3.3 最优回归设计

王正明　易泰河

系统工程学院　军事建模与仿真系

2019 年 12 月 20 日

引言

思考

将一幅画固定在墙上，至少需要几颗钉子？如果给三颗钉子，应该如何钉？给四颗呢？

例

假设消耗的能量 y 与跑步距离 x 存在指数关系：

$$y = A\exp\{Bx\}, \qquad x \in [5\text{km}, 10\text{km}].$$

A 和 B 为待估计的参数．设计试验以高效地得到 A 和 B，并设计一次验证模型的跑步试验．

引言

思考

将一幅画固定在墙上，至少需要几颗钉子？如果给三颗钉子，应该如何钉？给四颗呢？

例

假设消耗的能量 y 与跑步距离 x 存在指数关系:

$$y = A\exp\{Bx\}, \qquad x \in [5\text{km}, 10\text{km}].$$

A 和 B 为待估计的参数．设计试验以高效地得到 A 和 B，并设计一次验证模型的跑步试验．

引言

例 (称重问题)

用一架精度为 σ 的天平称 4 个不同的物体, 称重方案:

序号	左侧	右侧	差异
1	1 2 3 4	空	y_1
2	1 2	3 4	y_2
3	1 3	2 4	y_3
4	1 4	2 3	y_4

每一个物体的测量精度都能达到 $\sigma^2/4$. 能否找到一种精度更高的称重方案?

知识回顾

- 称 $\xi_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_n\}$ 为一个试验次数为 n 的精确设计, $\boldsymbol{x}_i$ 为它的支撑点或谱点. ξ_n 的矩阵表示

$$\boldsymbol{D}_{\xi_n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

- 离散设计: $\xi_k = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$.
- 称试验区域 $\mathcal{X}$ 上的概率分布 ξ 为设计, 给定试验次数 n 后从 ξ 中抽取 n 个样本作为试验方案.
- 也称 ξ_n 为确定性设计, ξ_k 和 ξ 为近似设计.

知识回顾

- 线性回归模型: $y = \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\beta} + \varepsilon$.

- 线性模型 $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$ 的参数估计:

$$\begin{cases} \widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}, \\ \widehat{\sigma}^2 = \dfrac{1}{n-m}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \end{cases}$$

- $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\right)$, 它的方差矩阵为 $\sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$.

- 预测值 $\hat{y}(\boldsymbol{x}) = \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\hat{\boldsymbol{\beta}}$ 的方差为 $\sigma^2 \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})$.

- 矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_i)$ 决定参数估计和预测的精度.

3.3 最优回归设计

线性回归模型 $y = \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 中:

- 精确设计 $\xi_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_n\}$ 的信息矩阵定义为

$$\boldsymbol{M}(\xi_n) = \frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_i).$$

- 离散设计 $\xi_k = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$ 的信息矩阵定义为

$$\boldsymbol{M}(\xi_k) = \sum_{i=1}^{k} p_i\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_i).$$

线性回归模型 $y = \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 中:

- 精确设计 $\xi_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_n\}$ 的信息矩阵定义为

$$\boldsymbol{M}(\xi_n) = \frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_i).$$

- 离散设计 $\xi_k = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$ 的信息矩阵定义为

$$\boldsymbol{M}(\xi_k) = \sum_{i=1}^{k} p_i\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_i).$$

定义 (设计的信息矩阵)

线性回归模型 $y = \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 中, 称试验区域 $\mathcal{X}$ 上的概率分布 ξ 为一个设计, 其信息矩阵定义为

$$\boldsymbol{M}(\xi) := \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\mathrm{d}\xi,$$

称满足 $\det(\boldsymbol{M}(\xi)) \neq 0$ 的设计 ξ 为非奇异的.

- 一个好的设计应使 $M(\xi)$ 达到“最大”.

定义 (设计的信息矩阵)

线性回归模型 $y = \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 中, 称试验区域 $\mathcal{X}$ 上的概率分布 ξ 为一个设计, 其信息矩阵定义为

$$\boldsymbol{M}(\xi) := \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\mathrm{d}\xi,$$

称满足 $\det(\boldsymbol{M}(\xi)) \neq 0$ 的设计 ξ 为非奇异的.

