Discrete Choice Theory and Travel Demand Modelling

The Multinomial Logit Model

Anders Karlström

Division of Transport and Location Analysis, KTH

Jan 21, 2013

Introduction

This lecture:

- Basics of Discrete Choice Theory
- Application: Travel Demand Modelling
- Literature: Koppelman and Bhat (2006) ch. 1-4 (parts)

What is a model?

What is a model?

- Mathematical formulation of a relationship between causes and effects:

$$f : A \rightarrow B$$

What is a model?

What is a model?

- Mathematical formulation of a relationship between causes and effects:
$$f : A \to B$$
- Used for prediction of the future

What is a model?

What is a model?

- Mathematical formulation of a relationship between causes and effects:
$$f : A \to B$$
- Used for prediction of the future
  - Deterministic: predicts the effects with certainty

What is a model?

What is a model?

- Mathematical formulation of a relationship between causes and effects:

$$f : A \to B$$

- Used for prediction of the future
  - Deterministic: predicts the effects with certainty
  - Probabilistic: predicts probabilities of different effects

What is a model?

- Mathematical formulation of a relationship between causes and effects:

$$f : A \rightarrow B$$

- Used for prediction of the future
  - Deterministic: predicts the effects with certainty
  - Probabilistic: predicts probabilities of different effects
- Discrete Choice Models (DCM) are probabilistic models

What is a model?

What is a model?

- Mathematical formulation of a relationship between causes and effects:

$$f : A \to B$$

- Used for prediction of the future
  - Deterministic: predicts the effects with certainty
  - Probabilistic: predicts probabilities of different effects
- Discrete Choice Models (DCM) are probabilistic models
  - Binomial=only two alternatives

What is a model?

What is a model?

- Mathematical formulation of a relationship between causes and effects:
$$f : A \to B$$
- Used for prediction of the future
  - Deterministic: predicts the effects with certainty
  - Probabilistic: predicts probabilities of different effects
- Discrete Choice Models (DCM) are probabilistic models
  - Binomial=only two alternatives
  - Multinomial=more than two alternatives

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
    - Marketing: the choice of brand

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president
  - Economics: labour market participation

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president
  - Economics: labour market participation
  - Medicine: response to treatment

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president
  - Economics: labour market participation
  - Medicine: response to treatment
  - Transportation: the choice of travel mode

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president
  - Economics: labour market participation
  - Medicine: response to treatment
  - Transportation: the choice of travel mode
  - Planning: the choice of residential area (where to live)

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president
  - Economics: labour market participation
  - Medicine: response to treatment
  - Transportation: the choice of travel mode
  - Planning: the choice of residential area (where to live)
  - etc...

Discrete Choice Models (DCM)

- Predicts the probability that an alternative is chosen, given certain values of input variables
- The alternatives are given in advance – the *choice set*
- DCM are used in many contexts:
  - Marketing: the choice of brand
  - Politics: the choice of president
  - Economics: labour market participation
  - Medicine: response to treatment
  - Transportation: the choice of travel mode
  - Planning: the choice of residential area (where to live)
  - etc...
- Nobel prize in 2000 to Dan McFadden and Jim Heckman, see http://nobelprize.org/nobel_prizes/economics/laureates/2000/

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:
    - type, e.g. logit

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:
    - type, e.g. logit
    - structure, e.g. multinomial

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:
    - type, e.g. logit
    - structure, e.g. multinomial
    - explanatory variables, e.g. time and cost

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:
  - type, e.g. logit
  - structure, e.g. multinomial
  - explanatory variables, e.g. time and cost
- Estimation of model parameters

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:
  - type, e.g. logit
  - structure, e.g. multinomial
  - explanatory variables, e.g. time and cost
- Estimation of model parameters
- Application of the model to new data

Aim and goal of the model

What is the idea with a model?

- To *understand*, *characterise* and *predict* the behaviour of individuals in terms of choice

How?

