

# 贝叶斯线性回归

## Bayesian Linear Regression

汇报人：庄亮亮

温州大学 数理学院

2020 年 11 月

# 目录

## 1 Introduction

## 2 Bayesian Inference for Linear Regression

## 3 Prediction

## 4 Model Selection and Checking

## 5 Robust Regression

## 6 Analysis of Variance

## 7 Ridge Regression for Multicollinearity

## 8 Regression with Autoregressive Errors

## 简介

- 线性回归是为**因变量**和一个或多个**解释变量**之间的关系建模的最常见的统计方法之一。
- 本章可以作为我们将在后面章节中讨论的更新和更复杂的建模方法的良好起点。

## 一般形式

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$(\beta_0, \beta_1, \dots, \beta_p)$  为未知模型参数;  $\varepsilon$  是误差变量。

## 矩阵形式

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

其中,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

假设  $\varepsilon$  服从  $N(0, \sigma^2 \mathbf{I})$ , 方差  $\sigma^2$  未知。

## 频率派观点

### 似然函数

$$L(\beta, \sigma^2 | \mathbf{X}, \mathbf{y}) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \right]$$

分别对  $\beta, \sigma^2$  进行求导：

$$\begin{aligned} S_1(\beta, \sigma^2) &= \frac{\partial \log L}{\partial \beta} = -\frac{1}{2\sigma^2} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta) \\ S_2(\beta, \sigma^2) &= \frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta). \end{aligned}$$

假设  $\mathbf{X}^T \mathbf{X}$  满秩，令上述式子为 0，得到估计值：

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad \hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta}).$$

但是这的  $\hat{\sigma}^2$  是有偏的，无偏估计如下：

$$S^2 = \frac{1}{n - p - 1} (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta})$$

# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression**
- 3 Prediction
- 4 Model Selection and Checking
- 5 Robust Regression
- 6 Analysis of Variance
- 7 Ridge Regression for Multicollinearity
- 8 Regression with Autoregressive Errors

## 贝叶斯派观点

在贝叶斯分析中，方差参数的逆起着重要作用，称为精度  
 $\tau = \sigma^{-2}$ 。

根据模型的假设，我们有

$$\mathbf{y} \mid \boldsymbol{\beta}, \tau \sim N(\mathbf{X}\boldsymbol{\beta}, \tau^{-1}\mathbf{I})$$

我们进一步假设， $\boldsymbol{\beta}$  和  $\tau$  是独立的。因此，未知参数的联合后验分布为

$$\pi(\boldsymbol{\beta}, \tau \mid \mathbf{X}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \tau \mid \mathbf{X}, \mathbf{y})p(\boldsymbol{\beta})p(\tau)$$

其中  $p(\boldsymbol{\beta})$  和  $p(\tau)$  是参数  $\boldsymbol{\beta}$  和  $\tau$  的先验。后验分布的封闭形式只有在某些受限制的先验分布下才能得到。

## 贝叶斯派观点

### 先验的选取

- 已知先验信息——将其纳入先验分布中。
- 未知先验信息——先验分布能够保证对推断有较小的影响，例如， $p(\beta) \propto 1$ 。非信息先验的**主要缺点**是它对参数的变换不是不变的，更多讨论见附录 B。



## 例子

假设模型是一个隐高斯模型，典型的先验例子为：

$$\beta \sim N_{p+1}(\mathbf{c}_0, \mathbf{V}_0), \quad \tau \sim \text{Gamma}(a_0, b_0)$$

$c_0, V_0, a_0, b_0$  已知，通常假设  $V_0$  是对角矩阵。

虽然在这些先验下后验是难以处理的，但可以直接构建一个 Gibbs 采样算法并使用 MCMC 实现 (Gelman et al., 2014)。具体来说，算法在两个条件分布之间进行迭代：

$$\begin{cases} \pi(\beta \mid \mathbf{X}, \mathbf{y}, \tau) \propto L(\beta, \tau \mid \mathbf{X}, \mathbf{y})p(\beta) \\ \pi(\tau \mid \mathbf{X}, \mathbf{y}, \beta) \propto L(\beta, \tau \mid \mathbf{X}, \mathbf{y})p(\tau) \end{cases}$$

