The logit model

Derivation and specification

Michel Bierlaire

Mathematical Modeling of Behavior

Outline

The logit model

Probability for individual n to choose alternative i within the set $\mathcal{C}_n$:

$$P(i|\mathcal{C}_n) = \frac{e^{\mu_n V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{\mu_n V_{jn}}}.$$

where

$$V_{in} = \sum_{k=1}^{K} \beta_k z_{ink},$$

where

- $\mathcal{C}_n$ is the set of alternatives for individual n,
- z_{in} are the attributes of alternative i for individual n, and
- μ_n and β_k, $k = 1, \ldots, K$ are parameters to be estimated from data.

Example: transportation mode choice in Switzerland

Three alternatives

- Car,
- public transportation (pt),
- Slow modes (sm).

Six attributes

- Travel cost (car and pt).
- Travel time (car and pt).
- Waiting time (pt).
- Distance (sm).

Example: n is Priya

Who is Priya?

- German speaking.
- Age: 44.
- Gender: female.
- Subscription: no GA.
- Socio-prof. category: manager.
- Trip purpose: work.
- Car availability: yes.

Attributes

- Car cost: $z_{\text{car,Priya},1} = 0.34$ CHF.
- Time by car: $z_{\text{car,Priya},2} = 3.0$ min.
- pt cost: $z_{\text{pt,Priya},1} = 3.0$ CHF.
- Time by pt: $z_{\text{pt,Priya},2} = 17.0$ min.
- Waiting time: $z_{\text{pt,Priya},3} = 1.0$ min.
- Distance: $z_{\text{sm,Priya},4} = 2.5$ km.

Example: n is Priya

Utility: function of the attributes

$$V_{\text{pt,Priya}} = -(\beta_{0,\text{Priya}} + \text{cost}_{\text{pt,Priya}} + \beta_{t,\text{Priya}}\ \text{time}_{\text{pt,Priya}} + \beta_{w,\text{Priya}}\ \text{waiting time}_{\text{pt,Priya}}).$$

Opposite of the generalized cost.

Example of parameters

$$\begin{aligned} V_{\text{pt,Priya}} &= -14.64 - \text{cost}_{\text{pt,Priya}} - 0.303\ \text{time}_{\text{pt,Priya}} - 1.97\ \text{waiting time}_{\text{pt,Priya}} \\ &= -24.8. \end{aligned}$$

Example: n is Priya

Utility: function of the attributes

$$
\begin{aligned}
V_{\text{pt,Priya}} &= -24.8, \\
V_{\text{car,Priya}} &= 10.4 - \text{cost}_{\text{car,Priya}} - 1.18\ \text{time}_{\text{car,Priya}} \\
&= 6.52, \\
V_{\text{sm,Priya}} &= -6.16\ \text{distance}_{\text{sm,Priya}} \\
&= -15.4.
\end{aligned}
$$

Example: *n* is Priya

Logit model: probability for Priya to choose the car

$$\mathcal{C}_{\text{Priya}} = \{\text{car}, \text{pt}, \text{sm}\}, \ \mu_{\text{Priya}} = 0.0347.$$

$$\begin{aligned}
\mu_{\text{Priya}} V_{\text{pt},\text{Priya}} &= 0.0347 \cdot (-24.8) = -0.859, \\
\mu_{\text{Priya}} V_{\text{car},\text{Priya}} &= 0.0347 \cdot 6.52 = 0.226, \\
\mu_{\text{Priya}} V_{\text{sm},\text{Priya}} &= 0.0347 \cdot (-15.4) = -0.534.
\end{aligned}$$

Example: *n* is Priya

Logit model: probability for Priya to choose the public transportation

$$\mathcal{C}_{\text{Priya}} = \{\text{car}, \text{pt}, \text{sm}\}.$$

$$\begin{aligned} P_{\text{Priya}}(\text{pt}|\mathcal{C}_{\text{Priya}}) &= \frac{e^{-0.859}}{e^{-0.859} + e^{0.226} + e^{-0.534}} \\ &= 0.187. \end{aligned}$$

Example: n is Priya

Logit model: probability for Priya to choose the car

$$\mathcal{C}_{\text{Priya}} = \{\text{car}, \text{pt}, \text{sm}\}.$$

$$\begin{aligned} P_{\text{Priya}}(\text{car}|\mathcal{C}_{\text{Priya}}) &= \frac{e^{0.226}}{e^{-0.859} + e^{0.226} + e^{-0.534}} \\ &= 0.554. \end{aligned}$$

Example: n is Priya

Logit model: probability for Priya to choose a slow mode

$$\mathcal{C}_{\text{Priya}} = \{\text{car}, \text{pt}, \text{sm}\}.$$

$$\begin{aligned} P_{\text{Priya}}(\text{sm}|\mathcal{C}_{\text{Priya}}) &= \frac{e^{-0.534}}{e^{-0.859} + e^{0.226} + e^{-0.534}} \\ &= 0.259. \end{aligned}$$

Example: n is Priya

Logit model for Priya

$$
\begin{aligned}
P_{\text{Priya}}(\text{pt}) &= 18.7\%,\\
P_{\text{Priya}}(\text{car}) &= 55.4\%,\\
P_{\text{Priya}}(\text{sm}) &= 25.9\%.
\end{aligned}
$$

Example: n is Mateo

Who is Mateo?

- French speaking.
- Age: 35.
- Gender: male.
- Subscription: GA.
- Socio-prof. category: craftman.
- Trip purpose: not work.
- Car availability: no.

Attributes

- Car cost: irrelevant.
- Time by car: irrelevant.
- pt cost: $z_{\text{pt,Mateo},1} = 0.0$ CHF.
- Time by pt: $z_{\text{pt,Mateo},2} = 17.0$ min.
- Waiting time: $z_{\text{pt,Mateo},3} = 1.0$ min.
- Distance: $z_{\text{sm,Mateo},4} = 2.5$ km.

Example: *n* is Mateo

Utility: function of the attributes

$$V_{\text{pt,Mateo}} = -(\beta_{0,\text{Mateo}} + \text{cost}_{\text{pt,Mateo}} + \beta_{t,\text{Mateo}}\ \text{time}_{\text{pt,Mateo}} + \beta_{w,\text{Mateo}}\ \text{waiting time}_{\text{pt,Mateo}}).$$

Example of parameters

$$\begin{aligned} V_{\text{pt,Mateo}} &= -11.5 - \text{cost}_{\text{pt,Mateo}} - 0.166\ \text{time}_{\text{pt,Mateo}} - 0.19\ \text{waiting time}_{\text{pt,Mateo}} \\ &= -17.5. \end{aligned}$$

Example: n is Mateo

Utility: function of the attributes

$$
\begin{aligned}
V_{\text{pt,Mateo}} &= -17.5, \\
V_{\text{sm,Mateo}} &= -10.6 \text{ distance}_{\text{sm,Mateo}} \\
&= -26.5.
\end{aligned}
$$

Example: *n* is Mateo

Logit model: scaled utility for Mateo

$$\mathcal{C}_{\text{Mateo}} = \{\text{pt}, \text{sm}\}, \;\; \mu_n = 0.0521.$$

$$\begin{aligned} \mu_{\text{Mateo}} V_{\text{pt,Mateo}} &= 0.0521 \cdot -17.5 = -0.912, \\ \mu_{\text{Mateo}} V_{\text{sm,Mateo}} &= 0.0521 \cdot -26.5 = -1.38. \end{aligned}$$

Example: n is Mateo

Logit model: choice probabilities for Mateo

$$\mathcal{C}_{\text{Mateo}} = \{\text{pt}, \text{sm}\}.$$

$$\begin{aligned}
P_{\text{Mateo}}(\text{pt}|\mathcal{C}_{\text{Mateo}}) &= \frac{e^{-0.912}}{e^{-0.912} + e^{-1.38}} \\
&= 0.616, \\
P_{\text{Mateo}}(\text{slow modes}|\mathcal{C}_{\text{Mateo}}) &= \frac{e^{-1.38}}{e^{-0.912} + e^{-1.38}} \\
&= 0.384.
\end{aligned}$$

Example: n is Mateo

Logit model for Mateo

$$
\begin{aligned}
P_{\text{Mateo}}(\text{pt}) &= 61.6\%, \\
P_{\text{Mateo}}(\text{car}) &= 0\%, \\
P_{\text{Mateo}}(\text{sm}) &= 38.4\%.
\end{aligned}
$$

How does it work?

