



# An Introduction to Discrete Choice Theory (for ABM)

Discrete choice for ABM

---

Debraj Roy

June 6, 2024

Computational Science Lab

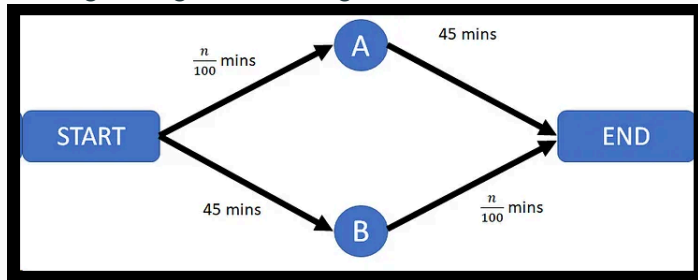
# Table of contents

1. Introduction
2. Linear probability model
3. Logit Model
4. Generalised Extreme Value
5. Probit Model
6. Summary & Applying Discrete Choice

# Introduction

# Discrete Choice

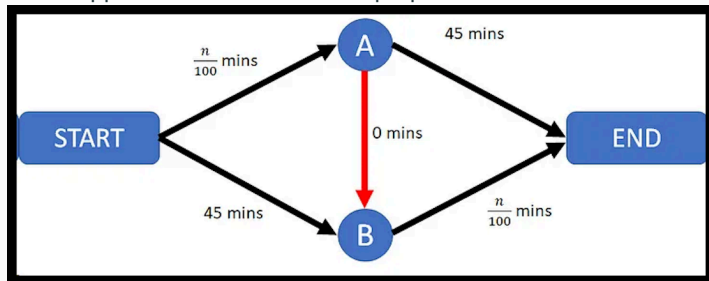
Suppose 4000 commuters need to get from START to END every morning, through the following network:



We assume that they each act independently (without cooperation) to minimise their own travel time.

# Braess Paradox

Now suppose that a new road is proposed. Should it be built?



Braess, D., A. Nagurney, T. Wakolbinger. 2005. On a paradox of traffic planning. *Transportation Sci.* 39 446–450.

# Why Discrete Choice?

Discrete Choice Theory/Modelling tries to model the way individuals (or collectives) make a choice from a selection (two or more) of discrete alternatives.

Applied in:

# Why Discrete Choice?

Discrete Choice Theory/Modelling tries to model the way individuals (or collectives) make a choice from a selection (two or more) of discrete alternatives.

Applied in:

**economics** : What car to buy?



# Why Discrete Choice?

Discrete Choice Theory/Modelling tries to model the way individuals (or collectives) make a choice from a selection (two or more) of discrete alternatives.

Applied in:

**economics** : What car to buy?

**transportation** : Which mode of transport to take?





# Why Discrete Choice?

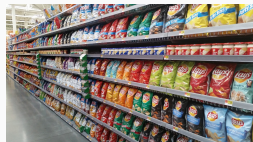
Discrete Choice Theory/Modelling tries to model the way individuals (or collectives) make a choice from a selection (two or more) of discrete alternatives.

Applied in:

**economics** : What car to buy?

**transportation** : Which mode of transport to take?

**business** : What people buy?



# Probabilities?

Who owns an electric vehicle, how will that change in the future?

Assumption that education level is important

Sample of 600 people randomly selected. Asked if they own an EV and how many years of education they have.

Table 1: Electric Vehicle Ownership

EV	low	med	high	
Y	10	100	120	230
N	140	200	30	370
	150	300	150	600

# Probabilities?

Dependent variable:

$$y = \begin{cases} 1 & \text{owner} \\ 2 & \text{not owner} \end{cases}$$

Independent variable (education level):

$$x = \begin{cases} 1 & \text{low} \\ 2 & \text{med} \\ 3 & \text{high} \end{cases}$$

# Probabilities?

We have a sample and use the sample population to estimate population.

$$\hat{p}(y = 1) = p(y = 1)$$

Joint probabilities:

$$\hat{p}(y = 1, x = 2) = \frac{100}{600} = 0.1667$$

Marginal probabilities:

$$\hat{p}(y = 1) = \sum_{k=1}^3 \hat{p}(1, k) = \frac{10}{600} + \frac{100}{600} + \frac{120}{600} = 0.383$$

Conditional probabilities:

$$\begin{aligned}\hat{p}(y = 1|x = 2) &= \hat{p}(y = 1, x = 2)/\hat{p}(x = 2) \\ &= 0.1667/0.5 = 0.333\end{aligned}$$

$$\hat{p}(y = 1|x = 1) = 0.067$$

$$\hat{p}(y = 1|x = 3) = 0.8$$

This is a simple model to predict  $y$  given  $x$ . Note this assumes things don't change.

