第 11 章　二值选择模型

11.1　二值选择模型

有时被解释变量 y 是离散的，而非连续的，称为**离散选择模型**(discrete choice model)或**定性反应模型**(qualitative response model)。

最常见的离散选择模型是**二值选择行为**(binary choices)。

比如：考研或不考研；就业或待业；买房或不买房；买保险或不买保险；贷款申请被批准或拒绝；出国或不出国；回国或不回

国；战争或和平；生或死。

由于被解释变量为虚拟变量，取值为 0 或 1，故通常不宜进行 OLS 回归。

假设个体只有两种选择，比如$y=1$(考研)或$y=0$(不考研)。

最简单的建模方法为**线性概率模型**(Linear Probability Model，简记 LPM)：

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i \quad (i = 1, \cdots, n) \tag{11.1}$$

其中，解释变量$\boldsymbol{x}_i \equiv (x_{i1}\ x_{i2}\ \cdots\ x_{iK})'$，而参数$\boldsymbol{\beta} \equiv (\beta_1\ \beta_2\ \cdots\ \beta_K)'$

线性概率模型的优点是，计算方便(y为虚拟变量并不影响 OLS 估计)，且容易得到边际效应(即回归系数)。

缺点是，明知被解释变量y的取值非 0 即 1，但根据线性概率模型所作的预测值却可能出现$\hat{y}>1$或$\hat{y}<0$的不现实情形，参见图 11.1。

故一般只将 LPM 作为粗略的参考。

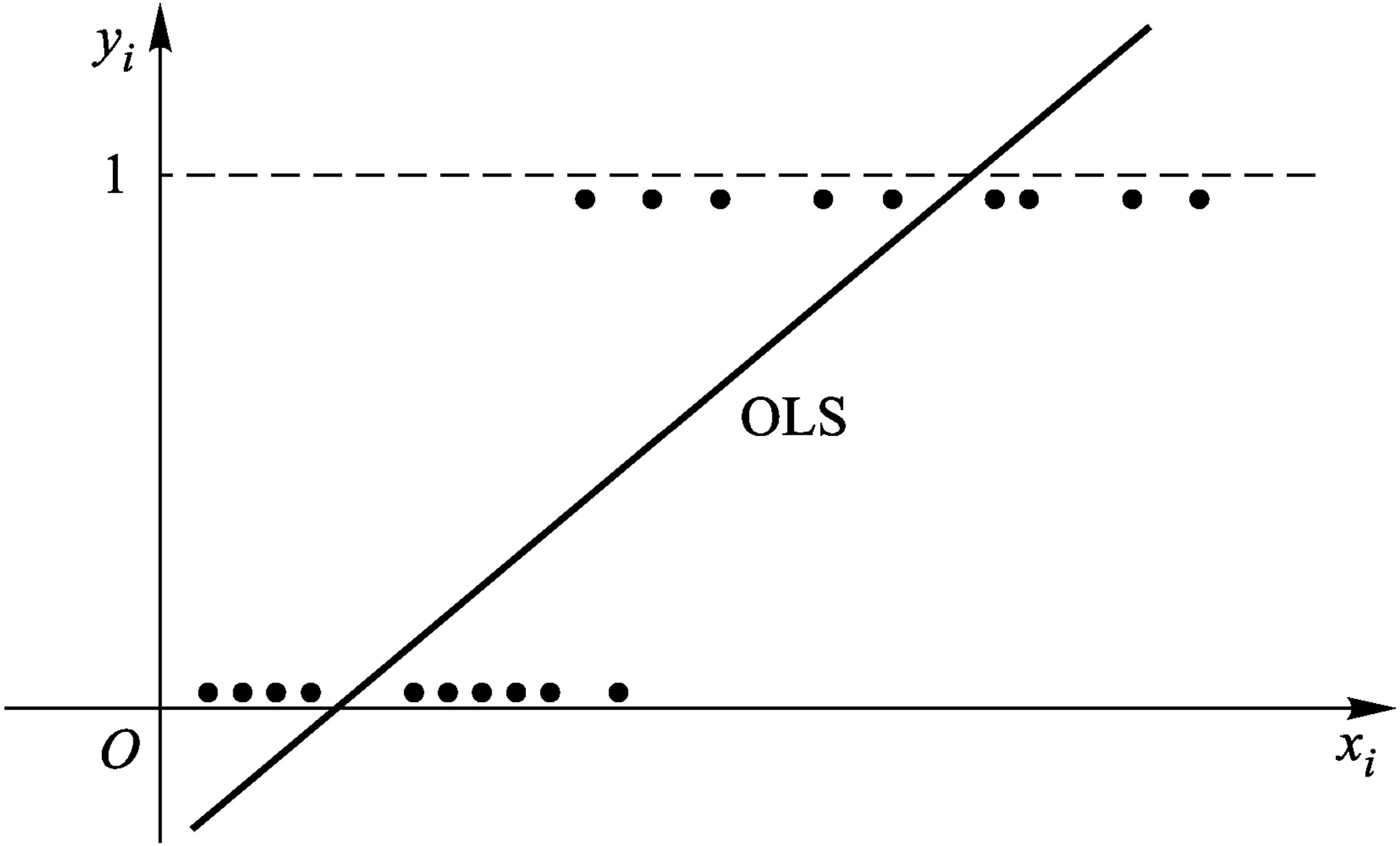

图 11.1　线性概率模型

为使y的预测值总是介于$[0,1]$之间，在给定$\boldsymbol{x}$的情况下，考虑y的两点分布概率：

$$\begin{cases} \mathrm{P}(y=1\,|\,\boldsymbol{x}) = F(\boldsymbol{x},\boldsymbol{\beta}) \\ \mathrm{P}(y=0\,|\,\boldsymbol{x}) = 1-F(\boldsymbol{x},\boldsymbol{\beta}) \end{cases} \tag{11.2}$$

函数$F(\mathbf{x},\boldsymbol{\beta})$称为**连接函数**(link function)，它将解释变量$\mathbf{x}$与被解释变量y连接起来。

由于y的取值要么为0，要么为1，故y肯定服从两点分布。

连接函数的选择具有一定的灵活性。

通过选择合适的连接函数$F(\boldsymbol{x},\boldsymbol{\beta})$(比如，某随机变量的累积分布函数)，可以保证$0 \le \hat{y} \le 1$

可将 $\hat{y}$ 理解为"$y=1$"发生的条件概率，因为

$$\mathrm{E}(y \mid \boldsymbol{x}) = 1 \cdot \mathrm{P}(y=1 \mid \boldsymbol{x}) + 0 \cdot \mathrm{P}(y=0 \mid \boldsymbol{x}) = \mathrm{P}(y=1 \mid \boldsymbol{x}) \quad (11.3)$$

如果 $F(\boldsymbol{x}, \boldsymbol{\beta})$ 为标准正态的累积分布函数(cdf)，则

$$\mathrm{P}(y=1 \mid \boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{\beta}) = \Phi(\boldsymbol{x}'\boldsymbol{\beta}) \equiv \int_{-\infty}^{\boldsymbol{x}'\boldsymbol{\beta}} \phi(t)\, dt \quad (11.4)$$

其中，$\phi(\cdot)$ 与 $\Phi(\cdot)$ 分别为标准正态的密度函数与累积分布函数。

此模型称为 Probit。

如果$F(\boldsymbol{x}, \boldsymbol{\beta})$为“逻辑分布”(logistic distribution)的累积分布函数，则

$$\mathrm{P}(y=1 \mid \boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{\beta}) = \Lambda(\boldsymbol{x}'\boldsymbol{\beta}) \equiv \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}'\boldsymbol{\beta})} \qquad (11.5)$$

