交通大数据

二元Logistic回归

- 郭延永、徐铖铖、李豪杰
- guoyanyong@seu.edu.cn

Binary Logistic Regression

- Logistic regression, which is useful for predicting **a binary** dependent variable Y （0,1） as a function of predictor variables
- As an example, consider a model of a driver's stated propensity to exceed the speed limit on a highway as a function of various exogenous factors.
- The goal of logistic regression, much like linear regression, is to identify the best fitting model that describes the relationship between **a binary dependent** variable and a set of independent or **explanatory variables**.

Binary Logistic Regression

- The logit is the LN (to base) of the odds, or likelihood ratio, that the dependent variable is 1, such that

$$\mathrm{logit}(p_i) = \ln\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$$

The Eq. (2) can be written as:

$$P(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})}$$

Odds （机会比率）

$$\frac{p}{1-p}$$

- where P is the probability of event occurred (y=1), and 1-P is the probability of event not occurred (y=0), β0 is the model constant and the βi are the parameter estimates for the independent variables (the set of independent variables).
- The probability P ranges from 0 to 1, while the natural logarithm LN(P/(1 – P)) ranges from negative infinity to positive infinity.

Binary Logistic Regression

- Using a logistic regression model, the probability of exceeding the speed limit under certain traffic conditions can be estimated using the following equation

$$P(y_i = 1) = \frac{1}{1 + e^{-x'_i\beta}} \quad (i = 1, 2, \ldots, n)$$

- where P(yi = 1) denotes the probability of exceeding the speed limit, and$-x'_i\boldsymbol{\beta}$ is the multiple linear combination of explanatory variables:

$$x_i'\beta = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$$

- This model is estimable by standard maximum likelihood methods, with the likelihood function given as:

$$L(\beta) = \prod_{i=1}^{n}\left[\left(\frac{1}{1+e^{-x_i'\beta}}\right)^{y_i}\left(1-\frac{1}{1+e^{-x_i'\beta}}\right)^{(1-y_i)}\right]$$

Binary Logistic Regression

- The interpretation of variable parameter in logistic regression:

$$Y = \text{logit}(P) = LN\left(\frac{P_i}{1-P_i}\right) = B_0 + B_i \cdot X_i$$

$$\left(\frac{P_i}{1-P_i}\right) = EXP^{B_0+B_iX_i} = EXP^{B_0}EXP^{B_iX_i}$$

- The fundamental equation for the logistic regression shows that when the value of an independent variable increases by one unit, and all other variables are held constant, the new probability ratio $[P_i/(1 - P_i)]$ is given as

$$\left(\frac{P_i}{1-P_i}\right)^* = EXP^{B_0}EXP^{B_i(X_i+1)} = EXP^{B_0}EXP^{B_iX_i}EXP^{B_i} = \left(\frac{P_i}{1-P_i}\right)EXP^{B_i}$$

Binary Logistic Regression

$$\left(\frac{P_i}{1-P_i}\right) = EXP^{B_0+B_iX_i} = EXP^{B_0}EXP^{B_iX_i}$$

$$\left(\frac{P_i}{1-P_i}\right)^* = EXP^{B_0}EXP^{B_i(X_i+1)} = EXP^{B_0}EXP^{B_iX_i}EXP^{B_i} = \left(\frac{P_i}{1-P_i}\right)EXP^{B_i}$$

Thus when independent variable X_i increases by one unit, with all other factors remaining constant, the odds $[P_i/(1-P)]$ increases by a factor EXP^{B_i}. The factor EXP^{B_i} is called the odds ratio (OR) and ranges from zero to positive infinity. It indicates the relative amount by which the odds of the outcome increases (OR > 1) or decreases (OR < 1) when the value of the corresponding independent variable increases by 1 unit.

Binary Logistic Regression

- To provide some insight into the implications of parameter estimation results, elasticities are computed to determine the marginal effects of the independent variables.
- Elasticities provide an estimate of the impact of a variable on the probability of exceeding the speed limit and are interpreted as the effect of a 1% change in the variable on the probability P_i.
- For example, an elasticity of –1.32 is interpreted to mean that a 1% increase in the variable reduces the probability of exceeding the speed limit by 1.32%.

Binary Logistic Regression

- Elasticities are the correct way of evaluating the relative impact of each variable in the model. Elasticity of probability P_i is defined as

$$E_{x_{ik}}^{P_i} = \frac{\partial P_i}{\lambda_i} \times \frac{x_{ik}}{\partial x_{ik}} = \left(1 - P_i\right) \beta_k x_{ik}$$

where E represents the elasticity, x_{ik} is the value of the kth independent variable for observation i, β_k is the estimated parameter for the kth independent variable and λ_i is the expected frequency for observation i.

