同濟大學
TONGJI UNIVERSITY

多元离散选择模型理论

同济大学
交通运输工程学院
叶 昕 教授

同济大学交通运输工程学院
SCHOOL OF TRANSPORTATION ENGINEERING
TONGJI UNIVERSITY

学习目的

- 了解多元选择模型的推导
- 掌握多元Logit模型的概率和拟合度计算公式
- 掌握判断选择模型中理论不可辨认系数的方法
- 理解多元Logit模型变量设定的常用规则
- 了解基于多元Logit模型的边际效应和弹性分析
- 理解估计方式选择模型的数据采集方法

二元选择推广到三元选择

- 二元选择模型：

$U_1 = V_1 + \varepsilon_1$，$U_2 = V_2 + \varepsilon_2$

根据效用最大化原理：

$P(y = 1) = P(U_1 > U_2)$

$P(y = 2) = P(U_2 > U_1)$

- 推广到三元选择：

$U_1 = V_1 + \varepsilon_1$，$U_2 = V_2 + \varepsilon_2$，$U_3 = V_3 + \varepsilon_3$

根据效用最大化原理：

$P(y = 1) = P(U_1$ 是三个随机效用中最大的$) = P(U_1 > U_2, U_1 > U_3) = P[U_1 > \max(U_2, U_3)]$

Gumbel分布的重要属性之二

- 两个相互独立的服从Gumbel分布且等方差的随机变量的较大值仍然服从Gumbel分布
- 如果$x \sim G(\mu_1, \beta)$, $y \sim G(\mu_2, \beta)$, $z = \max(x, y)$, 那么 $z \sim G\left(\beta \cdot \ln\left[\exp\left(\frac{\mu_1}{\beta}\right) + \exp\left(\frac{\mu_2}{\beta}\right)\right], \beta\right)$

证明过程

$$F_z(z) = P[\max(x,y) < z] = P(x < z, y < z) = P(x < z)P(y < z)$$

$$= F_x(z)F_y(z) = \exp\left[-\exp\frac{-(z-\mu_1)}{\beta}\right]\exp\left[-\exp\frac{-(z-\mu_2)}{\beta}\right]$$

$$= \exp\left[-\exp\left(\frac{\mu_1 - z}{\beta}\right) - \exp\left(\frac{\mu_2 - z}{\beta}\right)\right]$$

$$= \exp\left[-\exp\left(\frac{\mu_1}{\beta}\right)\exp\left(\frac{-z}{\beta}\right) - \exp\left(\frac{\mu_2}{\beta}\right)\exp\left(\frac{-z}{\beta}\right)\right]$$

令 $k = \exp\left(\frac{\mu_1}{\beta}\right) + \exp\left(\frac{\mu_2}{\beta}\right)$,

$$F_z(z) = \exp\left[-k \cdot \exp\left(\frac{-z}{\beta}\right)\right] = \exp\left[-\exp\left(-\frac{z - \beta\ln(k)}{\beta}\right)\right]$$

所以z，即max(x, y)，同样服从Gumbel分布

$$z \sim G\left(\beta \cdot \ln\left[\exp\left(\frac{\mu_1}{\beta}\right) + \exp\left(\frac{\mu_2}{\beta}\right)\right], \beta\right)$$

该属性可推广到多个随机变量

- J个相互独立的服从Gumbel分布且等方差的随机变量的最大值仍然服从Gumbel分布
- 如果$\varepsilon_1 \sim G(\mu_1, \beta)$, $\varepsilon_2 \sim G(\mu_2, \beta)$,…, $\varepsilon_J \sim G(\mu_J, \beta)$, $z = \max(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$, 那么

$$z \sim G\{\beta \cdot \ln\left[\sum_{j=1}^{J} \exp\left(\frac{\mu_j}{\beta}\right)\right], \beta\}$$

多元logit模型的推导 [1]

- Gumbel分布的线性变换：如果$x \sim G(\mu, \beta)$, $\alpha x + \gamma \sim G(\alpha\mu + \gamma, \alpha\beta)$ (其中，α和γ为常数)

- $U_1 = V_1 + \varepsilon_1$，$U_2 = V_2 + \varepsilon_2$,…, $U_J = V_J + \varepsilon_J$

$P(y = 1) = P(U_1 \text{ is the maximum})$

$= P(U_1 > U_2, U_1 > U_3 \ldots, U_1 > U_J)$

$= P[U_1 > \max(U_2, U_3, \ldots, U_J)]$

$U_j = V_j + \varepsilon_j$

Because $\varepsilon_j \sim \mathrm{G}(0,1)$, $U_j \sim \mathrm{G}(V_j, 1)$, $U_1 \sim \mathrm{G}(V_1, 1)$