其中，函数$\Lambda(\cdot)$的定义为$\Lambda(z) \equiv \dfrac{\exp(z)}{1+\exp(z)}$。

此模型称为 Logit。

逻辑分布的密度函数关于纵轴对称，期望为 0，方差为$\pi^2/3$(大于标准正态的方差)，具有厚尾(fat tails)，参见图 11.2。

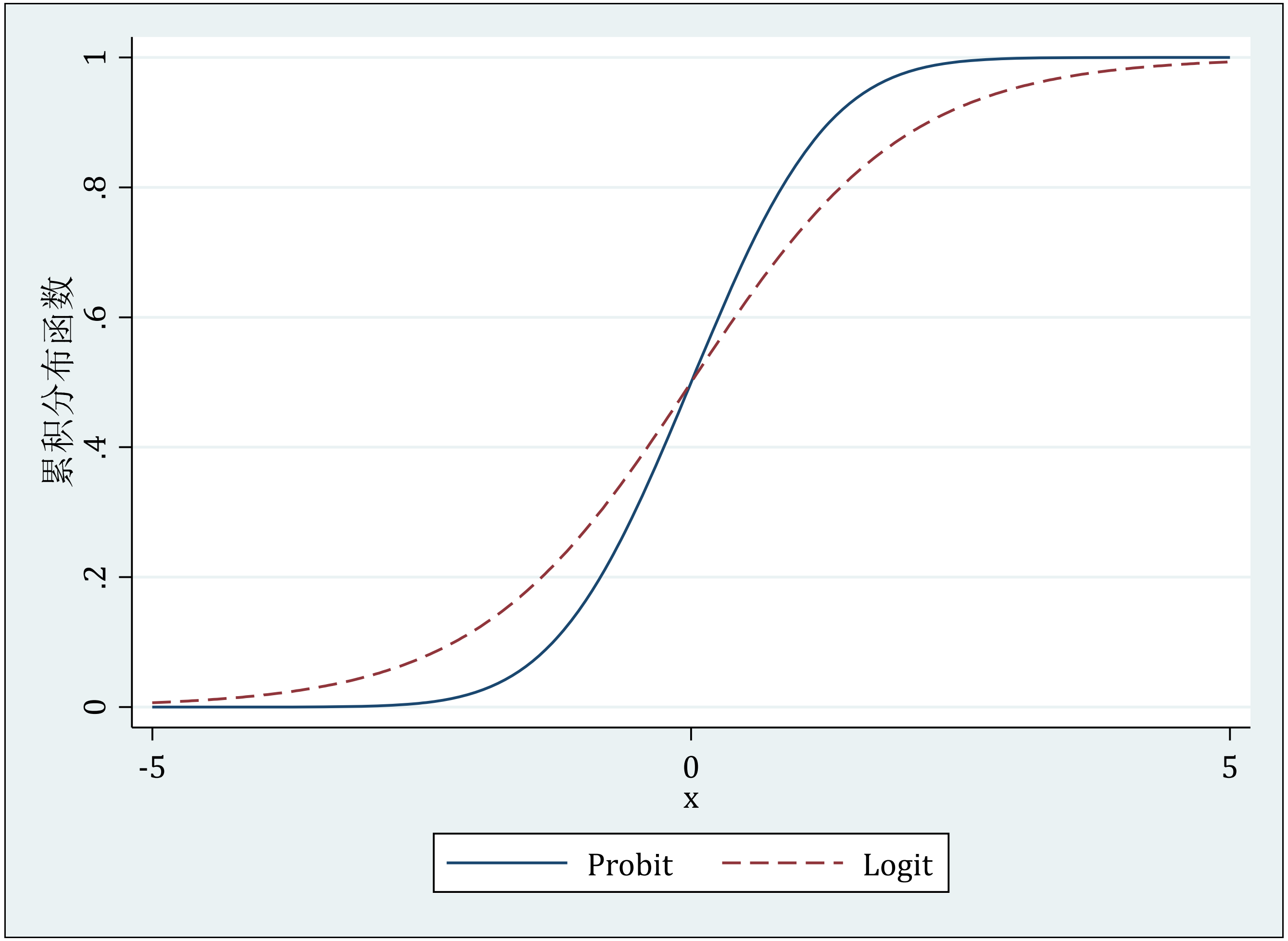

图 11.2　标准正态分布与逻辑分布的累积分布函数

在实践中，Probit 与 Logit 都很常用，二者的估计结果(比如边际效应)也通常很接近。

Logit 模型的优势在于，逻辑分布的累积分布函数有解析表达式(而标准正态分布没有)，故计算 Logit 更为方便

Logit 的回归系数更容易解释其经济意义。

11.2 最大似然估计的原理

Probit 与 Logit 模型在本质上都是非线性模型，无法通过变量转换而变为线性模型。

对于非线性模型，常使用**最大似然估计法**(Maximum Likelihood Estimation，简记 MLE 或 ML)。回顾概率统计中的最大似然估计。

假设随机变量y的概率密度函数为$f(y;\theta)$，其中θ为未知参数。为了估计θ，从y的总体中抽取样本容量为n的随机样本$\{y_1,\cdots,y_n\}$。

假设$\{y_1,\cdots,y_n\}$为 iid，则样本数据的联合密度函数为

$$f(y_1;\theta)\,f(y_2;\theta)\cdots f(y_n;\theta)=\prod_{i=1}^{n} f(y_i;\theta) \tag{11.6}$$

其中，$\prod_{i=1}^{n}$ 表示连乘。

在抽样之前，$\{y_1, \cdots, y_n\}$为随机向量。

抽样之后，$\{y_1, \cdots, y_n\}$就有了特定的样本值。

可将样本的联合密度函数视为在给定$\{y_1, \cdots, y_n\}$情况下，未知参数θ的函数。定义**似然函数**(likelihood function)为

$$L(\theta; y_1, \cdots, y_n) = \prod_{i=1}^{n} f(y_i; \theta) \qquad (11.7)$$

似然函数与联合密度函数完全相等，只是θ与$\{y_1, \cdots, y_n\}$的角色互换，即把θ作为自变量，而视$\{y_1, \cdots, y_n\}$为给定。

将乘积的形式转化为求和的形式，得到**对数似然函数**(log-likelihood function)：

$$\ln L(\theta; y_1, \cdots, y_n) = \sum_{i=1}^{n} \ln f(y_i; \theta) \qquad (11.8)$$

最大似然估计法的思想：给定样本取值后，该样本最有可能来自参数θ为何值的总体。

寻找$\hat{\theta}_{\mathrm{ML}}$，使得观测到样本数据的可能性最大，即最大化对数似然函数：

$$\max_{\theta} \ \ln L(\theta; y_1, \cdots, y_n) \qquad (11.9)$$

假设存在唯一内点解，则此无约束极值问题的一阶条件为

$$\frac{\partial \ln L(\theta; y_1, \cdots, y_n)}{\partial \theta} = 0 \qquad (11.10)$$

求解此一阶条件，即可得到最大似然估计量$\hat{\theta}_{\mathrm{ML}}$。

例 假设$y \sim N(\mu, \sigma^2)$，其中σ^2已知，得到一个样本容量为 1 的样本$y_1 = 2$，求对μ的最大似然估计。

根据正态分布的密度函数可知，此样本的似然函数为

$$L(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(2-\mu)^2}{2\sigma^2}\right\} \qquad (11.11)$$

此似然函数在$\hat{\mu} = 2$处取最大值，参见图 11.3。

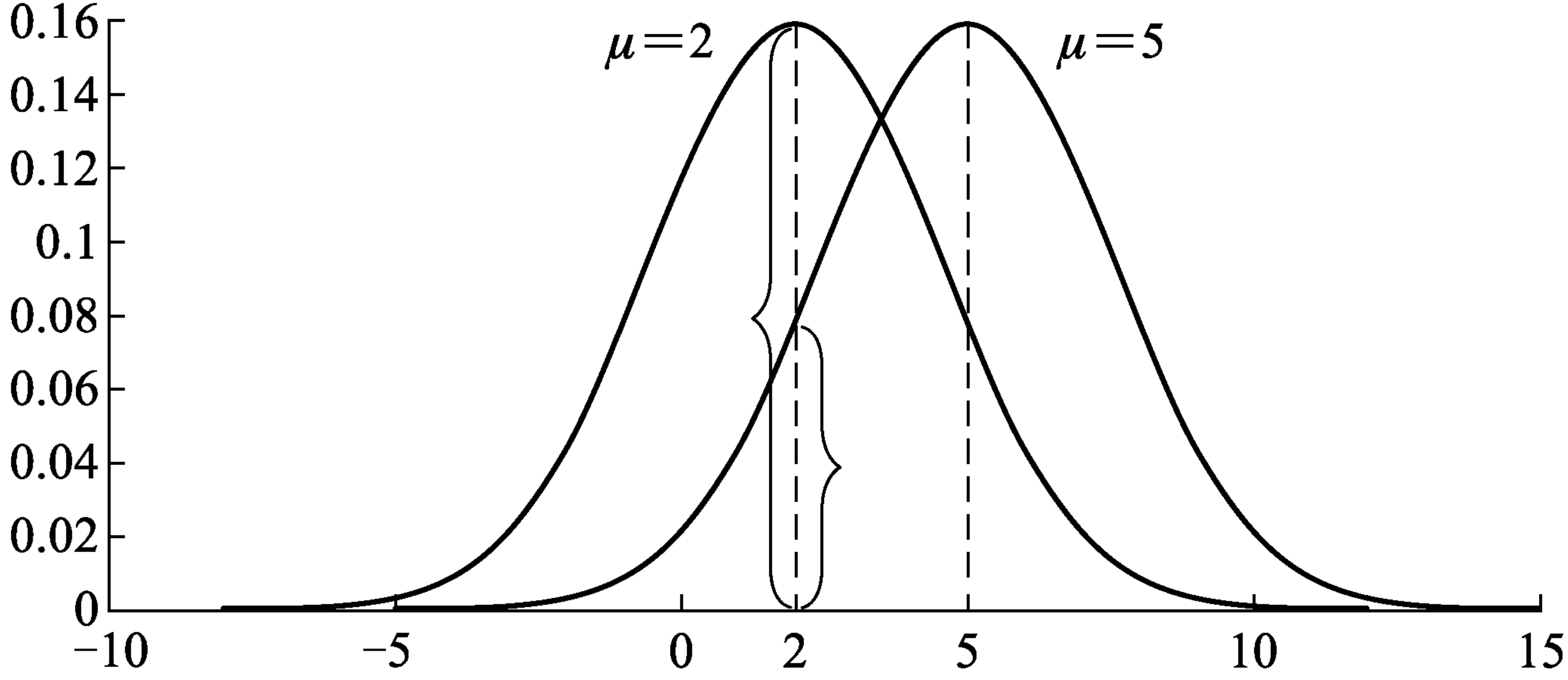

图 11.3　选择参数使观测到样本的可能性最大

例(非正式)　某人操一口浓重的四川口音，则判断他最有可能来自四川。

在一定的正则条件(regularity conditions)下，MLE 估计量具有以下良好的大样本性质，可照常进行大样本统计推断。

(1) $\hat{\theta}_{\mathrm{ML}}$为一致估计，即$\mathrm{plim}\,\hat{\theta}_{\mathrm{ML}} = \theta$。

