第4章　参数估计

吴密霞

北京工业大学统计与数据科学系

E-mail: *wumixia@bjut.edu.cn*

教材

- 吴密霞, 王松桂. 2024. 线性模型引论（第2版）, 科学出版社.

线性模型

- 考虑线性模型：

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{4.1}$$

这里$\boldsymbol{y}$为$n \times 1$ 观测向量, $\mathbf{X}$为$n \times p$的设计矩阵,$\boldsymbol{\beta}$为$p \times 1$未知参数向量, $\boldsymbol{e}$为随机误差, 且满足 Gauss-Markov假设:

$$E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{e}) = \sigma^2 \mathbf{I}_n,$$

其中σ^2为误差方差, $\sigma^2 > 0$. 称模型为Gauss-Markov模型.

  - 当$\mathrm{rk}(\mathbf{X}) = r \leq p$时，称其为降秩线性模型
  - 当$\mathrm{rk}(\mathbf{X}) = p$ 时，称其为满秩线性模型

最小二乘法的思想

β的真值应该使误差向量$\boldsymbol{e} = \boldsymbol{y} - \mathbf{X}\beta$ 的长度平方

$$Q(\beta) = \|\boldsymbol{e}\|^2 = \|\boldsymbol{y} - \mathbf{X}\beta\|^2 = (\boldsymbol{y} - \mathbf{X}\beta)'(\boldsymbol{y} - \mathbf{X}\beta)$$

达到最小.

应该通过极小化$Q(\beta)$来求β的估计. 注意到

$$Q(\beta) = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta,$$

利用矩阵微商公式(见第二章)

$$\frac{\partial \boldsymbol{y}'\mathbf{X}\beta}{\partial \beta} = \mathbf{X}'\boldsymbol{y}, \qquad \frac{\partial \beta'\mathbf{X}'\mathbf{X}\beta}{\partial \beta} = 2\mathbf{X}'\mathbf{X}\beta,$$

于是

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\boldsymbol{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

令其等于0, 得到正则方程

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\boldsymbol{y}. \tag{4.2}$$

因为$\mathbf{X}'\boldsymbol{y} \in \mathcal{M}(\mathbf{X}') = \mathcal{M}(\mathbf{X}'\mathbf{X})$, 于是正则方程(4.2) 是相容的.

根据定理2.3.3, 正则方程(4.2)的解为

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{y},$$

这里$(\mathbf{X}'\mathbf{X})^{-}$是$\mathbf{X}'\mathbf{X}$的任意一个广义逆.

- 根据函数极值理论知：$\widehat{\beta}$只是函数$Q(\beta)$的驻点. 下证$\widehat{\beta}$使$Q(\beta)$达到最小. 事实上, 对任意一个β, 都有

$$
\begin{aligned}
Q(\beta) &= \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \\
&= \|\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2 \\
&= \|\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|^2 + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})
\end{aligned}
$$

由于第二项总是非负的, 于是

$$
Q(\boldsymbol{\beta}) \geq \|\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|^2 = Q(\widehat{\boldsymbol{\beta}}).
$$

此式表明, $\widehat{\beta}$确使$Q(\beta)$达到最小.

- 现证使$Q(\beta)$达到最小的必是$\widehat{\beta}$. 事实上,

$$Q(\beta) \geq \|\mathbf{y} - \mathbf{X}\widehat{\beta}\|^2 = Q(\widehat{\beta})$$

等号成立, 当且仅当

$$(\widehat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\widehat{\beta} - \beta) = \mathbf{0},$$

等价地

$$\mathbf{X}(\widehat{\beta} - \beta) = \mathbf{0}.$$

不难证明, 上式又等价于

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}'\mathbf{y},$$

这就证明了, 使$Q(\beta)$达到最小值的点必为正则方程的解

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}.$$

- 若rk$(\mathbf{X}) = p$, 则$\mathbf{X}'\mathbf{X}$可逆, $\boldsymbol{\beta}$可估,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

是$\boldsymbol{\beta}$的无偏估计, 称其为$\boldsymbol{\beta}$ 的最小二乘(least squares, LS)估计.

- 若rk$(\mathbf{X}) < p$, 则$\widehat{\boldsymbol{\beta}}$不再是$\boldsymbol{\beta}$的无偏估计,

$$E(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \neq \boldsymbol{\beta}.$$

此时$\boldsymbol{\beta}$的线性无偏估计根本不存在, 故称$\boldsymbol{\beta}$不可估.

反证法. 若存在$p \times n$矩阵$\mathbf{A}$, 使得对一切$\boldsymbol{\beta}$成立, 都有$E(\mathbf{A}\boldsymbol{y}) = \mathbf{A}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \Longleftrightarrow$ $\mathbf{A}\mathbf{X} = \mathbf{I}_p$, 这就与rk$(\mathbf{A}\mathbf{X}) \leq$ rk$(\mathbf{X}) < p =$ rk$(\mathbf{I}_p)$ 相矛盾.

可估函数

若$\text{rk}(\mathbf{X}) < p$, 则$\beta$不可估. 此时可考虑$\beta$的线性组合$\boldsymbol{c}'\beta$.

定义4.1.1

若存在$n \times 1$ 向量$\boldsymbol{a}$, 使得

$$E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{c}'\beta$$

对一切β 成立, 则称$\boldsymbol{c}'\beta$是可估函数(estimable function).

定理4.1.1

$\boldsymbol{c}'\beta$是可估函数$\Longleftrightarrow \boldsymbol{c} \in \mathcal{M}(\mathbf{X}')$.

证明 $\boldsymbol{c}'\beta$ 是可估函数$\Longleftrightarrow$存在$\boldsymbol{a}_{n\times 1}$, 使得对一切$\beta$都有$E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{c}'\beta$

$$\Longleftrightarrow \boldsymbol{a}'\mathbf{X}\beta = \boldsymbol{c}'\beta, \text{对一切}\beta \Longleftrightarrow \boldsymbol{c} = \mathbf{X}'\boldsymbol{a}.$$

定理4.1.1表明:

- 使$\boldsymbol{c}'\boldsymbol{\beta}$可估的全体$p \times 1$向量$\boldsymbol{c}$构成子空间$\mathcal{M}(\mathbf{X}')$;
- 若$\boldsymbol{c}_1'\boldsymbol{\beta}$和$\boldsymbol{c}_2'\boldsymbol{\beta}$均可估, 则任意线性组合$\alpha_1\boldsymbol{c}_1'\boldsymbol{\beta} + \alpha_2\boldsymbol{c}_2'\boldsymbol{\beta}$也可估.
- 若$\boldsymbol{c}_1$和$\boldsymbol{c}_2$为线性无关, 则称可估函数$\boldsymbol{c}_1'\boldsymbol{\beta}$ 和$\boldsymbol{c}_2'\boldsymbol{\beta}$ 是线性无关的.
- 线性无关的可估函数组最多含有rk$(\mathbf{X}) = r$ 个可估函数.
- 对于任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$, $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$与$(\mathbf{X}'\mathbf{X})^-$的选择无关, 是唯一的,

$$\boldsymbol{c}'\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\boldsymbol{y} = \boldsymbol{a}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\boldsymbol{y} = \boldsymbol{a}'\mathbf{X}(\mathbf{X}'\mathbf{X})^+\mathbf{X}'\boldsymbol{y}$$

- $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$为$\boldsymbol{c}'\boldsymbol{\beta}$ 的无偏估计, $E(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}) = \boldsymbol{a}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{a}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{c}'\boldsymbol{\beta}$.

定义4.1.2

对可估函数$\boldsymbol{c}'\beta$, 称$\boldsymbol{c}'\widehat{\beta}$为$\boldsymbol{c}'\beta$的LS估计.

- 记$\mathbf{C} = (\boldsymbol{c}_1, \cdots, \boldsymbol{c}_m)'$ 是一个$m \times p$的常数矩阵. 若$\boldsymbol{c}_1'\beta, \cdots, \boldsymbol{c}_m'\beta$ 皆可估, 则称$\mathbf{C}\beta$可估. 易证

$$\mathbf{C}\beta\text{可估} \Longleftrightarrow \mathcal{M}(\mathbf{C}') \subseteq \mathcal{M}(\mathbf{X}').$$

- 若$\mathbf{C}\beta$ 可估, 则$\mathbf{C}\beta$ 的LS估计为$\mathbf{C}\widehat{\beta}$.

记 $\mathbf{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)', \quad \boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)' = \mathbf{X}\beta.$

Guass-Markov模型的分量形式:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, \qquad i = 1, \cdots, n,$$

其中，$\boldsymbol{x}_i$为$\mathbf{X}$的第i行向量, e_i为误差向量的第i个分量,

$$E(e_i) = 0, \quad \mathrm{Cov}(e_i) = \sigma^2, \quad \mathrm{Cov}(e_i, e_j) = 0, (i \neq j).$$

- $\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta}$, $i = 1, \cdots, \mu_n$ 是n个可估函数;
- 任意可估函数都是$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ 的线性函数，即 $\boldsymbol{c}'\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{\mu}$
- $\boldsymbol{\mu}$的LS估计为$\widehat{\boldsymbol{\mu}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{y} = \mathsf{P}_{\mathbf{X}}\,\boldsymbol{y}$.