- Note that elasticities are computed for each observation i. It is common to report a single elasticity as the average elasticity over all i.

Binary Logistic Regression

- For indicator variables, a pseudo-elasticity is computed to estimate an approximate elasticity of the variables. The pseudo-elasticity gives the **incremental change** in probability by changes in the indicator variables.
- The pseudo-elasticity for indicator variables, is computed as

$$E_i = \left[\frac{\mathrm{EXP}[\Delta(x'\beta)][1 + \mathrm{EXP}(x_i\beta_i)]}{\mathrm{EXP}[\Delta(x'\beta)][\mathrm{EXP}(x_i\beta_i)] + 1} - 1 \right]$$

- To determine if the estimated parameter is significantly different from zero, the test statistic t*, which is approximately t distributed, is

$$t^* = \frac{\beta - 0}{\mathrm{S.E.}(\beta)}$$

Binary Logistic Regression

- The likelihood ratio test statistic is

$$X^2 = -2[LL(\beta_R) - LL(\beta_u)]$$

where $LL(\beta_R)$ is the log likelihood at convergence of the "restricted" model (sometimes considered to have all parameters in β equal to 0, or just to include the constant term, to test overall fit of the model), and $LL(\beta_U)$ is the log likelihood at convergence of the unrestricted model.

- The χ^2 statistic is χ^2 distributed with the degrees of freedom equal to the difference in the numbers of parameters in the restricted and unrestricted model (the difference in the number of parameters in the β_R and the β_U parameter vectors.

Binary Logistic Regression

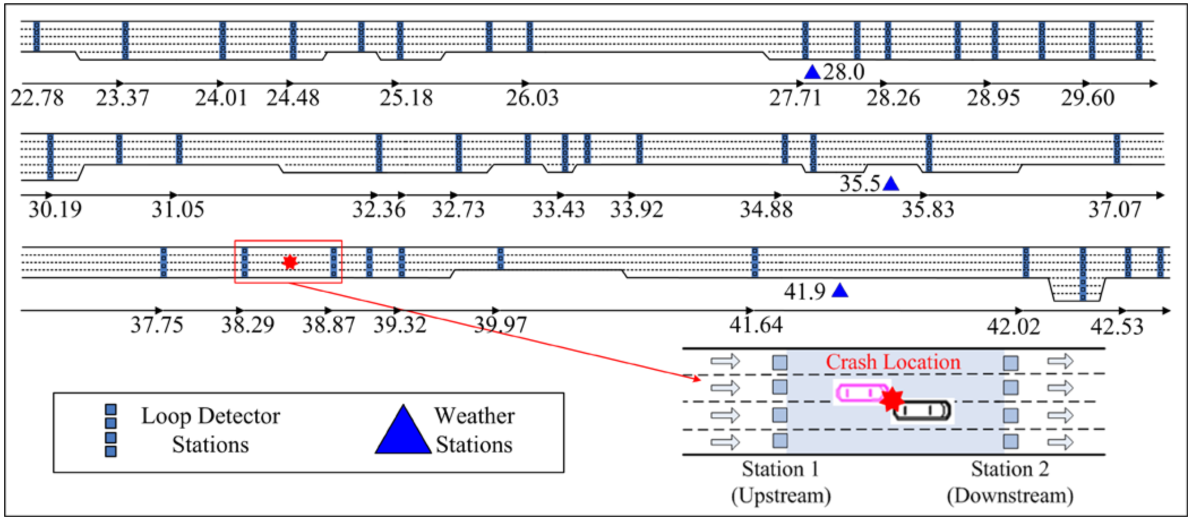
- Another measure of overall model fit is the ρ^2 statistic. The ρ^2 statistic is

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)}$$

where LL(β) is the log likelihood at convergence with parameter vector β and LL(0) is the initial log likelihood (with all parameters set to zero).

- The perfect model would have a likelihood function equal to one (all selected alternative outcomes would be predicted by the model with probability one, and the product of these across the observations would also be one) and the log likelihood would be zero, giving a ρ^2 of one.
- Thus the ρ^2 statistic is between zero and one and the closer it is to one, the more variance the estimated model is explaining.