Let $\mathrm{k} = \ln\left[\sum_{j=2}^{J} \exp(V_j)\right]$

$\max(U_2, \ldots, U_J) \sim G(k, 1)$

Rewrite $\max(U_2, \ldots, U_J)$ as $k + \eta$, where $\eta \sim \mathrm{G}(0, 1)$

多元logit模型的推导 [2]

$$P\left[U_1 > \max(U_2, \ldots, U_J)\right] = P(V_1 + \varepsilon_1 > k + \eta)$$

$$= P(\eta - \varepsilon_1 < V_1 - k); (\eta - \varepsilon_1) \sim \text{Logistic}(0, 1)$$

$$P(y = 1) = P\left[U_1 > \max(U_2, \ldots, U_J)\right]$$

$$= \frac{1}{1+\exp[-(V_1-k)]}$$

$$= \frac{1}{1 + \exp\left(\ln\left[\sum_{j=2}^{J} \exp(V_j)\right] - V_1\right)}$$

$$= \frac{\exp(V_1)}{\exp(V_1) + \left[\sum_{j=2}^{J} \exp(V_j)\right]}$$

$$= \frac{\exp(V_1)}{\sum_{j=1}^{J} \exp(V_j)}$$

多元logit模型的概率计算公式

- 不失一般性，$P(y=k)=\dfrac{\exp(V_k)}{\sum_{j=1}^{J}\exp(V_j)}$
- 因此，二元logit模型是多元logit模型的一个特例
- 类似地，在实践中常将系统部分参数化为解释变量的线性组合：$V_j = X_j\beta_j$

多元logit模型的对数似然函数

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

多元logit模型的极大似然估计特性 [1]

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

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

上式中，当 $k=1$时, $\delta_{ik}=1$; 当$k>1$, $\delta_{ik}=0$

$P_i(1)$代表第i个个体选择第1个选项的概率

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

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

多元logit模型的极大似然估计特性 [2]

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

- 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

多元logit模型的成功应用案例

- **Mcfadden 教授在70年代成功预测BART在通勤市场的份额 (6.3% vs. 6.2%)**

Table 1. Prediction Success Table, Journey-to-Work

(Pre-BART Model and Post-BART Choices)

Cell Counts	Predicted Choices				
Actual Choices	Auto Alone	Carpool	Bus	BART	Total
Auto Alone	255.1	79.1	28.5	15.2	378
Carpool	74.7	37.7	15.7	8.9	137
Bus	12.8	16.5	42.9	4.7	77
BART	9.8	11.1	6.9	11.2	39
Total	352.4	144.5	94.0	40.0	631
Predicted Share	55.8%	22.9%	14.9%	6.3%	
(Std. Error)	*(11.4%)*	*(10.7%)*	*(3.7%)*	*(2.5%)*	
Actual Share	59.9%	21.7%	12.2%	6.2%	

模型的统计推断和拟合度指标

- t检验针对单个模型系数
- 似然比检验针对多个模型系数
- 拟合度指标：

似然比指数（**Likelihood ratio index**）

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

调节似然比指数（**Adjusted likelihood ratio index**）

$$adj.\rho^2(0)=1-\frac{LL(\hat{\beta})-K}{LL(0)} \qquad adj.\rho^2(\mathrm{c})=1-\frac{LL(\hat{\beta})-K}{LL(c)}$$

$$LL(0)=\sum_{i=1}^{n}\ln\left(\frac{1}{J}\right) \qquad LL(c)=\sum_{i=1}^{n}\sum_{j=1}^{J}y_{ij}\ln(s_j) \qquad \text{其中} \quad s_j=\sum_{i=1}^{n}y_{ij}/n$$

为什么？

模型系数的可辨认性 (Identification)

- 模型系数在某些情况下具有不可辨认性 (Unidentification)
  - 理论上的不可辨认: 由于参数向量β中存在多余参数而造成的
  - 实证上的不可辨认: 由于样本量过小，或系数所对应的变量在样本中缺乏变化而造成的
- 系数不可辨认会导致MLE过程不收敛
- 理论上的不可辨认性可由分析概率函数表达式得知

选择模型中理论上不可辨认的系数 [1]

- 二元选择模型中，$V_1 = \beta_0 + X_1\beta_1$，$V_2 = \gamma_0 + X_2\gamma_1$，其中仅有一个常数项可以被辨认。因为，

$$P(y=1) = \frac{\exp(\beta_0 + X_1\beta_1)}{\exp(\beta_0 + X_1\beta_1) + \exp(\gamma_0 + X_2\gamma_1)} = \frac{\exp[(\beta_0 - \gamma_0) + X_1\beta_1]}{\exp[(\beta_0 - \gamma_0) + X_1\beta_1] + \exp(X_2\gamma_1)}$$

