

# 应用时间序列分析

## 课堂讲义

原著: 何书元  
制作: 李东风

2011年秋季学期



# 目 录

<b>第一章 时间序列</b>	<b>1</b>
1.1 时间序列的分解	2
1.1.1 时间序列概念	2
1.1.2 时间序列的分解	3
1.1.3 时间序列和随机过程	6
1.2 平稳序列	7
1.2.1 平稳序列及其自协方差函数	7
1.2.2 白噪声	10
1.2.3 正交平稳序列	12
1.3 线性平稳序列和线性滤波	13
1.3.1 有限运动平均	13
1.3.2 线性平稳序列	14
1.3.3 时间序列的线性滤波	17
1.4 正态时间序列和随机变量的收敛性	19
1.4.1 随机向量的数学期望和方差	19
1.4.2 正态平稳序列	20
1.5 严平稳序列及其遍历性	22
1.6 Hilbert空间中的平稳序列	24
1.6.1 Hilbert空间	24
1.6.2 内积的连续性	26
1.6.3 复值随机变量	28
1.7 平稳序列的谱函数	30
1.7.1 平稳序列的谱函数	30
1.7.2 线性滤波与谱	32
1.7.3 采样定理	34
1.8 离散谱序列及其周期性	36
<b>第二章 自回归模型</b>	<b>39</b>
2.1 推移算子和常系数差分方程	40
2.1.1 推移算子	40
2.1.2 常系数齐次线性差分方程	41
2.1.3 非齐次线性差分方程	44
2.2 自回归模型及其平稳性	45
2.2.1 特例: $AR(1)$	45
2.2.2 一般 $AR(p)$	46

2.2.3	平稳解和通解	46
2.3	AR( $p$ )序列的谱密度和Yule-Walker方程	52
2.3.1	AR( $p$ )序列的谱密度	52
2.3.2	Yule-Walker方程	54
2.3.3	自协方差函数的周期性	56
2.3.4	自协方差函数的正定性	58
2.3.5	时间序列的可完全预测性	60
2.4	平稳序列的偏相关系数和Levinson递推公式	62
2.4.1	最优线性预测	62
2.4.2	最小相位性	64
2.4.3	Levinson递推公式	64
2.4.4	偏相关系数	65
2.5	AR( $p$ )序列举例	68
<b>第三章</b>	<b>滑动平均模型与自回归滑动平均模型</b>	<b>73</b>
3.1	滑动平均模型	74
3.1.1	模型引入	74
3.1.2	MA( $q$ )模型和MA( $q$ )序列	74
3.1.3	最小序列	77
3.1.4	MA( $q$ )系数的递推计算	77
3.1.5	MA( $q$ )模型举例	78
3.2	自回归滑动平均模型	81
3.2.1	ARMA( $p, q$ )模型及其平稳解	81
3.2.2	ARMA( $p, q$ )序列的自协方差函数	82
3.2.3	ARMA( $p, q$ )模型的可识别性	83
3.2.4	ARMA序列的谱密度和可逆性	87
3.2.5	例子	88
3.3	广义ARMA模型和ARIMA( $p, d, q$ )模型介绍	89
3.3.1	广义ARMA模型	89
3.3.2	求和ARIMA( $p, d, q$ )模型	90
3.3.3	单位根过程	90
3.3.4	分数差分ARFIMA( $p, d, q$ )模型	90
<b>第四章</b>	<b>均值和自协方差函数的估计</b>	<b>93</b>
4.1	均值的估计	94
4.1.1	相合性	94
4.1.2	中心极限定理	96
4.1.3	收敛速度	97
4.1.4	$\bar{X}_N$ 的模拟计算	97

4.2	自协方差函数的估计	99
4.2.1	自协方差估计公式及正定性	99
4.2.2	$\hat{\gamma}_k$ 的相合性	100
4.2.3	$\hat{\gamma}_k$ 的渐近分布	101
4.2.4	模拟计算	105
4.3	白噪声检验	106
4.3.1	白噪声的 $\chi^2$ 检验	106
4.3.2	样本自相关置信区间检验法	107
<b>第五章</b>	<b>时间序列的预报</b>	<b>109</b>
5.1	最佳线性预测的基本性质	110
5.1.1	最佳线性预测	110
5.1.2	Hilbert空间中的投影	117
5.1.3	最佳预测	121
5.2	非决定性平稳序列及其Wold表示	123
5.2.1	非决定性平稳序列	123
5.2.2	Wold表示定理	127
5.2.3	Kolmogorov公式	136
5.2.4	最佳预测和最佳线性预测相等的条件	137
5.3	时间序列的递推预测	139
5.3.1	时间序列的递推预测	139
5.3.2	正态时间序列的区间预测	143
5.3.3	平稳序列的递推预测	143
5.4	ARMA( $p, q$ )序列的递推预测	145
5.4.1	AR( $p$ )序列的预测	145
5.4.2	MA( $q$ )序列的预测	148
5.4.3	ARMA( $p, q$ )序列的预测	149
5.4.4	ARMA( $p, q$ )序列多步预测	152
5.4.5	ARIMA序列预测	152
<b>第六章</b>	<b>ARMA模型的参数估计</b>	<b>155</b>
6.1	AR( $p$ )模型的参数估计	156
6.1.1	Yule-Walker估计	156
6.1.2	最小二乘估计	157
6.1.3	最大似然估计	162
6.1.4	AR( $p$ )模型定阶	163
6.1.5	模型拟合检验	166
6.1.6	AR谱密度估计	167
6.2	MA( $q$ )模型的参数估计	168

6.2.1	MA( $q$ )模型的矩估计及其计算	168
6.2.2	MA( $q$ )模型的逆相关函数法	170
6.2.3	MA( $q$ )模型的新息估计方法	172
6.2.4	MA( $q$ )模型的定阶方法	174
6.2.5	MA( $q$ )模型的拟合检验	175
6.2.6	MA谱密度估计	175
6.3	ARMA( $p, q$ )模型的参数估计	176
6.3.1	ARMA( $p, q$ )模型的矩估计方法	176
6.3.2	ARMA( $p, q$ )模型的自回归逼近法	177
6.3.3	正态时间序列的似然函数	178
6.3.4	ARMA( $p, q$ )模型的最大似然估计	179
6.3.5	ARMA( $p, q$ )模型的检验	183
6.3.6	ARMA( $p, q$ )模型的定阶方法	184
6.3.7	ARMA( $p, q$ )模型的谱密度估计	185

## 课程介绍

### 课程介绍

- 研究沿时间记录的数据。
- 主要针对序列相关性。
- 主要使用二阶矩、矩阵运算。
- 也用少部分的基础泛函分析、复变函数。
- 有广泛应用但本课程主要是打基础。

### 时间序列分析的应用领域

- 经济与金融，比如宏观经济数据序列、证券价格、指数等。
- 工程，比如机械振动、无线电信号、音频信号。
- 科学研究，比如全球气温、人体生理节律、化学反应时间序列等。
- 社会，研究逐年、逐月记录的数据，如交通流量、幸福程度变化等。

### 参考书目

- 教材：何书元《应用时间序列分析》，北京大学出版社，2003
- 参考：P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods. Springer-Verlag, 1987. （有中译本：田铮，《时间序列的理论与方法》，高等教育出版社，第二版，2001）
- P.S.P. Cowpertwait and A.V. Metcalfe, Introductory Time Series with R. Springer, 2009
- J.D. Cryer and Kung-Sik Chan, Time Series Analysis with Applications in R, 2nd Ed, Springer, 2008.
- 肖枝洪，郭明月(2009) 时间序列分析与SAS应用，武汉大学出版社
- 李正辉, 李庭辉(2010) 时间序列分析实验，中国统计出版社





# 第一章 时间序列

§1.1 时间序列的分解

§1.1.1 时间序列概念

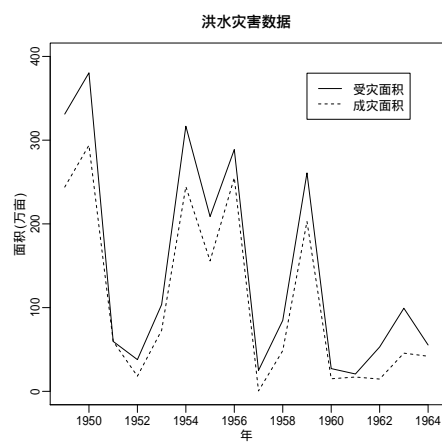
时间序列定义

- 时间序列定义: 按时间顺序排列的随机变量序列。
- 观测样本: 时间序列各随机变量的观测样本。一定是有限多个。
- 一次实现(一条轨道): 时间序列的一组实际观测。
- 时间序列分析的任务: 数据建模, 解释、控制或预报。
- 记号:  $\{X_t\}$ ,  $\{x_t\}$ ,  $X(t)$ ,  $x(t)$ .

例子: 北京地区洪涝灾害数据(单位: 万亩)

年代	序号	受灾面积	成灾面积
1949	1	331.12	243.96
1950	2	380.44	293.90
1951	3	59.63	59.63
1952	4	37.89	18.09
1953	5	103.66	72.92
1954	6	316.67	244.57
1955	7	208.72	155.77
1956	8	288.79	255.22
1957	9	25.00	0.50
1958	10	84.72	48.59
1959	11	260.89	202.96
1960	12	27.18	15.02
1961	13	20.74	17.09
1962	14	52.99	14.66
1963	15	99.25	45.61
1964	16	55.36	41.90

例子: 北京地区洪涝灾害数据(图)



## §1.1.2 时间序列的分解

## 时间序列的分解

•

$$X_t = T_t + S_t + R_t, \quad t = 1, 2, \dots$$

• 趋势项、季节项、随机项

• 分解

• 单周期季节项:  $S(t+s) = S(t), \forall t$ . 只需要  $S_1, \dots, S_s$  且可设

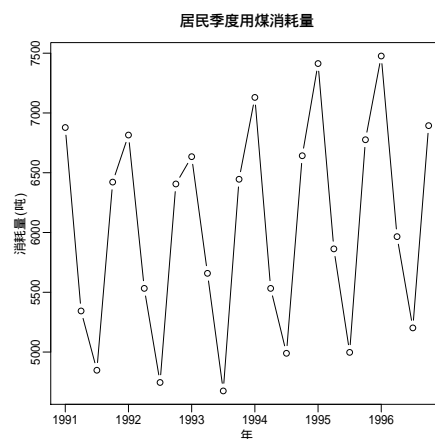
$$\sum_{j=1}^s S_j = 0$$

• 随机项: 可设  $ER_t = 0, \forall t$ .

## 例子: 居民用煤消耗季度值

年	1 季度	2 季度	3 季度	4 季度	年平均
1991	6878.4	5343.7	4847.9	6421.9	5873.0
1992	6815.4	5532.6	4745.6	6406.2	5875.0
1993	6634.4	5658.5	4674.8	6445.5	5853.3
1994	7130.2	5532.6	4989.6	6642.3	6073.7
1995	7413.5	5863.1	4997.4	6776.1	6262.6
1996	7476.5	5965.5	5202.1	6894.1	6384.5
季平均	7058.1	5649.3	4909.6	6597.7	

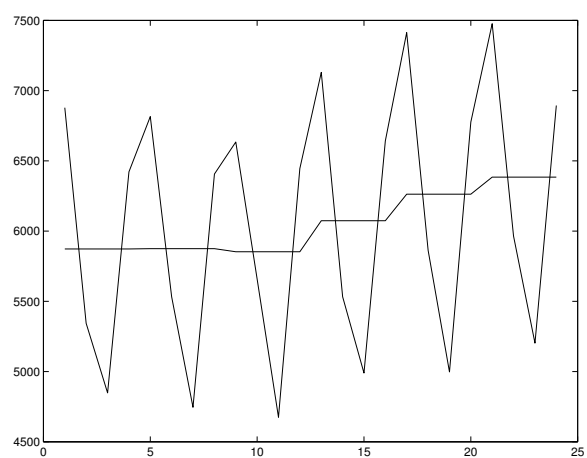
## 用煤消耗时间序列图



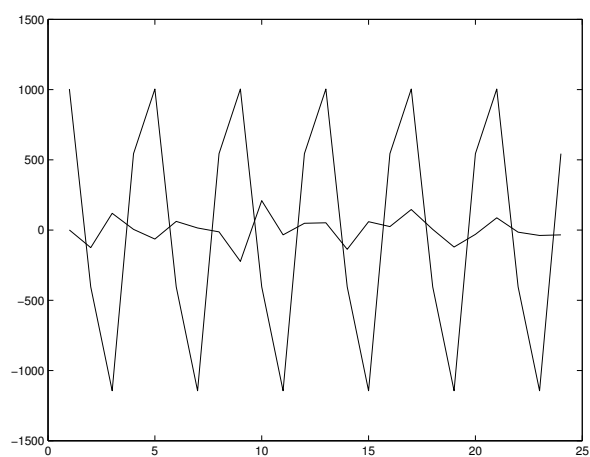
## 分解方法

- 估计趋势 $\{\hat{T}_t\}$ 后,  $X_t - \hat{T}_t$ 主要由季节项和随机项组成, 季节项可以用 $X_t - \hat{T}_t$  每个季节平均得到;
- 趋势估计可用:
  - 每年平均;
  - 线性回归拟合直线;
  - 二次曲线回归;
  - 滑动平均估计;
  - 得到 $\hat{T}_t$ 和 $\hat{S}_t$ 后可以再从 $X_t - \hat{S}_t$  估计新的 $\hat{T}_t$ 。
- 季节项估计还可以直接列入模型用哑变量表示(注意共线问题)。
- 用 $\hat{T}_t + \hat{S}_t$ 拟合或预报 $y_t$ 。
- 还可以建立趋势、季节项、随机项的状态空间模型。
- 演示: `atsa01.r::demo.coal()`。

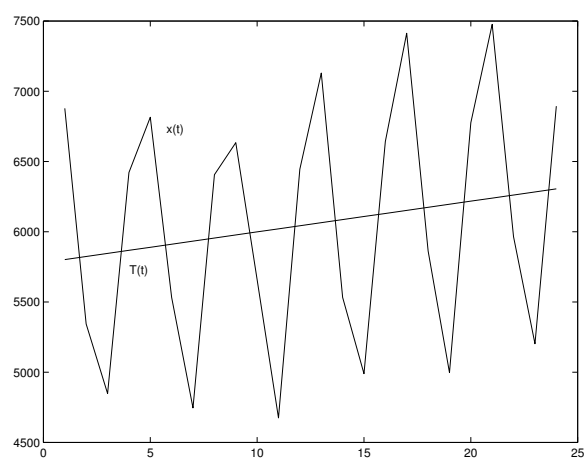
## 用煤消耗时间序列图(分段趋势)



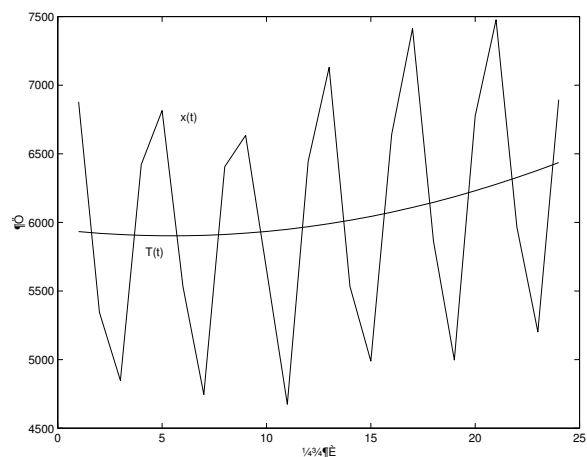
用煤消耗时间序列分解图：季节项和随机项



用煤消耗时间序列图(线性趋势)



用煤消耗时间序列图(二次曲线趋势)



### §1.1.3 时间序列和随机过程

#### 随机过程

- 随机过程的概念: 指标集  $\mathcal{T}$ 。  $\{X_t : t \in \mathcal{T}\}$  称为随机过程。
- 时间序列:  $\mathcal{T}$  为全体整数或正整数时, 随机过程称为随机序列; 把整数下标看成时间则称随机序列为时间序列。
- 连续时随机过程、连续时时间序列:  $\mathcal{T}$  为全体实数或全体非负实数时称随机过程为连续时随机过程。下标看成时间时称为连续时的时间序列。
- 离散采样: 连续时的时间序列记录下来就变成了离散时间。
- 平稳性: 去除趋势项和季节项后的随机部分经常具有平稳性。

## §1.2 平稳序列

### 平稳序列——引言

- 序列分解中趋势和季节部分一般可以用非随机函数描述（但也可以用随机模型）。
- 随机项通常是平稳的。表现：水平没有明显变化；方差没有明显变化；相关性结构不随时间变化。
- 独立序列不能预报；平稳序列可以用历史值预报。
- 记号：
  - $\mathbb{Z}$ —所有整数的集合；。
  - $\mathbb{N}_+$ —所有正整数的集合；。
  - $\mathbb{N}$ —表示 $\mathbb{Z}$ 或 $\mathbb{N}_+$ 。

### §1.2.1 平稳序列及其自协方差函数

#### 平稳序列定义

- **定义2.1** 如果时间序列 $\{X_t\} = \{X_t : t \in \mathbb{N}\}$  满足
  - (1) 对任何 $t \in \mathbb{N}$ ,  $EX_t^2 < \infty$ ;
  - (2) 对任何 $t \in \mathbb{N}$ ,  $EX_t = \mu$ ;
  - (3) 对任何 $t, s \in \mathbb{N}$ ,  $E[(X_t - \mu)(X_s - \mu)] = \gamma_{t-s}$ , 则称 $\{X_t\}$  是平稳时间序列, 简称为平稳序列. 称实数列 $\{\gamma_t\}$ 为 $\{X_t\}$ 的自协方差函数.
- 性质：
  - 期望、方差与 $t$ 无关。
  - 时间平移不影响两时刻的相关系数。
  - 又称平稳序列为二阶矩平稳序列, 还称为宽平稳序列。

#### 自协方差函数性质

- (1) 对称性:  $\gamma_k = \gamma_{-k}, \forall k \in \mathbb{Z}$ .
- (2) 非负定性:

$$\Gamma_n = (\gamma_{k-j})_{k,j=1}^n = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix}$$

非负定( $\forall n \in \mathbb{N}_+$ )。

- (3) 有界性:  $|\gamma_k| \leq \gamma_0, \forall k \in \mathbb{Z}$ .
- 任何满足上述三个性质的实数列都被称为**非负定序列**.
- 所以平稳序列的自协方差函数是非负定序列.
- 可以证明, 每个非负定序列都可以是一个平稳序列的自协方差函数(见[9]).

### 非负定性及随机变量的线性相关

- $\Gamma_n$  的特点: 设  $\alpha = (a_1, \dots, a_n)^T$ , 则

•

$$\alpha^T \Gamma_n \alpha = \text{Var} \left( \sum_{i=1}^n a_i X_i \right) \geq 0$$

- $\Gamma_n$  退化(不满秩)当且仅当存在  $\alpha \neq 0$  使得

$$\text{Var} \left( \sum_{i=1}^n a_i X_i \right) = 0$$

这时称随机变量  $X_1, \dots, X_n$  是**线性相关**的。即  $X_1, \dots, X_n$  的非零线性组合是退化随机变量。

- 如果  $X_1, \dots, X_n$  线性相关, 则  $m \geq n$  时  $X_1, \dots, X_m$  线性相关。

### Schwarz不等式

•

$$|E(XY)| \leq \sqrt{EX^2 EY^2}$$

•

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$$

- 推论:

$$E|X| \leq \sqrt{E|X|^2}$$

•

$$\gamma_t = \text{Cov}(Y_1, Y_{t+1}) \leq \gamma_0$$



**例2.1 平稳序列的线性变换**

- $\{X_t\}$  为平稳序列, 期望  $\mu$ , 自协方差函数  $\gamma(t)$ 。
- $Y_t = a + bX_t, t \in \mathbb{Z}$ 。
- $EY_t = a + b\mu$ 。
- $\text{Cov}(Y_s, Y_{s+t}) = b^2 \text{Cov}(X_t, X_{t+s}) = b^2 \gamma(t)$ 。
- 可见  $\{Y_t\}$  平稳。
- 若取

$$Y_t = \frac{X_t - \mu}{\sqrt{\gamma_0}}$$

则  $EY_t = 0, \text{Var}(Y_t) = 1$ , 称  $\{Y_t\}$  为  $\{X_t\}$  的标准化序列。

**自相关系数**

- **定义2.2** 设平稳序列  $\{X_t\}$  的标准化序列是  $\{Y_t\}$ .  $\{Y_t\}$  的自协方差函数

$$\rho_k = \gamma_k / \gamma_0, \quad k \in \mathbb{Z},$$

称为平稳序列  $\{X_t\}$  的自相关系数。

- 自相关系数  $\{\rho_t\}$  是满足  $\rho_0 = 1$  的自协方差函数, 从而也是非负定序列。

**例2.2 调和平稳序列**

- 设  $a, b$  是常数, 随机变量  $U$  在  $(-\pi, \pi)$  内均匀分布, 则

$$X_t = b \cos(at + U), \quad t \in \mathbb{Z}$$

是平稳序列。

•

$$\begin{aligned} EX_t &= \frac{1}{2\pi} \int_{-\pi}^{\pi} b \cos(at + u) du = 0, \\ E(X_t X_s) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} b^2 \cos(at + u) \cos(as + u) du \\ &= \frac{1}{2} b^2 \cos((t - s)a), \end{aligned}$$

- 这个平稳序列的观测样本和自协方差函数  $\gamma_k = 0.5b^2 \cos(ak)$  都是以  $a$  为角频率, 以  $2\pi/a$  为周期的函数。
- 这个例子告诉我们, 平稳序列也可以有很强的周期性。
- $\{X_t\}$  的一次实现是一个周期函数, 不表现出随机性。

## §1.2.2 白噪声

- **定义2.3** (白噪声) 设 $\{\varepsilon_t\}$  是一个平稳序列. 如果对任何 $s, t \in \mathbb{N}$ ,

$$\begin{aligned} E\varepsilon_t &= \mu, \\ \text{Cov}(\varepsilon_t, \varepsilon_s) &= \sigma^2 \delta_{t-s} = \begin{cases} \sigma^2, & t = s, \\ 0, & t \neq s, \end{cases} \end{aligned}$$

则称 $\{\varepsilon_t\}$ 是一个**白噪声**, 记做 $\text{WN}(\mu, \sigma^2)$ .

## 白噪声(续)

- 设 $\{\varepsilon_t\}$ 是白噪声,
- 当 $\{\varepsilon_t\}$  是独立序列时, 称 $\{\varepsilon_t\}$  是**独立白噪声**。
- 当 $\mu = 0$  时, 称 $\{\varepsilon_t\}$  为**零均值白噪声**。白噪声的另一种定义要求零均值, 本书中用到的白噪声一般都是零均值的。
- 当 $\mu = 0, \sigma^2 = 1$  时, 称 $\{\varepsilon_t\}$  为**标准白噪声**。
- 当 $\varepsilon_t$ 服从正态分布时, 称 $\{\varepsilon_t\}$ 是**正态白噪声**。正态白噪声总是独立白噪声。
- Kronecker函数 $\delta_t$ , 白噪声满足 $\text{Cov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \delta_{t-s}$ 。

## 例2.3 Poisson 过程

- 如果连续时的随机过程 $\{N(t) : t \in [0, \infty)\}$ 满足
- (1)  $N(0) = 0$ , 且对任何 $s \geq 0, t > 0$  和非负整数 $k$ ,

$$P(N(t+s) - N(s) = k) = \frac{(\lambda t)^k}{k!} \exp[-\lambda t], (\lambda > 0,)$$

- (2)  $\{N(t)\}$  有独立增量性: 对任何 $n > 1$  和 $0 \leq t_0 < t_1 < \cdots < t_n$ , 随机变量 $N(t_j) - N(t_{j-1})$ ,  $j = 1, 2, \cdots, n$  相互独立,
- 则称 $\{N(t)\}$  是一个强度 $\lambda$ 的**Poisson 过程**。
- $EN(t) = \lambda t, \text{Var}(N(t)) = \lambda t$ .

**Poisson白噪声**

- 定义

$$\varepsilon_n = N(n+1) - N(n) - \lambda, \quad n = 1, 2, \dots,$$

- $E\varepsilon_n = 0$ ,
- $\text{Var}(\varepsilon_n) = \lambda$ .
- $\{\varepsilon_n\}$  是独立白噪声, 称为**Poisson白噪声**.
- ave和std表示样本平均和样本标准差。

**布朗运动**

- 如果连续时的随机过程 $\{B(t) : t \in [0, \infty)\}$ 满足
- (1)  $B(0) = 0$ , 且对任何 $s \geq 0, t > 0$ ,  $B(t+s) - B(s)$ 服从正态分布 $N(0, t)$ ;
- (2)  $\{B(t)\}$  有独立增量性
- 则称 $\{B(t)\}$ 是一个标准布朗运动.
- 定义

$$\varepsilon_n = B(n+1) - B(n), \quad n = 1, 2, \dots,$$

则 $\{\varepsilon_n\}$  是一个标准正态白噪声.

**随机相位**

- $U_1, U_2, \dots$  iid  $U(0, 2\pi)$ .

- 

$$X_t = b \cos(at + U_t), \quad t \in \mathbb{Z},$$

(a, b为常数)

- 则 $X_t$  独立;
- $EX_t = 0, \text{Var}X_t = 0.5b^2$ .
- $X_t$ 是独立白噪声。
- 演示: atsa01.r::demo.harmonic().

## §1.2.3 正交平稳序列

## 正交和不相关

- 不相关:  $\text{Cov}(X, Y) = 0$ .
- 正交:  $E(XY) = 0$ .
- 对零均值随机变量正交与不相关等价。
- 对平稳列  $\{X_t\}$  和  $\{Y_t\}$ ,
  - 不相关:  $\text{Cov}(X_t, Y_s) = 0, \forall t, s$ .
  - 正交:  $E(X_t Y_s) = 0, \forall t, s$ .

## 定理2.2

- 平稳列  $\{X_t\}, \{Y_t\}$ , 自协方差函数  $\gamma_X(t), \gamma_Y(t)$ , 期望  $\mu_X, \mu_Y$ 。

$$Z_t = X_t + Y_t, t \in \mathbb{Z}$$

- (1) 如果  $\{X_t\}$  和  $\{Y_t\}$  正交, 则  $\{Z_t\}$  是平稳序列, 有自协方差函数

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k) - 2\mu_X\mu_Y, \quad k = 0, 1, 2, \dots$$

- (2) 如果  $\{X_t\}$  和  $\{Y_t\}$  不相关, 则  $\{Z_t\}$  是平稳序列, 有自协方差函数

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k), \quad k = 0, 1, 2, \dots$$

- (3) 当  $\mu_X = \mu_Y = 0$  且两序列正交时(2)的结论成立。

### §1.3 线性平稳序列和线性滤波

#### §1.3.1 有限运动平均

##### 有限运动平均

- 线性平稳序列是白噪声的线性组合得到的序列。
- 最简单的线性平稳序列是有限运动平均。
- 设  $\{\varepsilon_t\} = \{\varepsilon_t : t \in \mathbb{Z}\}$  是  $WN(0, \sigma^2)$ . 对于非负整数  $q$  和常数  $a_0, a_1, \dots, a_q$  ( $a_0 \neq 0, a_q \neq 0$ ), 我们称

$$X_t = \sum_{j=0}^q a_j \varepsilon_{t-j} = a_0 \varepsilon_t + a_1 \varepsilon_{t-1} + \dots + a_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}$$

是白噪声  $\{\varepsilon_t\}$  的(有限)运动平均, 简称为MA (Moving Average). 运动平均又称作滑动平均.

##### MA的平稳性

- $EX_t = 0$
- $EX_{t+k}X_t = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} a_j a_{j+k}, & 0 \leq k \leq q, \\ 0, & k > q \end{cases}$
- $\{X_t\}$  平稳。
- $\gamma_k = 0, \forall k > q$ , 称这样的序列为  $q$  相关的。
- 随机变量有限平均到随机变量无穷级数的推广需要概率论的极限理论。

##### 概率极限定理

- **定理3.1** (单调收敛定理) 如果非负随机变量序列单调不减:  $0 \leq \xi_1 \leq \xi_2 \leq \dots$ , 则当  $\xi_n \rightarrow \xi$ , a.s. 时, 有  $E\xi = \lim_{n \rightarrow \infty} E\xi_n$ .
- 广义随机变量.
- 对于任何时间序列  $\{Y_t\}$ , 利用单调收敛定理得到

$$\begin{aligned} E\left[\sum_{t=-\infty}^{\infty} |Y_t|\right] &= \lim_{n \rightarrow \infty} E\left[\sum_{t=-n}^n |Y_t|\right] \\ &= \lim_{n \rightarrow \infty} \sum_{t=-n}^n E|Y_t| = \sum_{t=-\infty}^{\infty} E|Y_t|. \end{aligned}$$

- **定理3.2** (控制收敛定理) 如果随机变量序列 $\{\xi_n\}$ 满足 $|\xi_n| \leq \xi_0$  a.s. 和 $E|\xi_0| < \infty$ , 则当 $\xi_n \rightarrow \xi$ , a.s.时,  $E|\xi| < \infty$  并且 $E\xi_n \rightarrow E\xi$ .

### §1.3.2 线性平稳序列

#### 线性平稳序列

- **绝对可和**: 如果实数列 $\{a_j\}$  满足

$$\sum_{j=-\infty}^{\infty} |a_j| < \infty,$$

则称 $\{a_j\}$ 是绝对可和的. 记 $\{a_j\} \in l_1$ .

- 注意:  $\{a_j\} \in l_1$  则  $\{a_j\} \in l_2$  (即 $\sum_j a_j^2 < \infty$ ).
- 对于绝对可和的实数列 $\{a_j\}$ , 定义零均值白噪声 $\{\varepsilon_t\}$ 的无穷滑动和如下

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}.$$

- 则 $\{X_t\}$ 是平稳序列。
- $EX_t = 0$ .
- $\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k}, k \in \mathbb{Z}$ .

#### 线性序列的a.s.收敛性

- $\{X_t\}$ 有定义?
- 设

$$\eta = \sum_{j=-\infty}^{\infty} |a_j| |\varepsilon_{t-j}|$$

则 $\eta$ 有定义 (允许取 $+\infty$ );

- 又由单调收敛定理得

$$E\eta = \sum_{j=-\infty}^{\infty} |a_j| E|\varepsilon_{t-j}| \leq \sigma \sum_{j=-\infty}^{\infty} |a_j| < \infty$$

- 因此 $\eta < \infty$ , a.s.(否则 $E\eta = +\infty$ )。
- 由 $\eta = \sum_{j=-\infty}^{\infty} |a_j| |\varepsilon_{t-j}| < \infty$ (a.s.), 在概率空间中除去一个零测集的意义下级数

$$\sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

处处绝对收敛(称为a.s.绝对收敛);

- 所以, 此级数a.s.收敛到一个随机变量 $\xi$ ,

$$\xi = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j} \quad (*)$$

在a.s.收敛意义下存在。

#### 线性序列的 $L^1$ 收敛性

(\*) 在 $L^1$ 意义下也是收敛的, 因为

$$\mathbb{E} \left| \sum_{|j|>N} a_j \varepsilon_{t-j} \right| \leq \sum_{|j|>N} |a_j| \mathbb{E} |\varepsilon_{t-j}| \leq \sigma \sum_{|j|>N} |a_j| \rightarrow 0$$

#### 线性序列的平稳性

由 $\sum_j |a_j| \mathbb{E} |\varepsilon_{t-j}| < \infty$ 及控制收敛定理得

$$\mathbb{E} \sum_j a_j \varepsilon_{t-j} = \sum_j a_j \mathbb{E} \varepsilon_{t-j} = 0$$

令

$$V = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |a_j| |a_l| |\varepsilon_{t-j}| |\varepsilon_{t+k-l}|$$

则

$$\begin{aligned} \mathbb{E} V &= \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |a_j| |a_l| \mathbb{E} |\varepsilon_{t-j}| |\varepsilon_{t+k-l}| \\ &\leq \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |a_j| |a_l| \sigma^2 = \sigma^2 \left( \sum_{j=-\infty}^{\infty} |a_j| \right)^2 < \infty \end{aligned}$$

由控制收敛定理得

$$\begin{aligned} \mathbb{E} X_t X_{t+k} &= \mathbb{E} \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j} \sum_{l=-\infty}^{\infty} a_l \varepsilon_{t+k-l} \\ &= \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_j a_l \mathbb{E} (\varepsilon_{t-j} \varepsilon_{t+k-l}) \\ &= \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k} \end{aligned}$$

即 $\{X_t\}$ 是平稳序列,  $\mathbb{E} X_t = 0$ ,  $\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k}$ .

### 均方意义的的线性序列

- 设  $\{a_j\} \in l_2$  则

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j} \quad (L^2) \quad (**)$$

也是平稳序列。期望为零，自协方差函数同上。

- $X_t$  定义的无穷级数是  $L^2$  收敛的。证明需要应用 Hilbert 空间性质。见 1.6 节例 6.3。
- 注意  $\{a_j\} \in L_1 \implies \{a_j\} \in L_2$ 。

### $l_1$ 系数的线性平稳列的自协方差函数的收敛性

当  $\{a_j\} \in l_1$  时

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$$

. 事实上,

$$\begin{aligned} \sum_{k=-\infty}^{\infty} |\gamma_k| &\leq \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |a_j a_{j+k}| \\ &= \sigma^2 \sum_{j=-\infty}^{\infty} |a_j| \sum_{k=-\infty}^{\infty} |a_{j+k}| = \sigma^2 \left( \sum_{j=-\infty}^{\infty} |a_j| \right)^2 < \infty \end{aligned}$$

### $l_2$ 系数的线性平稳列的自协方差函数的收敛性

当  $\{a_j\} \in l_2$  时, 有

**定理 3.3** 自协方差函数  $\lim_{k \rightarrow \infty} \gamma_k = 0$ .

证明略。

### 线性序列的应用

线性序列描述了自协方差函数衰减到零的时间序列。只要样本自协方差函数衰减到零就可以用线性序列来描述。

### 单边线性序列

- 

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}$$

- 称为单边运动平均(MA)，或单边无穷滑动和。



- 这样的  $X_t$  有因果性:  $X_t$  只受  $s \leq t$  的  $\varepsilon_s$  影响而不受  $t$  时刻以后的  $\varepsilon_s$  影响。

- 

$$\gamma_k = \begin{cases} \sum_{j=0}^{\infty} a_j a_{j+k}, & k \geq 0 \\ \gamma_{-k} & k < 0 \end{cases}$$

### §1.3.3 时间序列的线性滤波

#### 时间序列的线性滤波

- 对序列  $\{X_t\}$  进行滑动求和:

$$Y_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j}, \quad t \in \mathbb{Z}$$

称为对  $\{X_t\}$  进行 **线性滤波**。

- 其中绝对可和的  $\{h_j\}$  称为一个 **保时线性滤波器**。
- 如果输入信号  $\{X_t\}$  是平稳列则输出  $\{Y_t\}$  也是平稳列。

- 

$$\mu_Y = EY_t = \sum_{j=-\infty}^{\infty} h_j EX_{t-j} = \mu_X \sum_{j=-\infty}^{\infty} h_j$$

- 

$$\begin{aligned} \gamma_Y(n) &= \text{Cov}(Y_{n+1}, Y_1) \\ &= \sum_{j,k=-\infty}^{\infty} h_j h_k E[(X_{n+1-j} - \mu)(X_{1-k} - \mu)] \\ &= \sum_{j,k=-\infty}^{\infty} h_j h_k \gamma_{n+k-j} \end{aligned}$$

#### 矩形窗滤波器

- 取

$$h_j = \begin{cases} \frac{1}{2M+1}, & |j| \leq M \\ 0, & |j| > M. \end{cases}$$

- 则

$$Y_t = \frac{1}{2M+1} \sum_{j=-M}^M x_{t-j}$$

- 是  $X_t$  的滑动平均。可以平滑  $\{X_t\}$ , 抑制高频信号。
- 高频信号表现是粗糙和复杂的曲线, 低频信号表现为缓慢和光滑的变化。

## 例3.1 余弦波信号的滤波

•

$$X_t = S_t + \varepsilon_t = b \cos(\omega t + U) + \varepsilon_t, \quad t \in \mathbb{Z}$$

- $U \sim U(0, 2\pi)$ ,  $\{\varepsilon_t\}$  零均值平稳,  $U$  与  $\{\varepsilon_t\}$  独立。
- 信号  $\{S_t\}$  方差  $b^2/2$ , 噪声  $\{\varepsilon_t\}$  方差  $\sigma^2$ , 信噪比  $b^2/(2\sigma^2)$ 。
- 作矩形窗滤波:

$$\begin{aligned} Y_t &= \frac{1}{2M+1} \sum_{j=-M}^M X_{t-j} \\ &= \frac{b \sin[\omega(M+0.5)]}{(2M+1) \sin(\omega/2)} \cos(\omega t + U) + \eta_t \\ \eta_t &= \frac{1}{2M+1} \sum_{j=-M}^M \varepsilon_{t-j} \end{aligned}$$

- 除了  $\eta_t$  项之外, 结果与原始信号相比只有幅度有了成比例变化。
- (技巧: 用复数的极坐标表示来推导有关三角函数求和的公式。)
- 演示: `atsa01.r::demo.mafilt()`。
- $\text{Var}(\eta_t) = \frac{\sigma^2}{2M+1}$ 。
- 新的信噪比为

$$\frac{b^2}{2\sigma^2} \frac{\sin^2[\omega(M+0.5)]}{(2M+1) \sin^2(\omega/2)}.$$

- 特别当  $\omega(M+0.5) = \pi/2$  时信噪比为

$$\frac{b^2 \omega}{2\pi \sigma^2 \sin^2(\omega/2)} > \frac{2b^2}{\pi \omega \sigma^2} = \frac{b^2}{2\sigma^2} \cdot \frac{4}{\pi \omega}$$

- 信噪比至少增大到  $4/(\pi \omega)$  倍,  $\omega$  越小信噪比提高越多。
- $\omega$  越小,  $M$  应越大。

## §1.4 正态时间序列和随机变量的收敛性

## §1.4.1 随机向量的数学期望和方差

## 随机向量的数学期望和方差

- 矩阵随机变量  $\mathbf{X} = (X_{i,j})_{m \times n}$ .
- $E\mathbf{X} = (EX_{i,j})_{m \times n} = (\mu_{ij})_{m \times n}$ .
- 随机向量  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ . 则

$$\Sigma_X = \text{Var}(\mathbf{X}) = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]$$

- $\Sigma_X$  对称半正定(非负定)。
- $\Sigma_X = E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X})E(\mathbf{X})^T$ .

## 随机向量线性变换

- $\mathbf{Y} = \mathbf{a} + B\mathbf{X}$  则
- $E\mathbf{Y} = \mathbf{a} + BE\mathbf{X}$ ,
- $\text{Var}(\mathbf{Y}) = B\Sigma_X B^T$ .

## 多维正态分布

- 称随机向量  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^T$  服从  $m$  维(或多维)正态分布, 如果存在  $m$  维常数列向量  $\mu$ ,  $m \times n$  常数矩阵  $B$  和 iid 的标准正态随机变量  $X_1, X_2, \dots, X_n$  使得  $\mathbf{Y} = \mu + B\mathbf{X}$ .
- 这时  $E\mathbf{Y} = \mu$ ,  $\Sigma = \text{Var}(\mathbf{Y}) = BB^T$ .
- $\mathbf{Y}$  的特征函数为

$$\phi_Y(\mathbf{t}) = \exp \left[ i\mathbf{t}^T \mu - \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t} \right]. \quad (4.6)$$

这是多维正态分布的等价定义。

- 多元正态分布记为  $\mathbf{Y} \sim N(\mu, \Sigma)$ .  $\mathbf{Y}$  的分布完全由  $\mu, \Sigma$  决定。
- 当  $\Sigma > 0$  时,  $\mathbf{Y}$  有密度

$$p(\mathbf{y}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu) \right\}$$

- 若  $|\Sigma| = 0$ , 则  $\mathbf{Y}$  的分量由两部分  $Y_1$  和  $Y_2$  组成,  $\text{Var}(Y_1) > 0$ ,  $Y_2$  为  $Y_1$  的线性组合。(递推证明)

### 多维正态分布的充要条件

- 定理4.1  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)^T \sim N(\boldsymbol{\mu}, \Sigma)$  的充分必要条件是
- 对任何  $\mathbf{a} = (a_1, a_2, \dots, a_n)^T \in \mathbb{R}^n$

$$Y = \mathbf{a}^T \boldsymbol{\xi} \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \Sigma \mathbf{a}). \quad (4.7)$$

- 证明(板书): 必要性: 由(4.6)得Y的特征函数为

$$\begin{aligned} \phi(t) &= E \exp(itY) = E \exp[i(t\mathbf{a}^T)\boldsymbol{\xi}] \\ &= \exp \left[ it\mathbf{a}^T \boldsymbol{\mu} - \frac{1}{2} t^2 \mathbf{a}^T \Sigma \mathbf{a} \right] \end{aligned} \quad (4.8)$$

由一元正态分布的性质知  $Y \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \Sigma \mathbf{a})$ 。

- 充分性: 若(4.7)成立, 则(4.8)成立, 取  $t = 1$ , 对任意  $\mathbf{a}$  有

$$E \exp(i\mathbf{a}^T \boldsymbol{\xi}) = \exp \left( i\mathbf{a}^T \boldsymbol{\mu} - \frac{1}{2} \mathbf{a}^T \Sigma \mathbf{a} \right).$$

即  $\boldsymbol{\xi}$  的特征函数为(4.6)。

### §1.4.2 正态平稳序列

#### 正态平稳序列

- 定义4.2 对于时间序列  $\{X_t\}$ , 如果对任何  $n \geq 1$  和  $t_1, t_2, \dots, t_n \in \mathbb{Z}$ , 有
- $(X(t_1), X(t_2), \dots, X(t_n))$  服从多元正态分布,
- 则称  $\{X_t\}$  是正态时间序列.
- 特别当  $\{X_t\}$  还是平稳序列时, 又称为正态平稳序列.