# Probability Model

Conditional probabilities:

$$\hat{p}(y = 1|x = 1) = \pi_1 = 0.067$$

$$\hat{p}(y = 1|x = 2) = \pi_2 = 0.333$$

$$\hat{p}(y = 1|x = 3) = \pi_3 = 0.8$$

Future level of education changes (0.1, 0.6, 0.3)

$$\begin{aligned} p(y = 1) &= \sum_i p(y = 1|x = i)p(x = i) \\ &= 0.1\pi_1 + 0.6\pi_2 + 0.3\pi_3 \\ &= 0.4467 \end{aligned}$$

So if education level changes from (0.25,0.5,0.25) to (0.1,0.6,0.3) then ownership increases from 38.33% to 44.67%

Ben-Akiva and Lerman (Ben-Akiva and Lerman 1985) define a choice as a sequence of steps:

- Definition of the choice problem
- Generation of Alternatives
- Evaluation of attributes of alternatives
- Choice/Selection
- Implementation

Choice set is finite, selection is made based on *attributes* of choices.  
*Decision rule* defines how the choice is made from attributes.

# Discrete Choice Modelling

There is a decision maker (agent) and a choice set of discrete options. There are other theories that deal with continuous choice options, but the constraints on the choice set discrete choice are:



There is a decision maker (agent) and a choice set of discrete options. There are other theories that deal with continuous choice options, but the constraints on the choice set discrete choice are:

- Choices are mutually exclusive (you cannot choose two options together)

# Discrete Choice Modelling

There is a decision maker (agent) and a choice set of discrete options. There are other theories that deal with continuous choice options, but the constraints on the choice set discrete choice are:

- Choices are mutually exclusive (you cannot choose two options together)
- The Choice Set must be exhaustive - for your model to work

# Discrete Choice Modelling

There is a decision maker (agent) and a choice set of discrete options. There are other theories that deal with continuous choice options, but the constraints on the choice set discrete choice are:

- Choices are mutually exclusive (you cannot choose two options together)
- The Choice Set must be exhaustive - for your model to work
- Number of alternatives must be finite

Note that there is freedom in designing the choice set and that will greatly impact the derived model. If two options  $A$  and  $B$  are similar they can be grouped.

## Attributes

Each alternative in the choice set has a number of attributes that determine the decision. (e.g., price, travel time, etc.)

## Decision Rule

A way to weigh up the alternatives and decide one option:

- Dominance
- Satisfaction
- Utility
- etc.

## Assumption

A decision maker evaluates each alternative using a utility function, then chooses the option that has the maximal utility.

if  $U(car) > U(bus)$  then *car*

Assume utility is a linear combination of the parameters

$$U(car) = \beta_1(cost) + \beta_2(traveltime) + \beta(parkingcost) + \dots$$

Parameters represent tastes - which might vary per person (agent), or by type of agent. These include socio-economic characteristics (e.g., age, income, gender)

# Discrete Choice Theory

Table 2: Choices for travelling to work

Alternatives	time (t)	cost (p)	comfort (o)
car	$t_c$	$c_c$	$o_c$
bus	$t_b$	$c_b$	$o_b$
walk	$t_w$	$c_w$	$o_w$

$$U_i = U(t_i, c_i, o_i)$$

Only one alternative is possible, and we pick  $i$  such that  $U_i > U_k \forall k$

# Discrete Choice Theory - Utility Function

More generally Assume additive utility function of the form:

$$U_i = -\beta_t t_i - \beta_c c_i + \beta_o o_i$$

Generally, a decision maker  $n$  makes a choice among  $J$  alternatives. The decision maker picks alternative  $i \in J$  where the utility is maximised:

$$U_{ni} > U_{nj} \quad \forall j \in J \quad j \neq i$$

## Estimating the parameters

In discrete choice theory typically the researcher/modeller is trying to estimate which attributes are most important to the decision maker in the decision process.

