

## 第 12 章 多值选择模型

### 12.1 多项 **Logit** 与多项 **Probit**

假设可供个体选择的方案为  $y = 1, 2, \dots, J$ , 其中  $J$  为正整数。

个体  $i$  选择方案  $j$  的效用为

$$U_{ij} = \mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_{ij} \quad (i = 1, \dots, n; \quad j = 1, \dots, J)$$

解释变量  $\mathbf{x}_i$  只随个体  $i$  而变, 不随方案  $j$  而变 (比如, 个体的性

别、年龄、收入等特征), 称为“只随个体而变”(case-specific) 或“不随方案而变”(alternative-invariant)。

系数 $\boldsymbol{\beta}_j$ 表明,  $\mathbf{x}_i$ 对效用 $U_{ij}$ 的作用取决于方案 $j$ 。

个体 $i$ 选择方案 $j$ , 当且仅当方案 $j$ 的效用高于所有其他方案:

$$\begin{aligned} P(y_i = j | \mathbf{x}_i) &= P(U_{ij} \geq U_{ik}, \forall k \neq j) \\ &= P(U_{ik} - U_{ij} \leq 0, \forall k \neq j) \\ &= P(\varepsilon_{ik} - \varepsilon_{ij} \leq \mathbf{x}_i' \boldsymbol{\beta}_j - \mathbf{x}_i' \boldsymbol{\beta}_k, \forall k \neq j) \end{aligned}$$

假设 $\{\varepsilon_{ij}\}$ 为 iid 且服从 I 型极值分布(type I extreme value

distribution), 可证明:

$$P(y_i = j | \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta}_j)}{\sum_{k=1}^J \exp(\mathbf{x}_i' \boldsymbol{\beta}_k)}$$

选择各方案的概率之和为 1, 即  $\sum_{j=1}^J P(y_i = j | \mathbf{x}_i) = 1$ 。

这是二值选择 Logit 模型向多值选择模型的推广。

无法同时识别所有的系数  $\boldsymbol{\beta}_k$ ,  $k = 1, \dots, J$ 。

如将  $\boldsymbol{\beta}_k$  变为  $\boldsymbol{\beta}_k^* = \boldsymbol{\beta}_k + \boldsymbol{\alpha}$  ( $\boldsymbol{\alpha}$  为常数向量), 不影响模型拟合。

通常将某方案(比如, 方案 1)作为“参照方案”(base category), 令相应系数  $\boldsymbol{\beta}_1 = \mathbf{0}$ , 可得

$$P(y_i = j | \mathbf{x}_i) = \begin{cases} \frac{1}{1 + \sum_{k=2}^J \exp(\mathbf{x}_i' \boldsymbol{\beta}_k)} & (j = 1) \\ \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta}_j)}{1 + \sum_{k=2}^J \exp(\mathbf{x}_i' \boldsymbol{\beta}_k)} & (j = 2, \dots, J) \end{cases}$$

“ $j=1$ ”所对应的方案为参照方案。

此模型为“多项 logit”(multinomial logit), 可用 MLE 估计。

个体  $i$  的似然函数为

$$L_i(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J) = \prod_{j=1}^J [\text{P}(y_i = j | \mathbf{x}_i)]^{I(y_i=j)}$$

其对数似然函数为

$$\ln L_i(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J) = \sum_{j=1}^J I(y_i = j) \cdot \ln \text{P}(y_i = j | \mathbf{x}_i)$$

如果假设  $\{\varepsilon_{i1}, \dots, \varepsilon_{iJ}\}$  服从  $J$  维正态分布，则可得“多项 probit” (multinomial probit) 模型，但涉及高维积分，不易计算。

## 12.2 条件 **Logit** 模型

有些解释变量既随个体而变，也随方案而变。

比如，在选择交通工具时，乘车时间既因个体而异(不同个体的行车路线可能不同)，也因交通工具而异(坐公交的时间与乘出租车的时间不同)。

这种解释变量称为“随方案而变”(alternative-specific)，既包括同时随方案与个体而变的变量，也包括随方案而变但不随个体而变的变量(比如，选择加入某协会时，所有人交的会费都一样)。