#### 正态平稳列

- $\{X_t : t \in \mathbb{N}_+\}$  是正态时间序列  $\iff$  对任何正整数  $m$ ,  $(X_1, X_2, \dots, X_m)$  服从  $m$  维正态分布;
- $\{X_t : t \in \mathbb{Z}\}$  是正态时间序列  $\iff$  对任何正整数  $m$ ,  $(X_{-m}, X_{-m+1}, \dots, X_m)$  服从  $2m+1$  维正态分布.
- 正态分布对线性运算的封闭性为其理论研究提供了便利.
- 另外, 正态分布和线性模型之间有一种内在的联系.

### 概率极限

- $\xi_n \sim F_n(x), \xi \sim F(x)$ 。
- 如果在 $F$ 的每个连续点 $x$  有 $F_n(x) \rightarrow F(x)$ , 则称 $\xi_n$ 依分布收敛到 $\xi$ , 记做 $\xi_n \xrightarrow{d} \xi$ ;
- 如果对任取 $\epsilon > 0$  有 $P(|\xi_n - \xi| \geq \epsilon) \rightarrow 0$ , 则称 $\xi_n$ 依概率收敛到 $\xi$ , 或称 $\xi_n$ 相合于 $\xi$ , 或 $\xi_n$ 弱收敛到 $\xi$ , 记做 $\xi_n \xrightarrow{P} \xi$ ;
- 如果 $E|\xi_n - \xi| \rightarrow 0$ , 则称 $\xi_n$   $L^1$ 收敛到 $\xi$  (很少用);
- 如果 $E|\xi_n - \xi|^2 \rightarrow 0$ , 则称 $\xi_n$   $L^2$ (均方)收敛到 $\xi$ , 记做 $\xi_n \rightarrow \xi (L^2)$ ;
- **定理4.2**  $L^2$ 收敛 $\Rightarrow L^1$ 收敛 $\Rightarrow$  依概率收敛 $\Rightarrow$  依分布收敛。
- (证明自学)

### 正态序列收敛定理

- **定理4.3** 如果正态序列 $\xi_n \sim N(\mu_n, \sigma_n^2), n \in \mathbb{N}$  依分布收敛到随机变量 $\xi$ , 则极限

$$\lim \mu_n = \mu, \lim \sigma_n^2 = \sigma^2$$

存在, 且 $\xi \sim N(\mu, \sigma^2)$ .

### 正态线性序列

- **定理4.4**  $\{\varepsilon_t\}$ 是正态 $WN(0, \sigma^2)$ 序列, 实数列 $\{a_j\}$ 绝对可和, 则线性序列

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

是零均值正态平稳列。

- 证明: 板书。
- 当 $\{a_j\} \in l_2$ 时结论仍成立。

## §1.5 严平稳序列及其遍历性

### 严平稳序列

- 随机向量同分布：联合分布函数相同。
- 时间序列 $\{X_t\}$ 与 $\{Y_t\}$ 同分布： $\forall n \in \mathbb{N}_+$ 和 $t_1, \dots, t_n \in \mathbb{Z}$ ,  $(X(t_1), \dots, X(t_n))^T$ 与 $(Y(t_1), \dots, Y(t_n))^T$ 分布。
- 严平稳： $\{X_t\}$ 对 $\forall n \in \mathbb{N}_+$ 和 $k \in \mathbb{Z}$ 都有

$$(X_1, X_2, \dots, X_n)^T \text{ 和 } (X_{1+k}, X_{2+k}, \dots, X_{n+k})^T \text{ 同分布.}$$

- 特点：分布平移不变。对任多元函数 $\phi(x_1, x_2, \dots, x_m)$ 有 $Y_t = \phi(X_{t+1}, \dots, X_{t+m})$ ,  $t \in \mathbb{Z}$ 仍是严平稳列。

### 严平稳与宽平稳

- 二阶矩有限的严平稳为宽平稳。
- 宽平稳一般不是严平稳。
- 正态平稳列既是宽平稳也是严平稳。
- 平稳序列—宽平稳序列—弱平稳序列。
- 严平稳序列—强平稳序列。

### 遍历性

- 时间序列一般只有一条轨道。
- 要用时间序列 $\{X_t\}$ 的一次实现 $x_1, x_2, \dots$ 推断 $\{X_t(\omega), t \in \mathbb{N}, \omega \in \Omega\}$ 的统计性质。
- 遍历性可以保证从一条轨道可以推断整体的统计性质。
- 如果严平稳序列是遍历的, 从它的一次实现 $x_1, x_2, \dots$  就可以推断出这个严平稳序列的所有有限维分布:

$$F(x_1, x_2, \dots, x_m) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_m \leq x_m), \quad m \in \mathbb{N}.$$

有遍历性的严平稳序列被称作**严平稳遍历序列**。

**遍历定理**

**定理5.1** 如果 $\{X_t\}$ 是严平稳遍历序列, 则有如下的结果:

1. 强大数律: 如果 $E|X_1| < \infty$  则

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_t = EX_1, a.s..$$

2. 对任何多元函数 $\phi(x_1, x_2, \dots, x_m)$ ,

$$Y_t = \phi(X_{t+1}, X_{t+2}, \dots, X_{t+m})$$

是严平稳遍历序列.

**线性平稳列的遍历定理**

**定理5.2** 如果 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$ , 实数列 $\{a_j\}$ 平方可和, 则线性平稳序列

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z},$$

是严平稳遍历的.

**例5.1**

对严平稳序列 $\{X_t\}$ , 定义严平稳序列

$$Y_t = I[X(t+t_1) \leq y_1, X(t+t_2) \leq y_2, \dots, X(t+t_m) \leq y_m], \quad t \in \mathbb{Z}.$$

这里 $I[A]$ 是事件 $A$ 的示性函数. 如果 $\{X_t\}$ 是遍历的, 由定理5.1 的(2)知道 $\{Y_t\}$ 也是遍历的, 并且有界. 利用定理5.1 的(1)(强大数律)得到

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n Y_t &= EY_0 \\ &= P(X(t_1) \leq y_1, X(t_2) \leq y_2, \dots, X(t_m) \leq y_m), \quad a.s.. \end{aligned}$$

这个例子说明, 在几乎必然的意义下,  $\{X_t\}$ 的每一次观测都可以决定 $\{X_t\}$ 的有限维分布.

## §1.6 Hilbert空间中的平稳序列

### §1.6.1 Hilbert空间

#### 无穷维函数空间的必要性

- 本节中设 $\{X_t\}$ 是平稳序列。
- $X_t$ 可以用线性组合 $\sum_{j=1}^k a_j x_{t-j}$ 预测。
- 应该推广到用无穷的线性组合 $\sum_{j=1}^{\infty} a_j x_{t-j}$ 预测。
- 这就需要研究这样的无穷线性组合的性质。
- 所有 $\{X_t\}$ 及其无穷线性组合构成一个无穷维函数空间，其中的元素是有二阶矩的随机变量。
- 下面从有限线性组合讲起。

#### 平稳列导出的线性空间

- 设 $\{X_t\}$ 是平稳序列。令

$$L^2(X) = \left\{ \sum_{j=1}^k a_j X(t_j) \mid a_j \in \mathbb{R}, t_j \in \mathbb{Z}, 1 \leq j \leq k, k \in \mathbb{N}_+ \right\}$$

- $\forall X, Y, Z \in L^2(X), a, b \in \mathbb{R}$ 有
  1.  $X + Y = Y + X \in L^2(X), (X + Y) + Z = X + (Y + Z);$
  2.  $0 \in L^2(X), X + 0 = X, X + (-X) = 0 \in L^2(X);$
  3.  $a(X + Y) = aX + aY \in L^2(X), (a + b)X = aX + bX, a(bX) = (ab)X.$
- 即 $L^2(X)$ 是一个线性空间。

#### $L^2(X)$ 的内积

- 定义 $\langle X, Y \rangle = E(XY),$
- 则

$$\begin{aligned} \langle X, Y \rangle &= \langle Y, X \rangle \\ \langle aX + bY, Z \rangle &= a\langle X, Z \rangle + b\langle Y, Z \rangle \end{aligned}$$

- $\langle X, X \rangle \geq 0$ , 并且 $\langle X, X \rangle = 0$ 当且仅当 $X = 0$ , a.s., 所以 $L^2(X)$  又是内积空间。



- 内积有Schwarz不等式

$$| \langle X, Y \rangle | \leq [\langle X, X \rangle \langle Y, Y \rangle]^{1/2}$$

- 但我们还需要对无穷线性组合的封闭性。

### $L^2(X)$ 上的距离

- 定义模

$$\|X\| = (\langle X, X \rangle)^{1/2}$$

- 距离

$$\|X - Y\| = (\langle X - Y, X - Y \rangle)^{1/2}$$

- 则  $\|X - Y\| = \|Y - X\| \geq 0$ , 且  $\|X - Y\| = 0$  当且仅当  $X = Y$ , a.s.
- Schwarz不等式可以写成

$$| \langle X, Y \rangle | \leq \|X\| \cdot \|Y\|$$

- 三角不等式(用Schwarz不等式可证明):

$$\|X - Y\| \leq \|X - Z\| + \|Z - Y\|$$

- $L^2 = \{Y : EY^2 < \infty\}$  也是内积空间和距离空间,  $L^2(X)$  是  $L^2$  的子空间。

### $L^2$ 中的极限

- 对有限维空间, 定义了内积已经足够。
- 对无穷维空间, 要考虑极限问题。
- **定义6.1** 对  $\xi_n \in L^2$ ,  $\xi_0 \in L^2$ :

- (1) 如果  $\lim_{n \rightarrow \infty} \|\xi_n - \xi_0\| = 0$ , 则称  $\xi_n$  在  $L^2$  中(或均方)收敛到  $\xi_0$ , 记做  $\xi_n \xrightarrow{\text{m.s.}} \xi_0$  或  $\xi_n \xrightarrow{L^2} \xi_0$ .
- (2) 如果当  $n, m \rightarrow \infty$  时,  $\|\xi_n - \xi_m\| \rightarrow 0$ , 则称  $\{\xi_n\}$  是  $L^2$  中的基本列或Cauchy列.

### Hilbert空间

- **定理6.1** 如果 $\{\xi_n\}$ 是 $L^2$ 中的基本列, 则(在a.s.的意义下)有惟一的 $\xi \in L^2$ 使得 $\xi_n \xrightarrow{m.s.} \xi$ .
- 证明略。
- **完备的内积空间**: 每个基本列都有极限在空间内的内积空间。又称Hilbert空间。
- $L^2$ 是Hilbert空间。
- 用 $\bar{L}^2(X)$ 表示 $L^2$ 中包含 $L^2(X)$ 的最小闭子空间, 则 $\bar{L}^2(X)$ 是Hilbert空间, 称为由平稳序列 $\{X_t\}$ 生成的Hilbert空间。

### §1.6.2 内积的连续性

#### 内积的连续性

- **定理6.2** (内积的连续性) 在内积空间中, 如果 $\|\xi_n - \xi\| \rightarrow 0, \|\eta_n - \eta\| \rightarrow 0$  则有
  - (1)  $\|\xi_n\| \rightarrow \|\xi\|$ ,
  - (2)  $\langle \xi_n, \eta_n \rangle \rightarrow \langle \xi, \eta \rangle$ .

#### 定理6.2证明

(1)

$$|\|\xi_n\| - \|\xi\|| \leq \|\xi_n - \xi\| \rightarrow 0$$

(2)

$$\begin{aligned} |\langle \xi_n, \eta_n \rangle - \langle \xi, \eta \rangle| &= |\langle \xi_n, \eta_n - \eta \rangle + \langle \xi_n - \xi, \eta \rangle| \\ &\leq |\langle \xi_n, \eta_n - \eta \rangle| + |\langle \xi_n - \xi, \eta \rangle| \\ &\leq \|\xi_n\| \|\eta_n - \eta\| + \|\xi_n - \xi\| \|\eta\| \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

### 例6.1

- $\mathbb{R}^n$ 是线性空间, 定义内积 $\langle a, b \rangle = a^T b$  则为内积空间。
- $\mathbb{R}^n$ 是完备的内积空间。
- $|a| = \sqrt{a^T a}$  为欧氏模。

**$n$ 维Hilbert空间**

- 设 $\{X_t\}$ 是零均值平稳列,  $\mathbf{X} = (X_1, \dots, X_n)$ 。
- 令 $L_n = \text{sp}\{X_1, \dots, X_n\} = \{\mathbf{a}^T \mathbf{X} : \mathbf{a} \in \mathbb{R}^n\}$
- 则 $L_n$ 是Hilbert空间, 称为由 $\mathbf{X}$ 生成的Hilbert空间。
- $L_n$ 是线性空间和内积空间易验证, 下面证明其完备性。

 **$L_n$ 的完备性证明**

- 先设 $\{X_t\}$ 是标准白噪声 $\text{WN}(0, 1)$ 。
- 对任何线性组合 $\xi_n = \mathbf{a}_n^T \mathbf{X}$ , 只要

$$\|\xi_n - \xi_m\|^2 = \|\mathbf{a}_n^T \mathbf{X} - \mathbf{a}_m^T \mathbf{X}\|^2 = (\mathbf{a}_n - \mathbf{a}_m)^T (\mathbf{a}_n - \mathbf{a}_m) \rightarrow 0,$$

- 由例6.1知道有 $\mathbf{a} \in \mathbb{R}^n$  使得

$$|\mathbf{a}_n - \mathbf{a}| \rightarrow 0$$

- 当 $n \rightarrow \infty$ , 取 $\xi = \mathbf{a}^T \mathbf{X}$  时,

$$\|\xi_n - \xi\|^2 = (\mathbf{a}_n - \mathbf{a})^T (\mathbf{a}_n - \mathbf{a}) \rightarrow 0.$$

- 可见 $\{X_t\}$ 是标准白噪声 $\text{WN}(0, 1)$ 时 $L_n$  是完备的。
- 对一般的零均值平稳序列, 可以设协方差阵 $\Gamma = \text{E}(\mathbf{X}\mathbf{X}^T)$ 的秩是 $m$ ,  $m \leq n$ .
- 对 $\Gamma$ 做特征值分解得

$$\Gamma = P^T \Lambda P$$

$$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_m, 0, \dots, 0\}$$

- 令

$$A = \text{diag}\{\lambda_1^{-1/2}, \dots, \lambda_m^{-1/2}, 1, \dots, 1\} \triangleq \Lambda^{-1/2}$$

$$\mathbf{Y} = A P \mathbf{X}, \quad \mathbf{X} = P^T A^{-1} \mathbf{Y}$$

- 则

$$\begin{aligned} \text{Var}(\mathbf{Y}) &= A P \text{Var}(\mathbf{X}) P^T A = A P P^T \Lambda P P^T A \\ &= \text{diag}\{1, \dots, 1, 0, \dots, 0\} \end{aligned}$$

- 因此 $Y_1, \dots, Y_m$ 是某零均值白噪声列的某段。 $\mathbf{X}$ 的线性组合即 $Y_1, \dots, Y_m$ 的线性组合。
- 因此 $L_n$ 是完备的。

$L^2$ 意义下的线性序列

- 考虑 $L^2$ 中的零均值白噪声列 $\{\varepsilon_t\}$ , 设 $\text{Var}(\varepsilon_t) = \sigma^2$ .
- 设 $\{a_j\} \in l_2$ .
- 令

$$\xi_n(t) = \sum_{j=-n}^n a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}$$

- 则 $\xi_n(t) \in L^2$ . 对 $m < n$ , 当 $m \rightarrow \infty$ 时

$$\begin{aligned} \|\xi_n - \xi_m\|^2 &= \left\| \sum_{m < |j| \leq n} a_j \varepsilon_{t-j} \right\|^2 \\ &= \sigma^2 \sum_{m < |j| \leq n} a_j^2 \leq \sigma^2 \sum_{|j| > m} a_j^2 \rightarrow 0 \end{aligned}$$

- 由 $L^2$ 完备性知存在 $X_t \in L^2$ 使得 $\xi_n(t) \xrightarrow{\text{m.s.}} X_t$ .
- 记 $\xi_n(t)$ 在 $L^2$ 中的极限 $X_t$ 为

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}$$

- 来证明 $\{X_t\}$ 平稳。由 $L^2$ 中内积连续性得

$$\mathbb{E}X_t = \lim_{n \rightarrow \infty} \langle \xi_n(t), 1 \rangle = \lim_n \mathbb{E}\xi_n(t) = 0$$

- 以及

$$\begin{aligned} \mathbb{E}X_t X_{t+k} &= \lim_n \langle \xi_n(t), \xi_n(t+k) \rangle \\ &= \lim_n \left\langle \sum_{j=-n}^n a_j \varepsilon_{t-j}, \sum_{l=-n}^n a_l \varepsilon_{t+k-l} \right\rangle \\ &= \lim_n \sum_{j=-n}^n \sum_{l=-n}^n a_j a_l \delta_{k-l+j} \sigma^2 = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+k} \end{aligned}$$

## §1.6.3 复值随机变量

## 复值随机变量

- $Z = X + iY$ 称为复值随机变量。 $\mathbb{E}Z = \mathbb{E}X + i\mathbb{E}Y$ .

•

$$\text{Cov}(Z_1, Z_2) = \mathbb{E}((Z_1 - \mathbb{E}Z_1)(Z_2 - \mathbb{E}Z_2)^*)$$

•

$$\text{Cov}(\mathbf{Z}_1, \mathbf{Z}_2) = \mathbb{E}((\mathbf{Z}_1 - \mathbb{E}\mathbf{Z}_1)(\mathbf{Z}_2 - \mathbb{E}\mathbf{Z}_2)^*)$$

$\mathbf{Z}^*$ 表示 $\mathbf{Z}$ 的共轭转置。

**复值随机变量Hilbert空间**

- $E|Z|^2 = EX^2 + EY^2 < \infty$ 时称 $Z$ 是二阶矩有限的复随机变量。
- 所有二阶矩有限复随机变量的集合 $H$ 在定义内积 $\langle X, Y \rangle = E(X\bar{Y})$ 后构成Hilbert空间。

**复值时间序列**

- 复值随机变量的序列 $\{Z_n\}$ 称为复时间序列。
- 若 $EZ_n = \mu$ ,  $\text{Cov}(Z_n, Z_m) = \gamma_{n-m}$ ,  $\forall n, m \in \mathbb{Z}$ , 则称 $\{Z_n\}$ 是复值平稳序列。
- $\gamma_{-k} = \gamma_k^*$ .
- 若复值零均值平稳列 $\{\varepsilon\}$ 满足 $\text{Cov}(\varepsilon_n, \varepsilon_m) = \sigma^2 \delta_{n-m}$ ,  $\forall n \neq m$ , 称 $\{\varepsilon_t\}$ 为复值零均值白噪声。

## §1.7 平稳序列的谱函数

### §1.7.1 平稳序列的谱函数

#### 时域和频域

- 遍历的时间序列可以从一次实现的时间分布进行统计分析，称为时域分析。
- 平稳时间序列的二阶性质也可以从其频率分解来研究，称为频域分析。
- 频谱的典型代表是声音。
- 演示1: 在Audacity中显示并播放笛子曲片段01music-demo2.wav。其中有一个C音主频率为523Hz，相应的片段存入了91mi1.wav; 我们人为用正弦波生成了频率为523Hz的声音文件，存入了91mi-pure.wav。在Audacity中聆听、看波形、频谱图。
- 在R中读入01music-demo2.wav，画C523Hz片段时间序列图、谱密度、动态谱峰图。生成523Hz正弦波，画图和谱密度估计。(atsa01.r::demo.flute())
- 演示2: 直接生成的两个单音的序列图和谱密度估计。复音。聆听。(atsa01.r::demo.song())

#### 音调和频率对照表

十二音阶的高音和低音频率相差一倍，2开12次方根为1.059463，是EF之间和BC之间的频率倍数；2开6次方根为1.122137，是CD, DE, FG, GA, AB之间的频率倍数。

音调	频率	频率	频率
C	262	523	1047
D	294	587	1175
E	330	659	1319
F	349	698	1397
G	392	784	1568
A	440	880	1760
B	494	988	1976

#### 谱函数定义

- 设平稳序列 $\{X_t\}$  有自协方差函数 $\{\gamma_k\}$ .

- 如果有 $[-\pi, \pi]$ 上的单调不减右连续的函数 $F(\lambda)$ 使得

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} dF(\lambda), \quad F(-\pi) = 0, \quad k \in \mathbb{Z}, \quad (7.1)$$

则称 $F(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma_k\}$ 的谱分布函数, 简称为谱函数;

- 如果有 $[-\pi, \pi]$ 上的非负函数 $f(\lambda)$ 使得

$$\gamma_k = \int_{-\pi}^{\pi} f(\lambda) e^{ik\lambda} d\lambda, \quad k \in \mathbb{Z}, \quad (7.2)$$

则称 $f(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma_k\}$ 的谱密度函数或功率谱密度, 简称为谱密度或功率谱.

- 谱反映了平稳序列的相关结构。
- 演示: 高频时间序列和低频时间序列的序列图和密度(atsa01.r::demo.song()).

### 谱函数与谱密度关系

- 定理7.1(Herglotz定理) 平稳序列的谱函数是惟一存在的。
- 若 $F(\lambda)$ 和 $f(\lambda)$ 同时存在则

$$F(\lambda) = \int_{-\pi}^{\lambda} f(s) ds \quad (7.3)$$

- 若 $F(\lambda)$ 绝对连续则其几乎处处导数 $F'(\lambda)$ 为谱密度。
- 若 $F(\lambda)$ 是连续函数, 除去有限点外导函数存在且连续则

$$f(\lambda) = \begin{cases} F'(\lambda) & \text{当 } F'(\lambda) \text{ 存在} \\ 0 & \text{当 } F'(\lambda) \text{ 不存在} \end{cases}$$

是谱密度。

### 线性平稳列的谱密度

- 定理7.2 如果 $\{\varepsilon_t\}$ 是 $WN(0, \sigma^2)$ , 实数列 $\{a_j\}$ 平方可和, 则线性平稳序列

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z},$$

有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} a_j e^{ij\lambda} \right|^2. \quad (7.4)$$

- 证明略。

### 两正交序列的谱

- **定理7.3** 设 $\{X_t\}$  和 $\{Y_t\}$ 是相互正交的零均值平稳序列,  $c$ 是常数, 定义

$$Z_t = X_t + Y_t + c, \quad t \in \mathbb{Z}.$$

1. 如果 $\{X_t\}$  和 $\{Y_t\}$ 分别有谱函数 $F_X(\lambda)$  和 $F_Y(\lambda)$ , 则平稳序列 $\{Z_t\}$  有谱函数 $F_Z(\lambda) = F_X(\lambda) + F_Y(\lambda)$ .
2. 如果 $\{X_t\}$  和 $\{Y_t\}$ 分别有谱密度 $f_X(\lambda)$  和 $f_Y(\lambda)$ , 则 $\{Z_t\}$  有谱密度 $f_Z(\lambda) = f_X(\lambda) + f_Y(\lambda)$ .

- **证明:** 主要由

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k)$$

及谱函数和谱密度定义可得。

### §1.7.2 线性滤波与谱

#### 线性滤波与谱

- 设平稳序列 $\{X_t\}$ 有谱函数 $F_X(\lambda)$ 和自协方差函数 $\{\gamma_k\}$ .  $H = \{h_j\}$ 是一个绝对可和的保时线性滤波器(参见§1.3).
- 当输入过程是 $\{X_t\}$ 时, 输出过程是

$$Y_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j}, \quad t \in \mathbb{Z}.a.s. \quad (7.5)$$

•

$$\gamma_Y(k) = \sum_{l,j=-\infty}^{\infty} h_l h_j \gamma(k+l-j). \quad (7.6)$$

求和意义为实数级数绝对收敛。

- 由控制收敛定理

$$\begin{aligned} \gamma_Y(k) &= \sum_{l,j=-\infty}^{\infty} h_l h_j \int_{-\pi}^{\pi} \exp(i(k+l-j)\lambda) dF_X(\lambda) \\ &= \int_{-\pi}^{\pi} \sum_{l,j=-\infty}^{\infty} h_l h_j \exp(i(l-j)\lambda) e^{ik\lambda} dF_X(\lambda) \\ &= \int_{-\pi}^{\pi} \left| \sum_{j=-\infty}^{\infty} h_j \exp(-ij\lambda) \right|^2 e^{ik\lambda} dF_X(\lambda) \\ &= \int_{-\pi}^{\pi} |H(e^{-i\lambda})|^2 e^{ik\lambda} dF_X(\lambda), \end{aligned} \quad (7.7)$$



- 其中

$$H(z) = \sum_{j=-\infty}^{\infty} h_j z^j, \quad |z| \leq 1. \quad (7.8)$$

- 线性滤波输出 $\{Y_t\}$ 的谱函数为

$$F_Y(\lambda) = \int_{-\pi}^{\lambda} |H(e^{-is})|^2 dF_X(s). \quad (7.9)$$

- 当 $\{X_t\}$ 有谱密度 $f_X(\lambda)$ 时

$$F_Y(\lambda) = \int_{-\pi}^{\lambda} |H(e^{-is})|^2 f_X(s) ds. \quad (7.10)$$

即 $\{Y_t\}$ 谱密度为

$$f_Y(\lambda) = |H(e^{-i\lambda})|^2 f_X(\lambda) \quad (7.11)$$

- 结论归纳成定理7.4.

#### 定理7.4

- **定理7.4** 设 $\{X_t\}$ 是平稳序列,  $H = \{h_j\}$ 是绝对可和的保时线性滤波器,  $\{Y_t\}$ 为滤波输出

$$Y_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j}, \quad t \in \mathbb{Z}.a.s. \quad (7.5)$$

$H(z)$ 是滤波器 $H$ 的特征多项式

$$H(z) = \sum_{j=-\infty}^{\infty} h_j z^j, \quad |z| \leq 1. \quad (7.8)$$

- (1) 如果 $\{X_t\}$ 有谱函数 $F_X(\lambda)$ , 则 $\{Y_t\}$ 有谱函数

$$F_Y(\lambda) = \int_{-\pi}^{\lambda} |H(e^{-is})|^2 dF_X(s). \quad (7.9)$$

- (2) 如果 $\{X_t\}$ 有谱密度 $f_X(\lambda)$ , 则 $\{Y_t\}$ 有谱密度

$$f_Y(\lambda) = |H(e^{-i\lambda})|^2 f_X(\lambda) \quad (7.11)$$

## §1.7.3 采样定理

## 连续时平稳过程

- 设 $\{X(t), t \in \mathbb{R}\}$ 是连续时的实值随机过程, 若

$$EX(t) \equiv \mu, \quad \forall t \in \mathbb{R}$$

$$\text{Cov}(X(t), X(s)) = \gamma(t-s), \quad t, s \in \mathbb{R}$$

则称 $\{X(t)\}$ 为连续时平稳过程。

- $\{\gamma(\tau), \tau \in \mathbb{R}\}$ 称为 $\{X(t)\}$ 的自协方差函数。

## 连续时平稳过程的谱

- 对连续时平稳过程 $\{X(t), t \in \mathbb{R}\}$ , 设自协方差函数为 $\{\gamma(\tau)\}$ , 若有 $(-\infty, \infty)$ 上的单调不减右连续的函数 $F(\lambda)$ 使得

$$\gamma(\tau) = \int_{-\infty}^{\infty} e^{i\tau\lambda} dF(\lambda), \quad F(-\infty) = 0, \quad \tau \in \mathbb{R}, \quad (1.1)$$

则称 $F(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma(\tau)\}$ 的谱函数, 在一定理论条件下也有类似于Herglotz定理的结论。

- 如果有 $(-\infty, \infty)$ 上的非负函数 $f(\lambda)$ 使得

$$\gamma(\tau) = \int_{-\infty}^{\infty} f(\lambda) e^{i\tau\lambda} d\lambda, \quad \tau \in \mathbb{R}, \quad (1.2)$$

则称 $f(\lambda)$ 是 $\{X_t\}$ 或 $\{\gamma(\tau)\}$ 的谱密度函数或功率谱密度, 简称为谱密度或功率谱。

## 采样

- 连续时的平稳过程只能以一定的时间间隔记录, 称为采样。
- 如何保证不损失信息?

## 采样定理

- 定理 设 $\{X(t), t \in \mathbb{R}\}$ 为连续时平稳过程, 有谱函数 $F(\lambda)$ 且谱函数满足

$$\int_{|\lambda| \geq 2\pi\omega_0} dF(\lambda) = 0$$

则

$$X_t = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2\omega_0}\right) \frac{\sin(2\pi\omega_0 t - n\pi)}{2\pi\omega_0 t - n\pi} \quad (L^2), \quad t \in \mathbb{R}$$

- 这时记  $Y_n = X\left(\frac{n}{2\omega_0}\right)$ ,  $n \in \mathbb{Z}$ , 则平稳时间序列  $\{Y_n\}$  是  $X(t)$  以  $\Delta = \frac{1}{2\omega_0}$  等间隔取样的结果。
- 从采样  $\{Y_n, n \in \mathbb{Z}\}$  可以恢复原信号  $\{X(t), t \in \mathbb{R}\}$ 。

### 上界频率

- $\omega_0$  称为上界频率, 采样间隔不大于  $\Delta = \frac{1}{2\omega_0}$  采样恢复原始信号。
- 如果采样间隔大于  $\Delta$ , 信号的高频部分有丢失和混杂入低频的问题。
- **声音信号采样** 数字音频信号以一定频率采样, 并把信号强度数字化为8位、16位等。采样频率越高、数字化位数越多则信号保真度越高。CD音质采样频率为44.1kHz(即采样间隔为  $\Delta = 1/44100$  秒), 信号强度用每声道16位记录。这样, 立体声音频每秒钟记录的数据比特数为  $44100 \times 16 \times 2 = 1411200$  (称为码率1411.2kbps)。每分钟约80MB数据。
- MP3编码使用比较低的采样速率和数字化位数, 并利用听觉心理学进行了压缩, 码率在32 ~ 320kbps。

## §1.8 离散谱序列及其周期性

## 简单的离散谱序列

- 随机变量 $\xi, \eta$ 。  $E\xi = 0, E\eta = 0; E(\xi\eta) = 0, E\xi^2 = E\eta^2 = \sigma^2$ 。常数 $\lambda_0 \in (0, \pi]$ 。

- 

$$Z_t = \xi \cos(\lambda_0 t) + \eta \sin(\lambda_0 t), t \in \mathbb{N}_+ \quad (8.1)$$

- 写成极坐标表示:

$$A = \sqrt{\xi^2 + \eta^2}, \quad \cos \theta = \frac{\xi}{A}, \sin \theta = \frac{\eta}{A} \quad (8.2)$$

$$Z_t = A \cos(t\lambda_0 - \theta) \quad (8.3)$$

- 其实现是一个相位为 $-\theta$ 的角频率为 $\lambda_0$ 的余弦函数的离散采样，表现并不随机，随机性表现在多个实现样本。
- 自协方差函数为

$$\gamma_{t-s} = E(Z_t Z_s) = \sigma^2 \cos((t-s)\lambda_0)$$

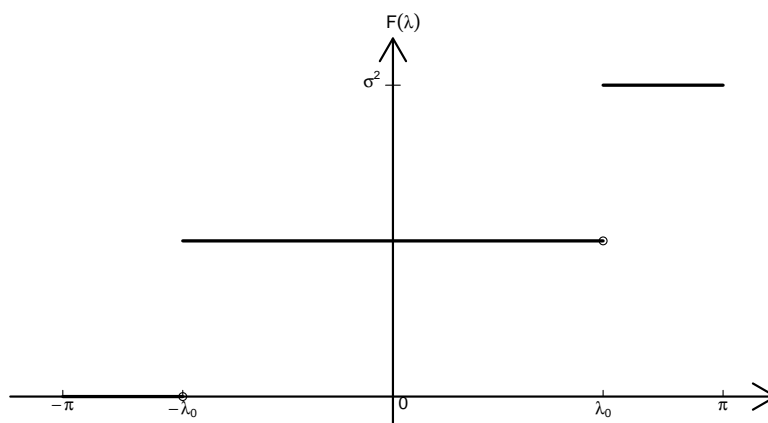
(用到 $\xi, \eta$ 的正交性)

## 简单离散谱序列的谱函数

- 若 $\lambda_0 \neq \pi$ , 谱函数为

$$F(\lambda) = \sigma^2 [0.5I_{[-\lambda_0, \pi]}(\lambda) + 0.5I_{[\lambda_0, \pi]}(\lambda)]$$

- 图像



- 这样的阶梯函数形式的谱函数表示在 $[-\pi, \pi]$ 上的一个测度, 此测度仅在 $-\lambda_0$ 和 $\lambda_0$ 两个点上有质量 $\sigma^2/2$ 。
- 由积分与测度的关系可知

$$\int_{-\pi}^{\pi} e^{ik\lambda} dF(\lambda) = \frac{\sigma^2}{2} [e^{-ik\lambda_0} + e^{ik\lambda_0}] = \gamma_k$$

- 若 $\lambda_0 = \pi$ , 则 $\gamma_k = \sigma^2 \cos(k\pi)$ ,  $F(\lambda) = \sigma^2 I_{\{\pi\}}(\lambda)$ 。

### 离散谱

- 阶梯函数的谱函数称为离散谱函数, 相应的平稳序列称为离散谱序列。
- 离散谱函数没有对应的谱密度, 但是可以逼近。令

$$f_n(\lambda) = n\sigma^2 I_{[-\lambda_0, -\lambda_0+1/(2n)] \cup [\lambda_0-1/(2n), \lambda_0]}(\lambda)$$

$$F_n(\lambda) = \int_{-\pi}^{\lambda} f_n(u) du$$

则 $F_n(\lambda)$ 为连续的折线函数,

$$F_n(\lambda) \rightarrow F(\lambda), \quad \lambda \neq \pm\lambda_0, \quad n \rightarrow \infty.$$

- $f_n(\lambda)$ 是某平稳列的谱密度。易证明

$$\int_{-\pi}^{\pi} e^{ik\lambda} f_n(\lambda) d\lambda \rightarrow \gamma_k, \quad n \rightarrow \infty$$

- 当 $n$ 很大时,  $f_n(\lambda)$ 对应的平稳列和 $\{Z_t\}$ 的表现已经很接近, 其轨道表现为近似周期函数形式。
- 推广来看, 如果某平稳列的谱密度在某处有很高的峰, 则此序列的轨道在峰对应的频率(角频率)处应该表现出周期性。比如, 月度数据的谱密度在角频率 $\frac{2\pi}{12}$ 处表现出高的峰值。
- 演示: `demo01.r::demo.month.spec()`。国际航空订票数据的随机项的谱密度在周期12处有显著峰值。
- 设 $\omega$ 为角频率,  $f$ 为频率,  $T$ 为周期, 则

$$\omega = 2\pi f, \quad T = \frac{1}{f}$$

- 考察数据的周期性变化是谱密度估计的重要应用之一。

### 多个频率成分的离散谱序列

- 设有 $2p$ 个随机变量 $\xi_k, \eta_k, k = 1, \dots, p$ , 所有 $2p$ 个两两正交, 满足

$$E\xi_j = E\eta_j = 0, \quad E\xi_j^2 = E\eta_j^2 = \sigma_j^2 \quad (8.6)$$

- 设 $\lambda_j \in (0, \pi], j = 1, \dots, p$ , 定义

$$\begin{aligned} Z_t &= \sum_{j=1}^p [\xi_j \cos(\lambda_j t) + \eta_j \sin(\lambda_j t)] \\ &= \sum_{j=1}^p A_j \cos(\lambda_j t - \theta_j), \quad t \in \mathbb{N}_+ \end{aligned} \quad (8.7)$$

- 其轨道表现为有 $p$ 个频率成分的非随机函数。
- $\{Z_t\}$ 为零均值平稳列,

$$\gamma_k = \sum_{j=1}^p \sigma_j^2 \cos(k\lambda_j), \quad k \in \mathbb{Z}. \quad (8.9)$$

### 离散谱函数

- 设所有 $\lambda_j \neq \pi$ ,  $\{Z_t\}$ 的谱函数表现为在 $\pm\lambda_j$ 处有跳跃 $\frac{\sigma_j^2}{2}$ 的阶梯函数, 表明谱的能量集中在这 $2p$ 个频率上。
- 也可以考虑可数个正交离散谱序列的叠加。
- 前一节讲谱时的单音例子就是离散谱的例子。
- 重复演示: 单音和复音的例子 `atsa01.r::demo.song()`。

## 第二章 自回归模型

## §2.1 推移算子和常系数差分方程

### §2.1.1 推移算子

#### 推移算子

- 对任何时间序列 $\{X_t\}$ 和无穷级数

$$\psi(z) = \sum_{j=-\infty}^{\infty} b_j z^j$$

- 只要级数

$$\sum_{j=-\infty}^{\infty} b_j X_{t-j}$$

在某种意义下收敛(例如a.s.收敛, 依概率收敛, 均方收敛),

- 就定义

$$\begin{aligned} \psi(\mathcal{B}) &= \sum_{j=-\infty}^{\infty} b_j \mathcal{B}^j, \\ \psi(\mathcal{B})X_t &= \sum_{j=-\infty}^{\infty} b_j \mathcal{B}^j X_t = \sum_{j=-\infty}^{\infty} b_j X_{t-j}. \end{aligned} \quad (1.1)$$

并且称 $\mathcal{B}$ 是时间 $t$ 的向后推移算子, 简称为**推移算子**.

- 显然 $\mathcal{B}X_t = X_{t-1}$ .
- $\mathcal{B}$ 确实是Hilbert空间 $\bar{L}^2(X)$ 上的一个算子, 这里我们只给出它的简单性质。

#### 推移算子的性质

1. 对和 $t$ 无关的随机变量 $Y$ , 有 $\mathcal{B}Y = Y$ .
2.  $\mathcal{B}^n(aX_t) = a\mathcal{B}^n X_t = aX_{t-n}$ .
3.  $\mathcal{B}^{n+m}X_t = \mathcal{B}^n(\mathcal{B}^m)X_t = X_{t-n-m}$ .
4. 对多项式 $\psi(z) = \sum_{j=0}^p c_j z^j$ , 有 $\psi(\mathcal{B})X_t = \sum_{j=0}^p c_j X_{t-j}$ .
5. 对于多项式 $\psi(z) = \sum_{j=0}^p c_j z^j$  和  $\phi(z) = \sum_{j=0}^q d_j z^j$  的乘积 $A(z) = \psi(z)\phi(z)$ , 有

$$A(\mathcal{B})X_t = \psi(\mathcal{B})[\phi(\mathcal{B})X_t] = \phi(\mathcal{B})[\psi(\mathcal{B})X_t].$$

6. 对于时间序列 $\{X_t\}$ ,  $\{Y_t\}$ , 多项式 $\psi(z) = \sum_{j=0}^p c_j z^j$ , 和随机变量 $U, V, W$ , 有

$$\psi(\mathcal{B})(UX_t + VY_t + W) = U\psi(\mathcal{B})X_t + V\psi(\mathcal{B})Y_t + W\psi(1).$$

**证明:** 板书。



## §2.1.2 常系数齐次线性差分方程

## 常系数齐次线性差分方程

- 给定 $p$ 个实数 $a_1, a_2, \dots, a_p, a_p \neq 0$ , 我们称

$$X_t - [a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p}] = 0, \quad t \in \mathbb{Z}, \quad (1.2)$$

为 $p$ 阶齐次常系数线性差分方程, 简称为齐次差分方程.

- 满足(1.2)的实数列(或复数列) $\{X_t\}$ 称为(1.2)的解.
- 满足(1.2)的实值(或复值)时间序列 $\{X_t\}$ 也称为(1.2)的解.

