MATH 363 Week 2 General Linear Model

GLM直观表示

> 形如：
>
> $$Y_i = \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i,\ i = 1, 2, \ldots, n$$
>
> 这样的就是 General Linear Model, 其中有 p 个自变量和 n 个因变量，且：
>
> 1. x_{ij} 是已知的 covariates
> 2. β_j 是未知的参数 parameters
> 3. ε_i 是随机误差 random errors
>
> 这些误差对于 $i \neq j$ 是独立的，并且 $\varepsilon_i \sim N(0, \sigma^2)$

GLM矩阵表示 Matrix Notation

> $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

其中，$\mathbf{Y}$ 是 $(n \times 1)$ 的因变量观测值矩阵，$\mathbf{X}$ 是 $(n \times p)$ 的自变量观测值矩阵（也可以被称为design matrix），$\boldsymbol{\beta}$ 是 $(p \times 1)$ 的未知数值的参数矩阵，$\boldsymbol{\varepsilon}$ 是 $(n \times 1)$ 的随机误差矩阵

关于 $\boldsymbol{\varepsilon}$，

1. 因为每一个随机误差都是独立同分布服从 $N(0, \sigma^2)$ 的随机变量，所以 $\boldsymbol{\varepsilon}$ 是服从多元正态分布的，且有 $E[\boldsymbol{\varepsilon}] = \vec{0}$ 以及方差矩阵 $Var[\boldsymbol{\varepsilon}] = \sigma^2 I_n$，其中，$I_n$ 是一个 $(n \times n)$ 的单位矩阵
2. 方差矩阵（方差-协方差矩阵）的元素 (i, j) 是 $Cov(X_i, X_j)$, 并且如果normal variables的协方差为0，则它们是独立的，反之对于其他分布则不能这么说

总之，

$$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 I_n)$$

如何估计GLM中的β

最小二乘法估计Least Squares estimation of β

我们必须要找到一个“最好的” $\hat{\beta}$ 来最小化误差平方和, 也就是:

$$S(\beta) = \sum_i^n \varepsilon_i^2 = \sum_i^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2$$

接下来对每一个 β_k 分别求偏导，就会得到 p 个这样的等式:

$$\frac{\partial S}{\partial \beta_k} = -2\sum_{i=1}^n x_{ik}[y_i - (\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip})]$$

对于每个 k, 我们都要有:

$$\sum_{i=1}^n x_{ik}[y_i - (\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip})] = 0$$

即,

> $$\sum_i^n x_{ik}y_i - \sum_i^n x_{ik}\underbrace{\sum_{j=1}^p x_{ij}\beta_p}_{x_i^T\boldsymbol{\beta}} = 0$$
>
> 也就是，
>
> $$\sum_i^n x_{ik}y_i - \sum_i^n x_{ik}x_i^T\boldsymbol{\beta} = 0$$
>
> 那么考虑到每一个 k，
>
> 我们则有：
>
> $$\mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}\beta = 0$$
>
> 假设 $\mathbf{X}^T\mathbf{X}$ 可逆，即有逆 $(\mathbf{X}^T\mathbf{X})^{-1}$，我们有：
>
> $$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Estimators估计量的性质

> Estimator是用来估计参数parameter的
>
> 1. Estimators是随机变量（因为不同的观测值会导致估计量值的变化），所以估计量有分布distribution
> 2. 无偏估计下， $E(\hat{\beta}) - \beta = 0$
> 3. 方差应该尽可能的小
> 4. 随着样本量增大，估计量的值会逐渐接近参数真实值

$\hat{\beta}$ 的分布（期望&方差）

> 线性Linear in $\hat{\beta}$:

$$\hat{\beta} = \underbrace{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}_{\text{常数矩阵}\mathbf{P}}\mathbf{Y} = \mathbf{P}\mathbf{Y}$$

$\hat{\beta}$ 的期望

由数学期望的线性性质，

$E(\hat{\beta}) = E(\mathbf{P}\mathbf{Y}) = \mathbf{P}E(\mathbf{Y}) = \mathbf{P}(\mathbf{X}\beta + E(\varepsilon)) = \mathbf{P}\mathbf{X}\beta = \beta$

这也同时说明 $\hat{\beta}$ 是对于参数 β 的无偏unbiased估计

$\hat{\beta}$ 的方差

方差的一个性质：若有常数矩阵 A, 则对于随机变量 Z:

$$Var(AZ) = AVar(Z)A^T$$

因此，结合方差的其他性质，我们有，

$Var(\hat{\beta}) = Var(\mathbf{P}\mathbf{Y}) = \mathbf{P}Var(\varepsilon)\mathbf{P}^T = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$

其中注意到，因为 $\mathbf{X}^T\mathbf{X}$ 是对称的 $(p \times p)$ 矩阵，所以它的逆矩阵也是对称的