个体  $i$  选择方案  $j$  的效用:

$$U_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij} \quad (i = 1, \dots, n; \quad j = 1, \dots, J)$$

解释变量  $\mathbf{x}_{ij}$  既随个体  $i$  而变, 也随方案  $j$  而变。

系数  $\boldsymbol{\beta}$  表明,  $\mathbf{x}_{ij}$  对效用  $U_{ij}$  的作用不依赖于方案  $j$ , 比如乘车时间依个体与方案而变, 但乘车时间太长所带来的负效用是一致的。

个体  $i$  选择方案  $j$  的概率:

$$P(y_i = j | \mathbf{x}_{ij}) = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta})}{\sum_{k=1}^J \exp(\mathbf{x}'_{ik}\boldsymbol{\beta})}$$

此模型称为“条件 logit”(conditional logit), 也称为“McFadden 选择模型”(McFadden's choice model)。

条件 Logit 模型的估计方法与多项 Logit 模型类似。

不同的是, 在条件 Logit 模型中, 由于系数  $\boldsymbol{\beta}$  不依赖于方案, 故不需要选择参照方案, 也不需要 will  $\boldsymbol{\beta}$  的某部分标准化为 0。



## 12.3 混合 Logit 模型

上面分别考虑了解释变量不随方案而变的多项 Logit 模型，以及解释变量随方案而变的条件 Logit 模型。

考虑这两种情况同时发生的混合情形。

假设个体  $i$  选择方案  $j$  的效用：

$$U_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma}_j + \varepsilon_{ij} \quad (i = 1, \dots, n; \quad j = 1, \dots, J)$$

解释变量  $\mathbf{x}_{ij}$  既随个体  $i$  而变，也随方案  $j$  而变；解释变量  $\mathbf{z}_i$  只

随个体  $i$  而变。个体  $i$  选择方案  $j$  的概率：

$$P(y_i = j | \mathbf{x}_{ij}, \mathbf{z}_i) = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_i\boldsymbol{\gamma}_j)}{\sum_{k=1}^J \exp(\mathbf{x}'_{ik}\boldsymbol{\beta} + \mathbf{z}'_i\boldsymbol{\gamma}_k)}$$

在文献中称为“混合 Logit”(mixed logit), Stata 仍称为条件 Logit。

为识别该模型，也需选择一个参照方案(比如方案 1)，令  $\boldsymbol{\gamma}_1 = \mathbf{0}$ 。

在多值选择模型中，被解释变量必然为“多项分布”(multinomial distribution)，故一般不必使用稳健标准误，使用普通标准误即可；这一点类似于二值选择模型。

如果数据为聚类样本，则应使用聚类稳健的标准误。

在多项 Logit 与混合 Logit 模型中，对参数估计值  $\hat{\boldsymbol{\beta}}_j$  的解释以参照方案为转移。

以多项 Logit 为例，假设“方案 1”或“方案  $j$ ” ( $j \neq 1$ ) 必然发生(二者必居其一)，则在此条件下，“方案  $j$ ”发生的条件概率为

$$P(y = j | y = 1 \text{ or } j) = \frac{P(y = j)}{P(y = 1) + P(y = j)} = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_j)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_j)}$$

上式与二值 Logit 具有完全相同的形式。

“几率比” (odds ratio)或“相对风险” (relative risk)为

$$\frac{P(y = j)}{P(y = 1)} = \exp(\mathbf{x}'_i \boldsymbol{\beta}_j)$$

对数几率比(log-odds ratio)为

$$\ln \left[ \frac{P(y = j)}{P(y = 1)} \right] = \mathbf{x}'_i \boldsymbol{\beta}_j$$

