Mode Choice — Models

Philip A. Viton

October 30, 2013
We consider an individual choosing among the available transportation modes for a specific trip purpose.

The set of available modes is called the individual’s *Choice Set*.

Our task is to understand an individual’s decision (behavior) — which mode she selects — in terms of *observables*: that is, determinants of behavior which we as analysts can observe, measure, and potentially change.
We assume that once she has decided on a mode, she utilizes *only* that mode. We will say that she consumes 1 unit of the mode.

This means that we will need to define “compound modes” so that if she decides to access transit by (say) walking, this will be a different mode than if she decides to get to the transit system by car.

This setup — the individual has a set of alternatives and must choose exactly one — is known as *Discrete Choice*.

It differs from a more common situation where the individual can choose any quantities from any of the available alternatives (think of grocery shopping: in the vegetable department you can choose as many as you want; you are not restricted to buying just one vegetable).
The Discrete Choice setup applies to many more settings than simply mode choice, and the methods we will be developing can be used, with only a change in terminology, to analyze a wide variety of choice settings.

Examples:

- Individual choice of a place (location) to live
- Individual choice of a vacation destination
- Choice of a car or major appliance model to purchase
- Choice of which university (from among those where you’ve been accepted) to attend
- Appellate judge: vote to affirm or overrule the lower court decision
- City council member: vote for or against a project/plan

(Note that in each of these examples we will assume that you can choose just one of the available alternatives).
We now focus explicitly on the mode-choice decision.

- Consider an individual making repeated mode choices in a situation where the observables — the observable characteristics of the modes and of the individual herself — do not change.
- Do we expect that she will make the same choice every time? That is, would we surprised if she made different choices on different days, even if the observable determinants of her decision were unchanged?
- Alternatively: suppose we knew all and only the observables. Are we confident that we could predict her choices each and every time?
Of course, the answer is No, and the reason is clear: in addition to the determinants of her decision that we can observe, the individual also takes into account factors that she knows, but which we (as analysts) do not observe.

For example, we may not know that her car is in the shop for repairs on one day (and that would incline her to use transit); or that she has a doctor’s appointment after work and needs to arrive promptly (and that would incline her to use her car).
We can summarize our analytical situation in two ways:

- There are determinants of her choice that we as analysts do not observe: that is, our knowledge of the determinants of her behavior is incomplete.
- From our perspective, there appears to be an element of randomness in her decision-making: when the observables do not change her behavior may change for no observable reason.

Note that this is not to say that our individual’s choice is random: it is just the way things appear to us, as analysts trying to understand (explain) her decision.

Our analysis will need to take these sources of variation into account if we are to make plausible inferences about her decisions.
The natural way to do this is to employ notions of probability. Probability is designed to handle the situation where there are several potential outcomes (here: choices) and we cannot say exactly which will occur.

By focussing on the probability that the individual makes a given choice we will be incorporating our lack of complete knowledge about out individual’s decision-making.
Individual Behavior (I)

We now describe the discrete choice setting more formally.

- Individual $i$ has $J$ modes indexed by $j$ ($j = 1, 2, \ldots, J$) to choose from. We say that her choice set is of size $J$ (is $J$-dimensional).
- There is a vector of observable characteristics of mode $j$ as applicable to individual $i$. Here, observable means observable to us: we can gather (or have gathered) data on these characteristics.
- We assume that there are $K$ observables and write them, for individual $i$ and mode $j$ as $x_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijK})$ Note that we are saying that all modes can be described by the same $K$ characteristics, though some modes may have values indicating that the characteristic may not be present or relevant.
- Individual $i$ is a characteristics-taker: her decision whether or not to use a particular mode will not alter their characteristics.
Individual Behavior (II)

We will assume that individual $i$ makes her choice in two steps:

1. She examines all the characteristics of each mode in her choice set, and constructs a single summary value for each mode. Note that this summary involves all the modal characteristics, including the ones that are not observable to us (the analysts).

2. She chooses the mode (for this trip) that has the highest summary number (ie is best for her).
Individual Behavior (III)

- We will say that individual $i$ summarizes all the characteristics of each available mode via a special function called a *utility* function. The summary value associated with a mode is her *utility* from that mode.

- If individual $i$ selects mode $j$ she gets utility $u_{ij}$ from it (this is sometimes called her *conditional utility*, since it is conditional on her having selected mode $j$).

- We will assume that utility function has the property that $u_{ij} > u_{ik}$ if and only if individual $i$ regards mode $j$ as better than mode $k$. We say that the utility function *represents* her preferences.

- It can be shown that under very general conditions a utility function that represents an individual’s preferences exists.