σ^2的LS估计

- 残差向量

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \mathbf{X}\widehat{\beta} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\boldsymbol{y},$$

作为误差向量的一个“估计”，它对研究关于误差假设的合理性起重要作用. 易证: $E(\widehat{\boldsymbol{e}}) = \mathbf{0}, \quad \mathrm{Cov}(\widehat{\boldsymbol{e}}) = \sigma^2(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})$.

- 基于 $\widehat{\boldsymbol{e}}$ 可构造 σ^2 的估计

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{n-r} = \frac{\|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2}{n-r},$$

这里$r = \mathrm{rk}(\mathbf{X})$. 通常称$\widehat{\sigma}^2$为$\sigma^2$的LS估计.

定理4.1.3

$\widehat{\sigma}^2$ 是σ^2 的无偏估计.

证明 因$\mathbf{I}_n - \mathbf{P_X}$为幂等阵, 于是

$$\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} = \boldsymbol{y}'(\mathbf{I}_n - \mathbf{P_X})\boldsymbol{y},$$

利用定理3.2.1和等式$(\mathbf{I}_n - \mathbf{P_X})\mathbf{X} = \mathbf{0}$, 可证得

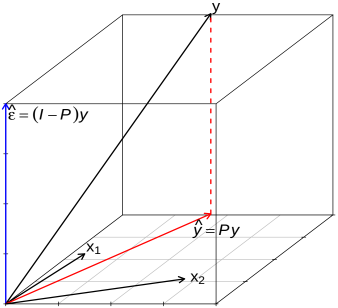
$$E(\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}) = (\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \mathbf{P_X})\mathbf{X}\boldsymbol{\beta} + \mathrm{tr}(\mathbf{I}_n - \mathbf{P_X})\mathrm{Cov}(\boldsymbol{y}) = \sigma^2\mathrm{tr}(\mathbf{I}_n - \mathbf{P_X}),$$

再利用迹和幂等阵的性质, 得

$$E(\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}) = \sigma^2[n - \mathrm{tr}(\mathbf{P_X})] = \sigma^2[n - \mathrm{rk}(\mathbf{X})].$$

定理得证.

LS估计$\mathbf{X}\widehat{\beta}$和残差向量$\widehat{\boldsymbol{e}}$分别为$\boldsymbol{y}$向$\mathcal{M}(\mathbf{X}')$和$\mathcal{M}(\mathbf{X}')$的正交补空间上的正交投影, 如图所示:



- 残差平方和 $\mathrm{SS}_e = \widehat{e}'\widehat{e}$可被表达为

$$\mathrm{SS}_e = \|y - \mathbf{X}\widehat{\beta}\|^2 = y'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})y = e'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})e$$

- 对任一可估函数$c'\beta$的LS估计 唯一, 因为

$$c'\widehat{\beta} = a'\mathbf{X}\widehat{\beta} = a'\mathbf{P}_{\mathbf{X}}y = a'\mathbf{X}(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'y$$

- 可估函数$c'\beta$的线性无偏估计有无穷多个.

事实上, 如果$a'y$为$c'\beta$的一个无偏估计，$b'\mathbf{X} = 0$, 则

$$E(a+b)'y = E(a'y) + E(b'y) = c'\beta + b'\mathbf{X}'\beta = c'\beta$$

Gauss-Markov定理

对任意的可估函数$\boldsymbol{c}'\boldsymbol{\beta}$, LS估计$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$为其惟一的最佳线性无偏(best linear unbiased, BLU)估计, 即在其线性无偏估计当中, $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$的方差最小.

证明 显然$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$为$\boldsymbol{c}'\boldsymbol{\beta}$的线性无偏估计. 现证$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$的方差最小.

$$\text{Var}(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}) = \text{Var}(\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{y}) = \sigma^2\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{c}.$$

由$\boldsymbol{c}'\boldsymbol{\beta}$的可估性, 知存在向量$\boldsymbol{\alpha}_{n\times 1}$, 使得$\boldsymbol{c} = \mathbf{X}'\boldsymbol{\alpha}$, 于是,

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{c} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\alpha} = \mathbf{X}'\boldsymbol{\alpha} = \boldsymbol{c},$$

从而得$\text{Var}(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{c}$.

另一方面, 设$\boldsymbol{a}'\boldsymbol{y}$为$\boldsymbol{c}'\beta$的任一无偏估计, 于是有$\mathbf{X}'\boldsymbol{a} = \boldsymbol{c}$. 从而

$$
\begin{aligned}
\mathsf{Var}(\boldsymbol{a}'\boldsymbol{y}) - \mathsf{Var}(\boldsymbol{c}'\widehat{\beta}) &= \sigma^2[\boldsymbol{a}'\boldsymbol{a} - \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{c}] \\
&= \sigma^2(\boldsymbol{a}' - \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}')(\boldsymbol{a} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{c}) \\
&= \sigma^2\|\boldsymbol{a} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{c}\|^2 \geq 0,
\end{aligned}
$$

并且等号成立$\Longleftrightarrow \boldsymbol{a}' = \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}X' \Longleftrightarrow \boldsymbol{a}'\boldsymbol{y} = \boldsymbol{c}'\widehat{\beta}$. 定理证毕.

这个定理奠定了LS估计在线性模型参数估计理论中的地位.

注 它所刻画的LS估计在线性无偏估计类中的最优性. 但在另一个度量估计优劣的标准（如MSE）, 它可能并不一定最优

推论4.1.1

设$\psi_i = \boldsymbol{c}_i'\boldsymbol{\beta}$, $i = 1, \cdots, k$都是可估函数, $\alpha_i, i = 1, \cdots, k$ 是实数, 则$\psi = \sum_{i=1}^{k} \alpha_i \psi_i$也可估, 且$\widehat{\psi} = \sum_{i=1}^{k} \alpha_i \widehat{\psi}_i = \sum_{i=1}^{k} \alpha_i \boldsymbol{c}'\widehat{\boldsymbol{\beta}}$ 是ψ的BLU估计.

推论4.1.2

设$\boldsymbol{c}'\boldsymbol{\beta}$和$\boldsymbol{d}'\boldsymbol{\beta}$是两个可估函数,则

$$\mathrm{Var}(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^- \boldsymbol{c},$$

$$\mathrm{Cov}(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}, \boldsymbol{d}'\widehat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^- \boldsymbol{d},$$

并且上述两式与所含广义逆的选择无关.

对于可估函数$\mathbf{C}\beta$, 其线性无偏估计类可表示为

$$\mathcal{D} = \{\mathbf{A}y;\ \mathbf{A}\text{满足条件}\mathbf{A}\mathbf{X} = \mathbf{C}\}.$$

注意到 $\mathrm{Cov}(\mathbf{A}y) = \sigma^2\mathbf{A}\mathbf{A}'$, $\mathrm{cov}(\mathbf{C}\widehat{\beta}) = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}'$,

$$\mathrm{cov}(\mathbf{A}y) - \mathrm{cov}(\mathbf{C}\widehat{\beta}) = \sigma^2\mathbf{A}\left(\mathbf{I}_p - \mathbf{X}(\mathbf{X}'\mathbf{X})^-)\mathbf{X}'\right)\mathbf{A}' \geq \mathbf{0}.$$（半正定）

定理4.1.4

对于可估函数$\mathbf{C}\beta$, LS估计$\mathbf{C}\widehat{\beta}$为其唯一的最小协方差阵线性无偏（minimum dispersion linear unbiased, MDLU）估计.

- 若$\mathbf{A}y$为可估函数$\mathbf{C}\beta$的MDLU估计, 则其分量也是MVLU的.

定理4.1.5

对正态线性模型

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \quad \boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

设$\boldsymbol{c}'\boldsymbol{\beta}$为任一可估函数, 则

(1) LS估计$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$是$\boldsymbol{c}'\boldsymbol{\beta}$的极大似然(maximum likelihood, ML)估计, 且

$$\boldsymbol{c}'\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{c}'\boldsymbol{\beta},\ \sigma^2\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{c});$$

(2) $(n-r)\widehat{\sigma}^2/n$为$\sigma^2$ 的ML估计, 且$(n-r)\widehat{\sigma}^2/\sigma^2 \sim \chi^2_{n-r}$;

(3) $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$ 与 $\widehat{\sigma}^2$ 相互独立, 这里 $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\boldsymbol{y}$, $r = \mathrm{rk}(\mathbf{X})$.

证明 (1) 记$\boldsymbol{\mu} = \mathbf{X}\beta$,考虑$\boldsymbol{\mu}$和$\sigma^2$的似然函数

$$L(\boldsymbol{\mu}, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{\mu}\|^2\right\},$$

取对数, 略去常数项, 得

$$\log L(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{\mu}\|^2.$$

对均值向量$\boldsymbol{\mu}$的LS估计$\widehat{\boldsymbol{\mu}} = \mathbf{X}\widehat{\beta}$, 我们有

$$\|\boldsymbol{y} - \widehat{\boldsymbol{\mu}}\|^2 = \|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2 = \min\|\boldsymbol{y} - \mathbf{X}\beta\|^2 = \min_{\boldsymbol{\mu}=\mathbf{X}\beta}\|\boldsymbol{y} - \boldsymbol{\mu}\|^2.$$

于是, 对每一个固定的σ^2,

$$\log L(\widehat{\boldsymbol{\mu}}, \sigma^2) \geq \log L(\boldsymbol{\mu}, \sigma^2),$$

而

$$\log L(\widehat{\boldsymbol{\mu}}, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|\boldsymbol{y} - \widehat{\boldsymbol{\mu}}\|^2,$$

在$\tilde{\sigma}^2 = \frac{1}{n}\|\boldsymbol{y} - \widehat{\boldsymbol{\mu}}\|^2$ 达到最大. 于是

$$\widehat{\boldsymbol{\mu}} = \mathbf{X}\widehat{\boldsymbol{\beta}}, \quad \tilde{\sigma}^2 = \frac{1}{n}\|\boldsymbol{y} - \widehat{\boldsymbol{\mu}}\|^2$$

分别为$\boldsymbol{\mu}$ 和σ^2 的ML估计.

