The multinomial logit model

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

The multinomial logit model

Outline

- The basics
- Model properties
- Numerical example

The basics

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

The basics

Multinomial logit model

- Abbreviated as the MNL model, also referred to as the conditional logit model
- Most basic random utility model
- Typically used as a starting point
- Many extensions exist

The basics

Econometric model

MULTINOMIAL LOGIT MODEL

- Endogenous variable y_{ni} (binary)
- Exogeneous variables $x_{ni1}, x_{ni2}, \ldots$
- Coefficients $\alpha_{i0}, \alpha_{i1}, \alpha_{i2}, \ldots$
- $y_{ni} = \begin{cases} 1, & \text{if } V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \text{ for all } j \\ 0, & \text{otherwise} \end{cases}$
- $V_{ni} = \alpha_{i0} + \alpha_{i1} x_{ni1} + \alpha_{i2} x_{ni2} + \ldots$
- $\varepsilon_{ni} \sim \text{Gumbel}(0, \mu)$

LINEAR REGRESSION MODEL

- Endogenous variable y_n (continuous)
- Exogeneous variables $x_{n1}, x_{n2}, \ldots$
- Coefficients $\alpha_0, \alpha_1, \alpha_2, \ldots$
- $y_n = V_n + \varepsilon_n$
- $V_n = \alpha_0 + \alpha_1 x_{n1} + \alpha_2 x_{n2} + \ldots$
- $\varepsilon_n \sim \text{Normal}(0, \sigma)$

Indices: Observation n – Alternative i,j

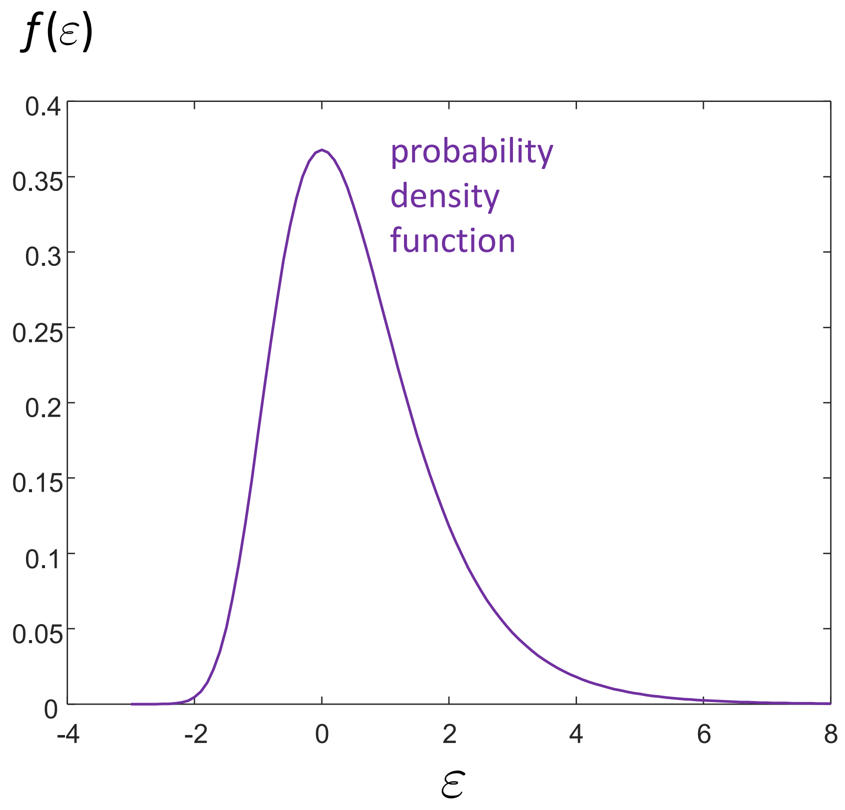
The basics

Core assumptions on random components

- ε_{ni}'s are Gumbel distributed
  - Gumbel is also known as Extreme Value type I (EV1) distribution
  - Variance is inversely related to scale parameter μ:

$$\text{Var}(\varepsilon_{ni}) = \frac{\pi^2}{6\mu^2}$$

- ε_{ni}'s are independent and identically distributed (IID)
  - No correlations across alternatives and observations
  - Extent of noise across alternatives and observations is the same