## 齐次线性差分方程的解

- (1.2)的解 $\{X_t\}$ 可以由它的 $p$ 个初值 $X_0, X_1, \dots, X_{p-1}$ 逐步递推得到:

$$\begin{aligned} X_t &= [a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p}], \quad t \geq p, \\ X_{t-p} &= \frac{1}{a_p} [x_t - a_1 X_{t-1} - \dots - a_{p-1} X_{t-p+1}], \quad t - p < 0 \end{aligned}$$

- 若初值是随机变量则递推得到的 $\{X_t\}$ 是时间序列。
- 用推移算子把差分方程写成

$$A(\mathcal{B})X_t = 0, \quad t \in \mathbb{Z}, \quad \text{其中 } A(z) = 1 - \sum_{j=1}^p a_j z^j. \quad (1.3)$$

- $A(z)$ 称为差分方程(1.2)的特征多项式。
- 解有线性性质:  $\{X_t\}, \{Y_t\}$ 是解则

$$\xi X_t + \eta Y_t$$

也是解。

## 差分方程基础解

- 设多项式 $A(z)$ 有 $k$ 个互不相同的零点 $z_1, z_2, \dots, z_k$ , 其中 $z_j$ 是 $r(j)$ 重零点.
- 可以证明对每一 $z_j$ 有

$$A(\mathcal{B})t^l z_j^{-t} = 0, \quad l = 0, 1, \dots, r(j) - 1 \quad (1.4)$$

• 证明:

$$A(z) = \prod_{j=1}^k (1 - z_j^{-1}z)^{r(j)}$$

$$A(\mathcal{B}) = \prod_{j=1}^k (1 - z_j^{-1}\mathcal{B})^{r(j)}$$

只要证明

$$(1 - z_j^{-1}\mathcal{B})^{l+1}(t^l z_j^{-t}) = 0, \quad l = 0, 1, \dots, r(j) - 1 \quad (1.5)$$

归纳法。  $l = 0$  时

$$(1 - z_j^{-1}\mathcal{B})z_j^{-t} = z_j^{-t} - z_j^{-1}z_j^{-(t-1)} = 0$$

设对  $0 \leq l \leq m - 1$  已经证明

$$(1 - z_j^{-1}\mathcal{B})^{l+1}t^l z_j^{-t} = 0$$

则对  $l = m$  有

$$\begin{aligned} & (1 - z_j^{-1}\mathcal{B})^{m+1}t^m z_j^{-t} \\ &= (1 - z_j^{-1}\mathcal{B})^m \left( t^m z_j^{-t} - (t-1)^m z_j^{-(t-1)} \right) \\ &= (1 - z_j^{-1}\mathcal{B})^m (t^m - (t-1)^m) z_j^{-t} \\ &= (1 - z_j^{-1}\mathcal{B})^m (c_1 t^{m-1} + c_2 t^{m-2} + \dots + c_m) z_j^{-t} \\ &= 0 \end{aligned}$$

于是(1.5)成立, 从而(1.4)成立。把基础解线性组合可以得到齐次线性差分方程的通解。

### 齐次线性差分方程的通解

**定理1.1** 设  $A(z)$  有  $k$  个互不相同的零点  $z_1, z_2, \dots, z_k$ , 其中  $z_j$  是  $r(j)$  重零点。则

$$t^l z_j^{-t}, \quad l = 0, 1, \dots, r(j) - 1, \quad j = 1, 2, \dots, k \quad (1.6)$$

是(1.2)的  $p$  个解; 而且, (1.2)的任何解  $\{X_t\}$  都可以写成这  $p$  个解的线性组合:

$$X_t = \sum_{j=1}^k \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z}, \quad (1.7)$$

其中的随机变量  $U_{l,j}$  可以由  $\{X_t\}$  的初值  $X_0, X_1, \dots, X_{p-1}$  惟一决定。(1.7)称为齐次线性差分方程(1.2)的**通解**。此定理关于时间序列叙述, 实际上对差分方程的复数或实数列解也是成立的。

- (1.7)中 $z_j$ 可以是虚数。记

$$U_{l,j} = V_{l,j} e^{i\theta_{l,j}}, \quad z_j = \rho_j e^{i\lambda_j}$$

则

$$\begin{aligned} \operatorname{Re}(U_{l,j} t^l z_j^{-t}) &= \operatorname{Re}(V_{l,j} t^l \rho_j^{-t} e^{-i(\lambda_j t - \theta_{l,j})}) \\ &= V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_{l,j}) \end{aligned}$$

- 差分方程(1.2)的实值解可以表示为

$$\sum_{j=1}^k \sum_{l=0}^{r(j)-1} V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_{l,j}), \quad t \in \mathbb{Z}$$

- $\{V_{l,j}, \theta_{l,j}\}$ 可以由初值 $X_0, X_1, \dots, X_{p-1}$ 唯一决定。

### 通解的收敛性

- 如果差分方程的特征多项式 $A(z)$ 的根都在单位圆外:  $|z_j| > 1, \quad j = 1, 2, \dots, k$ , 或 $A(z) \neq 0, \quad \forall |z| \leq 1$ ,
- 取 $1 < \alpha < \min\{|z_j| : j = 1, \dots, k\}$ , 则

$$t^l |z_j|^{-t} = t^l (\alpha/|z_j|)^t \alpha^{-t} = o(\alpha^{-t})$$

- 于是方程的任意解 $X_t$ 满足

$$|X_t| = o(\alpha^{-t}) \quad a.s., \quad t \rightarrow \infty \quad (1.9)$$

- 称 $X_t$ 以负指数阶收敛到零。
- 如果特征多项式有单位根 $z_j = \exp(i\lambda_j)$ , 则方程有一个周期解

$$X_t = a \cos(\lambda_j t), \quad t \in \mathbb{Z}$$

- 如果单位圆内有根 $z_j = \rho_j \exp(i\lambda_j), \rho_j < 1$ , 则方程有一个爆炸解(发散解)

$$X_t = a \left( \frac{1}{\rho_j} \right)^t \cos(\lambda_j t), \quad t \in \mathbb{Z}$$

## §2.1.3 非齐次线性差分方程

## 非齐次线性差分方程及其通解

- 设 $\{Y_t\}$ 为实值时间序列。

$$A(\mathcal{B})X_t = Y_t, \quad t \in \mathbb{Z} \quad (1.10)$$

- 满足(1.10)的时间序列 $\{X_t\}$ 称为(1.10)的解。
- 如果有(1.10)的某个解(称为特解) $\{X_t^{(0)}\}$ , 则通解可以写成

$$X_t = X_t^{(0)} + \sum_{j=1}^k \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z} \quad (1.11)$$

## §2.2 自回归模型及其平稳性

### §2.2.1 特例: AR(1)

#### AR(1)例子

•

$$X_t = aX_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \{\varepsilon_t\} \sim \text{WN}(0, \sigma^2) \quad (2.1)$$

- 从初值 $X_0$ 出发。 $a$ 越小，初值影响减小越快。
- $|a|$ 接近于1时，初值和前面的 $\varepsilon_{t-j}$ 影响减小越慢，序列振荡。
- 只要 $|a| < 1$ ，序列最终可以稳定下来。称系统(2.1)是稳定的。
- 如果 $a = \pm 1$ 则序列振荡越来越大，呈爆炸型。
- $|a| > 1$ 时序列也不能稳定。 $|a| \geq 1$ 时称(2.1)是非稳定的。(演示)

#### AR(1)的差分方程及平稳解

$A(z) = 1 - az$ 是差分方程(2.1)的特征多项式。 $z_1 = \frac{1}{a}$ 是特征根。

稳定的充分必要条件是 $|a| < 1$ ，或 $|z_1| > 1$ ，即特征根都在单位圆外。

- 当 $|a| < 1$ 时下面的线性序列有定义:

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j}, \quad t \in \mathbb{Z} \quad (2.2)$$

•

$$\begin{aligned} aX_{t-1} + \varepsilon_t &= a \sum_{j=0}^{\infty} a^j \varepsilon_{t-1-j} + \varepsilon_t \\ &= \sum_{i=1}^{\infty} a^i \varepsilon_{t-i} + \varepsilon_t \quad (i = j + 1) \\ &= X_t \end{aligned}$$

- 于是平稳序列(2.2)是非齐次差分方程(2.1)的解，称为平稳解。
- (2.1)的通解为

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j} + \xi a^t, \quad t \in \mathbb{Z}$$

- 当  $t \rightarrow \infty$  时(2.1)的所有解a.s.收敛到平稳解(2.2)。收敛速度是负指数速度  $|a|^t$ 。
- 平稳解可以看成系统(2.1)处于稳定状态的情况。
- 特征根  $\frac{1}{a}$  离单位圆越远, 稳定性越好。

### §2.2.2 一般AR( $p$ )

#### AR( $p$ )模型

- **定义2.1**(AR( $p$ )模型) 如果  $\{\varepsilon_t\}$  是白噪声  $WN(0, \sigma^2)$ , 实数  $a_1, a_2, \dots, a_p$  ( $a_p \neq 0$ ) 使得多项式  $A(z)$  的零点都在单位圆外:

$$A(z) = 1 - \sum_{j=1}^p a_j z^j \neq 0, \quad |z| \leq 1, \quad (2.4)$$

则称 $p$ 阶差分方程

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (2.5)$$

是一个 $p$ 阶自回归模型, 简称为AR( $p$ )模型.

- 满足AR( $p$ )模型(2.5)的平稳时间序列  $\{X_t\}$  称为(2.5)的**平稳解**或AR( $p$ )序列.
- 称  $\mathbf{a} = (a_1, a_2, \dots, a_p)^T$  是AR( $p$ )模型的自回归系数.
- 称条件(2.4) 是**稳定性条件**或**最小相位条件**.
- $A(z)$ 称为模型(2.5)的特征多项式. 模型可写成

$$A(\mathcal{B})X_t = \varepsilon_t, \quad t \in \mathbb{Z} \quad (2.6)$$

### §2.2.3 平稳解和通解

#### AR( $p$ )的平稳解

- 设多项式  $A(z)$  的互异根是  $z_1, z_2, \dots, z_k$ .
- 取  $1 < \rho < \min\{|z_j|\}$ .
- $A^{-1}(z) = 1/A(z)$  是  $\{z : |z| \leq \rho\}$  内的解析函数.
- 从而有Taylor级数

$$A^{-1}(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq \rho. \quad (2.7)$$

- 由级数(2.7)在 $z = \rho$ 的收敛性得到 $|\psi_j \rho^j| \rightarrow 0$ , 当 $j \rightarrow \infty$ .
- 于是由

$$\psi_j = o(\rho^{-j}), \text{ 当 } j \rightarrow \infty \quad (2.8)$$

知道 $\{\psi_j\}$ 绝对可和. 而且,  $\min\{|z_j|\}$ 越大,  $\psi_j$ 趋于零越快.

- 令

$$A^{-1}(\mathcal{B}) = \sum_{j=0}^{\infty} \psi_j \mathcal{B}^j.$$

- 如果 $\{X_t\}$ 是(2.6)的平稳解, 则形式地

$$X_t = A^{-1}(\mathcal{B})A(\mathcal{B})X_t = A^{-1}(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}. \quad (*)$$

- 可见平稳解如果存在必然为

$$X_t = A^{-1}(\mathcal{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (2.9)$$

- $\{\psi_j\}$ 称为平稳序列 $\{X_t\}$ 的Wold系数。
- (\*)式不严格, 我们给出两个引理把(\*)严格化。

### 线性滤波的交换率

**引理1:** 设 $\{a_k\}, \{b_j\}$ 为两个绝对可和的实数列, 则实数列

$$d_m = \sum_{j=-\infty}^{\infty} a_j b_{m-j}$$

绝对可和, 记

$$A(z) = \sum_k a_k z^k, \quad B(z) = \sum_j b_j z^j, \quad D(z) = \sum_m d_m z^m$$

- (1) 若 $\{y_t\}$ 为有界的数列:  $|y_t| \leq M, t \in \mathbb{Z}$ , 则

$$A(\mathcal{B})[B(\mathcal{B})y_t] = B(\mathcal{B})[A(\mathcal{B})y_t] = D(\mathcal{B})y_t$$

- (2) 若 $\{X_t\}$ 为平稳列, 则

$$A(\mathcal{B})[B(\mathcal{B})X_t] = B(\mathcal{B})[A(\mathcal{B})X_t] = D(\mathcal{B})X_t, \text{ a.s.}$$

证明: (1)

$$\begin{aligned} & \sum_k \sum_j |a_k| |b_j| |y_{t-k-j}| \\ & \leq M \left( \sum_k |a_k| \right) \left( \sum_j |b_j| \right) < \infty \end{aligned} \quad (*1)$$

所以

$$\begin{aligned} A(\mathcal{B})B(\mathcal{B})y_t &= \sum_k a_k \sum_j b_j y_{t-k-j} \\ &= \sum_k \sum_j a_k b_j y_{t-k-j} \\ &= \sum_k \sum_m a_k b_{m-k} y_{t-m} \quad (\text{令 } m = k + j, j = m - k) \\ &= \sum_m \sum_k a_k b_{m-k} y_{t-m} \quad (\text{无穷级数次序交换用到}(*1)\text{式}) \\ &= \sum_m d_m y_{t-m} \end{aligned}$$

同样可证  $B(\mathcal{B})A(\mathcal{B})y_t = D(\mathcal{B})y_t$ 。

(2) 定义允许取 $+\infty$ 值的随机变量

$$V = \sum_k \sum_j |a_k| |b_j| |X_{t-k-j}|$$

则

$$EV \leq \sqrt{E(X_1^2)} \left( \sum_k |a_k| \right) \left( \sum_j |b_j| \right) < +\infty$$

所以

$$0 \leq V < +\infty, \quad \text{a.s.}$$

因此以概率1成立:

$$\sum_k a_k \sum_j b_j X_{t-k-j} \quad \text{绝对收敛}$$

级数可交换次序, 同上可证明

$$A(\mathcal{B})B(\mathcal{B})X_t = B(\mathcal{B})A(\mathcal{B})X_t = D(\mathcal{B})X_t$$



## 线性滤波的逆

**引理2** 设实系数多项式  $A(z) = \sum_{j=0}^p \phi_j z^j$  满足最小相位条件:

$$A(z) \neq 0, \quad \forall |z| \leq 1$$

则存在  $\delta > 0$  使

$$A^{-1}(z) \triangleq \frac{1}{A(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1 + \delta$$

且有

(1) 若  $\{y_t\}$  为有界数列, 则

$$y_t = A^{-1}(\mathcal{B})A(\mathcal{B})y_t = A(\mathcal{B})A^{-1}(\mathcal{B})y_t$$

(2) 若  $\{X_t\}$  为平稳列, 则

$$X_t = A^{-1}(\mathcal{B})A(\mathcal{B})X_t = A(\mathcal{B})A^{-1}(\mathcal{B})X_t, \quad \text{a.s.}$$

**证明:** 设多项式  $A(z)$  的根为  $z_1, \dots, z_p$ , 取

$$1 < 1 + \delta < \min_j |z_j|$$

则  $A(z)$  在  $|z| \leq 1 + \delta$  无零点,  $A^{-1}(z)$  在  $|z| \leq 1 + \delta$  解析, 可以展开为 Taylor 级数

$$A^{-1}(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1 + \delta$$

由于  $\sum_{j=0}^{\infty} \psi_j (1 + \delta)^j$  收敛所以  $\psi_j (1 + \delta)^j \rightarrow 0$  当  $j \rightarrow \infty$ , 即  $\psi_j = o((1 + \delta)^{-j})$  ( $j \rightarrow \infty$ )。由此知  $\{\psi_j\}$  绝对可和。定义  $\phi_j = 0$ , 对  $j < 0$  或  $j > p$ , 定义  $\psi_j = 0$  对  $j < 0$ , 注意到

$$A(z)A^{-1}(z) = 1$$

令  $d_0 = 1, d_m = 0$  对  $m \neq 0$ , 由引理1可得所需结论。

**AR(p)的平稳解及通解定理****定理2.1**

1. 由(2.9)定义的时间序列  $\{X_t\}$  是  $\text{AR}(p)$  模型(2.5)的唯一(a.s.意义)平稳解;
2.  $\text{AR}(p)$  的模型的通解有如下形式

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \sum_{j=1}^k \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z}. \quad (2.10)$$

## 定理2.1证明

•

$$X_t = A^{-1}(\mathcal{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (2.9)$$

- 由引理2知(2.9)是平稳序列, 且(2.9)定义的 $\{X_t\}$ 满足

$$A(\mathcal{B})X_t = A(\mathcal{B})A^{-1}(\mathcal{B})\varepsilon_t = \varepsilon_t$$

即(2.9)定义的 $\{X_t\}$ 是模型(2.5)的解。

- 再来证明(2.9)是唯一的平稳解。
- 如果(2.5)另有平稳解 $\{Y_t\}$ , 则 $A(\mathcal{B})Y_t = \varepsilon_t$ , 由引理2得

$$Y_t = A^{-1}(\mathcal{B})A(\mathcal{B})Y_t = A^{-1}(\mathcal{B})\varepsilon_t = X_t.$$

即平稳解唯一。

- 由非齐次线性常系数差分方程的知识立即可得定理的第二条结论。

## Wold系数的递推公式

- 记 $a_0 = -1$ 则 $A(z) = -\sum_{j=0}^p a_j z^j$ ,

$$1 = A(z)A^{-1}(z) = -\sum_{m=0}^{\infty} \left( \sum_{j=0}^p a_j \psi_{m-j} \right) z^m$$

- 故 $\psi_0 = 1$ ,  $\sum_{j=0}^p a_j \psi_{m-j} = 0$ ,  $m > 0$ 。
- 于是

$$\psi_m = \sum_{j=1}^p a_j \psi_{m-j}, \quad m = 1, 2, \dots$$

$$\psi_m = 0, \quad m < 0$$

## 通解与平稳解的关系

- $AR(p)$ 的通解 $\{Y_t\}$ 与平稳解有如下关系

$$|Y_t - X_t| = \left| \sum_{j=1}^k \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t} \right| = o(\rho^{-t}), \text{ a.s. } t \rightarrow \infty$$

其中 $1 < \rho < \min\{|z_j|\}$ 。

- 根离单位园越远, 稳定下来的速度越快。
- 可以用此事实作为模拟产生 $AR(p)$ 序列的理论基础。

**AR序列的模拟**

- 取 $x_{1-p} = \cdots = x_0 = 0$ , 生成 $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$ .
- 迭代得到 $x_t = \varepsilon_t + \sum_{j=0}^p a_j x_{t-j}, j = 1, 2, \dots, n_0 + n$ .
- 取 $y_t = x_{t+n_0}, t = 1, 2, \dots, n$ .
- $n_0$ 取50即可, 但特征根接近单位圆时要取大的 $n_0$ 。

## §2.3 AR(p)序列的谱密度和Yule-Walker方程

### §2.3.1 AR(p)序列的谱密度

#### AR(p)序列的自协方差

- 因为AR(p)的平稳解为

$$X_t = A^{-1}(\mathcal{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

由线性平稳列性质知 $\{X_t\}$ 为零均值, 自协方差函数为

$$\gamma_k = E(X_{t+k}X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \quad k = 0, 1, \dots \quad (3.1)$$

- 设 $1 < \rho < \min\{|z_j|\}$ , 则 $\psi_j = o(\rho^{-j})$ , 有

$$\begin{aligned} |\gamma_k| &\leq \sigma^2 \sum_{j=0}^{\infty} |\psi_j| \cdot |\psi_{j+k}| \leq \sigma^2 \left( \sum_{j=0}^{\infty} |\psi_j| \right) \left( \sum_{l=k}^{\infty} |\psi_l| \right) \\ &\leq c_0 \sum_{l=k}^{\infty} \rho^{-l} \leq c_1 \rho^{-k} \end{aligned} \quad (3.2)$$

即 $\{\gamma_k\}$ 负指数衰减。

- $\{X_t\}$ 序列前后的相关减少很快, 称为时间序列的**短记忆性**。
- 特征根离单位圆越远 $\{\gamma_k\}$ 衰减越快。

#### AR(p)的谱密度

- 由线性平稳列的谱密度公式得平稳解的谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \psi_j e^{ij\lambda} \right|^2$$

- 而 $\sum \psi_j z^j = 1/A(z)$ 所以

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|A(e^{i\lambda})|^2} \quad (3.3)$$

- $f(\lambda)$ 是一个恒正的偶函数。
- 如果 $A(z)$ 有靠近单位圆的根 $\rho_j e^{i\lambda_j}$ 则 $|A(e^{i\lambda_j})|$ 会接近零, 造成谱密度在 $\lambda = \lambda_j$ 处有一个峰值。

## 谱密度的自协方差函数反演公式

- 谱密度的定义是满足

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$$

的非负可积函数。上式是一个Fourier级数系数的公式(差一个常数)。

- 在 $\{\gamma_k\}$ 满足一定条件下 $f(\lambda)$ 必存在且可表成 $\{\gamma_k\}$ 的Fourier级数。
- **定理3.1** 如果平稳序列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$  绝对可和:  $\sum |\gamma_k| < \infty$ , 则 $\{X_t\}$ 有谱密度

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\lambda}. \quad (3.4)$$

由于谱密度是实值函数, 所以(3.4)还可以写成

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \cos(k\lambda) = \frac{1}{2\pi} \left[ \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(k\lambda) \right].$$

**证明:** 因为 $\{\gamma_k\}$ 绝对可和所以(3.4)右边绝对一致收敛,  $f(\lambda)$ 连续。于是积分与级数可交换:

$$\int_{-\pi}^{\pi} f(\lambda) e^{ij\lambda} d\lambda = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \int_{-\pi}^{\pi} e^{-i(k-j)\lambda} d\lambda = \gamma_j.$$

还要验证 $f(\lambda)$ 非负。若 $X_1, \dots, X_N$ 为 $\{X_t\}$ 的观测值,

$$I_N(\lambda) = \frac{1}{2\pi N} \left| \sum_{j=1}^N X_j e^{ij\lambda} \right|^2, \quad \lambda \in [-\pi, \pi]$$

称为 $X_1, \dots, X_N$ 的**周期图**。

令 $f_N(\lambda) = \mathbb{E} I_N(\lambda)$ , 则 $f_N(\lambda) \geq 0$ , 于是

$$\begin{aligned} 0 \leq f_N(\lambda) &= \frac{1}{2\pi N} \sum_{k=1}^N \sum_{j=1}^N \gamma_{k-j} e^{-i(k-j)\lambda} \\ &= \frac{1}{2\pi N} \sum_{m=1-N}^{N-1} (N - |m|) \gamma_m e^{-im\lambda} \\ &= \frac{1}{2\pi} \sum_{m=1-N}^{N-1} \gamma_m e^{-im\lambda} - \frac{1}{2\pi N} \sum_{m=1-N}^{N-1} |m| \gamma_m e^{-im\lambda}. \end{aligned} \quad (*)$$

由Kronecker引理知后一项趋于0, 于是

$$f(\lambda) = \lim_{N \rightarrow \infty} f_N(\lambda) \geq 0.$$

附注：(\*)式的二重求和的简化

•

$$\begin{aligned} & \frac{1}{2\pi N} \sum_{k=1}^N \sum_{j=1}^N \gamma_{k-j} e^{-i(k-j)\lambda} \\ &= \frac{1}{2\pi N} \sum_{k=1}^N \sum_{m=k-N}^{k-1} \gamma_m e^{-im\lambda} \quad (\text{令 } m = k - j) \end{aligned}$$

- 交换 $m$ 与 $k$ 的求和次序。因为关于 $m$ 的条件为

$$k - N \leq m \leq k - 1$$

所以 $k \leq m + N$ ,  $k \geq m + 1$ , 求和变为

•

$$\frac{1}{2\pi N} \sum_{m=1-N}^{N-1} \sum_{k=\max(m+1,1)}^{k=\min(m+N,N)} \gamma_m e^{-im\lambda}$$

- 因为 $m \geq 0$ 时 $k$ 的求和从 $m+1$ 到 $N$ 有 $N-m$ 项,  $m < 0$ 时 $k$ 的求和从1到 $m+N$ 有 $N-|m|$ 项, 所以求和变为

$$\frac{1}{2\pi N} \sum_{m=1-N}^{N-1} (N-|m|) \gamma_m e^{-im\lambda}$$

AR( $p$ )谱密度的自协方差函数表示

推论3.2 AR( $p$ )的平稳解序列 $\{X_t\}$ 有谱密度

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\lambda} = \frac{\sigma^2}{2\pi} \frac{1}{|A(e^{i\lambda})|^2}.$$

### §2.3.2 Yule-Walker方程

白噪声列与平稳解的关系

- $A(\mathcal{B})X_t = \varepsilon_t$ 的平稳解为

$$X_t = A^{-1}(\mathcal{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

- 对 $k \geq 1$ 由控制收敛定理得

$$E(X_t \varepsilon_{t+k}) = \sum_{j=0}^{\infty} \psi_j E(\varepsilon_{t-j} \varepsilon_{t+k}) = 0$$

即 $X_t$ 与未来的输入不相关。

- 如果 $\{\varepsilon_t\}$ 是独立白噪声则 $X_t$ 与未来的输入独立。

**Yule-Walker方程**

- 对  $n \geq p$ , 把  $X_t, X_{t+1}, \dots, X_{t+n-1}$  的递推式写成矩阵形式得

- 

$$\begin{pmatrix} X_t \\ X_{t+1} \\ \vdots \\ X_{t+n-1} \end{pmatrix} = \begin{pmatrix} X_{t-1} & X_{t-2} & \cdots & X_{t-n} \\ X_t & X_{t-1} & \cdots & X_{t+1-n} \\ \vdots & \vdots & & \vdots \\ X_{t+n-2} & X_{t+n-3} & \cdots & X_{t-1} \end{pmatrix} \mathbf{a}_n + \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t+1} \\ \vdots \\ \varepsilon_{t+n-1} \end{pmatrix}, \quad (3.6)$$

- 其中

$$\mathbf{a}_n = (a_{n,1}, a_{n,2}, \dots, a_{n,n})^T \triangleq (a_1, a_2, \dots, a_p, 0, \dots, 0)^T$$

- 定义  $\{X_t\}$  的自协方差矩阵

$$\Gamma_n = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} \quad \text{和} \quad \boldsymbol{\gamma}_n = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}, \quad (3.8)$$

- 在(3.6)两边同时乘上  $X_{t-1}$  后取数学期望, 利用  $X_t$  与未来输入的不相关性有

$$\boldsymbol{\gamma}_n = \Gamma_n \mathbf{a}_n, \quad n \geq p.$$

- 即

$$\gamma_k = a_1 \gamma_{k-1} + a_2 \gamma_{k-2} + \cdots + a_p \gamma_{k-p}, \quad k \geq 1$$

自协方差列  $\{\gamma_k\}$  在  $k \geq 1$  时满足 AR(p) 模型的齐次差分方程

$$A(\mathcal{B})\gamma_k = 0, \quad k \geq 1$$

- 对 $\gamma_0$ ,

$$\begin{aligned}
 \gamma_0 &= EX_t^2 = E \left( \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t \right)^2 \\
 &= E \left( \sum_{j=1}^p a_j X_{t-j} \right)^2 + E \varepsilon_t^2 \\
 &= \mathbf{a}_n^T \Gamma_n \mathbf{a}_n + \sigma^2 \\
 &= \mathbf{a}_n^T \boldsymbol{\gamma}_n + \sigma^2, \quad n \geq p \\
 &= a_1 \gamma_1 + a_2 \gamma_2 + \cdots + a_p \gamma_p + \sigma^2
 \end{aligned}$$

- 定理3.3(Yule-Walker方程)  $AR(p)$ 序列的自协方差函数满足

$$\boldsymbol{\gamma}_n = \Gamma_n \mathbf{a}_n, \quad \gamma_0 = \boldsymbol{\gamma}_n^T \mathbf{a}_n + \sigma^2, \quad n \geq p, \quad (3.9)$$

- 即

$$\gamma_k = a_1 \gamma_{k-1} + a_2 \gamma_{k-2} + \cdots + a_p \gamma_{k-p}, \quad k \geq 1 \quad (\dagger)$$

$$A(\mathcal{B})\gamma_k = 0, \quad k \geq 1$$

$$\sigma^2 = \gamma_0 - a_1 \gamma_1 - a_2 \gamma_2 - \cdots - a_p \gamma_p = A(\mathcal{B})\gamma_0 \quad (\ddagger)$$

- 特别地, 当 $n = p$ 时

$$\Gamma_p \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix} \quad (3.9')$$

记 $\phi_0 = 1, \phi_1 = -a_1, \dots, \phi_p = -a_p$ , 则 $A(z) = \sum_{j=0}^p \phi_j z^j$ , AR模型可写成 $\sum_{j=0}^p \phi_j X_{t-j} = \varepsilon_t$ 。

Yule-Walker方程可写成

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_p \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \vdots & & \vdots \\ \gamma_p & \gamma_{p-1} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

### §2.3.3 自协方差函数的周期性

自协方差函数的周期性



- 对 $k < 0$ 定义 $\psi_k = 0$ 。
- **推论3.4** AR(p)序列的自协方差函数 $\{\gamma_k\}$ 满足和AR(p)模型 $A(\mathcal{B})X_t = \varepsilon_t$ 相应的差分方程

$$\gamma_k - (a_1\gamma_{k-1} + a_2\gamma_{k-2} + \cdots + a_p\gamma_{k-p}) = \sigma^2\psi_{-k}, \quad k \in \mathbb{Z}$$

- **证明:**  $k \geq 0$ 时即定理结论。对 $k < 0$ ,

$$\begin{aligned} & \gamma_k - (a_1\gamma_{k-1} + a_2\gamma_{k-2} + \cdots + a_p\gamma_{k-p}) \\ &= \mathbb{E} \left[ X_{t-k} \left( X_t - \sum_{j=1}^p a_j X_{t-j} \right) \right] \\ &= \mathbb{E}(X_{t-k}\varepsilon_t) = \mathbb{E} \left[ \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-k-j} \varepsilon_t \right] = \sigma^2 \psi_{-k} \quad \square \end{aligned}$$

- 设 $A(z)$ 有 $p$ 个互异根 $z_j = \rho_j e^{i\lambda_j}, j = 1, \dots, p$ , 推导可得

$$\begin{aligned} \gamma_t &= A^{-1}(\mathcal{B})\sigma^2\psi_{-t} \\ &= \sigma^2 \sum_{j=1}^p c_j A^{-1}(z_j^{-1})z_j^{-t} \\ &= \sigma^2 \sum_{j=1}^p A_j \rho_j^{-t} \cos(\lambda_j t + \theta_j), \quad t \geq 0 \end{aligned} \quad (3.12)$$

- 可见如果 $\{z_j\}$ 中有靠近单位圆的复根则 $\{\gamma_k\}$ 的衰减振荡特性会显现出来。

### 例3.1

- AR(4)模型1:

$$z_1, z_2 = 1.09e^{\pm i\pi/3}, \quad z_3, z_4 = 1.098e^{\pm i2\pi/3}$$

周期为 $2\pi/(\pi/3) = 6$ 和 $2\pi/(2\pi/3) = 3$ 。

- AR(4)模型2:

$$z_1, z_2 = 1.264e^{\pm i\pi/3}, \quad z_3, z_4 = 1.273e^{\pm i2\pi/3}$$

- AR(4)模型3:

$$z_1, z_2 = 1.635e^{\pm i\pi/3}, \quad z_3, z_4 = 1.647e^{\pm i2\pi/3}$$

- 见演示: AR roots demo。

## §2.3.4 自协方差函数的正定性

## 自协方差函数的正定性

- AR( $p$ )平稳解唯一故自协方差函数可被自回归系数和白噪声方差唯一决定。
- 反之, 若 $\Gamma_p$ 正定则根据Yule-Walker方程可以从 $\gamma_0, \gamma_1, \dots, \gamma_p$ 解出 $a_1, \dots, a_p, \sigma^2$ :

$$\mathbf{a}_p = \Gamma_p^{-1} \boldsymbol{\gamma}_p, \quad \sigma^2 = \gamma_0 - \boldsymbol{\gamma}_p^T \Gamma_p^{-1} \boldsymbol{\gamma}_p. \quad (3.13)$$

- **定理3.5** 设 $\Gamma_n$ 是平稳序列 $\{X_t\}$ 的 $n$ 阶自协方差矩阵,  $\gamma_0 > 0$ 。

1. 如果 $\{X_t\}$ 的谱密度 $f(\lambda)$ 存在, 则对 $n \geq 1$ ,  $\Gamma_n$ 正定;
2. 如果 $\lim_{k \rightarrow \infty} \gamma_k = 0$ , 则对 $n \geq 1$ ,  $\Gamma_n$ 正定。

## 引理

**引理:** 对实平稳列 $\{X_t\}$ , 设其自协方差阵为 $\Gamma_n$ ,  $n \in \mathbb{N}$ ; 设其谱函数为 $F(\lambda)$ 。对 $\forall \mathbf{b} = (b_1, \dots, b_n) \in R^n$  有

$$\mathbf{b}^T \Gamma_n \mathbf{b} = \int_{-\pi}^{\pi} \left| \sum_{j=1}^n b_j e^{ij\lambda} \right|^2 dF(\lambda)$$

若 $\{X_t\}$ 有谱密度 $f(\lambda)$ 则

$$\mathbf{b}^T \Gamma_n \mathbf{b} = \int_{-\pi}^{\pi} \left| \sum_{j=1}^n b_j e^{ij\lambda} \right|^2 f(\lambda) d\lambda$$

**证明:**

$$\begin{aligned} \mathbf{b}^T \Gamma_n \mathbf{b} &= \sum_{j=1}^n \sum_{k=1}^n b_j b_k \gamma_{j-k} \\ &= \sum_{j=1}^n \sum_{k=1}^n b_j b_k \int_{-\pi}^{\pi} e^{i(j-k)\lambda} dF(\lambda) \\ &= \int_{-\pi}^{\pi} \sum_{j=1}^n \sum_{k=1}^n b_j b_k e^{ij\lambda} e^{-ik\lambda} dF(\lambda) \\ &= \int_{-\pi}^{\pi} \left| \sum_{j=1}^n b_j e^{ij\lambda} \right|^2 dF(\lambda) \quad \square \end{aligned}$$

**定理3.5证明**

(1) 对  $\mathbf{b} = (b_1, \dots, b_n)^T$ ,  $\sum_{j=1}^n b_j z^{j-1}$  至多有  $n-1$  个零点。  $\gamma_0 = \int_{-\pi}^{\pi} f(\lambda) d\lambda > 0$ , 于是

$$\mathbf{b}^T \Gamma_n \mathbf{b} = \int_{-\pi}^{\pi} \left| \sum_{j=1}^n b_j e^{ij\lambda} \right|^2 f(\lambda) d\lambda > 0$$

(2) 用反证法。设  $\Gamma_n$  正定,  $\det(\Gamma_{n+1}) = 0$  和  $EX_t = 0$  (非零均值情况只要减去均值). 定义

$$\mathbf{X}_n = (X_1, X_2, \dots, X_n)^T$$

对任何实向量  $\mathbf{b} = (b_1, b_2, \dots, b_n)^T \neq 0$  有

$$E(\mathbf{b}^T \mathbf{X}_n)^2 = \mathbf{b}^T \Gamma_n \mathbf{b} > 0,$$

且由  $|\Gamma_{n+1}| = 0$  知存在  $\mathbf{a} = (a_1, a_2, \dots, a_{n+1})^T \neq 0$ ,  $a_{n+1} \neq 0$  使得

$$E(\mathbf{a}^T \mathbf{X}_{n+1})^2 = \mathbf{a}^T \Gamma_{n+1} \mathbf{a} = 0.$$

于是

$$\mathbf{a}^T \mathbf{X}_{n+1} = a_1 X_1 + a_2 X_2 + \dots + a_{n+1} X_{n+1} = 0$$

a.s. 成立,  $X_{n+1}$  可以由  $\mathbf{X}_n$  线性表示:

$$X_{n+1} = -\frac{a_n}{a_{n+1}} X_n - \frac{a_{n-1}}{a_{n+1}} X_{n-1} - \dots - \frac{a_1}{a_{n+1}} X_1, \quad \text{a.s.},$$

利用  $\{X_t\}$  的平稳性知道

$$X_t = -\frac{a_n}{a_{n+1}} X_{t-1} - \frac{a_{n-1}}{a_{n+1}} X_{t-2} - \dots - \frac{a_1}{a_{n+1}} X_{t-n}, \quad \text{a.s.}, \quad t \in \mathbb{Z}$$

递推知对任何  $k \geq 1$ ,  $X_{n+k}$  可以由  $X_1, X_1, \dots, X_n$  线性表示, 即有实向量  $\boldsymbol{\alpha} \triangleq \boldsymbol{\alpha}^{(k)} \triangleq (\alpha_1, \dots, \alpha_n)^T$  使得

$$X_{n+k} = (\boldsymbol{\alpha}^{(k)})^T \mathbf{X}.$$

$X_{n+k}$  被  $\mathbf{X}$  线性表示, 说明  $X_{n+k}$  与  $X_1, \dots, X_n$  有强的相关, 而定理假设是  $\gamma_k \rightarrow 0$ , 又说明  $X_{n+k}$  与  $X_1, \dots, X_n$  的相关性要趋于零, 这就会有矛盾, 下面把矛盾严格表述。

用  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  表示  $\Gamma_n$  的特征值, 则有正交矩阵  $T$  使得

$$T \Gamma_n T^T = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

用  $|\boldsymbol{\alpha}^{(k)}|$  表示  $\boldsymbol{\alpha}^{(k)}$  的欧氏模, 则有

$$\begin{aligned}
\gamma_0 &= E X_{n+k}^2 = E((\boldsymbol{\alpha}^{(k)})^T \mathbf{X})^2 = (\boldsymbol{\alpha}^{(k)})^T \Gamma_n \boldsymbol{\alpha}^{(k)} \\
&= ((\boldsymbol{\alpha}^{(k)})^T T^T)(T \Gamma_n T^T)(T \boldsymbol{\alpha}^{(k)}) \\
&\geq \lambda_1 (T \boldsymbol{\alpha}^{(k)})^T (T \boldsymbol{\alpha}^{(k)}) = \lambda_1 |\boldsymbol{\alpha}^{(k)}|^2.
\end{aligned}$$

即有  $|\boldsymbol{\alpha}^{(k)}| \leq \sqrt{\gamma_0/\lambda_1} < \infty$ .