对任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$, 存在$\boldsymbol{\alpha} \in R^n$, 使得$\boldsymbol{c}'\boldsymbol{\beta} = \boldsymbol{\alpha}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\alpha}'\boldsymbol{\mu}$, 由ML估计的不变性, $\boldsymbol{c}'\boldsymbol{\beta}$的ML估计为

$$\boldsymbol{\alpha}'\widehat{\boldsymbol{\mu}} = \boldsymbol{\alpha}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'\widehat{\boldsymbol{\beta}}.$$

又因$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$为$\boldsymbol{y}$的线性函数, 而$\boldsymbol{y} \sim N_n(\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{I}_n)$, 依定理3.3.4知,

$$\boldsymbol{c}'\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta},\ \sigma^2\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{c}).$$

由$\boldsymbol{c}'\boldsymbol{\beta}$可估, 得$\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \boldsymbol{c}'$, (1)得证.

(2) 的第一条结论已证. 因为$\mathsf{P}_{\mathbf{X}}\mathbf{X} = \mathbf{X}$, 所以

$$\frac{(n-r)\widehat{\sigma}^2}{\sigma^2} = \frac{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{\sigma^2} = \frac{\boldsymbol{y}'(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}})\boldsymbol{y}}{\sigma^2} = \frac{\boldsymbol{e}'(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}})\boldsymbol{e}}{\sigma^2} = \boldsymbol{z}'(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}})\boldsymbol{z},$$

其中$\boldsymbol{z} = \boldsymbol{e}/\sigma \sim N_n(\mathbf{0},\ \mathbf{I}_n)$. 由$\mathbf{I}_n - \mathsf{P}_{\mathbf{X}}$的幂等性及

$$\mathrm{rk}(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}}) = \mathrm{tr}(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}}) = n - \mathrm{tr}(\mathsf{P}_{\mathbf{X}}) = n - \mathrm{rk}(\mathbf{X}) = n - r,$$

利用定理3.4.3, 即得$(n-r)\widehat{\sigma}^2/\sigma^2 \sim \chi^2_{n-r}$.

定理4.1.5的证明

(3) 注意到$c'\widehat{\beta}$与$\widehat{\sigma}^2$ 分别为正态向量y的线性型和二次型, 和

$$c'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}}) = \mathbf{0},$$

由正态向量的线性型和二次型独立性定理3.5.1 推得(3).

LS估计的最优性质

定理4.1.6

对于正态线性模型$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$, $\boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$,

(1) $T_1 = \boldsymbol{y}'\boldsymbol{y}$和$\boldsymbol{T}_2 = \mathbf{X}'\boldsymbol{y}$为完全充分统计量,

(2) 对任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$, $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$为其惟一的最小方差无偏(minimum variance unbiased, MVU)估计, $\widehat{\sigma}^2$为σ^2的惟一MVU估计.

证明 观测向量$\boldsymbol{y}$的概率密度函数为

$$
\begin{aligned}
f(y) &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})\right\} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{y}'\boldsymbol{y} + \frac{1}{\sigma^2}\boldsymbol{y}'\mathbf{X}\boldsymbol{\beta} - \frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right\},
\end{aligned}
$$

记$\theta_1 = -\frac{1}{2\sigma^2}$, $\boldsymbol{\theta}_2 = \frac{\boldsymbol{\beta}}{\sigma^2}$,它们是所谓的自然参数, 则上式可改写为

$$f(\mathbf{y}) = \frac{1}{(-\pi)^{\frac{n}{2}}}\theta_1^n \exp\left\{\frac{1}{4\theta_1}\boldsymbol{\theta}_2'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}_2\right\}\exp\{\theta_1 T_1 + \boldsymbol{\theta}_2 \boldsymbol{T}_2\}.$$

将$f(\mathbf{y})$表成了指数族的自然形式，其参数空间为

$$\Theta = \left\{\begin{pmatrix}\theta_1\\ \boldsymbol{\theta}_2\end{pmatrix}; \quad \theta_1 < 0, \quad \boldsymbol{\theta}_2 \in R^p\right\}.$$

依定理2.2 (陈希孺, 1999)知:

$$T_1 = \mathbf{y}'\mathbf{y}\text{和}\boldsymbol{T}_2 = \mathbf{X}'\mathbf{y}\text{为完全充分统计量}.$$

对任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$,其LS估计

$$\boldsymbol{c}'\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{T}_2,$$

误差方差σ^2的LS估计

$$\widehat{\sigma}^2 = (T_1 - \boldsymbol{T}_2(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{T}_2)/(n-r),$$

它们都是完全充分统计量的函数. 同时我们知道, 它们都是无偏估计, 依Lehmann-Scheffe定理立得$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$和$\widehat{\sigma}^2$分别是$\boldsymbol{c}'\boldsymbol{\beta}$和$\sigma^2$的惟一MVU估计. 定理证毕.

线性模型$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$下, 可估函数$\boldsymbol{c}'\boldsymbol{\beta}$的LS 估计$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$的性质

- MVU性:

若$\boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ (正态线性模型), 则任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$的LS估计$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$在 其所有的(线性的和非线性)无偏估计类中方差最小.

- BLU性:

若$E(\boldsymbol{e}) = \mathbf{0}$, $\mathrm{cov}(\boldsymbol{e}) = \sigma^2\mathbf{I}_n$ （Guass-Markov模型下）, 则可估函数$\boldsymbol{c}'\boldsymbol{\beta}$的LS估计$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$在 线性无偏估计类中方差最小性.

定理4.1.7

对于正态线性模型$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$, $\boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$, 其任意可估函数$\mathbf{C}\boldsymbol{\beta}$的ML估计$\mathbf{C}\widehat{\boldsymbol{\beta}}$为其无偏估计类中协方差阵最小(minimum dispersion unbiased, MDU)估计. 当$\text{rk}(\mathbf{X}) = p$时, $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$的协方差阵正好等于Cramér-Rao 下界

$$-\left[\frac{\partial^2 \ln L(\boldsymbol{\mu}, \sigma^2)}{\partial^2 \boldsymbol{\beta}}\right]^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1},$$

这里$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$.

例子 设μ为一物体的重量, 现对该物体测量n次, 其测量值记为$y_1, \cdots, y_n$. 若$y_i \sim N(\mu, \sigma^2)$, 则$\bar{y} = \sum y_i/n$为$\mu$MDU估计.

分块LS估计

线性模型的分块运算主要应用于以下两种情形：

- 对参数分块

  当模型往往包含冗余参数（nuisance parameter）时, 感兴趣参数的LS估计的降秩表达

- 对样本分块

  Meta分析或大规模数据中, 由于隐私保护或计算机存储运算能力的限制, 多地存储的子数据集不能被传输到一台中央机器来分析，需要分块运算.

对线性模型$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$ 参数分块

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}, \quad \mathrm{E}(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{cov}(\boldsymbol{e}) = \sigma^2\mathbf{I}_n.$$

$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ 的LS解

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}^{-} \begin{pmatrix} \mathbf{X}_1'\boldsymbol{y} \\ \mathbf{X}_2'\boldsymbol{y} \end{pmatrix}.$$

- 当$\mathrm{rk}(\mathbf{X}) = p$时, $\boldsymbol{\beta}_1$和$\boldsymbol{\beta}_2$皆可估, 其LS估计为

$$\widehat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\boldsymbol{y}, \quad \widehat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'(\boldsymbol{y} - \mathbf{X}_2\widehat{\boldsymbol{\beta}}_2),$$

这里 $\mathbf{N}_{\mathbf{X}_1} = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1{}'\mathbf{X}_1)^{-}\mathbf{X}_1{}'$.

- 假定$\mathbf{X}_2$是列满秩, 并且$\mathbf{X}_1$的列与$\mathbf{X}_2$的列线性无关, 即

$$\mathcal{M}(\mathbf{X}_1) \cap \mathcal{M}(\mathbf{X}_2) = \{\mathbf{0}\}, \quad \mathrm{rk}(\mathbf{X}_2) = p_2.$$

定理4.2.2

$$\widehat{\beta}_2 = (\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\boldsymbol{y}, \quad \mathrm{cov}(\widehat{\beta}_2) = \sigma^2(\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}$$

$$\widehat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'(\boldsymbol{y} - \mathbf{X}_2\widehat{\beta}_2)$$

对任一可估函数$\boldsymbol{c}'\beta_1$, $\mathrm{var}(\boldsymbol{c}'\widehat{\beta}_1) = \sigma^2\boldsymbol{c}'\mathbf{M}\boldsymbol{c}$, 这里,

$$\mathbf{M} = (\mathbf{X}_1'\mathbf{X}_1)^{-} + (\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\mathbf{X}_2\mathbf{X}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-}.$$

参数分块的LS估计的直观解释

- $\widehat{\beta}_2 = (\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{N}_{\mathbf{X}_1}\boldsymbol{y}$ 为约简模型

$$\mathbf{N}_{\mathbf{X}_1}\boldsymbol{y} = \mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2\beta_2 + \varepsilon \tag{4.3}$$

  的LS估计, 其中$\varepsilon = \mathbf{N}_{\mathbf{X}_1}\boldsymbol{e}$.