(2) $\hat{\theta}_{\mathrm{ML}}$服从渐近正态分布。

(3) 在大样本下，$\hat{\theta}_{\mathrm{ML}}$是最有效率的估计(渐近方差最小)。

由于模型存在非线性，故最大似然估计通常没有解析解，而只能寻找“数值解”(numerical solution)。

在实践中，一般使用“迭代法”(iteration)进行数值求解。

常用的迭代法为**高斯−牛顿法**(Gauss-Newton method)。

MLE 的一阶条件可以归结为求非线性方程 $f(x)=0$的解。

假设 $f(x)$的导数 $f'(x)$处处存在，参见图 11.4。

记该方程的解为x^*，满足 $f(x^*)=0$。

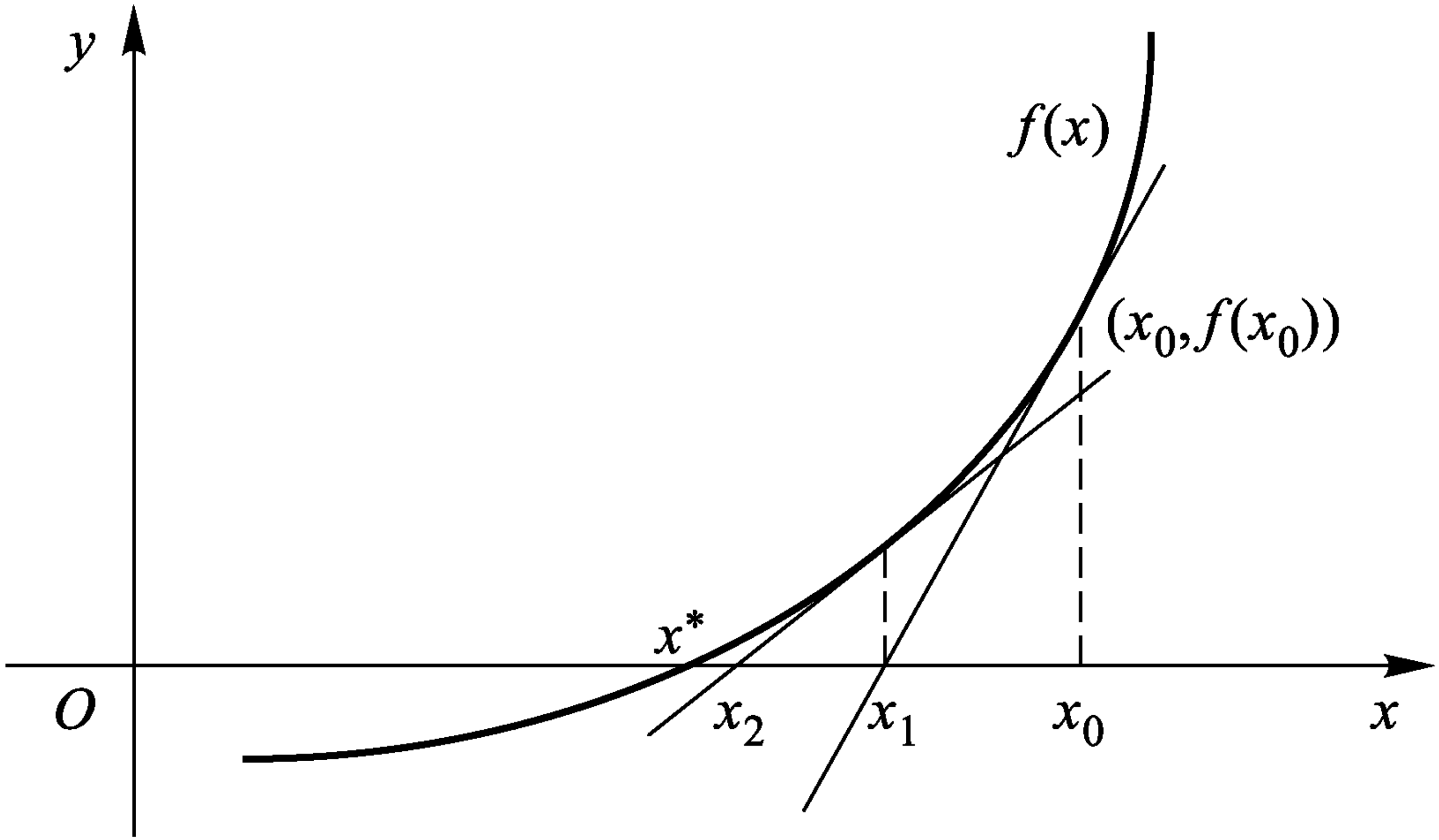

图 11.4　高斯-牛顿法

高斯-牛顿法之所以常用，原因之一是它的收敛速度很快，是二次的。

比如，如果本次迭代的误差为 0.1，则下次迭代的误差约为 0.1^2，而下下次迭代的误差约为 0.1^4，等等。

如果初始值 x_0 选择不当，也可能出现迭代不收敛的情形。

使用牛顿法得到的可能只是“局部最大值”(local maximum)，而非“整体最大值”(global maximum)。

MLE 很容易应用于多参数的情形。

假设随机变量y的概率密度函数为$f(y;\boldsymbol{\theta})$，其中$\boldsymbol{\theta}=(\theta_1 \quad \theta_2)'$，则对数似然函数为

$$\ln L(\boldsymbol{\theta}; y_1, \cdots, y_n) = \sum_{i=1}^{n} \ln f(y_i; \boldsymbol{\theta}) \qquad (11.12)$$

此最大化问题的一阶条件为

$$\begin{cases} \dfrac{\partial \ln L(\boldsymbol{\theta}; y_1, \cdots, y_n)}{\partial \theta_1} = 0 \\ \\ \dfrac{\partial \ln L(\boldsymbol{\theta}; y_1, \cdots, y_n)}{\partial \theta_2} = 0 \end{cases} \qquad (11.13)$$

求解此联立方程组(11.13)，即可得到最大似然估计量$\hat{\theta}_{1,\,\mathrm{ML}}$与$\hat{\theta}_{2,\,\mathrm{ML}}$。

高斯-牛顿法也适用于多元函数$f(\mathbf{x})=0$的情形，只要在上述迭代过程中，将切线替换为(超)切平面即可。

11.3　二值选择模型的 MLE 估计

以 Logit 模型为例，将 MLE 应用于二值选择模型。对于样本数据$\{\mathbf{x}_i, y_i\}_{i=1}^{n}$，根据方程(11.5)，第 i 个观测数据的概率为

$$f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = \begin{cases} \Lambda(\boldsymbol{x}_i'\boldsymbol{\beta}), & \text{若 } y_i = 1 \\ 1-\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta}), & \text{若 } y_i = 0 \end{cases} \tag{11.14}$$

其中，$\Lambda(z) \equiv \dfrac{\exp(z)}{1+\exp(z)}$为逻辑分布的累积分布函数。上式可更紧凑地写为

$$f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = \left[\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right]^{y_i} \left[1-\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right]^{1-y_i} \quad (11.15)$$

如果$y_i = 1$，则$1 - y_i = 0$，故上式等于$\left[\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right]$；反之，如果$y_i = 0$，则$1 - y_i = 1$，故上式等于$\left[1-\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right]$。

将上式取对数可得

$$\ln f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = y_i \ln\left[\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right] + (1-y_i)\ln\left[1-\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right] \quad (11.16)$$

假设样本中的个体相互独立，则整个样本的对数似然函数为

$$\ln L(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} y_i \ln\left[\Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right] + \sum_{i=1}^{n} (1 - y_i) \ln\left[1 - \Lambda(\boldsymbol{x}_i'\boldsymbol{\beta})\right] \tag{11.17}$$

把对数似然函数对 $\boldsymbol{\beta}$ 求偏导，即可得到最大化的一阶条件。

满足此一阶条件的估计量即为 MLE 估计量，记为 $\hat{\boldsymbol{\beta}}_{\text{ML}}$。

根据 MLE 的一般理论，$\hat{\boldsymbol{\beta}}_{\text{ML}}$ 为 $\boldsymbol{\beta}$ 的一致估计量，服从渐近正态分布，且在大样本下具有最小渐近方差。

11.4 边际效应

对于线性模型，回归系数β_k的经济意义十分明显，就是解释变量x_k对被解释变量y的边际效应(marginal effects)。

在非线性模型中，估计量$\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$一般并非边际效应。

以 Probit 为例，计算解释变量x_k的边际效应：

$$\frac{\partial \mathrm{P}(y=1 \mid \boldsymbol{x})}{\partial x_k} = \frac{\partial \Phi(\boldsymbol{x}'\boldsymbol{\beta})}{\partial x_k} = \frac{\partial \Phi(\boldsymbol{x}'\boldsymbol{\beta})}{\partial(\boldsymbol{x}'\boldsymbol{\beta})} \cdot \frac{\partial(\boldsymbol{x}'\boldsymbol{\beta})}{\partial x_k} = \phi(\boldsymbol{x}'\boldsymbol{\beta}) \cdot \beta_k \tag{11.18}$$