- 一个好的设计应使 $\boldsymbol{M}(\xi)$ 达到 "最大".

例 (一元一次线性回归)

设 $x \in [-1, 1]$, 求线性模型 $y = \beta_0 + \beta_1 x + \varepsilon$ 中精确设计 $\xi_n = \{x_1, x_2, \cdots, x_n\}$ 和密度函数为 $p(x)$ 的设计 ξ 的信息矩阵.

- 精确设计 $\xi_n = \{x_1, x_2, \cdots, x_n\}$ 的信息矩阵为

$$\begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n} x_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix}.$$

- 密度函数为 $p(x)$ 的设计 ξ 的信息矩阵为

$$\begin{bmatrix} 1 & \int_{-1}^{1} x p(x) \mathrm{d}x \\ \int_{-1}^{1} x p(x) \mathrm{d}x & \int_{-1}^{1} x^2 p(x) \mathrm{d}x \end{bmatrix}.$$

例 (一元一次线性回归)

设 $x \in [-1, 1]$, 求线性模型 $y = \beta_0 + \beta_1 x + \varepsilon$ 中精确设计 $\xi_n = \{x_1, x_2, \cdots, x_n\}$ 和密度函数为 $p(x)$ 的设计 ξ 的信息矩阵.

- 精确设计 $\xi_n = \{x_1, x_2, \cdots, x_n\}$ 的信息矩阵为

$$\begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^n x_i \\ \frac{1}{n}\sum_{i=1}^n x_i & \frac{1}{n}\sum_{i=1}^n x_i^2 \end{bmatrix}.$$

- 密度函数为 $p(x)$ 的设计 ξ 的信息矩阵为

$$\begin{bmatrix} 1 & \int_{-1}^1 xp(x)\mathrm{d}x \\ \int_{-1}^1 xp(x)\mathrm{d}x & \int_{-1}^1 x^2 p(x)\mathrm{d}x \end{bmatrix}.$$

例 (二元一次线性回归)

设 $(x_1, x_2) \in [-1, 1]^2$, 讨论线性模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

中精确设计 ξ_n 和密度函数为 $p(x_1, x_2)$ 的设计 ξ 的信息矩阵.

- $M(\xi_n) = \begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^n x_{i1} & \frac{1}{n}\sum_{i=1}^n x_2 \\ \frac{1}{n}\sum_{i=1}^n x_{i1} & \frac{1}{n}\sum_{i=1}^n x_{i1}^2 & \frac{1}{n}\sum_{i=1}^n x_{i1}x_{i2} \\ \frac{1}{n}\sum_{i=1}^n x_{i2} & \frac{1}{n}\sum_{i=1}^n x_{i1}x_{i2} & \frac{1}{n}\sum_{i=1}^n x_{i2}^2 \end{bmatrix}$
- $M(\xi) =$

$$\begin{bmatrix} 1 & \int x_1 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ \int x_1 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_1^2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_1 x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ \int x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_1 x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_2^2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \end{bmatrix}$$

例 (二元一次线性回归)

设 $(x_1, x_2) \in [-1, 1]^2$, 讨论线性模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

中精确设计 ξ_n 和密度函数为 $p(x_1, x_2)$ 的设计 ξ 的信息矩阵.

- $$\boldsymbol{M}(\xi_n) = \begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n} x_{i1} & \frac{1}{n}\sum_{i=1}^{n} x_2 \\ \frac{1}{n}\sum_{i=1}^{n} x_{i1} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{i2} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{i2} & \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 \end{bmatrix}$$
- $\boldsymbol{M}(\xi) =$
$$\begin{bmatrix} 1 & \int x_1 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ \int x_1 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_1^2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_1 x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ \int x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_1 x_2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 & \int x_2^2 p(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \end{bmatrix}$$

信息矩阵的性质

以 Ξ 表示所有设计组成的集合，Ξ_n 表示支撑点数为 n 的设计组成的集合，$\mathcal{M} = \{\boldsymbol{M}(\xi) : \xi \in \Xi\}$:

(1) 任意设计 ξ 的信息矩阵 $\boldsymbol{M}(\xi)$ 都是非负定的;

(2) Ξ 是凸集，$\mathcal{M}$ 是一个闭凸集;

(3) 如果 $n < m$，则 $\det(\boldsymbol{M}(\xi)) = 0$ 对任意 $\xi \in \Xi_n$ 都成立;

(4) 任给 $\xi \in \Xi$，存在 $\tilde{\xi} \in \Xi_n$，$n \leq m(m+1)/2 + 1$，使得 $\boldsymbol{M}(\xi) = \boldsymbol{M}(\tilde{\xi})$.

