Mixed Logit

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

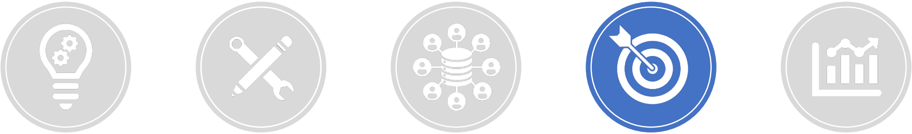
Interpretation & application

Mixed Logit

Introduction

- Key example of a model allowing for continuous random heterogeneity
- Very powerful model, widely used in academia and practice
- This session is more **advanced** and **theoretical**, due to the very nature of the model
- You do not necessarily need to understand all the mathematical detail
- Important to understand:
  - MMNL is a **complex** model
  - analyst decisions have major impacts on results

Overview



Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Overview

Setting the scene

- Two treatments, with a simple time/money trade-off
- Lower income people have higher cost sensitivity and lower time sensitivity
- With deterministic heterogeneity, can calculate probability for treatment choices for both groups of patients
  - these are probabilities *conditional* on observing income, and hence the time & cost sensitivity according to the model
- The probabilities in the population are distributed according to the size of the two groups of patients
- But we also know the location of each person!

Treatment	1	2
Wait (days)	28	14
Cost (£)	100	250

Low income (60% of sample)		
β_t	-0.04	
β_c	-0.01	
V	-2.12	-3.06
P	0.72	0.28

High income (40% of sample)		
β_t	-0.06	
β_c	-0.005	
V	-2.18	-2.09
P	0.48	0.52

Overview

Distribution of probabilities

- Let β_n give the (vector of) sensitivities for person n
- If we "know" β_n, we can calculate probabilities
  - e.g. with linear-in-attributes MNL, we have $P_{in}\left(\beta_n, x\right) = \frac{e^{\beta_n x_{in}}}{\sum_{j=1}^{J} e^{\beta_n x_{jn}}}$
- In models with deterministic heterogeneity, we observe the source of heterogeneity
  - can calculate person-specific probabilities and show distribution of P across sample
- Problem: with random heterogeneity, only have the distribution
  - not each person's location on that distribution

Overview

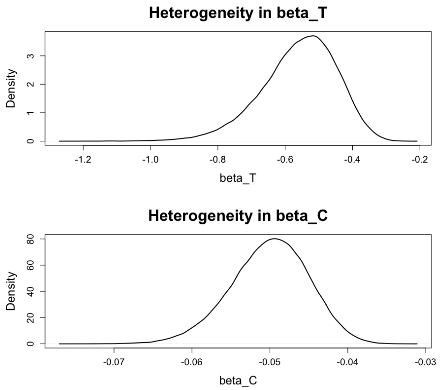
Mixed Logit in a nutshell

- Let $P_{in}(\beta_n, x)$ again be the probability of person n choosing alternative i
- Value of β_n is now not "observed", but only known up to a probability
- β_n follows a continuous (multivariate) distribution over individuals
  - $\beta_n \sim f(\beta_n \mid \Omega)$
- We know from MNL that if β_n varies across people, then so do the probabilities
- Mixed Logit
  - probabilities follow a continuous distribution across individuals

Overview

Illustration

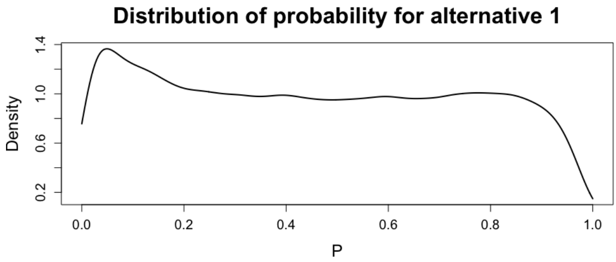
- Negative lognormal distribution for waiting time and cost coefficient
- Ensures purely negative response to time and cost
- What does this mean for the choice probabilities?



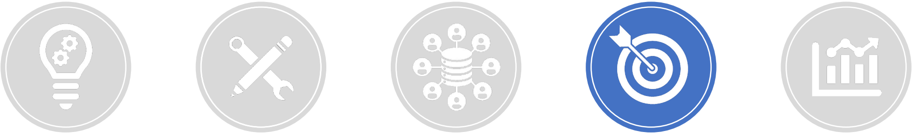
Overview

Resulting probability for alternative 1

Treatment	1	2
Wait (days)	28	14
Cost (£)	100	250



Model specification



Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Model specification

Key decisions

- An analyst needs to decide:
  - which model parameters follow random distributions
  - what distributions are used
  - whether univariate or multivariate distributions are used
- These decisions have major impacts on model results and interpretation