- Observations of individual behaviour
- Definition of a model:
    - type, e.g. logit
    - structure, e.g. multinomial
    - explanatory variables, e.g. time and cost
- Estimation of model parameters
- Application of the model to new data
- This lecture gives the basics of the definition of a logit model

T

ravel demand

- Travel is the result of several choices:

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)
  - decision(s) of what route to follow

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)
  - decision(s) of what route to follow
- Each one of these decisions has to be modelled explicitly

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)
  - decision(s) of what route to follow
- Each one of these decisions has to be modelled explicitly
  - here we concentrate on "how" – mode

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)
  - decision(s) of what route to follow
- Each one of these decisions has to be modelled explicitly
  - here we concentrate on "how" – mode
    - in the project, also "where" – destination

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)
  - decision(s) of what route to follow
- Each one of these decisions has to be modelled explicitly
  - here we concentrate on "how" – mode
    - in the project, also "where" – destination
- Choices between several "discrete" (separate, non-overlapping) alternatives

T

ravel demand

- Travel is the result of several choices:
  - decision to travel (at all)
  - decision of what time to depart
  - decision of where to go - destination
  - decision of how to go (what mode of transport)
  - decision(s) of what route to follow
- Each one of these decisions has to be modelled explicitly
  - here we concentrate on "how" – mode
    - in the project, also "where" – destination
- Choices between several "discrete" (separate, non-overlapping) alternatives
- The order of the decisions can be modelled in several different ways

Simple illustrative example

- Choice of transport mode made by students of this course (fictive)
- Choice set: $M = \{\text{bicycle}, \text{t-bana}\}$
- Random sample of 30 students

	Programme			
	Trafiktek. $x = 1$	Stadsplan. $x = 2$	Master $x = 3$	
Bicycle ($y = 1$)	1	10	2	13
T-bana ($y = 2$)	5	4	8	17
	6 20%	14 45%	10 35%	30 100 %

Simple illustrative example

- Dependent variable (choice)

$$y = \begin{cases} 1 & \text{if bicycle} \\ 2 & \text{if t-bana} \end{cases}$$

- Independent (explanatory) variable

$$x = \begin{cases} 1 & \text{if Trafiktek.} \\ 2 & \text{if Stadsplan.} \\ 3 & \text{if Master} \end{cases}$$

Simple illustrative example

- Estimate the probability to choose bicycle, $\widehat{P}\left(y=1\right)$
- The *joint* probability of choosing bicycle *and* following the master programme is
$$\widehat{P}\left(y=1,x=3\right)=2/30=0.07$$
- The *marginal* probability of choosing bicycle is the *sum* of the joint probabilities over *all values* of the explanatory variable:

$$\widehat{P}(y=1)=\sum_{k=1}^{3}\widehat{P}(y=1,x=k)=\frac{1}{30}+\frac{10}{30}+\frac{2}{30}=\frac{13}{30}=0.43$$

Simple illustrative example

- Conditional probability

$$P(y = i | x = k) = \frac{P(y = i, x = k)}{P(x = k)}, \ P(x = k) > 0$$

- Conditional probability of choosing bicycle given student from the master program

$$\widehat{P}(y = 1 | x = 3) = \frac{2/30}{10/30} = 0.2$$

Simple illustrative example

- In the same way we obtain

$$\widehat{P}(y = 1|x = 1) = \widehat{\pi}_1 = 0.17$$
$$\widehat{P}(y = 1|x = 2) = \widehat{\pi}_2 = 0.71$$
$$\widehat{P}(y = 1|x = 3) = \widehat{\pi}_3 = 0.2$$

- $P(y = i|x = k) = \pi_k$ is the behavioral model, here we estimate it from the sample $(\widehat{\pi}_k)$

Simple illustrative example

Assumptions:

- Only the programme affiliation matters for mode choice!
- Only these two alternatives available
- Preferences are stable over time
- Then we can forecast the modal shares for another distribution of students