3.3 最优回归设计

3.3.1 设计的信息矩阵

3.3.2 优良性准则

3.3.3 等价性定理

3.3.4 D 最优设计的迭代求解

定义 (最优设计)

设 $\varPhi : \mathcal{M} \mapsto \mathbb{R}^+$ 满足 $\boldsymbol{M}_1 \geq \boldsymbol{M}_2 \Rightarrow \varPhi(\boldsymbol{M}_1) \leq \varPhi(\boldsymbol{M}_2)$. 若存在设计 $\xi^* \in \varXi$ 使得

$$\varPhi\left(\boldsymbol{M}(\xi^*)\right) = \inf\left\{\varPhi\left(\boldsymbol{M}(\xi)\right) : \xi \in \varXi\right\},$$

则称 ξ^* 为 $\varPhi$ 最优设计($\varPhi$-optimal design).

- 一般不能保证 $\varPhi(\boldsymbol{M}_1) \leq \varPhi(\boldsymbol{M}_2) \Rightarrow \boldsymbol{M}_1 \geq \boldsymbol{M}_2$, 因此选择的最优性准则 $\varPhi$ 应具有一定的统计意义.
- 为简单起见, 记 $\varPhi(\xi) = \varPhi(\boldsymbol{M}(\xi))$.

定义 (最优设计)

设 $\Phi : \mathcal{M} \mapsto \mathbb{R}^+$ 满足 $\boldsymbol{M}_1 \geq \boldsymbol{M}_2 \Rightarrow \Phi(\boldsymbol{M}_1) \leq \Phi(\boldsymbol{M}_2)$. 若存在设计 $\xi^* \in \varXi$ 使得

$$\Phi\left(\boldsymbol{M}(\xi^*)\right) = \inf\left\{\Phi\left(\boldsymbol{M}(\xi)\right) : \xi \in \varXi\right\},$$

则称 ξ^* 为 Φ 最优设计(Φ-optimal design).

- 一般不能保证 $\Phi(\boldsymbol{M}_1) \leq \Phi(\boldsymbol{M}_2) \Rightarrow \boldsymbol{M}_1 \geq \boldsymbol{M}_2$, 因此选择的最优性准则 Φ 应具有一定的统计意义.
- 为简单起见, 记 $\Phi(\xi) = \Phi(\boldsymbol{M}(\xi))$.

定义 (D 最优设计)

取 $\Phi_D(\xi) := \det\left(\boldsymbol{M}^{-1}(\xi)\right)$, 称相应的最优设计为 D 最优设计(D-optimal design).

给定置信概率 α, 最小二乘估计 $\hat{\boldsymbol{\beta}}$ 的置信椭球体

$$\left\{\boldsymbol{\beta} \in \mathbb{R}^m : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{M}^{-1}(\xi)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \le c_\alpha\right\}$$

的体积 $V(\xi) \propto \left[\det\left(\boldsymbol{M}^{-1}(\xi)\right)\right]^{\frac{1}{2}}$, $\det\left(\boldsymbol{M}^{-1}(\xi)\right)$ 越小 $\hat{\boldsymbol{\beta}}$ 的精度越高.

定义 (D 最优设计)

取 $\Phi_D(\xi) := \det\left(\boldsymbol{M}^{-1}(\xi)\right)$, 称相应的最优设计为 D 最优设计(D-optimal design).