- Where does the logit model come from?
  - With two alternatives.
  - With multiple alternatives.
- How do we specify the utility functions?
  - What variables can be involved?
  - How do we come up with a fonctional form?
  - How do we derive a different model for different individuals?
- How do we estimate the parameters?

Outline

- Introduction
- Binary logit model
- Logit with multiple alternatives
- Normalization
- Alternative Specific Constants
- Coefficients of attributes
- Taste heterogeneity
- Heteroscedasticity
- Nonlinear specifications

The binary logit model

Two alternatives: $\mathcal{C}_n = \{i, j\}$

$$
\begin{aligned}
U_{in} &= V_{in} + \varepsilon'_{in}, \\
U_{jn} &= V_{jn} + \varepsilon'_{jn}.
\end{aligned}
$$

Main issue

- Utility is latent, not observed.
- Only the choice is observed.
- More complicated than linear regression.
- How do we know the “zero” of utility?
- How do we know the units of utility?

Binary choice model

Choice model

$$P_n(i|\{i,j\}) = \Pr(U_{in} \geqslant U_{jn}).$$

Invariant to shifts

$$P_n(i|\{i,j\}) = \Pr(U_{in} + \eta \geqslant U_{jn} + \eta), \ \forall \eta \in \mathbb{R}.$$

Invariant to scale

$$P_n(i|\{i,j\}) = \Pr(\mu U_{in} \geqslant \mu U_{jn}), \ \forall \mu \in \mathbb{R}, \mu > 0.$$

Binary choice model

Choice model

$$\begin{array}{rcl} P_n(i|\{i,j\}) & = & \Pr(U_{in} \geqslant U_{jn}) \\ & = & \Pr(V_{in} + \varepsilon'_{in} \geqslant V_{jn} + \varepsilon'_{jn}) \\ & = & \Pr(V_{in} - V_{jn} \geqslant \varepsilon'_{jn} - \varepsilon'_{in}) \\ & = & \Pr\left(\varepsilon'_n \leqslant V_{in} - V_{jn}\right), \end{array}$$

where $\varepsilon'_n = \varepsilon'_{jn} - \varepsilon'_{in}$.

Note

- For binary choice, it would be sufficient to make assumptions about $\varepsilon'_n = \varepsilon'_{jn} - \varepsilon'_{in}$.
- But we want to generalize later on.

Error term

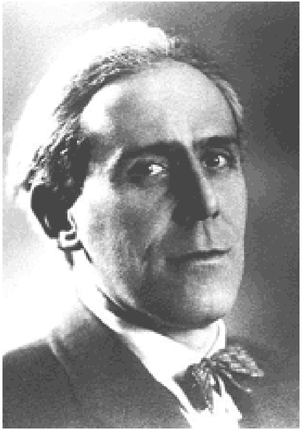
Assumptions about the random variables ε'_{in} and ε'_{jn}

ε'_{in} and ε'_{jn} are the maximum of many r.v. capturing unobserved attributes (e.g. mood, experience), measurement and specification errors.

Gumbel theorem

The maximum of many i.i.d. random variables approximately follows an Extreme Value distribution: $\text{EV}(\eta, \mu)$, with $\mu > 0$.

Extreme value distribution



Emil Julius Gumbel (1891–1966)

- father of extreme value theory,
- politically involved left-wing pacifist in Germany,
- strongly against right wing's campaign of organized assassination (1919),
- first German professor to be expelled from university under the pressure of the Nazis,
- 1932: he left Heidelberg to Paris, where he met Borel and Fréchet,
- 1940: he had to escape to New-York, where he continued his fight against Nazism by helping the US secret service.

The Extreme Value distribution EV(η, μ)

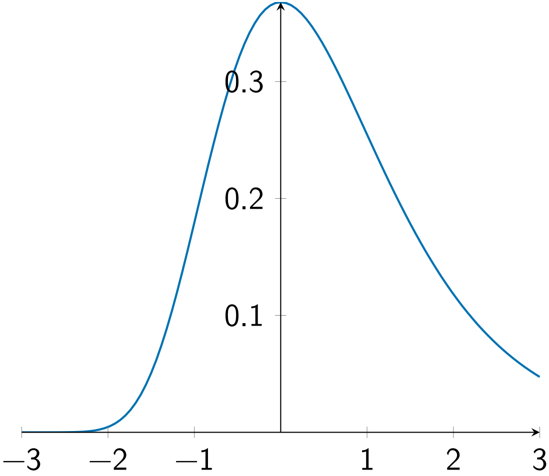
Probability density function (pdf)

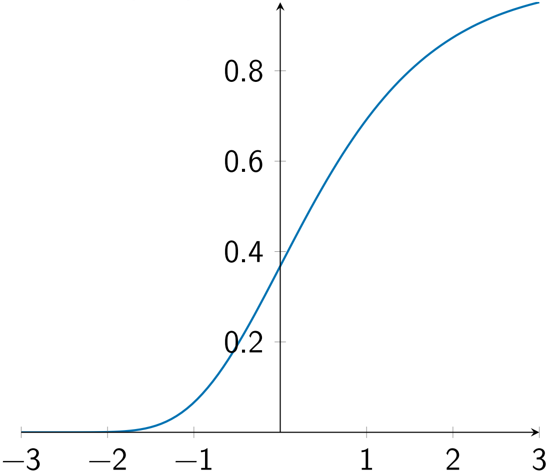
$$f(t) = \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}}.$$

Cumulative distribution function (CDF)

$$\begin{aligned} P(\varepsilon \leqslant c) = F(c) &= \int_{-\infty}^{c} f(t)dt \\ &= e^{-e^{-\mu(c-\eta)}}. \end{aligned}$$

The Extreme Value distribution


pdf EV(0,1)


CDF EV(0,1)

The Extreme Value distribution

Properties

$$\varepsilon \sim \mathrm{EV}(\eta, \mu).$$

- Mode: η.
- Mean: $E[\varepsilon] = \eta + \frac{\gamma}{\mu}$ where γ is Euler's constant.
- Variance: $\mathrm{Var}[\varepsilon] = \frac{\pi^2}{6\mu^2}$.

Euler's constant

$$\gamma = -\int_0^\infty e^{-x} \ln x \, dx \approx 0.5772.$$

The Extreme Value distribution

Properties

- Let $\varepsilon \sim EV(\eta, \mu)$, $\alpha > 0$ and $\beta \in \mathbb{R}$. Then

$$\alpha\varepsilon + \beta \sim EV(\alpha\eta + \beta, \mu/\alpha).$$

- In particular, if $\varepsilon \sim$ EV(0, 1), then,

$$\alpha\varepsilon + \beta \sim EV(\beta, 1/\alpha).$$

- Therefore, if $\varepsilon \sim$ EV(0, 1), then, using $\alpha = 1/\mu$ and $\beta = \eta$,

$$\varepsilon' = \eta + \frac{1}{\mu}\varepsilon \sim \mathrm{EV}(\eta, \mu).$$

The Extreme Value distribution

Properties

Let $\varepsilon_1 \sim EV(\eta_1, \mu)$ and $\varepsilon_2 \sim EV(\eta_2, \mu)$

$$\varepsilon = \varepsilon_1 - \varepsilon_2 \sim \text{Logistic}(\eta_1 - \eta_2, \mu),$$

that is

$$F_\varepsilon(x) = \frac{1}{1 + \exp(-\mu(x - (\eta_1 - \eta_2)))}.$$

Note: the two EV distributions must have the same scale μ.