## INLA

①  $\pi(\tau | \mathbf{X}, \mathbf{y})$  近似:

$$\tilde{\pi}(\tau | \mathbf{X}, \mathbf{y}) \propto \frac{\pi(\boldsymbol{\beta}, \tau, \mathbf{X}, \mathbf{y})}{\tilde{\pi}(\boldsymbol{\beta} | \tau, \mathbf{X}, \mathbf{y})} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*(\tau)} \quad (1)$$

②  $\tilde{\pi}(\beta_j | \mathbf{X}, \mathbf{y})$  近似:

$$\begin{aligned} \tilde{\pi}(\beta_j | \mathbf{X}, \mathbf{y}) &= \int \tilde{\pi}(\beta_j | \tau, \mathbf{X}, \mathbf{y}) \tilde{\pi}(\tau | \mathbf{X}, \mathbf{y}) d\tau \\ &\approx \sum_k \tilde{\pi}(\beta_j | \tau_k, \mathbf{X}, \mathbf{y}) \tilde{\pi}(\tau_k | \mathbf{X}, \mathbf{y}) \Delta_k \end{aligned} \quad (2)$$

## 具体过程

- 1 求得  $\tilde{\pi}(\tau | \mathbf{X}, \mathbf{y})$ , 找到其峰值点并定位一组点  $\{\tau_1, \dots, \tau_K\}$
- 2 代入计算  $\tilde{\pi}(\tau_1 | \mathbf{X}, \mathbf{y}), \dots, \tilde{\pi}(\tau_K | \mathbf{X}, \mathbf{y})$
- 3 使用简易拉普拉斯近似求  $\tilde{\pi}(\beta_j | \tau_k, \mathbf{X}, \mathbf{y}), k = 1, \dots, K, \circ$
- 4 计算  $\tilde{\pi}(\beta_j | \mathbf{X}, \mathbf{y}) = \sum_{k=1}^K \tilde{\pi}(\beta_j | \tau_k, \mathbf{X}, \mathbf{y}) \tilde{\pi}(\tau_k | \mathbf{X}, \mathbf{y}) \Delta_k$  其中  $\Delta_k$  为权重。

## 其他

从 INLA 过程中获得的近似后验边际可用于计算相关数据：

- 1 后验均值、方差和分位数；
- 2 偏差信息标准 (DIC)、边际概率（比较和验证模型中常用）。

# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression
- 3 Prediction**
- 4 Model Selection and Checking
- 5 Robust Regression
- 6 Analysis of Variance
- 7 Ridge Regression for Multicollinearity
- 8 Regression with Autoregressive Errors

# Posterior predictive distribution

新的数据：解释变量为  $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_p)^T$ ，预测  $\tilde{y}$ 。

## 贝叶斯预测

基于后验预测分布  $p(\tilde{y}|\mathbf{y})$ ，令  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau)$ 。

$$\begin{aligned} p(\tilde{y} | \mathbf{y}) &= \frac{p(\tilde{y}, \mathbf{y})}{p(\mathbf{y})} = p(\mathbf{y})^{-1} \int p(\tilde{y} | \boldsymbol{\theta}) p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= p(\mathbf{y})^{-1} \int p(\tilde{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) p(\mathbf{y}) d\boldsymbol{\theta} \\ &= \int p(\tilde{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \end{aligned}$$

大多数回归模型的后验预测分布没有解析形式。在传统的贝叶斯分析中，预测可以通过后验预测模拟进行，即从  $p(\tilde{y} | \mathbf{y})$  中随机抽取样本。

# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression
- 3 Prediction
- 4 Model Selection and Checking**
- 5 Robust Regression
- 6 Analysis of Variance
- 7 Ridge Regression for Multicollinearity
- 8 Regression with Autoregressive Errors