使用了微分的链式法则(chain rule)，且假定x_k为连续变量。

由于 Probit 与 Logit 所使用的分布函数不同，故其参数估计值并不直接可比。

需要分别计算二者的边际效应，然后进行比较。

由表达式(11.18)可知，对于非线性模型而言，边际效应通常不是常数，它随着解释向量**x**而变。

由于非线性模型的边际效应一般不是常数，故存在不同的边际效应概念。常用的边际效应概念包括：

(1) **平均边际效应**(average marginal effect)，即分别计算在每个样本观测值上的边际效应，然后进行简单算术平均。

(2) **样本均值处的边际效应**(marginal effect at mean)，即计算在 $\mathbf{x} = \bar{\mathbf{x}}$ 处的边际效应。

(3) **某代表值处的边际效应**(marginal effect at a representative value)，即给定 $\mathbf{x}^*$，计算在 $\mathbf{x} = \mathbf{x}^*$ 处的边际效应。

以上三种边际效应的计算结果可能有较大差异。传统上，常计算样本均值处 $\mathbf{x} = \bar{\mathbf{x}}$ 的边际效应，因为计算方便。

但在非线性模型中，样本均值处的个体行为并不等于样本中个体的平均行为。

对于政策分析而言，使用平均边际效应(Stata 的默认方法)，或在某代表值处的边际效应通常更有意义。

11.5 回归系数的经济意义

既然$\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$并非边际效应，那么它究竟有什么含义？

对于 Logit 模型，记事件发生的概率为$p \equiv \mathrm{P}(y=1\,|\,\boldsymbol{x})$，则事件不发生的概率为$1-p=\mathrm{P}(y=0\,|\,\boldsymbol{x})$。

由于$p=\dfrac{\exp(\boldsymbol{x}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}'\boldsymbol{\beta})}$，$1-p=\dfrac{1}{1+\exp(\boldsymbol{x}'\boldsymbol{\beta})}$，故事件发生与不发生的几率为

$$\frac{p}{1-p}=\exp(\boldsymbol{x}'\boldsymbol{\beta}) \qquad (11.19)$$

$\frac{p}{1-p}$称为**几率**(odds)或**相对风险**(relative risk)。

例如，在一个检验药物疗效的随机实验中，“$y=1$”表示“生”，而“$y=0$”表示“死”。

如果几率为 2，则意味着存活的概率是死亡概率的两倍。

对方程(11.19)两边取对数可得

$$\ln\left(\frac{p}{1-p}\right)=\boldsymbol{x}'\boldsymbol{\beta}=\beta_1 x_1+\cdots+\beta_K x_K \qquad (11.20)$$

$\ln\left(\dfrac{p}{1-p}\right)$称为**对数几率**(log-odds)，而上式右边为线性函数。

回归系数$\hat{\beta}_j$表示解释变量x_j增加一个微小量引起对数几率的边际变化。

由于取对数意味着百分比的变化，故可把$\hat{\beta}_j$视为半弹性(semi-elasticity)，即x_j增加一单位引起几率$\left(\dfrac{p}{1-p}\right)$的变化百分比。

比如，$\hat{\beta}_j = 0.12$，意味着x_j增加一单位引起几率增加 12%。

以上解释隐含地假设x_j为连续变量。

如果x_j为离散变量(比如，性别、子女数)，则可使用另一解释方法。

假设x_j增加一单位，从x_j变为x_j+1，记事件发生概率p的新值为p^*，则新几率与原几率的比值可写为(此处无法使用微积分)

$$\frac{\frac{p^*}{1-p^*}}{\frac{p}{1-p}}=\frac{\exp[\beta_1+\beta_2x_2+\cdots+\beta_j(x_j+1)+\cdots+\beta_Kx_K]}{\exp(\beta_1+\beta_2x_2+\cdots+\beta_jx_j+\cdots+\beta_Kx_K)}=\exp(\beta_j) \tag{11.21}$$

有些研究者偏好计算$\exp(\hat{\beta}_j)$，它表示解释变量x_j增加一单位引起几率的变化倍数。

Stata 称$\exp(\hat{\beta}_j)$为**几率比**(odds ratio)。

例 $\hat{\beta}_j = 0.12$，则$\exp(\hat{\beta}_j) = e^{0.12} = 1.13$，故当$x_j$增加一单位时，新几率是原几率的 1.13 倍，或增加 13%，因为$\exp(\hat{\beta}_j) - 1 = 1.13 - 1 = 0.13$。

如果$\hat{\beta}_j$较小，则$\exp(\hat{\beta}_j) - 1 \approx \hat{\beta}_j$(将$\exp(\hat{\beta}_j)$泰勒展开)，此时以上两种方法等价。

如果x_j至少必须变化一个单位(比如性别、婚否等虚拟变量，以及年龄，子女个数等)，则应使用$\exp(\hat{\beta}_j)$。

对于 Probit 模型，无法对其系数$\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$进行类似的解释。这是 Probit 模型的劣势。

11.6 拟合优度

如何衡量(非线性)二值模型的拟合优度呢？

由于不存在平方和分解公式，故无法计算R^2。

Stata 仍然汇报一个**准**R^2(Pseudo R^2)，由 McFadden (1974)所提出，其定义为

$$\text{准}R^2 \equiv \frac{\ln L_0 - \ln L_1}{\ln L_0} \qquad (11.22)$$

其中，$\ln L_1$为原模型的对数似然函数之最大值，而$\ln L_0$为以常数项为唯一解释变量的对数似然函数之最大值。

由于y为离散的两点分布，似然函数的最大可能值为 1(即取值概率为 1)，故对数似然函数的最大可能值为 0，记为$\ln L_{\max}$。

显然，$0 \geq \ln L_1 \geq \ln L_0$，而$0 \leq \text{准}R^2 \leq 1$，参见图 11.5。

由于 $\ln L_{\max} = 0$ ，故可将“准R^2”写为

$$\text{准}R^2 = \frac{\ln L_1 - \ln L_0}{\ln L_{\max} - \ln L_0} \qquad (11.23)$$

分子为加入解释变量后，对数似然函数的实际增加值 $(\ln L_1 - \ln L_0)$；而分母为最大可能增加值 $(\ln L_{max} - \ln L_0)$。

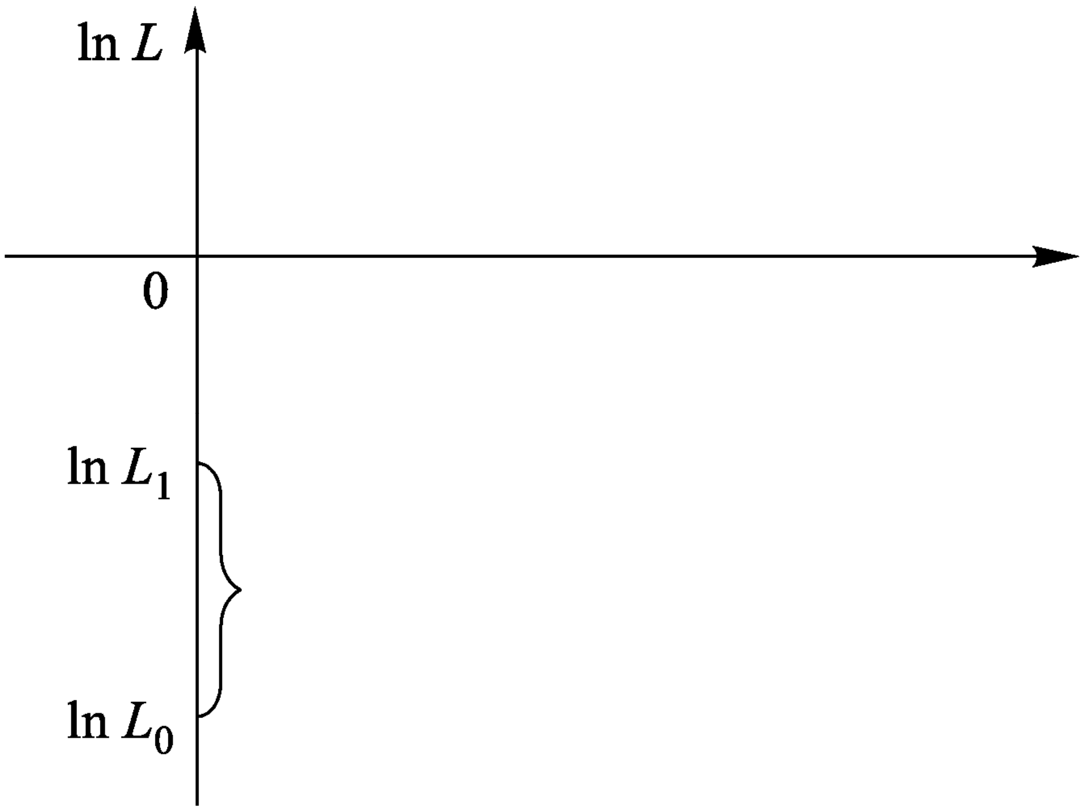

图 11.5　准 R^2 的计算

判断拟合优度的另一方法是计算**正确预测百分比**(percent correctly predicted)。

如果发生概率的预测值$\hat{y} \geq 0.5$，则认为其预测$y = 1$；反之，则认为其预测$y = 0$。

将预测值与实际值(样本数据)进行比较，即可计算正确预测的百分比。

11.7 准最大似然估计

使用 MLE 的前提是对总体的分布函数作具体的假定。

比如，Probit 与 Logit 模型分别假设被解释变量y的两点分布概率由标准正态或逻辑分布的累积分布函数所给出。

但此分布函数的设定可能不正确，即存在“设定误差”(specification error)。

定义 使用不正确的分布函数所得到的最大似然估计量，称为**准最大似然估计**(Quasi MLE，简记 QMLE)或**伪最大似然估计**(Pseudo MLE)。