The Extreme Value distribution

Properties

- Let $\varepsilon_1 \sim EV(\eta_1, \mu)$ and $\varepsilon_2 \sim EV(\eta_2, \mu)$ independent. Then,

$$\varepsilon = \max(\varepsilon_1, \varepsilon_2) \sim EV\left(\frac{1}{\mu}\ln(e^{\mu\eta_1} + e^{\mu\eta_2}), \mu\right).$$

- Let $\varepsilon_i \sim EV(\eta_i, \mu)$, $i = 1, \ldots, J$ independent. Then,

$$\varepsilon = \max(\varepsilon_1, \ldots, \varepsilon_J) \sim EV\left(\frac{1}{\mu}\ln\sum_{i=1}^{J} e^{\mu\eta_i}, \mu\right).$$

- The sum of two EV r.v. is **not** an EV r.v.

Modeling assumptions

Distributions

- ε'_{in} and ε'_{jn} are i.i.d. EV(η_n, μ_n).
- $\eta_n, \mu_n \in \mathbb{R}$, $\mu_n > 0$.
- i.i.d. $=$ independent and identically distributed.
- i.i.d. across i.
- Independent across n.

Modeling assumptions

Change of variables: isolate the parameters

$$\begin{aligned}\varepsilon'_{in} &= \eta_n + \frac{1}{\mu_n}\varepsilon_{in},\\ \varepsilon'_{jn} &= \eta_n + \frac{1}{\mu_n}\varepsilon_{jn},\end{aligned}$$

where ε_{in}, $\varepsilon_{jn} \sim \text{EV}(0, 1)$.

Binary logit model

Specification

If the model is specified as

$$\begin{array}{rcl} U_{in} & = & V_{in} + \eta_n + \frac{1}{\mu_n} \varepsilon_{in}, \\ U_{jn} & = & V_{jn} + \eta_n + \frac{1}{\mu_n} \varepsilon_{jn}, \end{array}$$

we can assume w.l.o.g. that ε_{in}, ε_{jn} are i.i.d. $\sim$ EV(0, 1), across both i and n.

Binary logit model

Choice model

$$\begin{aligned} P_n(i|\{i,j\}) &= \Pr\left(U_{in} \geqslant U_{jn}\right) \\ &= \Pr\left(V_{in} + \eta_n + \frac{1}{\mu_n}\varepsilon_{in} \geqslant V_{jn} + \eta_n + \frac{1}{\mu_n}\varepsilon_{jn}\right) \\ &= \Pr\left(\frac{1}{\mu_n}(\varepsilon_{jn} - \varepsilon_{in}) \leqslant V_{in} + \cancel{\eta_n} - V_{jn} - \cancel{\eta_n}\right), \\ &= \Pr(\varepsilon_{jn} - \varepsilon_{in} \leqslant \mu_n V_{in} - \mu_n V_{jn}). \end{aligned}$$

Property of EV

$$\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in} \sim \text{Logistic}(0, 1).$$

The Logistic distribution: Logistic(η,μ)

Probability density function (pdf)

$$f(t) = \frac{\mu e^{-\mu(t-\eta)}}{(1+e^{-\mu(t-\eta)})^2}.$$

Cumulative distribution function (CDF)

$$P(c \geqslant \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt = \frac{1}{1+e^{-\mu(c-\eta)}}.$$

with $\mu > 0$.

Binary logit model

Choice model

$$P_n(i|\{i,j\}) = \Pr\left(\varepsilon_n \leqslant \mu_n V_{in} - \mu V_{jn}\right) = F_\varepsilon(\mu_n V_{in} - \mu_n V_{jn}).$$

The binary logit model

$$P_n(i|\{i,j\}) = \frac{1}{1 + e^{-\mu_n(V_{in} - V_{jn})}} = \frac{e^{\mu_n V_{in}}}{e^{\mu_n V_{in}} + e^{\mu_n V_{jn}}}.$$

The binary logit model

Key element of the specification

$$\mu_n V_{in}.$$

Comments

- η_n does not play any role in the model.
- The units of V_{in} must be fixed. The model must be normalized.
- Before doing it, we extend the model to more than two alternatives.

Outline

Introduction

Binary logit model

Logit with multiple alternatives

Normalization

Alternative Specific Constants

Coefficients of attributes

Taste heterogeneity

Heteroscedasticity

Nonlinear specifications

Multiple alternatives

Choice set: $\mathcal{C}_n = \{1, \ldots, J_n\}$

$$\begin{aligned} U_{1n} &= V_{1n} + \varepsilon_{1n}, \\ &\vdots \\ U_{J_n n} &= V_{J_n n} + \varepsilon_{J_n n}. \end{aligned}$$

Choice set

Universal choice set: $\mathcal{C}$

- All potential alternatives for the population.
- Alternatives relevant to the analyst.

Mode choice

- driving alone,
- walking,
- biking,
- subway,
- bus rapid transit,
- light rail,
- ride hailing,
- on demand transit,
- bicycle/e-scooter sharing,
- car sharing,
- air taxi.

Choice set

Individual's choice set: $\mathcal{C}_n$

- No driver license.
- No auto available.
- Awareness of public transportation.
- Public transportation unreachable.
- Walking not an option for long distance.

Mode choice

- ~~driving alone~~,
- ~~walking~~,
- biking,
- ~~subway~~,
- ~~bus rapid transit~~,
- ~~light rail~~,
- ride hailing,
- on demand transit,
- bicycle/e-scooter sharing,
- car sharing,
- air taxi.

Choice set

Choice set generation is tricky

- How to model “awareness”?
- What does “long distance” exactly mean?
- What does “unreachable” exactly mean?

We assume here deterministic rules

- Car is available if *n* has a driver license and a car is available in the household.
- Walking is available if trip length is shorter than 4km.

Availability conditions

$$\delta_{in} = \begin{cases} 1 & \text{if } i \in \mathcal{C}_n, \\ 0 & \text{otherwise.} \end{cases} \quad \text{or} \quad \ln \delta_{in} = \begin{cases} 0 & \text{if } i \in \mathcal{C}_n, \\ -\infty & \text{otherwise.} \end{cases}$$

Choice model

$$P_n(i|\mathcal{C}_n) = P_n(i|\delta_n, \mathcal{C}) = \Pr(U_{in} + \ln \delta_{in} \geqslant U_{jn} + \ln \delta_{jn}).$$

Error terms

Logit: same assumptions as for binary logit

ε_{in} are

- independent and
- identically distributed,
- extreme value EV(η_n, μ_n).

Comments

- i.i.d. across i,
- independent across n.

The logit model: derivation

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geqslant \max_{j \in \mathcal{C}_n \setminus \{i\}} U_{jn}) = \Pr(V_{in} + \varepsilon_{in} \geqslant \max_{j \in \mathcal{C}_n \setminus \{i\}} V_{jn} + \varepsilon_{jn}).$$

Best alternative different from i

$$U_{-in} = \max_{j \in \mathcal{C}_n \setminus \{i\}} U_{jn} = \max_{j \in \mathcal{C}_n \setminus \{i\}} (V_{jn} + \varepsilon_{jn}).$$

Binary choice model

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geqslant U_{-in}).$$

The logit model

Property of Extreme Value distribution

$$U_{-in} = V_{-in} + \varepsilon_{-in}$$

where

$$V_{-in} = \frac{1}{\mu_n} \ln \sum_{j \in \mathcal{C}_n \setminus \{i\}} e^{\mu_n V_{jn}},$$

and

$$\varepsilon_{-in} \sim \text{EV}(0, \mu_n).$$

The logit model

Binary logit

$$P(i|\mathcal{C}_n) = \frac{e^{\mu_n V_{in}}}{e^{\mu_n V_{in}} + e^{\mu_n V_{-in}}}$$

Therefore...

$$\begin{aligned}
V_{-in} &= \frac{1}{\mu_n} \ln \sum_{j \in \mathcal{C}_n \setminus \{i\}} e^{\mu_n V_{jn}}, \\
e^{\mu_n V_{-in}} &= e^{\mu_n \frac{1}{\mu_n} \ln \sum_{j \in \mathcal{C}_n \setminus \{i\}} e^{\mu_n V_{jn}}} = \sum_{j \in \mathcal{C}_n \setminus \{i\}} e^{\mu_n V_{jn}}, \\
P(i|\mathcal{C}_n) &= \frac{e^{\mu_n V_{in}}}{e^{\mu_n V_{in}} + \sum_{j \in \mathcal{C}_n \setminus \{i\}} e^{\mu_n V_{jn}}} = \frac{e^{\mu_n V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{\mu_n V_{jn}}}.
\end{aligned}$$

The logit model

$$P(i|\mathcal{C}_n) = \frac{e^{\mu_n V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{\mu_n V_{jn}}}.$$

where

$$V_{in} = \sum_k \beta_k z_{ink},$$

where z_{in} is the vector of attributes of alternative i for individual n.

