let C be a choice correspondence that satisfies properties α and β. There is one and only one preference relation that rationalizes C*

Proof. *The fact that there is such a preference relation we have already proved. We will prove uniqueness by contradiction. Say $\succeq_1$and $\succeq_2$ both rationalized C ,and $\succeq_1 \neq \succeq_2$. Without loss of generality, this implies that there exists an x and y such that $x \succeq_1 y$ but not $x \succeq_2 y$. But the former statement implies that $x \in C(\{x, y\})$ while the latter implies $x \notin C(\{x, y\})$, a contradiction.*

■