STAT243 Lecture 9.1 Monte Carlo Considerations

1 回归示例动机 (Motivating Example)

考虑一个**线性回归模型**：

$$Y = X\beta + \varepsilon$$

其中：

- $Y = (y_1, y_2, \ldots, y_n)$ 为响应变量；
- X 为 $n \times p$ 的设计矩阵；
- $\varepsilon \sim N(0, \sigma^2 I)$。

标准最小二乘估计为：

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y$$

根据经典假设（线性期望与独立同方差误差），可得：

$$\begin{aligned} E\hat{\beta} &= \beta \\ \mathrm{Var}(\hat{\beta}) = E((\hat{\beta} - E\hat{\beta})^2) &= \sigma^2 (X^\top X)^{-1} \\ \mathrm{MSPE}(Y^*) = E(Y^* - \hat{Y})^2) &= \sigma^2 (1 + X^{*\top} (X^\top X)^{-1} X^*). \end{aligned}$$

其中 Y^* 为新样本的预测值。

1.1 当模型假设不成立时

如果：

- 均值与 X 的关系**非线性**；
- 误差项**非独立或异方差**；
- 或者我们使用了**稳健估计方法**（例如抗离群点估计）——

则上述解析结果不再适用。
此时我们可使用 **Monte Carlo 模拟** 来研究估计量的性质。

1.2 Monte Carlo 回归模拟的基本思路

若生成 m 个独立样本集 $Y_1, \ldots, Y_m$ 来估计 β，并在第 i 个样本集上得到 $\hat{\beta}_i$，则：

- 估计期望：

$$\widehat{E[\hat{\beta}]} = \bar{\hat{\beta}} = \frac{1}{m} \sum_{i=1}^{m} \hat{\beta}_i$$

- 估计方差：

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \frac{1}{m} \sum_{i=1}^{m} (\hat{\beta}_i - \bar{\hat{\beta}})^2$$

Monte Carlo 方法能帮助我们评估在“非标准假设”下的估计稳定性与偏差。

2 蒙特卡洛方法基础 (Monte Carlo Basics)

2.1 核心思想 (Overview)

我们通常希望估计某个期望：

$$\phi = E_f[h(Y)]$$

其中 $Y \sim f$，$h(Y)$ 为感兴趣的函数。

2.1.1 典型示例

若 $h(Y) = I(Y \leq y)$，则：

$$\phi = P(Y \leq y) = E_f[I(Y \leq y)]$$

表示累积分布函数 $F(y)$。

2.2 蒙特卡洛估计量 (Monte Carlo Estimator)

若从分布 f 独立采样 $Y_1, \ldots, Y_m$，则：

$$\hat{\phi} = \frac{1}{m} \sum_{i=1}^{m} h(Y_i)$$

根据大数定律：

$$\hat{\phi} \xrightarrow[m \to \infty]{} E_f[h(Y)] = \phi$$

在实际模拟中，Y_i 通常是一个完整的数据集，而非单个观测值。
若每个样本有 n 个观测，则 $Y_i = (Y_{i1}, \ldots, Y_{in})$。

2.3 回归问题中的对应关系 (Back to Regression Example)

若我们希望检验回归估计量是否有偏：

$$\phi = E[\hat{\beta}], \quad h(Y) = \hat{\beta}(Y)$$

则 Monte Carlo 估计为：

$$\hat{\phi} = \frac{1}{m} \sum_{i=1}^{m} \hat{\beta}_i$$

若要估计方差：

$$\phi = \mathrm{Var}(\hat{\beta}) = E[(\hat{\beta} - E\hat{\beta})^2]$$

则：

$$\hat{\phi} = \frac{1}{m} \sum_{i=1}^{m} (\hat{\beta}_i - \bar{\hat{\beta}})^2$$

2.4 置信区间覆盖率 (Confidence Interval Coverage)

我们可进一步评估置信区间的覆盖概率：

$$h(Y) = I[\beta \in CI(Y)]$$

Monte Carlo 估计：

$$\hat{\phi} = \frac{1}{m}\sum_{i=1}^{m} I[\beta \in CI(Y_i)]$$

理想情况下，$\hat{\phi} \approx 1-\alpha$（例如 0.95）。

覆盖率偏低的原因：

1. 估计的方差过小；
2. 估计量存在偏差。

---

3 模拟误差 (Simulation Uncertainty)

3.1 Simulation uncertainty

因为 $\hat{\phi}$ 是 $h(Y_i)$ 的样本均值，
其方差为：

$$\mathrm{Var}(\hat{\phi}) = \frac{\sigma^2}{m}, \quad \sigma^2 = \mathrm{Var}(h(Y))$$

样本估计为：

$$\hat{\sigma}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(h(Y_i) - \hat{\phi})^2$$

从而：

$$\widehat{\mathrm{Var}}(\hat{\phi}) = \frac{\hat{\sigma}^2}{m} = \frac{1}{m(m-1)}\sum_{i=1}^{m}(h(Y_i) - \hat{\phi})^2$$

其阶为 $O(1/m)$，标准误差随 $1/\sqrt{m}$ 减小。

3.2 区分模拟误差与统计误差

> ⚠ **Remark**
>
> **Monte Carlo 不确定性 ≠ 统计不确定性**
>
> - Monte Carlo 不确定性：来自模拟随机数的误差
> - 统计不确定性：来自数据生成过程的不确定性

我们可以通过增大 m 来减少 Monte Carlo 方差，但其收敛速度为 $\propto 1/\sqrt{m}$。

3.3 回归示例中的模拟方差 (Regression Example)

可关心的 Monte Carlo 方差包括：

- $\widehat{\mathrm{Var}}(\hat{E}[\hat{\beta}] - \beta)$：偏差估计的不确定性；
- $\widehat{\mathrm{Var}}(\widehat{\mathrm{Var}}(\hat{\beta}))$：方差估计的不确定性；
- $\widehat{\mathrm{Var}}(\widehat{\mathrm{MSPE}}(Y^*))$：预测误差估计的不确定性。

在每种情况下，$\widehat{\mathrm{Var}}()$ 表示估计模拟方差。

---

4 方差缩减技巧 (Variance Reduction, Optional)

4.1 常见方法

- **重要性抽样 (Importance Sampling)**
- **对照变量 (Control Variates)**
- **反对称抽样 (Antithetic Sampling)**

后两种在实际中使用较少，但思想上可减少模拟方差。

4.2 分层模拟 (Stratified Simulation)

若能将总体划分为若干已知比例的分层，可在各层内分别估计均值 μ_h，再根据层权重加权组合。
这种做法能减少层间抽样带来的方差。

4.3 Rao-Blackwell化 (Rao-Blackwellization)

若希望计算：

$$E[h(X)], \quad X = (X_1, X_2)$$

根据迭代期望：

$$E[h(X)] = E[E(h(X) \mid X_2)]$$

若能解析计算 $E(h(X) \mid X_2)$，则无需为 X_1 额外引入随机性。
Rao-Blackwell 化的估计为：

$$\hat{\mu}_{RB} = \frac{1}{m} \sum_{i=1}^{m} E[h(X) \mid X_{2,i}]$$

其方差更小，因为：

$$\mathrm{Var}(E[h(X) \mid X_2]) < \mathrm{Var}(h(X))$$

这是由于：

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

因此，Rao-Blackwell化本质上利用了**条件期望降低方差**的思想。