Outline

Choosing the units

Issue

- As utility is latent, the units are arbitrary.
- We need to choose the units.
- We first introduce a specification that is convenient to interpret.

Context

- Utility contains a cost/price variable (in CHF, say).
- We constrain its coefficient to be -1.
- Utility = opposite of generalized cost.
- Units: CHF.

Example

Setting $\beta_c = -1$

$$V_{in} = -\text{cost}_{in} + \beta_t \text{time}_{in} + \beta_h \text{direct}_{in}.$$

Interpretation of the coefficients

- Willingness to pay for an increase of the variable.
- β_t: transforms minutes into CHF: value of time (opposite).
- β_h: transforms the feature of direct service into CHF.

Logit model

Moneymetric utility function

$$V_{in} = -\text{cost}_{in} + \sum_k \beta_k z_{ink}.$$

Choice model

$$P_n(i|\mathcal{C}) = \frac{e^{\mu_n V_{in}}}{\sum_{j \in \mathcal{C}} e^{\mu_n V_{in}}} = \frac{e^{-\mu_n \text{cost}_{in} + \sum_k \mu_n \beta_k z_{ink}}}{\sum_{j \in \mathcal{C}} e^{-\mu_n \text{cost}_{jn} + \sum_k \mu_n \beta_k z_{jnk}}}.$$

Outline

- Introduction
- Binary logit model
- Logit with multiple alternatives
- Normalization
- Alternative Specific Constants
- Coefficients of attributes
- Taste heterogeneity
- Heteroscedasticity
- Nonlinear specifications

Relax the i.i.d. assumption

i.i.d. assumption

- Same η for all alternatives i: relaxed in this lecture.
- Same μ for all alternatives i: relaxed in the lecture on mixtures.
- Independence across alternatives i: relaxed in the lecture on MEV models.
- Independence across observations n: relaxed in the lecture on panel data.

Motivation

	iPhone	Samsung
Price	\$900	\$900
Cellular	5G	5G
Capacity	128GB	128GB
Display	6.6”	6.6”
Market shares US	60 %	27%

Conclusion: we cannnot assume $\mathrm{E}[\varepsilon_{in}] = \mathrm{E}[\varepsilon_{jn}]$.

Source: gs.statcounter.com/vendor-market-share/mobile

Motivation

- How can we explain the difference?
- Only the names of the alternatives are different.
- It means that the names have an important behavioral influence.
- In marketing, it is called “brand equity”.

Modeling

Alternative Specific Constants

- We define new variables:

$$z_{i,\text{iPhone}} = \begin{cases} 1 & \text{if } i \text{ is iPhone,} \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{i,\text{Samsung}} = \begin{cases} 1 & \text{if } i \text{ is Samsung,} \\ 0 & \text{otherwise.} \end{cases}$$

- We include them with coefficients in the utility function.

Alternative Specific Constants

Original formulation

$$U_{in} = V_{in} + \eta_n + \frac{1}{\mu_n}\varepsilon_{in}.$$

New formulation

$$U_{in} = V_{in} + \left(\beta_{\text{iPhone}} z_{i,\text{iPhone}} + \beta_{\text{Samsung}} z_{i,\text{Samsung}}\right) + \eta_n + \frac{1}{\mu_n}\varepsilon_{in}.$$

Alternative Specific Constants

New formulation

$$U_{\text{iPhone},n} = V_{\text{iPhone},n} + (\beta_{\text{iPhone}} + \eta_n) + \frac{1}{\mu_n}\varepsilon_{\text{iPhone},n}.$$
$$U_{\text{Samsung},n} = V_{\text{Samsung},n} + (\beta_{\text{Samsung}} + \eta_n) + \frac{1}{\mu_n}\varepsilon_{\text{Samsung},n}.$$

Location parameters

$$\eta_{\text{iPhone},n} = \beta_{\text{iPhone}} + \eta_n$$
$$\eta_{\text{Samsung},n} = \beta_{\text{Samsung}} + \eta_n$$

Alternative Specific Constants

New formulation

$$U_{\text{iPhone},n} = V_{\text{iPhone},n} + \eta_{\text{iPhone},n} + \frac{1}{\mu_n}\varepsilon_{\text{iPhone},n}.$$

$$U_{\text{Samsung},n} = V_{\text{Samsung},n} + \eta_{\text{Samsung},n} + \frac{1}{\mu_n}\varepsilon_{\text{Samsung},n}.$$

Comments

- The use of ASCs relaxes the assumption that the location parameters are the same across alternatives.
- Abuse of language: the coefficients of the ASCs are often called ASCs themselves.
- In practice, the numbering of alternatives is often consistent with their names (iPhone is always alternative 1, etc.) But it is not required.
- There is an identification issue.

Identification

Issue

For any value of η_n,

$$P_n(i|\mathcal{C}_n) = \Pr(V_{in} + \eta_n + \eta_i + \frac{1}{\mu_n}\varepsilon_{in} \geqslant V_{jn} + \eta_n + \eta_j + \frac{1}{\mu_n}\varepsilon_{jn}, \ \forall j \in \mathcal{C}_n).$$

Normalization

- Select any arbitrary alternative j, and normalize $\eta_n = -\eta_j$.
- It is equivalent to normalize the ASC of alternative j to zero.
- If there are J_n alternatives, only $J_n - 1$ ASCs are identified.

Example: binary choice

Normalization of the iPhone constant

$$U_{\text{iPhone},n} = V_{\text{iPhone},n} + \frac{1}{\mu_n}\varepsilon_{\text{iPhone},n}.$$

$$U_{\text{Samsung},n} = V_{\text{Samsung},n} + \eta_{\text{Samsung},n} + \frac{1}{\mu_n}\varepsilon_{\text{Samsung},n}.$$

Normalization of the Samsung constant

$$U_{\text{iPhone},n} = V_{\text{iPhone},n} + \eta_{\text{iPhone},n} + \frac{1}{\mu_n}\varepsilon_{\text{iPhone},n}.$$

$$U_{\text{Samsung},n} = V_{\text{Samsung},n} + \frac{1}{\mu_n}\varepsilon_{\text{Samsung},n}.$$

Motivation

	iPhone	Samsung
Price	$900	$900
Cellular	5G	5G
Capacity	128GB	128GB
Display	6.6"	6.6"
Market shares (US)	60 %	27%
Market shares (Europe)	36 %	31%
Market shares (CH)	58 %	27%

Conclusion: ASC may vary across *n*. Heterogenous population.

Source: gs.statcounter.com/vendor-market-share/mobile

Modeling heterogeneity

Behavioral assumption

- Individuals have different taste parameters.
- This is why $\eta_{\text{iPhone},n}$ and $\beta_{\text{Samsung},n}$ carry an index *n*.
- In practice, it is in general statistically impossible to estimate a different β per individual.
- Instead, the heterogeneity is captured by socio-economic characteristics:

$$\beta_{\text{iPhone,n}} = \beta_{\text{iPhone},n}(\text{country}_n; \theta)$$

Modeling heterogeneity

ASC function of the socio-economic characteristics

$$\eta_{in} = \theta_{i1}\ s_n + \theta_{i2}\ (1 - s_n),$$

where $s_n = 1$ if n resides in the USA, 0 otherwise.