另一方面

$$\begin{aligned}
\gamma_0 &= E((\boldsymbol{\alpha}^{(k)})^T \mathbf{X} \cdot X_{n+k}) = (\boldsymbol{\alpha}^{(k)})^T E(\mathbf{X} X_{n+k}) \\
&= (\boldsymbol{\alpha}^{(k)})^T (\gamma_{n+k-1}, \gamma_{n+k-2}, \dots, \gamma_k)^T \\
&\leq |\boldsymbol{\alpha}^{(k)}| \left( \sum_{j=0}^{n-1} \gamma_{j+k}^2 \right)^{1/2} \\
&\leq (\gamma_0/\lambda_1)^{1/2} \left( \sum_{j=0}^{n-1} \gamma_{k+j}^2 \right)^{1/2} \\
&\rightarrow 0. \quad \text{当 } k \rightarrow \infty
\end{aligned}$$

这与  $\gamma_0 > 0$  矛盾, 故  $\det(\Gamma_{n+1}) = 0$  不成立. □

### 线性平稳序列的自协方差的正定性

- **推论3.6** (系数平方可和的)线性平稳序列的自协方差阵总是正定的。称其自协方差函数  $\{\gamma_j, j \in \mathbb{Z}\}$  是正定序列。

有限个频率的离散谱序列的轨道是周期函数, 可以用有限个历史值的线性组合无误差地预报整个序列。

### §2.3.5 时间序列的可完全预测性

#### 随机变量的线性相关性和线性预测

- 对于方差有限的随机变量  $Y_1, Y_2, \dots, Y_n$ , 如果有不全为零的常数  $b_1, \dots, b_n$  使得

$$\text{Var}\left(\sum_{j=1}^n b_j Y_j\right)^2 = 0,$$

则称随机变量  $Y_1, Y_2, \dots, Y_n$  是**线性相关**的, 否则称为线性无关的。

- 线性相关时, 存在常数  $b_0$  使得  $\sum_{j=1}^n b_j Y_j = b_0$  a.s. 成立。
- 并且当  $b_n \neq 0$  时,  $Y_n$  可以由  $Y_1, Y_2, \dots, Y_{n-1}$  线性表示:

$$Y_n = a_0 + a_1 Y_{n-1} + \dots + a_n Y_1$$

这时我们称  $Y_n$  可以由  $Y_1, Y_2, \dots, Y_{n-1}$  **完全线性预测**。

## 平稳序列的可完全预测性

- 对于平稳序列 $\{X_t\}$ ,  $X_{t-1}, \dots, X_{t-n}$ 的一个带截距的线性组合为 $b_1 X_{t-1} + \dots + b_n X_{t-n} - b_0$ , 这 $n$ 个变量线性无关当且仅当

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^n b_j X_{t-j} - b_0\right) &= \text{Var}\left(\sum_{j=1}^n b_j X_{t-j}\right) \\ &= \mathbf{b} \Gamma_n \mathbf{b} > 0 \end{aligned}$$

即 $\Gamma_n$ 正定。

- 反之, 若 $\Gamma_n$ 正定而 $\Gamma_{n+1}$ 不满秩, 则 $X_t$ 可以被 $X_{t-1}, \dots, X_{t-n}$ 完全线性预测。
- 线性平稳列不能完全线性预测。
- 有限个频率成分的离散谱序列可完全线性预测。

## §2.4 平稳序列的偏相关系数和Levinson递推公式

### §2.4.1 最优线性预测

#### 最优线性预测

- 设  $X_1, X_2, \dots, X_n, Y$  为随机变量。考虑估计问题

$$L(Y|X_1, \dots, X_n) \triangleq \arg \min_{\hat{Y}=a_0+a_1X_1+\dots+a_nX_n} E(Y - \hat{Y})^2$$

称  $L(Y|X_1, \dots, X_n)$  为  $Y$  关于  $X_1, \dots, X_n$  的最优线性估计。

- $L(Y|X_1, \dots, X_n)$  是  $Y$  在空间  $\mathcal{L}(1, X_1, \dots, X_n)$  上的投影。
- 下面推导  $L(Y|X_1, \dots, X_n)$  的公式。

#### 最优线性估计公式

记  $\mathbf{X} = (X_1, \dots, X_n)^T$ , 令  $\boldsymbol{\xi} = \mathbf{X} - E\mathbf{X}$ ,  $\eta = Y - EY$ 。设  $\Sigma_{\mathbf{X}} \triangleq \text{Var}(\mathbf{X}) = \text{Var}(\boldsymbol{\xi})$  正定。记  $\Sigma_{\mathbf{X}, Y} = \text{Cov}(\mathbf{X}, Y)$ 。对  $a_0, a_1, \dots, a_n \in R$ , 记  $\mathbf{a} = (a_1, \dots, a_n)^T$ , 有

$$\begin{aligned} & E(Y - (a_0 + a_1X_1 + \dots + a_nX_n))^2 \\ &= E(\eta - (a_1\xi_1 + \dots + a_n\xi_n))^2 + (EY - a_0 - \mathbf{a}^T E\mathbf{X})^2 \end{aligned}$$

已知  $a_1, \dots, a_n$  后取  $a_0 = EY - \mathbf{a}^T E\mathbf{X}$  就可以使上式后一项为零, 所以不妨设  $E\mathbf{X} = 0, EY = 0$ 。

这时

$$\begin{aligned} & E(Y - (a_1X_1 + \dots + a_nX_n))^2 \\ &= \text{Var}(Y) + \mathbf{a}^T \Sigma_{\mathbf{X}} \mathbf{a} - 2\mathbf{a}^T \Sigma_{\mathbf{X}, Y} \end{aligned}$$

令关于  $\mathbf{a}$  的导数等于零求得最小值点为

$$\mathbf{a} = \Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}, Y}$$

估计误差的最小值为

$$E(Y - \mathbf{a}\mathbf{X})^2 = \text{Var}(Y) - \Sigma_{\mathbf{X}, Y}^T \Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}, Y}$$

当  $|\Gamma_n| = 0$  时, 最优线性估计也存在。

#### 平稳序列的最优线性预测

- 设 $\{X_t\}$ 为零均值平稳列。考虑用 $X_1, \dots, X_n$ 的线性组合预测 $X_{n+1}$ 。设 $\Gamma_n > 0$ ，则

$$\begin{aligned}
 & L(X_{n+1}|X_n, X_{n-1}, \dots, X_1) \\
 &= \left[ \text{Var} \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix} \right]^{-1} \text{Cov} \left( \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix}, X_{n+1} \right) \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix}^T \\
 &= (\Gamma_n^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix})^T \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix} \\
 &\triangleq a_{n1}X_n + a_{n2}X_{n-1} + \dots + a_{nn}X_1 \\
 &\triangleq \mathbf{a}_n^T (X_n, X_{n-1}, \dots, X_1)^T
 \end{aligned}$$

- 称Yule-Walker方程

$$\Gamma_n \begin{pmatrix} a_{n1} \\ \vdots \\ a_{nn} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}$$

简记为

$$\Gamma_n \mathbf{a}_n = \boldsymbol{\gamma}_n$$

的解 $\mathbf{a}_n$ 为 $\{\gamma_k\}$ 的 $n$ 阶Yule-Walker系数。

•

$$L(X_{n+1}|X_n, \dots, X_1) = \mathbf{a}_n^T (X_n, \dots, X_1)^T$$

- 由平稳性

$$L(X_t|X_{t-1}, \dots, X_{t-n}) = \mathbf{a}_n^T (X_{t-1}, \dots, X_{t-n})^T$$

- 最小的线性预测方差为

$$\begin{aligned}
 \sigma_n^2 &\triangleq \text{E}(X_{n+1} - (a_{n1}X_n + \dots + a_{nn}X_1))^2 \\
 &= \text{Var}(X_{n+1}) - \boldsymbol{\gamma}_n^T \Gamma_n \boldsymbol{\gamma}_n \\
 &= \gamma_0 - \boldsymbol{\gamma}_n^T \mathbf{a}_n \\
 &= \gamma_0 - a_{n1}\gamma_1 - \dots - a_{nn}\gamma_n
 \end{aligned}$$

- 由平稳性

$$\text{E}(X_t - (a_{n1}X_{t-1} + \dots + a_{nn}X_{t-n}))^2 = \sigma_n^2$$

### §2.4.2 最小相位性

#### Yule-Walker系数的最小相位性

- 如果 $\{\gamma_k\}$ 是某AR( $p$ )序列的自协方差函数, 则 $p$ 阶的Yule-Walker方程解出的Yule-Walker系数就是AR模型的自回归系数, 所以满足最小相位性:

$$A(z) = 1 - \sum_{j=1}^p a_j z^j \neq 0, \quad \text{对 } |z| \leq 1$$

- 对于一般的平稳列有:

- **定理4.1** 如果实数列 $\gamma_k, k = 0, 1, \dots, n$ 使得

$$\Gamma_{n+1} \triangleq \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-1} \\ \vdots & \vdots & & \vdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_0 \end{pmatrix} > 0$$

则解出的 $n$ 阶Yule-Walker系数 $a_n$ 满足最小相位条件:

$$1 - \sum_{j=1}^n a_{nj} z^j \neq 0, \quad |z| \leq 1.$$

- 证明自学。
- 线性平稳列的自协方差列正定, 所以其任意 $n$ 阶Yule-Walker系数都满足最小相位条件。

### §2.4.3 Levinson递推公式

#### Levinson递推公式

**定理4.2** 如果 $\Gamma_{n+1}$ 正定, 则 $\gamma_k, k = 0, 1, \dots, n$ 的 $1, 2, \dots, n, n+1$ 阶Yule-Walker系数 $\{a_{ij}, i = 1, \dots, n+1, j = 1, \dots, i\}$ 和均方误差 $\sigma_k^2$ 可以如下递推计算:

$$\begin{aligned} \sigma_0^2 &= \gamma_0 \\ a_{1,1} &= \gamma_1 / \gamma_0 \\ \sigma_k^2 &= \sigma_{k-1}^2 (1 - a_{k,k}^2) \\ a_{k+1,k+1} &= \frac{\gamma_{k+1} - a_{k,1}\gamma_k - a_{k,2}\gamma_{k-1} - \cdots - a_{k,k}\gamma_1}{\gamma_0 - a_{k,1}\gamma_1 - a_{k,2}\gamma_2 - \cdots - a_{k,k}\gamma_k} \\ a_{k+1,j} &= a_{k,j} - a_{k+1,k+1}a_{k,k+1-j}, \quad 1 \leq j \leq k \end{aligned} \quad (4.4)$$

其中

$$\sigma_k^2 = E(X_{k+1} - (a_{k,1}X_{k-1} + \cdots + a_{k,k}X_1))^2 \quad (4.5)$$

是用 $X_k, X_{k-1}, \dots, X_1$ 预测 $X_{k+1}$ 的均方误差。



## Levinson公式的记忆方法

- 回忆§2.3中的

$$\gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = 0, \quad k \geq 1 \quad (\dagger)$$

$$\sigma^2 = \gamma_0 - \sum_{j=1}^p a_j \gamma_{0-j} \quad (\ddagger)$$

- 在 $a_{k+1,k+1}$ 的递推公式中, 想象 $p = k, a_j = a_{k,j}, j = 1, \dots, p$ 。
  - 则递推公式分子是( $\dagger$ )的左边(用 $k+1$ 代替 $k$ )。
  - 分母是( $\ddagger$ )的右边。
- $a_{k+1,j}, 1 \leq j \leq k$ 的公式为

$$\begin{pmatrix} a_{k+1,1} \\ \vdots \\ a_{k+1,k} \end{pmatrix} = \begin{pmatrix} a_{k,1} \\ \vdots \\ a_{k,k} \end{pmatrix} - a_{k+1,k+1} \begin{pmatrix} a_{k,k} \\ \vdots \\ a_{k,1} \end{pmatrix}$$

## §2.4.4 偏相关系数

## 偏相关系数

- 定义4.1 如果 $\Gamma_n$ 正定, 称 $a_{n,n}$ 为 $\{X_t\}$ 或 $\{\gamma_k\}$ 的 $n$ 阶偏(自)相关系数。
- 

$$a_{n,n} = \text{Corr}[X_1 - L(X_1|X_2, \dots, X_n), \\ X_{n+1} - L(X_{n+1}|X_2, \dots, X_n)]$$

- 即 $a_{n,n}$ 为 $X_1$ 和 $X_{n+1}$ 扣除 $X_2, \dots, X_n$ 的线性影响后的相关系数。

## AR序列的偏相关系数

- 设 $\{X_t\}$ 是AR( $p$ )序列。其自协方差函数正定。
- 由Yule-Walker方程(3.9)知其 $n$ 阶( $n \geq p$ )Y-W系数为

$$\begin{aligned} \mathbf{a}_n &= (a_1, \dots, a_p, 0, \dots, 0)^T \\ &= (a_{n,1}, a_{n,2}, \dots, a_{n,n})^T, \quad n \geq p \end{aligned} \quad (4.6)$$

- 其偏相关系数满足

$$a_{n,n} = \begin{cases} a_p \neq 0, & n = p \\ 0, & n > p \end{cases} \quad (4.7)$$

称为偏相关系数 $p$ 后截尾。

- 反之, 如果一个零均值平稳列偏相关系数 $p$ 后截尾, 则它必是AR( $p$ )序列。
- 偏相关截尾隐含要求自协方差列正定。

### AR序列的充要条件

- **定理4.3** 设零均值平稳列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$ 是正定序列, 则 $\{X_t\}$ 是AR( $p$ )序列的充分必要条件是, 它的偏相关系数 $\{a_{n,n}\}$   $p$ 后截尾。
- 证明只要证明充分性。
- 记 $(a_{p,1}, \dots, a_{p,p}) = (a_1, \dots, a_p)$ , 令 $\varepsilon_t = X_t - \sum_{j=1}^p a_j X_{t-j}$ , 只要证明 $\{\varepsilon_t\}$ 是白噪声。
- 最小相位性由定理4.1给出。

**证明:** 记 $\mathbf{a}_p = (a_{p,1}, \dots, a_{p,p}) = (a_1, \dots, a_p)$ , 由Levinson公式和 $a_{p+k,p+k} = 0 (k > 0)$ 得

$$\begin{aligned} a_{p+1,j} &= a_{p,j} - a_{p+1,p+1} a_{p,p+1-j} = a_j, & 1 \leq j \leq p \\ a_{p+k,j} &= a_{p+k-1,j} = \dots = a_{p,j} = a_j, & k \geq 2, 1 \leq j \leq p \\ a_{p+k,j} &= a_{p+k-1,j} = 0 & p < j \leq p+k \end{aligned}$$

即 $n \geq p$ 时

$$\mathbf{a}_n = (a_{n,1}, a_{n,2}, \dots, a_{n,n}) = (a_1, a_2, \dots, a_p, 0, \dots, 0)$$

注意 $\mathbf{a}_n$ 是Y-W方程的解, 即

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}$$

可写成

$$\begin{aligned}\gamma_k &= a_1\gamma_{k-1} + a_2\gamma_{k-2} + \cdots + a_p\gamma_{k-p} \\ &= \sum_{j=1}^p a_j\gamma_{k-j}, \quad k \geq 1\end{aligned}\quad (*)$$

由定理4.1知 $A(z) = 1 - \sum_{j=1}^p a_j z^j$ 满足最小相位条件。  
令

$$\varepsilon_t = X_t - \sum_{j=1}^p a_j X_{t-j}, \quad t \in \mathbb{Z}$$

则 $\{\varepsilon_t\}$ 是平稳序列, 满足 $E\varepsilon_t = 0$ ,  $\text{Var}(\varepsilon_t) = \sigma_p^2 > 0$  (因为 $\{\gamma_k\}$ 为正定序列所以 $\{X_t\}$ 不是可完全线性预测的)。

下面只要证明 $\{\varepsilon_t\}$ 是白噪声。

$\forall t > s$ 有

$$\begin{aligned}E(\varepsilon_t X_s) &= E\left[\left(X_t - \sum_{j=1}^p a_j X_{t-j}\right) X_s\right] \\ &= \gamma_{t-s} - \sum_{j=1}^p a_j \gamma_{t-s-j} \\ &= 0 \quad (\text{由}(*))\end{aligned}$$

所以 $t > s$ 时

$$E(\varepsilon_t \varepsilon_s) = E\left[\varepsilon_t \left(X_s - \sum_{j=1}^p a_j X_{s-j}\right)\right] = 0$$

即 $\{\varepsilon_t\}$ 是 $\text{WN}(0, \sigma_p^2)$ , 且 $a_1, a_2, \dots, a_p$ 满足最小相位条件。证毕。

### 本节内容的应用意义

- 有了观测样本 $x_1, x_2, \dots, x_N$ 可以估计样本自协方差函数:

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

- 有了 $\{\hat{\gamma}_k\}$ 可以计算各阶偏相关系数的估计 $\{a_{k,k}\}$ 。
- 如果发现样本偏相关系数呈现截尾性则可以拟合AR模型。
- 定理4.1保证当 $\hat{\Gamma}_{p+1}$ 正定时得到的模型系数满足最小相位条件。
- 最小相位条件保证系统是稳定的, 预测有意义。
- 真实模型为 $\text{AR}(p)$ 时 $\hat{\Gamma}_{p+1}$  a.s.正定。

## §2.5 AR(p)序列举例

## 例5.1

$|a| < 1$ , AR(1)模型

$$\begin{aligned} X_t &= aX_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z} \\ \{\varepsilon_t\} &\sim \text{WN}(0, \sigma^2) \end{aligned}$$

有平稳解

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j}$$

自协方差函数

$$\begin{aligned} \gamma_0 &= \sigma^2 \sum_{j=0}^{\infty} a^{2j} = \frac{\sigma^2}{1-a^2} \\ \gamma_k &= a\gamma_{k-1} = \cdots = a^k \gamma_0, \quad k \geq 1 \end{aligned} \quad (5.2)$$

自相关系数

$$\rho_k = \frac{\gamma_k}{\gamma_0} = a^k \quad (2.1)$$

谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - ae^{i\lambda}|^2} \quad (2.2)$$

$$= \frac{\sigma^2}{2\pi} [1 + a^2 - 2a \cos \lambda]^{-2}, \quad \lambda \in [-\pi, \pi] \quad (2.3)$$

演示:  $a = \pm 0.85$  时序列的演示。

比较:  $a = 0.85$

$a = -0.85$

(1) 数据表现出趋势性, 相邻的数据差别不大;	(1) 数据上下摆动, 趋势性不明显;
(2) (1)中的现象在 $\{\rho_k\}$ 得到体现: 相邻随机变量正相关	(2) (1)中的现象在 $\{\rho_k\}$ 得到体现: 相邻随机变量负相关
(3) $\rho_k$ 单调减少趋于0	(3) $\rho_k$ 正负交替趋于0
(4) 谱密度的能量集中在低频, $f(\lambda) < f(0), \lambda \in (0, \pi]$ , 数据无周期现象, 周期 $T = \frac{2\pi}{0} = \infty$	(4) 谱密度能量集中在高频, $f(\lambda) < f(\pi), \lambda \in [0, \pi)$ , 数据有周期现象, 周期 $T = \frac{2\pi}{\pi} = 2$
(5) 偏相关系数 $a_{1,1} = 0.85$ , $a_{k,k} = 0$ , 当 $k > 1$	(5) 偏相关系数 $a_{1,1} = -0.85$ , $a_{k,k} = 0$ , 当 $k > 1$
(6) 随 $a$ 接近于0, 以更快的速度收敛到0	(6) 上述性质随 $a$ 接近-1变得更明显, 随 $a$ 接近0变得不明显

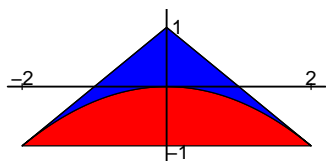
**AR(2): 稳定性条件**

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

$$A(z) = 1 - a_1 z - a_2 z^2 \neq 0, |z| \leq 1$$

稳定性条件为:

$$a_2 \pm a_1 < 1, \quad |a_2| < 1$$



(蓝色: 实根; 红色: 复根)

**AR(2): 自相关系数**

- 设 $A(z)$ 的根为 $z_1 = b_1 e^{i\lambda_1}$ ,  $z_2 = b_2 e^{i\lambda_2}$ 。
- 由Y-W方程

$$\rho_0 = 1,$$

$$\rho_1 = \frac{a_1}{1 - a_2}$$

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2}, \quad k \geq 1$$

- $a_{11} = \rho_1, \quad a_{2,2} = a_2, \quad a_{k,k} = 0 (k \geq 3)$ 。

**AR(2)例子: 稳定域和允许域**

- 

$$\mathcal{A} = \{(a_1, a_2) : a_2 \pm a_1 < 1, |a_2| < 1\}$$

称为AR(2)的稳定域。

- 从Y-W方程可以用 $\rho_1, \rho_2$ 表示 $a_1, a_2$ :

$$a_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}, \quad a_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

- 反之有

$$\rho_1 = \frac{a_1}{1 - a_2}, \quad \rho_2 = a_2 + \frac{a_1^2}{1 - a_2}$$

- $(a_1, a_2) \in \mathcal{A} \iff (\rho_1, \rho_2) \in$

$$\mathcal{C} = \{(\rho_1, \rho_2) : \rho_1^2 < (1 + \rho_2)/2, |\rho_1| < 1, |\rho_2| < 1\}$$

- $\mathcal{C}$ 称为AR(2)的允许域。

### AR(2)的谱密度与特征根

- 特征根与系数有如下关系

$$\begin{aligned} 1 - a_1 z - a_2 z^2 &= (1 - \frac{z}{z_1})(1 - \frac{z}{z_2}) \\ a_1 &= \frac{1}{z_1} + \frac{1}{z_2}, \quad a_2 = -\frac{1}{z_1} \frac{1}{z_2} \\ z_1 z_2 &= \frac{1}{(-a_2)}, \quad z_1 + z_2 = \frac{a_1}{(-a_2)} \end{aligned}$$

•

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|a_2| |e^{i\lambda} - b_1 e^{i\lambda_1}|^2 \cdot |e^{i\lambda} - b_2 e^{i\lambda_2}|^2}$$

### AR(2)实特征根时的表现

- 当 $a_1^2 + 4a_2 \geq 0$ 时(蓝色区域)特征方程有两个实根。
- 当 $a_2 > 0$ 时 $z_1$ 和 $z_2$ 异号, 设 $z_1 < -1, z_2 > 1$ 。  $f(\lambda)$ 在0和 $\pi$ 处有峰值。由 $z_1 + z_2 = \frac{(-a_1)}{a_2}$ 知
  - $a_1 > 0$ 时 $z_1 + z_2 < 0$ , 正根 $z_2$ 离单位圆更近,  $f(\lambda)$ 在0点最大。 $\rho_k$ 都为正数, 震荡衰减。 $\{X_t\}$ 表现出相邻点的正相关。
  - $a_1 < 0$ 时 $z_1 + z_2 > 0$ , 负根 $z_1$ 离单位圆更近,  $f(\lambda)$ 在 $\pi$ 点最大。 $\rho_k$ 主要呈现正负交替衰减。 $\{X_t\}$ 也表现出正负振荡。
  - $a_1 = 0$ 时 $z_1 = -z_2$ , 离单位圆一样近,  $f(\lambda)$ 在0,  $\pi$ 一样高。 $\rho_k$ 都非负, 表现出振荡衰减。 $\{X_t\}$ 表现出正负振荡。
- 当 $a_2 < 0$ 有复根或实根; 为实根时 $z_1$ 与 $z_2$ 同号, 与 $a_1$ 同号。
  - $a_1 > 0$ 时有两个正根,  $f(\lambda)$ 只在0点有峰值。 $\rho_k$ 单调衰减。 $\{X_t\}$ 表现出相邻点的正相关。
  - $a_1 < 0$ 时有两个负根,  $f(\lambda)$ 只在 $\pi$ 点有峰值。 $\rho_k$ 正负交替衰减。 $\{X_t\}$ 表现出正负振荡。

**AR(2)虚特征根时的表现**

- $z_1, z_2$ 是虚根 $\iff a_1^2 + 4a_2 < 0$ (红色区域)。

$$z_1, z_2 = be^{\pm i\lambda_0}, \quad b > 1, \lambda_0 \neq 0, \pi$$

- 由(3.12)得

$$\rho_k = \frac{\cos(k\lambda_0 + \theta_0)}{b^k \cos(\theta_0)}, \quad k \geq 0 \quad (5.3)$$

这时 $\{\rho_k\}$ 振荡衰减, 振荡角频率为 $\lambda_0$ 。

- $\{X_t\}$ 也呈现出在频率 $\lambda_0$ 处的振荡。





### 第三章 滑动平均模型与自回归滑动平均模型

### §3.1 滑动平均模型

#### §3.1.1 模型引入

##### $q$ 步相关

- 平稳序列 $\{X_t\}$ 的自协方差函数若满足 $\gamma_q \neq 0, \gamma_k = 0, k > q$ , 则称 $\{X_t\}$ 是 $q$ 步相关的。
- 有限项的线性平稳列具有 $q$ 步相关性, 称为滑动平均模型。

##### 滑动平均模型的例子

- 每隔2小时记录的化学反应数据时间序列 $\{x_t, t = 1, \dots, 197\}$ 。
- 一阶差分得

$$y_t = x_t - x_{t-1}, t = 2, \dots, 197$$

- $\{y_t\}$ 的样本自相关系数列呈现截尾性。
- 可以拟合

$$Y_t = \varepsilon_t + \hat{b}\varepsilon_{t-1}, t \in \mathbb{Z} \quad (1.1)$$

- 模型特点是 $\{\gamma_k\}$  1步截尾。

#### §3.1.2 MA( $q$ )模型和MA( $q$ )序列

##### MA( $q$ )模型和MA( $q$ )序列

- 定义1.1 设 $\{\varepsilon_t\}$  是 $WN(0, \sigma^2)$ , 如果实数 $b_1, b_2, \dots, b_q$  ( $b_q \neq 0$ )使得

$$B(z) = 1 + \sum_{j=1}^q b_j z^j \neq 0, |z| < 1,$$

则称

$$X_t = \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (1.2)$$

是 $q$ 阶滑动平均模型, 简称为MA( $q$ )模型;

- 称由(1.2)决定的平稳序列 $\{X_t\}$ 是滑动平均序列, 简称为MA( $q$ )序列.
- 如果进一步要求多项式 $B(z)$ 在单位圆上也没有零点:  $B(z) \neq 0$  当 $|z| \leq 1$ , 则称(1.2)是可逆的MA( $q$ )模型, 称相应的平稳序列是可逆的MA( $q$ )序列.

## MA的特征

- 低阶的MA与AR相比较光滑(滑动平均), 振荡较轻。
- 稳定性较好。
- 高阶的MA可以模拟AR的特征。
- 用推移算子把模型写为

$$X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z} \quad (1.3)$$

- 对于可逆MA,  $B^{-1}(z)$ 有Taylor展式

$$B^{-1}(z) = \sum_{j=0}^{\infty} \phi_j z^j, \quad |z| \leq 1 + \delta \quad (\delta > 0)$$

所以

$$\varepsilon_t = B^{-1}(\mathcal{B})X_t = \sum_{j=0}^{\infty} \phi_j X_{t-j} \quad (1.4)$$

## MA序列的自协方差函数

- 记  $b_0 = 1$ , 则对MA( $q$ )序列有  $EX_t = 0$ ,

$$\begin{aligned} \gamma_k &= E(X_t X_{t+k}) \\ &= \begin{cases} \sigma^2 \sum_{j=0}^{q-k} b_j b_{j+k}, & 0 \leq k \leq q \\ 0, & k > q \end{cases} \end{aligned} \quad (1.5)$$

## MA序列的谱密度

- 定理1.1 MA( $q$ )序列  $\{X_t\}$  的自协方差函数是 $q$ 步截尾的:

$$\gamma_q = \sigma^2 b_q \neq 0, \quad \gamma_k = 0, \quad |k| > q. \quad (1.6)$$

并且有谱密度

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2 \\ &= \frac{1}{2\pi} \sum_{k=-q}^q \gamma_k e^{-ik\lambda}, \quad \lambda \in [-\pi, \pi]. \end{aligned} \quad (1.7)$$

- (1.7)的第一条可以用线性平稳列的定理或用谱密度定义直接证明。第二条用自协方差绝对可和时谱密度公式得到也可以用谱密度定义得到。

**MA( $q$ )序列的充要条件**

- MA( $q$ )序列是自协方差函数 $q$ 步截尾的, 反之若平稳列 $\{X_t\}$ 自协方差函数 $q$ 步截尾则其必为MA( $q$ )序列。
- **定理1.3** 设零均值平稳序列 $\{X_t\}$ 有自协方差函数 $\{\gamma_k\}$ , 则 $\{X_t\}$ 是MA( $q$ )序列的充分必要条件是

$$\gamma_q \neq 0, \gamma_k = 0, |k| > q.$$

- 只需要证明充分性。
- 证明需要一个复变函数引理。

**引理1.2**

- **引理1.2** 设实常数 $\{c_j\}$ 使得 $c_q \neq 0$  和

$$g(\lambda) = \frac{1}{2\pi} \sum_{j=-q}^q c_j e^{-ij\lambda} \geq 0, \lambda \in [-\pi, \pi],$$

则有惟一的实系数多项式:

$$B(z) = 1 + \sum_{j=1}^q b_j z^j \neq 0, |z| < 1, b_q \neq 0. \quad (1.8)$$

使得

$$g(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2.$$

这里 $\sigma^2$ 为某个正常数.

- **证明:** 令 $G(z) = \sum_{j=-q}^q c_j z^{j+q}$  则 $G(z)$ 的 $2q$ 个根中 $z_1$ 是根必有 $z_1^{-1}$ 也是根。 $z_1 \neq \pm 1$ 时 $z_1^{-1} \neq z_1$ 。把这样成对的根只取其中一个且要求不在单位圆内即可组成 $B(\cdot)$ 。

**定理1.3证明**

- 由自协方差绝对可和时谱密度公式得

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-q}^q \gamma_k e^{-ik\lambda}$$

- 由引理,

$$f(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{-i\lambda})|^2, \quad B(z) \text{ 单位园内没有根.}$$

- 如果 $B(z)$ 在单位圆内和单位圆上都没有根, 则可定义 $\varepsilon_t = B^{-1}(\mathcal{B})X_t$ , 用线性滤波的谱密度公式可得 $\{\varepsilon_t\}$ 的谱密度是白噪声谱密度。
- 单位圆上可能有根的一般情况可以用Hilbert空间预测的方法证明。(见Brockwell & Davis proposition 3.2.1, P.89, 中译本P.70, 但那里的MA没有根的条件; 或参考谢衷洁《时间序列分析》, 北京大学出版社1990, P92.)。

### §3.1.3 最小序列

#### 最小序列

- **定义1.2** 设 $\{X_t : t \in \mathbb{Z}\}$ 是平稳序列. 用 $H_x$ 表示 $\{X_t\}$ 产生的Hilbert空间, 用 $H_x(s)$ 表示由 $\{X_t : t \neq s\}$ 产生的Hilbert空间. 如果

$$H_x \neq H_x(s)$$

对某个 $s \in \mathbb{Z}$ 成立, 则称 $\{X_t\}$ 是最小序列。

- 对平稳序列, 可以证明对某个 $s$ 成立 $H_x(s) \neq H_x$ 则对所有的 $s$ 成立 $H_x(s) \neq H_x$ 。
- 说明最小序列的每个 $X_t$ 都含有其他 $X_s$ 中没有的信息。
- 可完全线性预测的平稳序列(某 $\Gamma_n$ 不满秩)非最小序列。
- **定理1.4**([7]) 设平稳序列 $\{X_t\}$ 有谱密度 $f(\lambda)$ , 则 $\{X_t\}$ 是最小序列的充分必要条件是

$$\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)} < \infty. \quad (1.10)$$

- 谱密度连续恒正的平稳列是最小序列。
- AR( $p$ )序列是最小序列。
- 可逆的MA( $q$ )序列的谱密度连续有正下界所以是最小序列。
- 单位圆上有根的MA( $q$ )序列其谱密度 $f(\lambda)$ 中有

$$|1 - e^{i(\lambda - \lambda_j)}|^2 = O(|\lambda - \lambda_j|^2)$$

成分所以其倒数不可积, 因此不可逆的MA( $q$ )序列不是最小序列。

- 最小序列一定不能 $n$ 步完全线性预测, 所以其自协方差列正定。

### §3.1.4 MA( $q$ )系数的递推计算

#### MA( $q$ )系数的计算

- MA( $q$ )序列的系数 $(b_1, b_2, \dots, b_q)$ 及 $\sigma^2$  可以被 $\gamma_0, \gamma_1, \dots, \gamma_q$  唯一确定。
- 可以用文献[5]方法计算模型参数。

记

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{q \times q}, \quad C = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{q \times 1} \quad (1.11)$$

$$\Omega_k = \begin{pmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_k \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{k+1} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_q & \gamma_{q+1} & \cdots & \gamma_{q+k-1} \end{pmatrix}, \quad \gamma_q = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{pmatrix}$$

则有:

$$\mathbf{b}_q = \frac{1}{\sigma^2}(\gamma_q - A\Pi C), \quad \sigma^2 = \gamma_0 - C^\tau \Pi C, \quad (1.12)$$

其中

$$\Pi = \lim_{k \rightarrow \infty} \Omega_k \Gamma_k^{-1} \Omega_k^\tau. \quad (1.13)$$

公式(1.12),(1.13)为以后用观测样本估计 $MA(q)$ 模型的参数打下了基础.

### §3.1.5 $MA(q)$ 模型举例

#### $MA(1)$

- 可逆 $MA(1)$

$$X_t = \varepsilon_t + b\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2), \quad |b| < 1$$

- 自协方差和自相关

$$\begin{cases} \gamma_0 = \sigma^2(1 + b^2) \\ \gamma_1 = \sigma^2 b \\ \gamma_k = 0, \quad k \geq 2 \end{cases}$$

$$\begin{cases} \rho_1 = \frac{b}{1+b^2} \\ \rho_k = 0, \quad k \geq 2 \end{cases}$$

- 谱密度

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi} |1 + be^{i\lambda}|^2 \\ &= \frac{\sigma^2}{2\pi} (1 + b^2 + 2b \cos \lambda), \quad \lambda \in [-\pi, \pi] \end{aligned}$$

- 偏相关系数不截尾:

$$a_{k,k} = -\frac{(-b)^k(1-b^2)}{(1-b^{2k+2})}, \quad k \geq 1$$

- 逆表示

$$\varepsilon_t = \sum_{j=0}^{\infty} (-b)^j X_{t-j}$$

### MA(2)

- 可逆MA(2)

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}, \quad t \in \mathbb{Z}$$

- $B(z) = 1 + b_1z + b_2z^2 \neq 0, |z| \leq 1$ .

- 可逆域:

$$\begin{aligned} & \{(b_1, b_2) : B(z) \neq 0, |z| \leq 1\} \\ &= \{(b_1, b_2) : b_2 \pm b_1 > -1, |b_2| < 1\} \end{aligned}$$

- 自协方差

$$\begin{aligned} \gamma_0 &= \sigma^2(1 + b_1^2 + b_2^2) \\ \gamma_1 &= \sigma^2(b_1 + b_1b_2) \\ \gamma_2 &= \sigma^2b_2 \\ \gamma_k &= 0, \quad k > 2 \end{aligned}$$

- 自相关系数

$$\begin{aligned} \rho_1 &= \frac{b_1 + b_1b_2}{1 + b_1^2 + b_2^2}, \quad \rho_2 = \frac{b_2}{1 + b_1^2 + b_2^2} \\ \rho_k &= 0, \quad k > 2 \end{aligned}$$

- 谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 + b_1e^{i\lambda} + b_2e^{i2\lambda}|^2$$

- MA(2)的实际例子:

$$X_t = \varepsilon_t - 0.36\varepsilon_{t-1} + 0.85\varepsilon_{t-2}$$

- 特征根为  $1.084652e^{\pm i1.374297}$ 。

- 

$$\gamma_0 = \sigma^2(1 + b_1^2 + b_2^2) = 7.4084$$

$$\gamma_1 = \sigma^2(b_1 + b_1b_2) = -2.664$$

$$\gamma_2 = \sigma^2b_2 = 3.4$$

$$\gamma_k = 0, \quad k > 2$$

- $(\rho_1, \rho_2) = (-0.3596, 0.4589)$ 。

- 谱密度及系数的求解见演示。



## §3.2 自回归滑动平均模型

### §3.2.1 ARMA( $p, q$ )模型及其平稳解

#### ARMA模型

- **定义2.1** 设 $\{\varepsilon_t\}$ 是 $WN(0, \sigma^2)$ , 实系数多项式 $A(z)$ 和 $B(z)$ 没有公共根, 满足 $b_0 = 1$ ,  $a_p b_q \neq 0$ 和

$$\begin{aligned} A(z) &= 1 - \sum_{j=1}^p a_j z^j \neq 0, |z| \leq 1, \\ B(z) &= \sum_{j=0}^q b_j z^j \neq 0, |z| < 1, \end{aligned} \quad (2.1)$$

就称差分方程:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (2.2)$$

是一个自回归滑动平均模型, 简称为ARMA( $p, q$ )模型. 称满足(2.2)的平稳序列 $\{X_t\}$ 为平稳解或ARMA( $p, q$ )序列.

#### ARMA模型平稳解

- 模型写成

$$A(\mathcal{B})X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z} \quad (2.3)$$

- $A^{-1}(z)B(z)$ 在 $|z| < \rho$ 解析( $1 < \rho < \min\{|z_j|\}$ ,  $\{z_j\}$ 为 $A(z)$ 的所有根), 可以Taylor展开

$$\Psi(z) \triangleq A^{-1}(z)B(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq \rho \quad (2.4)$$

- 易见 $\psi_j = o(\rho^{-j})$ ,  $A^{-1}(\mathcal{B})B(\mathcal{B})\varepsilon_t = \Psi(\mathcal{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ 是线性平稳列。两边用 $A(\mathcal{B})$ 作用, 根据§2.2补充的引理1, 2知

$$A(\mathcal{B})\Psi(\mathcal{B})\varepsilon_t = A(\mathcal{B})A^{-1}(\mathcal{B})B(\mathcal{B})\varepsilon_t = B(\mathcal{B})\varepsilon_t$$

即 $\Psi(\mathcal{B})\varepsilon_t$ 是ARMA( $p, q$ )模型(2.2)的解。

- 反之, 若 $\{Y_t\}$ 是(2.2)的一个平稳解, 在(2.2)两边作用 $A^{-1}(\mathcal{B})$ 即得

$$A^{-1}(\mathcal{B})A(\mathcal{B})Y_t = Y_t = A^{-1}(\mathcal{B})B(\mathcal{B})\varepsilon_t = \Psi(\mathcal{B})\varepsilon_t$$

- 即

$$X_t = A^{-1}(\mathcal{B})B(\mathcal{B})\varepsilon_t = \Psi(\mathcal{B})\varepsilon_t \quad (2.6)$$

是ARMA( $p, q$ )模型(2.2)的唯一平稳解。

- 称(2.6)中的 $\{\psi_j\}$ 为 $\{X_t\}$ 的Wold系数。
- **定理2.1** 由(2.6)定义的平稳序列 $\{X_t\}$ 是ARMA( $p, q$ )模型(2.2)的惟一平稳解。

### ARMA模型方程的通解

- 模型(2.2)的任意解可以写成

$$Y_t = X_t + \sum_{j=1}^k \sum_{l=0}^{r(j)-1} V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_{l,j}), t \in \mathbb{Z}, \quad (2.7)$$

其中 $\{X_t\}$ 为平稳解(2.6),  $z_1, z_2, \dots, z_k$ 为 $A(z)$ 的全体互不相同的零点,  $z_j = \rho_j e^{i\lambda_j}$  有重数 $r(j)$ . 随机变量 $V_{j,l}, \theta_{l,j}$  由 $Y_0 - X_0, Y_1 - X_1, \dots, Y_{p-1} - X_{p-1}$ 惟一决定.

### ARMA模型的模拟生成

- 

$$|Y_t - X_t| \leq \sum_{j=1}^k \sum_{l=0}^{r(j)-1} |V_{l,j}| t^l \rho_j^{-t}, t \rightarrow \infty \quad (2.8)$$

- 可以据此模拟ARMA模型: 取初值 $Y_{-(p-1)} = \dots = Y_{-1} = Y_0 = 0$ , 递推得

$$Y_t = \sum_{j=1}^p a_j Y_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}, t = 1, 2, \dots, m+n$$

当 $m$ 较大时取后一段 $Y_t, t = m+1, m+2, \dots, m+n$ 作为ARMA( $p, q$ )模型的模拟数据。

- 当 $A(z)$ 有靠近单位圆的根时 $m$ 要取得较大。

### §3.2.2 ARMA( $p, q$ )序列的自协方差函数

#### ARMA序列的自协方差函数

- $\{\gamma_k\}$ 可由Wold系数表示:

$$\gamma_k = \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \quad k = 0, 1, 2, \dots \quad (2.10)$$

- 由于 $\psi_j = o(\rho^{-j}), j \rightarrow \infty$ , 由(2.10)可得 $\gamma_k = o(\rho^{-j}), j \rightarrow \infty$ 。

## ARMA序列的Wold系数递推公式

- 记  $b_j = 0, j < 0$  或  $j > q, b_0 = 1; \psi_j = 0, j < 0$ 。
- 由参数  $\mathbf{a}_p = (a_1, \dots, a_p)^T, \mathbf{b}_p = (b_1, \dots, b_q)^T$  计算  $\{\psi_j\}$  时可以递推

$$\psi_j = \begin{cases} 1, & j = 0, \\ b_j + \sum_{k=1}^p a_k \psi_{j-k}, & j = 1, 2, \dots \end{cases} \quad (2.11)$$

## Wold系数递推公式的证明

- 记  $A(z) = 1 - \sum_{j=1}^p a_j z^j = \sum_{j=0}^p \phi_j z^j$ 。注意

$$\begin{aligned} A(z)\Psi(z) &= \sum_{k=0}^p \phi_k z^k \sum_{j=0}^{\infty} \psi_j z^j \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^p \phi_k \psi_{j-k} z^j = B(z) \end{aligned}$$

- 比较系数得

$$\begin{aligned} \sum_{k=0}^p \phi_k \psi_{j-k} &= b_j, \quad j \geq 1 \\ \psi_j &= \sum_{k=1}^p a_k \psi_{j-k} + b_j, \quad j \geq 1 \end{aligned}$$

- 即(2.11)成立。

## §3.2.3 ARMA(p, q)模型的可识别性

## 可识别性

- 我们将证明: 由ARMA(p, q)模型的自协方差函数  $\{\gamma_k\}$  可以决定ARMA(p, q)模型的参数

$$(\mathbf{a}^T, \mathbf{b}^T, \sigma^2) = (a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2)$$

- 引理2.2 设  $\{X_t\}$  是(2.2)的平稳解。如果又有白噪声  $\{\eta_t\}$  和实系数多项式  $C(\mathcal{B}), D(\mathcal{B})$  使得

$$C(\mathcal{B})X_t = D(\mathcal{B})\eta_t, \quad t \in \mathbb{Z},$$

成立, 则  $C(z)$  的阶数  $\geq p$ ,  $D(z)$  的阶数  $\geq q$ 。

- 这主要因为我们要求多项式  $A(z)$  和  $B(z)$  互素。

## ARMA序列的Y-W方程

- ARMA模型的平稳解为

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

所以

$$E(\varepsilon_{t+k} X_t) = 0, \quad k > 0$$

- 类似AR模型可推导ARMA模型的Y-W方程: 在模型方程两边同乘以 $X_{t-k}$ 求期望得

$$E(X_t X_{t-k}) = \sum_{j=1}^p a_j E(X_{t-j} X_{t-k}) + \sum_{j=0}^q b_j E(\varepsilon_{t-j} X_{t-k})$$

即

$$\begin{aligned} \gamma_k &= \sum_{j=1}^p a_j \gamma_{k-j} + \sum_{j=0}^q b_j E(\varepsilon_{t-j} \sum_{l=0}^{\infty} \psi_l \varepsilon_{t-k-l}) \\ &= \sum_{j=1}^p a_j \gamma_{k-j} + \sum_{j=0}^q b_j \psi_{j-k}, \quad k \in \mathbb{Z} \end{aligned}$$

- 当 $q < k$ 时 $\psi_{j-k} = 0, j = 0, 1, \dots, q$ , 上式为

$$\gamma_k = \sum_{j=1}^p a_j \gamma_{k-j}, \quad k \geq q+1$$

- 总之

$$\gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = \begin{cases} \sigma^2 \sum_{j=\max(0,k)}^q b_j \psi_{j-k}, & k < q \\ \sigma^2 b_q, & k = q \\ 0 & k > q \end{cases} \quad (2.14)$$

- 对 $k > q$ 的Y-W方程可以写成矩阵形式:

$$\begin{bmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{bmatrix} = \begin{bmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2} \\ \vdots & \vdots & & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \quad (2.15)$$

- 对比AR( $p$ )的Y-W方程, 相当于 $\Gamma_p$ 的 $(i, j)$ 元素写成 $\gamma_{i-j}$ 后给所有 $\gamma$ 的下标都加上 $q$ 。

- 把系数矩阵记为 $\Gamma_{p,q}$ :

$$\begin{aligned}\Gamma_{p,q} &= (\gamma_{|q+i-j|})_{i,j=1,2,\dots,p} \\ &= \begin{bmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2} \\ \vdots & \vdots & & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{bmatrix}\end{aligned}$$

- 只要 $\Gamma_{p,q}$ 可逆则可解出 $a_1, \dots, a_p$ 。
- 解出 $a_1, \dots, a_p$ 后令

$$Y_t = A(\mathcal{B})X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}$$

- 则 $\{Y_t\}$ 是一个MA( $q$ )序列, 其自协方差函数为 $q$ 步截尾, 且

$$\begin{aligned}\gamma_y(k) &= E(Y_t Y_{t-k}) \\ &= \sum_{j=0}^p \sum_{l=0}^p \phi_j \phi_l E(X_{t-j} X_{t-l}) \\ &= \sum_{j=0}^p \sum_{l=0}^p \phi_j \phi_l \gamma_{k+l-j}, \quad 0 \leq k \leq q\end{aligned}$$

- 可以用 § 3.1 的方法唯一解出 $b_1, \dots, b_q, \sigma^2$ 。
- 于是, 只要 $\Gamma_{p,q}$ 可逆, 则ARMA( $p, q$ )序列的自协方差函数和ARMA( $p, q$ )模型型的参数 $(\mathbf{a}_p^T, \mathbf{b}_q^T, \sigma^2)$  相互惟一决定。

#### ARMA模型中AR部分的参数求解

- 如果 $\Gamma_{p,q}$ 可逆则由(2.15)可以解出 $a_1, \dots, a_p$ 。
- **定理2.3** 设 $\{\gamma_k\}$ 为ARMA( $p, q$ )序列 $\{X_t\}$ 的自协方差函数列, 则 $m \geq p$ 时 $\Gamma_{m,q}$ 可逆。

**证明:** 用反证法然后由引理2.2导出矛盾。

设 $\Gamma_{m,q}$  ( $m \times m$  矩阵) 不满秩, 则存在 $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{m-1})^T \neq 0$  使得 $\Gamma_{m,q}\boldsymbol{\beta} = 0$ , 即

$$\sum_{l=0}^{m-1} \beta_l \gamma_{q+k-l} = 0, \quad k = 0, 1, \dots, m-1 \quad (2.18)$$

注意当  $k \geq m$  时  $q + k - l > q$ , 所以这时  $\gamma_{q+k-l} = \sum_{j=1}^p a_j \gamma_{q+k-l-j}$ , 所以取  $k = m$  有

$$\begin{aligned} \sum_{l=0}^{m-1} \beta_l \gamma_{q+k-l} &= \sum_{l=0}^{m-1} \beta_l \sum_{j=1}^p a_j \gamma_{q+k-l-j} \\ &= \sum_{j=1}^p a_j \sum_{l=0}^{m-1} \beta_l \gamma_{q+(k-j)-l} \\ &= 0 \quad (\text{由(2.18)及 } 0 \leq k-j = m-j \leq m-1) \end{aligned}$$

递推得上式当  $k > m$  时也成立。因此

$$\sum_{l=0}^{m-1} \beta_l \gamma_{q+k-l} = 0, \quad k \geq m.$$

令  $Y_t = \sum_{l=0}^{m-1} \beta_l X_{t-l}$  则  $\{Y_t\}$  是零均值平稳列, 利用

$$E(Y_t X_{t-q-k}) = \sum_{l=0}^{m-1} \beta_l \gamma_{q+k-l} = 0, \quad k \geq 0$$

可知  $\{Y_t\}$  的自协方差  $q-1$  步截尾,  $\{Y_t\}$  是  $MA(q-1)$  序列, 存在  $\{\eta_t\} \sim WN(0, s^2)$  使得

$$\sum_{l=0}^{m-1} \beta_l X_{t-l} = \sum_{j=0}^{q-1} d_j \eta_{t-j}$$

与引理2.2矛盾。 □

### ARMA模型的一个充分条件

- **定理2.4** 设零均值平稳序列  $\{X_t\}$  有自协方差函数  $\{\gamma_k\}$ . 又设实数  $a_1, a_2, \dots, a_p$  ( $a_p \neq 0$ ) 使得  $A(z) = 1 - \sum_{j=1}^p a_j z^j$  满足最小相位条件, 另外

$$\gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = \begin{cases} c \neq 0, & k = q, \\ 0, & k > q, \end{cases} \quad (2.19)$$

则  $\{X_t\}$  是一个  $ARMA(p', q')$  序列, 其中  $p' \leq p, q' \leq q$ .