该条件概率不依赖于任何其他方案。如果将多值选择模型中的任何两个方案单独挑出来，都是二值 logit 模型。

此假定称为“无关方案的独立性”(Independence of Irrelevant Alternatives, 简记 IIA)。

类似地, 条件 logit 模型也服从 IIA 假定。

在实践中, 如果不同方案之间很类似, 则 IIA 假设不一定满足。这是多项 Logit、条件 logit 与混合 Logit 模型的共同缺点。

比如, 假设共有三个备选的交通方式, 即自驾车、乘红色公共汽车(red bus)、乘蓝色公共汽车(blue bus)。

根据“无关选择的独立性”假定, 在给定选择“自驾车”或“乘红色公共汽车”条件下, 选择“自驾车”的条件概率与是否存在

“乘蓝色公共汽车”的选择无关。

由于“蓝色公共汽车”与“红色公共汽车”仅颜色不同，故加上“蓝色公共汽车”的选择之后，对“自驾车”概率没有影响，但将使“红色公共汽车”概率降低一半。

因此，引入“蓝色公共汽车”后，上述条件概率将增大，与“无关选择的独立性”假定相矛盾。称为“红车蓝车问题”(red bus-blue bus problem)。

对于 IIA 假定，检验方法之一为豪斯曼检验。

如果 IIA 假定成立，则去掉某个方案不影响对其他方案参数的一致估计，只是降低效率。

故在 IIA 成立情况下，去掉某个方案后子样本的系数估计值(记为  $\hat{\beta}_R$ )与全样本的系数估计值(记为  $\hat{\beta}_F$ )没有系统差别。

Hausman and McFadden (1984)提出统计量：

$$(\hat{\beta}_R - \hat{\beta}_F)' \left[ \widehat{\text{Var}(\hat{\beta}_R)} - \widehat{\text{Var}(\hat{\beta}_F)} \right]^{-1} (\hat{\beta}_R - \hat{\beta}_F)' \xrightarrow{d} \chi^2(m)$$

其中， $m$  等于  $\hat{\beta}_R$  的维度。

检验 IIA 假定的方法之二为 Small and Hsiao (1985)所提出。

Cheng and Long (2007)通过蒙特卡罗方法发现，这两个检验的小样本性质都不好，结论只有参考价值)。

## 12.4 嵌套 Logit

多项 Logit、条件 Logit 与混合 Logit 都须满足 IIA 假定。在实践中，如果方案比较类似，则 IIA 假设可能不满足。



解决方法之一是，把相似的方案归入一组，允许同组内的方案相关，但不同组的方案相互独立。

比如，在交通工具的选择上，可把“公交车、地铁”归入“公共交通组”，而把“自驾车、出租车”归入“私人交通组”，形成嵌套式(nested)的树形结构。

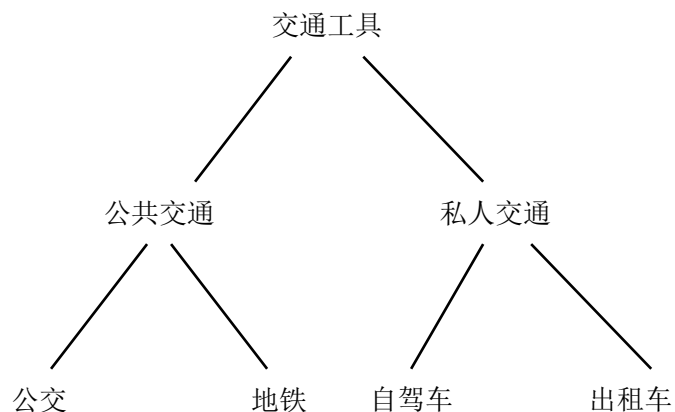


图 12.1 嵌套 Logit 的树形结构

图 12.1 中的“交通工具”可视为此树形结构的“根部”(root), 第一层(level 1)的“公共交通、私人交通”为“树干”(limb), 而第二层(level 2)的“公交、地铁、自驾车、出租车”为“树枝”(branch)。

可将此树形结构视为“决策树”(decision tree), 但关键在于它的统计性质, 即在第二层的每组内部允许扰动项相关, 比如  $(\varepsilon_{i, bus}, \varepsilon_{i, subway})$  相关,  $(\varepsilon_{i, car}, \varepsilon_{i, taxi})$  相关; 但这两组之间不相关, 即  $(\varepsilon_{i, bus}, \varepsilon_{i, subway})$  与  $(\varepsilon_{i, car}, \varepsilon_{i, taxi})$  不相关。