# 模型选择

## Model Selection by DIC

在第一章中，我们讨论了贝叶斯模型的选择和检验。这里，主要在 INLA 框架下使用空气污染数据进行分析。

- ①  $AIC = -2 \log p(\mathbf{y} \mid \hat{\theta}_{mle}) + 2k$
- ②  $DIC = -2 \sum_{i \in \mathcal{I}} \log \{p(y_i \mid \beta_j, \tau)\} + constant$

inlas 中，设置 `control.compute = list(dic = TRUE)` 即可。



## 模型拟合检验

在统计分析中，模型拟合检验非常重要。在贝叶斯分析中，模型评估通常是：

- 1 后验预测模型检验 (Posterior Predictive Model Checking)
- 2 留一交叉验证预测模型检验 (Leave-one out Cross-Validation Model Checking)

# 模型拟合检验

## Posterior Predictive Model Checking

重复观察  $y_i^*$  的后验预测分布，它的密度函数是：

$$p(y_i^* | \mathbf{y}) = \int p(y_i^* | \theta) p(\theta | \mathbf{y}) d\theta$$

相应的后验预测  $p$  值为

$$p(y_i^* \leq y_i | \mathbf{y})$$

# 代码实现

```
1 > inla.pred <- inla(formula, data = ..., control.
    predictor = list(link = 1, compute = TRUE)
    )
2 > post.predicted.pval <- vector(mode = "numeric",
    length = nrow(usair))
3 > for(i in (1:nrow(usair))) {
4   post.predicted.pval[i] <-
5     nla.pmarginal(q=usair$S02[i], marginal =
6     inla.pred$marginals.fitted.values[[i]])
7 }
8 > hist(post.predicted.pval, main="", breaks = 10,
    xlab="Posterior predictive p-value")
```

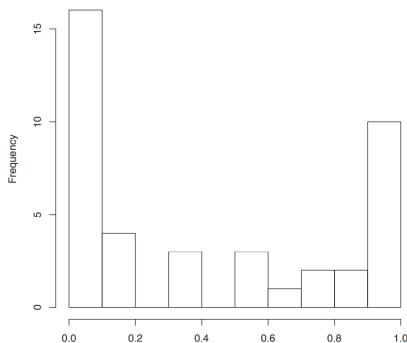


图 1: Posterior predictive p-value.

很多后验预测  $p$  值都接近于 0 或 1。然而，解释后验预测  $p$  值的一个缺点是，即使数据来自真实的模型，它们也不能具有均匀分布。

# 模型拟合检验

## Cross-Validation Model Checking

评价模型的优劣的两个量：

### ① Conditional Predictive Ordinate (CPO)

$$\text{CPO}_i = p(y_i | \mathbf{y}_{-i})$$

极端值表示一个令人惊讶的观察结果；

### ② Probability Integral Transform (PIT)

$$\text{PIT}_i = p(y_i^* \leq y_i | \mathbf{y}_{-i})$$

柱状图看起来不是均匀分布，表明不适合当前的模型。

注意， $PIT$  和后验预测  $p$  值之间的唯一区别是： $PIT$  是基于  $y_{-i}$  计算的，而不是基于  $y$ 。

# 模型拟合检验

## Cross-Validation Model Checking

如果我们把所有  $CPO$  值的乘积看作一个“伪边际似然”，这就给出了一个交叉验证的拟合度量。Geisser 和 Eddy(1979) 提出的对数拟边际似然 ( $LPML$ ):

$$LPML = \log \left\{ \prod_{i=1}^n p(y_i | \mathbf{y}_{-i}) \right\} = \sum_{i=1}^n \log p(y_i | \mathbf{y}_{-i}) = \sum_{i=1}^n \log CPO_i \quad (3)$$

通常被用作  $DIC$  的另一种测量方法。Draper and Krnjajic (2007, Sec. 4.1) 表明对于近似高斯后验,  $DIC$  近似于  $LPML$ 。