给定置信概率 α, 最小二乘估计 $\hat{\boldsymbol{\beta}}$ 的置信椭球体

$$\left\{\boldsymbol{\beta} \in \mathbb{R}^m : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{M}^{-1}(\xi)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq c_\alpha\right\}$$

的体积 $V(\xi) \propto \left[\det\left(\boldsymbol{M}^{-1}(\xi)\right)\right]^{\frac{1}{2}}$, $\det\left(\boldsymbol{M}^{-1}(\xi)\right)$ 越小 $\hat{\boldsymbol{\beta}}$ 的精度越高.

例 (一元一次线性回归)

设 $x \in [-1, 1]$, 限制在精确设计中求线性模型 $y = \beta_0 + \beta_1 x + \varepsilon$ 的 D 最优设计.

- 信息矩阵

$$M(\xi_n) = \frac{1}{n}\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

  的行列式为 $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.
- 若试验次数 $n = 2$, 则 D 最优设计为 $\xi_2 = \{-1, 1\}$.
- 若试验次数 $n = 3$, 则 D 最优设计为 $\xi_3 = \{-1, -1, 1\}$ 或 $\{-1, 1, 1\}$.

例 (一元一次线性回归)

设 $x \in [-1, 1]$, 限制在精确设计中求线性模型 $y = \beta_0 + \beta_1 x + \varepsilon$ 的 D 最优设计.

- 信息矩阵

$$\boldsymbol{M}(\xi_n) = \frac{1}{n} \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

  的行列式为 $\frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \bar{x})^2$.

- 若试验次数 $n = 2$, 则 D 最优设计为 $\xi_2 = \{-1, 1\}$.
- 若试验次数 $n = 3$, 则 D 最优设计为 $\xi_3 = \{-1, -1, 1\}$ 或 $\{-1, 1, 1\}$.

例 (称重问题)

用一架精度为 σ 的天平称 4 个不同的物体, 称重方案:

序号	左侧	右侧	差异
1	1 2 3 4	空	y_1
2	1 2	3 4	y_2
3	1 3	2 4	y_3
4	1 4	2 3	y_4

每一个物体的测量精度都能达到 $\sigma^2/4$. 能否找到一种精度更高的称重方案?

定义 (G 最优设计)

称 $d(\boldsymbol{x},\xi) := \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi)\boldsymbol{f}(\boldsymbol{x})$ 为设计 ξ 的**标准化方差**, 称

$$\Phi_G(\xi) := \sup\left\{d(\boldsymbol{x},\xi) : \boldsymbol{x} \in \mathcal{X}\right\}$$

为 G **最优准则**, 相应的最优设计称为 G **最优设计**.

点 $\boldsymbol{x} \in \mathcal{X}$ 处的响应预测值 $\hat{y}(\boldsymbol{x}) = \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\hat{\boldsymbol{\beta}}$ 的方差

$$\mathrm{Var}\left(\hat{y}(\boldsymbol{x})\right) = \frac{\sigma^2}{n}\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi)\boldsymbol{f}(\boldsymbol{x}).$$

G 最优设计使得模型的最大预测方差达到最小.

定义 (G 最优设计)

称 $d(\boldsymbol{x},\xi) := \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi)\boldsymbol{f}(\boldsymbol{x})$ 为设计 ξ 的标准化方差, 称

$$\Phi_G(\xi) := \sup\{d(\boldsymbol{x},\xi) : \boldsymbol{x}\in\mathcal{X}\}$$

为 G 最优准则, 相应的最优设计称为 G 最优设计.

点 $\boldsymbol{x}\in\mathcal{X}$ 处的响应预测值 $\hat{y}(\boldsymbol{x}) = \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\hat{\boldsymbol{\beta}}$ 的方差

$$\mathrm{Var}\left(\hat{y}(\boldsymbol{x})\right) = \frac{\sigma^2}{n}\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi)\boldsymbol{f}(\boldsymbol{x}).$$

G 最优设计使得模型的最大预测方差达到最小.

- 取向量 $\boldsymbol{c} \in \mathbb{R}^m$, 称 $\Phi_C(\xi) := \boldsymbol{c}^{\mathrm{T}} \boldsymbol{M}^{-1}(\xi)\boldsymbol{c}$ 为 C 最优准则, 它使得参数线性组合 $\boldsymbol{c}^T\boldsymbol{\beta}$ 的最优无偏估计的方差达到最小.