- 二元选择模型中，$U_1 = X_1\beta_1 + \varepsilon_1$，$U_2 = X_2\beta_2 + \varepsilon_2$，$\varepsilon_1$和$\varepsilon_2$的标准差$\sigma$是不可辨认的系数。因为，

$$P(y = 1) = P(X_1\beta_1 + \sigma\varepsilon_1 > X_2\beta_2 + \sigma\varepsilon_2) = P[X_1(\beta_1/\sigma) + \varepsilon_1 > X_2(\beta_2/\sigma) + \varepsilon_2]$$

选择模型中理论上不可辨认的系数 [2]

- 多元选择模型中，如果有J个选项对应于J个效用方程，其中一个效用方程中的常数项β_0不可辨认，在MLE过程中需要固定为0。为什么？
- 如以下多元选择模型例子：

$$V_{auto} = \alpha_1 + \beta_1 * AutoTime + \beta_2 * ParkingCost + \gamma_1 * Income$$

$$V_{rail} = \alpha_2 + \beta_1 * RailTime + \gamma_2 * Income$$

$$V_{bus} = \beta_1 * BusTime + \gamma_3 * Income$$

这里，出行时间为一般属性，为所有选项共有；

而停车费为驾驶方式特有属性；

收入为出行者社会经济属性。这里所有的系数都可辨认吗？

选择模型中理论上不可辨认的系数 [3]

- 这里用V'代表效用方程中“收入”以前的部分：

$$P(\text{Auto}) = \frac{\exp(V_{Auto}' + \gamma_1 Income)}{\exp(V_{Auto}' + \gamma_1 Income) + \exp(V_{Rail}' + \gamma_2 Income) + \exp(V_{Bus}' + \gamma_3 Income)}$$

$$= \frac{\exp(V_{Auto}')}{\exp(V_{Auto}') + \exp[V_{Rail}' + (\gamma_2 - \gamma_1) Income] + \exp[V_{Bus}' + (\gamma_3 - \gamma_1) Income]}$$

- 仅能辨认2个系数
- 同一个社会经济属性变量不能同时出现在所有的效用方程中
- 其中一个系数需要被标准化（normalize）为0

选择模型中理论上不可辨认的系数 [4]

- 效用方程中不能设置完美线性相关的两个或者多个变量，否则他们的系数将不可辨认
- 典型的例子：一系列的哑变量不可同时置入一个效用方程
  - 例如：如果哑变量male和female被同时置入方程

$V = \beta_0 + \beta_1 * male + \beta_2 * female = \beta_0 + \beta_1 * male + \beta_2 * (1 - male) = \beta_0 + \beta_2 + (\beta_1 - \beta_2) * male$，因此仅能辨认$(\beta_1 - \beta_2)$，其中一项需要被标准化为0

  - 如果一系列哑变量 LowIncome, MidIncome, HighIncome同时置入：$V = \beta_0 + \beta_1 *$ LowIncome $+ \beta_2 *$ MidIncome $+ \beta_3 *$ HighIncome，会有什么问题？

选择模型的变量设定 (specification) [1]

- 以出行方式选择模型为例：
  - 时间和金钱可以被设定为“一般”变量，公共交通到、离站时间，等候时间为公交方式“一般”变量，出行者个人属性为“选项特定”变量，模型设定如下:

$$V_{auto} = \alpha_1 + \beta_1 * InvehTime_{auto} + \beta_2 * ParkingCost_{auto} + \gamma_1 * Male$$

$$V_{rail} = \alpha_2 + \beta_1 * InvehTime_{rail} + \beta_2 * Fare_{rail} + \beta_3 * AccessTime_{rail} + \beta_4 * WaitTime_{rail} + \gamma_2 * Male$$

$$V_{bus} = \beta_1 * InvehTime_{bus} + \beta_2 * Fare_{bus} + \beta_3 * AccessTime_{bus} + \beta_4 * WaitTime_{rail}$$

  - 出行时间价值可以用β_1 / β_2来表示，从而给时间定价
  - 譬如，$\beta_1 = -0.05$分钟$^{-1}$， $\beta_2 = -0.10$元$^{-1}$ ，

模型给出的出行时间价值就是**0.05/0.10*60 = 30** 元/小时

选择模型的变量设定 (specification) [2]

- “一般”变量的系数可以是不相等的，譬如：
  - 跟私家车行程时间相比，出行者可能更不能忍受公交车内的行程时间
  - 跟在地铁站点等候相比，出行者可能更不能忍受公交车站点的等候时间
- 在模型中可以设定不同的系数值，如下：

$$V_{auto} = \alpha_1 + \beta_1 * InvehTime_{auto} + \beta_2 * ParkingCost_{auto} + \gamma_1 * Male$$

$$V_{rail} = \alpha_2 + \beta_5 * InvehTime_{rail} + \beta_2 * Fare_{rail} + \beta_3 * AccessTime_{rail} + \beta_4 * WaitTime_{rail} + \gamma_2 * Male$$

$$V_{bus} = \beta_6 * InvehTime_{bus} + \beta_2 * Fare_{bus} + \beta_3 * AccessTime_{bus} + \beta_7 * WaitTime_{rail}$$