## Observations

The researcher observes some attributes of the alternatives  $x_{nj}$ , and some attributes of the decision maker labelled  $s_n$ . A function is then defined that relates these attributes to the utility of alternative  $j$

$$V_{nj} = V(x_{nj}, s_n)$$

This is the representative utility



## Missing information

The researcher will use data to estimate things from a sample of the population (data set), there will be some attributes that the researcher cannot measure. The random utility model explicitly adds these unobserved (or error) terms as  $\epsilon_{nj}$ . These are the terms that impact the utility but are not captured by  $V_{nj}$

More generally the sources of uncertainty might be:

- Unobserved attributes
- Unobserved taste variations
- Measurement error

# Random Utility Model

Assumes that instead of utility being a real value, it is instead defined by a function and is subject to noise/uncertainty.

Dependent variable is latent and the expected utility of choice  $j$  for decision maker  $n$  is defined as:

$$U_{nj} = V_{nj} + \epsilon_{nj}$$

Where  $\epsilon_{nj}$  captures all the factors that impact utility but are not included in  $V_{nj}$  – the unknowns.

$\epsilon_{nj}$  is just defined as the difference between  $V_{nj}$  and  $U_{nj}$ . Therefore the characteristics (e.g., the distribution) of  $\epsilon_{nj}$  depend on the researchers specification of  $V_{nj}$ .

$\epsilon_{nj} \forall j$  is not known and therefore treated as random. The joint density of the random vector is denoted as  $f(\epsilon_n)$ , where:

$$\epsilon_n = \langle \epsilon_{n1}, \dots, \epsilon_{nJ} \rangle$$

The the probability that decision maker  $n$  chooses alternative  $i$

$$\begin{aligned}P_{ni} &= \text{Prob}(U_{ni} > U_{nj} \quad \forall j \neq i) \\&= \text{Prob}(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \quad \forall j \neq i) \\&= \text{Prob}(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i)\end{aligned}$$

$$\begin{aligned} P_{ni} &= \text{Prob}(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) \\ &= \int_{\epsilon} I(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) f(\epsilon_n) \delta \epsilon_n \end{aligned}$$

$I(x)$  is the indicator function which is 1 if the predicate  $x$  is true, otherwise 0.

This is a multidimensional integral over the density of the unobserved portion of utility. Different discrete choice models are obtained from different assumptions about the distribution of the unobserved portion of utility

$$P_{ni} = \int_{\epsilon} I(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) f(\epsilon_n) \delta \epsilon_n$$

The integral takes a closed form only for certain  $f(\epsilon_n)$ .

Different forms of choice model work by assuming that the unobserved portion of the utility follows different distributions:

**Logit:** Extreme value distribution (iid)

**Nested Logit:** Generalised extreme value

**Mixed Logit:** Extreme value distribution (iid) and something else

**Probit:** Multi-variate normal

Only nested logit and logit have closed form solutions for the integral - the others are estimated numerically.

## Meaning of choice - Example

$$V_c = \alpha T_c + \beta M_c$$

$$V_b = \alpha T_b + \beta M_b$$

Where  $T$  and  $M$  represent the time and money for each mode of transport car ( $c$ ) and bus ( $b$ ).  $\alpha$  and  $\beta$  are estimated by the modeller.

Suppose,  $V_c = 3$  and  $V_b = 4$ , based on the observed factors bus is the better option. The final choice depends on the unobserved factors - if they are sufficiently better for bus i.e.,  $\epsilon_c - \epsilon_b < 1$  then  $U_b > U_c$ .