- 称 $\Phi_A(\xi) := \mathrm{tr}\{\boldsymbol{M}^{-1}(\xi)\}$ 为 A 准则或 MV 准则, 它使参数估计的方差之和最小.

- 记 $\lambda_{\min}(\boldsymbol{M}(\xi))$ 为矩阵 $\boldsymbol{M}(\xi)$ 的特征值的最小值, 称 $\Phi_E(\xi) := \lambda_{\min}^{-1}(\boldsymbol{M}(\xi))$ 为 E 准则, 该准则使置信椭球体的最长轴最小.

- 取向量 $\boldsymbol{c} \in \mathbb{R}^m$, 称 $\Phi_C(\xi) := \boldsymbol{c}^{\mathrm{T}} \boldsymbol{M}^{-1}(\xi)\boldsymbol{c}$ 为 C 最优准则, 它使得参数线性组合 $\boldsymbol{c}^T\boldsymbol{\beta}$ 的最优无偏估计的方差达到最小.

- 称 $\Phi_A(\xi) := \mathrm{tr}\{\boldsymbol{M}^{-1}(\xi)\}$ 为 A 准则或 MV 准则, 它使参数估计的方差之和最小.

- 记 $\lambda_{\min}(\boldsymbol{M}(\xi))$ 为矩阵 $\boldsymbol{M}(\xi)$ 的特征值的最小值, 称 $\Phi_E(\xi) := \lambda_{\min}^{-1}(\boldsymbol{M}(\xi))$ 为 E 准则, 该准则使置信椭球体的最长轴最小.

- 取向量 $\boldsymbol{c} \in \mathbb{R}^m$, 称 $\Phi_C(\xi) := \boldsymbol{c}^{\mathrm{T}} \boldsymbol{M}^{-1}(\xi)\boldsymbol{c}$ 为 C 最优准则, 它使得参数线性组合 $\boldsymbol{c}^T\boldsymbol{\beta}$ 的最优无偏估计的方差达到最小.

- 称 $\Phi_A(\xi) := \mathrm{tr}\{\boldsymbol{M}^{-1}(\xi)\}$ 为 A 准则或 MV 准则, 它使参数估计的方差之和最小.

- 记 $\lambda_{\min}(\boldsymbol{M}(\xi))$ 为矩阵 $\boldsymbol{M}(\xi)$ 的特征值的最小值, 称 $\Phi_E(\xi) := \lambda_{\min}^{-1}(\boldsymbol{M}(\xi))$ 为 E 准则, 该准则使置信椭球体的最长轴最小.

例 (一元一次线性回归)

$x \in [-1, 1]$, 讨论线性模型 $y = \beta_0 + \beta_1 x + \varepsilon$ 的只做两次试验的 A、G、C 以及 E 最优设计.

例 (一元多项式回归)

一元线性回归模型 $y=\beta_0+\beta_1 x+\cdots+\beta_d x^d+\varepsilon$ 的 D-最优设计的支撑点是方程

$$(1-x^2)\frac{\mathrm{d}P_d(x)}{\mathrm{d}x}=0$$

的根, 即由两个端点 -1 和 1, 以及 d 阶 Legender 多项式的导数的根组成, 一共 $d+1$ 个点. Legender 多项式定义如下:

$$\begin{cases} P_0(x)=1, \qquad P_1(x)=x, \\ (d+1)P_{d+1}(x)=(2d+1)xP_d(x)-dP_{d-1}(x). \end{cases}$$

例 (一元多项式回归 Cont.)

当 $d = 1, 2, \cdots, 6$ 时, 一元线性回归模型

$$y = \beta_0 + \beta_1 x + \cdots + \beta_d x^d + \varepsilon$$

的 D-最优设计的支撑点:

d	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	-1						1
2	-1			0			1
3	-1		-0.4472		0.4472		1
4	-1		-0.6547	0	0.6547		1
5	-1	-0.7651	-0.2852		0.2852	0.7651	1
6	-1	-0.8302	-0.4688	0	0.4688	0.8302	1

例 (多元一次线性回归模型)

考虑 p 个因子的一阶线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon.$$

(Heiligers, 1992) 和 (Chen, 2003) 指出, 如果试验区域 $\mathcal{X}$ 为 $\mathbb{R}^p$ 中的凸集, 则多元线性回归模型的离散或精确 D 最优设计的设计点都在 $\mathcal{X}$ 的顶点上.