- $\widehat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'(\boldsymbol{y} - \mathbf{X}_2\widehat{\beta}_2)$ 为将$\widehat{\beta}_2$ 替换原模型 $\boldsymbol{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \boldsymbol{e}$ 的β_2后所得模型

$$\tilde{\boldsymbol{y}} = \mathbf{X}_1\beta_1 + \boldsymbol{e} \tag{4.4}$$

  的LS解，其中 $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{X}_2\widehat{\beta}_2$.

如果$\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, 则

$$\mathbf{N}_{\mathbf{X}_1}\mathbf{X}_2 = (\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1})\mathbf{X}_2 = \mathbf{X}_2.$$

推论4.2.1

假设模型$\boldsymbol{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \boldsymbol{e}$满足$\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, $\mathrm{rk}(\mathbf{X}_2) = p_2$, 则可估函数$\boldsymbol{c}'\beta_1$和$\beta_2$的LS估计为

$$\boldsymbol{c}'\widehat{\beta}_1 = \boldsymbol{c}'(\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'\boldsymbol{y}, \quad \widehat{\beta}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\boldsymbol{y},$$

且$\mathrm{var}(\boldsymbol{c}'\widehat{\beta}_1) = \sigma^2\boldsymbol{c}'(\mathbf{X}_1'\mathbf{X}_1)^{-}\boldsymbol{c}'$, $\mathrm{cov}(\widehat{\beta}_2) = \sigma^2(\mathbf{X}_2'\mathbf{X}_2)^{-1}$.

对于多分块模型：

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \cdots + \mathbf{X}_k\boldsymbol{\beta}_k + \boldsymbol{e}, \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{cov}(\boldsymbol{e}) = \sigma^2\mathbf{I}_n,$$

当$\mathbf{X}_i$ 列满秩，且$\mathbf{X}_i'\mathbf{X}_j = \mathbf{0}$, $(i \neq j)$ 时, $(\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_k')'$的LS估计为

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \vdots \\ \widehat{\boldsymbol{\beta}}_k \end{pmatrix} = \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{y} \\ \vdots \\ (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'\boldsymbol{y} \end{pmatrix}$$

此时$\boldsymbol{\beta}_i$的LS估计等价于子模型

$$\boldsymbol{y} = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{e}$$

的的LS估计.

例4.2.1 考虑线性回归模型

$$\boldsymbol{y} = \mathbf{1}_n\beta_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e}, \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \operatorname{cov}(\boldsymbol{e}) = \sigma^2\mathbf{I}_n,$$

其中β_0为截距项, $\boldsymbol{\beta}$为回归系数. β_0 和$\boldsymbol{\beta}_1$的LS估计

$$\widehat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'/n)\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'/n)\boldsymbol{y} = (\mathbf{X}_C'\mathbf{X}_C)^{-1}\mathbf{X}_C'\boldsymbol{y},$$

$$\widehat{\beta}_0 = \mathbf{1}_n'(\boldsymbol{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}}_1)/n = \bar{y} - \bar{\boldsymbol{x}}_1'\widehat{\boldsymbol{\beta}}_1,$$

这里$\mathbf{X}_C = (\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'/n)\mathbf{X}_1$ 为$\mathbf{X}_1$的（列）中心化后的矩阵.

中心化的线性模型

$$\boldsymbol{y} = \mathbf{1}_n\alpha + \mathbf{X}_C\boldsymbol{\beta}_1 + \boldsymbol{e}, \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \operatorname{cov}(\boldsymbol{e}) = \sigma^2\mathbf{I}_n,$$

这里, $\alpha = \beta_0 + \bar{\boldsymbol{x}}'\boldsymbol{\beta}_1$.

因为 $\mathbf{1}_n'\mathbf{X}_C = \mathbf{0}$, 于是$\alpha$ 和$\boldsymbol{\beta}_1$ 的LS估计

$$\widehat{\alpha} = \bar{y},$$

$$\widehat{\boldsymbol{\beta}}_1 = (\mathbf{X}_C'\mathbf{X}_C)^{-1}\mathbf{X}_C'\boldsymbol{y}.$$

注 $\widehat{\boldsymbol{\beta}}_1$ 中心化的线性模型与非中心化线性模型下回归系数$\boldsymbol{\beta}_1$的LS估计相同, 皆为$\widehat{\boldsymbol{\beta}}_1$.

- 对样本分块

假设观测数据被分块存储在K台机器上, 全模型可被分块为

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_K \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_K \end{pmatrix}. \tag{4.5}$$

假设$\mathrm{rk}(\mathbf{X}_k) = p$, 则全模型$\boldsymbol{\beta}$的LS估计为

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{k=1}^{K} \mathbf{X}_k' \mathbf{X}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k' \boldsymbol{y}_k$$

假设观测数据被分块存储在K台机器上，全模型可被分块为

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_K \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{pmatrix} \beta + \begin{pmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_K \end{pmatrix}. \tag{4.6}$$

假设$\mathrm{rk}(\mathbf{X}_k) = p$，则全模型$\beta$的LS估计为

$$\hat{\beta} = \left(\sum_{k=1}^{K} \mathbf{X}_k' \mathbf{X}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k' \boldsymbol{y}_k$$

子模型

$$\boldsymbol{y}_k = \mathbf{X}_k \beta_k + \boldsymbol{e}_k, \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{cov}(\boldsymbol{e}) = \sigma^2 \mathbf{I}_{n_k} \tag{4.7}$$

的LS估计$\hat{\beta}_k = (\mathbf{X}_k' \mathbf{X}_k)^{-1} \mathbf{X}_k' \boldsymbol{y}_k$.

记$\mathbf{V}_k = \mathbf{X}_k'\mathbf{X}_k$. 注意到

$$\widehat{\beta} = \left(\sum_{k=1}^{K} \mathbf{X}_k'\mathbf{X}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k'\boldsymbol{y}_k = \left(\sum_{k=1}^{K} \mathbf{V}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{V}_k\hat{\beta}_k,$$

于是, 在计算全模型β的LS估计时，各台机器只需传输子模型的LS估计$\hat{\beta}_k$和矩阵$\mathbf{V}_k$到中央机器上即可，无需传输原始数据.

关于分布式统计推断的研究综述，参阅

Gao, Y., Liu, W. D., Wang, H. S., Wang, X. Z., X. Z., Yan, Y. B, Zhang, R. Q. A review of distributed statistical inference. Statistical Theory and Related Fields, 2022, 6: 89-99.

线性模型

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \text{cov}(\boldsymbol{e}) = \sigma^2 \mathbf{I}_n,$$

约束条件：

$$\mathbf{H}\boldsymbol{\beta} = \boldsymbol{d}, \quad \text{即 } \boldsymbol{h}_i'\boldsymbol{\beta} = d_i, \quad i = 1, \cdots, k.$$

其中rk$(\mathbf{H}) = k$, $\mathcal{M}(\mathbf{H}') \subset \mathcal{M}(\mathbf{X}')$, $\boldsymbol{d} \in \mathcal{M}(\mathbf{H})$,

$$\mathbf{H} = \begin{pmatrix} \boldsymbol{h}_1' \\ \vdots \\ \boldsymbol{h}_k' \end{pmatrix}, \qquad \boldsymbol{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix}. \tag{4.8}$$

问题: 在k个条件下求$\boldsymbol{\beta}$使$Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2$达到最小值

为了应用Lagrange乘子法, 构造 辅助函数

$$\begin{aligned} F(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + 2\sum_{i=1}^{k} \lambda_i (\boldsymbol{h}_i' \boldsymbol{\beta} - d_i) \\ &= \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + 2\boldsymbol{\lambda}'(\mathbf{H}\boldsymbol{\beta} - \boldsymbol{d}) \\ &= (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}'(\mathbf{H}\boldsymbol{\beta} - \boldsymbol{d}), \end{aligned}$$

其中$\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)'$为Lagrange乘子. 令$\frac{\partial F(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \mathbf{0}$, 得

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\boldsymbol{y} - \mathbf{H}'\boldsymbol{\lambda}. \tag{4.9}$$

因为$\mathcal{M}(\mathbf{H}') \subset \mathcal{M}(\mathbf{X}')$, 所以上方程关于$\boldsymbol{\beta}$是相容的, 其解

$$\widehat{\boldsymbol{\beta}}_{\mathbf{H}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{y} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'\widehat{\boldsymbol{\lambda}}_{\mathbf{H}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'\widehat{\boldsymbol{\lambda}}_{\mathbf{H}}.$$

代入$\mathbf{H}\beta = \boldsymbol{d}$, 得$\boldsymbol{d} = \mathbf{H}\widehat{\beta}_{\mathbf{H}} = \mathbf{H}\widehat{\beta} - \mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'\widehat{\boldsymbol{\lambda}}_{\mathbf{H}}$, 等价地

$$\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'\widehat{\boldsymbol{\lambda}}_{\mathbf{H}} = (\mathbf{H}\widehat{\beta} - \boldsymbol{d}). \tag{4.10}$$