The basics

Utility components

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

V_{ni}

- Deterministic part
- Represents explained choice behaviour, accounts for variables included in the model
- Relative size depends on importance of included variables

ε_{ni}

- Random part
- Represents unexplained choice behaviour, accounts for all variables not included in the model
- Relative size is given by its standard deviation, $\sigma = \frac{\pi}{\mu\sqrt{6}}$

The basics

Large scale parameter μ

- Choice behaviour is more deterministic (i.e., attribute changes matter more)
- Typically means that choices are forecast with higher reliability
- But very large scale can also indicate issues with data or overfitting

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

The basics

Small scale parameter μ

- Choice behaviour is more random (high error variance)
- Typically means that choices are forecast less well
- Can indicate problems with data and/or model specification

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

The basics

Multinomial logit formula

- The probability that decision-maker *n* chooses alternative *i* among alternatives $1,\ldots,J$ equals

$$
\begin{aligned}
P_{ni} &= \Pr\left(y_{ni}=1\right) \\
&= \Pr\left(U_{ni} > U_{nj}, \text{ for all } j \neq i\right) \\
&= \Pr\left(V_{ni}+\varepsilon_{ni} > V_{nj}+\varepsilon_{nj}, \text{ for all } j \neq i\right) \\
&= \Pr\left(\varepsilon_{nj}-\varepsilon_{ni} < V_{ni}-V_{nj}, \text{ for all } j \neq i\right) \\
&= \frac{e^{\mu V_{ni}}}{\sum_{j=1}^{J} e^{\mu V_{nj}}} = \frac{\exp(\mu V_{ni})}{\sum_{j=1}^{J} \exp(\mu V_{nj})} = \frac{\exp(\mu V_{ni})}{\exp(\mu V_{n1})+\ldots+\exp(\mu V_{nJ})}
\end{aligned}
$$

where $e \approx 2.71828$ is Euler's number

The basics

Scale parameter μ is not identifiable

- Scale parameter is data set specific but cannot be separately estimated
- Common normalisation:

$$\mu = 1 \quad \text{such that} \quad \varepsilon \sim \text{Gumbel}(0,1)$$

The basics

Normalising scale does not mean that error variance is fixed

- Scale parameter is absorbed into coefficients of V
- In the multinomial logit formula, we can replace $\mu V_{ni} = \mu\left(\alpha_{i0} + \alpha_{i1}x_{ni1} + \alpha_{i2}x_{ni2} + \ldots\right)$ with

$$V_{ni} = \beta_{i0} + \beta_{i1}x_{ni1} + \beta_{i2}x_{ni2} + \ldots, \qquad \beta_{ik} = \mu\alpha_{ik}$$

- Estimated coefficients β_{ik} jointly capture preferences (α_{ik}) and scale (μ)
- Scale should be considered when combining data sets (see Week 6)

Since scale parameter μ is data set specific, estimated coefficients β_{ik} are not comparable across studies

Model properties

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Model properties

Only differences in utility matter

- Choice probabilities are not affected when the same number is added to all utilities

- The probability that decision-maker *n* chooses alternative *i* among alternatives $1,\ldots,J$ equals

$$
\begin{aligned}
P_{ni} &= \Pr\left(V_{ni}+\varepsilon_{ni}+a > V_{nj}+\varepsilon_{nj}+a, \text{ for all } j \neq i\right) \\
&= \Pr\left(V_{ni}+\varepsilon_{ni} > V_{nj}+\varepsilon_{nj}, \text{ for all } j \neq i\right) \\
&= \Pr\left(\varepsilon_{nj}-\varepsilon_{ni} < V_{ni}-V_{nj}, \text{ for all } j \neq i\right) \\
&= \frac{\exp(V_{ni})}{\sum_{j=1}^{J}\exp(V_{nj})} = \frac{1}{\sum_{j=1}^{J}\exp(\underbrace{V_{nj}-V_{ni}}_{\text{utility difference}})}
\end{aligned}
$$

Model properties

Only differences in utility matter – Example

- Assume the following utilities:
  - $V_{\text{Drug A}} = 2$
  - $V_{\text{Drug B}} = 3$

  $\Delta = 1$

- Probability that Drug A is chosen:

$$P_{\text{Drug A}} = \frac{\exp(2)}{\exp(2)+\exp(3)} = \frac{\exp(0)}{\exp(0)+\exp(3-2)} = \frac{1}{1+\exp(1)} = \boxed{0.27}$$