## **Linear probability model**



# Types of Choice Model

The models can be classified according to the number of alternatives:

- Binary (Binomial) choice models (2 alternatives)
- Multinomial choice models ( $> 2$  alternatives)
  - Assume no correlation in unobserved factors
  - Allow correlation in unobserved factors

We will only look at the binary choice in lectures - but you can look in detail at other options.

$$y = \beta_0 + \beta_1 x + \epsilon; i.i.d(0, \sigma^2)$$

$$y = \begin{cases} 0 & \text{not} \\ 1 & \text{college} \end{cases}$$

$$E(y|x) = \sum_i P(y = y_i|x) y_i = P(y = 1|x)$$

$$E(y|x) = \beta_0 + \beta_1 x = P(y = 1|x)$$

# Problems with LPM

Suppose: You want to predict if an individual will go to college based on their parental wage. You collect data and come up with a linear probability model

$$y = 0.3 + 0.2lp_{wage} + \epsilon; i.i.d(0, \sigma^2)$$

$$P(y = 1|x) = 0.3 + 0.2lp_{wage}$$

when  $lp_{wage} = -5$ , then  $P(y = 1|x) = -0.7$

when  $lp_{wage} = 10$ , then  $P(y = 1|x) = 2.3$

But probabilities  $\in [0, 1]$

Solution:

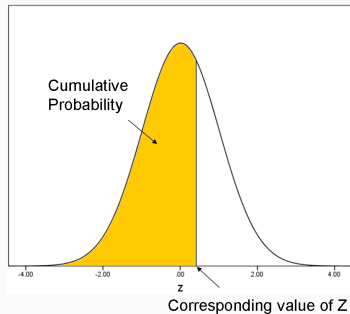
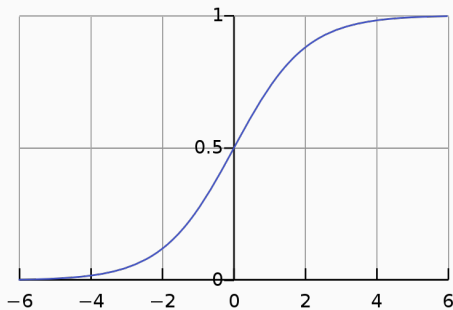
$$P(y = 1|x) = F(0.3 + 0.2/p_{wage}) \in [0, 1]$$

where  $F(-\infty) = 0$ , and  $F(+\infty) = 1$

Can you think of reasonable  $F()$ ?

# Problems with LPM

Solution:



## Taste variation (endogeneity)

Unobserved heterogeneity is typically about unobservable components of the effects that you are estimating. Unobserved heterogeneity might be that some people have higher returns (e.g., increases in wages) from going to college than others. Let the returns for person  $i$  be  $\beta + b_i$  with  $\mathbb{E}(b_i) = 0$ . We have

$$y_i = x_i(\beta + b_i) + w_i'\gamma + \epsilon_i,$$

where  $y_i$  is (typically, log) income,  $x_i$  is years of education, and  $w_i$  is a set of other controls. An example of endogeneity is when  $x_i$  is correlated with  $\epsilon_i$  (e.g., education is correlated with IQ, which is not among our other predictors).

If we estimate a single coefficient, we have

$$y_i = x_i\beta + w_i'\gamma + (\epsilon_i + bx_i) = x_i\beta + w_i'\gamma + \tilde{\epsilon}_i$$

See that the included variable  $x_i$  is correlated with the error term  $\tilde{\epsilon}_i$ ,

# Logit Model

The logit model is perhaps the most widely used discrete choice model. It assumes the  $\epsilon_{ni}$  is iid extreme value for all  $i$ .

## **Independent and Identically Distributed (iid)**

For each choice  $i$ , the unobserved components  $\epsilon_{ni}$  are assumed to follow the same distribution and are mutually independent.

The critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives.



Is this assumption restrictive?

Recall that  $\epsilon_{nj}$  is the difference between the real utility  $U_{nj}$  and the observed utility  $V_{nj}$  - as specified by the modeller.

The independence assumption, the error of alternative  $i$  provides no information about the error of  $j$ . OR...