由于超球体和超立方体都是凸集, 因此这两种试验区域的 D 最优设计点都在其顶点上. 构造离散或精确 D 最优设计时, 只需在全体顶点的集合中搜索即可.

例 (多元一次线性回归模型 cont.)

当试验区域为超球体 $\mathcal{X}=\left\{\boldsymbol{x}\in\mathbb{R}^p:\sum_{i=1}^{p}x_i^2\leq 1\right\}$ 时，设超球体内嵌正多面体的顶点个数为 n，并记为 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$. 记

$$U_n=\begin{Bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n\\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n}\end{Bmatrix}.$$

- 则当 $n>p$ 时，不管是否存在截距项 β_0，U_n 都是多元一次线性回归模型的 D 最优设计.
- 由于超球体内嵌正多面体有无穷个，而且顶点个数只要大于 p 即可，因此多元线性模型的离散 D 最优设计不唯一.

例 (多元一次线性回归模型 cont.)

当试验区域为超立方体 $\mathcal{X}=\{\boldsymbol{x}\in\mathbb{R}^p:|x_i|\le 1, i=1,2,\cdots,p\}$ 时，设 $S=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_s:\boldsymbol{v}_i\in\mathbb{R}^p, i=1,2,\cdots,s\}$ 表示超立方体的全体顶点. 则

- 离散 D 最优设计为

$$\xi_{n,S}=\left\{\begin{matrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_s\\ \frac{1}{s} & \frac{1}{s} & \dots & \frac{1}{s}\end{matrix}\right\}.$$

- 试验次数为 n 的精确 D 最优设计为

$$\xi_{n,S}=\left\{\begin{matrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_s\\ \frac{n_1}{n} & \frac{n_1}{n} & \dots & \frac{n_s}{n}\end{matrix}\right\},\quad \max_{i,j}|n_i-n_j|\le 1, \sum_{i=1}^s n_i=n.$$

例 (多元二次线性回归)

考虑包含 p 个因子的二阶线性回归模型

$$y = \beta_0 + \sum_{i=1}^{p} \beta_j x_j + \sum_{1 \le i \le j \le p} \beta_{ij} x_i x_j + \varepsilon.$$

当试验区域为超球体时, 截距项会影响 D 最优设计.

- 当截距项不为 0 时, D 最优设计为: 中心点权重为 $1/[(p+1)(p+2)]$, 其余权重均匀分布在超球体的球面上;
- 当截距项为 0 时, D 最优设计为超球体球面上的均匀分布, 不包括中心点.

思考

将一幅画固定在墙上，至少需要几颗钉子？如果给三颗钉子，应该如何钉？给四颗呢？

课堂小结

- 设计的信息矩阵;
- 最优设计的概念;
- 最优设计的字母序准则: A 准则、C 准则、D 准则、G 准则.

3.3 最优回归设计

3.3.1 设计的信息矩阵

3.3.2 优良性准则

3.3.3 等价性定理

3.3.4 D 最优设计的迭代求解

设函数 $f: \mathbb{R}^p \mapsto \mathbb{R}$. $f(\boldsymbol{x})$ 在点 $\boldsymbol{x} \in \mathbb{R}^p$ 处沿着方向 $\boldsymbol{h} \in \mathbb{R}^p$ 的方向导数的定义为

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{h}} = \lim_{\alpha \to 0^+} \frac{1}{\alpha} \left[f(\boldsymbol{x} + \alpha \boldsymbol{h}) - f(\boldsymbol{x}) \right].$$

二元函数 $f(x_1, x_2)$ 沿着方向 $\boldsymbol{h} = (\cos\alpha, \sin\alpha)$ 的方向导数为

$$\frac{\partial f(x_1, x_2)}{\partial \boldsymbol{h}} = \frac{\partial f(x_1, x_2)}{\partial x_1} \cos\alpha + \frac{\partial f(x_1, x_2)}{\partial x_2} \sin\alpha.$$