- Assume the following utilities:
  - $V_{\text{Drug A}} = 5$
  - $V_{\text{Drug B}} = 6$

  $\Delta = 1$

- Probability that Drug A is chosen:

$$P_{\text{Drug A}} = \frac{\exp(5)}{\exp(5)+\exp(6)} = \frac{\exp(0)}{\exp(0)+\exp(6-5)} = \frac{1}{1+\exp(1)} = \boxed{0.27}$$

Model properties

Only differences in utility matter – Example

- Assume the following utilities:
  - $V_{\text{Drug A}} = 0$
  - $V_{\text{Drug B}} = 1$

$\Delta = 1$

- Probability that Drug A is chosen:

$$P_{\text{Drug A}} = \frac{\exp(0)}{\exp(0)+\exp(1)} = \boxed{0.27}$$

- Assume the following utilities:
  - $V_{\text{Drug A}} = -4$
  - $V_{\text{Drug B}} = -3$

$\Delta = 1$

- Probability that Drug A is chosen:

$$P_{\text{Drug A}} = \frac{\exp(-4)}{\exp(-4)+\exp(-3)} = \frac{\exp(0)}{\exp(0)+\exp(-3--4)} = \frac{1}{1+\exp(1)} = \boxed{0.27}$$

Model properties

Independence of Irrelevant Alternatives (IIA)

- The relative probability of someone choosing between two options is independent of other alternatives in the choice set

- Assume the following utilities:
  - $V_{\text{Drug A}} = 2$
  - $V_{\text{Drug B}} = 3$

- Relative probabilities of Drugs A and B:

$$\frac{P_{\text{Drug A}}}{P_{\text{Drug B}}} = \frac{\frac{\exp(2)}{\exp(2)+\exp(3)}}{\frac{\exp(3)}{\exp(2)+\exp(3)}} = 0.37$$

- Assume the following utilities:
  - $V_{\text{Drug A}} = 2$
  - $V_{\text{Drug B}} = 3$
  - $V_{\text{Drug C}} = 5$

- Relative probabilities of Drugs A and B:

$$\frac{P_{\text{Drug A}}}{P_{\text{Drug B}}} = \frac{\frac{\exp(2)}{\exp(2)+\exp(3)+\exp(5)}}{\frac{\exp(3)}{\exp(2)+\exp(3)+\exp(5)}} = 0.37$$

Model properties

Is IIA realistic?

- Are relative choice probabilities always independent of the alternatives?

- IIA is a strong assumption and may not hold in practice
- Extensions of the MNL model exist to relax the IIA assumption

Numerical example

Numerical example

Assumptions

- Assume the following utilities:
  - $V_{\text{Drug A}} = 2$
  - $V_{\text{Drug B}} = 3$

- What is the probability that a decision-maker chooses Drug A?

$$P_{\text{Drug A}} = \Pr\left(U_{\text{Drug A}} > U_{\text{Drug B}}\right)$$
$$= \Pr\left(2 + \varepsilon_{\text{Drug A}} > 3 + \varepsilon_{\text{Drug B}}\right)$$

- Assume that $\varepsilon_{\text{Drug A}}, \varepsilon_{\text{Drug B}} \sim \text{Gumbel}(0,1)$

Numerical example

Spreadsheet

- See `MNL probabilities.xlsx`

P23

	A	C	D	E	G	H	I	K	L
1								Analytical choice probabilities	
2								27%	73%
3									
4								Simulated choice probabilities	
5		Scale (mu)	1					27%	73%
6									
7			Drug A			Drug B		Drug A	Drug B
8	n	V	eps	U=V+eps	V	eps	U=V+eps	choice	choice
9	1	2	1.966	3.966	3	2.729	5.729	0	1
10	2	2	0.183	2.183	3	2.837	5.837	0	1
11	3	2	-0.429	1.571	3	0.272	3.272	0	1
12	4	2	1.034	3.034	3	-1.304	1.696	1	0
13	5	2	0.717	2.717	3	-1.290	1.710	1	0
14	6	2	0.395	2.395	3	-0.497	2.503	0	1
15	7	2	1.076	3.076	3	0.685	3.685	0	1
16	8	2	1.206	3.206	3	0.979	3.979	0	1
17	9	2	-0.003	1.997	3	0.767	3.767	0	1
18	10	2	-1.157	0.843	3	-0.255	2.745	0	1