准最大似然估计是否一定不一致？

不一定！

例如，假设线性模型的扰动项服从正态分布，则 MLE 估计量与 OLS 估计量完全相同，而 OLS 估计量的一致性并不依赖于关于分布函数的具体假设。

关于 QMLE 估计量的标准误可分为以下两种情况考虑。

(1) 如果 QMLE 为一致估计量，考虑到可能存在对分布函数的设定误差，故应使用稳健标准误(robust standard errors)，即相对于模型设定稳健的标准误。

此稳健标准误与异方差稳健的标准误是一致的，因为扰动项方差是否相同也是一种模型设定。

(2) 如果 QMLE 估计量不一致，则即使采用稳健标准误也无济于事。

此时，QMLE 估计量$\hat{\boldsymbol{\beta}}_{\text{QML}} \xrightarrow{p} \boldsymbol{\beta}^* \neq \boldsymbol{\beta}$，故首先应担心估计量的一致性

。

稳健标准误只是更精确估计了一个错误的**准真实参数**(pseudo true parameter)$\boldsymbol{\beta}^*$，而且通常不知道$\boldsymbol{\beta}^*$的经济意义。

在这种情况下，稳健标准误只是一致地估计了一个不一致估计量的方差(a consistent estimator of the variance of an inconsistent estimator)。

具体到二值选择模型(Probit 或 Logit 模型)，只要条件期望函数 $\mathrm{E}(y \mid \boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{\beta})$ 设定正确，则 MLE 估计就是一致的。

由于两点分布的特殊性，在 iid 的情况下，只要 $\mathrm{E}(y \mid \boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{\beta})$ 成立，则稳健标准误就等于 MLE 的普通标准误。

因此，如果认为模型设定正确，就没有必要使用稳健标准误(但使用稳健标准误也没有错)。

反之，如果模型设定不正确(即 $\mathrm{E}(y \mid \boldsymbol{x}) \neq F(\boldsymbol{x}, \boldsymbol{\beta})$)，则 Probit 与 Logit 模型并不能得到对系数 $\boldsymbol{\beta}$ 的一致估计，使用稳健标准误也就没有太大意义。

对于二值选择模型，使用普通标准误或稳健标准误都可以(文献中尚无定论)。

11.8 三类渐近等价的大样本检验

在计量经济学中，经常使用以下三类在大样本下渐近等价的统计检验。

考虑以下线性回归模型：

$$y_i = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, \cdots, n) \tag{11.24}$$

其中，解释变量$\boldsymbol{x} \equiv (x_1\ x_2\ \cdots\ x_K)'$，而参数$\boldsymbol{\beta} \equiv (\beta_1\ \beta_2\ \cdots\ \beta_K)'$。

考虑检验以下原假设：

$$H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad (11.25)$$

其中，$\boldsymbol{\beta}_0$已知，共有 K 个约束。

(1) **沃尔德检验**(Wald Test)：沃尔德检验通过考察$\boldsymbol{\beta}$的无约束估计量$\hat{\boldsymbol{\beta}}$与$\boldsymbol{\beta}_0$的距离来进行检验。

如果H_0正确，则$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$的绝对值不应该很大。

由于$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$为多维向量，故使用以下二次型：

$$W \equiv (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' [\mathrm{Var}(\hat{\boldsymbol{\beta}})]^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(K) \qquad (11.26)$$

在大样本下，此 Wald 统计量服从渐近$\chi^2(K)$分布，其中K为约束条件的个数(在此为解释变量个数)。

第 5-6 章所介绍的单一系数t检验(在大样本下使用标准正态进行检验)、联合线性假设的F检验(大样本下可使用χ^2分布进行检验)都是 Wald 检验。

(2) **似然比检验**(Likelihood Ratio Test，简记LR)：

似然比检验通过比较无约束估计量$\hat{\boldsymbol{\beta}}$与有约束估计量$\hat{\boldsymbol{\beta}}^*$的差别来进行检验。

无约束的似然函数最大值$\ln L(\hat{\boldsymbol{\beta}})$比有约束的似然函数最大值$\ln L(\hat{\boldsymbol{\beta}}^*)$更大

因为在无约束条件下的参数空间Θ比有约束条件下(即H_0成立时)参数的取值范围更大，参见图 11.6。

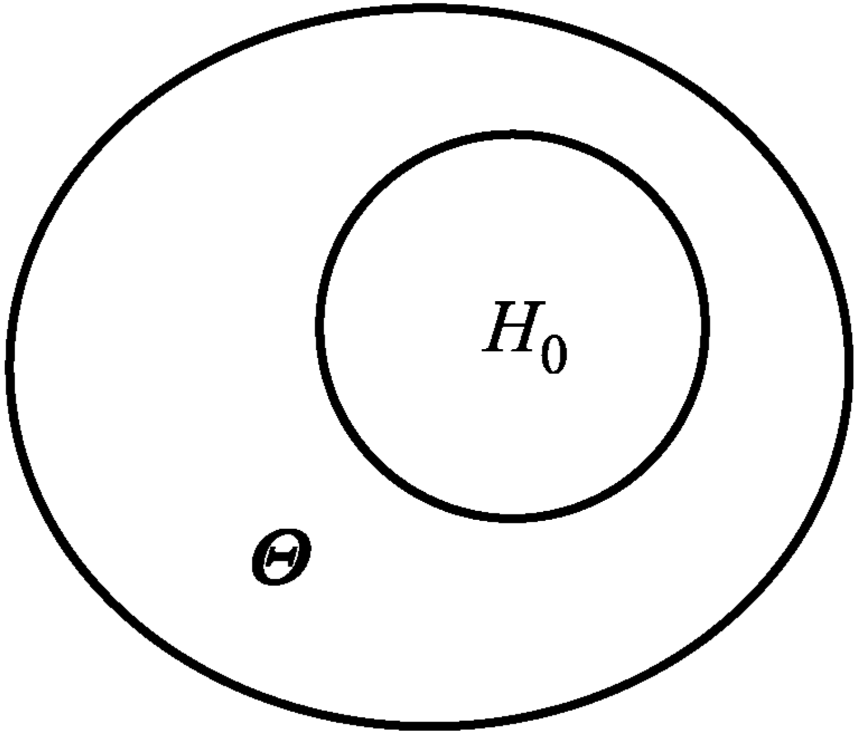

图 11.6　无约束与有约束的参数空间

如果H_0正确，则$[\ln L(\hat{\boldsymbol{\beta}}) - \ln L(\hat{\boldsymbol{\beta}}^*)]$不应该很大。

在此例中，有约束的估计量$\hat{\boldsymbol{\beta}}^* = \boldsymbol{\beta}_0$。

LR统计量为

$$LR \equiv -2\ln\left[\frac{L(\hat{\boldsymbol{\beta}}^*)}{L(\hat{\boldsymbol{\beta}})}\right] = 2\left[\ln L(\hat{\boldsymbol{\beta}}) - \ln L(\hat{\boldsymbol{\beta}}^*)\right] \xrightarrow{d} \chi^2(K) \tag{11.27}$$

在大样本下，LR统计量也服从渐近$\chi^2(K)$分布。

第 5 章 F 统计量的另一表达式 $F=\dfrac{(\text{SSR}^{*}-\text{SSR})/(K-1)}{\text{SSR}/(n-K)}$，就可看成是依据似然比原理而设计的。

在进行 Probit 或 Logit 回归时，Stata 会汇报一个似然比统计量，检验除常数项外所有参数的联合显著性，即考察原模型与只有常数项模型的似然函数最大值之比(准 R^2 的计算也基于此)。

(3) **拉格朗日乘子检验**(Lagrange Multiplier Test，简记 LM)：

Wald 检验只考察无约束估计量 $\hat{\boldsymbol{\beta}}$，LR 检验同时考察无约束估计量 $\hat{\boldsymbol{\beta}}$ 与有约束估计量 $\hat{\boldsymbol{\beta}}^{*}$。

LM 检验则只考察有约束估计量 $\hat{\boldsymbol{\beta}}^{*}$。

考虑以下有约束条件的对数似然函数最大化问题：

$$\begin{aligned} &\max_{\tilde{\boldsymbol{\beta}}} \ \ln L(\tilde{\boldsymbol{\beta}}) \\ &s.t. \ \tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 \end{aligned} \qquad (11.28)$$

$\tilde{\boldsymbol{\beta}}$为在最大化过程中假想的参数$\boldsymbol{\beta}$取值(hypothetical value)。

对于约束极值问题，可引入以下拉格朗日乘子函数：

$$\max_{\tilde{\boldsymbol{\beta}}, \boldsymbol{\lambda}} \ \ln L(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\lambda}'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \qquad (11.29)$$