*The modeller specifies  $V_{nj}$  sufficiently well that the unobserved part  $\epsilon_{nj}$  is essentially white noise.*

# IID - when does this not hold

One example of when this fails:

We have a bag containing two coins, with different probability of heads  $p_1$  and  $p_2$ . Pick a coin and toss it  $N$  times,  $X_i$  is indicator function of heads on the  $i^{th}$  toss.

$$\begin{aligned}P\{X_i = 1\} &= \frac{1}{2}P\{X_i = 1|coin_1\} + \frac{1}{2}P\{X_i = 1|coin_2\} \\&= \frac{p_1 + p_2}{2}\end{aligned}$$

$$\begin{aligned}P\{X_i = 1, X_j = 1\} &= \frac{P\{X_i = 1, X_j = 1|coin_1\} + P\{X_i = 1, X_j = 1|coin_2\}}{2} \\&= \frac{p_1^2 + p_2^2}{2}\end{aligned}$$

$$P\{X_i = 1, X_j = 1\} \neq P\{X_i = 1\}P\{X_j = 1\}$$

They are not independent!

**Homework:** Under what condition do we get independence? Can you think of an intuitive explanation?

The reason for the Logit Model being so popular is simply one of convenience. This is at the cost of the assumption - unobserved factors could be related between choice.

E.g., People who don't chose bus because of overcrowding, might have the same reaction to rail travel.

$$U_{nj} = V_{nj} + \epsilon_{nj}$$

Logit model assumes each  $\epsilon_{nj}$  identically distributed extreme value. (Gumbel, Type I and extreme value). The extreme value distribution gives slightly fatter tails than a normal, which means it can include outliers. Usually, however, the difference between extreme value and independent normal errors is indistinguishable empirically.

Standard (maximum) Gumbel distribution is  $\mu = 0$  and  $\beta = 1$

$$f(x) = e^{-x}e^{-e^{-x}}$$

See:

<https://www.itl.nist.gov/div898/handbook/eda/section3/eda366g.htm>

Then the density for each unobserved component of utility is:

$$f(\epsilon_{nj}) = e^{-\epsilon_{nj}} e^{-e^{-\epsilon_{nj}}}$$

Cumulative

$$F(\epsilon_{nj}) = e^{-e^{-\epsilon_{nj}}}$$

Recall that the difference between the two random terms is crucial:

In the case of gumbel  $\epsilon_{nji}$

$$\epsilon_{nji} = \epsilon_{nj} - \epsilon_{ni}$$

Actually follows a logistic distribution.

$$F(\epsilon_{nji}) = \frac{e^{\epsilon_{nji}}}{1 + e^{\epsilon_{nji}}}$$

# Derive Logit Model

$$\begin{aligned}P_{ni} &= \text{Prob}(V_{ni} + \epsilon_{ni} > \epsilon_{nj} + V_{nj} \quad \forall j \neq i) \\&= \text{Prob}(\epsilon_{nj} < \epsilon_{ni} + V_{ni} - V_{nj} \quad \forall j \neq i)\end{aligned}$$

if  $\epsilon_{ni}$  is considered given, the above expression is the cumulative distribution for  $\epsilon_{nj}$  evaluated at  $\epsilon_{ni} + V_{ni} - V_{nj}$ . This cumulative distribution over all  $j \neq i$  is the product of all individual cumulative distributions:

$$P_{ni}|\epsilon_{ni} = \prod_{j \neq i} e^{-e^{(\epsilon_{ni} + V_{ni} - V_{nj})}}$$

# Derive Logit Model

$\epsilon_{ni}$  isn't given, so choice probability is integral of  $P_{ni}|\epsilon_{ni}$  over all values of  $\epsilon_{ni}$  weighted by density.

$$P_{ni} = \int \left( \prod_{j \neq i} e^{-e^{(\epsilon_{ni} + V_{ni} + V_{nj})}} \right) e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}}$$

The integral leads to a closed form expression - which is why we assume the Gumbel.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$



# Derive Logit Model

$\epsilon_{ni}$  isn't given, so choice probability is integral of  $P_{ni}|\epsilon_{ni}$  over all values of  $\epsilon_{ni}$  weighted by density.

$$P_{ni} = \int \left( \prod_{j \neq i} e^{-e^{(\epsilon_{ni} + V_{ni} + V_{nj})}} \right) e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}}$$

The integral leads to a closed form expression - which is why we assume the Gumbel.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