Specification of the utility

$$\begin{aligned} U_{in} &= V_{in} + \theta_{i1}\ s_n + \theta_{i2}\ (1 - s_n) + \frac{1}{\mu}\varepsilon_{in}, \\ &= V_{in} + (\theta_{i1} - \theta_{i2})\ s_n + \theta_{i2} + \frac{1}{\mu}\varepsilon_{in}, \\ &= V_{in} + \theta_{i,\text{diff}}\ s_n + \theta_{i2} + \frac{1}{\mu}\varepsilon_{in}. \end{aligned}$$

Modeling heterogeneity

Notes

- Including a socio-economic characteristic in the utility captures an heterogenous alternative specific constant.
- The constant normalized to zero must not be segmented. Therefore, no socio-economic characteristic should appear alone in the specification.
- There are two possible specifications:

$$
\begin{aligned}
U_{in} &= V_{in} + \theta_{i1}\, s_n + \theta_{i2}\, (1 - s_n) + \frac{1}{\mu}\varepsilon_{in}, \\
U_{in} &= V_{in} + \theta_{i,\text{diff}}\, s_n + \theta_{i2} + \frac{1}{\mu}\varepsilon_{in}.
\end{aligned}
$$

The first is easier to interpret. The second is more appropriate to combine several characteristics.

Modeling heterogeneity

Curse of dimensionality

- Several socio-economic characteristics can be used to segment the population.
- However, the number of segments may become too large.
- For example, if K is the number of binary socio-economic characteristics, the number of segments is 2^K.

Practical simplification

- Include a reference parameter.
- Include one parameter for each binary socio-economic characteristic, capturing the difference with the reference one.
- Total number of parameters: $1 + K$.

Modeling heterogeneity

Example

Three binary characteristics: s_1, s_2, s_3.

Combinatorial specification: $2^3 = 8$ parameters

$$\begin{aligned}
&\theta_1 s_1 s_2 s_3 + \theta_2 s_1 s_2 (1 - s_3) + \theta_3 s_1 (1 - s_2) s_3 + \theta_4 s_1 (1 - s_2)(1 - s_3) + \\
&\theta_5 (1 - s_1) s_2 s_3 + \theta_6 (1 - s_1) s_2 (1 - s_3) + \\
&\theta_7 (1 - s_1)(1 - s_2) s_3 + \theta_8 (1 - s_1)(1 - s_2)(1 - s_3).
\end{aligned}$$

Practical simplification: $1 + 3 = 4$ parameters

$$\theta_{\text{ref}} + \theta_1 s_1 + \theta_2 s_2 + \theta_3 s_3.$$

Modeling heterogeneity

Practical simplification: $1 + 3 = 4$ parameters

$$\theta_{\text{ref}} + \theta_1 s_1 + \theta_2 s_2 + \theta_3 s_3.$$

Value per segment

s_1	s_2	s_3	θ	s_1	s_2	s_3	θ
0	0	0	θ_{ref}	1	0	0	$\theta_{\text{ref}} + \theta_1$
0	0	1	$\theta_{\text{ref}} + \theta_3$	1	0	1	$\theta_{\text{ref}} + \theta_1 + \theta_3$
0	1	0	$\theta_{\text{ref}} + \theta_2$	1	1	0	$\theta_{\text{ref}} + \theta_1 + \theta_2$
0	1	1	$\theta_{\text{ref}} + \theta_2 + \theta_3$	1	1	1	$\theta_{\text{ref}} + \theta_1 + \theta_2 + \theta_3$

Extension to all parameters

Alternative Specific Constants

- Qualitative variables.
- Alternative specific coefficient.
- Heterogeneity: coefficients vary across individuals.

This potentially applies to all coefficients

Outline

Quantitative attributes

Numerical and continuous

- $z_{ink} \in \mathbb{R}, \forall i, n, k$.
- Associated with a specific unit.
- Vary across both i and n.

Examples

- Auto in-vehicle time (in minutes).
- Transit in-vehicle time (in minutes).
- Auto out-of-pocket cost (in cents).
- Transit fare (in cents).
- Walking time to the bus stop (in minutes).

Straightforward modeling

Quantitative attributes

- β transforms the unit of the corresponding attribute into utility units.
- Example: consider two moneymetric specifications:

$$\begin{array}{rcl} \mu V_{in} & = & -\mu \text{cost} + \mu\beta \text{TT}_{in} + \cdots \\ \mu V_{in} & = & -\mu \text{cost} + \mu\beta' \text{TT}'_{in} + \cdots \end{array}$$

- If TT_{in} is a number of minutes, the unit of β is CHF/minute.
- If TT'_{in} is a number of hours, the unit of β' is CHF/hour.
- Both models are equivalent, but the estimated value of the coefficient will be different

$$\beta \text{TT}_{in} = \beta' \text{TT}'_{in} \Longrightarrow \frac{\beta'}{\beta} = \frac{\text{TT}_{in}}{\text{TT}'_{in}} = 60.$$

Qualitative attributes

Examples

- Level of comfort for the train.
- Reliability of the bus.
- Color.
- Shape.
- etc.

Modeling

Identify all possible levels of the variable

- Comfortable,
- Rather comfortable,
- Not comfortable.

Introduce a 0/1 attribute for each level k

$$z_k \in \{0, 1\}, \forall k = 1, \ldots, K,$$

$$\sum_{k=1}^{K} z_k = 1.$$

Example with $K = 3$

- z_{c} for comfortable.
- z_{rc} for rather comfortable.
- z_{nc} for not comfortable.

	z_{c}	z_{rc}	z_{nc}
comfortable	1	0	0
rather comfortable	0	1	0
not comfortable	0	0	1

Modeling

Utility function with ASC

$$
\begin{aligned}
z_{nc} &= 1 - z_{\text{c}} - z_{\text{rc}}. \\
\mu V_i &= -\mu(\text{cost}_i + \beta_{\text{c}} z_{\text{c}} + \beta_{\text{rc}} z_{\text{rc}} + \beta_{\text{nc}} z_{\text{nc}} + \beta_i) \\
&= -\mu(\text{cost}_i + \beta_{\text{c}} z_{\text{c}} + \beta_{\text{rc}} z_{\text{rc}} + \beta_{\text{nc}}(1 - z_{\text{c}} - z_{\text{rc}}) + \beta_i) \\
&= -\mu(\text{cost}_i + (\beta_{\text{c}} - \beta_{\text{nc}}) z_{\text{c}} + (\beta_{\text{rc}} - \beta_{\text{nc}}) z_{\text{rc}} + \beta_{\text{nc}} + \beta_i) \\
&= -\mu(\text{cost}_i + \beta'_{\text{c}} z_{\text{c}} + \beta'_{\text{rc}} z_{\text{rc}} + \beta'_i)
\end{aligned}
$$

Comments

- Index *n* omitted.
- Coefficient of one level is confounded with the ASC.
- In the presence of an ASC, only $K-1$ binary variables can be included.

Alternative specific coefficients

Generic vs alternative specific

$$\begin{array}{rcl} \mu_n V_{\text{car},n} & = & -\mu_n \text{cost} + \mu_n \beta_t \text{TT}_{\text{car},n} + \cdots \\ \mu_n V_{\text{bus},n} & = & -\mu_n \text{cost} + \mu_n \beta_t \text{TT}_{\text{bus},n} + \cdots \end{array}$$

or

$$\begin{array}{rcl} \mu_n V_{\text{car},n} & = & -\mu_n \text{cost} + \mu_n \beta_{t,\text{car}} \text{TT}_{\text{car},n} + \cdots \\ \mu_n V_{\text{bus},n} & = & -\mu_n \text{cost} + \mu_n \beta_{t,\text{bus}} \text{TT}_{\text{bus},n} + \cdots \end{array}$$

Modeling assumption: a minute has/has not the same value whether it is incurred on the auto or bus mode.

Outline

- Introduction
- Binary logit model
- Logit with multiple alternatives
- Normalization
- Alternative Specific Constants
- Coefficients of attributes
- Taste heterogeneity
- Heteroscedasticity
- Nonlinear specifications

Modeling heterogeneity

Behavioral assumption

- Individuals have different taste parameters.
- The difference is explained by **continuous** socio-economic characteristics.