# 贝叶斯残差分析

## Bayesian Residual Analysis

$$r_i = \varepsilon_i(\beta) = y_i - x_i^T \beta, \quad i = 1, \dots, n \quad (4)$$

称为“现实残差”，与经典的残差估计  $y_i - x_i^T \hat{\beta}$  不同。这些残差的先验分布为  $N(0, \sigma^2)$ ，它们的后验分布可由  $\beta$  和  $\sigma^2$  的后验分布得到。

# 贝叶斯残差分析

## 残差图

检验残差图是回归分析中模型诊断的一种常用方法。

- 1 从  $\beta$  和  $\sigma^2$  的后验分布中生成样本；
- 2 将这些样本代入 (4) 中，得到残差后验分布的样本。计算  $r_i$  的后验均值或中位数；
- 3 将这些残差与拟合值进行绘图，可以检测出异常值，也可以揭示模型假设（如正态性或方差齐性假设）。

INLA 中，使用 `bri.lmresid.plot` 生成贝叶斯残差图。



# 贝叶斯残差分析

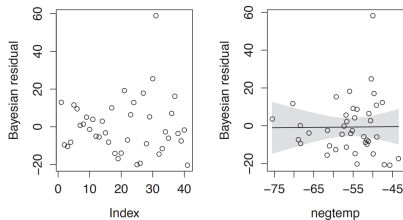


图 2: 贝叶斯残差图

# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression
- 3 Prediction
- 4 Model Selection and Checking
- 5 Robust Regression**
- 6 Analysis of Variance
- 7 Ridge Regression for Multicollinearity
- 8 Regression with Autoregressive Errors

# 稳健回归

基于正态分布的统计推断往往容易受到异常值的影响。Lange 等人 (1989) 引入了  $t$  分布误差回归模型，作为传统正态模型的稳健扩展。可用于研究非正态的残差，以及存在极端点的情况。

## $t$ 分布误差回归模型

假设响应变量  $Y$  服从  $t$  分布  $t(\mu, \sigma, \nu)$ ，其概率密度函数为：

$$f(y \mid \mu, \sigma, \nu) = \frac{v^{v/2} \Gamma((v+1)/2)}{\sigma \sqrt{\pi} \Gamma(v/2)} \left\{ v + \left( \frac{y - \mu}{\sigma} \right)^2 \right\}^{-(v+1)/2}$$

$\mu$  是位置参数,  $\sigma$  是扩散参数,  $\nu$  是自由度,  $\Gamma(\cdot)$  是伽马函数。  
 $t(\mu_i, \sigma, \nu)$  可以看作是正态分布和伽马分布的混合。

如果一个潜在变量  $\eta_i \sim \Gamma(v/2, v/2)$ ,  $y_i \mid \eta_i \sim N(\mu_i, \sigma^2/\eta_i)$ , 那么  $y_i \sim t(\mu_i, \sigma, \nu)$ 。

## $t$ 分布误差回归模型

$$\begin{cases} y_i \sim t(\mu_i, \sigma, \nu) \\ \mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n \end{cases} \quad (5)$$

参数  $\sigma$  是未知的扩散参数，它模拟了  $y_i$  的方差。在 INLA 中，我们建立了精确参数  $\tau = \sigma^{-2}$  的模型。

在 INLA 中，通过指定 `family= "T"` 实现。

# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression
- 3 Prediction
- 4 Model Selection and Checking
- 5 Robust Regression
- 6 Analysis of Variance**
- 7 Ridge Regression for Multicollinearity
- 8 Regression with Autoregressive Errors

# 方差分析

用来确定两组或更多组数据之间的差异或相似程度的方法。方差分析是**带有分类变量回归**的一个特殊情况。

## 单因素情况

水平为  $i = 1, 2, \dots, a$ , 假设响应变量  $Y$  为正态分布, 且该变量只影响  $Y$  的均值, 则模型为

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad (6)$$

其中  $y_{ij}$  是  $i^{th}$  因子水平的观察值  $j^{th} (j = 1, \dots, n_i)$ , 且误差为  $\varepsilon_{ij} \sim N(0, \sigma^2)$ 。