# Derive Logit Model

From the  $ij$ -response set  $\mathcal{R}^{J_i}$  the probability that player  $i \in N$  will choose action  $j$  is given by:

$$\begin{aligned}\sigma_{ij}(\bar{u}_i) &= P[\hat{u}_{ij} \geq \hat{u}_{ik} \quad \forall k = 1, \dots, J_i \quad k \neq j] \\ &= P[\bar{u}_{ij} + \mathcal{E}_{ij} \geq \bar{u}_{ik} + \mathcal{E}_{ik} \quad \forall k = 1, \dots, J_i \quad k \neq j] \\ &= P[\mathcal{E}_{ik} \leq \bar{u}_{ij} - \bar{u}_{ik} + \mathcal{E}_{ij} \quad \forall k = 1, \dots, J_i \quad k \neq j] \\ &= P[\mathcal{E}_{i1} \leq \bar{u}_{ij} - \bar{u}_{i1} + \mathcal{E}_{ij}, \dots, \mathcal{E}_{i(j-1)} \leq \bar{u}_{ij} - \bar{u}_{i(j-1)} + \mathcal{E}_{ij}, \\ &\quad \mathcal{E}_{i(j+1)} \leq \bar{u}_{ij} - \bar{u}_{i(j+1)} + \mathcal{E}_{ij}, \dots, \mathcal{E}_{iJ_i} \leq \bar{u}_{ij} - \bar{u}_{iJ_i} + \mathcal{E}_{ij}] \\ &= \int_{-\infty}^{+\infty} \left[ \prod_{\substack{k=1 \\ k \neq j}}^{J_i} P[\mathcal{E}_{ik} \leq \bar{u}_{ij} - \bar{u}_{ik} + \varepsilon_{ij}] \right] f_i(\varepsilon_{ij}) \, d\varepsilon_{ij}\end{aligned}$$

with pdf  $f_i(\varepsilon_{ij}) = \frac{dF_i(\varepsilon_{ij})}{d\varepsilon_{ij}} = \lambda \exp(-\exp(-\lambda\varepsilon_{ij}-\gamma)) \exp(-\lambda\varepsilon_{ij}-\gamma)$ , so we have

# Derive Logit Model

$$\sigma_{ij}(\bar{u}_i) = \int_{-\infty}^{+\infty} \left[ \prod_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik} + \varepsilon_{ij}) - \gamma)) \right] \lambda \exp(-\exp(-\lambda\varepsilon_{ij} - \gamma)) \exp(-\lambda\varepsilon_{ij} - \gamma) d\varepsilon_{ij}$$

Let  $z = \exp(-\lambda\varepsilon_{ij} - \gamma)$ , such that the above equation becomes

$$\begin{aligned} \sigma_{ij}(\bar{u}_i) &= \int_0^{+\infty} \left[ \prod_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik}))z) \right] \exp(-z) dz \\ &= \int_0^{+\infty} \exp\left(-z \left\{ \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik})) \right\}\right) \exp(-z) dz \\ &= \int_0^{+\infty} \exp\left(-z \left\{ 1 + \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik})) \right\}\right) dz \\ &= -\frac{1}{1 + \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik}))} \exp\left(-z \left\{ 1 + \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik})) \right\}\right) \Big|_0^{+\infty} \\ &= 0 - \left(-\frac{1}{1 + \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(-\lambda(\bar{u}_{ij} - \bar{u}_{ik}))}\right) = \frac{1}{1 + \frac{1}{\exp(\lambda\bar{u}_{ij})} \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \exp(\lambda\bar{u}_{ik})} \\ &= \frac{\exp(\lambda\bar{u}_{ij})}{\sum_{k=1}^{J_i} \exp(\lambda\bar{u}_{ik})} \end{aligned}$$

# Limitations of Logit Model

1. Taste Variation (possible but limited)
2. Independence of Irrelevant Alternatives (IIA)
3. Assumes repeated choice has no correlation in observed factors

The value or importance that decision makers place on each attribute of the alternatives varies, in general, over decision makers. Logit can do taste variation that varies systematically with respect to observed variables - but not random variation.

# Independence of Irrelevant Alternatives

If a choice problem is changed to include a new alternative, we need to know how that alternative will impact existing ratios.

If  $A$  is preferred to  $B$  out of the choice set  $\{A, B\}$ , introducing a third option  $X$ , expanding the choice set to  $\{A, B, X\}$ , must not make  $B$  preferable to  $A$ .

