《线性回归》
—线性回归(2)

杨　　瑛

清华大学　数学科学系
Email: yangying@mail.tsinghua.edu.cn
Tel: 62796887

2019.03.14

主要内容：线性模型(2)

1. **多重线性模型的LSE和MLE**
   - 多重线性模型的LSE
   - 协方差矩阵
   - 不对单个变量回归的原因
   - 线性模型参数的LSE
   - 线性模型参数的MLE

数据结构和模型

- 设$(\mathbf{Y}_i, \mathbf{X}_i), 1 \le i \le n,$ 是n的独立同分布的观测数据，$\mathbf{Y}_i$是响应变量，$\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})^T$是$p \times 1$的协变量。
- 对于这些数据，建立多重线性回归模型：

$$\begin{aligned} \mathbf{Y}_i &= \beta_0 + \mathbf{X}_i^T \beta + \epsilon_i \\ &= \beta_0 + \sum_{k=1}^{p} X_{ik}\beta_k + \epsilon_i, 1 \le i \le n, \end{aligned} \tag{1}$$

其中β_0和β是未知参数（向量），ϵ_i是不可观测的随机误差。通常假定：

- ✓**(C1)** $E[\epsilon_i] = 0$;
- ✓**(C2)** $\mathrm{Var}(\epsilon_i) = \sigma^2, \sigma^2 > 0$是未知的;
- ✓**(C3)** $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0, 1 \le i \ne j \le n$。

数据结构和模型(续)

- 为了紧凑期间，假定$X_{i1} \equiv 1$, 则模型(1)可以改写为：

$$\mathbf{Y}_i = \mathbf{X}_i^T \beta + \epsilon_i, 1 \leq i \leq n, \tag{2}$$

其中β是$p \times 1$的未知参数向量，ϵ_i是不可观测的随机误差。

- 模型(2)还可以写为更为紧凑的矩阵形式：

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \tag{3}$$

其中$\mathbf{Y}$是$n \times 1$的列向量，$\mathbf{X}$是$n \times p$的矩阵，θ是$p \times 1$ 的列向量，ϵ 是$n \times 1$的随机误差向量。

数据结构和模型(续)

- 当假定随机误差$(\epsilon_1, \cdots, \epsilon_n) \sim N(\mathbf{0}, \Sigma)$时，模型(3)通常写为：

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N(\mathbf{0}, \Sigma), \tag{4}$$

其中$\mathbf{Y}$是$n \times 1$的列向量，$\mathbf{X}$是$n \times p$的矩阵，θ是$p \times 1$ 的列向量，ϵ 是$n \times 1$的随机误差向量, $\mathbf{0}$是所有元素为0的$n \times 1$向量，Σ是$n \times n$的正定矩阵。

协方差矩阵的作用：

- Σ用来刻画随机误差的相关性。特别地，如果随机误差是iid的$N(0,\sigma^2)$随机变量时，$\Sigma=\sigma^2 I_n$, I_n是$n\times n$的单位阵。
- Σ还有其它选择，例如，
  - ✓ $\Sigma=\text{diag}(\sigma_1^2,\cdots,\sigma_n^2)$;
  - ✓ Σ的主对角线元素是σ^2, 次对角线的元素是ρ $(0<\rho<1)$, 其余元素皆为0;
  - ✓ Σ的主对角线的元素都是σ^2, 第j次对角线的元素是ρ^j $(0<\rho<1), j=1,\cdots,n-1$.
- 尝试理解不同协方差矩阵的含义。

多重回归与简单回归
不对单个变量回归的原因：

- 当有一个响应变量，多个协变量时，是做多个简单回归还是做多重回归？
- 下面的一个人造的例子要说明为什么多重回归不能简单地被几个简单的回归过程所代替。

Example

假设有两个协变量x_1, x_2, 它们的观察结果如下:

x_1	0	1	2	3	0	1	2	3
x_2	-1	0	1	2	1	2	3	4
y	1	2	3	4	-1	0	1	2

y与x_1和x_2的关系时怎样的？

Example

图1的左侧: 我们将y的值与协变量x_1和x_2的对应值作图。随后，我们(在三维空间中)找到了一个平面，它与这8个点完全吻合:

$$y = 2x_1 - x_2, \quad (\hat{\sigma}^2 = 0)$$

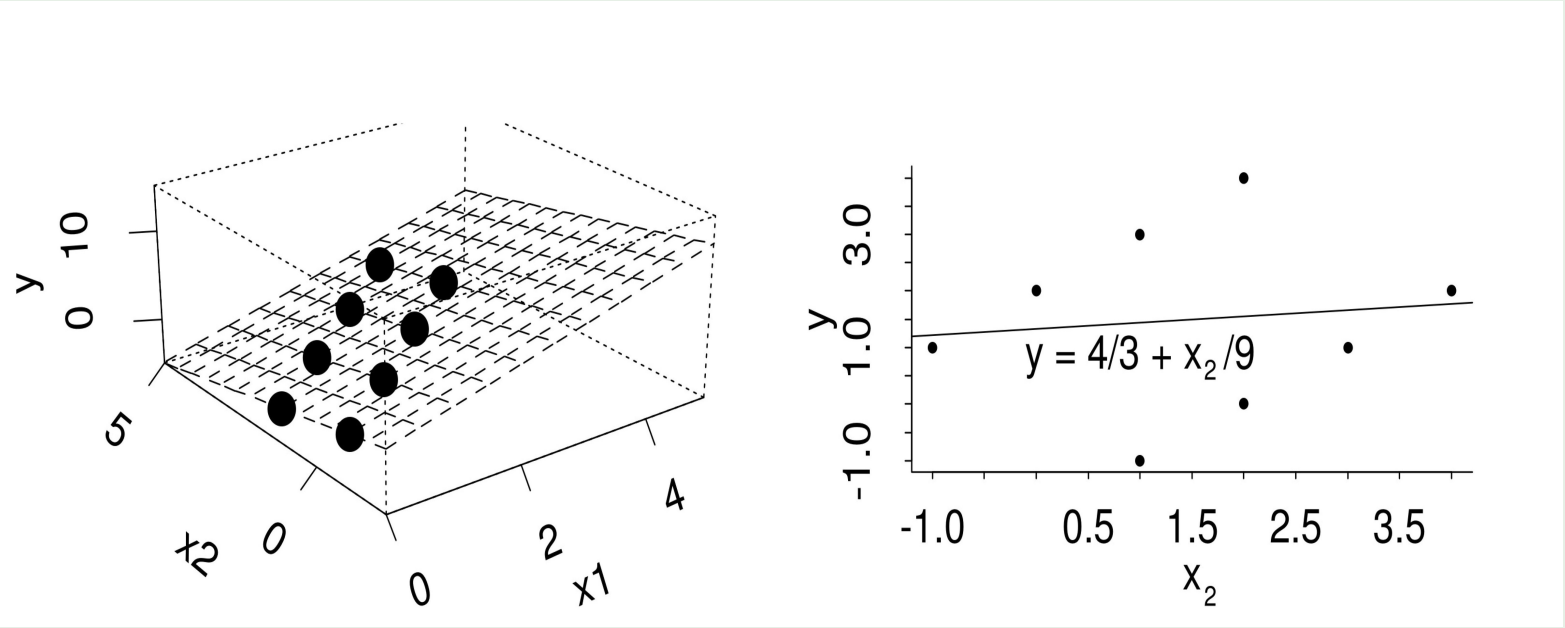

Figure: 多重回归与简单回归

Example

这里的系数(2和-1)告诉我们，如果固定其中的一个变量，让另外一个协变量改变一个单位，系数刚好是y的改变量。

我们得出这样的结论: y**随着**x_2**的增加而减小，(**x_2**越大**$\Rightarrow$ y**越小)。**

图1的右边: 我们简单地y回归x_2, 不包含协变量x_1。最小二乘法得到回归直线为:

$$\boxed{y = \frac{1}{9}x_2 + \frac{4}{3}, \qquad (\hat{\sigma}^2 = 1.72).}$$

我们得出这样的结论:x_2**增加时，**y**也增加(**x_2**增加**$\Rightarrow$ y**增加)。**

在模型中包含或者不包含x_1，回归的结果是有差异的。y依赖于x_2的行为之所以存在这种差异，是因为协变量x_1和x_2具**有很强的相关性**。即，当x_2增加时，x_1也增加。

最小二乘估计

- 对于模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N(\mathbf{0}, \Sigma)$$

未知参数向量θ可以通过极小化误差平方和来实现，即：

$$\min_{\theta} \|\mathbf{Y} - \mathbf{X}\theta\|^2 = \min_{\theta} (\mathbf{Y} - \mathbf{X}\theta)^T (\mathbf{Y} - \mathbf{X}\theta)$$

$\Longrightarrow$

$$(\mathbf{X}^T\mathbf{X})\theta = \mathbf{X}^T\mathbf{Y}. \tag{5}$$

如果$\mathbf{X}^T\mathbf{X}$可逆，则$\Longrightarrow \widehat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. 这就是所谓的最小二乘方法【黑板】.
如果$\mathbf{X}^T\mathbf{X}$不可逆，则可以利用矩阵的广义逆得到

$$\Longrightarrow \widehat{\theta} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{Y}.$$

σ^2的估计：

当$\Sigma = \sigma^2 I_n$时，利用残差平方和可以估计σ^2:

$$\widehat{\sigma}^2 = \frac{1}{n-p}\|\mathbf{Y} - \mathbf{X}\widehat{\theta}\|^2$$

是σ^2的无偏估计【黑板】.

最小二乘估计的几何解释：

- 【黑板】

线性模型参数的MLE：

- 对于模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N(\mathbf{0}, \sigma^2 I_n)$$

未知参数向量θ和σ^2可以通过MLE来实现。
似然函数为：

$$\ell(\theta, \sigma^2 | \mathbf{X}, \mathbf{Y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\|\mathbf{Y} - \mathbf{X}\theta\|^2/(2\sigma^2)\right\}$$

$$\Longrightarrow \widehat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

这是θ的MLE【黑板】

σ^2的MLE：

σ^2的MLE为:

$$\widehat{\sigma}^2 = \frac{1}{n}\|\mathbf{Y} - \mathbf{X}\widehat{\theta}\|^2$$

是σ^2的MLE【黑板】.

线性模型参数的MLE：

- 对于模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N(\mathbf{0}, \Sigma^2)$$

其中，θ是未知参数向量，Σ是未知的协方差矩阵。

- 思考题：
如何求出θ和Σ的MLE？

进一步阅读内容：

关于线性模型参数的LSE和MLE的更多细节，

请仔细阅读：

G. A. F. Seber and A. J. Lee. (2003). Linear Regression Analysis. 2 nd Ed. p.35-57中的内容。

【第六讲结束】