



同濟大學
TONGJI UNIVERSITY

二元离散选择模型的估计

同濟大學
交通運輸工程學院
叶昕 教授





- 理解模型参数估计的基本概念
- 理解极大似然估计的思想
- 了解获得极大似然估计的数值方法
- 理解极大似然估计值的统计特性
- 理解极大似然法估计模型的统计推断
- 掌握应用统计软件**SPSS**估计二元**logit**模型及汇总模型输出结果



- 如何估计**Logistic**回归模型中的变量系数？
- 参数估计：根据从总体中抽取的样本估计总体分布中包含的未知参数
- 常用的参数估计方法：
 - 最小二乘法
 - 极大似然法
 - 矩估计法
 - 贝叶斯估计法

如何获得Logistic回归模型的参数?



- 在线性回归模型中，可以使用最小二乘法，估计模型参数

➤ 系数估计值: $\hat{\beta} = (X'X)^{-1}X'y$

- Logistic回归模型给出选择的概率值，可以利用极大似然法进行模型参数估计

极大似然法 (Maximum Likelihood Estimate, MLE)



- 如果 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 是一系列从总体中抽取的样本
- 总体符合某概率分布，其概率分布或密度函数可表达为 $f(\mathbf{x}, \beta)$
- 似然函数: $L(\beta) = \prod_{i=1}^n f(x_i, \beta)$
- 对数似然函数: $LL(\beta) = \ln[\prod_{i=1}^n f(x_i, \beta)] = \sum_{i=1}^n \ln[f(x_i, \beta)]$
- 极大似然估计值: $\hat{\beta} = \operatorname{argmax}[LL(\beta)]$

MLE例子1 (连续变量, 有收敛形式)



- If x_1, x_2, \dots, x_n follow the same distribution
- Probability density function:
$$f(x) = (\theta + 1)x^\theta, 0 < x < 1;$$
$$f(x) = 0, \text{ otherwise.}$$
- $L(\theta) = \prod_{i=1}^n [(\theta + 1) x_i^\theta] = (\theta + 1)^n (\prod_{i=1}^n x_i)^\theta$
- $LL(\theta) = n \ln(\theta + 1) + \theta \sum_{i=1}^n \ln(x_i)$
- To maximize $LL(\theta)$, let $d[LL(\theta)]/d\theta = n/(\theta+1) + \sum_{i=1}^n \ln(x_i) = 0$
- $\hat{\theta} = -1 - \frac{n}{\sum_{i=1}^n \ln(x_i)}$

MLE例子2 (离散变量, 有收敛形式)



- If x_1, x_2, \dots, x_n follow Poisson distribution
- Probability function:
 $P(x) = \exp(-\lambda) \cdot \lambda^x / x!$, x is non-negative integer.
- $LL(\lambda) = \sum_{i=1}^n [-\lambda + x_i \ln(\lambda) - \ln(x_i!)]$
 $= -\lambda n + \sum_{i=1}^n [x_i \ln(\lambda) - \ln(x_i!)]$
- To maximize $LL(\lambda)$, let $d[LL(\lambda)]/d\lambda = -n + (1/\lambda) \cdot \sum_{i=1}^n x_i = 0$
- $\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$



MLE例子3 (无收敛形式)

- If x_1, x_2, \dots, x_n follow Weibull distribution
- Probability density function: $f(x) = k \cdot x^{k-1} \cdot \exp(-x^k)$
- $L(k) = \prod_{i=1}^n [k \cdot x_i^{k-1} \cdot \exp(-x_i^k)]$
- $LL(k) = \sum_{i=1}^n [\ln(k) + (k-1) \ln(x_i) - x_i^k]$
 $= n \cdot \ln(k) + (k-1) \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n x_i^k$
- To maximize $LL(k)$, let $d[LL(k)]/dk = n/k + \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \ln(x_i) x_i^k = 0$
- There is no closed form for \hat{k}