如果所有方案的扰动项都互不相关, 则回到多项 logit 或条件 logit 的情形。

更一般地,假设在第一层共有  $J$  个树干备选,而树干  $j$  包含  $K_j$  个树枝(不同树干包含的树枝数可以不同),分别以  $j1, \dots, jK_j$  来指代。个体  $i$  选择树干  $j$  树枝  $k$  方案的效用为(略去个体  $i$  下标)

$$U_{jk} = \mathbf{x}'_{jk} \boldsymbol{\beta} + \mathbf{z}'_j \boldsymbol{\gamma}_j + \varepsilon_{jk} \quad (j = 1, \dots, J; k = 1, \dots, K)$$

其中,解释变量  $\mathbf{z}_j$  只随树干而不随树枝方案而变(故其系数  $\boldsymbol{\gamma}_j$  可随  $j$  而变),而  $\mathbf{x}_{jk}$  同时随树干与树枝方案而变(故其系数  $\boldsymbol{\beta}$  没有下标,不随  $j, k$  而变)。

McFadden (1978) 假定扰动项  $\varepsilon_{jk}$  服从“广义极值分布” (Generalized Extreme Value distribution, 简记 GEV), 其 cdf 为

$$F(\boldsymbol{\varepsilon}) = \exp \left[ -G(e^{-\varepsilon_{11}}, \dots, e^{-\varepsilon_{1K_1}}; \dots; e^{-\varepsilon_{J1}}, \dots, e^{-\varepsilon_{JK_J}}) \right]$$

其中, 函数  $G(\cdot)$  的形式为

$$G(\boldsymbol{w}) = G(w_{11}, \dots, w_{1K_1}; \dots; w_{J1}, \dots, w_{JK_J}) = \sum_{j=1}^J \left( \sum_{k=1}^{K_j} w_{jk}^{1/\tau_j} \right)^{\tau_j}$$

上式右边的  $\left( \sum_{k=1}^{K_j} w_{jk}^{1/\tau_j} \right)^{\tau_j}$ , 在形式上为不变替代弹性函数

(Constant Elasticity of Substitution, CES)。

参数  $\tau_j = \sqrt{1 - \text{Corr}(\varepsilon_{jk}, \varepsilon_{jl})}$ ，与  $(\varepsilon_{jk}, \varepsilon_{jl})$  的相关系数呈反向变动的关系，称为“不相似参数” (dissimilarity parameter)，在 Stata 中记为 tau。  $0 \leq \tau_j \leq 1$ 。

如果  $\tau_j = 1$  ( $j = 1, \dots, J$ )，则所有方案的扰动项均不相关，满足 IIA 假定，回到多项 logit、条件 logit 或混合 logit 的情形。

选择树干  $j$  树枝  $k$  方案的概率为：

$$p_{jk} = p_j \times p_{k|j} = \frac{\exp(\mathbf{z}'_j \boldsymbol{\gamma} + \tau_j I_j)}{\sum_{m=1}^J \exp(\mathbf{z}'_m \boldsymbol{\gamma} + \tau_m I_m)} \times \frac{\exp(\mathbf{x}'_{jk} \boldsymbol{\beta}_j / \tau_j)}{\sum_{l=1}^{K_j} \exp(\mathbf{x}'_{jl} \boldsymbol{\beta}_j / \tau_j)}$$

其中， $p_j$  为选择树干  $j$  的概率， $p_{k|j}$  为在选择树干  $j$  的情况下，选择树枝  $k$  的条件概率；而  $I_j$  的定义为

$$I_j \equiv \ln \left[ \sum_{l=1}^{K_j} \exp(\mathbf{x}'_{jl} \boldsymbol{\beta}_j / \tau_j) \right]$$

其中， $I_j$  称为“包含价值”(inclusive value)或“对数和”(log-sum)。

根据此式写出整个样本的对数似然函数，称为“全信息最大似然估计”(Full Information Maximum Likelihood，简记 FIML)。

另一方法为分步对  $p_j$  与  $p_{k|j}$  进行 MLE 估计，称为“有限信息最大似然估计”(Limited Information Maximum Likelihood，简记 LIML)，但不如 FIML 有效率。

进行 FIML 估计后，可对联合假设“ $H_0: \tau_1 = \dots = \tau_J = 1$ ”进行似然比检验。

如果接受原假设，则 IIA 假定成立，不需要使用嵌套 logit，可直接使用多项 logit、条件 logit 或混合 logit 模型。