Simple illustrative example

- Shares before: 20% Trafiktek., 45% Stadsplan., 35% Master
- Future shares: 40% Trafiktek., 20% Stadsplan., 40% Master
- Probability of choosing bicycle under new scenario

$$\begin{aligned} P(y=1) &= \sum_{k=1}^{3} P(y=1|x=k)P(x=k) \\ &= 0.4\widehat{\pi}_1 + 0.2\widehat{\pi}_2 + 0.4\widehat{\pi}_3 \\ &= 0.29 \end{aligned}$$

Simple illustrative example

- $P(x = k)$ can easily be obtained and forecasted
- $P(y = i|x = k)$: simple behavioral model
- This lecture focuses on behavioral models, namely the probability of choosing an alternative given a set of available alternatives $M = \{\text{bicycle}, \text{t-bana}\}$ and the explanatory variable x (Programme)
- In this example, $P(\text{bicycle}|M, x)$ or simpler $P(y = 1|x)$

Discrete Choice Framework

- Decision-maker
  - individual (person/household)
  - socio-economic characteristics (age, gender, education etc.)
- A number of alternatives, i.e. the choice set
  - could be specific to the individual or household, i.e. dependent on the index t: $C_t = \{1, 2, \dots, J_t\}$ with J_t alternatives
- Attributes of each alternative (price, quality)
- A decision rule, e.g. utility maximisation

Utility

- Each discrete alternative is associated with a *utility* U_i

Utility

- Each discrete alternative is associated with a *utility* U_i
- Utility could be represented by any monotone function (strictly increasing or decreasing in its variables)

Utility

- Each discrete alternative is associated with a *utility* U_i
- Utility could be represented by any monotone function (strictly increasing or decreasing in its variables)
- The utility is determined by properties of both the *alternative* and the individual *decision-maker*

Utility

- Each discrete alternative is associated with a *utility* U_i
- Utility could be represented by any monotone function (strictly increasing or decreasing in its variables)
- The utility is determined by properties of both the *alternative* and the individual *decision-maker*
- The relative importance of each of the properties is determined by parameters

Utility

- Each discrete alternative is associated with a *utility* U_i
- Utility could be represented by any monotone function (strictly increasing or decreasing in its variables)
- The utility is determined by properties of both the *alternative* and the individual *decision-maker*
- The relative importance of each of the properties is determined by parameters
- The alternative with the highest utility, relative to all the other alternatives, is chosen

Utility

- Utility differences between alternatives determine which alternative is chosen

Utility

- Utility differences between alternatives determine which alternative is chosen
- Only differences matter: $U_i - U_j, j \neq i$

Utility

- Utility differences between alternatives determine which alternative is chosen
- Only differences matter: $U_i - U_j, j \neq i$
- The choice is invariant to translation and scale

Utility

- Utility differences between alternatives determine which alternative is chosen
- Only differences matter: $U_i - U_j, j \neq i$
- The choice is invariant to translation and scale
  - translation: $U_i + const - (U_j + const) = U_i - U_j$

Utility

- Utility differences between alternatives determine which alternative is chosen
- Only differences matter: $U_i - U_j, j \neq i$
- The choice is invariant to translation and scale
  - translation: $U_i + const - (U_j + const) = U_i - U_j$
  - scale: $\alpha U_i - \alpha U_j = \alpha(U_i - U_j) \lesseqgtr 0 \quad iff \quad U_i - U_j \lesseqgtr 0, \alpha > 0$

Deterministic utility maximisation

- Individual t *always* chooses alternative i if $U_{it} \geqslant U_{jt}, \forall j \in C_t$
- Example of deterministic choice, $C = \{\text{bicycle}, \text{t-bana}\}$

If $U_{\text{bicycle}} > U_{\text{t-bana}} \Rightarrow P(\text{bicycle}) = 1$
If $U_{\text{bicycle}} < U_{\text{t-bana}} \Rightarrow P(\text{bicycle}) = 0$

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*
- Randomness has several sources