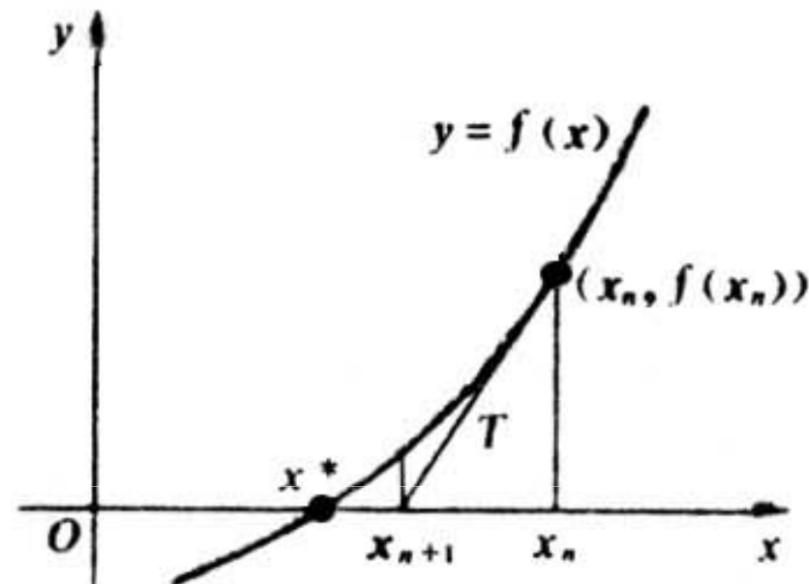


- 牛顿法 (Newton's method)
 - 切线法
- 拟牛顿法 (Quasi-Newton methods)
 - 割线法

牛顿法 (Newton's method)



- 从初始值 x_0 开始, 利用切线不断逼近解 x^*
- 迭代公式:



$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \quad n = 0, 1, \dots$$

- 缺点: 每次迭代都需要计算二阶导数, 比较耗时

拟牛顿法 (Quasi-Newton methods)

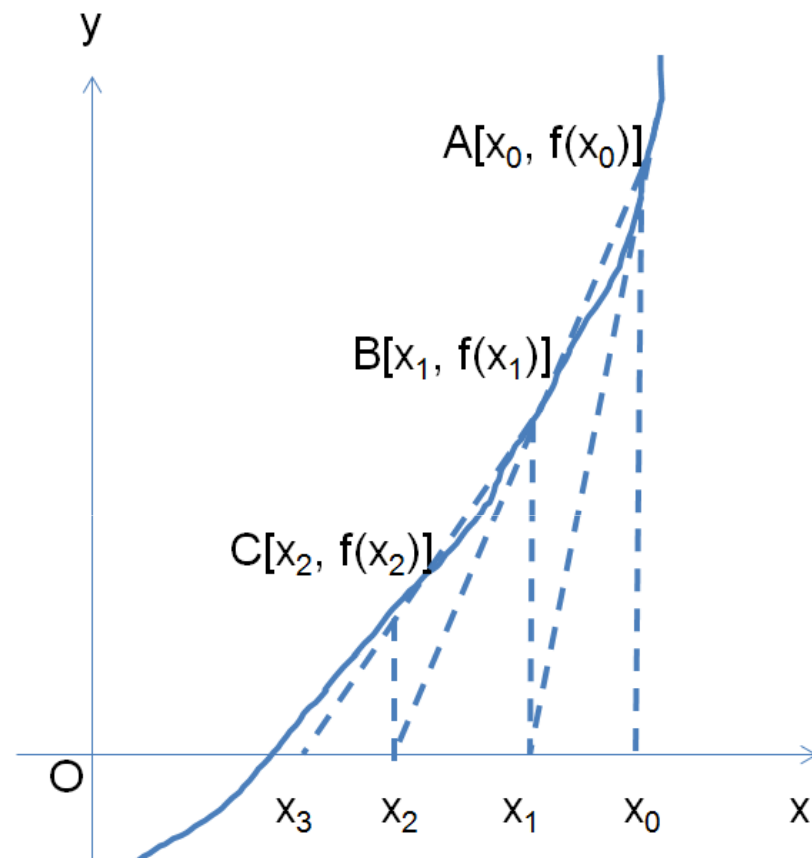


■ 割线法

- 从初始点 x_0 开始, 利用切线法先找到 x_1
- 利用两点之间的割线逼近解 x^*
- 割线方程: $[Y - f(x_1)]/(X-x_1) = [f(x_1) - f(x_0)]/(x_1-x_0)$