选择模型的变量设定 (specification) [3]

- 建模过程中的艺术性和科学性
  - 建模的过程表达建模者的主观意图
  - 但模型估计结果需要接受统计检验
- 模型系数需要有合理的正负号
- 在一定情况下（如， 出于特殊研究兴趣，但样本量比较小或变量在样本中变化较小），可保留不显著系数
- 两个在样本中高度相关的解释变量（如，出行时间和距离）同时置入一个效用函数，他们的系数往往不显著或呈现出不合理性；此时应有取舍，保留更有意义的变量
- 没有必要一味追求模型的高拟合度，建模过程中应更关注关键变量系数估计值的合理性，这才是模型具有预测能力的关键

基于选择模型的边际效应分析

$$P_i = \frac{\exp(V_i)}{\sum_{j=1}^{J}\exp(V_j)}$$ ，这里i, j代表选项

如果 $V_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \ldots + \beta_K X_{Ki}$

- 选择模型中系数的解释不像线性回归模型系数的解释那么直观，因为解释变量线性地影响效用值，而非线性地影响选择概率
- 边际效应 （Marginal Effect）

直接效应 = $\frac{\partial P_i}{\partial X_{ki}} = \beta_k \times (P_i) \times (1 - P_i)$

交叉效应 = $\frac{\partial P_j}{\partial X_{ik}} = -\beta_k \times (P_i) \times (P_j) \quad \forall i \neq j$

边际效应总和为0。

$$\begin{aligned}\sum_{\forall j}\frac{\partial P_j}{\partial X_{ik}} &= \frac{\partial P_i}{\partial X_{ik}} + \sum_{\forall j \neq i}\frac{\partial P_j}{\partial X_{ik}} \\ &= \beta_k P_i(1 - P_i) - \sum_{\forall j \neq i}\beta_k P_i P_j \\ &= \beta_k P_i(1 - P_i) - \beta_k P_i \sum_{\forall j \neq i} P_j \\ &= \beta_k P_i(1 - P_i) - \beta_k P_i(1 - P_i) = 0\end{aligned}$$

基于选择模型的弹性分析

- 弹性定义：

$$\text{Elasticity} = \frac{\text{Percentage Change in Probability}}{\text{Percentage Change in Attribute}}$$

$$= \frac{(P_2 - P_1)/P_1}{(X_2 - X_1)/X_1} = \frac{\Delta P / P_1}{\Delta X / X_1}$$

$$\eta_X^P = \frac{\left(\frac{\partial P}{P}\right)}{\left(\frac{\partial X}{X}\right)} = \left(\frac{\partial P}{\partial X}\right) \times \left(\frac{X}{P}\right)$$

直接弹性 = $\eta_{X_{ik}}^{P_i} = \beta_k P_i (1 - P_i) \left(\frac{X_{ik}}{P_i}\right) = \beta_k X_{ik} (1 - P_i)$

交叉弹性 = $\eta_{X_{ik}}^{P_j} = \left(-\beta_k P_i P_j\right) \left(\frac{X_{ik}}{P_j}\right) = -\beta_k X_{ik} P_i$

多元logit模型给出的交叉弹性对于每一个选项来说是相等的。

估计方式选择模型需要采集的数据

- OD之间被选择的出行方式调查
- OD之间可选择出行方式的服务水平
- 出行者的人口及社会经济属性

OD之间被选择的出行方式

- 可以从传统的家庭OD出行调查中提取
  - 通常需要区分出行目的（如通勤、购物、娱乐等）
  - 样本中需要有不同的OD，否则会造成系数在实证上的不可辨认性
- 也可以针对某个出行目的，做专门的出行方式调查

OD之间可选择出行方式的服务水平

- 传统方式：
  - 需要获取整个城市的道路和公交网络数据
  - 通过最短程路径分析，计算每个OD对之间的花费
- 可尝试的新方式：
  - 利用网络地图服务，获得相关OD对之间不同出行方式的花费

出行者的人口及社会经济属性

- 出行者的性别、年龄、学历、职业、收入、家庭状况等变量
- 这个数据是可以选择的
- 但社会人群的差异化正受到越来越多的关注

选择模型在交通行为其他方面的应用

- 长期选择
  - 居住地选择
- 中期选择
  - 交通工具选择
- 短期选择
  - 出行目的地选择
  - 出行时间段选择
  - 出行路径选择

同濟大學
TONGJI UNIVERSITY

谢谢大家！

同济大学交通运输工程学院
SCHOOL OF TRANSPORTATION ENGINEERING
TONGJI UNIVERSITY