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*
- Randomness has several sources
  - the researcher (us!) does not have complete information about all individuals' preferences

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*
- Randomness has several sources
  - the researcher (us!) does not have complete information about all individuals' preferences
  - heterogeneity of preferences – people do not like the same things, or change their minds from one time to another

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*
- Randomness has several sources
  - the researcher (us!) does not have complete information about all individuals' preferences
  - heterogeneity of preferences – people do not like the same things, or change their minds from one time to another
  - incomplete information about the alternatives

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*
- Randomness has several sources
  - the researcher (us!) does not have complete information about all individuals' preferences
  - heterogeneity of preferences – people do not like the same things, or change their minds from one time to another
  - incomplete information about the alternatives
  - randomness in the attributes of the alternatives (i.e. travel time, comfort)

Random utility

- We cannot determine the utility of everybody exactly – we allow for *random variation*
- Randomness has several sources
  - the researcher (us!) does not have complete information about all individuals' preferences
  - heterogeneity of preferences – people do not like the same things, or change their minds from one time to another
  - incomplete information about the alternatives
  - randomness in the attributes of the alternatives (i.e. travel time, comfort)
  - measurement errors

Random utility

- We divide the utility in a *deterministic* part V and a *random error* ε:

$$U_{it} = V_{it} + \varepsilon_{it}$$

where i indexes alternative and t individual.

Random utility

- We divide the utility in a *deterministic* part V and a *random error* ε:

$$U_{it} = V_{it} + \varepsilon_{it}$$

  where i indexes alternative and t individual.
- V_{it} is a function of observable variables:

Random utility

- We divide the utility in a *deterministic* part V and a *random error* ε:

$$U_{it} = V_{it} + \varepsilon_{it}$$

  where i indexes alternative and t individual.
- V_{it} is a function of observable variables:
  - attributes of the alternatives

Random utility

- We divide the utility in a *deterministic* part V and a *random error* ε:

$$U_{it} = V_{it} + \varepsilon_{it}$$

  where i indexes alternative and t individual.
- V_{it} is a function of observable variables:
  - attributes of the alternatives
  - preferences of the individual

Random utility

- We divide the utility in a *deterministic* part V and a *random error* ε:

$$U_{it} = V_{it} + \varepsilon_{it}$$

  where i indexes alternative and t individual.
- V_{it} is a function of observable variables:
  - attributes of the alternatives
  - preferences of the individual
- The assumed distribution and correlation structure of ε determines the type of model: probit, logit, nested logit etc.

Random utility

Choice probability:

$$\begin{aligned} P(\text{bicycle}|C) &= P(U_{\text{bicycle}} > U_{\text{t-bana}}) \\ &= P(V_{\text{bicycle}} + \varepsilon_{\text{bicycle}} > V_{\text{t-bana}} + \varepsilon_{\text{t-bana}}) \\ &= P(V_{\text{bicycle}} - V_{\text{t-bana}} > \varepsilon_{\text{t-bana}} - \varepsilon_{\text{bicycle}} = \varepsilon) \end{aligned}$$

Normal distribution

- Normal distribution: the *sum* of many independent, identically distributed (i.i.d.) random variables is approximately distributed Normal:

$$\varepsilon \sim N\left(\mu, \sigma^2\right)$$

  with mean μ and variance σ^2.

- This leads to the *probit model*, which does not have an analytic ("closed") form, and is therefore computationally cumbersome (multi-dimensional numeric integration),
  - both for estimation of parameters and forecasting

Gumbel distribution

- Gumbel distribution/"Extreme-Value type I": The *maximum* of many i.i.d. random variables (scaled appropriately) is Extreme-Value distributed
- If two random variables U_1 and U_2 are distributed Gumbel(V_i, η) with cumulative distribution function (c.d.f.)