**证明:** 设  $Y_t = A(\mathcal{B})X_t = X_t - \sum_{j=1}^p a_j X_{t-j}$ . 则  $\{Y_t\}$  是零均值平稳序列, 满足

$$E(Y_t X_{t-k}) = \gamma_k - \sum_{j=1}^p a_j \gamma_{k-j} = \begin{cases} c \neq 0, & k = q, \\ 0, & k > q. \end{cases}$$

所以有

$$\begin{aligned}\gamma_y(k) &= E(Y_t Y_{t-k}) = E\left[Y_t \left(X_{t-k} - \sum_{j=1}^p a_j X_{t-k-j}\right)\right] \\ &= \begin{cases} c \neq 0, & k = q, \\ 0, & k > q. \end{cases}\end{aligned}$$

说明 $\{Y_t\}$ 的自协方差函数是 $q$ 后截尾的.

由定理1.3知道,  $\{Y_t\}$ 为一个MA( $q$ )序列, 即存在单位圆内没有根的 $q$ 阶实系数多项式 $B(z)$ 使得 $B(0) = b_0 = 1$  和

$$A(\mathcal{B})X_t = Y_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}, \quad (2.20)$$

其中 $\{\varepsilon_t\}$ 是WN( $0, \sigma^2$ ).

如果 $A(z)$ 和 $B(z)$ 没有公因子, 上述模型就是所需要的ARMA( $p, q$ )模型. 否则设公因子是 $C(z)$ , 则有 $A(z) = C(z)A'(z)$ ,  $B(z) = C(z)B'(z)$ . 这时(2.20)变成

$$C(\mathcal{B})A'(\mathcal{B})X_t = C(\mathcal{B})B'(\mathcal{B})\varepsilon_t.$$

两边乘 $C^{-1}(\mathcal{B})$ (显然 $C(z)$ 也满足最小相位条件) 后得到所需ARMA( $p', q'$ )模型:  
 $A'(\mathcal{B})X_t = B'(\mathcal{B})\varepsilon_t$ .  $\square$

### §3.2.4 ARMA序列的谱密度和可逆性

有理谱密度

- 由于ARMA序列的 $\{\gamma_k\}$ 绝对可和, 以及平稳解的线性序列表达式, 可得ARMA( $p, q$ )序列(2.6)有谱密度

$$\begin{aligned}f(\lambda) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\lambda} \\ &= \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \psi_j e^{ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \left| \frac{B(e^{i\lambda})}{A(e^{i\lambda})} \right|^2\end{aligned} \quad (2.21)$$

形如(2.21)的谱密度被称为有理谱密度.

可逆的ARMA模型

- **定义2.2** 在ARMA( $p, q$ )模型的定义2.1 中, 如果进一步要求 $B(z)$ 在单位圆上无根:

$$B(z) = 1 + \sum_{j=1}^q b_j z^j \neq 0, |z| \leq 1 \quad (2.22)$$

则称ARMA( $p, q$ )模型(2.2)为可逆的ARMA模型, 称相应的平稳解为可逆的ARMA( $p, q$ )序列.

- 从定理1.4(最小序列的谱条件)知道可逆的ARMA( $p, q$ )序列是最小序列.
- 对于可逆的ARMA( $p, q$ )模型(2.3) 由于 $B^{-1}(z)A(z)$ 在 $\{z : |z| \leq \rho\}$  ( $\rho > 1$ )内解析, 所以有Taylor展式:

$$B^{-1}(z)A(z) = \sum_{j=0}^{\infty} \varphi_j z^j, \quad |z| \leq \rho, \quad (2.23)$$

其中 $|\varphi_j| = o(\rho^{-j})$ , 当 $j \rightarrow \infty$ , 从而可以定义 $B^{-1}(\mathcal{B})A(\mathcal{B}) = \sum_{j=0}^{\infty} \varphi_j \mathcal{B}^j$ . 在(2.3)两边乘以 $B^{-1}(\mathcal{B})$ , 得到:

$$\varepsilon_t = B^{-1}(\mathcal{B})A(\mathcal{B})X_t = \sum_{j=0}^{\infty} \varphi_j X_{t-j}, \quad t \in \mathbb{Z}. \quad (2.24)$$

(2.24)是(2.6) 的逆转形式, 表明可逆ARMA( $p, q$ )序列和它的噪声序列可以相互线性表示.

### §3.2.5 例子

#### ARMA例2.1

- ARMA(4,2):

$$\begin{aligned} a_1 &= -0.9, & a_2 &= -1.4, \\ a_3 &= -0.7, & a_4 &= -0.6; \\ b_1 &= 0.5, & b_2 &= -0.4. \end{aligned} \quad (2.25)$$

- $A(z)$ 的根为 $1.1380e^{\pm 2.2062i}$ ,  $1.1344e^{\pm 1.4896i}$ ,  $B(z)$ 的两个实根为 $2.3252, -1.0752$ .
- 此时间序列有两个频率成分。(演示)

#### ARMA例2.2

- 设 $\{X_t\} \sim \text{ARMA}(2,2)$ , 已知

$$(\gamma_0, \gamma_1, \dots, \gamma_4) = (4.61, -1.06, 0.29, 0.69, -0.12)$$

- 要反解ARMA参数。
- 见演示。



### §3.3 广义ARMA模型和ARIMA(p, d, q)模型介绍

#### §3.3.1 广义ARMA模型

##### 广义ARMA模型

- 设  $A(z) = 1 - \sum_{j=1}^p a_j z^j$ ,  $B(z) = 1 + \sum_{j=1}^q b_j z^j$  是两个没有公共根的实系数多项式,  $a_p b_q \neq 0$ ,  $\{\varepsilon_t\}$  是  $WN(0, \sigma^2)$ . 如果不对  $A(z)$ ,  $B(z)$  的根做任何其他限制, 则称差分方程:

$$A(\mathcal{B})X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}, \quad (3.1)$$

为广义ARMA(p, q)模型, 称满足(3.1)的  $\{X_t\}$  为广义ARMA(p, q)序列.

- 如果  $A(z)$  在单位圆上有根, 可以证明(3.1)没有平稳解.
- 如果  $A(z)$  在单位圆上没有根, 则有  $0 < \rho_1 < 1 < \rho_2$  使得复变函数  $B(z)/A(z)$  在圆环

$$D = \{z : \rho_1 \leq |z| \leq \rho_2\} \quad (3.2)$$

内解析.

- 于是  $B(z)/A(z)$  有Laurent级数展开

$$A^{-1}(z)B(z) = \sum_{j=-\infty}^{\infty} c_j z^j, \quad z \in D, \quad (3.3)$$

- 存在  $\rho > 1$  使  $c_j = o(\rho^{-|j|})$ ,  $j \rightarrow \pm\infty$ .
- 这样, 从(3.1)可以得到惟一的平稳解

$$X_t = A^{-1}(\mathcal{B})B(\mathcal{B})\varepsilon_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \quad (3.4)$$

- 如果  $A(z)$  在单位圆内有根, 由(3.4)定义的平稳序列是白噪声的双边无穷滑动和。
- 这个平稳序列不是合理的, 因为  $t$  时的观测受到了  $t$  以后的干扰的影响.
- 再由差分方程的理论知道, 这时(2.2)的其他解都随着时间的增加而加速振荡. 为此, 人们把这时的广义ARMA模型称为**爆炸模型**.

### §3.3.2 求和ARIMA( $p, d, q$ )模型

#### 求和ARIMA( $p, d, q$ )模型

- ARIMA( $p, d, q$ )模型是AR部分有单位特征根(即1)的广义ARMA模型,  $d$ 为单位特征根的重数。
- 除了单位根以外, 要求AR部分根都在单位圆外, MA部分单位圆内没有根。
- ARIMA( $p, d, q$ )序列是 $d$ 阶差分后服从ARMA( $p, q$ )模型的非平稳时间序列。
- 设 $d$ 是一个正整数, 如果

$$Y_t = (1 - \mathcal{B})^d X_t = \sum_{k=0}^d C_d^k (-1)^k X_{t-k}, \quad t \in \mathbb{Z}, \quad (3.5)$$

是一个ARMA( $p, q$ )序列, 则称 $\{X_t\}$ 是一个求和ARIMA( $p, d, q$ )序列. 简称为ARIMA( $p, d, q$ )序列, 其中 $C_d^k$ 是二项式系数.

- 设 $\{X_t\}$ 为ARIMA( $p, d, q$ )序列,  $(1 - \mathcal{B})^d X_t = Y_t$ , 模型通解为

$$X_t = C_0 + C_1 t + \cdots + C_{d-1} t^{d-1} + \sum_{n_{d-1}=1}^t \cdots \sum_{n_1=1}^{n_2} \sum_{j=1}^{n_1} Y_j.$$

- 实际问题中有许多数据经过一或两次差分后会稳定下来. 差分运算是对数据进行预处理的常用方法之一.

### §3.3.3 单位根过程

#### 单位根过程

- ARIMA( $p, 1, q$ )模型称为单位根过程, 相应的时间序列被称为单位根序列。
- 单位根过程与有一个AR部分特征根 $|z_j| > 1$ 但 $|z_j|$ 十分接近于1的平稳ARMA序列很难区分。
- 单位根过程与带有线性趋势的模型不同. 单位根过程数据没有固定走势。

### §3.3.4 分数差分ARFIMA( $p, d, q$ )模型

#### 长记忆和短记忆

- ARMA序列自协方差函数负指数衰减, 是短记忆的。
- 离散谱序列自协方差含不衰减到0, 是长记忆的。

- 其它的平稳序列如何区分长记忆还是短记忆? 若存在  $d < \frac{1}{2}$ , 使得

$$\gamma_k \sim k^{2d-1}, \quad k \rightarrow \infty$$

则称相应的序列为长记忆序列。

- 即  $\gamma_k \sim \frac{1}{k^\alpha}, \alpha > 0$ ,  $\gamma_k$  以幂函数速度趋于零。

### 长记忆序列

- 对于  $d \neq 0$ ,  $d \in (-0.5, 0.5)$ ,  $(1-z)^{-d}$  有Taylor展开公式

$$(1-z)^{-d} = \sum_{j=0}^{\infty} \pi_j z^j, \quad |z| \leq 1, \quad (3.13)$$

- $\{\pi_j\}$  平方可和。
- 定义线性平稳序列

$$X_t = (1 - \mathcal{B})^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \quad (3.16)$$

- 则  $\{X_t\}$  是模型

$$(1 - \mathcal{B})^d X_t = \varepsilon_t, \quad t \in \mathbb{Z}, \quad (3.17)$$

的惟一平稳解. 人们称(3.17)是ARFIMA(0, d, 0)模型.

- 类似可定义ARFIMA(p, d, q)模型。
- 对ARFIMA(0, d, 0)序列

$$\gamma_k \sim k^{2d-1} \frac{\sigma^2 \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}, \quad \text{当 } k \rightarrow \infty. \quad (3.20)$$

即长记忆。

- 当  $d \in (-0.5, 0)$  时

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty.$$

- 当  $d \in (0, 0.5)$  时

$$\sum_{k=0}^{\infty} |\gamma_k| = \infty.$$

- ARFIMA(0, d, 0)的偏相关函数满足

$$a_{k,k} = \frac{d}{k-d}, \quad k = 1, 2, \dots$$

可以用于d的估计。

- 一般ARFIMA(p, d, q)的讨论略过。



## 第四章 均值和自协方差函数的估计

### §4.1 均值的估计

#### 均值、自协方差估计的作用

- AR, MA, ARMA模型的参数可以由自协方差函数唯一确定。
- 有了样本之后, 可以先估计均值和自协方差函数。
- 然后由均值和自协方差函数解出模型参数。
- 均值和自协方差可以用矩估计法求。
- 还要考虑相合性、渐近分布、收敛速度等问题。

#### 均值估计公式

- 设  $x_1, x_2, \dots, x_N$  是平稳列  $\{X_t\}$  的观测。
- $\mu = EX_t$  的点估计为

$$\hat{\mu} = \bar{x}_N = \frac{1}{N} \sum_{k=1}^N x_k$$

- 把观测样本看成随机样本时记做大写的  $X_1, X_2, \dots, X_N$ 。

#### §4.1.1 相合性

##### 相合性

- 设统计量  $\hat{\theta}_N$  是  $\theta$  的估计. 在统计学中有如下的定义
  1. 如果  $E\hat{\theta}_N = \theta$ , 则称  $\hat{\theta}_N$  是  $\theta$  的 **无偏估计**.
  2. 如果当  $N \rightarrow \infty$ ,  $E\hat{\theta}_N \rightarrow \theta$ , 则称  $\hat{\theta}_N$  是  $\theta$  的 **渐近无偏估计**.
  3. 如果  $\hat{\theta}_N$  依概率收敛到  $\theta$ , 则称  $\hat{\theta}_N$  是  $\theta$  的 **相合估计**.
  4. 如果  $\hat{\theta}_N$  a.s.收敛到  $\theta$ , 则称  $\hat{\theta}_N$  是  $\theta$  的 **强相合估计**.
- 一般情况下, 无偏估计比有偏估计来得好. 对于由(1.1)定义的  $\bar{X}_N$ , 有

$$E\bar{X}_N = \frac{1}{N} \sum_{k=1}^N EX_k = \frac{1}{N} \sum_{k=1}^N \mu = \mu.$$

所以,  $\bar{X}_N$  是均值  $\mu$  的无偏估计.

## 均值估计的相合性

- 好的估计量起码应当是相合的. 否则, 估计量不收敛到要估计的参数, 它无助于实际问题的解决.
- 对于平稳序列 $\{X_t\}$ , 如果它的自协方差函数 $\{\gamma_k\}$ 收敛到零, 则

$$\begin{aligned}
 E(\bar{X}_N - \mu)^2 &= E\left[\frac{1}{N} \sum_{k=1}^N (X_k - \mu)\right]^2 \\
 &= \frac{1}{N^2} E\left[\sum_{j=1}^N \sum_{k=1}^N (X_k - \mu)(X_j - \mu)\right] = \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \gamma_{k-j} \\
 &= \frac{1}{N^2} \sum_{j=1}^N \sum_{m=1-j}^{N-j} \gamma_m \quad (\text{令 } m = k - j) \\
 &= \frac{1}{N^2} \sum_{m=-N+1}^{N-1} \sum_{j=\max(1-m, 1)}^{\min(N-m, N)} \gamma_m = \frac{1}{N^2} \sum_{m=-N+1}^{N-1} (N - |m|) \gamma_m \\
 &\leq \frac{1}{N} \sum_{m=-N}^N |\gamma_m| \rightarrow 0, \quad \text{当 } N \rightarrow \infty.
 \end{aligned} \tag{1.2}$$

即 $\bar{X}_N$  均方收敛到 $\mu$ .

利用切比雪夫不等式

$$\Pr(|\bar{X}_N - \mu| \geq \delta) \leq \frac{E(\bar{X}_N - \mu)^2}{\delta^2} \rightarrow 0, \quad (\delta > 0)$$

得到 $\bar{X}_N$  依概率收敛到 $\mu$ . 于是 $\bar{X}_N$ 是 $\mu$ 的相合估计.

## 均值估计性质

- **定理1.1** 设平稳序列 $\{X_t\}$ 有均值 $\mu$ 和自协方差函数 $\{\gamma_k\}$ , 则
  1.  $\bar{X}_N$ 是 $\mu$  的无偏估计.
  2. 如果 $\gamma_k \rightarrow 0$ , 则 $\bar{X}_N$  是 $\mu$ 的相合估计.
  3. 如果 $\{X_t\}$ 还是严平稳遍历序列, 则 $\bar{X}_N$ 是 $\mu$ 的强相合估计.
- 第三条结论利用§1.5 的遍历定理5.1可得.
- 一般地, 任何强相合估计一定是相合估计.
- 线性平稳列的均值估计是相合估计. (第一章定理3.3)
- ARMA模型的均值估计是相合估计.

## §4.1.2 中心极限定理

## 独立同分布样本的中心极限定理

- 若  $X_1, X_2, \dots, X_N$  iid  $\sim (\mu, \sigma^2)$ , 则  $\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} N(0, \sigma^2)$ .
- 可以据此计算  $\mu$  的95%置信区间。

$$[\bar{X}_N - 1.96\sigma/\sqrt{N}, \bar{X}_N + 1.96\sigma/\sqrt{N}]. \quad (1.3)$$

其中的1.96也经常用2近似代替。

## 平稳列的均值估计的中心极限定理

- **定理1.2** 设  $\{\varepsilon_t\}$  是独立同分布的  $WN(0, \sigma^2)$ . 线性平稳序列  $\{X_t\}$  由

$$X_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k}, \quad t \in \mathbb{Z}, \quad (1.5)$$

定义, 其中  $\{\psi_k\}$  平方可和. 如果  $\{X_t\}$  的谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=-\infty}^{\infty} \psi_k e^{-ik\lambda} \right|^2 \quad (1.6)$$

在  $\lambda = 0$  连续, 并且  $f(0) \neq 0$ , 则当  $N \rightarrow \infty$  时,

$$\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} N(0, 2\pi f(0))$$

## 推论

- 当  $\{\psi_k\}$  绝对可和时,  $f(\lambda)$  连续.
- **推论1.3** 如果  $\sum_{k=-\infty}^{\infty} |\psi_k| < \infty$  和  $\sum_{k=-\infty}^{\infty} \psi_k \neq 0$  成立, 则当  $N \rightarrow \infty$  时

$$\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} N(0, 2\pi f(0))$$

并且

$$2\pi f(0) = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j. \quad (1.7)$$

- 中心极限定理成立时可以构造  $\mu$  的渐近置信区间或对  $\mu$  作假设检验。



## §4.1.3 收敛速度

## 收敛速度

- 相合的估计量的渐近性质除了是否服从中心极限定理外, 还包括这个估计量的收敛速度.
- 收敛速度的描述方法之一是所谓的重对数律.
- 重对数律成立时, 得到的收敛速度的阶数一般是

$$O\left(\sqrt{\frac{2\ln\ln N}{N}}\right).$$

- 除了个别的情况, 这个阶数一般不能再被改进.
- **定理1.4** 设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$ , 线性平稳序列 $\{X_t\}$ 由(1.5)定义, 谱密度 $f(0) \neq 0$ . 当以下的条件之一成立时:

1. 当 $k \rightarrow \pm\infty$ ,  $\psi_{|k|}$  以负指数阶收敛于0;
2. 谱密度 $f(\lambda)$ 在 $\lambda = 0$ 连续, 并且 $E|\varepsilon_t|^r < \infty$  对某个 $r > 2$  成立,

则有重对数律

$$\limsup_{N \rightarrow \infty} \sqrt{\frac{N}{2\ln\ln N}} (\bar{X}_N - \mu) = \sqrt{2\pi f(0)}, \quad a.s. \quad (1.8)$$

$$\liminf_{N \rightarrow \infty} \sqrt{\frac{N}{2\ln\ln N}} (\bar{X}_N - \mu) = -\sqrt{2\pi f(0)}, \quad a.s. \quad (1.9)$$

- 易见重对数率满足时 $(\bar{X}_n - \mu) \cdot o(1) = o(\sqrt{\frac{\ln\ln N}{N}})$ ,  $\sqrt{\frac{N}{2\ln\ln N}} (\bar{X}_n - \mu) / o(1)$  不收敛。

§4.1.4  $\bar{X}_N$  的模拟计算

## AR(2)的均值计算

- 令

$$A(z) = (1 - \rho e^{i\theta} \cdot z)(1 - \rho e^{-i\theta} \cdot z)$$

考虑AR(2)模型

$$\begin{aligned} A(\mathcal{B})X_t &= \varepsilon_t \\ X_t &= 2\rho \cos \theta X_{t-1} - \rho^2 X_{t-2} + \varepsilon_t \end{aligned}$$

- 为模拟方便设 $\{\varepsilon_t\}$  iid  $\sim N(0, \sigma^2)$ 。

•

$$\bar{x}_N = \frac{1}{N} \sum_{t=1}^N x_t, \quad \bar{\varepsilon}_N = \frac{1}{N} \sum_{t=1}^N \varepsilon_t$$

- 模型方程两边求平均

$$\begin{aligned} \bar{x}_N &= 2\rho \cos \theta \frac{1}{N} \sum_{k=0}^{N-1} x_k - \rho^2 \frac{1}{N} \sum_{k=-1}^{N-2} x_k + \bar{\varepsilon}_N \\ &= 2\rho \cos \theta \bar{X}_N - \rho^2 \bar{X}_N + \bar{\varepsilon}_N \\ &\quad + 2\rho \cos \theta \frac{1}{N} (x_0 - x_N) - \rho^2 \frac{1}{N} (x_{-1} + x_0 - x_{N-1} - x_{N-2}) \\ &\approx 2\rho \cos \theta \bar{X}_N - \rho^2 \bar{X}_N + \bar{\varepsilon}_N \\ \bar{x}_N &\approx \frac{1}{A(1)} \bar{\varepsilon}_N = \frac{1}{1 - 2\rho \cos \theta + \rho^2} \bar{\varepsilon}_N \end{aligned}$$

- $N$ 较大时  $\bar{x}_N$  和  $\bar{\varepsilon}_N$  成正比。

#### 估计收敛性的模拟

- 为了观察  $N \rightarrow \infty$  时  $\bar{x}_N$  的收敛可以模拟  $L$  个值然后观察  $\bar{x}_N, N = n_0, n_0 + 1, \dots, L$  的变化。
- 为了研究固定  $N$  情况下  $\bar{X}_N$  的精度以至于抽样分布, 可以进行  $M$  次独立的随机模拟, 得到  $M$  个  $\bar{X}_N$  的观测值。这种方法对于难以得到估计量的理论分布的情况是很有用的。

## §4.2 自协方差函数的估计

### §4.2.1 自协方差估计公式及正定性

自协方差函数估计公式

•

$$\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} (x_j - \bar{x}_N)(x_{j+k} - \bar{x}_N), \quad 0 \leq k \leq N-1, \quad (2.2)$$

$$\hat{\gamma}_{-k} = \hat{\gamma}_k$$

- 样本自相关系数(ACF)估计为

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \quad |k| \leq N-1 \quad (2.3)$$

- $k$ 较大时参与平均的项减少所以 $\hat{\gamma}_k$ 估计误差会随 $k$ 增大而变大。
- 估计 $\gamma_k$ 一般不使用除以 $N-k$ 的估计形式:

$$\hat{\gamma}_k = \frac{1}{N-k} \sum_{j=1}^{N-k} (x_j - \bar{x}_N)(x_{j+k} - \bar{x}_N) \quad (2.4)$$

因为:

- 我们不对大的 $k$ 计算 $\hat{\gamma}_k$ ;
- 更重要的是只有除以 $N$ 的估计式才是正定的。

样本自协方差的正定性

- 只要观测 $x_1, x_2, \dots, x_N$ 不全相同则

$$\hat{\Gamma}_N = (\hat{\gamma}_{k-j})_{k,j=1,2,\dots,N}$$

正定。

- 令 $y_j = x_j - \bar{x}_N$ , 记

$$A = \begin{pmatrix} 0 & \cdots & 0 & y_1 & y_2 & \cdots & y_{N-1} & y_N \\ 0 & \cdots & y_1 & y_2 & y_3 & \cdots & y_N & 0 \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ y_1 & \cdots & y_{N-1} & y_N & 0 & \cdots & 0 & 0 \end{pmatrix} \quad (2.5)$$

则

$$\hat{\Gamma}_N = \frac{1}{N} A A^T$$

- 只要 $y_j$ 不全是零则 $A$ 满秩。
- 事实上, 设 $y_1 = \cdots = y_{k-1} = 0, y_k \neq 0$ , 则 $A$ 矩阵左面会出现一个以 $y_k$ 值开始非零的斜面, 显然是满秩的。
- 故 $x_1, \dots, x_N$ 不全相同时 $\hat{\Gamma}_N$ 正定。
- $\hat{\Gamma}_n (1 \leq n \leq N)$ 作为 $\hat{\Gamma}_N$ 的主子式也是正定的。

### §4.2.2 $\hat{\gamma}_k$ 的相合性

#### $\hat{\gamma}_k$ 的相合性

- **定理2.1** 设平稳序列的样本自协方差函数 $\hat{\gamma}_k$ 由(2.2)或(2.4)定义.
  1. 如果当 $k \rightarrow \infty$ 时 $\gamma_k \rightarrow 0$ , 则对每个确定的 $k$ ,  $\hat{\gamma}_k$ 是 $\gamma_k$ 的渐近无偏估计:
 
$$\lim_{N \rightarrow \infty} E\hat{\gamma}_k = \gamma_k.$$
  2. 如果 $\{X_t\}$ 是严平稳遍历序列, 则对每个确定的 $k$ ,  $\hat{\gamma}_k$ 和 $\hat{\rho}_k$ 分别是 $\gamma_k$ 和 $\rho_k$ 的强相合估计:

$$\lim_{N \rightarrow \infty} \hat{\gamma}_k = \gamma_k, a.s., \quad \lim_{N \rightarrow \infty} \hat{\rho}_k = \rho_k, a.s..$$

#### 定理2.1证明

下面只对由(2.2)定义的样本自协方差函数证明定理2.1, 对由(2.4)定义的 $\hat{\gamma}_k$ 的证明是一样的.

设 $\mu = EX_1$ , 则 $\{Y_t\} = \{X_t - \mu\}$ 是零均值的平稳序列. 利用

$$\bar{Y}_N = \frac{1}{N} \sum_{j=1}^N Y_j = \bar{X}_N - \mu$$

得到

$$\begin{aligned} \hat{\gamma}_k &= \frac{1}{N} \sum_{j=1}^{N-k} (Y_j - \bar{Y}_N)(Y_{j+k} - \bar{Y}_N) \\ &= \frac{1}{N} \sum_{j=1}^{N-k} [Y_j Y_{j+k} - \bar{Y}_N(Y_{j+k} + Y_j) + \bar{Y}_N^2]. \end{aligned} \quad (2.7)$$

利用(1.2)得到 $E\bar{Y}_N^2 = E(\bar{X}_N - \mu)^2 \rightarrow 0$ . 利用Schwarz不等式得到

$$E|\bar{Y}_N(Y_{j+k} + Y_j)| \leq [E\bar{Y}_N^2 E(Y_{j+k} + Y_j)^2]^{1/2} \leq [4E\bar{Y}_N^2 \gamma_0]^{1/2} \rightarrow 0.$$

所以当  $N \rightarrow \infty$ ,

$$E\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} E(Y_j Y_{j+k}) + o(1) = \frac{N-k}{N} \gamma_k + o(1) \rightarrow \gamma_k.$$

强相合性的证明: 用遍历定理得到

$$\begin{aligned} \bar{Y}_N &\rightarrow EY_1 = 0, \text{ a.s.}, \\ \frac{1}{N} \sum_{j=1}^{N-k} (Y_{j+k} + Y_j) &= \frac{1}{N} \left( \sum_{j=1}^N Y_j - \sum_{j=1}^k Y_j + \sum_{j=1}^{N-k} Y_j \right) \\ &= \bar{Y}_N - \frac{1}{N} \sum_{j=1}^k Y_j + \frac{N-k}{N} \bar{Y}_{N-k} \rightarrow 0, \text{ a.s.}, \end{aligned}$$

于是, 从(2.7)式 可知

$$\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} Y_j Y_{j+k} + o(1) \rightarrow E(Y_1 Y_{1+k}) = \gamma_k, \text{ a.s.}$$

### §4.2.3 $\hat{\gamma}_k$ 的渐近分布

$\hat{\gamma}_k$  的渐近分布——线性平稳列

- 只考虑线性序列。
- 设  $\{\varepsilon_t\}$  是 4 阶矩有限的独立同分布的  $WN(0, \sigma^2)$  ( $\sigma^2 > 0$ ), 实数列  $\{\psi_k\}$  平方可和。
- 线性平稳序列

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (2.8)$$

- $\{X_t\}$  有自协方差函数

$$\gamma_k = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} \quad (2.9)$$

- $\{X_t\}$  有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{ij\lambda} \right|^2. \quad (2.10)$$

- 设自协方差函数列  $\{\gamma_k\}$  平方可和。

- 设 $\{W_t\}$ 为独立同分布 $N(0,1)$ 。

- 令

$$\mu_4 = E\varepsilon_1^4, \quad M_0 = \frac{1}{\sigma^2}(\mu_4 - \sigma^4)^{1/2} > 0$$

- 定义正态时间序列

$$\xi_j = (M_0 \gamma_j) W_0 + \sum_{t=1}^{\infty} (\gamma_{t+j} + \gamma_{t-j}) W_t, \quad j \geq 0 \quad (2.11)$$

$$R_j = \sum_{t=1}^{\infty} (\rho_{t+j} + \rho_{t-j} - 2\rho_t \rho_j) W_t, \quad j \geq 1, \quad (2.12)$$

### 样本自协方差和自相关的中心极限定理

- **定理2.2** 设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$ , 满足 $\mu_4 = E\varepsilon_1^4 < \infty$ . 如果线性平稳序列(2.8)的谱密度(2.10)平方可积:

$$\int_{-\pi}^{\pi} f(\lambda)^2 d\lambda < \infty,$$

则对任何正整数 $h$ , 当 $N \rightarrow \infty$ 时, 有以下结果

1.  $\sqrt{N}(\hat{\gamma}_0 - \gamma_0, \hat{\gamma}_1 - \gamma_1, \dots, \hat{\gamma}_h - \gamma_h)$  依分布收敛到 $(\xi_0, \xi_1, \dots, \xi_h)$ .
  2.  $\sqrt{N}(\hat{\rho}_1 - \rho_1, \hat{\rho}_2 - \rho_2, \dots, \hat{\rho}_h - \rho_h)$  依分布收敛到 $(R_1, R_2, \dots, R_h)$ .
- 可以据此构造 $\gamma_k$ 和 $\rho_k$ 的近似的区间估计和近似的假设检验。

### 自相关检验的例子

- **例2.1**(接第三章例1.1) 对 $MA(q)$ 序列 $\{X_t\}$ , 利用定理2.2 得到, 只要当 $m > q$ :  $\sqrt{N}\hat{\rho}_m$ 依分布收敛到 $R_m$ 的分布。

- 

$$R_m = \sum_{t=1}^{\infty} (\rho_{t+m} + \rho_{t-m} - 2\rho_t \rho_m) W_t, \quad m \geq q+1$$

注意 $m \geq q+1$ 时 $\rho_m = 0, \rho_{t+m} = 0$ ,  $\rho_{t-m}$ 中的 $t-m$ 应属于 $[-q, q]$ , 所以令 $l = t-m$ 有

$$R_m = \sum_{l=-q}^q \rho_l W_{l+m}$$

- $R_m$ 为期望为0, 方差为 $1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_q^2$ 的正态分布。

- 在假设  $H_0$ :  $\{X_t\}$  是  $MA(q)$  下, 对  $m > q$  有

$$\Pr \left( \frac{\sqrt{N}|\hat{\rho}_m|}{\sqrt{1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_q^2}} \geq 1.96 \right) \approx 0.05.$$

- 现在用  $\{X_t\}$  表示第三章例1.1中差分后的化学浓度数据, 在  $H_0$ :  $\{X_t\}$  是  $MA(q)$  下, 用  $\hat{\rho}_k$  代替真值  $\rho_k$  后分别对  $q = 0, 1$  计算出

$$T_q(m) = \frac{\sqrt{N}\hat{\rho}_{m+q}}{\sqrt{1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + \cdots + 2\hat{\rho}_q^2}}, \quad m = 1, \dots, 6.$$

$m =$	1	2	3	4	5	6
$q = 0$	-5.778	0.281	-0.951	-0.121	-1.071	-0.116
$q = 1$	0.243	-0.821	-0.104	-0.925	-0.100	1.631

- 在  $q = 0$  的假设下,  $|T_0(1)| = 5.778 > 1.96$ , 所以应当否定  $q = 0$ .
- 实际工作中人们还计算概率

$$p = P(|\sqrt{N}\hat{\rho}_1| \geq | - 5.778 |),$$

并且把  $p$  称为检验的  $p$  值. 明显  $p$  值越小, 数据提供的否定原假设的依据越充分. 现在在  $H_0$  下,  $\sqrt{N}\hat{\rho}_1$  近似服从标准正态分布, 所以  $p$  值几乎是零, 因而必须拒绝  $\{X_t\}$  是  $MA(0)$  的假设.

- 取  $q = 1$  时  $|T_1(m)| < 1.96 (1 \leq m \leq 6)$ , 所以不能拒绝  $\{X_t\}$  是  $MA(1)$  的假设.

### 谱密度平方可积的充要条件

- 对于实际工作者来讲谱密度平方可积的条件通常很难验证, 于是希望能把定理2.2中谱密度平方可积的条件改加在自协方差函数  $\{\gamma_k\}$  的收敛速度上.
- 定理2.3 对任一平稳序列  $\{X_t\}$ , 它的自协方差函数平方可和的充分必要条件是它的谱密度平方可积.
- 这个结论主要是利用实变函数论中Fourier级数的理论, 只有证明  $f(\lambda) \geq 0$  时用了周期图(如P.67定理3.1的证明, 那里  $\{\gamma_k\}$  绝对可和)。证明略。
- 推论2.4 设  $\{\varepsilon_t\}$  是独立同分布的白噪声  $WN(0, \sigma^2)$ , 满足  $\mu_4 = E\varepsilon_t^4 < \infty$ . 如果线性平稳序列(2.8)的自协方差函数平方可和:  $\sum_k \gamma_k^2 < \infty$ , 则定理2.2中的结论成立.

$\psi_k$ 快速收敛条件下的中心极限定理

- 定理2.2 要求白噪声的方差有4阶矩. 下面关于线性平稳序列的样本自相关系数的中心极限定理不要求噪声项的4阶矩有限.
- **定理2.5**([26]) 设 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$ , 线性平稳序列 $\{X_t\}$ 由(2.8)定义. 如果自协方差函数 $\{\gamma_k\}$ 平方可和, 并且对某个常数 $\alpha > 0.5$ ,

$$m^\alpha \sum_{|k| \geq m} \psi_k^2 \rightarrow 0, \quad m \rightarrow \infty, \quad (2.13)$$

则对任何正整数 $h$ , 当 $N \rightarrow \infty$ 时

$$\sqrt{N}(\hat{\rho}_1 - \rho_1, \hat{\rho}_2 - \rho_2, \dots, \hat{\rho}_h - \rho_h)$$

依分布收敛到

$$(R_1, R_2, \dots, R_h).$$

- ARMA序列的 $\{\psi_j\}$ 满足(2.13)所以ARMA序列的白噪声列是独立同分布序列时定理2.5结论成立。

## 关于独立同分布列的中心极限定理

- **推论2.6.** 如果 $\{X_t\}$ 是独立同分布的白噪声,

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x}_N)(x_{t+k} - \bar{x}_N)}{\sum_{t=1}^N (x_t - \bar{x}_N)^2}$$

是样本自相关系数, 则对任何正整数 $h$

1.

$$\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_h)$$

依分布收敛到多元标准正态分布 $N(0, I_h)$ . 这里,  $I_h$ 是 $h \times h$ 的单位矩阵.