Example: one characteristic

$$\mu_n V_{in} = -\mu_n \text{cost}_{in} + \mu_n \beta_n z_{in} + \cdots$$

where

$$\beta_n = \beta_n(\text{income}_n).$$

Modeling heterogeneity

Interaction

Typical definition of β_n:

$$\beta_n = \beta \text{ income}_n.$$

$$\mu_n V_{in} = -\mu_n \text{cost}_{in} + \mu_n \beta \text{ income}_n z_{in} + \cdots$$
$$\mu_n V_{in} = -\mu_n \text{cost}_{in} + \mu_n \beta x_{in} + \cdots$$

where

$$x_{in} = \text{income}_n z_{in}.$$

Modeling heterogeneity

Example: several characteristics

$$\beta_n = \beta_n(\text{income}_n, \text{age}_n).$$

Interaction

$$\beta_n = \beta \text{ income}_n \text{ age}_n.$$

$$\begin{aligned}
\mu_n V_{in} &= -\mu_n \text{cost}_{in} + \mu_n \beta_n z_{in} + \cdots \\
&= -\mu_n \text{cost}_{in} + \mu_n \beta \text{ income}_n \text{ age}_n z_{in} + \cdots \\
&= -\mu_n \text{cost}_{in} + \mu_n \beta x_{in} + \cdots
\end{aligned}$$

where $x_{in} = \text{income}_n \text{age}_n z_{in}$.

Modeling heterogeneity

Functional forms

$$\begin{aligned} \beta_n &= \beta \text{ income}_n, \\ \beta_n &= \beta / \text{income}_n, \\ \beta_n &= \beta \log(\text{income}_n), \end{aligned}$$

etc.

Qualitative characteristics

Examples

- Sex.
- Education.
- Professional status.
- etc.

Modeling heterogeneity

Behavioral assumption

- Individuals have different taste parameters.
- The difference is explained by **qualitative/discrete** socio-economic characteristics.

Example: one characteristic

$$\mu_n V_{in} = -\mu_n \text{cost}_{in} + \mu_n \beta_n z_{in} + \cdots$$

where

$$\beta_n = \beta_n(\text{education}_n).$$

Modeling heterogeneity

Segmentation

- Assume that there are K levels for the qualitative variable (e.g. education).
- They characterize K segments in the population.
- Define

$$\delta_{kn} = \begin{cases} 1 & \text{if individual } n \text{ is associated with segment } k, \\ 0 & \text{otherwise.} \end{cases}$$

- Introduce a parameter β^k for each level and define

$$\beta_n = \sum_{k=1}^{K} \beta^k \delta_{kn}.$$

Modeling heterogeneity

Segmentation

$$\mu_n V_{in} = -\mu_n \mathrm{cost}_{in} + \mu_n \beta_n z_{in} + \cdots$$

$$\mu_n V_{in} = -\mu_n \mathrm{cost}_{in} + \mu_n \sum_{k=1}^{K} \beta^k \delta_{kn} z_{in} + \cdots$$

$$\mu_n V_{in} = -\mu_n \mathrm{cost}_{in} + \mu_n \sum_{k=1}^{K} \beta^k x_{ink} + \cdots$$

where

$$x_{ink} = \delta_{kn} z_{in}.$$

Segmentation with several variables

Example

- Gender (M, F).
- House location (metro, suburb, perimeter areas).
- 6 segments: (M, m), (M, s), (M, p), (F, m), (F, s), (F, p).

Segmentation

Specification

$$\beta_{M,m} TT_{M,m} + \beta_{M,s} TT_{M,s} + \beta_{M,p} TT_{M,p} +$$
$$\beta_{F,m} TT_{F,m} + \beta_{F,s} TT_{F,s} + \beta_{F,p} TT_{F,p} +$$

$TT_i = TT$ if indiv. belongs to segment i, and 0 otherwise.

Remarks

- For a given individual, exactly one of these terms is non zero.
- The number of segments grows exponentially with the number of variables.
- Practical solution: see the discussion about segmented ASCs.

Outline

Motivation

Different variance

- Scale parameter μ_n may vary across n.
- It means that the variance of the error terms vary across n.
- People may have different levels of knowledge (e.g. taxi drivers).
- Different sources of data:
  - RP / SP
  - Different locations.
  - Different points in time.

Heteroscedasticity

Data

- K groups of data.
- Each individual n belongs to exactly one group k.
- Characterized by indicators:

$$\delta_{kn} = \begin{cases} 1 & \text{if } n \text{ belongs to group k,} \\ 0 & \text{otherwise,} \end{cases}$$

and $\sum_k \delta_{kn} = 1$, for all n.

Each individual is observed at most once. This assumption will be relaxed in the lecture on panel data.

Heteroscedasticity

Assumption: variance of error terms is different across groups

- Define one scale parameter per group: μ_k, $k = 1, \ldots, K$.
- Each individual is associated with the parameter of her group:

$$\mu_n = \sum_{k=1}^{K} \delta_{kn} \mu_k.$$

Heteroscedasticity

Moneymetric utility: $\beta_c = -1$

$$\mu_1 V_{in_1} = -\mu_1 \text{cost}_{in_1} + \sum_k \mu_1 \beta_k z_{in_1 k},$$
$$\mu_2 V_{in_2} = -\mu_2 \text{cost}_{in_2} + \sum_k \mu_2 \beta_k z_{in_2 k}.$$

Comment

- Estimate K scale parameters, one per group.

Outline

- Introduction
- Binary logit model
- Logit with multiple alternatives
- Normalization
- Alternative Specific Constants
- Coefficients of attributes
- Taste heterogeneity
- Heteroscedasticity
- Nonlinear specifications

Nonlinear transformations of the variables

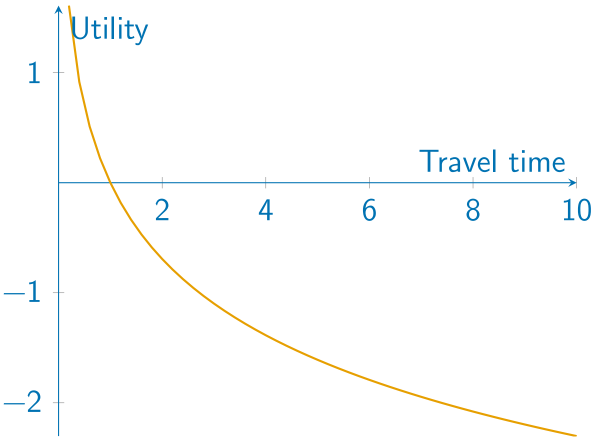
Example with travel time

- Compare a trip of 5 min with a trip of 10 min.
- Compare a trip of 120 min with a trip of 125 min.
- Utility difference: $\beta_T \cdot$ 5 min, in both cases.

Behavioral assumption

One more minute of travel is not perceived the same way for short trips as for long trips.

Nonlinear transformations of the variables



Nonlinear transformations of the variables

Assumption 1: the marginal impact of travel time is constant

$$V_{in} = \beta_T \text{time}_{in} + \cdots$$

Assumption 2: the marginal impact of travel time decreases with travel time

$$V_{in} = \beta_T \ln(\text{time}_{in}) + \cdots$$

Nonlinear transformations of the variables

Data can be preprocessed to account for nonlinearities

$$V_{in} = V(h(z_{in}, s_n)) = \sum_k \beta_k (h(z_{in}, s_n))_k$$

Note

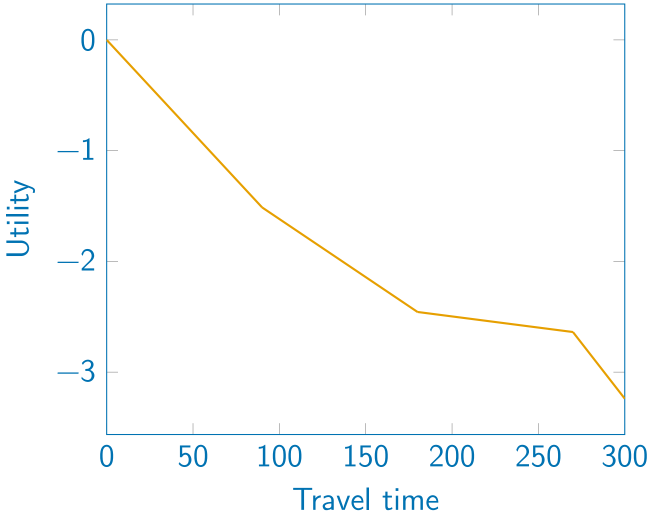
Interactions between attributes and socio-economic characteristics are a special case of *h*.