Sheet1

Numerical example

Deterministic part of utility

- $V_{\text{Drug A}} = 2$
- $V_{\text{Drug B}} = 3$

Analytical choice probabilities	
27%	73%

Simulated choice probabilities	
27%	73%

Scale (mu)	1

	Drug A			Drug B			Drug A	Drug B
n	V	eps	U=V+eps	V	eps	U=V+eps	choice	choice
1	2	1.966	3.966	3	2.729	5.729	0	1
2	2	0.183	2.183	3	2.837	5.837	0	1
3	2	-0.429	1.571	3	0.272	3.272	0	1
4	2	1.034	3.034	3	-1.304	1.696	1	0
5	2	0.717	2.717	3	-1.290	1.710	1	0
6	2	0.395	2.395	3	-0.497	2.503	0	1
7	2	1.076	3.076	3	0.685	3.685	0	1
8	2	1.206	3.206	3	0.979	3.979	0	1
9	2	-0.003	1.997	3	0.767	3.767	0	1
10	2	-1.157	0.843	3	-0.255	2.745	0	1

Numerical example

Calculation of choice probabilities via multinomial logit formula

- $P_{\text{Drug A}} = e^{\mu V_{\text{Drug A}}} / (e^{\mu V_{\text{Drug A}}} + e^{\mu V_{\text{Drug B}}})$
- $P_{\text{Drug B}} = e^{\mu V_{\text{Drug B}}} / (e^{\mu V_{\text{Drug A}}} + e^{\mu V_{\text{Drug B}}})$

K2 =EXP(D5*C9)/(EXP(D5*C9)+EXP(D5*G9))

									Analytical choice probabilities		
									27%	73%	
									Simulated choice probabilities		
		Scale (mu)	1						27%	73%	
		Drug A				Drug B			Drug A	Drug B	
n		V	eps	U=V+eps		V	eps	U=V+eps	choice	choice	
1		2	1.966	3.966		3	2.729	5.729	0	1	
2		2	0.183	2.183		3	2.837	5.837	0	1	
3		2	-0.429	1.571		3	0.272	3.272	0	1	
4		2	1.034	3.034		3	-1.304	1.696	1	0	
5		2	0.717	2.717		3	-1.290	1.710	1	0	
6		2	0.395	2.395		3	-0.497	2.503	0	1	
7		2	1.076	3.076		3	0.685	3.685	0	1	
8		2	1.206	3.206		3	0.979	3.979	0	1	
9		2	-0.003	1.997		3	0.767	3.767	0	1	
10		2	-1.157	0.843		3	-0.255	2.745	0	1	

Sheet1

Numerical example

Calculation of choice probabilities via simulation

- $\varepsilon_{\text{Drug A}}, \varepsilon_{\text{Drug A}} \sim \text{Gumbel}(0,1)$
- Simulated draws from standard Gumbel distribution

D9 =-LN(-LN(RAND()))/D5

	A	C	D	E	G	H	I	K	L
1								Analytical choice probabilities	
2								27%	73%
3									
4								Simulated choice probabilities	
5		Scale (mu)	1					27%	73%
6									
7		Drug A			Drug B			Drug A	Drug B
8	n	V	eps	U=V+eps	V	eps	U=V+eps	choice	choice
9	1	2	1.966	3.966	3	2.729	5.729	0	1
10	2	2	0.183	2.183	3	2.837	5.837	0	1
11	3	2	-0.429	1.571	3	0.272	3.272	0	1
12	4	2	1.034	3.034	3	-1.304	1.696	1	0
13	5	2	0.717	2.717	3	-1.290	1.710	1	0
14	6	2	0.395	2.395	3	-0.497	2.503	0	1
15	7	2	1.076	3.076	3	0.685	3.685	0	1
16	8	2	1.206	3.206	3	0.979	3.979	0	1
17	9	2	-0.003	1.997	3	0.767	3.767	0	1
18	10	2	-1.157	0.843	3	-0.255	2.745	0	1