这是一个关于$\widehat{\boldsymbol{\lambda}}_{\mathbf{H}}$的线性方程组. 因为$\mathbf{H}$ 的秩为k, 且$\mathcal{M}(\mathbf{H}') \subset \mathcal{M}(\mathbf{X}')$, 于是$\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'$跟所包含广义逆的选择无关. 故可证它是$k \times k$的可逆矩阵, 因而(4.10)有唯一解

$$\widehat{\boldsymbol{\lambda}}_{\mathbf{H}} = (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{H}\widehat{\beta} - \boldsymbol{d}).$$

将$\widehat{\boldsymbol{\lambda}}_{\mathbf{H}}$代入(4.9)得到

$$\widehat{\beta}_{\mathbf{H}} = \widehat{\beta} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{H}\widehat{\beta} - \boldsymbol{d}).$$

约束最小二乘估计的性质

约束最小二乘估计$\widehat{\beta}_{\mathbf{H}}$ 为是线性约束$\mathbf{H}\beta = \boldsymbol{d}$ 下β 的最小二乘解, 即

(a) $\mathbf{H}\widehat{\beta}_{\mathbf{H}} = \boldsymbol{d}$;

(b) 对一切满足$\mathbf{H}\beta = \boldsymbol{d}$ 的β, 都有$\|\boldsymbol{y} - \mathbf{X}\beta\|^2 \geq \|\boldsymbol{y} - \mathbf{X}\widehat{\beta}_{\mathbf{H}}\|^2$.

证明 (a)显然成立. (b)的证明关键:

$$(\boldsymbol{y} - \mathbf{X}\widehat{\beta})'\mathbf{X}(\widehat{\beta} - \beta) = \boldsymbol{y}(\mathbf{I}_n - \mathsf{P}_{\mathbf{X}})\mathbf{X}(\widehat{\beta} - \beta) = \mathbf{0},$$

$$(\widehat{\beta} - \widehat{\beta}_{\mathbf{H}})'\mathbf{X}'\mathbf{X}(\widehat{\beta}_{\mathbf{H}} - \beta) = \widehat{\boldsymbol{\lambda}}'(H\widehat{\beta}_{\mathbf{H}} - \mathbf{H}\beta) = \widehat{\boldsymbol{\lambda}}'(\boldsymbol{d} - \boldsymbol{d}) = 0$$

$$\begin{aligned}
\Longrightarrow \|\boldsymbol{y} - \mathbf{X}\beta\|^2 &= \|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2 + (\widehat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\widehat{\beta} - \beta) \\
&= \|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2 + \|\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{\mathbf{H}})\|^2 + \|\mathbf{X}(\widehat{\beta}_{\mathbf{H}} - \beta)\|^2 \\
&\geq \|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2 + \|\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{\mathbf{H}})\|^2.
\end{aligned}$$

定理4.3.1

对于线性模型，设$\mathbf{H}$为$k\times p$矩阵，rk$(\mathbf{H})=k$，$\mathcal{M}(\mathbf{H}')\subset\mathcal{M}(\mathbf{X}')$，且$\mathbf{H}\beta=\boldsymbol{d}$ 相容(即$\boldsymbol{d}\in\mathcal{M}(\mathbf{H})$)，则

(1) $\widehat{\beta}_{\mathbf{H}}=\widehat{\beta}-(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{H}\widehat{\beta}-\boldsymbol{d})$ 为β在线性约束条件$\mathbf{H}\beta=\boldsymbol{d}$下的约束LS解，$\mathbf{H}\widehat{\beta}_{\mathbf{H}}$为$\mathbf{H}\beta$的约束LS估计，这里$\widehat{\beta}=(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{y}$.

(2) 若rk$(\mathbf{X})=p$，则

$$\widehat{\beta}_{\mathbf{H}}=\widehat{\beta}-(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\widehat{\beta}-\boldsymbol{d})$$

为β的约束LS估计，这里$\widehat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$.

例4.3.1 在天文测量中, 对天空中三个星位点构成$\triangle ABC$的三个内角$\theta_1, \theta_2, \theta_3$进行测量. 用线性模型表示:

$$\begin{cases} y_1 = \theta_1 + e_1, \\ y_2 = \theta_2 + e_2, \\ y_3 = \theta_3 + e_3, \\ \theta_1 + \theta_2 + \theta_3 = \pi, \end{cases}$$

其中$e_i, i = 1, 2, 3$ 表示测量误差. 假设它们满足Guass-Markov假设, 这就是一个带有约束条件的线性模型. 令

$\boldsymbol{y} = (y_1, y_2, y_3)'$, $\boldsymbol{\beta} = (\theta_1, \theta_2, \theta_3)'$, $\mathbf{X} = \mathbf{I}_3$, $\mathbf{H} = (1, 1, 1)$, $b = \pi$.

θ_i的约束最小二乘估计为

$$\widehat{\boldsymbol{\theta}}_i = y_i - \frac{1}{3}(y_1 + y_2 + y_3 - \pi), \ i = 1, 2, 3.$$

定理4.3.2

在定理4.3.1假设下, 在参数区域$\mathbf{H}\beta = \boldsymbol{d}$上,

$$\widehat{\sigma}^2_{\mathbf{H}} = \|\boldsymbol{y} - \mathbf{X}\widehat{\beta}_{\mathbf{H}}\|^2/(n-r+k)$$

是σ^2的无偏估计.

证明 注意到$E\|\boldsymbol{y} - \mathbf{X}\widehat{\beta}_{\mathbf{H}}\|^2 = E\|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2 + E\|\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{\mathbf{H}})\|^2$, 其中,
$E\|\boldsymbol{y} - \mathbf{X}\widehat{\beta}\|^2 = (n-r)\sigma^2$,

$$\begin{aligned} E\|\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{\mathbf{H}})\|^2 &= E(\mathbf{H}\widehat{\beta} - \boldsymbol{d})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^-\mathbf{H}')^{-1}(\mathbf{H}\widehat{\beta} - \boldsymbol{d}) \\ &= \mathrm{tr}[(\mathbf{H}(\mathbf{X}'\mathbf{X})^-\mathbf{H}')^{-1}\mathrm{Cov}(\mathbf{H}\widehat{\beta})] = k\sigma^2, \end{aligned}$$

定理得证.

定理4.3.3

对于正态Guass-Markov线性模型，若定理4.3.1假设成立，则$\mathbf{H}\widehat{\beta}_{\mathbf{H}}$和$(n-r+k)\widehat{\sigma}^2_{\mathbf{H}}/n$分别为$\mathbf{H}\beta$和$\sigma^2$在线性约束$\mathbf{H}\beta=\boldsymbol{d}$下的约束ML估计.

证明 记$\boldsymbol{\mu}=\mathbf{X}\beta$, 考虑$\boldsymbol{\mu}$和$\sigma^2$的对数似然函数

$$\log L(\boldsymbol{\mu},\sigma^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{\mu}\|^2. \tag{4.11}$$

不难看出：在$\mathbf{H}\beta=\boldsymbol{d}$下$\boldsymbol{\mu}=\mathbf{X}\beta$的ML估计为

$$\widehat{\boldsymbol{\mu}}_{\mathbf{H}}=\arg\min_{\substack{\boldsymbol{\mu}=\mathbf{X}\beta\\ \mathbf{H}\beta=\boldsymbol{d}}}\|\boldsymbol{y}-\boldsymbol{\mu}\|^2=\mathbf{X}\widehat{\beta}_{\mathbf{H}},$$

将其带入(4.11), 极大化$\log L(\widehat{\boldsymbol{\mu}}_{\mathbf{H}},\sigma^2)$ 得 $\widetilde{\sigma}^2_{\mathbf{H}}=\frac{\|\boldsymbol{y}-\mathbf{X}\widehat{\beta}_{\mathbf{H}}\|^2}{n}=(n-r+k)\widehat{\sigma}^2_{\mathbf{H}}/n.$

线性模型

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{\Sigma} \tag{4.12}$$

其中$\boldsymbol{\Sigma} > \mathbf{0}$. 考虑参数$\boldsymbol{\beta}$, σ^2的估计问题.

- 采用模型变换.

用$\boldsymbol{\Sigma}^{-1/2}$ 左乘模型(4.12), 得Guass-Markov模型

$$\widetilde{\boldsymbol{y}} = \widetilde{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{u},$$

其中, $\widetilde{\boldsymbol{y}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}, \quad \widetilde{\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2}\mathbf{X}, \quad \boldsymbol{u} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{e},$

$$E(\boldsymbol{u}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{u}) = \sigma^2\mathbf{I}_n.$$

- 针对变换后的模型, 采用LS法得

$$\beta^* = (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^-\widetilde{\mathbf{X}}'\widetilde{\boldsymbol{y}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

$$\sigma^{2*} = \|\widetilde{\boldsymbol{y}} - \widetilde{\mathbf{X}}\beta^*\|^2/(n-r) = (\boldsymbol{y} - \mathbf{X}\beta^*)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mathbf{X}\beta^*)/(n-r).$$

称β^*和σ^{2*}为模型(4.12)的广义最小二乘(GLS)估计.