- Individual $i$ chooses her best (most-preferred) mode: she is a *utility-maximizer*. 
Structure of Utility

As we have argued, individual choice depends on:

- Factors observable to the analyst (e.g., prices or travel times of the modes).
- Factors unobservable to the analyst, but known to the decision-maker (e.g., whether the individual’s last experience with a particular mode was good/bad).
In light of this, we now assume that mode $j$’s utility for individual $i$:

- Has an observable (*systematic*) component that depends on the observable factors influencing her choice. We write this as $v_{ij}(x_{ij})$ or $v_{ij}$ for short.

- Has an unobservable (*idiosyncratic*) component representing determinants of choice that we as analysts do not observe (but that are known to the decision-maker).

- Following our earlier suggestion that from our perspective these unobservables induce an appearance of randomness in the individual’s decision-making, we represent the idiosyncratic component by a random variable $\eta_{ij}$.
So utility $u_{ij}$ is represented from the analyst’s perspective as:

$$u_{ij} = v_{ij} + \eta_{ij}$$

where $v_{ij}$ is the systematic part of utility and $\eta_{ij}$ is the idiosyncratic part.

The random variable $\eta_{ij}$ represents the analyst’s ignorance of all factors influencing individual $i$: it does not imply that the individual behaves randomly (from his/her own perspective).
Implications for the Study of Choice

From the analyst’s perspective:

- (Conditional) utility for \(i\) if mode \(j\) is selected:

\[ u_{ij} = \nu_{ij} + \eta_{ij} \]

- Because of the random component \(\eta_{ij}\), the utility that \(i\) gets if she chooses mode \(j\), \(u_{ij}\), is also a random variable.

- Thus, the event "for individual \(i\), the utility of mode \(j\) is highest" is also a random variable.

- This means that we can analyze only the probability that mode \(j\) maximizes \(i\)'s utility.

- This will be individual \(i\)'s choice probability for mode \(j\).
We now study what this means in more detail.

- We begin with the situation where the individual faces a *binary choice*: her choice set has two elements (modes; say, auto and transit).
- Mode 1 has utility (for individual $i$) of:
  \[ u_{i1} = v_{i1} + \eta_{i1} \]
- And similarly for mode 2:
  \[ u_{i2} = v_{i2} + \eta_{i2} \]
Because the individual is a utility maximizer, she will choose the mode that gives her the highest total utility.

And since utility $u_{ij}$ is a random variable, we focus on the probability that mode 1 (say) is best for individual $i$:

$$P_{i1} = \Pr[u_{i1} \geq u_{i2}]$$

$P_{i1}$ is individual $i$’s choice probability for mode 1. Clearly, in a binary choice setting, $P_{i2} = 1 - P_{i1}$.

We will assume that ties (the case $u_{i1} = u_{i2}$) are settled in favor of mode 1. If the $\eta$’s are continuous random variables, the event of a tie has probability zero, so this is a harmless assumption.
We now derive an expression for the choice probability $P_{i1}$ in a binary choice setting. We have:

\[
P_{i1} = \Pr[u_{i1} \geq u_{i2}]
\]

\[
= \Pr[v_{i1} + \eta_{i1} \geq v_{i2} + \eta_{i2}]
\]

\[
= \Pr[v_{i2} + \eta_{i2} \leq v_{i1} + \eta_{i1}]
\]

\[
= \Pr[\eta_{i2} - \eta_{i1} \leq v_{i1} - v_{i2}]
\]

where we have collected the random variables (the $\eta$’s) on the left-hand side of the inequality, and the non-random terms — the $v$’s — on the right-hand side. So in summary:

\[
P_{i1} = \Pr[\eta_{i2} - \eta_{i1} \leq v_{i1} - v_{i2}]
\]
We recognize $P_{ij}$ as the cumulative distribution function (cdf) of the random variable $\eta_i^2 - \eta_i^1$ evaluated at the “point” $v_i^1 - v_i^2$. If $\eta_i^2 - \eta_i^1$ has cdf $G(\cdot)$ then we can write:

$$P_{i1} = G(v_i^1 - v_i^2)$$

Alternatively, if $g(\cdot)$ is the probability density (frequency) function (pdf) of the random variable $\eta_i^2 - \eta_i^1$ then

$$P_{i1} = \int_{-\infty}^{v_i^1 - v_i^2} g(\eta_i^2 - \eta_i^1).$$

That is, the choice probability is the area under the probability density function from $-\infty$ to $v_i^1 - v_i^2$, the difference in the systematic utilities.
Binary Choice — Probabilities (cdf)

\[ P_{i1} = G(\eta_{i2} - \eta_{i1}) \]

\[ V_{i1} - V_{i2} \]
Binary Choice — Probabilities (cdf)

\[ P_{i1} = \int_{-\infty}^{v_{i1}-v_{i2}} g(\eta_{i2}-\eta_{i1}) \]

\[ V_{i1} - V_{i2} \]
To make these expressions usable in practice, we now need to say something more specific about:

- The (joint) distribution of the random variables $\eta_{i1}$ and $\eta_{i2}$.
- The way in which the systematic utility ($v_{ij}$) depends on the observables ($x_{ij}$).
We begin with the distributions of the random variables. As regards this, there are two important observations:

1. *Any* assumption about the joint distribution of the two random (idiosyncratic) elements $\eta_{ij}$ will yield a choice model. There is usually no theoretical basis for preferring one distributional assumption over another.