- ## ■ 迭代公式:
- $$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}$$
- $$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f'(x_n) - f'(x_{n-1})} f'(x_n), n = 1, 2, \dots$$

- ## ■ 仅计算二阶导数一次



牛顿法在多维空间的推广



- 在多维空间中，二阶导数对应海森矩阵：

$$H(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- 在多维空间中，一阶导数对应梯度：

$$g(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

- 迭代公式： $\mathbf{x}_{n+1} = \mathbf{x}_n - [H(\mathbf{x}_n)]^{-1}g(\mathbf{x}_n)$,

$$n = 0, 1, \dots$$

拟牛顿法在多维空间的推广



- **Broyden–Fletcher–Goldfarb–Shanno (BFGS) 算法**
- 可视为割线法在多维空间的推广，广泛应用于求解无约束的最优化问题
- 在极大似然估计过程中，海森矩阵通常仅计算两次（在开始和结束时各计算一次）

一阶和二阶导数的数值法逼近



- 很多计算机软件（如**Matlab**，**R**，**Gauss**等）可以根据用户提供的似然函数，给出极大似然估计值
- 利用数值法逼近一阶和二阶导数
- $f'(x) \approx [f(x+h) - f(x-h)]/(2h)$
- $f''(x) \approx [f(x-h) + f(x+h) - 2f(x)]/(h^2)$
- 由于计算机存储空间有限，**h**并不是越小越好
- 需要做多次尝试，获得最精确的逼近值
- 提供导数的解析表达式可以减少计算时耗

极大似然估计值的统计特性



- 一致性 (Consistency)

- 当样本量 $n \rightarrow +\infty$, $\hat{\beta} \rightarrow \beta$

- 有效性 (Efficiency)

- 在所有的无偏估计中具有最小的方差

- 渐近正态性 (Asymptotic Normality)

- $\hat{\beta} \sim N[\beta, V]$

- $V \approx -[H(\hat{\beta})]^{-1} = -\left[\frac{\partial^2 LL(\hat{\beta})}{\partial \hat{\beta}^2}\right]^{-1}$

极大似然法估计模型的统计推断



- **t检验推断单个估计值是否等于某常数c**
 - $t\text{-test} = |\hat{\beta} - c| / \text{std}(\hat{\beta})$ ，符合t分布
 - 当样本量n足够大(即自由度足够大时), t分布接近于正态分布
 - **p-值 (显著性水平) = $2 \times [1 - \Phi(t\text{-test})]$**
- **似然比检验两个嵌套模型**
 - **$2 \times (LL_U - LL_R) \sim \text{Chi}^2(\text{DF})$**
 - **LL_U** : 不受约束模型收敛时的对数极大似然值
 - **LL_R** : 受约束模型收敛时的对数极大似然值
 - **DF**: 两个模型之间的参数数量差

模型拟合度指标



- 似然比指数 (Likelihood ratio index)

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

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

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

- **LL(0)**: 模型无参数时的对数极大似然值
- **LL(c)**: 模型仅有常数参数时的对数极大似然值
- $LL(\hat{\beta})$: 模型收敛时的对数极大似然值
- **K**: 两个模型之间的参数数量差

二元logit模型的MLE



- 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_i 和选择变量 y_i ($i = 1, 2, \dots, n$)
- 根据二元logit模型：

$$P(y_i = 1) = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \quad P(y_i = 0) = \frac{1}{1 + \exp(X_i\beta)}$$

- 似然函数：

$$L(\beta) = \prod_{i=1}^n \left\{ \left[\frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \right]^{y_i} \cdot \left[\frac{1}{1 + \exp(X_i\beta)} \right]^{1-y_i} \right\}$$

- 对数似然函数：

$$LL(\beta) = \sum_{i=1}^n \left\{ y_i \ln \left[\frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \right] + (1 - y_i) \ln \left[\frac{1}{1 + \exp(X_i\beta)} \right] \right\}, \quad \hat{\beta} = \operatorname{argmax}[LL(\beta)]$$