**For Logit:**

$$\begin{aligned}\frac{P_{ni}}{P_{nk}} &= \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} \\ &= \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni} - V_{nk}}\end{aligned}$$

The ratio only depends on  $i$  and  $k$  nothing else. The relative odds of choosing  $i$  over  $k$  is the same no matter what other alternatives are.

# Independence of Irrelevant Alternatives - example

The red-bus, blue-bus example.

Assume you have two choices, blue bus (bb) or car (c) which are equally likely.

$$P_{nc} = P_{nbb} = \frac{1}{2}$$

In this case the ratio of the probabilities is:

$$\frac{P_c}{P_{bb}} = 1$$

# Independence of Irrelevant Alternatives - example

Now assume that a red bus, which in terms of attributes and preference is exactly the same for the decision maker. So

$$\frac{P_{rb}}{P_{bb}} = 1$$

And due to the IIA assumption

$$\frac{P_c}{P_{rb}} = 1$$

Therefore,

$$P_{nc} = P_{nbb} = P_{nr} = \frac{1}{3}$$

A more reasonable assumption would be the probability of taking any bus is equal to taking the car, then which bus is irrelevant.

$$P_{nbb} = P_{nr} = \frac{1}{4} \quad P_{nc} = \frac{1}{2}$$



# Proportional Substitution

If we change the attribute of a particular alternative  $i$  so that  $i$  becomes more attractive. The probability increase in  $i$  results in a reduction in other choices  $j \neq i$ .

The logit model assumes that this reduction happens equally across all options. Changing  $i$  will impact the probability of  $k$  and  $l$  equally:

$$\frac{P_{nk}^1}{P_{nl}^1} = \frac{P_{nk}^0}{P_{nl}^0}$$

The ratio of the probabilities for  $k$  and  $l$  is the same after the change in  $j$  (1) as it was before the change

# **Generalised Extreme Value**

In generalized extreme value (GEV) models the unobserved portion of the utility for all alternatives are jointly distributed as a generalised extreme value (not Type I).

# Nested Logit Model

Nested Logit is most common GEV.

Nested Logit works by dividing the choice set into subsets called  nests .  
The nests are chosen such that:

- Choices within a nest are assumed to follow IIA.
- For choices in two different nests IIA does not hold

# Nested Logit Model - an example

Consider a choice between car alone, **car pool**, **bus** or **rail**.

**Assume that the original choice is**

$$P_{nc} = 0.4 \quad P_{np} = 0.1 \quad P_{nb} = 0.3 \quad P_{nr} = 0.2$$

**Then we examine the change in choice as we remove one option.**

Table 3: Change in probabilities when one choice removed

Choice	original	car	pool	Bus	Rail
Car	.40	–	.45	.52	.48
Pool	.10	.20	–	.13	.12
Bus	.30	.48	.33	–	.40
Rail	.20	.32	.22	.35	–

# Nested Logit Model - an example

Table 4: Change in probabilities when one choice removed

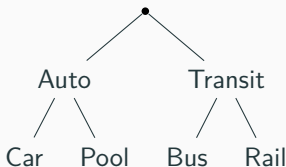
Choice	original	car	pool	Bus	Rail
Car	.40	–	.45 (+12.5%)	.52 (+30%)	.48 (+20%)
Pool	.10	.20 (+100%)	–	.13 (+30%)	.12(+20%)
Bus	.30	.48 (+60%)	.33 (+10%)	–	.40 (+33%)
Rail	.20	.32 (+60%)	.22(+10%)	.35(+70%)	–

Notice that the probabilities for rail and bus always increase by the same rate when one of the other alternatives is removed - IIA holds between these alternatives, so Train and Bus can be mapped to one *nest* called [transit](#). The same thing for car and pool - put these into a nest called [auto](#)

# Nested Logit Model - an example

You can visualise these nested choices as a tree. From this it's clear that the decision maker actually makes a two-phase choice.

First the choice of whether by car or public transport, then the choice of own car or pool and bus or rail.



In generalized extreme value (GEV) models the unobserved portion of the utility for all alternatives are jointly distributed as a generalised extreme value (not Type I).