Binary Logistic Regression



Symbol	Variables
$DetOcc_u$	Average 30-second detector occupancy at the upstream station (%)
$SpdDev_u$	Std. dev. of 30-second mean speeds at the upstream station (mile/h)
$SpdDev_d$	Std. dev. of 30-second mean speeds at the downstream station (mile/h)
$OccDif_d$	Average absolute difference in 30-second detector occupancies between adjacent lanes at the downstream station (%)
$Width_s$	Road surface width (ft)
$Width_o$	1 = if outer Shoulder width > 10ft; 0 = otherwise
Curve	1 = Curve section; 0 = otherwise

2021

2

Binary Logistic Regression

Crash (KA, BC, and PDO) vs. Non-crash					
Parameter	Estimate	Std. Error	Wald χ^2	Pr>Chisq	Elasticity
$DetOcc_u$	0.069	0.007	103.180	<.0001	0.484 (0.353[b])
$SpdDev_u$	0.046	0.016	8.510	0.004	0.175 (0.11)
$SpdDev_d$	0.047	0.017	7.615	0.006	0.177 (0.106)
$OccDif_d$	0.093	0.011	71.219	<.0001	0.271 (0.282)
$DetDist_{u\text{-}d}$	0.770	0.090	73.463	<.0001	0.403 (0.296)
$Width_s$	-0.040	0.008	25.216	<.0001	-2.042 (0.223)
$Width_o$	-0.683	0.150	20.718	<.0001	-0.336 (0.021)
Curve	0.352	0.121	8.521	0.004	0.171 (0.01)
Intercept	-1.979 (-5.397[a])	0.427	21.512	<.0001	—

Summary statistics:

$-2LL(c) = 5292.567; -2LL(\beta) = 4760.989$

$-2[LL(c) - LL(\beta)] = 531.578$ (8 df); $P < 0.0001$

Multinomial Logit Regression

- Multinomial logit modeling: given choice set, modeling choice behavior
- Binary logit is a special case of multinomial logit
- Definition of Multinomial Logit

$$P_n(i|C_n) = \frac{e^{V_{in}}}{e^{V_{in}} + e^{V_{jn}}}$$

Binary Logit Model

$$P_n(i|C_n) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}}$$

Multinomial Logit Model

Multinomial Logit Regression

- Estimation of Multinomial Logit

  - 如果从人群中抽取一个随机样本，记录每个人的解释变量X_{ik}和选择哑变量y_{ik} (i = 1, 2, … n; k = 1, 2, … J)
  - 根据多元logit模型：$P(y_{ik}=1)=\dfrac{\exp(X_{ik}\beta_k)}{\sum_{j=1}^{J}\exp(X_{ij}\beta_j)}$
  - 似然函数：

$$L(\beta)=\prod_{i=1}^{n}\left\{\prod_{k=1}^{J}\left[\frac{\exp(X_{ik}\beta_k)}{\sum_{j=1}^{J}\exp(X_{ij}\beta_j)}\right]^{y_{ik}}\right\}$$

  - 对数似然函数：$LL(\beta)=\sum_{i=1}^{n}\sum_{k=1}^{J}y_{ik}\ln\left[\dfrac{\exp(X_{ik}\beta_k)}{\sum_{j=1}^{J}\exp(X_{ij}\beta_j)}\right]$

Multinomial Logit Regression

$$LL(\beta)=\sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik}\ln\left[\frac{\exp(X_{ik}\beta_k)}{\sum_{j=1}^{J}\exp(X_{ij}\beta_j)}\right]=\sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik}\left\{X_{ik}\beta_k-\ln\left[\sum_{j=1}^{J}\exp(X_{ij}\beta_j)\right]\right\}$$

$$\frac{\partial LL(\beta)}{\partial \beta_1}=\sum_{i=1}^{n}\sum_{k=1}^{J} y_{ik}\left\{X_{i1}\delta_{ik}-X_{i1}\left[\frac{\exp(X_{i1}\beta_1)}{\sum_{j=1}^{J}\exp(X_{ij}\beta_j)}\right]\right\}=\sum_{i=1}^{n}\left\{X_{i1}\sum_{k=1}^{J} y_{ik}[\delta_{ik}-P_i(1)]\right\}=0$$

上式中，当 k = 1时, δ_{ik} =1; 当k > 1, δ_{ik} = 0

$P_i(1)$代表第i个个体选择第1个选项的概率

$$\sum_{k=1}^{J} y_{ik}[\delta_{ik}-P_i(1)]=y_{i1}[1-P_i(1)]+\sum_{k=2}^{J} y_{ik}[-P_i(1)]=y_{i1}-P_i(1)\sum_{k=1}^{J} y_{ik}=y_{i1}-P_i(1)$$

$$\sum_{i=1}^{n}\{X_{i1}[y_{i1}-P_i(1)]\}=0$$

Multinomial Logit Regression

$$\sum_{i=1}^{n}\{X_{i1}[y_{i1}-P_i(1)]\}=0$$

- X_{i1}可以是任何的选项特定变量，也可以是常数1，对应的系数是选项1的特定常数β_0。此时，

$$\sum_{i=1}^{n}[y_{i1}-P_i(1)]=0 \Rightarrow \sum_{i=1}^{n}y_{i1}=\sum_{i=1}^{n}P_i(1)$$