定理4.4.1

对任一可估函数$\boldsymbol{c}'\beta$, GLS估计$\boldsymbol{c}'\beta^*$为其惟一的BLU估计, 其方差为$\sigma^2\boldsymbol{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\boldsymbol{c}$; σ^{2*}为σ^2的无偏估计, 其中

$$\beta^* = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

$$\sigma^{2*} = (\boldsymbol{y} - \mathbf{X}\beta^*)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mathbf{X}\beta^*)/(n-r).$$

定理4.4.2

在线性模型$\boldsymbol{y} = \mathbf{X}\beta + \boldsymbol{e}$下, 若$\boldsymbol{e} \sim N(0, \sigma^2\boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}(> 0)$ 已知, 则

(1) 对任一可估函数$\boldsymbol{c}'\beta$, $\boldsymbol{c}'\beta^*$为$\boldsymbol{c}'\beta$的ML估计和唯一MVU估计, 且

$$\boldsymbol{c}'\beta^* \sim N(\boldsymbol{c}'\beta, \sigma^2\boldsymbol{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\boldsymbol{c}).$$

(2) $\frac{n-r}{n}\sigma^{2*}$为σ^2的ML估计, σ^{2*}为σ^2的唯一MVU估计, 且

$$(n-r)\sigma^{2*}/\sigma^2 \sim \chi^2_{n-r}.$$

(3) $\boldsymbol{c}'\beta^*$与σ^{2*}相互独立.

对于线性模型

$$\boldsymbol{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{e},\qquad E(\boldsymbol{e})=\mathbf{0},\qquad \mathrm{Cov}(\boldsymbol{e})=\sigma^2\boldsymbol{\Sigma}>\mathbf{0},$$

可估函数$\boldsymbol{c}'\boldsymbol{\beta}$ 的LS估计和GLS估计

$$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}=\boldsymbol{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{y},\ \ \boldsymbol{c}'\boldsymbol{\beta}^*=\boldsymbol{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}.$$

- 两者都是无偏的, $E(\boldsymbol{c}'\widehat{\boldsymbol{\beta}})=\boldsymbol{c}'\boldsymbol{\beta}=E(\boldsymbol{c}'\boldsymbol{\beta}^*)$;
- GLS估计的方差更小, 由定理4.4.1, 直接有

$$\mathrm{Var}(\boldsymbol{c}'\boldsymbol{\beta}^*)\leq\mathrm{Var}(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}).$$

例 用一种精密仪器在两个实验室对同一个物μ分别进行测量, 假设数据模型如下

$$y_{1i} = \mu + e_{1i}, \quad i = 1, \cdots, n_1,$$
$$y_{2i} = \mu + e_{2i}, \quad i = 1, \cdots, n_2.$$

由于两个实验室的客观条件及精密仪器的精度不同, 故测量误差的方差不等. 设

$$\mathrm{Var}(e_{1i}) = \sigma_1^2 \neq \mathrm{Var}(e_{2i}) = \sigma_2^2.$$

记 $\boldsymbol{y} = (y_{11}, \cdots, y_{1n_1}, y_{21}, \cdots, y_{2n_2})'$,

$$\boldsymbol{e} = (e_{11}, \cdots, e_{1n_1}, e_{21}, \cdots, e_{2n_2})'.$$

则

$$\text{Cov}(\boldsymbol{e}) = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{n_1} & 0 \\ 0 & \sigma_2^2 \mathbf{I}_{n_2} \end{pmatrix} = \sigma_2^2 \begin{pmatrix} \theta \mathbf{I}_{n_1} & 0 \\ 0 & \mathbf{I}_{n_2} \end{pmatrix} = \sigma_2^2 \boldsymbol{\Sigma}.$$

假设$\theta = \sigma_1^2/\sigma_2^2$ 已知, 则 μ 的GLS估计

$$\mu^* = \left(\frac{n_1}{\theta} + n_2\right)^{-1} \left(\frac{\sum_{i=1}^{n_1} y_{1i}}{\theta} + \sum_{i=1}^{n_2} y_{2i} \right) = v_1 \bar{y}_1 + v_2 \bar{y}_2,$$

这里, $v_1 = 1 - v_2$,

$$v_2 = \frac{w_2}{w_1 + w_2}, \quad w_i = \frac{1}{\text{Var}(\bar{y}_i)} = \frac{n_i}{\sigma_i^2}.$$

线性模型

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e} \quad E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{\Sigma}, \tag{4.13}$$

如果$|\boldsymbol{\Sigma}| = 0$, 则称该模型为奇异线性模型.

引理4.5.1

对于线性模型(4.13), 不管$\boldsymbol{\Sigma} > \mathbf{0}$或$\boldsymbol{\Sigma} \geq \mathbf{0}$, $\boldsymbol{y} \in \mathcal{M}(\boldsymbol{\Sigma}, \mathbf{X})$总是成立.

引理4.5.2

记$\mathbf{T} = \boldsymbol{\Sigma} + \mathbf{X}\mathbf{U}\mathbf{X}'$, 其中$\mathbf{U} \geq \mathbf{0}$, $\mathrm{rk}(\mathbf{T}) = \mathrm{rk}(\boldsymbol{\Sigma}, \mathbf{X})$, 则总有

(1) $\mathcal{M}(\mathbf{T}) = \mathcal{M}(\boldsymbol{\Sigma}, \mathbf{X})$.

(2) $\mathbf{X}'\mathbf{T}^-\mathbf{X}, \mathbf{X}'\mathbf{T}^-\boldsymbol{y}$和$(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})\mathbf{T}^-(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$都与广义逆$\mathbf{T}^-$的选择无关.

引理4.5.3

对于线性模型(4.13),可估函数$\boldsymbol{c}'\boldsymbol{\beta}$的一个无偏估计$\boldsymbol{a}'\boldsymbol{y}$为BLU估计，当且仅当它满足

$$\text{Cov}(\boldsymbol{a}'\boldsymbol{y}, \boldsymbol{b}'\boldsymbol{y}) = 0,$$

这里$\boldsymbol{b}'\boldsymbol{y}$为零的任一无偏估计，即$E(\boldsymbol{b}'\boldsymbol{y}) = 0$.

证明 设$\boldsymbol{l}'\boldsymbol{y}$为$\boldsymbol{c}'\boldsymbol{\beta}$的任一无偏估计，则$\boldsymbol{l}$一定可表示为$\boldsymbol{l} = \boldsymbol{a} + \boldsymbol{b}$，对某个满足$\mathbf{X}'\boldsymbol{b} = \mathbf{0}$的$\boldsymbol{b}$. 于是

$$\text{Var}(\boldsymbol{l}'\boldsymbol{y}) = \text{Var}(\boldsymbol{a}'\boldsymbol{y}) + \text{Var}(\boldsymbol{b}'\boldsymbol{y}) + 2\text{Cov}(\boldsymbol{a}'\boldsymbol{y}, \boldsymbol{b}'\boldsymbol{y}). \tag{4.14}$$

充分性得证.

下面用反证法来证明必要性.

设$\boldsymbol{a}'\boldsymbol{y}$为$\boldsymbol{c}'\beta$的BLU估计. 若存在一个$\boldsymbol{b}_0$, 满足$\mathbf{X}'\boldsymbol{b}_0 = \mathbf{0}$, 但有

$$\mathrm{Cov}(\boldsymbol{a}'\boldsymbol{y}, \boldsymbol{b}_0'\boldsymbol{y}) = \boldsymbol{d} \neq \mathbf{0},$$

不妨设$d < 0$. 若不然, 只需取$-\boldsymbol{b}_0$代替$\boldsymbol{b}_0$, 就可化为$d < 0$的情形.

用$\boldsymbol{b} = \alpha \boldsymbol{b}_0$代替(4.14) 中的$\boldsymbol{b}$,

$$f(\alpha) = \mathrm{Var}(\boldsymbol{l}'\boldsymbol{y}) - \mathrm{Var}(\boldsymbol{a}'\boldsymbol{y}) = \alpha(\alpha \boldsymbol{b}_0'\boldsymbol{\Sigma}\boldsymbol{b}_0 + 2d).$$

易证当$\alpha \in (0, -2b/\boldsymbol{b}_0'\boldsymbol{\Sigma}\boldsymbol{b}_0)$时, $f(\alpha) < 0$, 即

$$\mathrm{Var}(\boldsymbol{l}'\boldsymbol{y}) < \mathrm{Var}(\boldsymbol{a}'\boldsymbol{y}),$$

这与$\boldsymbol{a}'\boldsymbol{y}$为BLU估计相矛盾. 引理得证.

定理4.5.1

对于线性模型(4.13)和任一可估函数$c'\beta$有

(1) $c'\beta^* = c'(\mathbf{X}'\mathbf{T}^-\mathbf{X})^-\mathbf{X}'\mathbf{T}^-y$为$c'\beta$的BLU估计；

(2) $\mathrm{Var}(c'\beta^*) = \sigma^2 c'\left[(\mathbf{X}'\mathbf{T}^-\mathbf{X})^- - \mathbf{U}\right]c.$

(3) $\sigma^{2^*} = (y - \mathbf{X}\beta^*)'\mathbf{T}^-(y - \mathbf{X}\beta^*)/(\mathrm{rk}(\mathbf{T}) - \mathrm{rk}(\mathbf{X}))$ 为σ^2的无偏估计.