Sheet1

$$\varepsilon_{\text{Drug A}} = \frac{-\ln(-\ln(u))}{\mu}$$

$$\varepsilon_{\text{Drug B}} = \frac{-\ln(-\ln(u))}{\mu}$$

where u is a random number between 0 and 1

Numerical example

Simulated random utilities

- $U_{\text{Drug A}} = V_{\text{Drug A}} + \varepsilon_{\text{Drug A}}$
- $U_{\text{Drug B}} = V_{\text{Drug B}} + \varepsilon_{\text{Drug B}}$

E9 =C9+D9

Analytical choice probabilities	
27%	73%

Simulated choice probabilities	
27%	73%

Scale (mu): 1

	Drug A			Drug B			Drug A	Drug B
n	V	eps	U=V+eps	V	eps	U=V+eps	choice	choice
1	2	1.966	3.966	3	2.729	5.729	0	1
2	2	0.183	2.183	3	2.837	5.837	0	1
3	2	-0.429	1.571	3	0.272	3.272	0	1
4	2	1.034	3.034	3	-1.304	1.696	1	0
5	2	0.717	2.717	3	-1.290	1.710	1	0
6	2	0.395	2.395	3	-0.497	2.503	0	1
7	2	1.076	3.076	3	0.685	3.685	0	1
8	2	1.206	3.206	3	0.979	3.979	0	1
9	2	-0.003	1.997	3	0.767	3.767	0	1
10	2	-1.157	0.843	3	-0.255	2.745	0	1

Sheet1

Numerical example

Simulated choices

- $y_{\text{Drug A}} = 1$ (Drug A is chosen) if $U_{\text{Drug A}} > U_{\text{Drug B}}$ $\quad (y_{\text{Drug A}} = 0$ otherwise)
- $y_{\text{Drug B}} = 1$ (Drug B is chosen) if $U_{\text{Drug B}} > U_{\text{Drug A}}$ $\quad (y_{\text{Drug B}} = 0$ otherwise)

K9 =IF(E9>I9,1,0)

	Drug A	Drug B
Analytical choice probabilities	27%	73%
Simulated choice probabilities	27%	73%

Scale (mu): 1

	Drug A			Drug B			Drug A	Drug B
n	V	eps	U=V+eps	V	eps	U=V+eps	choice	choice
1	2	1.966	3.966	3	2.729	5.729	0	1
2	2	0.183	2.183	3	2.837	5.837	0	1
3	2	-0.429	1.571	3	0.272	3.272	0	1
4	2	1.034	3.034	3	-1.304	1.696	1	0
5	2	0.717	2.717	3	-1.290	1.710	1	0
6	2	0.395	2.395	3	-0.497	2.503	0	1
7	2	1.076	3.076	3	0.685	3.685	0	1
8	2	1.206	3.206	3	0.979	3.979	0	1
9	2	-0.003	1.997	3	0.767	3.767	0	1
10	2	-1.157	0.843	3	-0.255	2.745	0	1

Sheet1

Numerical example

Simulated choice probabilities

- $P_{\text{Drug A}} = \text{average}(y_{\text{Drug A}})$
- $P_{\text{Drug B}} = \text{average}(y_{\text{Drug B}})$

K9 =AVERAGE(K9:K1008)

	A	B	C	D	E	F	G	H	I	J	K	L
1											Analytical choice probabilities	
2											27%	73%
3												
4											Simulated choice probabilities	
5			Scale (mu)	1							27%	73%
6												
7			Drug A				Drug B				Drug A	Drug B
8	n		V	eps	U=V+eps		V	eps	U=V+eps		choice	choice
9	1		2	1.966	3.966		3	2.729	5.729		0	1
10	2		2	0.183	2.183		3	2.837	5.837		0	1
11	3		2	-0.429	1.571		3	0.272	3.272		0	1
12	4		2	1.034	3.034		3	-1.304	1.696		1	0
13	5		2	0.717	2.717		3	-1.290	1.710		1	0
14	6		2	0.395	2.395		3	-0.497	2.503		0	1
15	7		2	1.076	3.076		3	0.685	3.685		0	1
16	8		2	1.206	3.206		3	0.979	3.979		0	1
17	9		2	-0.003	1.997		3	0.767	3.767		0	1
18	10		2	-1.157	0.843		3	-0.255	2.745		0	1

Sheet1