二元logit模型的MLE例子



样本数据如下：

Observation Number	Auto Time	Transit Time	Chosen Alternative	Numeric Code
1	52.9	4.4	Transit	1
2	4.1	28.5	Transit	1
3	4.1	86.9	Auto	0
4	56.2	31.6	Transit	1
5	51.8	20.2	Transit	1
6	0.2	91.2	Auto	0
7	27.6	79.7	Auto	0
8	89.9	2.2	Transit	1
9	41.5	24.5	Transit	1
10	95.0	43.5	Transit	1
11	99.1	8.4	Transit	1
12	18.5	84.0	Auto	0
13	82.0	38.0	Auto	0
14	8.6	1.6	Transit	1
15	22.5	74.1	Auto	0
16	51.4	83.8	Auto	0
17	81.0	19.2	Transit	1
18	51.0	85.0	Auto	0
19	62.2	90.1	Auto	0
20	95.1	22.2	Transit	1
21	41.6	91.5	Auto	0

使用SPSS估计二元logit模型 [1]



- 打开SPSS，键入数据
- SPSS提供logistics回归模型的估计功能

- 二元logit模型:

$$V_{\text{auto}} = \beta_{\text{auto}} + \beta * \text{Time}_{\text{auto}}$$

$$V_{\text{transit}} = \beta * \text{Time}_{\text{transit}}$$

- 对应的logistic回归模型:

$$V_{\text{auto}} = \beta_{\text{auto}} + \beta * (\text{Time}_{\text{auto}} - \text{Time}_{\text{transit}})$$

- 需要计算“驾驶”选择变量“AutoChoice”和时间差“TimeDiff”

The screenshot shows the SPSS data editor window for a file named *data_2. The menu bar includes options like 文件(E), 编辑(E), 视图(V), 数据(D), 转换(T), 分析(A), 直销(M), 图形(G), and 实用程序. The toolbar contains icons for file operations and data manipulation. The data grid shows 22 rows of data with the following columns: ID, AutoTime, TransitTime, and TransitChoice. The first row is selected, and the value 48.50 is visible in the ID column.

ID	AutoTime	TransitTime	TransitChoice
1	52.90	4.40	1.00
2	4.10	28.50	1.00
3	4.10	86.90	.00
4	56.20	31.60	1.00
5	51.80	20.20	1.00
6	.20	91.20	.00
7	27.60	79.70	.00
8	89.90	2.20	1.00
9	41.50	24.50	1.00
10	95.00	43.50	1.00
11	99.10	8.40	1.00
12	18.50	84.00	.00
13	82.00	38.00	.00
14	8.60	1.60	1.00
15	22.50	74.10	.00
16	51.40	83.80	.00
17	81.00	19.20	1.00
18	51.00	85.00	.00
19	62.20	90.10	.00
20	95.10	22.20	1.00
21	41.60	91.50	.00
22			

使用SPSS估计二元logit模型 [2]



*data_21.sav [数据集0] - IBM SPSS St

文件(F) 编辑(E) 视图(V) 数据(D) 转换(T) 分析(A) 直销(M) 图形(G) 实用程序(U) 窗口(W) 帮助

计算变量(C)...
对个案内的值计数(O)...
转换值(F)...
重新编码为相同变量(S)...
重新编码为不同变量(R)...

ID	Auto	e	AutoChoice	TT_DIFF
1	1.00	1.00	.00	48.50
2	2.00	1.00	.00	24.40

TransitTime TransitChoice AutoChoice TT_DIFF 变量 变量 变量 变量

计算变量

目标变量(T): AutoChoice = 数字表达式(E): 1 - TransitChoice

类型与标签(L)...

ID
AutoTime
TransitTime

函数组(G):

计算变量

目标变量(T): TimeDiff = 数字表达式(E): AutoTime - TransitTime

类型与标签(L)...

ID
AutoTime

函数组(G):

使用SPSS估计logistic回归模型 [3]



The screenshot shows the SPSS main window with a data view of 21 cases. The 'AutoChoice' variable is highlighted in the menu, and the '二元 Logistic...' option is selected.