推论4.5.1

对于线性模型(4.13) 和任一可估函数$c'\beta$, 若$\mathcal{M}(\mathbf{X}) \subset \mathcal{M}(\mathbf{\Sigma})$, 则

(1) $c'\beta^* = c'(\mathbf{X}'\mathbf{\Sigma}^-\mathbf{X})^-\mathbf{X}'\mathbf{\Sigma}^-y$ 为$c'\beta$的BLU估计；

(2) $\mathrm{Var}(c'\beta^*) = \sigma^2 c'(\mathbf{X}'\mathbf{\Sigma}^-\mathbf{X})^-c$, 并且所有表达式与所包含的广义逆选择无关.

Panel模型

$$y_{it} = \beta_0 + x_{it1}\beta_1 + \cdots + x_{itk}\beta_k + \mu_i + e_{it}, \tag{4.15}$$

$$i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T,$$

这里y_{it}为第i个个体在时刻t 的观测值, μ_i为第i个个体的效应, 与随机误差e_{it}为随机误差不相关, 且

$$E(e_{it}) = 0, \quad \text{Var}(e_{it}) = \sigma_e^2, \quad E(\mu_i) = 0, \quad \text{Var}(\mu_i) = \sigma_\mu^2.$$

记$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)'$, $\boldsymbol{x}_{it} = (x_{it1}, \cdots, x_{itk})'$,

$$\boldsymbol{y} = (y_{11}, \cdots, y_{1T}, y_{21}, \cdots, y_{2T}, \cdots, y_{NT})',$$

$$\mathbf{X} = (\boldsymbol{x}_{11}, \cdots, \boldsymbol{x}_{1T}, \boldsymbol{x}_{21}, \cdots, \boldsymbol{x}_{2T}, \cdots, \boldsymbol{x}_{NT})',$$

$$\boldsymbol{e} = (e_{11}, \cdots, e_{1T}, e_{21}, \cdots, e_{2T}, \cdots, e_{NT})'.$$

- 模型矩阵形式：

$$\boldsymbol{y} = \mathbf{1}_{NT}\beta_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u},$$

这里, $\boldsymbol{u} = (\mathbf{I}_N \otimes \mathbf{1}_T)\boldsymbol{\mu} + \boldsymbol{e}$. 易证协方差阵的谱分解：

$$\mathsf{Cov}(\boldsymbol{u}) = \sigma_1^2 \mathsf{P}_1 + \sigma_e^2 \mathbf{Q} + \sigma_1^2 \mathbf{J}_{NT},$$

这里,

$$\mathsf{P}_1 = (\mathbf{I}_N - \mathbf{J}_N) \otimes \mathbf{J}_T,$$

$$\mathbf{Q} = \mathbf{I}_N \otimes (\mathbf{I}_T - \mathbf{J}_T),$$

$$\mathbf{J}_k = \mathbf{1}_k \mathbf{1}_k' / k.$$

- P_1左乘模型的等号两边，得约见模型：

$$\mathsf{P}_1\boldsymbol{y} = \mathsf{P}_1\mathbf{X}\beta + \mathsf{P}_1\boldsymbol{u}.$$

称其的BLU估计为Between估计：

$$\widehat{\beta}_B = (\mathbf{X}'\mathsf{P}_1\mathbf{X})^{-1}\mathbf{X}'\mathsf{P}_1\boldsymbol{y}$$

- 用$\mathbf{Q} = \mathbf{I}_N \otimes (\mathbf{I}_T - \mathbf{J}_T)$左乘原模型得变换模型：

$$\mathbf{Q}\boldsymbol{y} = \mathbf{Q}\mathbf{X}\beta + \mathbf{Q}\boldsymbol{e}$$

的BLU估计分别为Within估计：

$$\widehat{\beta}_W = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\boldsymbol{y}$$

线性模型

$$y = \mathbf{X}\beta + e, \quad E(e) = \mathbf{0}, \quad \mathrm{Cov}(e) = \sigma^2\mathbf{\Sigma}, \tag{4.13}$$

这里$\mathbf{\Sigma} \geq \mathbf{0}$已知. 对任一可估函数$c'\beta$, 它的LS估计为

$$c'\widehat{\beta} = c'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'y.$$

研究LS估计关于协方差阵的稳健性, 即研究$\mathbf{\Sigma}$满足什么结构时,

$$c'\widehat{\beta} = c'\beta^*, \tag{4.14}$$

记$\mathbf{Z}$为$n\times(n-r)$且秩为$n-r$的矩阵, 满足$\mathbf{X}'\mathbf{Z}=\mathbf{0}$, 这里$r=\text{rk}(\mathbf{X})$.

定理4.6.1

对于线性模型(4.13)和任一可估函数$c'\beta$, (4.14)成立当且仅当下列条件之一成立.

(1) $\mathbf{X}'\mathbf{\Sigma}\mathbf{Z}=\mathbf{0}$;

(2) $\mathbf{\Sigma}=\mathbf{X}\mathbf{\Lambda}_1\mathbf{X}'+\mathbf{Z}\mathbf{\Lambda}_2\mathbf{Z}'$;

(3) $\mathbf{\Sigma}=\mathbf{X}\mathbf{D}_1\mathbf{X}'+\mathbf{Z}\mathbf{D}_2\mathbf{Z}'+\mathbf{I}_n$,

其中$\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, $\mathbf{D}_1$和$\mathbf{D}_2$为任意对称阵, 但使$\mathbf{\Sigma}\geq\mathbf{0}$.

证明留给学生自学.

例 误差均匀相关模型(error uniform correlation models).

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad E(\boldsymbol{e}) = \mathbf{0},$$

误差向量$\boldsymbol{e}$的协方差阵具有如下形式

$$\text{Cov}(\boldsymbol{e}) = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} = \sigma^2[\rho \mathbf{1}_n \mathbf{1}_n' + (1-\rho)\mathbf{I}_n].$$

若$\mathbf{X}$的第一列全为1, 即模型包含常数项, 则$\mathbf{X}'\text{Cov}(\boldsymbol{e})\mathbf{Z} = \mathbf{0}$, 故任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$的LS估计仍为BLU估计.

定理4.6.2

对于线性模型(4.13) 和任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$, (4.14)成立当且仅当下列条件之一成立.

(1) $\mathbf{\Sigma X} = \mathbf{XB}$, 对某矩阵$\mathbf{B}$;

(2) $\mathcal{M}(\mathbf{X})$由$\mathbf{\Sigma}$的$r = \mathrm{rk}(\mathbf{X})$个特征向量张成;

(3) $\mathbf{P_X\Sigma}$为对称阵, 其中$\mathbf{P_X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$.

证明留给学生自学.

例 考虑单向分类随机模型

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \cdots, a, \; j = 1, \cdots, b,$$

这里μ为固定效应, α_i为随机效应, e_{ij}为随机误差, $\mathrm{Cov}(\alpha_i, e_{ij}) = 0$, $\mathrm{Var}(\alpha_i) = \sigma_\alpha^2$, $\mathrm{Var}(e_{ij}) = \sigma_e^2$. 记$n = ab$, $\mathbf{X} = \mathbf{1}_n$, $\mathbf{U} = \mathbf{I}_a \otimes \mathbf{1}_b$.

矩阵形式

$$\boldsymbol{y} = \mathbf{X}\mu + \mathbf{U}\boldsymbol{\alpha} + \boldsymbol{e},$$

$$\mathrm{Cov}(\boldsymbol{y}) = \sigma_\alpha^2 \mathbf{U}\mathbf{U}' + \sigma_e^2 \mathbf{I}_n = \sigma_\alpha^2 (\mathbf{I}_a \otimes \mathbf{1}_b \mathbf{1}_b') + \sigma_e^2 \mathbf{I}_n \triangleq \boldsymbol{\Sigma}(\sigma^2).$$

根据矩阵Z的定义$\mathbf{X}'\mathbf{Z} = \mathbf{1}_n'\mathbf{Z} = \mathbf{0}$, 所以$\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} = \mathbf{0}$. 依定理4.6.1, 固定效应$\mu$的LS估计$\widehat{\mu} = \bar{y}_{..}$是$\mu$的BLU估计.

考虑一般线性模型

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \qquad E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}), \tag{4.18}$$

这里$\boldsymbol{y}$为$n \times 1$观测向量, $\mathbf{X}$为$n \times p$设计阵, $\boldsymbol{\beta}$为$p \times 1$ 未知参数向量, $\boldsymbol{e}$为$n \times 1$随机误差, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)$也是未知参数向量. 设$\boldsymbol{\Sigma}(\boldsymbol{\theta}) > 0$ 对一切$\boldsymbol{\theta}$成立.

- 当$\boldsymbol{\theta}$已知时, 对任一可估函数$\boldsymbol{c}'\boldsymbol{\beta}$, 当$\boldsymbol{\theta}$已知时, $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$就是它的GLS估计, 也是BLU估计, 其中,

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{y}.$$

- 当θ未知时, $c'\widehat{\beta}(\theta)$ 就不再是一个可行估计.
  此时, 若可得到θ的一个估计$\widehat{\theta}$, 将$\widehat{\beta}(\theta)$中θ用估计代替，便得到$c'\beta$的一个两步估计$c'\widehat{\beta}(\widehat{\theta})$.