模型经常写成:

$$y_{ij} = \mu_0 + \alpha_i + \varepsilon_{ij} \quad (7)$$

有  $a + 1$  参数需要估计。

# 方差分析

## 角约束

为了使模型可识别，我们需要对参数进行约束。统计文献中常用的约束是**角约束** (corner constraint)。在  $r \in \{1, 2, \dots, \alpha\}$  中一个水平  $r$  的效果被设置为 0:  $\alpha_r = 0$ 。这个水平  $r$  被称为因子  $a$  的参考水平。通常，我们使用第一个水平作为参考水平，即：

$$\alpha_1 = 0.$$

因此， $\mu_0$  成为第一水平的平均值。“INLA” 包中就使用了角约束。

# 方差分析

## 多因素方差分析

$2 \times 2$  阶乘设计：因子 A 的每个水平与因子 B 的每个水平有交互作用。假设每个 cell 都观察  $n$  次。观测  $y_{ijk}$  模型为：

$$y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (8)$$

$i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, n$ 。  $\mu_0$  是总平均值，  $\gamma_{ij}$  是 A 第  $i$  水平和 B 第  $j$  水平的交互效应，误差服从  $\varepsilon_{ijk} \sim N(0, \sigma^2)$ 。

模型包含  $1 + a + b + ab$  参数，但数据只提供  $ab$  样本均值。因此，我们需要给参数加上合理约束：

$$\alpha_1 = \beta_1 = \gamma_{11} = \dots = \gamma_{1b} = \gamma_{21} = \dots = \gamma_{a1} = 0 \quad (9)$$



# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression
- 3 Prediction
- 4 Model Selection and Checking
- 5 Robust Regression
- 6 Analysis of Variance
- 7 Ridge Regression for Multicollinearity**
- 8 Regression with Autoregressive Errors

## 多重共线性

如果数据中存在多重共线性，使用普通最小二乘的统计推断可能不可靠。

### 存在问题

- 1 偏回归系数可能无法精确估计，标准误差通常很高；
- 2 评估预测因子在解释响应变量引起的变化时的相对重要性变得困难。

### 解决方法

- 1 Principal component regression  
通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，在使用这些主成分作为回归变量进行回归。
- 2 Ridge regression  
通过对系数的大小施加惩罚来解决普通最小二乘的问题。

# 岭回归

## Ridge regression

对带惩罚项的残差平方和最小化来得到系数：

$$\min_{\beta} \{ (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \} + \lambda \beta^T \beta \quad (10)$$

其中  $\lambda \geq 0$  是一个控制收缩量的复杂参数 (complexity parameter) :  $\lambda$  的值越大, 收缩量越大。

利用最小二乘法得到的岭回归估计量：

$$\hat{\beta}_{\text{ridge}}(\lambda) = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (11)$$

这种方法会引起偏差 (随着  $\lambda$  的增加而增加), 但是会获得更精确的参数估计 (例如, 参数的方差更小)。

# 岭回归

## Ridge regression

事实上，岭回归与标准后验贝叶斯回归估计密切相关，只不过回归参数使用了特定的先验分布。

假设  $\beta_i$  的先验分布为  $\beta_j \sim N(0, \sigma_0^2/\lambda), j = 1, \dots, p$ ，其中  $\sigma_0$  已知。 $\lambda$  值很大的话，那么对应的先验集中在零附近，因此导致更大的收缩，趋近于零。

给定数据  $(X, y)$ ， $\beta$  后验分布的均值为：

$$(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (12)$$

# 目录

- 1 Introduction
- 2 Bayesian Inference for Linear Regression
- 3 Prediction
- 4 Model Selection and Checking
- 5 Robust Regression
- 6 Analysis of Variance
- 7 Ridge Regression for Multicollinearity
- 8 Regression with Autoregressive Errors**