- 不失一般性，含有选项特定常数的多元logit模型可以保证样本中某选项的概率累加值等于样本中的该选项被选择的总次数

Multinomial Logit Regression

- Example

	A	B	C	D	E	F	G	H	I	J	K	L
1	idcase	idalt	depvar	ic	oc	income	agehead	rooms	ncoast	scoast	mountn	valley
2	1	1	1	866	199.69	70000	25	6	1	0	0	0
3	1	2	0	962.64	151.72	70000	25	6	1	0	0	0
4	1	3	0	859.9	553.34	70000	25	6	1	0	0	0
5	1	4	0	995.76	505.6	70000	25	6	1	0	0	0
6	1	5	0	1135.5	237.88	70000	25	6	1	0	0	0
7	2	1	1	727.93	168.66	50000	60	5	0	1	0	0
8	2	2	0	758.89	168.66	50000	60	5	0	1	0	0
9	2	3	0	796.82	520.24	50000	60	5	0	1	0	0
10	2	4	0	894.69	486.49	50000	60	5	0	1	0	0
11	2	5	0	968.9	199.19	50000	60	5	0	1	0	0
12	3	1	1	599.48	165.58	40000	65	2	1	0	0	0
13	3	2	0	783.05	137.8	40000	65	2	1	0	0	0
14	3	3	0	719.86	439.06	40000	65	2	1	0	0	0
15	3	4	0	900.11	404.74	40000	65	2	1	0	0	0
16	3	5	0	1048.3	171.47	40000	65	2	1	0	0	0
17	4	1	0	835.17	180.88	20000	50	4	0	1	0	0
18	4	2	0	793.06	147.14	20000	50	4	0	1	0	0
19	4	3	0	761.25	483	20000	50	4	0	1	0	0

Nested Logit Model

- **Contents**
  - Motivation for Nested Logit model
  - Derivation of Nested Logit model
  - Estimation of Nested Logit model

Nested Logit Model

- **Motivation for Nested Logit model**
  - IIA Property of MNL model

ε_{jn} independent identically distributed (i.i.d.)

$\varepsilon_{jn} \sim ExtremeValue(0,\mu) \quad \forall j$

$$P_n(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}$$

so $$\frac{P(i|C_1)}{P(j|C_1)} = \frac{P(i|C_2)}{P(j|C_2)} \qquad \forall\, i, j, C_1, C_2$$

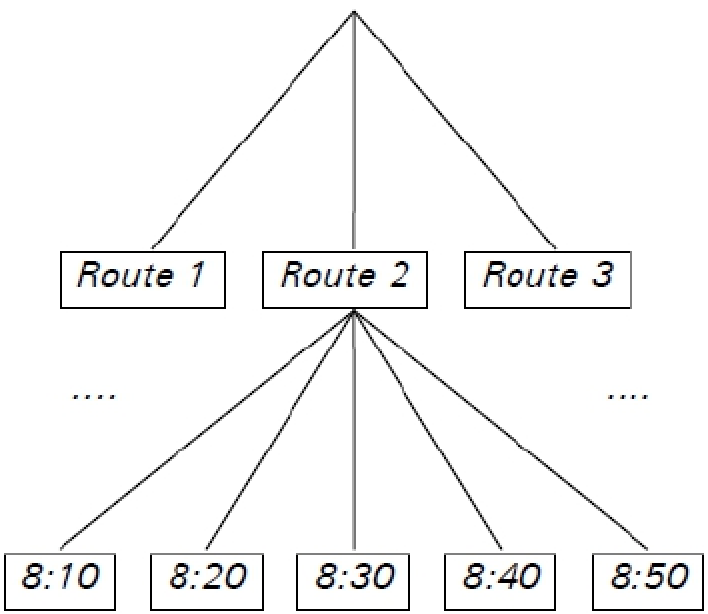
such that $i, j \in C_1$, $i, j \in C_2$, $C_1 \subseteq C_n$ and $C_2 \subseteq C_n$