ID	AutoTime	hoice	AutoChoice	TimeDiff	变量
1	1.00	52.90	1.00	.00	48.50
2	2.00	4.10	1.00	.00	-24.40
3	3.00	4.10	.00	1.00	-82.80
4	4.00	56.20	1.00	.00	24.60
5	5.00	51.80			31.60
6	6.00	.20			-91.00
7	7.00	27.60			-52.10
8	8.00	89.90			87.70
9	9.00	41.50			17.00
10	10.00	95.00			51.50
11	11.00	99.10			90.70
12	12.00	18.50			-65.50
13	13.00	82.00			44.00
14	14.00	8.60			7.00
15	15.00	22.50			-51.60
16	16.00	51.40			-32.40
17	17.00	81.00			61.80
18	18.00	51.00			-34.00
19	19.00	62.20			-27.90
20	20.00	95.10			72.90
21	21.00	41.60			61.80
22					-34.00

将选择变量
“AutoChoice”置入
为因变量

The screenshot shows the 'Logistic 回归' dialog box. The '因变量(D):' field contains 'AutoChoice'. The '协变量(C):' field contains 'TimeDiff'. The '方法(M):' dropdown is set to '进入'. The '选择变量(B):' field is empty.

将解释变量
“TimeDiff”置入
为协变量

SPSS 中 logistic 回归模型的输出结果



$$-2 \times LL(\hat{\beta})$$

模型汇总

	-2 对数似然值	Cox & Snell R 方	Nagelkerke R 方
1	12.332 ^a	.549	.733

a. 因为参数估计的更改范围小于 .001，所以估计在迭代次数 6 处终止。

MLE 过程中的迭代次数

时间变量系数估计值

方程中的变量

	B	S.E.	Wals	df	Sig.	Exp (B)
步骤 1 ^a TimeDiff	-.053	.021	6.620	1	.010	.948
常量	-.238	.750	.100	1	.752	.789

a. 在步骤 1 中输入的变量: TimeDiff.

常数项估计值

系数的显著性水平，一般小于 0.05 认为系数显著地不等于 0

似然比检验对于模型的整体评价



■ 似然比检验两个嵌套模型

- $2 \times (LL_U - LL_R) \sim \text{Chi}^2(\text{DF})$
- LL_U : 不受约束模型收敛时的对数极大似然值
- LL_R : 受约束模型收敛时的对数极大似然值
- DF : 两个模型之间的参数数量差

■ 在本例中:

- $LL(\beta) = -12.332/2 = -6.166$
- $LL(0) = 21 \cdot \ln(0.5) = -14.556$
- $LL(c) = 10 \cdot \ln(0.4762) + 11 \cdot \ln(1-0.4762) = -14.532$
- $\text{DF}(0) = 2, \text{DF}(c) = 1$
- $\text{Chi}^2\text{-Test}(0) = 2 \cdot (-6.166 + 14.556) = 16.780 > 5.99$ (模型显著)
- $\text{Chi}^2\text{-Test}(c) = 2 \cdot (-6.166 + 14.532) = 16.732 > 3.84$ (模型显著)

模型拟合度指标计算



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

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

■ 在本例中：

$$\rho^2(0) = 1 - \frac{-6.166}{-14.566} = 0.576 \quad \rho^2(c) = 1 - \frac{-6.166}{-14.532} = 0.576$$

$$adj.\rho^2(0) = 1 - \frac{-6.166 - 2}{-14.566} = 0.439 \quad adj.\rho^2(c) = 1 - \frac{-6.166 - 1}{-14.532} = 0.507$$

模型估计结果汇总



Variable	Variable Name	Coefficient Estimate	Asymptotic Std Error	t-statistic
1	Auto Constant	-0.2375	0.7505	-0.32
2	Travel Time	-0.0531	0.0206	-2.57

Summary Statistics

Number of Observations=21

Number of Cases=21

$L(0) = -14.556$

$L(c) = -14.532$

$L(\hat{\beta}) = -6.166$

$-2[L(0) - L(\hat{\beta})] = 16.780$

$-2[L(c) - L(\hat{\beta})] = 16.732$

$\rho^2 = 0.576$

$\hat{\rho}^2 = 0.439$



同濟大學
TONGJI UNIVERSITY

谢谢大家!