其中，$\boldsymbol{\lambda}$为拉格朗日乘子向量(Lagrange Multiplier)，其经济含义为约束条件(比如资源约束)的影子价格(shadow price)。

如果$\hat{\boldsymbol{\lambda}} = \mathbf{0}$,则此约束条件完全不起作用(可以无偿获取任意数量的资源)。

根据一阶条件(对$\tilde{\boldsymbol{\beta}}$求导)可知,

$$\hat{\boldsymbol{\lambda}} = \frac{\partial \ln L(\hat{\boldsymbol{\beta}}^*)}{\partial \tilde{\boldsymbol{\beta}}} \equiv \begin{pmatrix} \dfrac{\partial \ln L(\hat{\boldsymbol{\beta}}^*)}{\partial \tilde{\beta}_1} \\ \vdots \\ \dfrac{\partial \ln L(\hat{\boldsymbol{\beta}}^*)}{\partial \tilde{\beta}_K} \end{pmatrix} \tag{11.30}$$

最优的拉格朗日乘子向量$\hat{\boldsymbol{\lambda}}$等于对数似然函数在约束估计量$\hat{\boldsymbol{\beta}}^*$处的一阶偏导数(切线的斜率)。

如果$\hat{\boldsymbol{\lambda}} \approx \mathbf{0}$，则说明此约束条件不“紧”(tight)或不是“硬约束”(binding constraint)，加上这个约束条件并不会使似然函数的最大值下降很多，即原假设H_0很可能成立。

如果原假设H_0成立，则$(\hat{\boldsymbol{\lambda}} - \mathbf{0})$的绝对值不应很大。

以二次型来度量此距离，可得LM统计量：

$$LM \equiv \hat{\boldsymbol{\lambda}}' [\mathrm{Var}(\hat{\boldsymbol{\lambda}})]^{-1} \hat{\boldsymbol{\lambda}} \xrightarrow{d} \chi^2(K) \qquad (11.31)$$

其中，LM统计量也服从渐近$\chi^2(K)$分布，而$\mathrm{Var}(\hat{\boldsymbol{\lambda}})$为$\hat{\boldsymbol{\lambda}}$的协方差矩阵。

由于似然函数的一阶导 $\hat{\boldsymbol{\lambda}} = \dfrac{\partial \ln L(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}$ 被称为**得分函数**(score function)或**得分向量**(score vector)，故此检验也称为**得分检验**(score test)。

由于在无约束估计量 $\hat{\boldsymbol{\beta}}$ 处，$\dfrac{\partial \ln L(\hat{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}} = \mathbf{0}$ (MLE 的一阶条件)，故如果原假设 H_0 成立，则在约束估计量 $\hat{\boldsymbol{\beta}}^*$ 处，此得分向量也应该接近于 $\mathbf{0}$，即 $\dfrac{\partial \ln L(\hat{\boldsymbol{\beta}}^*)}{\partial \tilde{\boldsymbol{\beta}}} \approx \mathbf{0}$，而 LM 统计量反映的就是此接近程度。

在第 7-8 章，对异方差与自相关所进行的 nR^2 形式的检验都来自于 LM 检验的推导。

Wald 检验仅利用无约束估计的信息，*LM* 检验仅利用有约束估计的信息，而*LR*检验同时利用无约束与有约束估计的信息。

这三类检验在大样本下是渐近等价的，它们只是从不同的侧面去考察同一事物。

可以把这三类统计检验的思想画在同一张图上，参见图 11.7。

在实际应用中，究竟采取哪种检验常取决于“无约束估计”与“有约束估计”哪种更方便。

如果无约束估计更方便，则常使用 Wald 检验(比如，对线性回归系数的显著性检验)；如果有约束估计更方便，则常使用*LM* 检验(比如，对异方差、自相关的检验)；如果二者都方便，则可使用*LR*检验(比如，对非线性回归方程的显著性检验)。

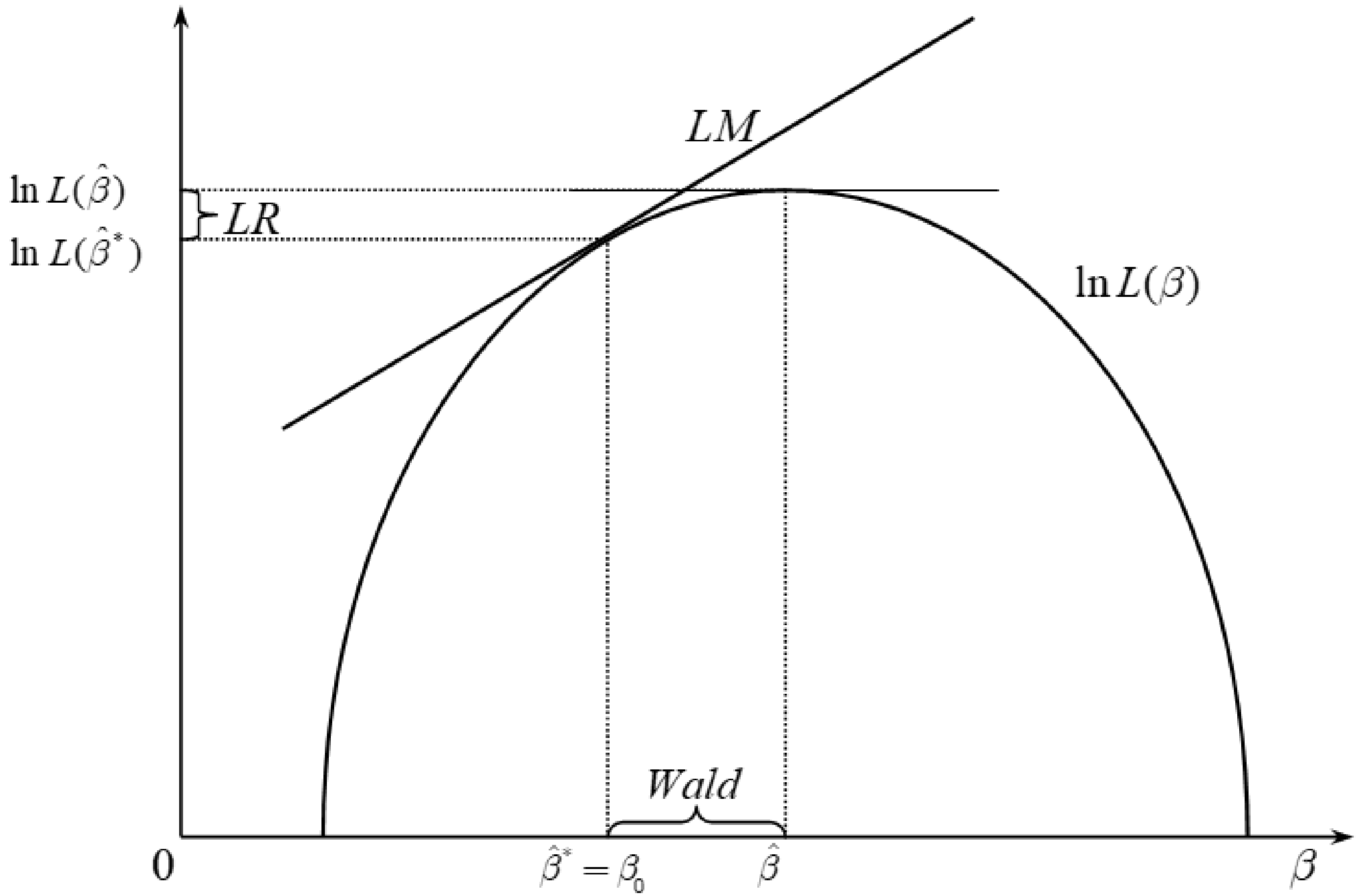

图 11.7　三类渐近等价的统计检验

11.9　二值选择模型的 Stata 命令与实例

二值模型的 Stata 命令为

```
. probit y x1 x2 x3,r
```
(probit 模型)

```
. logit y x1 x2 x3,r or
```
(logit 模型)

选择项“r”表示使用稳健标准误(默认为普通标准误)；

选择项“or”表示显示几率比(odds ratio)，而不显示回归系数。

完成 Probit 或 Logit 估计后，可进行预测，计算准确预测的百分比，或计算边际效应：

`. predict y1` (计算发生概率的预测值，记为 y1)

`. estat clas` (计算准确预测的百分比，clas 表示 classification)

`. margins,dydx(*)` (计算所有解释变量的平均边际效应；“*”代表所有解释变量)

`. margins,dydx(*) atmeans` (计算所有解释变量在样本均值处的边际效应)

`. margins,dydx(*) at(x1=0)` (计算所有解释变量在 $x1 = 0$ 处的平均边际效应)

`. margins,dydx(x1`　(计算解释变量 x1 的平均边际效应)

`. margins,eyex(*)`　(计算平均弹性，其中的两个“e”均指 elasticity)

`. margins,eydx(*)`　(计算平均半弹性，x变化一单位引起y变化百分之几)

`. margins,dyex(*)`　(计算平均半弹性，x变化 1%引起y变化几个单位)