Nested Logit Model

– **Motivation:** Capturing the correlation

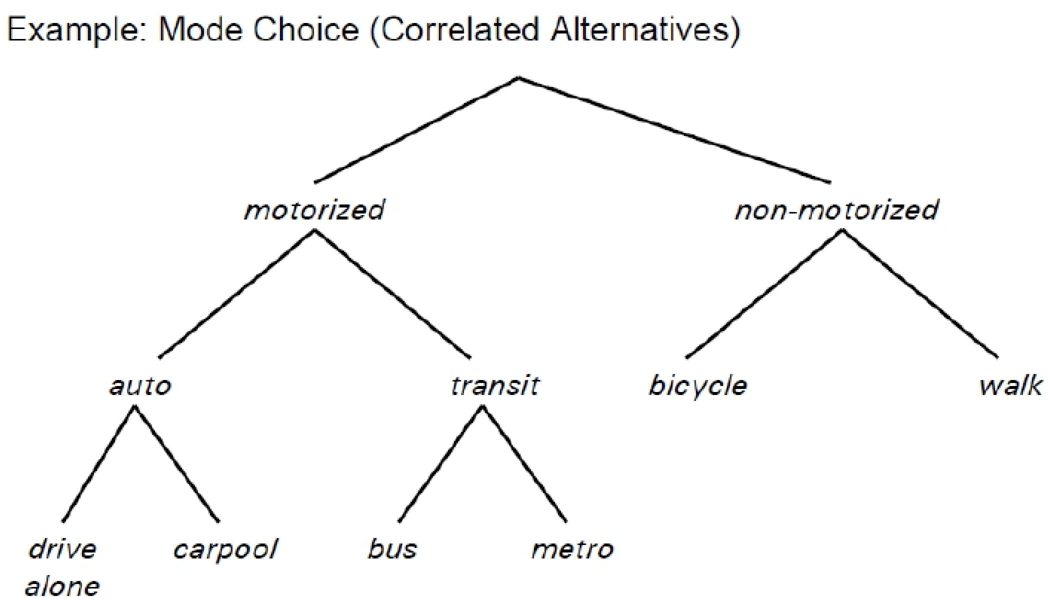
✓ Alternatives are correlated(e.g., red bus and blue bus)

✓ Multidimensional choices are considered (e.g., departure time and route)



Nested Logit Model

- **Derivation of Nested Logit model**
  - Tree Representation of Nested Logit

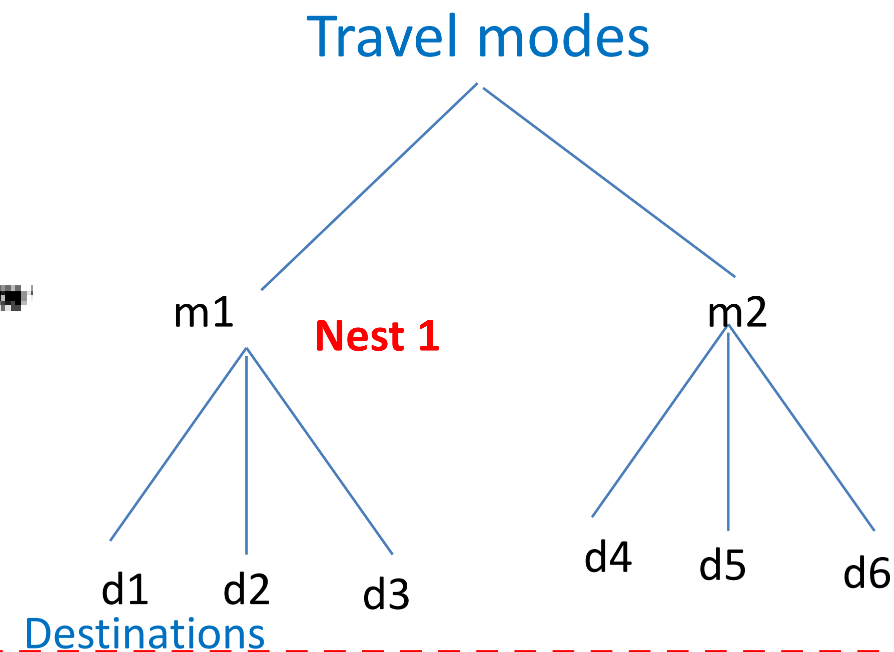


Nested Logit Model

– Assumptions of Nested Logit

$$U_{dm} = \tilde{V}_d + \tilde{V}_m + \tilde{V}_{dm} + \tilde{\varepsilon}_m + \tilde{\varepsilon}_{dm}$$

$$P_n(d) = P_n(m)P(d|m)$$



1. $\tilde{\varepsilon}_m$ and $\tilde{\varepsilon}_{dm}$ are independent for all $d \in \mathbf{D}_n$ and $m \in \mathbf{M}_n$.