定理4.7.1

对于线性模型(4.18), 假设e的分布关于原点是对称的. 设$\widehat{\theta} = \widehat{\theta}(y)$是$\theta$ 的一个估计, 它是y的偶函数且具有变换不变性. 设$c'\beta$为任一可估函数, 若$E(c'\widehat{\beta}(\widehat{\theta}))$存在, 则两步估计$c'\widehat{\beta}(\widehat{\theta})$是$c'\beta$的无偏估计, 即

$$E(c'\widehat{\beta}(\widehat{\theta})) = c'\beta.$$

证明留给学生自学.

例 Panel模型

$$y = \mathbf{1}_{NT}\beta_0 + \mathbf{X}\beta + u, \quad \text{Cov}(u) = \sigma_1^2\mathbf{P}_1 + \sigma_e^2\mathbf{Q} + \sigma_1^2\mathbf{J}_{NT},$$

应用Between估计和Within估计, 可得到σ_1^2和σ_e^2的无偏估计：

$$s_1^2 = \hat{u}_1'\mathbf{P}_1\hat{u}_1/(N-k-1),$$

$$s_2^2 = (y - \mathbf{X}\hat{\beta}_{\mathbf{W}})'\mathbf{Q}(y - \mathbf{X}\hat{\beta}_{\mathbf{W}})/(N(T-1)-k).$$

可证明两步估计

$$\hat{\beta}(s^2) = \left(\frac{\mathbf{B}}{s_1^2} + \frac{\mathbf{W}}{s_2^2}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{P}_1 y}{s_1^2} + \frac{\mathbf{X}'\mathbf{Q}y}{s_2^2}\right)$$

是β的无偏估计, 这里$s^2 = (s_1^2, s_2^2)'$.

- 两步估计$\widetilde{\boldsymbol{\beta}}$的均方误差矩阵(MSEM)

  为符号简单, 记$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}$,

$$\boldsymbol{\beta}^* = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y},$$
$$\widetilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{y}.$$

定理4.7.3

设$\boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2\boldsymbol{\Sigma})$, 则

$$\mathrm{MSEM}(\widetilde{\boldsymbol{\beta}}) = \mathsf{Cov}(\boldsymbol{\beta}^*) + E(\boldsymbol{b}\boldsymbol{b}'),$$

这里$\boldsymbol{b} = \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$.

MVU估计的判定定理(陈希孺, 1999)

参数θ 的一个无偏估计T 为MVU估计, 当且仅当对零的任一无偏估计U和一切$\theta \in \Theta$, 都有

$$\mathrm{Cov}(T, U) = 0,$$

这里Θ为参数空间, $\mathrm{var}(T) < \infty$, $\mathrm{var}(U) < \infty$.

这就是说,一个无偏估计要具有最小方差当且仅当它跟零的所有无偏估计都不相关.

- 协方差改进法(Rao,1967): 利用U_0 与T的相关性构造两者的一个线性组合, 使其为θ的无偏估计, 且比T具有更小方差.

定理4.8.1

设$\boldsymbol{\theta}$为$p \times 1$未知参数, 统计量$\boldsymbol{T}_1$和$\boldsymbol{T}_2$ 满足$E(\boldsymbol{T}_1) = \boldsymbol{\theta}$, $E(\boldsymbol{T}_2) = \mathbf{0}$,

$$\mathrm{Cov}\begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \triangleq \boldsymbol{\Sigma}.$$

假定$\boldsymbol{\Sigma} > \mathbf{0}$, $\boldsymbol{\Sigma}_{12} \neq \mathbf{0}$. 则在线性估计类

$$\mathbf{A} = \left\{\boldsymbol{T} = \mathbf{A}_1\boldsymbol{T}_1 + \mathbf{A}_2\boldsymbol{T}_2, \mathbf{A}_1 \text{ 和} \mathbf{A}_2 \text{ 为非随机阵}, \ E(\boldsymbol{T}) = \boldsymbol{\theta}\right\},$$

$\boldsymbol{\theta}$ 的BLU 估计为

$$\boldsymbol{\theta}^* = \boldsymbol{T}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{T}_2,$$

且$\mathrm{Cov}(\boldsymbol{\theta}^*) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \leq \boldsymbol{\Sigma}_{11} = \mathrm{Cov}(\boldsymbol{T}_1)$.

证明 将定理的条件用线性模型表示, 即为

$$\begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{0} \end{pmatrix} \boldsymbol{\theta} + \boldsymbol{e}, \qquad \boldsymbol{e} \sim (\mathbf{0}, \boldsymbol{\Sigma}),$$

易验证$\boldsymbol{\theta}^*$为该模型中$\boldsymbol{\theta}$的BLU估计. 其它结论显然, 定理证毕.

- 称估计$\boldsymbol{\theta}^* == \boldsymbol{T}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{T}_2$为协方差改进估计.
- 使用了$\boldsymbol{T}_2$和$\boldsymbol{T}_1$的相关性所带来附加信息, 对协方差阵的影响:

$$\mathrm{Cov}(\boldsymbol{\theta}^*) - \mathrm{Cov}(\boldsymbol{T}_1) = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \leq \mathbf{0}.$$

- 如果$\boldsymbol{\Sigma}_{12} = \mathbf{0}$, 这时$\boldsymbol{T}_2$与$\boldsymbol{T}_1$不相关, 自然$\boldsymbol{T}_2$也就没有任何改进$\boldsymbol{T}_1$的附加信息.

例 考虑一般线性回归模型

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \qquad E(\boldsymbol{e}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{\Sigma} > \mathbf{0},$$

这里rk$(\mathbf{X}) = p$. 取$\boldsymbol{T}_1 = \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$, $\boldsymbol{T}_2 = \mathbf{Z}'\boldsymbol{y}$, 这里$\mathbf{Z}$满足$\mathbf{X}'\mathbf{Z} = \mathbf{0}$, 且rk$(\mathbf{Z}) = n - p$, 则$E(\boldsymbol{T}_1) = E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, $E(\boldsymbol{T}_2) = \mathbf{0}$,

$$\mathrm{Cov}\begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix} = \sigma^2 \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} \\ \mathbf{Z}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z} \end{pmatrix}.$$

假定$\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} \neq \mathbf{0}$, 协方差改进估计

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{y} = \boldsymbol{\beta}^*,$$

即$\boldsymbol{\beta}$的BLU估计$\boldsymbol{\beta}^*$是从LS估计经过一次协方差改进得到的.

例 带线性约束的线性回归模型

$$\begin{cases} \boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, \qquad E(\boldsymbol{e}) = \mathbf{0}, \qquad \mathrm{Cov}(\boldsymbol{e}) = \sigma^2 \mathbf{I}_n, \\ \mathbf{H}\boldsymbol{\beta} = \boldsymbol{d}, \end{cases}$$

这里$\mathbf{H}$为$m \times p$ 矩阵, $\mathrm{rk}(\mathbf{H}) = m$, 且$\mathbf{H}\boldsymbol{\beta} = \boldsymbol{d}$是相容的. 取$\boldsymbol{T}_1 = \widehat{\boldsymbol{\beta}}$, $\boldsymbol{T}_2 = \mathbf{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{d}$. 在约束参数区域$\mathbf{H}\boldsymbol{\beta} = \boldsymbol{d}$上, $E(\boldsymbol{T}_2) = \mathbf{0}$,

$$\mathrm{Cov}\begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix} = \sigma^2 \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}' \\ \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}' \end{pmatrix}.$$

应用定理4.8.1, 得到协方差改进估计

$$\widehat{\boldsymbol{\beta}}_H = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{d}),$$

它正是$\boldsymbol{\beta}$的约束LS估计.

例 带随机形式附加信息的线性回归模型

$$\begin{aligned} \boldsymbol{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad E(\boldsymbol{e}) = \mathbf{0}, \qquad \mathrm{Cov}(\boldsymbol{e}) = \sigma^2 \mathbf{I}_n, \\ \boldsymbol{u} &= \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad E(\boldsymbol{\varepsilon}) = \mathbf{0}, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathbf{W}, \end{aligned}$$

这里$\mathbf{W} > \mathbf{0}$是已知矩阵, $\boldsymbol{\varepsilon}$和$\boldsymbol{e}$不相关. 取$\boldsymbol{T}_1 = \widehat{\boldsymbol{\beta}}$, $\boldsymbol{T}_2 = \boldsymbol{u} - \mathbf{H}\widehat{\boldsymbol{\beta}}$. 若$\sigma^2$ 已知, 则所得到的协方差改进估计

$$\boldsymbol{\beta}^*(\sigma^2) = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}' \left(\frac{\mathbf{W}}{\sigma^2} + \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}' \right)^{-1} (\mathbf{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{u}),$$

等价于

$$\boldsymbol{\beta}^*(\sigma^2) = \left(\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} + \mathbf{H}\mathbf{W}^{-1}\mathbf{H}' \right)^{-1} \left(\frac{\mathbf{X}'\boldsymbol{y}}{\sigma^2} + \mathbf{H}\mathbf{W}^{-1}\boldsymbol{u} \right).$$

这就是通常的混合估计.

小节

- 学习重点：最小二乘估计及其性质
- 需要掌握分块估计、约束估计、广义最小二乘估计、协方差改进估计的思想
- 了解最小二乘估计稳健性相关理论