2. 如果 $\mu_4 = E\varepsilon_t^4 < \infty$ , 则

$$\sqrt{N}(\hat{\gamma}_0 - \sigma^2, \hat{\gamma}_1, \dots, \hat{\gamma}_h)$$

依分布收敛到

$$\sigma^2(M_0 W_0, W_1, \dots, W_h).$$



**推论2.6证明**

- 对白噪声,  $\gamma_0 = \sigma^2$ ,

$$R_j = \sum_{t=1}^{\infty} (\rho_{t+j} + \rho_{t-j} - 2\rho_t \rho_j) W_t = W_j, j \geq 1$$

$$\xi_j = (M_0 \gamma_j) W_0 + \sum_{t=1}^{\infty} (\gamma_{t+j} + \gamma_{t-j}) W_t$$

$$= \gamma_0 W_j = \sigma^2 W_j$$

$$\xi_0 = (M_0 \gamma_0) W_0 + \sum_{t=1}^{\infty} (\gamma_{t+0} + \gamma_{t-0}) W_t$$

$$= M_0 \gamma_0 W_0 = \sigma^2 M_0 W_0$$

- 定理2.5的条件满足。第二条满足推论2.4的条件。

**§4.2.4 模拟计算****AR(2)模型实例**

- 首先用图形表示 $N$ 不同时 $\hat{\gamma}_k$ 的误差。
- 然后重复 $M = 1000$ 次计算1000个 $\hat{\gamma}_k$ 的标准差(称为标准误差)。发现 $N$ 增大时标准误差减小。
- 误差随 $N$ 减小的速度为 $N^{-1/2}$ 。
- 根离单位圆近的模式其估计标准误差大。

### §4.3 白噪声检验

#### §4.3.1 白噪声的 $\chi^2$ 检验

##### 白噪声的 $\chi^2$ 检验

- 若 $\{X_t\}$ 是独立同分布的白噪声, 根据推论2.6,  $N$ 足够大时

$$\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)$$

服从iid标准正态分布。于是

•

$$\hat{\chi}^2(m) \triangleq N(\hat{\rho}_1^2 + \hat{\rho}_2^2 + \dots + \hat{\rho}_m^2)$$

近似服从 $\chi^2(m)$ 分布。

- 要检验 $H_0: \{X_t\}$ 为独立同分布白噪声, 可以用否定域

$$\{\hat{\chi}^2(m) > \lambda_\alpha(m)\}$$

其中 $\lambda_\alpha(m)$ 为 $\chi^2(m)$ 的右侧 $\alpha$ 分位数。

- 检验的 $p$ 值为

$$\Pr(\chi^2(m) > \hat{\chi}^2(m)) = 1 - F_m(\hat{\chi}^2(m))$$

其中 $F_m(\cdot)$ 为 $\chi^2(m)$ 分布函数。

- 虽然 $H_0$ 要求 $\{X_t\}$ 独立同分布, 此方法可以用作一般性的白噪声检验。

#### AR(2)模拟数据的检验

- 对AR(2)模型取不同根离单位圆距离试验。根离单位圆越近与白噪声差别越大。
- 对MA(1)模型用不同 $b$ 模拟。 $b$ 接近于1时与白噪声差别变明显。
- 关于 $\hat{\chi}^2(m)$ 中项数 $m$ 的选取:  $m = 5$ 比 $m = 20$ 有效。注意以ARMA模型为例, 当 $k$ 较大时 $\rho_k$ 已经很小, 所以 $\rho_k^2$ 贡献不大, 取太大的 $m$ 容易使检验不敏感。
- 演示

### §4.3.2 样本自相关置信区间检验法

#### 样本相关系数置信区间检验法

- 当 $\{X_t\}$ 为独立同分布白噪声时 $\sqrt{N}(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)$  近似 $m$ 维标准正态分布。
- $\Pr(\sqrt{N}|\hat{\rho}_k| > 1.96) \approx 0.05$ .
- 如果超过5%的 $|\hat{\rho}_j| \geq 1.96/\sqrt{n}$  可否定 $H_0 : \{X_t\}$ 为独立同分布白噪声。
- 与 $\chi^2$ 检验理由类似,  $m$ 不应取太大。

#### 样本自相关置信区间检验法: 例子

- 对WN(0, 1)的检验可以比较成功。
- 对MA(1)的检验如果取 $m = 20$ 则很可能不成功, 因为一般只有 $\hat{\rho}_1$ 超过界限。
- 对AR的检验一般成功, 因为其相关系数不截尾。
- 演示



## 第五章 时间序列的预报

## §5.1 最佳线性预测的基本性质

### 预报问题

- 对于时间序列进行统计分析的主要目的之一是解决时间序列的预测问题.
- 任何时间序列 $\{X_t\}$ 都可以按第一章的(1.4)式分解成趋势项 $\{T_t\}$ 、季节项 $\{S_t\}$ 和随机项 $\{R_t\}$ 的和.
- 趋势项和季节项都可以被当做非随机的时间序列处理, 他们的预测问题往往是简单的.
- 随机项 $\{R_t\}$ 一般是平稳序列. 于是, 时间序列预测问题的重点应当是平稳序列.
- 本章主要讨论平稳序列的预测问题.
- 平稳序列的方差有限, 所以我们总是假设本章中的随机变量的方差有限.
- 由于平稳序列总是零均值平稳序列加上一个常数, 所以我们主要讨论零均值平稳序列的预测问题.

### §5.1.1 最佳线性预测

#### 最佳线性预测定义

- **定义1.1** 设 $Y$  和 $\mathbf{X} = (X_1, \dots, X_n)^T$ 是均值为零, 方差有限的随机变量(向量). 如果 $\mathbf{a} \in \mathbb{R}^n$ 使得对任何的 $\mathbf{b} \in \mathbb{R}^n$ , 有

$$E(Y - \mathbf{a}^T \mathbf{X})^2 \leq E(Y - \mathbf{b}^T \mathbf{X})^2.$$

则称 $\mathbf{a}^T \mathbf{X}$ 是用 $X_1, X_2, \dots, X_n$ 对 $Y$ 进行预测的**最佳线性预测**, 记做 $L(Y|\mathbf{X})$ 或 $\hat{Y}$ . 于是

$$\hat{Y} = L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X}. \quad (1.2)$$

- **定义1.2** 如果 $EY = b, E\mathbf{X} = \boldsymbol{\mu}$ , 定义

$$L(Y|\mathbf{X}) = L(Y - b|\mathbf{X} - \boldsymbol{\mu}) + b, \quad (1.3)$$

并称 $L(Y|\mathbf{X})$  是用 $X_1, X_2, \dots, X_n$ 对 $Y$ 进行预测时的最佳线性预测.

- 以下总设随机变量均值为零.
- 用 $\Gamma = E(\mathbf{X}\mathbf{X}^T)$ 表示 $\mathbf{X}$ 的协方差阵.
- 用 $\Sigma_{\mathbf{X}Y} = E(\mathbf{X}Y)$ 表示 $\mathbf{X}$ 和 $Y$ 的协方差向量.

## 性质1

- 性质1. 如果  $\mathbf{a} \in \mathbb{R}^n$ , 使得

$$\Gamma \mathbf{a} = \Sigma_{\mathbf{X}Y}, \quad (1.5)$$

则

$$L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X},$$

并且有

$$E(Y - L(Y|\mathbf{X}))^2 = EY^2 - E[L(Y|\mathbf{X})]^2 = EY^2 - \mathbf{a}^T \Gamma \mathbf{a}. \quad (1.6)$$

如果  $\Gamma$  和  $\Sigma_{\mathbf{X}Y}$  已知, 以  $\mathbf{a}$  为未知数的线性方程组(1.5)被称为预测方程.

证明: 对任何  $\mathbf{b} \in \mathbb{R}^n$ ,

$$\begin{aligned} & E(Y - \mathbf{b}^T \mathbf{X})^2 \\ &= E[Y - \mathbf{a}^T \mathbf{X} + (\mathbf{a}^T - \mathbf{b}^T) \mathbf{X}]^2 \\ &= E(Y - \mathbf{a}^T \mathbf{X})^2 + E[(\mathbf{a}^T - \mathbf{b}^T) \mathbf{X}]^2 + 2E[(\mathbf{a}^T - \mathbf{b}^T) \mathbf{X}(Y - \mathbf{a}^T \mathbf{X})] \\ &= E(Y - \mathbf{a}^T \mathbf{X})^2 + E[(\mathbf{a}^T - \mathbf{b}^T) \mathbf{X}]^2 + 2(\mathbf{a}^T - \mathbf{b}^T)[E(\mathbf{X}Y) - E(\mathbf{X}\mathbf{X}^T)\mathbf{a}] \\ &= E(Y - \mathbf{a}^T \mathbf{X})^2 + E[(\mathbf{a}^T - \mathbf{b}^T) \mathbf{X}]^2 \\ &\geq E(Y - \mathbf{a}^T \mathbf{X})^2. \end{aligned}$$

所以,  $\mathbf{a}^T \mathbf{X}$  是  $Y$  的最佳线性预测. 利用(1.5)得到

$$\begin{aligned} E[Y - L(Y|\mathbf{X})]^2 &= E(Y - \mathbf{a}^T \mathbf{X})^2 \\ &= EY^2 + \mathbf{a}^T E(\mathbf{X}\mathbf{X}^T)\mathbf{a} - 2\mathbf{a}^T E(\mathbf{X}Y) \\ &= EY^2 + \mathbf{a}^T \Gamma \mathbf{a} - 2\mathbf{a}^T \Gamma \mathbf{a} \\ &= EY^2 - \mathbf{a}^T \Gamma \mathbf{a}. \end{aligned}$$

- 注意:  $\mathbf{a}$  是预测方程的解等价于

$$E((Y - \mathbf{a}^T \mathbf{X})\mathbf{X}) = \Sigma_{\mathbf{X}Y} - \Gamma \mathbf{a} = 0$$

即  $Y - \mathbf{a}^T \mathbf{X}$  与  $\mathbf{X}$  正交。

- 性质说明  $Y - \mathbf{a}^T \mathbf{X}$  与  $\mathbf{X}$  正交则  $\mathbf{a}^T \mathbf{X} = L(Y|\mathbf{X})$ 。

## 性质2

## • 性质2.

1. 如果  $\Gamma = E(\mathbf{X}\mathbf{X}^T)$  可逆, 则  $\mathbf{a} = \Gamma^{-1}E(\mathbf{X}Y)$  使得  $L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X}$ 。
2. 预测方程  $\Gamma \mathbf{a} = E(\mathbf{X}Y)$  总有解。
3. 如果  $\det(\Gamma) = 0$ , 取正交矩阵  $A$  使得

$$A\Gamma A^T = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r, 0, \dots, 0), \quad \lambda_j > 0, j = 1, \dots, r.$$

定义  $\mathbf{Z} = A\mathbf{X} = (Z_1, Z_2, \dots, Z_n)^T = (Z_1, Z_2, \dots, Z_r, 0, \dots, 0)^T$  和  $\boldsymbol{\xi} = (Z_1, Z_2, \dots, Z_r)^T$ , 则  $E(\boldsymbol{\xi}\boldsymbol{\xi}^T)$  正定, 并且当取

$$\boldsymbol{\alpha} = [E(\boldsymbol{\xi}\boldsymbol{\xi}^T)]^{-1}E(\boldsymbol{\xi}Y) \quad (1.7)$$

时,  $L(Y|\mathbf{X}) = L(Y|\boldsymbol{\xi}) = \boldsymbol{\alpha}^T \boldsymbol{\xi}$ 。

- 性质2的第二条说明最佳线性预测总存在, 而且总可以由预测方程的解表示。
- 性质2的第三条说明当第一条不成立时,  $L(Y|\mathbf{X})$  可以通过  $\mathbf{X}$  的基表示。

**证明:** 仅需证明  $\det(\Gamma) = 0$  时第三和第二条成立。

$$\begin{aligned} E(\mathbf{Z}\mathbf{Z}^T) &= E(A\mathbf{X}\mathbf{X}^T A^T) = A\Gamma A^T \\ &= \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r, 0, \dots, 0) \end{aligned}$$

故  $Z_{r+1} = \dots = Z_n = 0$ 。且  $E(\boldsymbol{\xi}\boldsymbol{\xi}^T) = \text{diag}(\lambda_1, \dots, \lambda_r)$  是正定阵。当  $\boldsymbol{\alpha}$  按(1.7)定义时, 有

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \lambda_r \end{pmatrix} \boldsymbol{\alpha} = E(\boldsymbol{\xi}Y).$$

注意  $A\Gamma A^T$  与  $\lambda_1, \dots, \lambda_r$  的关系, 可以导出

$$\begin{aligned} A\Gamma A^T \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} &= \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} E(\boldsymbol{\xi}Y) \\ \mathbf{0} \end{pmatrix} = E\left(\begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{0} \end{pmatrix} Y\right) \\ &= E(\mathbf{Z}Y) = E(A\mathbf{X}Y) = AE(\mathbf{X}Y) = A\Sigma_{\mathbf{X}Y} \end{aligned}$$



两边同乘以  $A^T$ , 记  $\mathbf{a} = A^T \begin{pmatrix} \boldsymbol{\alpha} \\ 0 \end{pmatrix}$  有

$$\Gamma \mathbf{a} = \Sigma_{\mathbf{X}Y}$$

由性质1知

$$L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X} = (\boldsymbol{\alpha}^T, 0, \dots, 0) A \mathbf{X} = \boldsymbol{\alpha}^T \boldsymbol{\xi}$$

这在证明了第三条的同时也证明了第二条。

### 性质3

- **性质3.** 尽管  $\mathbf{a}$  由  $\Gamma \mathbf{a} = E(\mathbf{X}Y)$  决定时可以不惟一, 但  $L(Y|\mathbf{X})$  总是(a.s.)惟一的.
- **证明:** 预测方程总有解, 设  $\mathbf{a}$  为预测方程的一个解, 则由性质1对  $\forall \mathbf{b} \in \mathbb{R}^n$  有

$$E(Y - \mathbf{b}^T \mathbf{X})^2 = E(Y - \mathbf{a}^T \mathbf{X})^2 + E((\mathbf{a} - \mathbf{b})^T \mathbf{X})^2$$

若还有  $\mathbf{b}$  使得  $L(Y|\mathbf{X}) = \mathbf{b}^T \mathbf{X}$  则  $E(Y - \mathbf{b}^T \mathbf{X})^2 = E(Y - \mathbf{a}^T \mathbf{X})^2$ , 于是  $E((\mathbf{a} - \mathbf{b})^T \mathbf{X})^2 = 0$ , 即  $\mathbf{b}^T \mathbf{X} = \mathbf{a}^T \mathbf{X}$ , a.s.

### 性质4

- **性质4.**
  1. 如果  $E(\mathbf{X}Y) = 0$ , 则  $L(Y|\mathbf{X}) = 0$ .
  2. 如果  $Y = \sum_{j=1}^m b_j X_j$ , 则  $L(Y|\mathbf{X}) = Y$ .
- 这是线性预测的两个极端: 因变量和自变量不相关时线性预测无效; 因变量为自变量线性组合时可以完全线性预测。
- **证明:**
  1.  $\forall \mathbf{b}$

$$\begin{aligned} E(Y - \mathbf{b}^T \mathbf{X})^2 &= EY^2 + \mathbf{b}^T \Gamma \mathbf{b} - 2\mathbf{b}^T E(\mathbf{X}Y) \\ &= EY^2 + \mathbf{b}^T \Gamma \mathbf{b} \geq EY^2 = E(Y - 0)^2 \end{aligned}$$

所以  $L(Y|\mathbf{X}) = 0$ 。

2. 这时  $Y = \mathbf{b}^T \mathbf{X}$ ,  $E(Y - \mathbf{b}^T \mathbf{X})^2 = 0$ , 所以  $L(Y|\mathbf{X}) = \mathbf{b}^T \mathbf{X} = Y$ 。

## 性质5

- 设 $Y_1, Y_2, \dots, Y_m$ 是随机变量,  $b_j$ 是常数. 如果 $Y = \sum_{i=1}^m b_i Y_i$ , 则

$$L(Y|\mathbf{X}) = \sum_{j=1}^m b_j L(Y_j|\mathbf{X}).$$

- 性质5说明求最佳线性预测的运算 $L(\cdot|\mathbf{X})$ 是一种线性运算.
- 证明: 设 $\mathbf{a}_i$ 为 $\Gamma \mathbf{a}_i = \mathbf{E}(\mathbf{X}Y_i)$ 的解( $i = 1, 2, \dots, m$ ), 则 $L(Y_i|\mathbf{X}) = \mathbf{a}_i^T \mathbf{X}$ .  
取 $\mathbf{a} = \sum_{i=1}^m b_i \mathbf{a}_i$ , 则

$$\begin{aligned} \Gamma \mathbf{a} &= \Gamma \left( \sum_{i=1}^m b_i \mathbf{a}_i \right) = \sum_{i=1}^m b_i (\Gamma \mathbf{a}_i) = \sum_{i=1}^m b_i \mathbf{E}(\mathbf{X}Y_i) \\ &= \mathbf{E} \left( \mathbf{X} \sum_{i=1}^m b_i Y_i \right) = \mathbf{E}(\mathbf{X}Y) \end{aligned}$$

由性质1即知

$$L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X} = \sum_{i=1}^m b_i \mathbf{a}_i^T \mathbf{X} = \sum_{i=1}^m b_i L(Y_i|\mathbf{X})$$

## 性质6

- 性质6. 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ ,  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_m)^T$ . 如果 $E(\mathbf{X}\mathbf{Z}^T) = 0$ ( $\mathbf{X}$ 与 $\mathbf{Z}$ 不相关), 则有

$$L(Y|\mathbf{X}, \mathbf{Z}) = L(Y|\mathbf{X}) + L(Y|\mathbf{Z}).$$

- 证明: 记 $\boldsymbol{\xi} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix}$ , 记 $\Sigma_{XX} = \mathbf{E}(\mathbf{X}\mathbf{X}^T)$ ,  $\Sigma_{ZZ} = \mathbf{E}(\mathbf{Z}\mathbf{Z}^T)$ , 则

$$\Sigma \triangleq \mathbf{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T) = \begin{pmatrix} \Sigma_{XX} & \mathbf{0} \\ \mathbf{0} & \Sigma_{ZZ} \end{pmatrix}$$

设 $\mathbf{a}, \mathbf{b}$ 使得 $\Sigma_{XX}\mathbf{a} = \mathbf{E}(\mathbf{X}Y)$ ,  $\Sigma_{ZZ}\mathbf{b} = \mathbf{E}(\mathbf{Z}Y)$ , 则 $L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X}$ ,  $L(Y|\mathbf{Z}) = \mathbf{b}^T \mathbf{Z}$ , 取 $\mathbf{c} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$ , 则

$$\begin{aligned} \Sigma \mathbf{c} &= \begin{pmatrix} \Sigma_{XX}\mathbf{a} \\ \Sigma_{ZZ}\mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{E}(\mathbf{X}Y) \\ \mathbf{E}(\mathbf{Z}Y) \end{pmatrix} \\ &= \mathbf{E} \left( \begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} Y \right) = \mathbf{E}(\boldsymbol{\xi}Y) \end{aligned}$$

由性质1

$$\begin{aligned} L(Y|\mathbf{X}, \mathbf{Z}) &= L(Y|\boldsymbol{\xi}) = \mathbf{c}^T \boldsymbol{\xi} \\ &= \mathbf{a}^T \mathbf{X} + \mathbf{b}^T \mathbf{Z} = L(Y|\mathbf{X}) + L(Y|\mathbf{Z}) \end{aligned}$$

## 性质7

- 性质7 设  $\tilde{Y} = \mathbf{b}^T \mathbf{X}$  是  $\mathbf{X}$  的线性组合, 则  $\tilde{Y} = L(Y|\mathbf{X})$  的充分必要条件是

$$E(X_j(Y - \tilde{Y})) = 0, \quad 1 \leq j \leq n. \quad (1.8)$$

即

$$E(\mathbf{X}(Y - \mathbf{b}^T \mathbf{X})) = 0$$

- $L(Y|\mathbf{X})$  可以看成  $Y$  在  $\mathbf{X}$  张成的空间上的投影, 此性质即投影应满足的性质。
- 注意

$$E(\mathbf{X}(Y - \mathbf{b}^T \mathbf{X})) = E(\mathbf{X}Y) - \Gamma \mathbf{b}.$$

即残差与自变量正交等价于系数  $\mathbf{b}$  满足预测方程。

- 证明:

- 必要性: 设  $\tilde{Y}$  为  $L(Y|\mathbf{X})$ , 由性质2知存在  $\mathbf{a}$  满足预测方程, 由性质1和性质3知  $L(Y|\mathbf{X}) = \mathbf{a}^T \mathbf{X} = \mathbf{b}^T \mathbf{X}$ . 两边右乘以  $\mathbf{X}^T$  取期望得

$$\mathbf{a}^T \Gamma = \mathbf{b}^T \Gamma$$

注意  $\Gamma \mathbf{a} = E(\mathbf{X}Y)$  所以由上式得  $\Gamma \mathbf{b} = E(\mathbf{X}Y)$ , 即条件成立。

- 充分性: 条件成立时  $\mathbf{b}$  是预测方程的解, 由性质1即知  $\tilde{Y} = \mathbf{b}^T \mathbf{X}$  是最佳线性预测。

## 性质8

- 性质8. 如果

$$\hat{Y} = L(Y|X_1, X_2, \dots, X_n),$$

$$\tilde{Y} = L(Y|X_1, X_2, \dots, X_{n-1}),$$

则有

$$L(\hat{Y}|X_1, X_2, \dots, X_{n-1}) = \tilde{Y},$$

并且有

$$E(Y - \hat{Y})^2 \leq E(Y - \tilde{Y})^2. \quad (1.9)$$

- (1.9)表明在方差最小的意义下,  $\hat{Y}$  比  $\tilde{Y}$  要好. 这是由于  $X_1, X_2, \dots, X_n$  中包含的信息比  $X_1, X_2, \dots, X_{n-1}$  中包含的信息多的原因。

- 证明  $Y_0 \triangleq L(\hat{Y}|X_1, X_2, \dots, X_{n-1})$  是  $X_1, X_2, \dots, X_{n-1}$  的线性组合, 利用  $Y - \hat{Y}$ ,  $\hat{Y} - Y_0$  都和  $X_1, \dots, X_{n-1}$  正交, 得到

$$Y - Y_0 = (Y - \hat{Y}) + (\hat{Y} - Y_0)$$

和  $X_1, \dots, X_{n-1}$  正交. 利用性质7即知  $Y_0 = L(Y|X_1, \dots, X_{n-1}) = \tilde{Y}$ .

- $\hat{Y}$  是  $X_1, \dots, X_n$  对  $Y$  的最佳线性预测而  $\tilde{Y}$  是  $X_1, \dots, X_n$  对  $Y$  的一个线性组合所以有(1.9)成立。
- 这个性质实际是投影的性质。

### 性质9

- 性质9.(非零均值的最佳线性预测的意义) 如果  $EY = b$ ,  $E\mathbf{X} = \boldsymbol{\mu}$ , 按定义  $L(Y|\mathbf{X}) = b + L(Y - b|\mathbf{X} - \boldsymbol{\mu})$ . 事实上对任何  $c_0 \in \mathbb{R}$ ,  $\mathbf{c} \in \mathbb{R}^n$ ,

$$E[Y - L(Y|\mathbf{X})]^2 \leq E[Y - (c_0 + \mathbf{c}^T \mathbf{X})]^2. \quad (1.10)$$

- 证明 设  $L(Y|\mathbf{X}) = b + \mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu})$ , 则

$$\begin{aligned} & E\{Y - c_0 - \mathbf{c}^T \mathbf{X}\}^2 \\ &= E\{Y - b - \mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu}) + b + \mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu}) \\ &\quad - [c_0 + \mathbf{c}^T \boldsymbol{\mu} + \mathbf{c}^T(\mathbf{X} - \boldsymbol{\mu})]\}^2 \\ &= E\{[Y - b - \mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu})] + (b - c_0 - \mathbf{c}^T \boldsymbol{\mu}) + (\mathbf{a} - \mathbf{c})^T(\mathbf{X} - \boldsymbol{\mu})\}^2 \\ &= E[Y - b - \mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu})]^2 + (b - c_0 - \mathbf{c}^T \boldsymbol{\mu})^2 + (\mathbf{a} - \mathbf{c})^T \Gamma(\mathbf{a} - \mathbf{c}) \\ &\geq E[Y - b - \mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu})]^2 = E[Y - L(Y|\mathbf{X})]^2 \end{aligned}$$

### 性质10

- 性质10. 设  $\mathbf{X}$  和  $\mathbf{Y}$  分别是  $m$  和  $n$  维向量, 如果有实矩阵  $A, B$  使得  $\mathbf{X} = A\mathbf{Y}$ ,  $\mathbf{Y} = B\mathbf{X}$ , 则  $L(Z|\mathbf{X}) = L(Z|\mathbf{Y})$ .
- 如果  $\mathbf{X}$  和  $\mathbf{Y}$  能互相线性表示则其预报  $Z$  能达到的下界是相同的, 预报是一致的。
- 证明为习题。

### 预测计算例子

- 设已知 ARMA(4,2) 的参数。观测到  $\mathbf{x}_n = (x_1, \dots, x_n)^T$ ,  $n = 14$ 。
- 用  $\mathbf{x}_n$  预测  $x_{n+k}$ ,  $k = 1, 2, \dots, m$ ,  $m = 7$ 。

- 预测方程中 $\Gamma$ 为平稳序列的 $\Gamma_n$ 。
- 预测方程中 $\Sigma_{XY}$ 当 $Y = X_{n+k}$ 时为

$$\mathbf{g}_k = E(\mathbf{X}_n X_{n+k}) = (\gamma_{n+k-1}, \gamma_{n+k-2}, \dots, \gamma_k)^T$$

- 最佳线性预测为

$$\hat{X}_{n+k} \triangleq L(X_{n+k}|\mathbf{X}_n) = (\Gamma_n^{-1} \mathbf{g}_k)^T \mathbf{X}_n = \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{X}_n$$

- 预测方差为

$$\sigma^2(k) = \gamma_0 - (\Gamma_n^{-1} \mathbf{g}_k)^T \Gamma_n (\Gamma_n^{-1} \mathbf{g}_k) = \gamma_0 - \mathbf{g}_k^T \Gamma_n^{-1} \mathbf{g}_k$$

- 设 $\{X_t\}$ 为正态平稳列, 则 $X_{n+k} - \hat{X}_{n+k}$ 作为有限线性组合也是正态分布的。 $X_{n+k} - \hat{X}_{n+k} \sim N(0, \sigma^2(k))$ ,
- 可以构造 $X_{n+k}$ 的置信区间:

$$\Pr(|X_{n+k} - \hat{X}_{n+k}|/\sigma(k) \leq 1.96) = 0.95, \quad k = 1, 2, \dots, m$$

- 见演示。
- 对真实数据需要用 $x_1, x_2, \dots, x_N$ 估计 $\hat{\gamma}_k$ , 然后用令 $\mathbf{x}_n = (x_{N-n+1}, x_{N-n+2}, \dots, x_N)^T$ , 用 $\mathbf{x}_n$ 预报 $X_{N+k}$ ,  $k = 1, 2, \dots$ 。
- 数据先减去均值再估计并预测, 预测值要把均值加回去。

### §5.1.2 Hilbert空间中的投影

#### Hilbert空间中的投影

- 下面说明最佳线性预测实际上是Hilbert 空间中的投影.
- 用 $L^2$ 表示全体方差有限的随机变量构成的Hilbert 空间(参见§1.6).
- 设 $H$ 是 $L^2$ 的闭子空间,  $Y$ 属于 $L^2$ . 可以证明 $H$ 中存在惟一的 $\hat{Y}$ 使得

$$E(Y - \hat{Y})^2 = \inf_{\xi \in H} E(Y - \xi)^2 \quad (1.11)$$

- **定义1.3** 如果 $H$ 是 $L^2$ 的闭子空间,  $Y \in L^2$ ,  $\hat{Y} \in H$ 使得(1.11)成立, 则称 $\hat{Y}$ 是 $Y$ 在 $H$ 上的投影. 记做 $P_H(Y)$ , 并且称 $P_H$ 是投影算子.
- **定义1.4** 设 $Y \in L^2$ , 如果对 $H$ 中的任何 $\xi$ ,  $E(Y\xi) = 0$ , 则称 $Y$ 垂直于 $H$ .

## 投影存在唯一的证明

- 取  $Y_n \in H$  使

$$d = \inf_{\xi \in H} E(Y - \xi)^2 = \lim_{n \rightarrow \infty} E(Y - Y_n)^2$$

则  $(Y_n + Y_m)/2 \in H$ , 并且当  $n, m \rightarrow \infty$

•

$$\begin{aligned} & E(Y_n - Y_m)^2 \\ &= E[(Y_n - Y) - (Y_m - Y)]^2 \\ &\quad + E[(Y_n - Y) + (Y_m - Y)]^2 \\ &\quad - E[(Y_n + Y_m) - 2Y]^2 \\ &= 2E(Y_n - Y)^2 + 2E(Y_m - Y)^2 - 4E[(Y_n + Y_m)/2 - Y]^2 \\ &\leq 2E(Y_n - Y)^2 + 2E(Y_m - Y)^2 - 4d \\ &\rightarrow 2d + 2d - 4d = 0. \end{aligned} \tag{1.12}$$

- 于是,  $\{Y_n\}$  是  $H$  中的基本列, 从而有  $\hat{Y} \in H$  使得  $Y_n$  均方收敛到  $\hat{Y}$ .
- 由内积的连续性知道

$$E(Y - \hat{Y})^2 = \lim_{n \rightarrow \infty} E(Y - Y_n)^2 = d.$$

于是,  $\hat{Y}$  满足 (1.11).

- 如果又有  $\hat{\xi} \in H$  也使得 (1.11) 成立, 仿照 (1.12) 的推导得到

$$\begin{aligned} & E(\hat{Y} - \hat{\xi})^2 \\ &= E[(\hat{Y} - Y) - (\hat{\xi} - Y)]^2 \\ &\quad + E[(\hat{Y} - Y) + (\hat{\xi} - Y)]^2 - E[(\hat{Y} + \hat{\xi}) - 2Y]^2 \\ &= 2E(\hat{Y} - Y)^2 + 2E(\hat{\xi} - Y)^2 - 4E[(\hat{Y} + \hat{\xi})/2 - Y]^2 \\ &\leq 2d + 2d - 4d = 0. \end{aligned}$$

所以  $\hat{\xi} = \hat{Y}$ , a.s..

## 投影的垂直性(正交性)

- **定理1.1** 设  $Y \in L^2$ ,  $\hat{Y} \in H$ , 则  $\hat{Y} = P_H(Y)$  的充分必要条件是  $(Y - \hat{Y})$  垂直于  $H$ .

- **证明** 先证必要性. 设  $\hat{Y} = P_H(Y)$ . 对  $\forall \xi \in H$ , 我们证明

$$a \triangleq E[(Y - \hat{Y})\xi] = 0.$$

不妨设  $E\xi^2 = 1$ , 这时

$$\begin{aligned} d &\triangleq E(Y - \hat{Y})^2 \leq E(Y - \hat{Y} - a\xi)^2 \\ &= E(Y - \hat{Y})^2 + E(a\xi)^2 - 2aE[(Y - \hat{Y})\xi] \\ &= d + a^2 - 2a^2 = d - a^2 \end{aligned}$$

由此得到  $a = 0$ .

- 来证明充分性. 若  $\hat{Y} \in H$  使  $Y - \hat{Y} \perp H$ , 则对  $\forall \xi \in H$  有

•

$$\begin{aligned} E(Y - \xi)^2 &= E(Y - \hat{Y} + \hat{Y} - \xi)^2 \\ &= E(Y - \hat{Y})^2 + E(\hat{Y} - \xi)^2 + 2E[(Y - \hat{Y})(\hat{Y} - \xi)] \\ &= E(Y - \hat{Y})^2 + E(\hat{Y} - \xi)^2 \\ &\geq E(Y - \hat{Y})^2 \end{aligned}$$

- 即  $\hat{Y} = P_H(Y)$ .

### 最佳线性预报与投影的等价性

- 用  $L^2(\mathbf{X})$  表示  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  的元素和常数1生成的Hilbert空间. 它是  $X_1, X_2, \dots, X_n$  和常数1的线性组合的全体(参见第1章, 习题6.5).
- 设  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T = E\mathbf{X}$ .
- 对任何方差有限的随机变量  $Y$ , 设  $EY = b$ ,  $\hat{Y} = L(Y|\mathbf{X})$  由(1.3)式定义. 则有

$$Y - \hat{Y} = (Y - b) - L(Y - b|\mathbf{X} - \boldsymbol{\mu}).$$

- 利用性质7 知道

$$\begin{aligned} E[1 \cdot (Y - \hat{Y})] &= E(Y - b) - EL(Y - b|\mathbf{X} - \boldsymbol{\mu}) = 0, \\ E[X_i(Y - \hat{Y})] &= E[(X_i - \mu_i)(Y - \hat{Y})] + \mu_i E(Y - \hat{Y}) = 0. \end{aligned}$$

即得到  $(Y - \hat{Y}) \perp H \triangleq L^2(\mathbf{X})$ .

- 由定理1.1知道

$$L(Y|\mathbf{X}) = P_H(Y).$$

- 基于上述原因, 当 $H$ 是 $\{X_j : j \in T\}$ 和常数1生成的Hilbert空间, 我们也用

$$L(Y|1, X_j, j \in T) \quad \text{或} \quad L(Y|H)$$

表示 $P_H(Y)$ , 这里 $T$ 是一个可列的指标集.

- 下面记 $\|\xi\| = \sqrt{E\xi^2}$ ,  $\forall \xi \in L^2$ .  $\|\xi\|$ 是 $\xi$ 的长度。

### 投影算子的性质

**定理1.2** 设 $H, M$ 是 $L^2$ 的闭子空间,  $X, Y \in L^2$ ,  $a, b$ 是常数.

1.  $L(aX + bY|H) = aL(X|H) + bL(Y|H)$ ,
2.  $\|Y\|^2 = \|L(Y|H)\|^2 + \|Y - L(Y|H)\|^2$ ,
3.  $\|L(Y|H)\| \leq \|Y\|$ ,
4.  $Y \in H$  的充分必要条件是 $L(Y|H) = Y$ ,
5.  $Y$ 垂直于 $H$ 的充分必要条件是 $L(Y|H) = 0$ ,
6. 如果 $H$ 是 $M$ 的子空间, 则 $P_H P_M = P_M P_H = P_H$ , 并且对 $Y \in L^2$ ,

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|H)\|.$$

### 定理1.2证明

- (1) 设 $Z = aL(X|H) + bL(Y|H)$ 则 $Z \in H$ 。由

$$(aX + bY) - Z = a[X - L(X|H)] + b[Y - L(Y|H)]$$

看出 $(aX + bY) - Z \perp H$ 。所以

$$L(aX + bY|H) = Z = aL(X|H) + bL(Y|H)$$

这说明投影是线性算子。

- (2) 由于 $L(Y|H) \in H$ 而 $Y - L(Y|H) \perp H$ 所以

$$\begin{aligned} \|Y\|^2 &= E[(Y - L(Y|H)) + L(Y|H)]^2 \\ &= E[Y - L(Y|H)]^2 + E[L(Y|H)]^2 \\ &\quad + 2E[(Y - L(Y|H))L(Y|H)] \\ &= E[Y - L(Y|H)]^2 + E[L(Y|H)]^2 \\ &= \|Y - L(Y|H)\|^2 + \|L(Y|H)\|^2 \end{aligned}$$

- (3) 由(2)直接得到。



- (4) 必要性:  $Y \in H$  时取  $L(Y|H) = Y$  可得均方误差为0。  
充分性: 若  $L(Y|H) = Y$  则由于投影必须属于  $H$  所以  $Y \in H$ 。
- (5) 必要性: 若  $Y \perp H$  则  $0 \in H, Y - 0 \perp H$  所以  $L(Y|H) = 0$ 。  
充分性: 若  $L(Y|H) = 0$  则由  $Y - L(Y|H) \perp H$  知  $Y \perp H$ 。
- (6)  $\forall Y \in L^2$ , 设  $\xi = P_M(Y), \eta = P_H(Y)$ , 来证  $P_H(\xi) = \eta$ 。事实上,  $\eta \in H$  且

$$\xi - \eta = (Y - \eta) - (Y - \xi)$$

其中  $Y - \eta$  和  $Y - \xi$  都与  $H$  垂直, 所以  $P_H(\xi) = \eta$ , 即  $P_H P_M = P_H$ 。另外,  $H \subseteq M$  所以  $\eta \in M, P_M(\eta) = \eta$ , 即  $P_M P_H = P_H$ 。

由  $P_H(Y) \in M$  和  $P_M(Y)$  的定义马上可得

$$\|Y - P_H(Y)\|^2 \geq \|Y - P_M(Y)\|^2$$

### §5.1.3 最佳预测

#### 最佳预测

- 最佳线性预测只用了自变量的线性函数而未考虑其他函数。
- 设

$$M = \overline{\text{sp}}\{g(\mathbf{X}) : E g^2(\mathbf{X}) < \infty, g(\cdot) \text{ 是可测函数} \} \quad (1.14)$$

- 考虑用  $M$  中的元素逼近  $Y$ 。
- **定义1.5** 设  $M$  由(1.14)定义. 用  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  对  $Y$  进行预测时, 称

$$L(Y|M) \triangleq P_M(Y) \quad (1.15)$$

为  $Y$  的最佳预测。

- 最佳预测  $L(Y|M)$  实际上是概率论中的条件数学期望  $E(Y|\mathbf{X})$ 。
- $L^2(\mathbf{X})$  是  $M$  的子空间, 由定理1.2的(6)得

$$\|Y - L(Y|M)\| \leq \|Y - L(Y|\mathbf{X})\|$$

在预测均方误差最小的意义下最佳预测比最佳线性预测好。

- 但是由于  $M$  要比  $L^2(\mathbf{X})$  复杂很多, 实际计算最佳预测往往比计算最佳线性预测困难得多。
- 对于正态序列来讲, 最佳预测和最佳线性预测是一致的。

## 正态分布时最佳线性预测与最佳预测的等价性

- **定理1.3** 如果  $(X_1, X_2, \dots, X_n, Y)^T$  服从联合正态分布  $N(\boldsymbol{\mu}, \Sigma)$ ,  $M$  由(1.14)定义, 则

$$L(Y|M) = L(Y|X_1, X_2, \dots, X_n). \quad (1.16)$$

- **证明** 设  $\hat{Y} = L(Y|X_1, X_2, \dots, X_n)$ , 则  $(Y - \hat{Y})$  与  $\mathbf{X}$  正交. 由于  $E(Y - \hat{Y}) = 0$ , 所以  $Y - \hat{Y}$  与  $\mathbf{X}$  不相关. 由正态分布的性质知道,  $Y - \hat{Y}$  与  $\mathbf{X}$  独立, 从而和  $M$  中的任何随机变量独立. 对任何  $\xi \in M$ ,  $E[\xi(Y - \hat{Y})] = (E\xi)E(Y - \hat{Y}) = 0$ , 即  $Y - \hat{Y}$  垂直于  $M$ . 从  $\hat{Y} \in M$  和定理1.1 知道(1.16)成立.

## 例子

- **例1.2** 设随机变量  $\varepsilon, \eta$  独立, 都服从标准正态分布  $N(0, 1)$ ,
- 则  $E\eta^4 = 3$ .
- 取  $X = \eta, Y = (3\varepsilon^2 - \eta^2)\eta$ .
- 则  $EX = EY = 0, E(XY) = E(3\varepsilon^2\eta^2 - \eta^4) = 0$ . 从而  $L(Y|X) = 0$ .
- 计算

$$\begin{aligned} E(Y|X) &= E(3\varepsilon^2\eta - \eta^3|\eta) \\ &= 3\eta - \eta^3 = 3X - X^3 \end{aligned}$$

- 容易验证  $Y - (3\eta - \eta^3) = 3(\varepsilon^2 - 1)\eta$  垂直于

$$M = \bar{s}p\{g(X) : Eg^2(X) < \infty, g(x) \text{ 是可测函数}\}.$$

- 于是, 从  $3\eta - \eta^3 \in M$  知道:  $L(Y|M) = 3X - X^3$ .

## §5.2 非决定性平稳序列及其Wold表示

### §5.2.1 非决定性平稳序列

#### 非决定性平稳序列—介绍

- 对平稳序列, 考虑用所有的历史 $\{X_t, t \leq n\}$ 对 $X_{n+1}$ 进行最佳线性预测.
- 当预测误差是零时,  $X_{n+1}$ 的信息完全含在历史资料中. 这样的平稳序列被称为**决定性的**.
- 最小序列: 用 $\{X_s : s \neq t\}$ 预报 $X_t$ 误差不为零. 决定性序列不是最小序列.
- 实际问题中, 决定性平稳序列描述事物的发展没有新的信息出现.
- 实际问题中, 决定性平稳序列描述事物的发展没有新的信息出现.
- 如果用 $\{X_t, t \leq n\}$ 预测 $X_{n+1}$ 的误差不是零, 说明 $X_{n+1}$ 的信息不能由历史资料的线性组合及其极限完全确定, 我们称这种时间序列是**非决定性的**.
- 非决定性平稳序列描述事物的发展总伴随新的信息出现.
- 最小序列一定是非决定性的.
- 平稳序列的Wold定理表示告诉我们, 非决定性平稳序列总是可以分解成白噪声的单边滑动和加上一个决定性平稳序列.
- 从应用的角度讲, 非决定性平稳序列总是白噪声的单边滑动和加上一个离散谱序列.

#### 最佳线性预测均方误差的极限

- 设 $\{X_n : n \in \mathbb{Z}\}$ 是零均值平稳序列. 记

$$\mathbf{X}_{n,m} = (X_n, X_{n-1}, \dots, X_{n+1-m})^T,$$

这里 $n$ 表示向量的第一个脚标,  $m$ 表示向量的维数.

- 定义

$$\hat{X}_{n+1,m} = L(X_{n+1} | \mathbf{X}_{n,m}).$$

- 从最佳线性预测的性质8知道 $\sigma_{1,m}^2 = E(X_{n+1} - \hat{X}_{n+1,m})^2$  是 $m$ 的单调减函数, 于是定义

$$\sigma_1^2 \triangleq \lim_{m \rightarrow \infty} \sigma_{1,m}^2 < \infty.$$

- **定理2.1**  $\sigma_1^2 \triangleq \lim_{m \rightarrow \infty} \sigma_{1,m}^2$  与 $n$ 无关.