2. The terms $\tilde{\varepsilon}_{dm}$ are independent and identically Gumbel distributed with scale parameter μ^d

3. $\tilde{\varepsilon}_m$ is distributed so that $\max_{d \in \mathbf{D}_{nm}} U_{dm}$ is Gumbel distributed with scale parameter μ^m

Nested Logit Model

124

– Marginal choice probabilities for Nested Logit

$$P_n(m) = Pr\left(\max_{d\in \mathbf{D}_{nm}} U_{dm} \ge \max_{d\in \mathbf{D}_{nm'}} U_{dm'}, \forall m' \in \mathbf{M}_n, m' \neq m\right)$$

$$= Pr\left(\tilde{V}_m + \tilde{\varepsilon}_m + \max_{d\in \mathbf{D}_{nm}} [\tilde{V}_d + \tilde{V}_{dm} + \tilde{\varepsilon}_{dm}]\right.$$

$$\left.\ge \tilde{V}_{m'} + \tilde{\varepsilon}_{m'} + \max_{d\in \mathbf{D}_{nm'}} [\tilde{V}_d + \tilde{V}_{dm'} + \tilde{\varepsilon}_{dm'}], \forall m' \in \mathbf{M}_n, m' \neq m\right)$$

Since $\tilde{\varepsilon}_{dm}$ is by assumption Gumbel distributed with parameter μ^d, the term

$$\max_{d\in \mathbf{D}_{nm}} [\tilde{V}_d + \tilde{V}_{dm} + \tilde{\varepsilon}_{dm}] \qquad (10.34)$$

is also Gumbel distributed, but with parameters

$$\eta = \frac{1}{\mu^d} \ln \sum_{d\in \mathbf{D}_{nm}} e^{(\tilde{V}_d + \tilde{V}_{dm})\mu^d}, \qquad (10.35)$$

Nested Logit Model

$$
\begin{aligned}
P_n(m) &= Pr\left(\max_{d \in \mathbf{D}_{nm}} U_{dm} \geq \max_{d \in \mathbf{D}_{nm'}} U_{dm'}, \forall m' \in \mathbf{M}_n, m' \neq m\right) \\
&= Pr\Big(\tilde{V}_m + \tilde{\varepsilon}_m + \max_{d \in \mathbf{D}_{nm}} [\tilde{V}_d + \tilde{V}_{dm} + \tilde{\varepsilon}_{dm}] \\
&\qquad \geq \tilde{V}_{m'} + \tilde{\varepsilon}_{m'} + \max_{d \in \mathbf{D}_{nm'}} [\tilde{V}_d + \tilde{V}_{dm'} + \tilde{\varepsilon}_{dm'}], \forall m' \in \mathbf{M}_n, m' \neq m\Big)
\end{aligned}
$$

$$
\begin{aligned}
P_n(m) = Pr(&\tilde{V}_m + V'_m + \boxed{\tilde{\varepsilon}_m + \varepsilon'_m} \geq \tilde{V}_{m'} + V'_{m'} + \tilde{\varepsilon}_{m'} \\
&+ \varepsilon'_{m'}, \forall m' \in \mathbf{M}_n, m' \neq m),
\end{aligned}
$$

$$
V'_m = \frac{1}{\mu^d} \ln \sum_{d \in \mathbf{D}_{nm}} e^{(\tilde{V}_d + \tilde{V}_{dm})\mu^d}. \qquad \varepsilon'_m = \max_{d \in \mathbf{D}_{nm}} (\tilde{V}_d + \tilde{V}_{dm} + \tilde{\varepsilon}_{dm}) - V'_m
$$

Nested Logit Model

disturbance $\bar{\varepsilon}_m + \varepsilon'_m$ is by assumption iid Gumbel with a scale parameter μ^m for all $m \in \mathbf{M}_n$, and therefore

$$P_n(m) = \frac{e^{(\tilde{V}_m + V'_m)\mu^m}}{\sum_{m' \in \mathbf{M}_n} e^{(V_{m'} + V'_{m'})\mu^{m'}}}$$