函数极值点与方向导数的关系

- $\boldsymbol{x}^*$ 为 f 的极小值点, 当且仅当 $\frac{\partial f(\boldsymbol{x}^*)}{\partial \boldsymbol{h}} \geq 0$ 对任意 $\boldsymbol{h} \in \mathbb{R}^p$ 都成立;
- $\boldsymbol{x}^*$ 为 f 的极大值点, 当且仅当 $\frac{\partial f(\boldsymbol{x}^*)}{\partial \boldsymbol{h}} \leq 0$ 对任意 $\boldsymbol{h} \in \mathbb{R}^p$ 都成立.

- 假定 $\Phi : \Xi \mapsto \mathbb{R}^+$ 为凸函数, 即

$$\Phi\left(\alpha\xi + (1-\alpha)\eta\right) \leq \alpha\Phi(\xi) + (1-\alpha)\Phi(\eta),$$

- 要求 Φ 一阶可微. 称

$$F_{\Phi}(\xi,\eta) \;=\; \lim_{\alpha\to 0^+} \frac{1}{\alpha}\left[\Phi\left((1-\alpha)\xi + \alpha\eta\right) - \Phi(\xi)\right].$$

为$\Phi(\cdot)$ 在 ξ 处沿 η 方向的 F-导数.

- ξ^* 是 Φ 的最小值, 当且仅当 $F_{\Phi}(\xi^*,\eta) \geq 0$ 对任意设计 η 成立!

- 若 Φ 在 ξ 处可微, 则

$$F_\Phi\left(\xi, \sum w_i \eta_i\right) = \sum w_i F_\Phi(\xi, \eta_i),$$

其中 $\sum w_i = 1$.

- 以 $\delta_{\boldsymbol{x}}$ 表示在点 $\boldsymbol{x}$ 处权重为 1 的设计, $w(\boldsymbol{x})$ 表示设计 η 在点 $\boldsymbol{x}$ 处的权重则

$$F_\Phi(\xi, \eta) = \sum_{\boldsymbol{x}} w(\boldsymbol{x}) F_\Phi(\xi, \delta_{\boldsymbol{x}}).$$

- 称 $\phi(\boldsymbol{x}, \xi) := F_\Phi(\xi, \delta_{\boldsymbol{x}})$ 为敏感性函数.

定理 (等价性定理)

如果 Φ 为凸函数, 且在 Ξ 中的所有点处可微, 则下列命题等价:

(1) ξ^* 是 Φ-最优设计;

(2) 对任意 $\boldsymbol{x} \in \mathcal{X}$, 都有 $\phi(\boldsymbol{x}, \xi^*) \geq 0$;

(3) $\phi(\boldsymbol{x}, \xi^*)$ 在 ξ^* 的所有支撑点上都取最小值, 且最小值为 0.

- 应用到具体准则时, 需要验证其凸性和可微性, 并给出敏感性函数.

定理 (D 最优设计的等价性定理)

对于 D 最优设计而言, $-\log\det[\boldsymbol{M}(\xi)]$ 作为 Ξ 上的函数是一个凹函数, $\phi_D(\boldsymbol{x},\xi)=m-d(\boldsymbol{x},\xi)$, 且以下三个结论等价:

(1) ξ^* 是 D 最优设计, 即 $\det(\boldsymbol{M}(\xi^*))=\max_\xi\det(\boldsymbol{M}(\xi))$;

(2) ξ^* 是 G 最优设计, 即 $\max_{\boldsymbol{x}} d(\boldsymbol{x},\xi^*)=\min_\xi\max_{\boldsymbol{x}} d(\boldsymbol{x},\xi)$;

(3) ξ^* 满足 $\max_{\boldsymbol{x}} d(\boldsymbol{x},\xi^*)=m$, 且 $d(\boldsymbol{x},\xi^*)$ 在 ξ^* 的任一支撑点达到最大.

此外, 所有 D 最优设计有相同的信息矩阵, D 最优设计的线性组合还是 D 最优设计.