Model specification

Random parameters

- In theory, we should allow for random heterogeneity in all parameters
  - this would let the data speak
  - and avoid misattribution
- In practice, we need to consider empirical identification (data limitations) and computational costs
- Should think carefully which parameters are most likely to have variation in a population

Model specification

Distributional assumptions

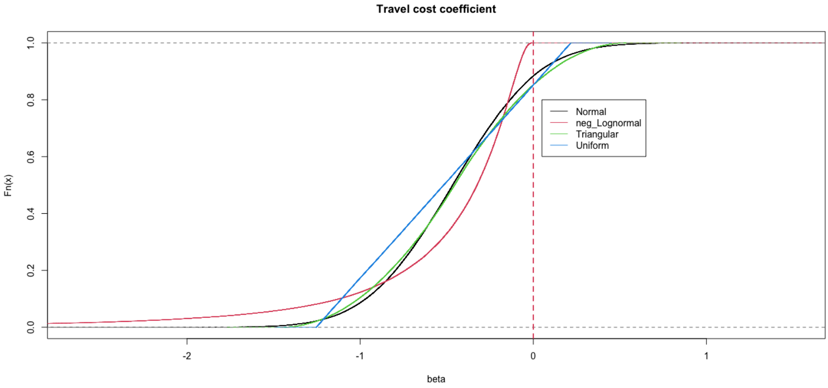
- Too many applications by default rely on Normal distributions
  - unbounded, and behaviourally not meaningful in many cases
  - problems in computing MRS/WTP
- Many other options exist
  - lognormal distribution (exponential of a Normal)
  - triangular distribution (sum of two independent uniforms with same support)
  - ...
- True shape can only be revealed by moving away from parametric distributions

Reference on inappropriate distributions: *Hess, S., Bierlaire, M. & Polak, J.W. (2005), Estimation of value of travel-time savings using Mixed Logit models, Transportation Research Part A, 39(2-3), pp. 221-236.*

Reference on flexible distributions: *Fosgerau, S. & Mabit, S. (2013), Easy and flexible mixture distributions, Economics Letters 120 (2), 206-210.*

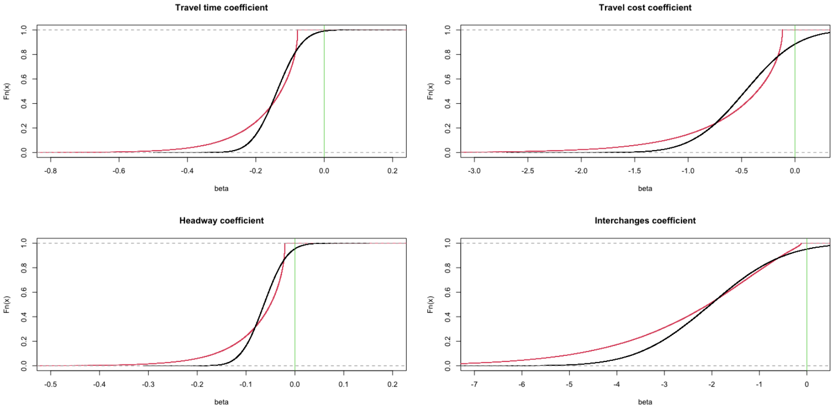
Model specification

Example with parametric distributions



Model specification

Non-parametric distribution confirms issues



Model specification

Multi-variate distributions

- The majority of applications rely on univariate distributions
- In practice, this may not be reasonable
  - people who care more about time may care less about cost, and vice versa
  - some people may be more sensitive overall than others
- Multi-variate distributions improve fit, reduce bias and allow model to allow for scale heterogeneity (but of course cannot disentangle it!)

Key reference: *Hess, S. & Train, K.E. (2017), Correlation and scale in mixed logit models, Journal of Choice Modelling, 23, pp. 1-8.*

Estimation

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Estimation

Recap on maximum likelihood estimation

- Log-likelihood: $LL\left(\Omega \mid x, Y\right) = \sum_{n=1}^{N} log\left(P_{jn_n^*}\left(\Omega, x\right)\right)$
- MLE: $\widehat{\Omega} = \underset{\Omega}{\arg\max}\, LL\left(\Omega \mid x, Y\right)$
- Optimisation requires $P_{jn_n^*}\left(\Omega, x\right), \ \forall n$
  - i.e. estimation requires us to calculate the probabilities of the choices in the data
- The issue now is how to calculate the probabilities for choices in a Mixed Logit model