MNL模型

Note that since the scale of $\bar{\varepsilon}_m + \varepsilon'_m$ is μ^m

Then

$$\frac{\mu^m}{\mu^d} \le 1.$$

必须满足

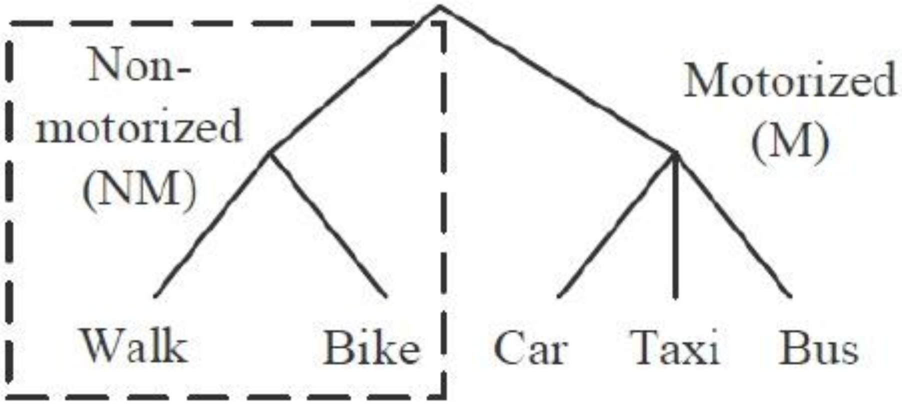
Nested Logit Model

$$
\begin{aligned}
\frac{\mu^m}{\mu^d} &= \left[\frac{\operatorname{var}(\tilde{\varepsilon}_{dm})}{\operatorname{var}(\tilde{\varepsilon}_{dm}) + \operatorname{var}(\tilde{\varepsilon}_m)}\right]^{1/2} \\
&= \left[1 - \frac{\operatorname{var}(\tilde{\varepsilon}_m)}{\operatorname{var}(\tilde{\varepsilon}_m) + \operatorname{var}(\tilde{\varepsilon}_{dm})}\right]^{1/2} \\
&= \left[1 - \frac{\operatorname{cov}(U_{dm}, U_{d'm})}{[\operatorname{var}(U_{dm}) \operatorname{var}(U_{d'm})]^{1/2}}\right]^{1/2} \\
&= \sqrt{1 - \operatorname{corr}(U_{dm}, U_{d'm})}.
\end{aligned}
$$

Dissimilarity parameter

Nested Logit Model

– Example

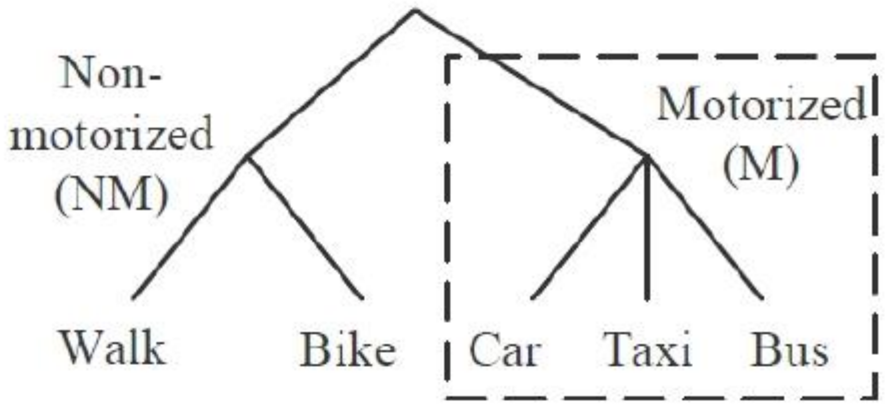


$$P(i \mid NM) = \frac{e^{\mu_{NM} V_i}}{e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}}} \qquad i = Walk, Bike$$

$$I_{NM} = \frac{1}{\mu_{NM}} \ln\left(e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}}\right)$$

Nested Logit Model

– Example

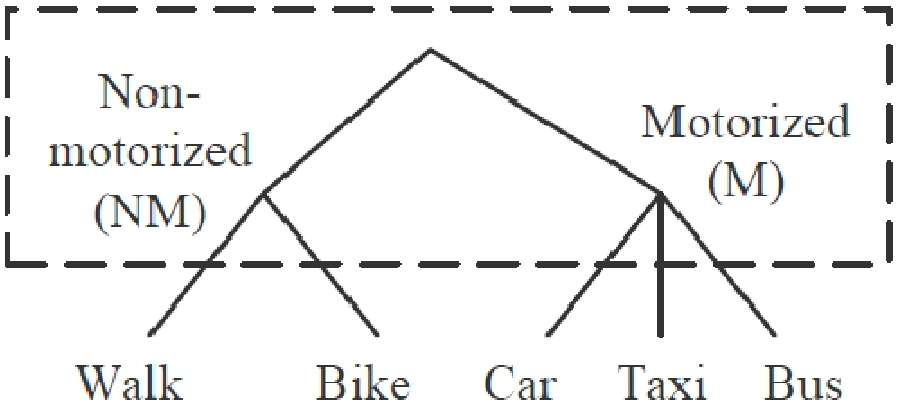


$$\mathrm{P}(i \mid M) = \frac{e^{\mu_M V_i}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}} \qquad i = Car, Taxi, Bus$$

$$I_M = \frac{1}{\mu_M} \ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})$$