2. However, largely for practical (often computational) reasons, the literature has been dominated by two assumptions.
Assume that $\eta_{i1}$ and $\eta_{i2}$ have independent $N(0, \frac{1}{2})$ distributions. (The second parameter is the variance, not the standard deviation).

Then $\eta_{i2} - \eta_{i1}$ has a $N(0, 1)$ distribution and:

$$P_{i1} = G(\nu_{i1} - \nu_{i2}) = \Phi(\nu_{i1} - \nu_{i2})$$

where $\Phi$ is the standard normal cumulative distribution function (cdf).

This defines the Binary Probit model of (mode) choice.
To evaluate mode choice probabilities according to this model, you need to use tables of the normal cdf $\Phi$, available in almost any statistics textbook. A useful relation is $\Phi(-a) = 1 - \Phi(a)$.

Examples: suppose $v_{i1} - v_{i2} = 1.3$. Then from tables of the normal distribution:

$$P_{i1} = \Phi(1.3)$$
$$= 0.9032$$

If $v_{i1} - v_{i2} = -0.5$ then:

$$P_{i1} = \Phi(-0.5)$$
$$= 1 - \Phi(0.5)$$
$$= 1 - 0.69146$$
$$= 0.3085$$
Binary Choice — Probit Model (III)

Note that these results make intuitive sense:

- When $v_{i1} - v_{i2} = 1.3$, we are saying that the observable characteristics for mode 1 appear better to individual $i$ than the observable characteristics of mode 2.

- This does not imply that she will select mode 1 (remember that her choice will also involve the unobservables); but it does suggest that the probabilities favor mode 1; in our case we calculate that in fact $P_{i1} = 0.9032$.

- On the other hand, when $v_{i1} - v_{i2} = -0.5$, we are saying that the observable characteristics for mode 2 appear better to individual $i$ than the observable characteristics of mode 1.

- This suggests that the probabilities favor mode 2; and in fact we calculate that here $P_{i1} = 0.3085$ and hence that $P_{i2} = 1 - P_{i1} = 1 - 0.3085 = 0.6915$. 

Philip A. Viton

October 30, 2013
Suppose on the other hand that $\eta_{i1}$ and $\eta_{i2}$ are independent random variates with the Type-1 Extreme Value (T1EV) distribution, whose frequency ($g$) and distribution ($G$) functions are:

$$g(\eta_{ij}) = e^{-\eta_{ij}} e^{-e^{-\eta_{ij}}} \quad \text{and} \quad G(\alpha) = e^{-e^{-\alpha}}$$

(also known as the Weibull or Gnedenko distribution).

Then it can be shown (see separate optional handout) that

$$P_{i1} = \frac{e^{\nu_{i1}}}{e^{\nu_{i1}} + e^{\nu_{i2}}} = \frac{1}{1 + e^{-(\nu_{i1} - \nu_{i2})}}$$

This defines the Binary Logit model of (mode) choice.
Note that the logit probabilities can be evaluated without tables — all you need is a calculator.

Examples: suppose \( v_{i1} - v_{i2} = 1.3 \), then:

\[
P_{i1} = \frac{1}{1 + e^{-(1.3)}}
\]

\[
= 0.785 835 0
\]

If \( v_{i1} - v_{i2} = -0.5 \) then:

\[
P_{i1} = \frac{1}{1 + e^{-(0.5)}}
\]

\[
= 0.377 540 7
\]

Note that the probit and logit models give different results for the choice probabilities: the probability model you select matters.
Multinomial Choice

- We now extend our work to the general case of a $J$-dimensional choice set (individual $i$ can choose from among $J$ available modes).
- This is known as multinomial choice.
- As before, $i$ will select mode $j$ if it yields the highest utility.
- Mode $j$ is best for $i$ if:
  - mode $j$ is better than mode 1 for $i$ AND
  - mode $j$ is better than mode 2 for $i$ AND
  - $\ldots$ AND $\ldots$
  - mode $j$ is better than mode $J$ for $i$

  (this is $J - 1$ conditions).
Multinomial Choice Probabilities (I)

Translating this into symbols:

- Mode $j$ is better than mode 1 if $u_{ij} \geq u_{i1}$, or if
  \[ v_{ij} + \eta_{ij} \geq v_{i1} + \eta_{i1} \]

- Mode $j$ is better than mode 2 if $u_{ij} \geq u_{i2}$, or if
  \[ v_{ij} + \eta_{ij} \geq v_{i2} + \eta_{i2} \]

- Mode $j$ is better than mode $k$ if $u_{ij} \geq u_{ik}$, or if
  \[ v_{ij} + \eta_{ij} \geq v_{ik} + \eta_{ik} \]
Multinomial Choice Probabilities (II)

So

\[ P_{ij} = \Pr [u_{ij} \geq u_i, u_{ij} \geq u_i, \ldots, u_{ij} \geq u_i] \]
\[ = \Pr [v_{ij} + \eta_{ij} \geq v_i, v_{ij} + \eta_{ij} \geq v_i, \ldots, \]
\[ v_{ij} + \eta_{ij} \geq v_j + \eta_{ij}, ] \]
\[ = \Pr [v_{i1} + \eta_{i1} \leq v_{ij} + \eta_{ij}, v_{i2} + \eta_{i2} \leq v_{ij} + \eta_{ij}, \ldots, \]
\[ v_{iJ} + \eta_{iJ} \leq v_{ij} + \eta_{ij} ] \]
\[ = \Pr [\eta_{i1} - \eta_{ij} \leq v_{ij} - v_{i1}, \eta_{i2} - \eta_{ij} \leq v_{ij} - v_{i2}, \ldots, \]
\[ \eta_{iJ} - \eta_{ij} \leq v_{ij} - v_{iJ} ] \]

\((J - 1 \text{ terms in each expression; decide ties in favor of mode } j)\)
To clean this up a bit, write the difference $\eta_{i1} - \eta_{ij}$ as $\eta_{i1j}$, or more generally, $\eta_{ik} - \eta_{ij} = \eta_{ikj}$.

And write $v_{ij} - v_{i1}$ as $v_{ij1}$, or more generally, $v_{ij} - v_{ik} = v_{ijk}$.

Then we can write:

$$P_{ij} = \Pr[\eta_{i1j} \leq v_{ij1}, \eta_{i2j} \leq v_{ij2}, \ldots, \eta_{ijJ} \leq v_{ijJ}]$$

We see that the choice probability is the joint cumulative distribution function of the random variables $(\eta_{i1j}, \eta_{i2j}, \ldots, \eta_{ijJ})$ evaluated at the “point” $(v_{ij1}, v_{ij2}, \ldots, v_{ijJ})$.

As before, any assumption about the joint distribution of the random variables (the $\eta_{ij}$ or the $\eta_{ikj}$) will give rise to a choice model.
Multinomial Probit Model

- If we assume that the \((\eta_{i1}, \eta_{i2}, \ldots, \eta_{iJ})\) are distributed as multivariate normal with mean vector 0 (this is a \(J\)-vector of 0’s) and a \(J \times J\) covariance matrix \(\Sigma\), this defines the *multinomial probit* model.

- The difficulty with this is that in order to evaluate these choice probabilities we need to perform a \(J - 1\)-dimensional integration of the multivariate normal distribution.

- For \(J\) greater than about 4, until very recently this has been regarded as effectively impossible.

- So for moderately high-dimensional choice sets, there has been very little use of the multinomial probit model.
On the other hand, assume that the $\eta_{ij}$ are distributed as independent T1EV random variates.

Then the choice probabilities are given by:

$$P_{ij} = \frac{e^{v_{ij}}}{\sum_{k=1}^{J} e^{v_{ik}}}$$

known as the (Multinomial) Logit model. (See separate handout for a derivation).

Note that, like the binary logit model, the choice probabilities can be evaluated using just a pocket calculator — no integrations necessary.

For this reason, the logit model has dominated the literature. But note that there is a loss of generality: the multinomial probit model allows for correlations among the unobserved characteristics of the modes while the multinomial logit model rules that out.
Structure of Systematic Utility

- Suppose mode $j$ has $K$ observable characteristics (as applicable to individual $i$): $x_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijK})$

- It is conventional to assume that systematic utility $v_{ij}$ is linear-in-parameters:

  $$v_{ij} = x_{ij} \beta = x_{ij1} \beta_1 + x_{ij2} \beta_2 + \cdots + x_{ijK} \beta_K$$

- So in the logit case the choice probabilities become:

  $$P_{ij} = \frac{e^{x_{ij} \beta}}{\sum_{k=1}^{J} e^{x_{ik} \beta}}$$

  *in which the only unknown is the weighting vector $\beta$.*

- So our remaining task is to say something about $\beta$. This is an estimation problem, and we turn next to strategies to allow us to do this.