以数据集 titanic.dta 为例，演示二值选择模型的 Stata 估计。

该数据集包括泰坦尼克号乘客的存活数据。

被解释变量为 *survive*(存活=1，死亡=0)。

解释变量包括 *child*(儿童=1，成年=0)，*female*(女性=1，男性=0)，*class1*(头等舱=1，其他=0)，*class2*(二等舱=1，其他=0)，*class3*(三等舱=1，其他=0)，*class4*(船员=1，其他=0)。

在生死的紧要关头，各色人等的存活概率几何？

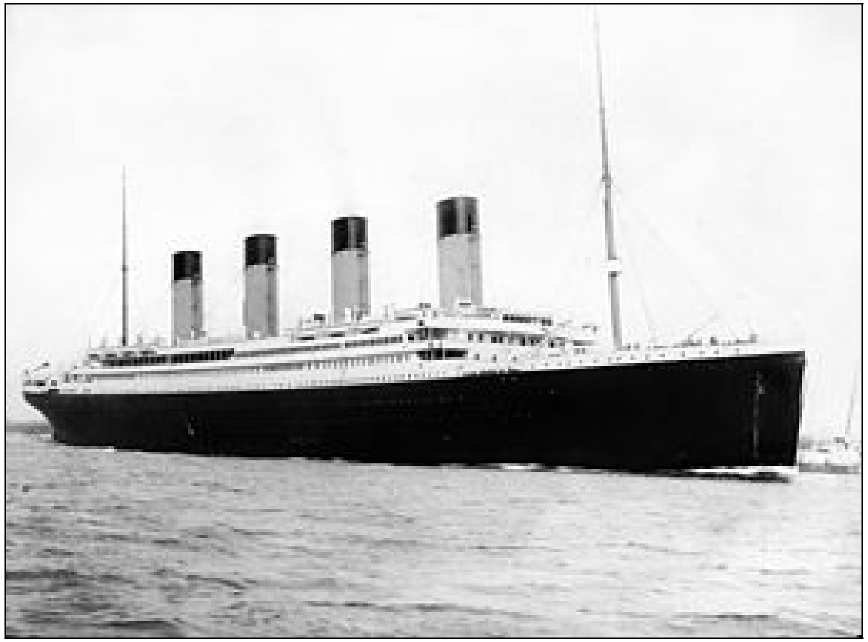

图 11.8 泰坦尼克号于 1914 年 4 月 10 日从英国南安普顿港出发

首先打开数据集，看一下原始数据。

```
. use titanic.dta,clear
. list
```

	class1	class2	class3	class4	child	female	survive	freq
1.	0	0	1	0	1	0	0	35
2.	0	0	1	0	1	1	0	17
3.	1	0	0	0	0	0	0	118
4.	0	1	0	0	0	0	0	154
5.	0	0	1	0	0	0	0	387
6.	0	0	0	1	0	0	0	670
7.	1	0	0	0	0	1	0	4
8.	0	1	0	0	0	1	0	13
9.	0	0	1	0	0	1	0	89
10.	0	0	0	1	0	1	0	3
11.	1	0	0	0	1	0	1	5
12.	0	1	0	0	1	0	1	11
13.	0	0	1	0	1	0	1	13
14.	1	0	0	0	1	1	1	1
15.	0	1	0	0	1	1	1	13
16.	0	0	1	0	1	1	1	14
17.	1	0	0	0	0	0	1	57
18.	0	1	0	0	0	0	1	14
19.	0	0	1	0	0	0	1	75
20.	0	0	0	1	0	0	1	192
21.	1	0	0	0	0	1	1	140
22.	0	1	0	0	0	1	1	80
23.	0	0	1	0	0	1	1	76
24.	0	0	0	1	0	1	1	20

原始数据只有 24 个观测值，但每个观测值可能重复多次；其重复次数以最后一列变量 `freq` 来表示。

比如，第一行数据显示，乘坐三等舱的男孩死亡者有 35 人；第二行数据显示，乘坐三等舱的女孩死亡者有 17 人；以此类推。

对于这种观测值重复的数据，在进行计算与估计时，必须以重复次数(`freq`)作为权重才能得到正确的结果。

其效果就相当于在数据文件中，将第一行数据重复 35 次，将第二行数据重复 17 次，以此类推(不同于以方差倒数为权重的加权最小二乘法)。

假设观测值的重复次数记录于变量 `freq`，在 Stata 中可通过在命令的最后加上“`[fweight=freq]`”来实现此加权计算或估计。其中，“`fweight`” 指 “frequency weight” (频数权重)。

首先看一下各变量的统计特征。

```
. sum [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
class1	2,201	.1476602	.3548434	0	1
class2	2,201	.1294866	.335814	0	1
class3	2,201	.3207633	.466876	0	1
class4	2,201	.40209	.4904313	0	1
child	2,201	.0495229	.2170065	0	1
female	2,201	.2135393	.4098983	0	1
survive	2,201	.323035	.4677422	0	1
freq	2,201	329.2726	250.0362	1	670

样本容量为2201(旅客与船员总人数)，而非24。

从变量survive的平均值可知,泰坦尼克号的平均存活率为0.32。

分别计算小孩、女士以及各等舱旅客的存活率。

```
. sum survive if child [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
survive	109	.5229358	.5017807	0	1

```
. sum survive if female [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
survive	470	.7319149	.4434342	0	1

```
. sum survive if class1 [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
survive	325	.6246154	.4849687	0	1

```
. sum survive if class2 [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
survive	285	.4140351	.493421	0	1

```
. sum survive if class3 [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
survive	706	.2521246	.4345403	0	1

```
. sum survive if class4 [fweight=freq]
```

Variable	Obs	Mean	Std. dev.	Min	Max
survive	885	.239548	.427049	0	1

小孩、女士、一等舱、二等舱的存活率分别为 0.52、0.73、0.62、0.41，高于平均存活率

三等舱、船员的存活率分别为 0.25、0.24，低于平均存活率。

下面进行回归分析。作为参照系，首先使用 OLS 估计线性概率模型。

```
. reg survive child female class1 class2 class3 [fweight=freq],r
```

由于所有乘客分为四类(class1-class4)，故只能放入三个虚拟变量(class1-class3)

将虚拟变量 class4(船员)作为参照类别，不放入回归方程。

```
Linear regression                               Number of obs     =      2,201
                                                F(5, 2195)        =     221.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2529
                                                Root MSE          =     .40474

------------------------------------------------------------------------------
             |               Robust
     survive | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       child |   .1812957   .0479499     3.78   0.000     .0872639    .2753275
      female |   .4906798   .0239292    20.51   0.000     .4437535    .5376061
      class1 |   .1755538   .0291386     6.02   0.000     .1184117     .232696
      class2 |  -.0105263   .0258402    -0.41   0.684       -.0612    .0401475
      class3 |  -.1311806   .0212996    -6.16   0.000      -.17295   -.0894112
       _cons |   .2267959   .0139872    16.21   0.000     .1993664    .2542254
------------------------------------------------------------------------------
```

儿童(child)、妇女(female)与头等舱旅客(class1)的存活概率均显著地更高，三等舱旅客(class3)的存活概率显著地更低，而二等舱旅客(class2)的存活概率与船员无显著差异。

其次，使用 Logit 进行估计：

```
. logit survive child female class1 class2 class3 [fweight=freq],nolog
```

其中，选择项“nolog”表示不显示 MLE 数值计算的迭代过程。

```
Logistic regression                                     Number of obs =  2,201
                                                        LR chi2(5)    = 559.40
                                                        Prob > chi2   = 0.0000
Log likelihood = -1105.0306                             Pseudo R2     = 0.2020

------------------------------------------------------------------------------
     survive | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       child |   1.061542   .2440257     4.35   0.000     .5832608    1.539824
      female |    2.42006   .1404101    17.24   0.000     2.144862    2.695259
      class1 |   .8576762   .1573389     5.45   0.000     .5492976    1.166055
      class2 |  -.1604188   .1737865    -0.92   0.356    -.5010342    .1801966
      class3 |  -.9200861   .1485865    -6.19   0.000     -1.21131   -.6288619
       _cons |  -1.233899   .0804946   -15.33   0.000    -1.391666   -1.076133
------------------------------------------------------------------------------
```

Logit 的估计结果在显著性方面与 OLS 完全一致。准R^2为 0.20。

检验整个方程显著性的LR统计量(`LR chi2(5)`)为 559.40，对应的p值为 0.000，故整个方程的联合显著性很高。

下面使用稳健标准误进行 Logit 估计。

```
. logit survive child female class1 class2 class3
[fweight=freq],nolog r
```

```
Logistic regression                                     Number of obs =  2,201
                                                        Wald chi2(5)  = 467.05
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -1105.0306                       Pseudo R2     = 0.2020

------------------------------------------------------------------------------
             |               Robust
     survive | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       child |   1.061542   .2767452     3.84   0.000     .5191318    1.603953
      female |    2.42006   .1363096    17.75   0.000     2.152898    2.687222
      class1 |   .8576762   .1475218     5.81   0.000     .5685387    1.146814
      class2 |  -.1604188   .1502193    -1.07   0.286    -.4548432    .1340056
      class3 |  -.9200861   .1621035    -5.68   0.000    -1.237803    -.602369
       _cons |  -1.233899   .0798876   -15.45   0.000    -1.390476   -1.077322
------------------------------------------------------------------------------
```

稳健标准误与普通标准误比较接近。

由于此回归中的解释变量均为虚拟变量，只能变化一个单位(从 0 变为 1)，为了便于解释回归结果，让 Stata 汇报几率比而非系数。

```
. logit survive child female class1 class2 class3 [fweight=freq],or nolog
```

```
Logistic regression                                     Number of obs =  2,201
                                                        LR chi2(5)    = 559.40
                                                        Prob > chi2   = 0.0000
Log likelihood = -1105.0306                             Pseudo R2     = 0.2020
```

survive	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
child	2.890826	.7054359	4.35	0.000	1.791872	4.663769
female	11.24654	1.579128	17.24	0.000	8.540859	14.80936
class1	2.357675	.3709541	5.45	0.000	1.732036	3.209306
class2	.851787	.1480291	-0.92	0.356	.6059037	1.197453
class3	.3984847	.0592095	-6.19	0.000	.2978068	.5331983
_cons	.2911551	.0234364	-15.33	0.000	.2486608	.3409114

Note: **_cons** estimates baseline odds.