Nested Logit Model

– Example



$$P(NM) = \frac{e^{\mu I_{NM}}}{e^{\mu I_{NM}} + e^{\mu I_{M}}}$$

$$P(M) = \frac{e^{\mu I_{M}}}{e^{\mu I_{NM}} + e^{\mu I_{M}}}$$

Nested Logit Model

– Example

- Calculation of choice probabilities

$$P(Bus) = P(Bus|M) \cdot P(M)$$

$$= \left[\frac{e^{\mu_M V_{Bus}}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}}\right] \cdot \left[\frac{e^{\mu I_M}}{e^{\mu I_{NM}} + e^{\mu I_M}}\right]$$

$$= \left[\frac{e^{\mu_M V_{Bus}}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}}\right] \cdot \left[\frac{e^{\frac{\mu}{\mu_M}\ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})}}{e^{\frac{\mu}{\mu_{NM}}\ln(e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}})} + e^{\frac{\mu}{\mu_M}\ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})}}\right]$$

Nested Logit Model

– Procedures of NL model specification

- Partition C_n into M non-overlapping nests:

$$C_{mn} \cap C_{m'n} = \varnothing \quad \forall\, m \neq m'$$

- Deterministic utility term for nest C_{mn}:

$$V_{C_{mn}} = \widetilde{V}_{C_{mn}} + \frac{1}{\mu_m} \ln \sum_{j \in C_{mn}} e^{\mu_m \widetilde{V}_{jn}}$$

- Model: $P(i \mid C_n) = P(C_{mn} \mid C_n) P(i \mid C_{mn}), i \in C_{mn} \subseteq C_n$

  where

$$P(C_{mn} \mid C_n) = \frac{e^{\mu V_{C_{mn}}}}{\sum_l e^{\mu V_{C_{ln}}}} \quad \text{and} \quad P(i \mid C_{mn}) = \frac{e^{\mu_m \widetilde{V}_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m \widetilde{V}_{jn}}}$$

Nested Logit Model

- **Estimation of Nested Logit model**
  - Simultaneous estimation

$$y_{in} = \begin{cases} 1 & \text{if observation } n \text{ chose alternative } i, \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathcal{L}^* = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{y_{in}},$$

Maximizing $\log(\mathcal{L}^*)$ Non-Concave

- Model: $P(i \mid C_n) = P(C_{mn} \mid C_n) P(i \mid C_{mn}), i \in C_{mn} \subseteq C_n$

  where

$$P(C_{mn} \mid C_n) = \frac{e^{\mu V_{C_{mn}}}}{\sum_{l} e^{\mu V_{C_{ln}}}} \quad \text{and} \quad P(i \mid C_{mn}) = \frac{e^{\mu_m \tilde{V}_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m \tilde{V}_{jn}}}$$

Mixed Logit Model

- Random Coefficients
  - The mixed logit probability can be derived from utility-maximizing behavior in several equivalent ways that provide different interpretations.
  - The most straightforward derivation, and most widely used in recent applications, is based on random coefficients.

Random

$$U_{nj} = \underbrace{\boxed{\beta_n'}}_{} x_{nj} + \varepsilon_{nj},$$

Heterogeneity in tastes on observed attributes

  - The coefficients vary over decision makers in the population with density f (β).

Mixed Logit Model

– Since the εnj ’s are iid extreme value, the probability conditional on βn is

$$L_{ni}(\beta_n) = \frac{e^{\beta_n' x_{ni}}}{\sum_j e^{\beta_n' x_{nj}}}.$$

– The Mixed logit probability is

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta)\, d\beta,$$

一般假设β服从正态分布

Mixed Logit Model

- **Error Components**
  - A mixed logit model can be used as simply representing error components that create **correlations among the utilities for different alternatives**

Random

$$U_{nj} = \alpha' x_{nj} + \boxed{\mu'_n} z_{nj} + \varepsilon_{nj},$$

α is a vector of fixed coefficients,
μ is a vector of random terms with zero mean,
and εnj is iid extreme value.

Mixed Logit Model

– Unobserved part of utility is

$$\eta_{nj} = \mu_n' z_{nj} + \varepsilon_{nj},$$

$$\Downarrow$$

$$\mathrm{Cov}(\eta_{ni}, \eta_{nj}) = E(\mu_n' z_{ni} + \varepsilon_{ni})(\mu_n' z_{nj} + \varepsilon_{nj}) = z_{ni}' W z_{nj},$$

***W* is the covariance of *μn*.**

– Error-component and random-coefficient specifications are formally equivalent.