- **证明** 设  $\mathbf{a} = (a_1, a_2, \dots, a_m)^T$  是预测方程(1.5)的解, 则  $\mathbf{a}$  和  $n$  无关. 由于

$$Y_n \triangleq X_{n+1} - \sum_{j=1}^m a_j X_{n+1-j} = X_{n+1} - \hat{X}_{n+1,m}, \quad n \in \mathbb{Z}, \quad (2.1)$$

是平稳序列, 所以  $\sigma_{1,m}^2 = EY_n^2 = EY_0^2$  与  $n$  无关. 最后  $\sigma_1^2 = \lim_{m \rightarrow \infty} \sigma_{1,m}^2$  与  $n$  无关.  $\square$

### 决定性与非决定性

- 对充分大的  $m$ ,  $L(X_{n+1}|\mathbf{X}_{n,m})$  表示用充分多的历史对未来  $X_{n+1}$  进行预测.
- $\sigma_{1,m}^2$  表示的是预测的均方误差.
- 当  $m \rightarrow \infty$  时,  $\sigma_{1,m}^2 \rightarrow 0$  说明  $X_{n+1}$  可以由所有历史  $X_n, X_{n-1}, \dots$  进行完全预测.
- 当  $\sigma_1^2 > 0$  说明  $X_{n+1}$  不可以由所有历史  $X_n, X_{n-1}, \dots$  的线性组合以及极限进行完全预测.
- **定义2.1** 设  $\{X_t\}$  是零均值平稳序列.
  1. 如果  $\sigma_1^2 = 0$ , 称  $\{X_t\}$  是决定性平稳序列;
  2. 如果  $\sigma_1^2 > 0$ , 称  $\{X_t\}$  是非决定性平稳序列, 并且称  $\sigma_1^2 = \lim_{m \rightarrow \infty} \sigma_{1,m}^2$  为  $\{X_t\}$  的一步(线性)预测的均方误差.

### 非零均值情况

- 对于平稳序列  $\{X_t\}$ , 如果  $EX_t = \mu$ , 引入  $\{Z_t\} = \{X_t - \mu\}$  和  $m$  维向量  $\boldsymbol{\mu}_m = (\mu, \dots, \mu)^T$ .
- 按照最佳线性预测的定义1.2,

$$\begin{aligned} \hat{X}_{n+1,m} &= \mu + L(X_{n+1} - \mu | \mathbf{X}_{n,m} - \boldsymbol{\mu}_m) \\ &= \mu + L(Z_{n+1} | \mathbf{Z}_{n,m}) = \mu + \hat{Z}_{n+1,m}. \end{aligned}$$

- 于是

$$E(Z_{n+1} - \hat{Z}_{n+1,m})^2 = E(X_{n+1} - \hat{X}_{n+1,m})^2.$$

即对  $X$  预报的均方误差等于对中心化得到的  $Z$  预报的均方误差。

- 因而, 当且仅当  $\{X_t - \mu\}$  是决定性平稳序列时, 称  $\{X_t\}$  是决定性平稳序列. 于是以后只需要讨论零均值的平稳序列.

**例2.1 可完全线性预测**

- 设平稳列  $\{X_t\}$  的  $n+1$  阶自协方差阵  $\Gamma_{n+1}$  退化,  $|\Gamma_n| > 0$ 。
- 则  $X_1, X_2, \dots, X_{n+1}$  线性相关,
- 所以  $X_{n+1}$  可以由  $X_n, X_{n-1}, \dots, X_1$  线性表示。
- 于是,  $L(X_{n+1}|X_n, \dots, X_1) = X_{n+1}$ 。
- 当  $m \geq n$  时,

$$L(X_{n+1}|X_n, \dots, X_{n-m+1}) = X_{n+1},$$

- 即有  $\sigma_{1,m}^2 = 0$ ,  $\{X_t\}$  是决定性平稳列。
- 最简单的决定性平稳列是  $X_t \equiv \xi$ ,  $\xi$  为随机变量。

**例2.2 离散谱序列**

- 设零均值随机变量  $\xi_j, \eta_k$  ( $j, k = 1, 2, \dots, p$ ) 两两正交, 满足

$$E(\xi_j^2) = E(\eta_j^2) = \sigma_j^2, \quad j = 1, 2, \dots \quad (2.2)$$

- 对确定的  $j$ , 定义简单离散谱序列

$$Z_j(t) = \xi_j \cos(t\lambda_j) + \eta_j \sin(t\lambda_j), \quad t \in \mathbb{Z}. \quad (2.3)$$

- 可以证明  $\{Z_j(t)\}$  是平稳序列。
- 事实上, 易见  $EZ_j(t) \equiv 0$ 。而

$$\begin{aligned} E[X_j(t)X_j(s)] &= E\xi_j^2 \cos(\lambda_j t) \cos(\lambda_j s) + E\eta_j^2 \sin(\lambda_j t) \sin(\lambda_j s) \\ &= \sigma_j^2 \cos((t-s)\lambda_j) \end{aligned}$$

只依赖于  $t-s$ 。

- $\{Z_j(t)\}$  的每一次实现是周期函数, 由 §2.3 的定理 3.7 知道  $\{Z_j(t)\}$  的 3 阶自协方差矩阵是退化的,
- 因此 (2.3) 是可完全线性预测的,  $X_n$  可以被  $X_{n-1}, X_{n-2}$  完全线性预测, 是决定性序列。
- 事实上容易证明

$$Z_j(t) = (2 \cos \lambda) Z_j(t-1) - Z_j(t-2)$$

- 定义离散谱序列

$$Z_t = \sum_{j=1}^p Z_j(t), \quad t \in \mathbb{Z}. \quad (2.4)$$

这是 $p$ 个简单离散谱序列的叠加.

- 由§2.3的定理3.7知道由(2.4)定义的离散谱序列也是决定性的.  $X_n$ 可以被 $X_{n-1}, X_{n-2}, \dots, X_{n-2p}$ 完全线性预测。

### 纯非决定性

- 决定性与非决定性取决于一步线性预报误差是否为零。
- 对非决定性序列, 用 $\{X_s, s \leq n\}$ 预报 $X_{n+k}$ 的误差会随 $k$ 增大而增大。
- 记

$$\sigma_{k,m}^2 = E[X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})]^2$$

则 $\sigma_{k,m}$ 也是 $m$ 的单调递减函数, 与 $n$ 无关。可定义

$$\sigma_k^2 = \lim_{m \rightarrow \infty} \sigma_{k,m}^2$$

- $\sigma_{k,m}^2$ 是 $k$ 的单调递增函数, 从而 $\sigma_k^2$ 是单调递增的:

$$\begin{aligned} \sigma_k^2 &= \lim_{m \rightarrow \infty} E(X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m}))^2 \\ &= \lim_{m \rightarrow \infty} E[X_{n+k-1} - L(X_{n+k-1}|X_{n-1}, X_{n-2}, \dots, X_{n-1-m})]^2 \\ &\geq \lim_{m \rightarrow \infty} E[X_{n+k-1} - L(X_{n+k-1}|X_n, X_{n-1}, \dots, X_{n-m-1})]^2 \\ &= \sigma_{k-1}^2. \end{aligned} \quad (2.6)$$

- 由最佳线性预测定义知

$$\begin{aligned} \sigma_{k,m}^2 &= E[X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})]^2 \\ &\leq E[X_{n+k} - 0]^2 = \gamma_0 \end{aligned}$$

所以 $\sigma_k^2 \leq \gamma_0$ 。

- $k \rightarrow \infty$ 时如果 $\sigma_k^2 \rightarrow \gamma_0$  则最佳线性预测与用平均值0预测效果相同, 没有作用。
- **定义2.2** 设 $\{X_t\}$ 是非决定性的平稳序列. 如果 $\lim_{k \rightarrow \infty} \sigma_k^2 = \gamma_0$ , 则称 $\{X_t\}$ 是**纯非决定性的**.

- 纯非决定性的平稳列不能作长期预报。与短记忆有关。
- 非决定性但不是纯非决定性的平稳列作长期预报是有意义的; 当然, 决定性序列可以精确地长期预报。
- 对纯非决定性的平稳序列, 有如下的结果:

$$\lim_{k \rightarrow \infty} \lim_{m \rightarrow \infty} E[L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})]^2 = 0. \quad (2.7)$$

- 实际上, 记  $\hat{X}_{n+k,m} = L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})$ . 由投影的正交性得

$$\sigma_{k,m}^2 = E(X_{n+k} - \hat{X}_{n+k,m})^2 = EX_{n+k}^2 - E\hat{X}_{n+k,m}^2.$$

于是得到

$$\lim_{k \rightarrow \infty} \lim_{m \rightarrow \infty} E\hat{X}_{n+k,m}^2 = \lim_{k \rightarrow \infty} \lim_{m \rightarrow \infty} (\gamma_0 - \sigma_{k,m}^2) = \gamma_0 - \gamma_0 = 0. \quad (2.8)$$

- 从(2.7)也可看出, 对于纯非决定性的平稳序列做长期或超长期预测是不合适的.

### §5.2.2 Wold表示定理

#### 线性闭包的等价定义

- **引理2.2** 设  $A$  为 Hilbert 空间  $H$  的子集, 记  $L_A$  为  $A$  的所有有限线性组合构成的子集, 记  $H_A$  为包含  $A$  的最小的闭子空间(称为由  $A$  生成的闭子空间). 则  $\forall \xi \in H_A$ , 必存在  $\xi_n \in L_A, n = 1, 2, \dots$  使得

$$\|\xi_n - \xi\| \rightarrow 0, \quad n \rightarrow \infty.$$

- **证明:** 易见  $L_A$  为  $H$  的子线性空间,  $H_A$  为  $H$  的子 Hilbert 空间,  $L_A \subset H_A$ . 记  $\bar{L}_A$  为  $L_A$  及  $L_A$  中元素的极限构成的集合, 因  $L_A \subset H_A$  及  $H_A$  是闭集可知  $\bar{L}_A \subset H_A$ .

另一方面, 可以证明  $\bar{L}_A$  是闭子空间, 由  $H_A$  的定义及  $A \subset L_A$  得  $H_A \subset \bar{L}_A$ .

#### 线性预测的极限

- 记  $H_n$  为  $X_n, X_{n-1}, \dots$  生成的闭子空间。
- $L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1})$  当  $m \rightarrow \infty$  时为  $L(X_{n+k}|H_n)$  (定理2.4)。
- **定理2.3** 设  $Y \in L^2, \xi \in H_n$ , 则  $\xi = L(Y|H_n)$  的充分必要条件是

$$Y - \xi \perp X_j, \quad j = n, n-1, n-2, \dots \quad (*)$$

• 证明:

- 必要性: 由定理1.1得到  $Y - \xi \perp H_n$  所以有(\*)。
- 充分性: 记  $A = \{X_n, X_{n-1}, \dots\}$ , 则由(\*)可知  $Y - \xi \perp L_A$ 。由引理2.2, 对  $\eta \in H_n$  有  $\eta_m \in L_A$  使  $\eta_m \rightarrow \eta$ , 由内积的连续性可得

$$E((Y - \xi)\eta) = \lim_{m \rightarrow \infty} E((Y - \xi)\eta_m) = 0$$

即  $Y - \xi \perp H_n$ , 由定理1.1即得  $\xi = L(X_{n+k}|H_n)$ 。

### 无穷维历史线性预报的意义

- $H_n$  为  $\{X_s, s \leq n\}$  所张成的子Hilbert空间,  $L(X_{n+k}|H_n)$  是一个投影。下面的定理说明这个投影是有穷维最佳线性预测的极限。
- 定理2.4 设  $\mathbf{X}_{n,m} = (X_n, X_{n-1}, \dots, X_{n-m+1})^T$ , 当  $m \rightarrow \infty$  时

$$L(Y|\mathbf{X}_{n,m}) \xrightarrow{m.s.} \hat{Y} \triangleq L(Y|H_n) \quad (2.11)$$

### 定理2.4证明

- 证明: 记  $\hat{Y}_m = L(Y|\mathbf{X}_{n,m})$ 。先证明  $\{\hat{Y}_m\}$  是  $H_n$  中基本列。显然  $\hat{Y}_m \in H_n$ , 设当  $m \rightarrow \infty$  时

$$\eta_m^2 \triangleq E(Y - \hat{Y}_m)^2 \rightarrow \eta^2 \quad (\text{注意单调性})$$

- 对  $m, k \rightarrow \infty$ , 注意  $\hat{Y}_m, \hat{Y}_{m+k}$  都和  $Y - \hat{Y}_{m+k}$  正交, 得

$$\begin{aligned} \|\hat{Y}_m - \hat{Y}_{m+k}\|^2 &= \|\hat{Y}_m - Y + Y - \hat{Y}_{m+k}\|^2 \\ &= \|\hat{Y}_m - Y\|^2 + \|Y - \hat{Y}_{m+k}\|^2 + 2\langle \hat{Y}_m - Y, Y - \hat{Y}_{m+k} \rangle \\ &= \eta_m^2 + \eta_{m+k}^2 - 2\langle Y, Y - \hat{Y}_{m+k} \rangle \\ &= \eta_m^2 + \eta_{m+k}^2 - 2\langle Y - \hat{Y}_{m+k}, Y - \hat{Y}_{m+k} \rangle \\ &= \eta_m^2 + \eta_{m+k}^2 - 2\eta_{m+k}^2 \rightarrow 0 \end{aligned}$$

- 因此  $\{\hat{Y}_m\}$  是  $H_n$  的基本列, 在  $H_n$  中存在唯一极限  $\xi$ 。
- 由内积的连续性, 对任何  $X_s, s \leq n$  有

$$\langle X_s, Y - \xi \rangle = \lim_{m \rightarrow \infty} \langle X_s, Y - L(Y|\mathbf{X}_{n,m}) \rangle = 0$$

- 由定理2.3得到  $\xi = L(Y|H_n)$ 。

□



## 最佳线性预报的方差

- 由内积连续性,

$$\begin{aligned}\sigma_1^2 &\triangleq \lim_{m \rightarrow \infty} \|X_{n+1} - L(X_{n+1} | \mathbf{X}_{n,m})\|^2 \\ &= \|X_{n+1} - L(X_{n+1} | H_n)\|^2 = \|X_1 - L(X_1 | H_0)\|^2\end{aligned}$$

- $\sigma_1^2 = 0 \leftrightarrow X_1 = L(X_1 | H_0)$ , 由定理1.2的(4),  $\sigma_1^2 = 0 \leftrightarrow X_1 \in H_0$ 。

- 类似地,

$$\begin{aligned}\sigma_k^2 &\triangleq \lim_{m \rightarrow \infty} \|X_{n+k} - L(X_{n+k} | \mathbf{X}_{n,m})\|^2 \\ &= \|X_{n+k} - L(X_{n+k} | H_n)\|^2 = \|X_k - L(X_k | H_0)\|^2\end{aligned}$$

- **定理2.5** 设  $\{X_t\}$  是零均值平稳列,

1.  $\{X_t\}$  是决定性序列当且仅当对某个  $n$  有

$$X_{n+1} \in H_n; \quad (2.14)$$

并且如果(2.14)对某个  $n$  成立则对所有  $n$  成立, 这时  $H_n = H_{n-1}, \forall n \in \mathbb{Z}$ 。

2.  $\{X_t\}$  是纯非决定性的当且仅当对某个  $n$ , 有

$$\sigma_k^2 = \|X_{n+k} - L(X_{n+k} | H_n)\|^2 \rightarrow \gamma_0, \quad k \rightarrow \infty \quad (2.15)$$

并且如果(2.15)对某个  $n$  成立则对所有  $n$  成立。

## Wold 表示定理

- **定理2.6**(Wold表示定理): 任一非决定性的零均值平稳列可以表示成

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} + V_t, \quad t \in \mathbb{Z} \quad (2.16)$$

其中

- (1)  $\varepsilon_t = X_t - L(X_t | X_{t-1}, X_{t-2}, \dots)$  是零均值白噪声, 满足

$$\begin{aligned}E\varepsilon_t^2 &= \sigma^2 > 0, \quad a_0 = 1 \\ a_j &= \langle X_t, \varepsilon_t / \sigma \rangle = E(X_t \varepsilon_{t-j}) / \sigma^2, \\ \sum_{j=0}^{\infty} a_j^2 &< \infty\end{aligned}$$

- (2)  $\{U_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, t \in \mathbb{Z}\}$  是与  $\{V_t\}$  正交的平稳列;
- (3) 定义  $H_\varepsilon(t) = \text{sp}\{\varepsilon_s : s \leq t\}$ ,  $H_U(t) = \text{sp}\{U_s : s \leq t\}$ , 则  $\forall t$

$$H_U(t) = H_\varepsilon(t)$$

- (4)  $\{U_t\}$  是纯非决定性的平稳序列, 有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} a_j e^{ij\lambda} \right|^2$$

- (5)  $\{V_t\}$  是决定性的平稳序列。对任何  $t \in \mathbb{Z}$ ,  $k \in \mathbb{N}$ ,  $V_t \in H_{t-k}$ 。

### Wold表示定理的意义

- **定义2.3** 在Wold表示定理中
  - (1) 称(2.16)是  $\{X_t\}$  的Wold表示;
  - (2) 称  $\{U_t\}$  是  $\{X_t\}$  的纯非决定性部分, 称  $\{V_t\}$  是  $\{X_t\}$  的决定性部分;
  - (3) 称  $\{a_j\}$  是  $\{X_t\}$  的Wold系数;
  - (4) 称一步预测误差  $\varepsilon_t = X_t - L(X_t | X_{t-1}, X_{t-2}, \dots)$  为  $\{X_t\}$  的新息序列;
  - (5) 称  $\sigma^2 = E\varepsilon_t^2$  为一步(线性)预测的均方误差。
- 这里新息的意思是完全不能被历史线性预测。
- 由Wold定理可知任何纯非决定性平稳序列可以表达为新息的单边滑动和。
- 事实上, 任何白噪声的单边滑动和(系数平方可和)一定是纯非决定性的, 但其中的白噪声不一定恰好是新息。

### ARMA序列的Wold表示

- **例2.3** 设  $\{X_t\}$  是  $\text{ARMA}(p, q)$  序列, 模型方程为

$$A(\mathcal{B})X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z},$$

- 设  $A^{-1}(z)B(z)$  有Taylor展开式

$$\Psi(z) = A^{-1}(z)B(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1$$

- 则

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \quad (2.19)$$

- 下面证明 $\{X_t\}$ 是纯非决定性的平稳序列, (2.19)是 $\{X_t\}$ 的Wold表示,  $\{\varepsilon_t\}$ 是 $\{X_t\}$ 的新息序列,  $\{\psi_j\}$ 是 $\{X_t\}$ 的Wold系数。
- 我们只对比较容易的可逆ARMA的情况证明。由(2.19)看出 $X_t \in H_\varepsilon(t)$ , 利用可逆性,  $\varepsilon_t = B^{-1}(\mathcal{B})A(\mathcal{B})X_t \in H_t$ , 所以 $H_t = H_\varepsilon(t)$ 。
- 于是

$$X_t - \varepsilon_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \in H_\varepsilon(t-1) = H_{t-1}$$

来证 $X_t - \varepsilon_t = L(X_t|H_{t-1})$ 。只要证明 $X_t - (X_t - \varepsilon_t) \perp H_{t-1}$ 。

- 事实上, 由于 $\varepsilon_t$ 与 $\varepsilon_{t-j}, j \geq 1$ 正交可知 $\varepsilon_t \perp H_\varepsilon(t-1) = H_{t-1}$ 。故

$$\begin{aligned} X_t - \varepsilon_t &= L(X_t|H_{t-1}) \\ \varepsilon_t &= X_t - L(X_t|H_{t-1}) \end{aligned}$$

即 $\{\varepsilon_t\}$ 是 $\{X_t\}$ 的新息列,  $\sigma^2 = E\varepsilon_t^2$  是一步预测均方误差, 在(2.19)两边同乘以 $\varepsilon_{t-j}$ 后取期望, 利用内积的连续性可得

$$\psi_j = \langle X_t, \varepsilon_{t-j} / \sigma \rangle$$

即 $\{\psi_j\}$ 是 $\{X_t\}$ 的Wold系数列。

### Wold定理的证明—I

- 取 $\varepsilon_t = X_t - L(X_t|H_{t-1})$ ,  $H_\varepsilon(t) = \bar{sp}\{\varepsilon_s : s \leq t\}$ , 易见 $\varepsilon_t \in H_t$ , 由 $H_t$ 的单调性可知 $\varepsilon_t \in H_s, \forall t < s$ , 因此 $H_\varepsilon(t) \subset H_t$ 。
- 来证明 $\{\varepsilon_t\}$ 是白噪声。
- 由定理2.4

$$\begin{aligned} L(X_t|H_{t-1}) &= \lim_{m \rightarrow \infty} L(X_t|X_{t-1}, \dots, X_{t-m}) \quad (L^2) \\ &= \lim_{m \rightarrow \infty} \mathbf{a}_m^T \mathbf{X}_{t-1,m}^T \end{aligned}$$

其中 $\mathbf{a}_m$ 是 $m$ 阶Y-W系数(预测方程的解), 不依赖于 $t$ ,

- 由内积的连续性

$$\begin{aligned} E\varepsilon_t^2 &= \lim_{m \rightarrow \infty} \|X_t - L(X_t|X_{t-1}, \dots, X_{t-m})\|^2 \\ &= \lim_{m \rightarrow \infty} (\gamma_0 - \mathbf{a}_m^T \Gamma_m \mathbf{a}_m) \quad (\text{与 } t \text{ 无关}) \\ &= \lim_{m \rightarrow \infty} \sigma_{1,m}^2 = \sigma^2 > 0 \quad (\text{由非决定性定义}) \end{aligned}$$

- 对  $s > t$ ,  $\varepsilon_s \perp H_{s-1} \supset H_t$  所以  $\varepsilon_s \perp \varepsilon_t, s > t$ 。即

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2), \quad \sigma^2 > 0$$

### Wold定理的证明—II

- 定义  $V_t = X_t - L(X_t|H_\varepsilon(t))$ , 则  $V_t \in H_t$ 。来证明  $\{\varepsilon_t\}$  和  $\{V_t\}$  正交。
- 由投影性质,  $V_t \perp H_\varepsilon(t)$ , 即  $V_t \perp \varepsilon_s, s \leq t$ 。
- 当  $s > t$  时, 注意  $\varepsilon_s \perp H_{s-1} \supset H_t$ , 而  $V_t \in H_t$  所以  $\varepsilon_s \perp V_t, s > t$ 。
- 于是  $\{\varepsilon_t\}$  和  $\{V_t\}$  正交。

### Wold定理的证明—III

- 来证明  $\{\varepsilon_t\}$  和  $\{X_t\}$  平稳相关。
- 当  $s > t$  时  $\varepsilon_s \perp H_t$  所以  $\langle \varepsilon_s, X_t \rangle = 0, s > t$ 。
- 当  $s \leq t$  时, 由定理2.4,

$$\begin{aligned} L(X_t|H_{t-1}) &= \lim_{m \rightarrow \infty} L(X_t|\mathbf{X}_{t-1,m}) \quad (L^2) \\ &= \lim_{m \rightarrow \infty} \mathbf{a}_m^T \mathbf{X}_{t-1,m} \end{aligned}$$

- 由内积的连续性,  $s \leq t$  时

$$\begin{aligned} \langle X_t, \varepsilon_s \rangle &= \lim_{m \rightarrow \infty} \langle X_t, X_s - \mathbf{a}_m^T \mathbf{X}_{t-1,m} \rangle \\ &= \gamma_{t-s} - \lim_{m \rightarrow \infty} (a_{m1}\gamma_{t-s+1} + \dots + a_{mm}\gamma_{t-s+m}) \end{aligned}$$

只依赖于  $t - s$ 。

- 所以  $\{\varepsilon_t\}$  和  $\{X_t\}$  平稳相关。

## Wold定理的证明—IV

- 由III的证明, 可以定义

$$a_j = \langle X_t, \varepsilon_{t-j} \rangle / \sigma^2 \quad (j \geq 0)$$

$a_j$ 与 $t$ 无关。且

$$\begin{aligned} a_0 &= \langle X_t, \varepsilon_t \rangle / \sigma^2 \\ &= \langle \varepsilon_t + L(X_t | H_{t-1}), \varepsilon_t \rangle / \sigma^2 \\ &= \langle \varepsilon_t, \varepsilon_t \rangle / \sigma^2 = 1 \end{aligned}$$

- 令 $U_t = L(X_t | H_{\varepsilon}(t))$ , 则 $V_t = X_t - L(X_t | H_{\varepsilon}(t)) = X_t - U_t$ ,  $X_t = U_t + V_t$ 。
- 来证明

$$U_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad (*)$$

- 定义 $U_{t,n} = L(X_t | \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-n})$ . 设 $U_{t,n} = \sum_{j=0}^n b_j \varepsilon_{t-j}$ , 则
- 对 $j = 0, 1, \dots, n$

$$\begin{aligned} \sigma^2 a_j &= \langle X_t, \varepsilon_{t-j} \rangle = \langle U_{t,n} + (X_t - U_{t,n}), \varepsilon_{t-j} \rangle \\ &= \langle U_{t,n}, \varepsilon_{t-j} \rangle = \sigma^2 b_j \end{aligned}$$

即 $b_j = a_j$ ,

$$U_{t,n} = L(X_t | \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-n}) = \sum_{j=0}^n a_j \varepsilon_{t-j}$$

- 注意 $U_{t,n}$ 是 $X_t$ 的投影所以 $\|U_{t,n}\|^2 \leq \|X_t\|^2 = \gamma_0$ , 所以

$$\|U_{t,n}\|^2 = \sigma^2 \sum_{j=0}^n a_j^2 \leq \gamma_0 < \infty$$

- 故 $\sum_{j=0}^{\infty} a_j^2 < \infty$ ,  $\sum_{j=0}^n a_j \varepsilon_{t-j}$ 均方收敛到 $\sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ 。
- 由定理2.4知 $U_{t,n}$ 均方收敛到 $U_t = L(X_t | H_{\varepsilon}(t))$ , 所以

$$U_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad (*)$$

- $\{U_t\}$ 是线性平稳列, 其谱密度立即可得(结论(4)的第二部分)。另外 $V_t = X_t - U_t$ 也是平稳列。
- 由于 $\{\varepsilon_t\}$ 与 $\{V_t\}$ 正交所以 $V_s \perp H_{\varepsilon}(t), \forall s, t \in \mathbb{Z}$ , 而 $U_t \in H_{\varepsilon}(t)$ 所以 $V_s \perp U_t$ ,  $\{V_t\}$ 与 $\{U_t\}$ 正交。
- 至此定理的(1)(2)已证明。

## Wold定理的证明—V

- 来证明第(3)条结论。定义 $H_U(t) = \text{sp}\{U_s : s \leq t\}$ , 来证明 $H_U(t) = H_\varepsilon(t)$ 。
- 显然 $U_t \in H_\varepsilon(t)$ 所以 $H_U(t) \subset H_\varepsilon(t)$ 。只要证明 $H_\varepsilon(t) \subset H_U(t)$ 。
- 注意 $\varepsilon_t \in H_t \subset \text{sp}\{U_s, V_s : s \leq t\}$ , 由引理2.2, 存在 $\xi_m \in L_{\{V_t, V_{t-1}, \dots, V_{t-m}\}}$ ,  $\eta_m \in L_{\{U_t, U_{t-1}, \dots, U_{t-m}\}}$ , 使

$$\|\xi_m + \eta_m - \varepsilon_t\|^2 \rightarrow 0$$

- 但前面已证明 $\{V_t\}$ 与 $\{\varepsilon_t\}$ 正交, 也与 $\{U_t\}$ 正交, 所以

$$\begin{aligned} & \|\xi_m + \eta_m - \varepsilon_t\|^2 \\ &= \|\xi_m\|^2 + \|\eta_m - \varepsilon_t\|^2 \\ &\geq \|\eta_m - \varepsilon_t\|^2 \end{aligned}$$

- 令 $m \rightarrow \infty$ 得 $\|\eta_m - \varepsilon_t\|^2 \rightarrow 0$ , 由引理2.2知 $\varepsilon_t \in H_U(t)$ 。
- 所以 $H_\varepsilon(t) \subset H_U(t)$ ,  $H_\varepsilon(t) = H_U(t)$ 。结论(3)证毕。

## Wold定理的证明—VI

- 来证明 $\{U_t\}$ 是纯非决定性的(结论(4))。
- 利用定理1.2(4)(5)

$$\begin{aligned} & L(U_{t+k}|H_U(t)) = L(U_{t+k}|H_\varepsilon(t)) \\ &= L\left[\sum_{j=0}^{k-1} a_j \varepsilon_{t+k-j} + \sum_{j=k}^{\infty} a_j \varepsilon_{t+k-j} | H_\varepsilon(t)\right] \\ &= \sum_{j=k}^{\infty} a_j \varepsilon_{t+k-j} \end{aligned}$$

- 于是

$$\begin{aligned} & \|U_{t+k} - L(U_{t+k}|H_U(t))\|^2 = \left\| \sum_{j=0}^{k-1} a_j \varepsilon_{t+k-j} \right\|^2 \\ &= \sigma^2 \sum_{j=0}^{k-1} a_j^2 \rightarrow \sigma^2 \sum_{j=0}^{\infty} a_j^2 = EU_t^2 \end{aligned}$$

- 按定义可知 $\{U_t\}$ 为纯非决定性的。
- 注意: 这个证明对一般单边线性序列不适用。

## Wold 定理的证明—VII

- 已证明  $\{V_t\}$  平稳, 来证明  $\{V_t\}$  是决定性的。用定理 2.5。
- 注意  $\varepsilon_{t-j} \in H_{t-j}$ ,

$$\begin{aligned} V_t &= X_t - U_t = X_t - \varepsilon_t - \sum_{j=1}^{\infty} a_j \varepsilon_{t-j} \\ &= L(X_t | H_{t-1}) - \sum_{j=1}^{\infty} a_j \varepsilon_{t-j} \end{aligned}$$

- 其中  $L(X_t | H_{t-1}) \in H_{t-1}$ ,  $\sum_{j=1}^{\infty} a_j \varepsilon_{t-j} \in H_{\varepsilon}(t-1) \subset H_{t-1}$ , 所以  $V_t \in H_{t-1}$ 。
- 注意  $H_{t-1} \subset \text{s}\bar{\text{p}}\{U_s, V_s : s \leq t-1\}$ , 由引理 2.2, 存在  $\xi_m \in L_{\{V_{t-1}, \dots, V_{t-m}\}}$ ,  $\eta_m \in L_{\{U_{t-1}, \dots, U_{t-m}\}}$ , 使

$$\|\xi_m + \eta_m - V_t\|^2 \rightarrow 0$$

- 已证明  $\{V_t\}$  与  $\{U_t\}$  正交所以  $\eta_m$  与  $\xi_m - V_t$  正交, 于是

$$\begin{aligned} \|\xi_m + \eta_m - V_t\|^2 &= \|\xi_m - V_t\|^2 + \|\eta_m\|^2 \\ &\geq \|\xi_m - V_t\|^2 \end{aligned}$$

- 故  $\|\xi_m - V_t\|^2 \rightarrow 0 (m \rightarrow \infty)$ , 由引理 2.2 知  $V_t \in H_V(t-1) = \text{s}\bar{\text{p}}\{V_s : s \leq t-1\}$ 。
- 由定理 2.5 知  $\{V_t\}$  为决定性的, 且  $H_V(t) = H_V(t-j), j \in \mathbb{Z}$ , 所以  $V_t \in H_V(t-j), t, j \in \mathbb{Z}$ 。

## 关于新息的讨论

- $\varepsilon_t = X_t - L(X_t | H_{t-1})$  是  $X_t$  提供的比  $X_{t-1}, X_{t-2}, \dots$  多的信息(线性意义下)。
- 可以证明

$$H_t = \text{s}\bar{\text{p}}(\varepsilon_t) \oplus H_{t-1}$$

- 这样

$$H_t = \bigoplus_{j=0}^{\infty} \text{s}\bar{\text{p}}(\varepsilon_{t-j}) \oplus H_{-\infty}$$

其中  $H_{-\infty} = \bigcap_t H_t$ 。  $V_t \in H_{-\infty}$ 。

- 事实上, 等式右侧两项正交且都属于 $H_t$ , 记 $\tilde{H}_t = \text{sp}(\{\varepsilon_t\} \cup H_{t-1})$ , 来证 $H_t \subset \tilde{H}_t$ 。
- 由引理2.2只要证明 $X_s \in \tilde{H}_t, s \leq t$ 。当 $s < t$ 时显然, 对 $X_t$ , 因为

$$X_t = \varepsilon_t + L(X_t|H_{t-1})$$

所以 $X_t$ 是 $\{\varepsilon_t\} \cup H_{t-1}$ 中元素的线性组合。

- 再来证明 $H_t = \text{sp}(\varepsilon_t) \oplus H_{t-1}$ 。
- $\forall \xi \in H_t = \tilde{H}_t$ , 设 $\xi_1 = L(\xi|\varepsilon_t)$ ,  $\xi_2 = L(\xi|H_{t-1})$ ,  $\xi_3 = \xi - \xi_1 - \xi_2$ 。
- 由引理2.2, 存在 $\alpha_m \in \mathbb{R}$ 和 $\eta_m \in H_{t-1}$ 使

$$\|\xi - (\alpha_m \varepsilon_t + \eta_m)\|^2 \rightarrow 0, \quad m \rightarrow \infty$$

- 但

$$\begin{aligned} & \|\xi - (\alpha_m \varepsilon_t + \eta_m)\|^2 \\ &= \|\xi_1 - \alpha_m \varepsilon_t\|^2 + \|\xi_2 - \eta_m\|^2 + \|\xi_3\|^2 \end{aligned}$$

令 $m \rightarrow \infty$ 即知 $\xi_3 = 0$ ,  $\xi = \xi_1 + \xi_2$ 。证毕。

### §5.2.3 Kolmogorov公式

#### 多步预报的均方误差

- 设 $\{X_t\}$ 是非决定性的平稳列, 由Wold表示定理(5),  $V_{t+n} \in H_t$ , 所以用无穷长历史进行的最佳线性预测为

$$\begin{aligned} L(X_{t+n}|H_t) &= L(U_{t+n}|H_t) + L(V_{t+n}|H_t) \\ &= L\left(\sum_{j=0}^{\infty} a_j \varepsilon_{t+n-j} | H_t\right) + V_{t+n} \\ &= \sum_{j=n}^{\infty} a_j \varepsilon_{t+n-j} + V_{t+n} \end{aligned} \quad (2.20)$$

- 称 $L(X_{t+n}|H_t)$ 是 $X_{t+n}$ 的 $n$ 步预报, 由Wold分解公式知预报误差为

$$X_{t+n} - L(X_{t+n}|H_t) = \sum_{j=0}^{n-1} a_j \varepsilon_{t+n-j} \quad (2.21)$$

- 预报的均方误差为

$$\sigma^2(n) = \sigma^2 \sum_{j=0}^{n-1} a_j^2 \quad (2.22)$$

- $n \rightarrow \infty$ 时 $\sigma^2(n) \rightarrow EU_t^2$ 。



## Kolmogorov公式

- **定理2.7** (Kolmogorov公式[7]) 设 $\{U_t\}$ 是非决定性平稳序列 $\{X_t\}$ 的纯非决定性部分,  $f(\lambda)$ 是 $\{U_t\}$ 的谱密度. 则有

$$\sigma^2 = E[X_t - L(X_t|H_{t-1})]^2 = 2\pi \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln f(\lambda) d\lambda\right). \quad (2.24)$$

- 公式(2.24)的证明需要较多解析函数的知识.
- 当 $\{U_t\}$ 是白噪声时, 公式(2.24)明显是成立的.
- 从Kolmogorov公式(2.24)看到, 如果 $\{X_t\}$ 是非决定性的, 则它的纯非决定性部分的谱密度 $f(\lambda)$ 必是 $\ln$ 可积的, 即

$$\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty. \quad (2.25)$$

## §5.2.4 最佳预测和最佳线性预测相等的条件

## 最佳预测

- 设 $\{X_t\}$ 是平稳序列, 用 $\mathcal{F}_t = \sigma\{X_t, X_{t-1}, \dots\}$ 表示由 $X_t, X_{t-1}, \dots$ 生成的 $\sigma$ -代数. 称条件数学期望

$$E(X_{t+k}|\mathcal{F}_t)$$

是用全体历史 $\{X_j : j \leq t\}$ 对 $X_{t+k}$ 进行预测时的**最佳预测**.

- 最佳预测是均方误差最小的, 这是因为条件数学期望 $E(X_{t+k}|\mathcal{F}_t)$ 是 $X_{t-1}, X_{t-2}, \dots$ 的函数, 二阶矩有限:

$$E[E(X_{t+k}|\mathcal{F}_t)]^2 \leq E[E(X_{t+k}^2|\mathcal{F}_t)] = EX_{t+k}^2 < \infty.$$

由概率论中数学期望性质可以证明对任意二阶矩有限的 $X_t, X_{t-1}, \dots$ 的函数 $\xi$ 有

$$E(X_{t+k} - \xi)^2 \geq E[X_{t+k} - E(X_{t+k}|\mathcal{F}_t)]^2 \quad (2.26)$$

## 最佳预测和最佳线性预测相等的条件

- 最佳预测一般比最佳线性预测好, 但是对纯非决定性序列如果其新息是独立序列则二者等价.
- **定理2.8** 设平稳序列 $\{X_t\}$ 有Wold表示

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \quad (2.27)$$

则

$$L(X_{t+n}|H_t) = E(X_{t+n}|\mathcal{F}_t), \quad n \geq 1, \quad t \in \mathbb{Z}, \quad (2.28)$$

成立的充分必要条件是

$$E(\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \dots) = 0, \quad t \in \mathbb{Z}. \quad (2.29)$$

- (2.29)的条件称为鞅差。

### 关于ARMA模型的推论

- **推论2.9** 设ARMA( $p, q$ )序列 $\{X_t\}$ 在例2.3中定义. 如果 $\{\varepsilon_t\}$ 是独立白噪声, 则用全体历史 $\{X_t, X_{t-1}, \dots\}$ 对 $X_{t+n}$ 进行预测时, 最佳预测和最佳线性预测相等.

### §5.3 时间序列的递推预测

#### §5.3.1 时间序列的递推预测

##### 时间序列的递推预测

- 第二章已经用Y-W方程给出了非决定性平稳序列的预测公式, 这里我们进行更深入的讨论。
- 非决定性平稳序列可以用Levinson递推得到预测系数和预测方差。
- 对AR( $p$ )模型只要使用自回归系数。
- 对MA模型和ARMA模型则只能递推得到系数。
- 这里研究可以化简MA模型和ARMA模型预测的公式, 以及非平稳时也可用的递推公式。
- 假设自协方差函数已知。
- 实际中可以用样本自协方差函数代替。

##### 递推预测的正交分解

- 设 $\{Y_t\}$ 是方差有限的零均值时间序列, 对任何正整数 $n$ , 用

$$L_n = \overline{\text{sp}}\{Y_1, Y_2, \dots, Y_n\}$$

表示 $Y_1, \dots, Y_n$ 的线性组合的全体.

- 定义 $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)^T$ ,

$$\hat{Y}_1 = 0, \hat{Y}_n = L(Y_n | \mathbf{Y}_{n-1}), n = 2, \dots \quad (3.1)$$

- 引入预测误差 $W_n$ 及其方差 $\nu_{n-1}$ 如下:

$$W_n = Y_n - \hat{Y}_n, \nu_{n-1} = EW_n^2. n = 1, 2, \dots \quad (3.2)$$

- 由最佳线性预测的性质7知道 $W_n$ 和 $L_{n-1}$ 中的任何随机变量正交, 并且 $W_n \in L_n$ .
- 于是 $\{W_n\}$ 是一个正交序列, 满足

$$E(W_n W_k) = \nu_{n-1} \delta_{n-k}.$$

这里 $\delta_t$ 是Kronecker 函数.

- 用

$$M_n \triangleq \overline{\text{sp}}\{W_1, W_2, \dots, W_n\}$$

表示  $W_1, W_2, \dots, W_n$  的线性组合全体.

- 则  $M_n \subset L_n$ .
- 对  $n \in \mathbb{N}$ , 我们用归纳法证明  $Y_n \in M_n$ .
- 首先  $Y_1 = W_1 \in M_1$ .
- 如果对  $k \leq n$  已经证明  $Y_k \in M_k$ , 注意  $\hat{Y}_{n+1} = L(Y_{n+1} | \mathbf{Y}_n) \in M_n$ , 于是

$$Y_{n+1} = (Y_{n+1} - \hat{Y}_{n+1}) + \hat{Y}_{n+1} = W_{n+1} + \hat{Y}_{n+1} \in M_{n+1}.$$

这就证明了对  $n \in \mathbb{N}$ ,  $Y_n \in M_n$  成立.

- 于是得到

$$\begin{aligned} L_n &= \overline{\text{sp}}\{Y_1, Y_2, \dots, Y_n\} \\ &= \overline{\text{sp}}\{W_1, W_2, \dots, W_n\} = M_n, \\ n &= 1, 2, \dots \end{aligned} \quad (3.3)$$

- §5.1 的性质10和(3.3)式告诉我们用  $\mathbf{W}_n = (W_1, W_2, \dots, W_n)^T$  对  $Y_{n+1}$  进行预测和用  $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)^T$  对  $Y_{n+1}$  进行预测是等价的.
- 由于  $\{W_t\}$  是正交序列, 所以用  $\mathbf{W}_n$  对  $Y_{n+1}$  进行预测有很多的方便.
- 类似于在正交基上的投影, 可以直接计算坐标.