Piecewise linear specification

Again: sensitivity to travel time varies with travel time

- Log transform is not the only specification.
- Another possibility: split the range of values of the variable.
- Each category is associated with a different coefficient.

Piecewise linear specification



Piecewise linear specification

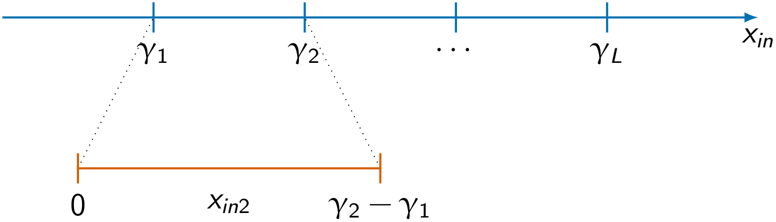
Procedure

- Select breakpoints $\gamma_1 < \gamma_2 < \ldots < \gamma_L$.
- Define new variables.

Piecewise linear specification

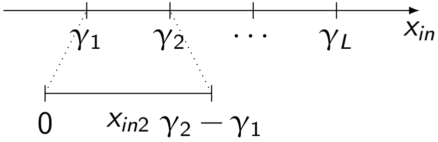
Piecewise linear specification

Piecewise linear specification



Piecewise linear specification

Formulation



$$x_{in1} = \begin{cases} x_{in} & \text{if } x_{in} < \gamma_1, \\ \gamma_1 & \text{otherwise.} \end{cases}$$

$$x_{in\ell} = \begin{cases} 0 & \text{if } x_{in} < \gamma_{\ell-1}, \\ x_{in} - \gamma_{\ell-1} & \text{if } \gamma_{\ell-1} \leqslant x_{in} < \gamma_\ell, \\ \gamma_\ell - \gamma_{\ell-1} & \text{otherwise.} \end{cases}$$

$$x_{in(L+1)} = \begin{cases} 0 & \text{if } x_{in} < \gamma_L, \\ x_{in} - \gamma_L & \text{otherwise.} \end{cases}$$

Piecewise linear specification

Equivalent formulations

$$x_{in1} = \begin{cases} x_{in} & \text{if } x_{in} < \gamma_1, \\ \gamma_1 & \text{otherwise.} \end{cases} \qquad \min(x_{in}, \gamma_1).$$

$$x_{in\ell} = \begin{cases} 0 & \text{if } x_{in} < \gamma_{\ell-1}, \\ x_{in} - \gamma_{\ell-1} & \text{if } \gamma_{\ell-1} \leqslant x_{in} < \gamma_\ell, \\ \gamma_\ell - \gamma_{\ell-1} & \text{otherwise.} \end{cases} \qquad \max(0, \min(x_{in} - \gamma_{\ell-1}, \gamma_\ell - \gamma_{\ell-1})).$$

$$x_{in(L+1)} = \begin{cases} 0 & \text{if } x_{in} < \gamma_L \\ x_{in} - \gamma_L & \text{otherwise} \end{cases} \qquad \max(0, x_{in} - \gamma_L).$$

Piecewise linear specification

Examples

$\gamma_1 = 90$, $\gamma_2 = 180$, $\gamma_3 = 270$.

x_{in}	50	100	200	300
x_{in1}	50	90	90	90
x_{in2}	0	10	90	90
x_{in3}	0	0	20	90
x_{in4}	0	0	0	30

$\gamma_1 = 1$, $\gamma_2 = 5$, $\gamma_3 = 10$.

x_{in}	0.5	4	8	12
x_{in1}	0.5	1	1	1
x_{in2}	0	3	4	4
x_{in3}	0	0	3	5
x_{in4}	0	0	0	2

Utility function

$$V_{in} = \sum_{\ell=1}^{L} \beta_\ell x_{in\ell}.$$

Box-Cox transforms

Box and Cox (1964)

$$V_{in} = \beta B(x_{in};\lambda) + \cdots$$

where

$$B(x_{in};\lambda) = \begin{cases} \dfrac{x_{in}^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln x_{in} & \text{if } \lambda = 0, \end{cases}$$

and $x_{in} > 0$.

Box-Cox transforms

Continuity

The function is continuous and differentiable in λ, because

$$\lim_{\lambda \to 0} \frac{x_{in}^{\lambda} - 1}{\lambda} = \ln x_{in}.$$

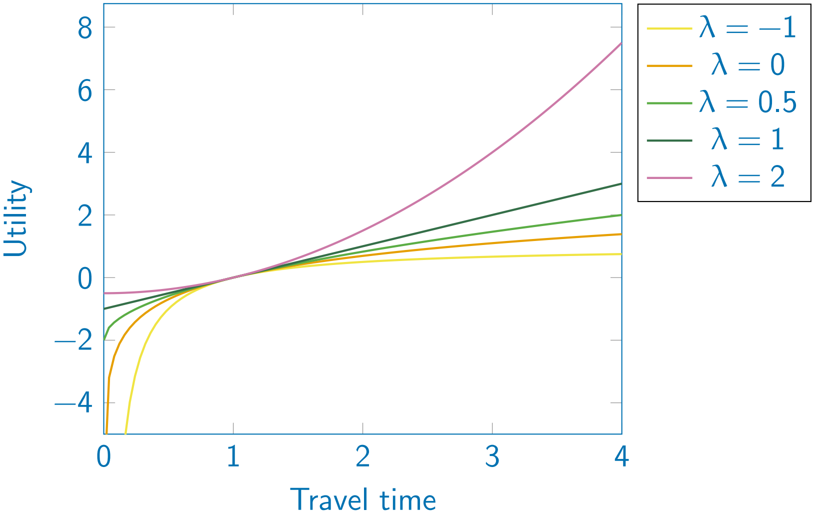
Box-Cox transforms

Convexity / concavity

$$\partial B(x_{in};\lambda)/\partial x_{in} = x^{\lambda-1},$$
$$\partial^2 B(x_{in};\lambda)/\partial x_{in}^2 = (\lambda-1)x^{\lambda-2}.$$

- B is convex if $\lambda \geqslant 1$.
- B is concave if $\lambda \leqslant 1$.

Box-Cox transforms



Nonlinear interactions

Taste heterogeneity

$$V_{in} = \beta_{in} z_{in} + \cdots$$

Linear interaction

$$\beta_{in} = \beta_i \text{ income}_n.$$

Nonlinear interaction

$$\beta_{in} = \beta_i \text{ income}_n^{\lambda}, \text{ where } \lambda = \frac{\partial \beta_{in}}{\partial \text{income}_n} \frac{\text{income}_n}{\beta_{in}}.$$

Nonlinear interactions

Remarks

- λ must be estimated.
- Use a reference value for the socio-economic characteristic:

$$\beta_{in} = \beta_i \left(\frac{\text{income}_n}{\text{refIncome}}\right)^{\lambda}.$$

- Reference value is arbitrary.
- Several (continuous) characteristics can be combined:

$$\beta_{in} = \beta_i \left(\frac{\text{income}_n}{\text{refIncome}}\right)^{\lambda_1} \left(\frac{\text{age}_n}{\text{refAge}}\right)^{\lambda_2}.$$

Summary

- ε_{in} i.i.d. EV(η, μ).
- Derivation: from binary logit to multiple alternatives.
- Identification issues due to the latent nature of utility.
- Normalization: η_n does not play any role.
- Normalization: $\beta_c = -1$: moneymetric specification.
- Alternative Specific Constants.
- Continuous/quantitative vs discrete/qualitative variables.
- Taste heterogeneity and interactions.
- Heteroscedasticity.
- Nonlinear specifications.

Appendices

- The binary probit model.
- Gumbel's theorem.