3.3 最优回归设计

3.3.1 设计的信息矩阵

3.3.2 优良性准则

3.3.3 等价性定理

3.3.4 D 最优设计的迭代求解

- 最优设计一般不存在解析表达式, 只能数值求解.
- 在等价性定理的基础上, 前苏联统计学家 Fedorov 给出了一个构造 D 最优设计的迭代算法.
- (Kernighan and Lin, 1970) 提出构造精确 D-最优设计 KL 算法.

Federov 迭代算法

Step 1 构造非奇异的初始设计 $\xi_0 = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$.

Step 2 对 $k = 0, 1, \cdots,$ 求

$$\begin{cases} \boldsymbol{x}_{n+k+1} = \arg\max\limits_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x}, \xi_k), \\ \alpha_k = \arg\max\limits_{\alpha \in [0,1]} \det \boldsymbol{M}(\xi_{k+1}(\alpha)). \end{cases}$$

这里 $\xi_{k+1}(\alpha) = (1-\alpha)\xi_k + \alpha\delta_{\boldsymbol{x}_{n+k+1}}$. 可以证明

$$\alpha_k = \frac{d(\boldsymbol{x}_{n+k+1}, \xi_k) - m}{[d(\boldsymbol{x}_{n+k+1}, \xi_k) - 1]m}.$$

- 当 $k \to \infty$ 时, 序列 ξ_k 收敛到 D 最优设计.

Federov 迭代算法

Step 1 构造非奇异的初始设计 $\xi_0 = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$.

Step 2 对 $k = 0, 1, \cdots,$ 求

$$\begin{cases} \boldsymbol{x}_{n+k+1} = \arg\max\limits_{\boldsymbol{x}\in\mathcal{X}} d(\boldsymbol{x}, \xi_k), \\ \alpha_k = \arg\max\limits_{\alpha\in[0,1]} \det \boldsymbol{M}(\xi_{k+1}(\alpha)). \end{cases}$$

这里 $\xi_{k+1}(\alpha) = (1-\alpha)\xi_k + \alpha\delta_{\boldsymbol{x}_{n+k+1}}$. 可以证明

$$\alpha_k = \frac{d(\boldsymbol{x}_{n+k+1}, \xi_k) - m}{[d(\boldsymbol{x}_{n+k+1}, \xi_k) - 1]m}.$$

- 当 $k \to \infty$ 时, 序列 ξ_k 收敛到 D 最优设计.

构造精确设计的 KL 算法

(1) 产生试验次数为 n_0 的确定性设计 ξ_0.

- 由部分希望试验的点和部分随机抽取的点组成.

构造精确设计的 KL 算法

(1) 产生试验次数为 n_0 的确定性设计 ξ_0.

(2) 添加设计点 $(n_0 < n)$ 或删除设计点 $(n_0 > n)$, 得到试验次数为 n 的初始设计 η_0.

- 前进法: 添加使设计 ξ_i 的标准化方差达到最大的点,

$$d(\boldsymbol{x}_l, \xi_i) = \max_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x}, \xi_i).$$

  当 $\boldsymbol{M}(\xi_i)$ 不可逆时, 以 $\boldsymbol{M}(\xi_i) + \varepsilon \boldsymbol{I}$ 来代替, 其中 $10^{-6} < \varepsilon < 10^{-4}$.

- 后退法: 删除使设计 ξ_i 的标准化方差达到最小的点,

$$d(\boldsymbol{x}_k, \xi_{i-1} \setminus \boldsymbol{x}_k) = \min_{\boldsymbol{x} \in \xi_{i-1}} d(\boldsymbol{x}, \xi_{i-1} \setminus \{\boldsymbol{x}\}).$$

构造精确设计的 KL 算法

(1) 产生试验次数为 n_0 的确定性设计 ξ_0.

(2) 添加设计点 $(n_0 < n)$ 或删除设计点 $(n_0 > n)$, 得到试验次数为 n 的初始设计 η_0.

(3) 对初始设计 η_0 的点进行替换, 直至收敛.
   - 替换的目的是使得信息矩阵的行列式增大, 由标准化方差的变动来实现.

总结

1. 线性模型的信息矩阵;
2. 最优设计的思想与概念;
3. 最优设计的常用准则;
4. 等价性定理;
5. 线性模型最优设计的求解算法.