Estimation

Econometrics

- Let $P_{in}(\beta_n, x)$ be probability of person n choosing alternative i
- We have a continuous distribution of β over individuals, $\beta_n \sim f(\beta_n \mid \Omega)$
- We do not know where on the distribution person n is
- Example for single choice task (in practice we will use repeated choices)
- Unconditional (on β_n) choice probability:

$$P_{in}(\Omega, x) = \int_{\beta_n} \left[P_{in}(\beta_n, x) \ f(\beta_n \mid \Omega) \right] \mathrm{d}\beta_n$$

- Probabilities given by an integral without a closed form solution
- Need to use approximation via numerical integration over distributions of β
  - often done using Monte Carlo simulation, giving us a *simulated log-likelihood*

Estimation

Numerical simulation and impact of simulation noise

- Monte Carlo simulation
  - approximate integral by averaging probabilities across large number of draws
- Computationally demanding
- But significant simulation error with low number of draws
  - Even more significant in more complex models
- Would translate into error in log-likelihood
  - has strong impact on parameter estimates!
  - bad idea to use high number of draws only for final model
- Use of quasi-random draws can help somewhat

Estimation

Parameters

- With MNL (and other fixed coefficients models), we estimate values of β
  - this includes constants, parameters multiplying attributes, interactions, etc
- The situation changes when we include random components in our model, such as random coefficients
- Example: cost coefficient (β_c) follows a random distribution
  - we do not obtain an estimate for β_c
  - we obtain estimates for the parameters of the distribution of β_c, e.g. mean and std dev
- We have $\beta_n \sim f\left(\beta_n \mid \Omega\right)$
  - Ω is a vector of parameters for the multivariate distribution of β in our data
  - we obtain estimates for Ω
  - for any elements of β that are not random, we obtain a point estimate

Illustrative example

Key concepts
& study plan

Experimental
design

Data collection
& processing

**Model specification
& estimation**

Interpretation
& application

Illustrative example

Application to Swiss VTT data

- Binary unlabelled public transport route choice, with alternatives described by travel time (TT), travel cost (TC), headway (HW), interchanges (CH)
- Simple linear in attributes utility function

$$V_{jnt} = \delta_j + \beta_{tt} TT_{jnt} + \beta_{tc} TC_{jnt} + \beta_{hw} HW_{jnt} + \beta_{ch} CH_{jnt}$$

- ASC normalisation: $\delta_2 = 0$
- For Mixed Logit, use negative lognormal distributions for the four β parameters, e.g.:

$$\beta_{tt} = -e^{\mu_{log(\beta_{tt})} + \sigma_{log(\beta_{tt})} \cdot \xi_{tt}}$$

- Ensures sign of β is purely negative
- $\xi_{tt} \sim N(0, 1)$, so sign of σ estimate irrelevant (it's not saying that the sd is negative!)

Illustrative example

Results

- Big improvement in model fit, and all standard deviations different from zero
- We will look at how to interpret these results later

```
Model name                       : MNL_swiss
Model description                : MNL model on Swiss route choice data
Estimation method                : bgw
Modelled outcomes                : 3492

LL(final)                        : -1665.62
Estimated parameters             : 5


Estimates (robust covariance matrix, 1-sided p-values):

    estimate std. error t-ratio p (1-sided)
asc1  -0.0159     0.0457    -0.3         0.4
b_tt  -0.0598     0.0067    -8.9       <2e-16 ***
b_tc  -0.1317     0.0236    -5.6        1e-08 ***
b_hw  -0.0374     0.0023   -16.2       <2e-16 ***
b_ch  -1.1521     0.0614   -18.8       <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Model name                       : MMNL_swiss_panel_all_negLN
Model description                : MMNL model with negative Lognormal distributions
Estimation method                : bgw
Modelled outcomes                : 3492

LL(final)                        : -1442.84
Estimated parameters             : 9


Estimates (robust covariance matrix, 1-sided p-values):

             estimate std. error t-ratio p (1-sided)
asc1           -0.039      0.071    -0.6       0.289
b_log_tt_mu    -1.985      0.110   -18.1      <2e-16 ***
b_log_tt_sig   -0.527      0.061    -8.7      <2e-16 ***
b_log_tc_mu    -0.961      0.179    -5.4       4e-08 ***
b_log_tc_sig   -0.940      0.069   -13.7      <2e-16 ***
b_log_hw_mu    -2.923      0.090   -32.4      <2e-16 ***
b_log_hw_sig    0.774      0.324     2.4       0.008 **
b_log_ch_mu     0.618      0.082     7.5       3e-14 ***
b_log_ch_sig   -0.887      0.128    -6.9       2e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
                                 LL par
MNL_swiss                  -1665.62   5
MMNL_swiss_panel_all_negLN -1442.84   9
Difference                   222.78   4

Likelihood ratio test-value:     445.56
Degrees of freedom:              4
Likelihood ratio test p-value:   3.96e-95
```