儿童的生存概率是成年人的近 3 倍(几率比为 2.89)。

妇女的存活概率是男人的 11 倍多(几率比为 11.25)。

头等舱旅客的存活概率是船员的 2.36 倍。

三等舱旅客的存活概率只是船员的 39.8%。

二等舱旅客的存活概率也略低于船员(几率比为 0.85)，但此差别在统计上不显著(p 值为 0.356)。

为了与 OLS 估计的回归系数比较，计算 Logit 模型的平均边际效应：

```
. margins,dydx(*)
```

```
Average marginal effects                          Number of obs = 2,201
Model VCE: OIM

Expression: Pr(survive), predict()
dy/dx wrt:  child female class1 class2 class3
```

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
child	.1732315	.0393799	4.40	0.000	.0960484	.2504147
female	.394926	.0171966	22.97	0.000	.3612214	.4286307
class1	.1399629	.0250922	5.58	0.000	.0907831	.1891427
class2	-.0261785	.0283616	-0.92	0.356	-.0817663	.0294093
class3	-.1501475	.0238334	-6.30	0.000	-.1968602	-.1034348

Logit 模型的平均边际效应与 OLS 回归系数相差不大。

为了演示目的，计算在样本均值处的边际效应。

```
. margins,dydx(*) atmeans
```

```
Conditional marginal effects                          Number of obs = 2,201
Model VCE: OIM

Expression: Pr(survive), predict()
dy/dx wrt:  child female class1 class2 class3
At: child  = .0495229 (mean)
    female = .2135393 (mean)
    class1 = .1476602 (mean)
    class2 = .1294866 (mean)
    class3 = .3207633 (mean)

-------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
       child |   .2223422   .0510772     4.35   0.000     .1222328    .3224516
      female |   .5068865   .0303542    16.70   0.000     .4473934    .5663797
      class1 |    .179642   .0332374     5.40   0.000     .1144979    .2447861
      class2 |     -.0336   .0363774    -0.92   0.356    -.1048983    .0376983
      class3 |  -.1927139   .0308186    -6.25   0.000    -.2531173   -.1323105
-------------------------------------------------------------------------------
```

计算 Logit 模型准确预测的比率:

```
. estat clas
```

```
Logistic model for survive

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       349           126  |        475
     -     |       362          1364  |       1726
-----------+--------------------------+-----------
   Total   |       711          1490  |       2201

Classified + if predicted Pr(D) >= .5
True D defined as survive != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   49.09%
Specificity                     Pr( -|~D)   91.54%
Positive predictive value       Pr( D| +)   73.47%
Negative predictive value       Pr(~D| -)   79.03%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)    8.46%
False - rate for true D         Pr( -| D)   50.91%
False + rate for classified +   Pr(~D| +)   26.53%
False - rate for classified -   Pr( D| -)   20.97%
--------------------------------------------------
Correctly classified                        77.83%
--------------------------------------------------
```

正确预测的比率为 $(349 + 1364)/2201 = 77.83\ \%$。

根据 Logit 模型的回归结果，　　预测每位乘客的存活概率，并记为变量 *prob*。

```
. predict prob
(option pr assumed; Pr(survive))
```

由此可考察给定某种特征旅客的生存概率。比如，计算 Ms. Rose (头等舱、成年、女性)的存活概率：

```
. list prob survive freq if class1==1 & child==0
& female==1
```

	prob	survive	freq
7.	.8853235	0	4
21.	.8853235	1	140

Ms. Rose 的存活概率高达 88.5%。

从频率上看，所有头等舱的 144 位成年女性中，只有 4 位死亡。

又比如，计算 Mr. Jack (三等舱、成年、男性)的存活概率：

```
. list prob survive freq if class3==1 & child==0
& female==0
```

	prob	survive	freq
5.	.1039594	0	387
19.	.1039594	1	75

Mr. Jack 的存活概率仅有 10.4%。

从频率上看，所有三等舱的 462 位成年男性中，只有 75 位生还。

类似地，可对此数据集进行 Probit 估计。

```
. probit survive child female class1 class2
class3 [fweight=freq],nolog
```

```
Probit regression                                       Number of obs =  2,201
                                                        LR chi2(5)    = 556.83
                                                        Prob > chi2   = 0.0000
Log likelihood = -1106.3142                             Pseudo R2     = 0.2011

------------------------------------------------------------------------------
     survive | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       child |   .5803382   .1377535     4.21   0.000     .3103463      .85033
      female |    1.44973   .0808635    17.93   0.000      1.29124    1.608219
      class1 |   .5399101   .0951552     5.67   0.000     .3534092    .7264109
      class2 |  -.0898158   .1028857    -0.87   0.383    -.2914681    .1118364
      class3 |  -.4875252   .0800342    -6.09   0.000    -.6443893   -.3306611
       _cons |  -.7530486   .0468804   -16.06   0.000    -.8449325   -.6611648
------------------------------------------------------------------------------
```

由于 Probit 与 Logit 模型的回归系数并不直接可比，下面考察 Probit 模型的平均边际效应及预测准确度。

```
. margins,dydx(*)
```

```
Average marginal effects                                    Number of obs = 2,201
Model VCE: OIM

Expression: Pr(survive), predict()
dy/dx wrt:  child female class1 class2 class3

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       child |   .1640035   .0386284     4.25   0.000     .0882932    .2397137
      female |   .4096934   .0177738    23.05   0.000     .3748574    .4445294
      class1 |   .1525785   .0262955     5.80   0.000     .1010403    .2041167
      class2 |  -.0253819   .0290666    -0.87   0.383    -.0823515    .0315876
      class3 |  -.1377745   .0223131    -6.17   0.000    -.1815075   -.0940416
------------------------------------------------------------------------------
```

```
. estat clas
```

Probit model for survive

Classified	True D	True ~D	Total
+	349	126	475
-	362	1364	1726
Total	711	1490	2201

Classified + if predicted Pr(D) >= .5
True D defined as survive != 0

Sensitivity	Pr(+| D)	49.09%
Specificity	Pr(-|~D)	91.54%
Positive predictive value	Pr(D| +)	73.47%
Negative predictive value	Pr(~D| -)	79.03%
False + rate for true ~D	Pr(+|~D)	8.46%
False - rate for true D	Pr(-| D)	50.91%
False + rate for classified +	Pr(~D| +)	26.53%
False - rate for classified -	Pr(D| -)	20.97%
Correctly classified		77.83%

Probit 模型的平均边际效应、准 R^2 与正确预测比率与 Logit 模型十分接近，故可视为基本等价(二者的估计系数虽有差距，但估计系数没有可比性)。

为了进一步验证这一点，下面使用 Probit 模型预测每位个体的存活概率，记为变量 *prob1*，并考察 *prob1* 与 *prob*(Logit 模型预测结果)的相关性。

```
. predict prob1
(option pr assumed; Pr(survive))

. corr prob prob1 [fweight=freq]
(obs=2201)
```

	prob	prob1
prob	1.0000	
prob1	0.9997	1.0000

Probit 与 Logit 模型对个体存活概率的预测结果相关系数高达0.9997，可以视为无差异。

11.10 其他离散选择模型

二值选择模型并非唯一的离散选择模型。其他离散选择模型还包括：

(1) 多值选择(multiple choices)：比如，对交通方式的选择(步行、骑车、自驾车、打的、地铁)，对不同职业的选择，对手机品牌的选择。

(2) 计数数据(count data)：有时被解释变量只能取非负整数。比如，企业在某段时间内获得的专利数；某人在一定时间内去医院看病的次数；某省在一年内发生煤矿事故的次数。

(3) 排序数据(ordered data)：有些离散数据有着天然的排序。比如，公司债券的评级(AAA, AA, A, B, C 级)，对“春节联欢晚会”的满意度(很满意、满意、不满意、很不满意)。

对于以上离散数据，一般也不宜直接进行 OLS 回归，主要估计方法仍为 MLE。

由于离散选择模型主要用于微观经济学的实证研究中，故是“微观计量经济学”(Microeconometrics)的重要组成部分。

除了离散数据外，微观计量经济学还关注的另一类数据类型为“受限被解释变量”(limited dependent variable)，即被解释变量的取值范围受到限制(包括断尾回归、归并回归与样本选择模型等)。

有关离散选择模型与受限被解释变量的具体介绍，请参见Wooldridge (2009)或陈强(2014, 2024)。