Appendix I: the probit model

Assumption: similar to linear regression

ε_{in} and ε_{jn} are the sum of many r.v. capturing unobserved attributes (e.g. mood, experience), measurement and specification errors.

Central limit theorem

The sum of many i.i.d. random variables approximately follows a normal distribution: $N(\eta, \sigma^2)$.

The normal distribution N(η, σ^2)

Probability density function (pdf)

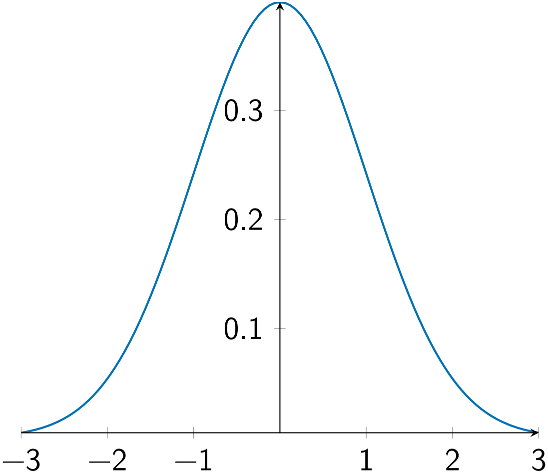
$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\eta}{\sigma}\right)^2}.$$

Cumulative distribution function (CDF)

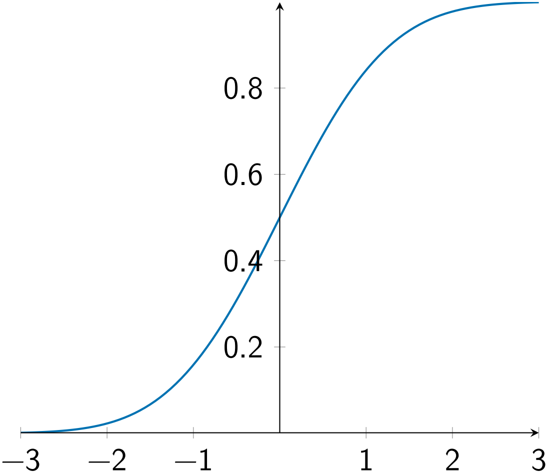
$$P(c \geqslant \varepsilon) = F(c) \quad = \quad \int_{-\infty}^{c} f(t) dt.$$

The normal distribution

pdf N(0,1)



CDF N(0,1)



The distribution

Assumptions

- ε_{in} and ε_{jn} are normally distributed, with variance σ_i^2 and σ_j^2, respectively, and covariance σ_{ij}.
- Note: identical distribution across n.
- If an alternative specific constant is in the model, their mean can be assumed to be any constant.
- $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$ is also normally distributed, with variance

$$\sigma^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}.$$

The binary probit model

Choice model

$$P_n(i|\{i,j\}) = \Pr\left(\varepsilon_n \leqslant V_{in} - V_{jn}\right) = F_\varepsilon(V_{in} - V_{jn}).$$

The binary probit model

$$P_n(i|\{i,j\}) = \Phi\left(\frac{V_{in} - V_{jn}}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(V_{in} - V_{jn})/\sigma} \exp\left(-\frac{1}{2}u^2\right) du.$$

Appendix II: Gumbel theorem

Motivation

- $X_1, \ldots, X_n$ i.i.d.
- $f_{X_i}(x) = f(x)$, $F_{X_i}(x) = F(x)$, $i = 1, \ldots, n$
- $X'_n = \max(X_1, \ldots, X_n)$.
- Applications:
  - rainfall,
  - floods,
  - earthquakes,
  - air pollution,
  - ...

Extreme value distribution

- $X'_n = \max(X_1, \ldots, X_n)$.
- $F_{X'_n} = F(x)^n$. Indeed

$$P(X'_n \leqslant x) = P(X_1 \leqslant x)P(X_2 \leqslant x)\ldots P(X_n \leqslant x).$$

- Warning: if $n \to \infty$

$$\lim_{n\to\infty} F_{X'_n}(x) = \begin{cases} 1 & \text{if } F(x) = 1, \\ 0 & \text{if } F(x) < 1. \end{cases}$$

Degenerate distribution (if you a die sufficiently many times, the maximum score will always be 6).

Extreme value distribution

- We want a limiting distribution which is non degenerate.
- Limiting distribution of some sequence of transformed “reduced” values.
- For instance $a_n X_n' + b_n$.
- a_n, b_n do not depend on x.
- CDF of limiting distribution: $G(x)$.
- Let’s identify desired properties.

Extreme value distribution

$$\begin{array}{ccc|l}
X_1 & \dots & X_n & \max(X_1, \dots, X_n) \\
X_{n+1} & \dots & X_{2n} & \max(X_{n+1}, \dots, X_{2n}) \\
\vdots & & \vdots & \\
X_{(i-1)n+1} & \dots & X_{in} & \max(X_{(i-1)n+1}, \dots, X_{in}) \\
\vdots & & \vdots & \\
X_{(N-1)n+1} & \dots & X_{Nn} & \max(X_{(N-1)n+1}, \dots, X_{Nn})
\end{array}$$

Two ways of seeing $\max(X_1, \dots, X_{Nn})$ when $n \to \infty$.

1. As a max of many X_i, the CDF should look like $G(a_N x + b_N)$.
2. The CDF of the max of each row is $G(x)$.
3. So the CDF of the max of all rows is $G(x)^N$.

Extreme value distribution

Stability postulate (Fréchet, 1927):

$$G(x)^N = G(a_N x + b_N).$$

We consider here the case $a_N = 1$ to obtain the so-called "type I extreme value distribution"

$$G(x)^N = G(x + b_N).$$

We have also

$$\begin{aligned} G(x)^{MN} &= G(x + b_N)^M &= G(x + b_N + b_M), \\ G(x)^{MN} &= G(x + b_{MN}). \end{aligned}$$

Extreme value distribution

Therefore

$$G(x + b_N + b_M) = G(x + b_{MN}),$$

that is

$$b_N + b_M = b_{MN},$$

so that b_N must be of the form

$$b_N = -\mu' \ln N,$$

and the stability postulate becomes

$$G(x)^N = G(x - \mu' \ln N).$$

Let's take the logarithm twice...

Extreme value distribution

$$G(x)^N = G(x - \mu' \ln N).$$

$$N \ln G(x) = \ln G(x - \mu' \ln N).$$

Warning: G is a CDF, so $G(x) \leqslant 1$ and $\ln G(x) \leqslant 0$, $\forall x$.

$$-N \ln G(x) = -\ln G(x - \mu' \ln N).$$

$$\ln N + \ln(-\ln G(x)) = \ln(-\ln G(x - \mu' \ln N)).$$

Define $h(x) = \ln(-\ln G(x))$ to obtain

$$\ln N + h(x) = h(x - \mu' \ln N).$$

h is affine.

Extreme value distribution

$$\begin{aligned}
\ln N + h(x) &= h(x - \mu' \ln N), \\
h(x) &= \alpha x + \beta, \\
h(0) &= \beta, \\
\ln N + \alpha x + \beta &= \alpha(x - \mu' \ln N) + \beta, \\
\alpha &= -\tfrac{1}{\mu'}.
\end{aligned}$$

Therefore

$$h(x) = h(0) - \frac{x}{\mu'}.$$

G is increasing in x (CDF), so h is decreasing in x. Therefore, $\mu' > 0$.

Extreme value distribution

$$h(x) = \ln(-\ln G(x)) = h(0) - \frac{x}{\mu'}.$$

$$-\ln G(x) = \exp\left(h(0) - \frac{x}{\mu'}\right) = \exp\left(-\frac{x - \mu' h(0)}{\mu'}\right).$$

$$G(x) = \exp\left(-\exp\left(-\frac{x - \mu' h(0)}{\mu'}\right)\right).$$

Let $\mu = 1/\mu'$ and $\eta = \mu' h(0) = \ln(-\ln G(0))/\mu$

$$G(x) = \exp\left(-\exp\left(-\mu(x - \eta)\right)\right).$$