$$F\left(U_i; V_i, \eta\right) = \exp\left(-\exp\left(-\eta\left(U_i - V_i\right)\right)\right)$$

  with location V_i and scale η, then the maximum of U_1 and U_2 is distributed Logistic:

$$P\left(U_1 \geqslant U_2\right) = \frac{\exp\left(\eta\left(V_1 - V_2\right)\right)}{1 + \exp\left(\eta\left(V_1 - V_2\right)\right)} = \frac{\exp\left(\eta V_1\right)}{\exp\left(\eta V_1\right) + \exp\left(\eta V_2\right)}$$

- This is the binomial logit model

Distribution of the random error

- The multinomial logit model:

$$P\left(U_i \geqslant \max_{j\in C}\left(U_j\right)\right) = F_{\max}\left(V_i;\eta\right) = \frac{\exp\left(\eta V_i\right)}{\sum_{j\in C}\exp\left(\eta V_j\right)}$$

- Normally, the scale parameter η is set to 1

Recapitulation

- We model the probability that a given individual chooses a given alternative among a set of available alternatives

Recapitulation

- We model the probability that a given individual chooses a given alternative among a set of available alternatives
- Under certain assumptions, this can be done using a multinomial logit model (MNL)

Recapitulation

- We model the probability that a given individual chooses a given alternative among a set of available alternatives
- Under certain assumptions, this can be done using a multinomial logit model (MNL)
- The probability of an alternative depends on socio-economic characteristics of the individual and the attributes of the available alternatives

Recapitulation

- We model the probability that a given individual chooses a given alternative among a set of available alternatives
- Under certain assumptions, this can be done using a multinomial logit model (MNL)
- The probability of an alternative depends on socio-economic characteristics of the individual and the attributes of the available alternatives
- With some more assumptions, the model can be used for predicting future or hypothetical choices

Back to example

- What is the probability that a student chooses t–bana if it costs 20 kr and bicycle is free?
- We assume that the deterministic utilities are:
  $V_{\text{t-bana}} = \beta_c \text{Cost}_{\text{t-bana}}$
  $V_{\text{bicycle}} = \beta_c \text{Cost}_{\text{bicycle}} + \beta_s \text{Student}$
  where $\beta_c = -0.1$ and $\beta_s = 0.05$
- $P(\text{t-bana}|\{\text{t-bana}, \text{bicycle}\}) = e^{(-0.1 \cdot 20)} / (e^{(-0.1 \cdot 20)} + e^{(0.05)}) \approx 0.11$
- How do you interpret the β_C and β_S values? Sign?
- Are any important variables omitted?

Back to example

Better formulation

$$V_{\text{tbana}} = ASC_{\text{t-bana}} + \beta_C \text{Cost}_{\text{t-bana}} + \beta_T \text{TravelTime}_{\text{t-bana}}$$
$$V_{\text{bicycle}} = \beta_C \text{Cost}_{\text{bicycle}} + \beta_T \text{TravelTime}_{\text{bicycle}} + \beta_S \text{Student}$$

How to determine the parameters

- What about unknown parameters (β-s and ASC)?
- Estimated from observed data
- Maximum-likelihood estimation
- Previous course, there is also a continuation course given next year

After this lecture you should know...

- ...how to interpret multinomial logit models

After this lecture you should know...

- ...how to interpret multinomial logit models
- ...which are the components of the model (decision-maker, alternatives, explanatory variables, etc.)

After this lecture you should know...

- ...how to interpret multinomial logit models
- ...which are the components of the model (decision-maker, alternatives, explanatory variables, etc.)
- ...which are the underlying assumptions

After this lecture you should know...

- ...how to interpret multinomial logit models
- ...which are the components of the model (decision-maker, alternatives, explanatory variables, etc.)
- ...which are the underlying assumptions
- During the course project you will be a user of a travel demand model (with given parameters)

Literature details

Koppelman and Bhat (2006), *A Self Instructive Course in Mode Choice Modeling: Multinomial and Nested Logit Models*

- Chapter 1: 1.1–1.2
- Chapter 2: all
- Chapter 3: all
- Chapter 4: 4.1