## 存在问题

在标准线性回归中，我们假设误差项  $\varepsilon$  服从  $N(0, \sigma^2 \mathbf{I}_n)$ 。但在许多应用中，也可能出现异方差或相关的情况。误差更普遍的假设为：

$$\varepsilon \sim N(0, \Sigma) \quad (13)$$

其中误差协方差矩阵  $\Sigma$  是对称且正定的。不同的对角线项在  $\Sigma$  中对应着非恒定的误差方差，而非零的不在对角线项的  $\Sigma$  中对应着有相关性的误差。

## 频率论分析

假设  $\Sigma$  已知，在频率统计中，广义最小二乘是极小化以下方程：

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \Sigma^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

解得：

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

协方差矩阵

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^T$$

协方差矩阵  $\Sigma$  在应用中通常是未知的，必须从数据中进行估计。但是，矩阵中有  $n(n+1)/2$  不同的元素。没有任何进一步的限制是不可能估计模型的。我们经常使用少量的参数来建模  $\Sigma$ 。

## 典型例子

在时间序列数据中，通过一段时间内的重复测量获得了一系列的观察结果。相邻的观测结果往往是相似的，观测值之间的这种相关性称为**自相关性**。忽略自相关的存在可能会对回归分析产生影响。

### 时间序列数据

考虑一个具有自回归误差的线性回归模型：

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_p x_{tp} + \varepsilon_t, \quad t = 1, \dots, n$$

其中  $E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2$ ，两个误差之间的相关性只取决于它们的时间间隔： $\text{Cor}(\varepsilon_t, \varepsilon_{t+s}) = \text{Cor}(\varepsilon_t, \varepsilon_{t-s}) = \rho_s$ ，其中  $\rho_s$  称为滞后  $s$  处的误差自相关。



# 时间序列数据

## first-order autoregressive process(AR(1))

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t, \quad t = 1, \dots, n \quad (14)$$

其中  $\eta_t \sim N(0, \sigma_\eta^2)$  独立。此时，误差协方差矩阵的结构如下：

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \quad (15)$$

其中,  $\sigma^2 = \sigma_n^2 / (1 - \rho^2)$ . 在这个模型中, 误差自相关  $\rho_s = \rho^s, s = 1, 2, \dots$  随着  $s$  的增加呈指数衰减到 0。

## 时间序列数据

### 拓展

高阶自回归模型可以直接由  $AR(1)$  模型推广而来。

其他常用的时间序列模型包括**移动平均 (MA) 过程**和 **ARMA 过程**。详见 Shumway 和 Stoffer(2011)。

# 时间序列数据

## 残差检验

设  $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n$  为回归模型的残差。残差的样本自相关函数定义为

$$\hat{\rho}(s) = \hat{\gamma}(s) / \hat{\gamma}(0) \quad (16)$$

其中,  $\hat{\gamma}(s) = n^{-1} \sum_{t=1}^{n-s} (\hat{\varepsilon}_{t+s} - \sum_{t=1}^n \hat{\varepsilon}_t / n) (\varepsilon_t - \sum_{t=1}^n \hat{\varepsilon}_t / n)$  为样本自协方差函数。

## 时间序列数据

一个  $AR(s)$  过程具有指数衰减的自相关函数和在第一个  $s$  时滞处具有非零峰值的偏自相关函数。

因此，在回归分析中，检查标准线性回归的残差自相关性和偏自相关性，可以帮助我们检验模型。

### Dubin–Watson 检验

$$d_s = \frac{\sum_{t=s+1}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-s})^2}{\sum_{t=s+1}^n \hat{\varepsilon}_t^2} \quad (17)$$

$H_0 : d_s = 0$ ，随机误差项不存在一阶序列相关。

在贝叶斯统计中，Dreze 和 Mouchart(1990) 使用经典的 Durbin-Watson 统计和检查残差的自相关图作为快速检查。

Bauwens 和 Rasquero(1993) 提出了两种残差自相关贝叶斯检验，检验误差自回归过程的参数的一个近似最高后验密度区域是否包含零假设。

# 致谢

致谢

谢谢大家。