This cumulative joint distributions allows for correlations over alternatives within nests.

$$\exp\left(-\sum_{k=1}^K \left(\sum_{j \in B_k} e^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right)$$

This distribution for the unobserved components of utility gives rise to the following choice probability for alternative  $i \in B_k$

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{\ell=1}^K \left(\sum_{j \in B_\ell} e^{V_{nj}/\lambda_\ell}\right)^{\lambda_\ell}}$$



# Probit Model

The probit models overcome the 3 problems of logit models: (taste variation, IIA, temporal correlation).

Probit models overcome this.

## **Probit Assumption**

A probit model assumes that the unobserved components of utility ( $\epsilon_{ni} \forall i$ ) are normally distributed.

Again utility is assumed to be observed and unobserved:

$$U_{nj} = V_{nj} + \epsilon_{nj} \quad \forall j$$

$\epsilon_n$  is then distributed normal with density:

$$\phi(\epsilon_n) = \frac{1}{(2\pi)^{J/2} |\Omega_n|^{1/2}} e^{-\frac{1}{2} \epsilon_n' \Omega_n^{-1} \epsilon_n}$$

where

$$\epsilon_n' = \langle \epsilon_{n1}, \dots, \epsilon_{nJ} \rangle \quad \Omega_n \text{ is the covariance matrix}$$

The choice probability is then

$$\begin{aligned} P_{ni} &= \text{Prob}(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \quad \forall j \neq i) \\ &= \int I(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \quad \forall j \neq i) \phi(\epsilon_n) \delta\epsilon_n \end{aligned}$$

Here there is no closed form solution, it's evaluated numerically in this case.

- Taste variation: Probit can do this.
- Substitution: estimating a full covariance matrix allows the model to represent any substitution pattern.
- IIA: Probit probabilities do not exhibit IIA.

$$U_{nj} = \beta_n x_{nj} + \epsilon_{nj}$$

$\beta_n$  is the vector of coefficients for decision maker  $n$  representing that person's tastes. Suppose the  $\beta_n$  is normally distributed in the population with mean  $b$  and covariance  $W$ :  $\beta_n \in N(b, W)$ . The goal is to estimate the parameters  $b$  and  $W$ .

The utility can be rewritten with  $x_{nj}$  decomposed into its mean and deviations from its mean:

$$U_{nj} = b_n x_{nj} + \beta_n \tilde{x}_{nj} + \epsilon_{nj}$$

The last two terms in the utility are random; denote their sum as  $\eta_{nj}$  to obtain  $U_{nj} = b_n x_{nj} + \eta_{nj}$ . The covariance of the  $\eta_{nj}$  's depends on  $W$  as well as the  $x_{nj}$  's, so that the covariance differs over decision makers.



## **Summary & Applying Discrete Choice**

## Mixed Logit

Allows taste variation, any substitution pattern and doesn't require IIA. It also does not require the unknowns are normally distributed.

The mixed logit requires simulation for estimating parameters and is still seen as 'the modern approach' despite being developed in the eighties.

The work comes about in estimating the values of  $\beta$  for the choice models. We find the values of  $\beta$  that best match the sample data we have collected.

There are different ways this can be done (provided in software normally):

- Maximum likelihood methods (numerical optimization)
- Bayesian Methods (Markov chain Monte Carlo)
- Simulation-based estimators (draw random variables to estimate integral)

We've discussed (briefly) most of the classic discrete choice models:

- Logit - simple form, disadvantages
- Nested Logit - IIA within a nest, but not between nests
- Probit - Normally distributed unknowns
- Mixed Logit - a very general form of choice model.

You can also consider cases of binary choice (2 alternatives) or multi-nomial choice.

Discrete Choice Theory provides a mature methodology for developing choice algorithms. If you are developing an ABM where to try to capture choice e.g.,

- Transport/Mobility modelling (Modal choice)
- Migration/Urban modelling (house choice)
- Economic Modelling (Purchase choice)
- Transportation (EV Charging station choice)

These models give you a way to develop a model of choice within an agent. You can consider to do this at a population level, or at an aggregate level, or at an individual level (depending on data)

## References



Ben-Akiva, Moshe E and Steven R Lerman (1985). *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press.