### 时间序列递推预测的定理

#### 定理3.1

- 设  $\{Y_t\}$  是零均值时间序列(不要求平稳!). 如果  $(Y_1, Y_2, \dots, Y_{m+1})^T$  的协方差矩阵

$$\left( E(Y_s Y_t) \right)_{1 \leq s, t \leq m+1} \quad (3.4)$$

正定,

- 则最佳线性预测

$$\begin{aligned} \hat{Y}_{n+1} &\triangleq L(Y_{n+1} | \mathbf{Y}_n), \quad n = 1, 2, \dots, m \\ &= \sum_{j=1}^n \theta_{n,j} W_{n+1-j} = \sum_{k=0}^{n-1} \theta_{n,n-k} W_{k+1} \\ &= \theta_{n,1} W_n + \theta_{n,2} W_{n-1} + \dots + \theta_{n,n} W_1 \\ &= \theta_{n,n} W_1 + \theta_{n,n-1} W_2 + \dots + \theta_{n,1} W_n \end{aligned} \quad (3.5)$$

- 其中的系数 $\{\theta_{n,j}\}$  和预测的均方误差 $\nu_n = EW_{n+1}^2$ 满足如下的递推公式.

•

$$\begin{cases} \nu_0 = EY_1^2, \\ \theta_{n,n-k} = [E(Y_{n+1}Y_{k+1}) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j]/\nu_k, \\ 0 \leq k \leq n-1, \\ \nu_n = EY_{n+1}^2 - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \end{cases} \quad (3.6)$$

其中 $\sum_{j=0}^{-1}(\cdot) \triangleq 0$ ,

- 递推的顺序是

$$\begin{array}{ccccccc} \nu_0 & & & & & & \\ \theta_{1,1} & \nu_1 & & & & & \\ \theta_{2,2} & \theta_{2,1} & \nu_2 & & & & \\ \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & \nu_3 & & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \end{array}$$

- 注意: 可以递推定义 $W_1 = Y_1$ ,  $W_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$ ,  $n = 1, 2, \dots$ 。

### 递推预测定理的证明

- **证明** 从自协方差矩阵(3.4)的正定性知道 $\nu_n = EW_{n+1}^2 > 0$ . 以下设 $0 \leq k \leq n-1$ .
- 在(3.5)两边同乘 $W_{k+1}$ 后求数学期望得到

$$\begin{aligned} E(\hat{Y}_{n+1}W_{k+1}) &= \sum_{j=1}^n \theta_{n,j} E(W_{n+1-j}W_{k+1}) \\ &= \theta_{n,n-k}\nu_k. \end{aligned} \quad (3.7)$$

- 注意到

$$\hat{Y}_{k+1} = \sum_{j=1}^k \theta_{k,j} W_{k+1-j} = \sum_{j=0}^{k-1} \theta_{k,k-j} W_{j+1},$$

- 再利用 $W_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$ 和 $W_{k+1}$ 垂直, 得到 $E(Y_{n+1}W_{k+1}) = E(\hat{Y}_{n+1}W_{k+1})$ ,

- 于是

$$\begin{aligned}
 \theta_{n,n-k} &= E(\hat{Y}_{n+1} W_{k+1}) / \nu_k \quad [\text{利用(3.7)}] \\
 &= E(Y_{n+1} W_{k+1}) / \nu_k \\
 &= E[Y_{n+1} (Y_{k+1} - \sum_{j=0}^{k-1} \theta_{k,k-j} W_{j+1})] / \nu_k \\
 &= [E(Y_{n+1} Y_{k+1}) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j] / \nu_k. \quad [\text{用(3.7)}]
 \end{aligned}$$

- 最后, 利用  $\nu_n = EW_{n+1}^2 = EY_{n+1}^2 - E\hat{Y}_{n+1}^2$  和(3.5)得到预测的均方误差公式:

$$\nu_n = EY_{n+1}^2 - \sum_{j=1}^n \theta_{n,j}^2 \nu_{n-j} = EY_{n+1}^2 - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j.$$

□

### 多步预报问题

- 下面考虑用  $\{Y_1, Y_2, \dots, Y_n\}$  预测  $Y_{n+k+1}$  的问题.
- 设  $(Y_1, Y_2, \dots, Y_{n+k+1})^T$  的自协方差矩阵正定.
- 仍记  $\hat{Y}_{n+k+1} = L(Y_{n+k+1} | \mathbf{Y}_{n+k})$ , 用  $W_j$  表示预测误差  $Y_j - L(Y_j | \mathbf{Y}_{j-1})$ ,
- 则有

$$\hat{Y}_{n+k+1} = \sum_{j=1}^{n+k} \theta_{n+k,j} W_{n+k+1-j}. \quad (3.8)$$

- 注意, 对  $j \geq 0$ ,  $W_{n+j+1}$  垂直于  $L_n$ ,  $W_{n-j} \in L_n$ . 根据§5.1 中最佳线性预测的性质4、5、8或定理1.2 得到

$$\begin{aligned}
 L(Y_{n+k+1} | \mathbf{Y}_n) &= L(\hat{Y}_{n+k+1} | \mathbf{Y}_n) \quad (\text{定理1.2(6)}) \\
 &= L\left[\sum_{j=1}^{n+k} \theta_{n+k,j} W_{n+k+1-j} \mid \mathbf{W}_n\right] \quad (\text{定理3.1}) \\
 &= L\left[\sum_{j=k+1}^{n+k} \theta_{n+k,j} W_{n+k+1-j} \mid \mathbf{W}_n\right] \quad (\text{定理1.2(5)}) \\
 &= \sum_{j=k+1}^{n+k} \theta_{n+k,j} W_{n+k+1-j}. \quad (\text{定理1.2(4)}) \quad (3.9)
 \end{aligned}$$

- 由投影的正交性, 得到预测的均方误差

$$\begin{aligned} E[Y_{n+k+1} - L(Y_{n+k+1}|\mathbf{Y}_n)]^2 &= EY_{n+k+1}^2 - E[L(Y_{n+k+1}|\mathbf{Y}_n)]^2 \\ &= EY_{n+k+1}^2 - \sum_{j=k+1}^{n+k} \theta_{n+k,j}^2 \nu_{n+k-j}, \end{aligned} \quad (3.10)$$

- 其中的系数 $\theta_{n+k,j}$ 和预测的均方误差 $\nu_{n+k-j}$ 可用递推公式(3.6)计算, 只不过因为 $Y_{n+1}, \dots, Y_{n+k}$ 未知所以 $W_{n+1}, \dots, W_{n+k}$ 不能计算。

### §5.3.2 正态时间序列的区间预测

#### 正态时间序列的区间预测

- 如果 $\{Y_t\}$ 是正态时间序列, 则 $\hat{Y}_{n+1}$ 也是最佳预测。
- $W_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$  作为 $Y_1, Y_2, \dots, Y_{n+1}$ 的线性组合服从正态分布 $N(0, \nu_n)$ 。
- 利用

$$\Pr(|Y_{n+1} - \hat{Y}_{n+1}|/\sqrt{\nu_n} \leq 1.96) = 0.95$$

可以得到

- $Y_{n+1}$ 的置信度为0.95的置信区间

$$[\hat{Y}_{n+1} - 1.96\sqrt{\nu_n}, \hat{Y}_{n+1} + 1.96\sqrt{\nu_n}].$$

### §5.3.3 平稳序列的递推预测

#### 平稳序列的递推预测

- 设 $\gamma_k = E(X_{t+k}X_t)$  是零均值平稳序列 $\{X_t\}$ 的自协方差函数,  $\Gamma_n$ 是 $\{X_t\}$ 的 $n$ 阶自协方差矩阵。
- 设 $\mathbf{X}_n = (X_1, X_2, \dots, X_n)^T$ ,  $Z_n = X_n - L(X_n|\mathbf{X}_{n-1})$ , 可以把定理3.1 改述如下。
- **推论3.2** 设 $\{X_t\}$ 是零均值平稳序列, 对任何 $n \in \mathbb{N}$ , 自协方差矩阵 $\Gamma_n$ 正定. 则最佳线性预测

$$\begin{aligned} \hat{X}_{n+1} &\triangleq L(X_{n+1}|\mathbf{X}_n) \\ &= \sum_{j=1}^n \theta_{n,j} Z_{n+1-j}, \quad n = 1, 2, \dots, \\ &= \sum_{k=0}^{n-1} \theta_{n,n-k} Z_{k+1} \end{aligned} \quad (3.11)$$

其中的系数 $\{\theta_{n,j}\}$  和预测的均方误差 $\nu_n = EZ_{n+1}^2$ 满足如下的递推公式。

•

$$\begin{cases} \nu_0 = \gamma_0 \\ \theta_{n,n-k} = [\gamma_{n-k} - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j] / \nu_k, \\ \quad 0 \leq k \leq n-1, \\ \nu_n = \gamma_0 - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \end{cases} \quad (3.12)$$

其中  $\sum_{j=0}^{-1}(\cdot) \triangleq 0$ ,

• 递推的顺序是

$$\begin{array}{ccccccc} & & & & & & \nu_0 \\ & & & & & & \theta_{1,1} & \nu_1 \\ & & & & & & \theta_{2,2} & \theta_{2,1} & \nu_2 \\ & & & & & & \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & \nu_3 \\ & & & & & & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

- 由于预测误差  $Z_n = X_n - L(X_n | \mathbf{X}_{n-1})$  和  $\mathbf{X}_{n-1}$  正交, 所以是不被  $\mathbf{X}_{n-1}$  包含的信息. 基于这个原因, 人们又称  $Z_n$  是 **样本新息**.
- 从 §5.2 的讨论知道,  $\nu_n = E[X_1 - L(X_1 | X_0, X_{-1}, \dots, X_{-n+1})]^2 \rightarrow \sigma^2$ , 当  $n \rightarrow \infty$ . 这里  $\sigma^2$  是用全体历史  $X_t, X_{t-1}, \dots$  预测  $X_{t+1}$  时的均方误差.  $\sigma^2 > 0$  表示  $\{X_t\}$  是非决定性的.



## §5.4 ARMA(p, q)序列的递推预测

### ARMA模型预测的意义

- 新息预测方法从理论上很完美, 对平稳列只需要知道自协方差函数列 $\{\gamma_k\}$ , 还可以预测非平稳列。
- 但是对样本数据协方差需要估计。注意新息预测系数中需要用到 $\gamma_n$ , 但是如果样本量不够的话估计 $\gamma_n$ 会产生很大误差, 导致预测误差很大。
- 如果我们知道序列服从ARMA模型, 就可以用比较少的自协方差估计得到模型参数估计, 用模型参数来得到预报公式, 这样的预报结果受随机误差影响比较小。
- 另外, ARMA模型中的白噪声如果是独立白噪声, 最佳线性预报还是最佳预报。这个结果虽然是对无穷长历史得到的但是样本量足够大时也近似成立。

#### §5.4.1 AR(p)序列的预测

##### AR(p)序列的一步预测

- 设 $\{X_t\}$ 满足AR(p)模型

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.1)$$

其中 $\{\varepsilon_t\}$ 是白噪声, 特征多项式 $A(z) = 1 - \sum_{j=1}^p a_j z^j \neq 0, |z| \leq 1$ .

- 考虑用 $\mathbf{X}_n = (X_1, X_2, \dots, X_n)^T$  预测 $X_{n+1}$ 的问题.
- 设 $\{\gamma_n\}$ 是 $\{X_t\}$ 的自协方差函数.
- 对于 $1 \leq n \leq p-1$ , 由最佳线性预测的性质1知道

$$\hat{X}_{n+1} = L(X_{n+1} | \mathbf{X}_n) = \gamma_n^T \Gamma_n^{-1} \mathbf{X}_n,$$

- 其中 $\Gamma_n$ 是 $\{X_t\}$ 的n阶自协方差矩阵,  $\gamma_n = (\gamma_n, \gamma_{n-1}, \dots, \gamma_1)^T$ . 预测的均方误差是

$$E(X_{n+1} - \hat{X}_{n+1})^2 = \gamma_0 - \gamma_n^T \Gamma_n^{-1} \gamma_n.$$

- 可以用Levinson递推公式计算系数和均方误差。
- 对于 $n \geq p$ , 由于当 $k \geq 1, X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ 与 $\varepsilon_{t+k}$  正交, 所以对 $n \geq p$

$$\begin{aligned} \hat{X}_{n+1} &= L(\varepsilon_{n+1} + \sum_{j=1}^p a_j X_{n+1-j} | \mathbf{X}_n) \\ &= \sum_{j=1}^p a_j X_{n+1-j} \end{aligned}$$

- 由于  $X_{n+1} - \hat{X}_{n+1}$  与  $X_n, \dots, X_1$  正交, 可见  $n \geq p$  时

$$\hat{X}_{n+1} = L(X_{n+1}|X_n, X_{n-1}, \dots, X_{n-p+1}).$$

#### AR(p)序列的多步预报

- 下面考虑用  $\mathbf{X}_n$  预测  $X_{n+k}$  ( $k \geq 1$ ) 的问题.
- 对于  $n \geq p$ , 对  $k$  用归纳法容易证明(习题4.1)

$$L(X_{n+k}|\mathbf{X}_n) = L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-p+1}).$$

- 当  $n \geq p$  时, 记

$$\hat{X}_{m,n} = \begin{cases} L(X_m|\mathbf{X}_n), & m > n \\ X_m, & m \leq n \end{cases}$$

- 则

$$\begin{aligned} L(X_{n+k}|\mathbf{X}_n) &= L\left(\sum_{j=1}^p a_j X_{n+k-j} + \varepsilon_{n+k}|\mathbf{X}_n\right) \\ &= L\left(\sum_{j=1}^p a_j X_{n+k-j}|\mathbf{X}_n\right) = \sum_{j=1}^p a_j \hat{X}_{n,n+k-j}, \\ k &= 1, 2, \dots \end{aligned}$$

- 可以递推计算  $\hat{X}_{m,n}, m > n$ .

#### 例4.1 AR(1)模型的预测

- 设  $X_n = a_1 X_{n-1} + \varepsilon_n$ . 则对任何  $m \in \mathbb{N}$ ,

$$\begin{aligned} L(X_{n+1}|X_n, X_{n-1}, \dots, X_{n-m+1}) &= a_1 X_n, \\ L(X_{n+2}|X_n, X_{n-1}, \dots, X_{n-m+1}) &= a_1 L(X_{n+1}|X_n) \\ &= a_1^2 X_n, \\ &\dots\dots\dots \\ L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m+1}) &= a_1^k X_n. \end{aligned}$$

- 当  $k \rightarrow \infty$ , 利用  $|a_1| < 1$  和控制收敛定理得到

$$\begin{aligned} &E[X_{n+k} - L(X_{n+k}|X_n, X_{n-1}, \dots, X_{n-m})]^2 \\ &= E(X_{n+k} - a_1^k X_n)^2 \\ &= E(X_k - a_1^k X_0)^2 \\ &\rightarrow \gamma_0 = EX_0^2, \text{ 当 } k \rightarrow \infty. \end{aligned}$$

- 说明AR(1)序列是纯非决定性的平稳序列. 实际上任何ARMA( $p, q$ )序列都是纯非决定性的(§5.2 例2.3).

#### 例4.2 降雨量预报

- 平均降雨量为 $\bar{X} = 540mm$ . 用 $X_t, t = 1, 2, \dots$ 表示该地区的逐年降雨量.
- 如果

$$Y_t = X_t - \bar{X}$$

满足AR(2)模型

$$Y_t = -0.54Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t, \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

- 给定观测 $X_1 = 560, X_2 = 470, X_3 = 580, X_4 = 496, X_5 = 576$ , 求 $X_8$ 的最佳线性预测.
- 首先中心化得,

$$Y_1 = 560 - 540 = 20, Y_2 = 470 - 540 = -70,$$

$$Y_3 = 580 - 540 = 40, Y_4 = 496 - 540 = -44,$$

$$Y_5 = 576 - 540 = 36.$$

- 用递推公式计算. 记 $\hat{Y}_j = L(Y_j | Y_1, Y_2, \dots, Y_5), 6 \leq j \leq 8$ , 则有

$$\hat{Y}_6 = -0.54 \times 36 + 0.3 \times (-44) = -32.64,$$

$$\hat{Y}_7 = -0.54\hat{Y}_6 + 0.3 \times 36 = 28.43$$

$$\hat{Y}_8 = -0.54\hat{Y}_7 + 0.3\hat{Y}_6 = -25.14$$

- 最后加上平均值得

$$\hat{X}_6 = 540 - 32.64 = 507.36$$

$$\hat{X}_7 = 540 + 28.43 = 568.43$$

$$\hat{X}_8 = 540 - 25.14 = 514.86.$$

- 在本例中, 特征多项式 $A(z) = 1 + 0.54z - 0.3z^2$ 有两个实根 $z_1 = -1.1355, z_2 = 2.9355$ . 最靠近单位圆的根的辐角是 $\pi$ , 所以序列有周期 $T = 2$ 的特性. 预测数据也体现了围绕均值540上下交替变化的特性.

§5.4.2 MA( $q$ )序列的预测MA( $q$ )序列的预测

- 设 $\{\varepsilon_t\}$ 是WN( $0, \sigma^2$ ), 实系数多项式  $B(z) = 1 + b_1z + \cdots + b_qz^q$ 在单位圆内无根:  $B(z) \neq 0, |z| < 1$ .
- 满足MA( $q$ )模型

$$X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.2)$$

的MA( $q$ )序列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$ 是 $q$ 后截尾的.

- 假设 $\sigma^2, b_1, b_2, \dots, b_q$ 已知, 我们考虑用 $X_1, X_2, \dots, X_n$ 预测 $X_{n+k}$ 的问题.
- 从(3.3)(新息预报)可以看出, 对 $n \geq 1$

$$L_n = \overline{\text{sp}}\{X_1, X_2, \dots, X_n\} = \overline{\text{sp}}\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n\}, \quad (4.3)$$

- 这里 $\hat{\varepsilon}_n = X_n - L(X_n | \mathbf{X}_{n-1})$ 是 $\{X_t\}$ 的逐步预测误差,  $\mathbf{X}_n = (X_1, \dots, X_n)^T$ .

MA( $q$ )序列的预测II

- 以下假定 $n \geq q$ .
- 由于 $\{\hat{\varepsilon}_t\}$ 是正交序列, 并由 $q$ 步截尾性可知 $X_{n+1}$ 与 $L_{n-q}$ 正交, 所以 $X_{n+1}$ 与 $\hat{\varepsilon}_{n-q}, \hat{\varepsilon}_{n-q-1}, \dots, \hat{\varepsilon}_1$ 正交, 根据最佳线性预测的性质6、4、10得到

$$\begin{aligned} L(X_{n+1} | \mathbf{X}_n) &= L(X_{n+1} | \hat{\varepsilon}_n, \dots, \hat{\varepsilon}_1) \quad (\text{性质10及(4.3)}) \\ &= L(X_{n+1} | \hat{\varepsilon}_n, \dots, \hat{\varepsilon}_{n-q+1}) + L(X_{n+1} | \hat{\varepsilon}_{n-q}, \dots, \hat{\varepsilon}_1) \quad (\text{性质6}) \\ &= L(X_{n+1} | \hat{\varepsilon}_n, \dots, \hat{\varepsilon}_{n-q+1}) \quad (\text{性质4}) \\ &= \sum_{j=1}^q \theta_{n,j} \hat{\varepsilon}_{n+1-j}. \quad (\text{新息预测公式}) \end{aligned} \quad (4.4)$$

- 预测的均方误差

$$\nu_n = E\hat{\varepsilon}_{n+1}^2 = \gamma_0 - \sum_{j=1}^q \theta_{n,j}^2 \nu_{n-j}. \quad (4.5)$$

- 这里的 $\{\theta_{n,j}\}, \nu_n$ 可以利用(3.12)进行递推计算, 但注意 $\{\gamma_k\}$ 是 $q$ 步截尾的.

§5.4.3 ARMA( $p, q$ ) 序列的预测ARMA( $p, q$ ) 序列的预测

- 对于满足ARMA( $p, q$ )模型

$$A(\mathcal{B})X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.7)$$

的ARMA序列 $\{X_t\}$ , 定义 $m = \max(p, q)$ 和(参见[14])

$$Y_t = \begin{cases} X_t/\sigma, & t = 1, 2, \dots, m, \\ A(\mathcal{B})X_t/\sigma, & t = m+1, \dots \end{cases} \quad (4.8)$$

- 则 $\{Y_t\}$ 由ARMA( $p, q$ )模型(4.7)的参数 $\beta = (a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q)^T$  和标准白噪声 $\{\varepsilon_t/\sigma\}$ 决定, 从而不依赖 $\sigma$ .
- 假设模型(4.7)中的参数已知, 我们考虑用 $X_1, X_2, \dots, X_n$ 对 $X_{n+k}$ 进行逐步预测的问题.
- 从 $Y_t$ 的定义知道, 对 $t \geq 1, Y_t \in L_t = \overline{\text{sp}}\{X_1, X_2, \dots, X_t\}$ , 并且 $X_1, X_2, \dots, X_m \in \overline{\text{sp}}\{Y_1, Y_2, \dots, Y_m\}$ .
- 容易看出, 对 $t > m$ ,

$$X_t = \sigma Y_t + \sum_{j=1}^p a_j X_{t-j} \in \overline{\text{sp}}\{Y_1, Y_2, \dots, Y_t\}.$$

- 于是再利用(3.3)(新息与原序列互相线性表示)得到

$$\begin{aligned} L_n &= \overline{\text{sp}}\{X_1, X_2, \dots, X_n\} \\ &= \overline{\text{sp}}\{Y_1, Y_2, \dots, Y_n\} = \overline{\text{sp}}\{W_1, W_2, \dots, W_n\}, \end{aligned}$$

其中 $W_t = Y_t - L(Y_t | \mathbf{Y}_{t-1})$ ,  $W_1 = Y_1$  是 $\{Y_t\}$ 的样本新息.

- 用 $\gamma_k$ 表示 $\{X_t\}$ 的自协方差函数, 取 $b_0 = 1, b_j = 0$ , 当 $j > q$ . 可以计算出

$$E(Y_s Y_t) = \begin{cases} \sigma^{-2} \gamma_{t-s}, & 1 \leq s \leq t \leq m, \\ \sigma^{-2} [\gamma_{t-s} - \sum_{j=1}^p a_j \gamma_{t-s-j}], & 1 \leq s \leq m < t, \\ \sum_{j=0}^q b_j b_{j+t-s}, & t \geq s > m, \end{cases} \quad (4.9)$$

其中的 $\{\gamma_k\}$  可用第三章的(2.10)和(2.11)计算.

- 定义 $\{X_t\}$ 的逐步预测误差

$$Z_t = X_t - L(X_t | \mathbf{X}_{t-1}), \quad Z_1 = X_1,$$

- 则对  $1 \leq t \leq m$ ,

$$\begin{aligned} W_t &= X_t/\sigma - L(X_t/\sigma | \mathbf{X}_{t-1}) \\ &= \sigma^{-1}[X_t - L(X_t | \mathbf{X}_{t-1})] = \sigma^{-1}Z_t \end{aligned}$$

- 对  $t \geq m+1$ ,

$$\begin{aligned} W_t &= \sigma^{-1}[X_t - \sum_{j=1}^p a_j X_{t-j} - L(X_t - \sum_{j=1}^p a_j X_{t-j} | \mathbf{X}_{t-1})] \\ &= \sigma^{-1}[X_t - L(X_t | \mathbf{X}_{t-1})] = \sigma^{-1}Z_t \end{aligned}$$

- 所以  $\{X_t\}$  和  $\{Y_t\}$  的预测误差之间总有以下的关系

$$Z_t = \sigma W_t, \quad EZ_t^2 = \sigma^2 EW_t^2, \quad t = 1, 2, \dots \quad (4.10)$$

- 以下仍用  $\nu_{t-1}$  表示  $EW_t^2$ , 就有  $EZ_t^2 = \sigma^2 \nu_{t-1}$ .
- 对于  $1 \leq n < m = \max(p, q)$ , 从逐步预测公式(3.5)得到

$$\hat{Y}_{n+1} = L(Y_{n+1} | \mathbf{Y}_n) = \sum_{j=1}^n \theta_{n,j} W_{n+1-j}.$$

于是对  $1 \leq n < m$ ,

$$\begin{aligned} \hat{X}_{n+1} &= L(X_{n+1} | \mathbf{X}_n) = L(\sigma Y_{n+1} | \mathbf{Y}_n) \\ &= \sigma \hat{Y}_{n+1} = \sum_{j=1}^n \theta_{n,j} \sigma W_{n+1-j} = \sum_{j=1}^n \theta_{n,j} Z_{n+1-j}. \end{aligned}$$

- 对于  $n \geq m$ , 利用ARMA序列的因果性即  $E(X_t \varepsilon_{t+k}) = 0, k = 1, 2, \dots$ , 得到

$$Y_{n+1} = \sigma^{-1} B(\mathcal{B}) \varepsilon_{n+1} = \sigma^{-1} \sum_{j=0}^q b_j \varepsilon_{n+1-j}$$

与  $\{X_j : 1 \leq j \leq n-q\}$  正交, 从而与

$$\overline{\text{sp}}\{Y_j : 1 \leq j \leq n-q\} = \overline{\text{sp}}\{W_j : 1 \leq j \leq n-q\}$$

中的任何随机变量正交.

- 利用最佳线性预测的性质6、4得到

$$L(Y_{n+1} | \mathbf{Y}_n) = \sum_{j=1}^q \theta_{n,j} W_{n+1-j}, \quad n \geq m = \max(p, q).$$

- 于是对  $n \geq m$ , 由  $Y$  和  $X$  的关系得到

$$\begin{aligned}\hat{X}_{n+1} &= L\left(\sum_{j=1}^p a_j X_{n+1-j} + \sigma Y_{n+1} | \mathbf{X}_n\right) \\ &= \sum_{j=1}^p a_j X_{n+1-j} + \sigma \sum_{j=1}^q \theta_{n,j} W_{n+1-j} \\ &= \sum_{j=1}^p a_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j} Z_{n+1-j}.\end{aligned}$$

- 总结上述推导得到:

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{n,j} Z_{n+1-j}, & 1 \leq n < m, \\ \sum_{j=1}^p a_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j} Z_{n+1-j}, & n \geq m. \end{cases} \quad (4.11)$$

- 预测的均方误差仍然是  $EW_{n+1}^2 = \sigma^2 \nu_n$ . 这里的  $\{\theta_{n,j}\}$ ,  $\nu_n = EW_{n+1}^2$  可以利用(4.9)和(3.6)进行递推计算.
- $Z_{n+1} = X_{n+1} - \hat{X}_{n+1}$  也可以递推计算.
- 因为  $\{Y_t\}$  和  $\sigma$  无关, 所以  $\theta_{n,k}$ ,  $\{W_n\}$  以及  $\nu_{n-1}$  都是和  $\sigma^2$  无关的量. 它们只依赖于参数  $\mathbf{a} = (a_1, a_2, \dots, a_p)^T$  和  $\mathbf{b} = (b_1, b_2, \dots, b_q)^T$ . 这个性质在研究ARMA模型的最大似然估计时将得到应用.

### 例4.3 ARMA序列的递推预报

- 接例1.1. 已计算§3.2例2.1的ARMA(4,2)的自协方差函数  $\gamma_0, \gamma_1, \dots, \gamma_{20}$ , 由此用(4.9)计算出  $E(Y_t Y_s)$ ,  $(1 \leq s, t \leq 21)$ .
- 用§4.3的公式(3.6)计算出新息预报系数  $\theta_{n,j}$  如下:

$n$	1	2	3	4	...	19	20
$\theta_{n,1}$	-0.226	-0.4017	-0.5705	0.1807	...	0.4875	0.489
$\theta_{n,2}$	0	-0.6865	-0.6353	-0.1597	...	-0.3937	-0.394
$\theta_{n,3}$	0	0	0.3699	-0.0000	...	-0.0001	0.000
$\theta_{n,4}$	0	0	0	-0.0000	...	0.0000	-0.000

- 该模型的观测数据  $x_1, \dots, x_{21}$  在例1.1中给出. 利用公式(4.11)和(3.6)可以计算出逐步预测  $\hat{X}_{k+1} = L(X_{k+1} | \mathbf{X}_k)$  和逐步预测的均方误差  $\nu_{k-1} = EW_k^2$  如下:

$j$	$\hat{X}_j$	$\nu_{j-1}$
1	0	6.670
2	0.104	6.330
3	0.070	2.505
4	-1.654	2.387
5	0.232	1.268
6	5.385	1.233
7	-1.788	1.142
8	-4.398	1.114
9	-0.837	1.086
10	0.839	1.069
11	2.259	1.056
12	-1.395	1.046
13	-2.354	1.038
14	0.467	1.031
15	2.585	1.026
16	1.069	1.022
17	-1.668	1.018
18	0.161	1.016
19	5.725	1.013
20	-0.229	1.011
21	-3.468	1.010

- 从以上数据看出 $\nu_k$ 收敛到 $\sigma^2 = 1$ 的速度是较理想的(参见习题4.4).
- 图5.4.1是观测 $x_t$ (虚线)和预测 $\hat{X}_t$ (实线)的数据图. 从图5.4.1看到逐步预测 $\hat{X}_t$ 可以理想地预测 $X_t$ 的走势, 这是和该ARMA序列的明显周期性有关的. 由于 $Z_k = X_k - \hat{X}_k$ 服从正态分布 $(0, \nu_{k-1})$ , 所以真值 $x_t$ 的置信度为0.95的置信下、上限分别是

$$\hat{X}_k - 1.96\sqrt{\nu_{k-1}}, \quad \hat{X}_k + 1.96\sqrt{\nu_{k-1}}.$$

图5.4.2的从上到下三条曲线分别是 $\{x_t\}$ 的置信上限,  $\{x_t\}$ 本身(虚线)和置信下限. 置信区间的长度 $3.92\sqrt{\nu_{k-1}}$ 随 $k$ 增加而减少, 最后稳定在3.92附近.

#### §5.4.4 ARMA( $p, q$ )序列多步预测

##### ARMA( $p, q$ )序列多步预测

- 设 $n > m$ ,

$$L(X_{n+k+1}|\mathbf{X}_n) = \begin{cases} \sum_{j=1}^p a_j L(X_{n+k+1-j}|\mathbf{X}_n) \\ + \sum_{j=k+1}^q \theta_{n+k,j} Z_{n+k+1-j}, & 0 \leq k < q, n > m \\ \sum_{j=1}^p a_j L(X_{n+k+1-j}|\mathbf{X}_n), & k \geq q, n > m. \end{cases}$$

#### §5.4.5 ARIMA序列预测

##### ARIMA序列预测

- 模型

$$A(\mathcal{B})(1 - \mathcal{B})^d X_t = B(\mathcal{B})\varepsilon_t, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2), \quad t \in N.$$



- 则  $Y_t = (1 - \mathcal{B})^d X_t$ ,  $t = d+1, d+2, \dots, n$  满足ARMA( $p, q$ )模型

$$A(\mathcal{B})Y_t = B(\mathcal{B})\varepsilon_t, \quad t \in N.$$

- 考虑用  $X_1, X_2, \dots, X_n$  预测  $X_{n+k}$  的问题.
- 首先利用ARMA序列的预测方法得到用  $Y_{d+1}, Y_{d+2}, \dots, Y_n$  预测

$$Y_{n+1}, Y_{n+2}, \dots, Y_{n+k}$$

时的最佳线性预测

$$\tilde{Y}_{n+j} = L(Y_{n+j} | Y_{d+1}, Y_{d+2}, \dots, Y_n), \quad j = 1, 2, \dots, k.$$

- 再由公式

$$(1 - \mathcal{B})^d \hat{X}_t = \tilde{Y}_t, \quad t = n+1, n+2, \dots, n+k$$

- 得到近似的递推公式:

$$\hat{X}_{n+k} = \tilde{Y}_{n+k} - \sum_{j=1}^d C_d^j (-1)^j \hat{X}_{n+k-j}, \quad k \geq 1,$$

其中  $\hat{X}_{n-j} = X_{n-j}$ , 当  $j \geq 0$ .



## 第六章 ARMA模型的参数估计

### §6.1 AR( $p$ )模型的参数估计

#### ARMA模型参数估计

- 对时间序列观测数据建立模型可以更深入了解数据的内在规律，也有利于预报。
- 估计模型的一般指导原则是其他条件相同时较简单的模型更有利。
- 建立模型，要兼顾对观测到的数据的拟合情况和模型的可推广性。
- 下面讨论相合性时，由§1.5定理5.2可知，ARMA模型中白噪声项如果是独立同分布的则该平稳列为严平稳遍历。

#### §6.1.1 Yule-Walker估计

##### Yule-Walker估计

- 对AR( $p$ )模型，需要估计系数 $a_1, a_2, \dots, a_p$ 和白噪声方差 $\sigma^2$ 。
- 设 $x_1, x_2, \dots, x_N$ 为观测数据，建模前一般先中心化：

$$y_t = x_t - \bar{x}_N, \quad t = 1, 2, \dots, N$$

- 假定 $\{y_1, y_2, \dots, y_N\}$ 是某个AR( $p$ )序列的观测样本。先假定 $p$ 已知。
- 如果序列的理论自协方差函数 $\{\gamma_k\}$ 已知则由Yule-Walker方程可以解出 $a_1, \dots, a_p$ 和 $\sigma^2$ ，见P.68定理3.3。
- 从观测样本可以估计样本自协方差 $\hat{\gamma}_k, k = 0, 1, \dots, p$ 。
- 在Y-W方程中用 $\hat{\gamma}_k$ 代替 $\gamma_k$ 可以解出 $\hat{a}_1, \dots, \hat{a}_p, \hat{\sigma}^2$ ，由§4.2 A的讨论知如果 $y_1, \dots, y_N$ 不全相同则 $\hat{\Gamma}_{p+1}$ 正定，Y-W方程的解存在唯一。
- 由§2.4定理4.1可知这样解出的 $\hat{a}_1, \dots, \hat{a}_p$ 满足最小相位条件。
- 为了避免计算 $\hat{\Gamma}_p$ 的逆可以用Levinson递推公式解Y-W方程得到 $\hat{a}_1, \dots, \hat{a}_p$ 和 $\hat{\sigma}^2$ 。
- 这样得到的AR( $p$ )参数估计叫做Y-W估计或Levinson估计。
- 优点是保证最小相位性，计算简单。

## Y-W估计的相合性

- 由§2.3定理3.5知AR(p)的 $\Gamma_p$ 正定, 所以 $a_1, \dots, a_p, \sigma^2$ 是 $\gamma_0, \gamma_1, \dots, \gamma_p$ 的连续函数。
- 如果 $\hat{\gamma}_k$ 是 $\gamma_k$ 的强相合估计, 则 $\hat{a}_1, \dots, \hat{a}_p, \hat{\sigma}^2$ 也是相应参数的强相合估计。
- 要使 $\hat{\gamma}_k$ 强相合, 由§4.2定理2.1, 一个充分条件是白噪声 $\{\varepsilon_t\}$ 独立同分布。这时ARMA模型严平稳遍历。
- **定理1.1** 如果AR(p)模型(1.1)中的 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$ ,  $E\varepsilon_t^4 < \infty$ , 则当 $N \rightarrow \infty$ 时
  1.  $\hat{\sigma}^2 \rightarrow \sigma^2, \hat{a}_j \rightarrow a_j, \text{ a.s.}, 1 \leq j \leq p$ .
  2.  $\sqrt{N}(\hat{a}_1 - a_1, \dots, \hat{a}_p - a_p)^T$  依分布收敛到 $p$ 维正态分布 $N(0, \sigma^2 \Gamma_p^{-1})$ 。
  3.  $\sqrt{N} \sup_{1 \leq j \leq p} |\hat{a}_j - a_j| = O(\sqrt{\ln \ln N}), \text{ a.s.}, \sqrt{N}|\hat{\sigma}^2 - \sigma^2| = O(\sqrt{\ln \ln N}), \text{ a.s.}$
- 证明:
  1. 由§4.2定理2.1知 $\hat{\gamma}_k$ 强相合, 再根据 $a_k, \sigma^2$ 为 $\{\gamma_k\}$ 的连续函数知 $\hat{a}_k, \hat{\sigma}^2$ 强相合。
  2. 证明略。
  3. 证明略。
- 当定理成立时参数估计有近似正态分布, 可以用来作置信区间和假设检验。
- 设 $\sigma_{jj}$ 为 $\sigma^2 \Gamma_p^{-1}$ 的第 $j \times j$ 元素, 则 $\hat{a}_j$ 近似服从 $N(a_j, \sigma_{jj}/N)$ , 于是 $a_j$ 的近似95% 置信区间为

$$[\hat{a}_j - 1.96\sqrt{\sigma_{jj}/\sqrt{N}}, \hat{a}_j + 1.96\sqrt{\sigma_{jj}/\sqrt{N}}]$$

其中 $\sigma_{jj}$ 可以用 $\hat{\sigma} \hat{\Gamma}_p^{-1}$ 的元素 $\hat{\sigma}_{jj}$ 代替。

## §6.1.2 最小二乘估计

## 最小二乘估计

- 最小二乘估计计算简单, 不需要计算自协方差。
- 最小二乘估计使残差平方和最小:

$$\min S(a_1, \dots, a_p) \triangleq \sum_{t=p+1}^N [y_t - (a_1 y_{t-1} + \dots + a_p y_{t-p})]^2$$

最小值点 $\hat{a}_1, \dots, \hat{a}_p$ 称为参数的最小二乘估计。

- 用线性模型理论, 写出最小二乘对应的正规方程为

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdots \\ y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} \end{pmatrix}, \\ \mathbf{a} &= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad (\mathbf{X}'\mathbf{X})\mathbf{a} = \mathbf{X}'\mathbf{Y} \end{aligned} \quad (1.12)$$

- 最小二乘估计 $\tilde{a}_1, \dots, \tilde{a}_p$ 是正规方程(1.12)的解。当 $p \times p$ 矩阵 $\mathbf{X}^T \mathbf{X}$ 正定时方程有唯一解

$$\tilde{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

依概率有界

- **定义1.1** 设 $\{\xi_n\}$ 是时间序列,  $\{c_n\}$ 是非零常数列, 如果对任意 $\varepsilon > 0$ , 存在正数 $M$ , 使得

$$\sup_n P(|\xi_n| > M) \leq \varepsilon$$

就称时间序列 $\{\xi_n\}$ 是依概率有界的, 记做 $\xi_n = O_p(1)$ 。如果 $\{\xi_n/c_n\} = O_p(1)$ , 就称 $\xi_n = O_p(c_n)$ 。

- 依概率有界就是除去一个概率任意小的集合后序列有界。
- 若 $\{\xi_n/c_n\} \xrightarrow{\text{Pr}} 0$ 则称 $\xi_n = o_p(c_n)$ 。
- **引理1:** 若 $\xi_n = o_p(c_n)$ 则 $\xi_n = O_p(c_n)$ 。
- 事实上, 按依概率收敛定义,  $\forall \delta > 0, \forall \varepsilon > 0$ , 存在 $N$ 使 $n > N$ 时

$$\Pr(|\xi_n/c_n| > \delta) < \varepsilon$$

取 $M \geq \delta$ 使得

$$\Pr(|\xi_n/c_n| > M) \leq \varepsilon, \quad n = 1, 2, \dots, N$$

于是

$$\Pr(|\xi_n/c_n| > M) \leq \varepsilon, \quad n = 1, 2, \dots$$

即 $\xi_n/c_n = O_p(1)$ 。

- **引理2** 如果 $\xi_n = O_p(1)$ 而 $c_n \rightarrow \infty$ 则 $\xi_n/c_n = o_p(1)$ 。
- 对随机变量 $\xi$ ,  $\xi = O_p(1)$ .
- **引理3** 若 $\{\xi_n\}$ 同分布, 则 $\xi_n = O_p(1)$ 。
- 事实上,  $\forall \varepsilon > 0, \exists M > 0$ 使 $\Pr(|\xi_1| > M) < \varepsilon$ 。由同分布性知 $\Pr(|\xi_t| > M) = \Pr(|\xi_1| > M) < \varepsilon$ 。
- **引理4**

$$O_p(1) \pm O_p(1) = O_p(1)$$

$$O_p(1) \cdot O_p(1) = O_p(1)$$

$$O_p(1) \pm o_p(1) = O_p(1)$$

$$O_p(1) \cdot o_p(1) = o_p(1).$$

- **引理5** 若 $\xi_n$ 依概率收敛到 $\xi$ 则 $\{\xi_n\} = O_p(1)$ 。
- 事实上 $\xi_n = \xi + (\xi_n - \xi) = O_p(1) + o_p(1)$ 。
- **引理6** 若 $\xi_n = O(1)$  a.s.则 $\xi_n = O_p(1)$ 。
- 事实上,  $\xi_n = O(1)$  a.s.则存在随机变量 $\xi > 0$ 使 $P(|\xi_n| \leq \xi) = 1$ 。  $\forall \varepsilon > 0, \exists M > 0$ 使 $P(\xi > M) < \varepsilon$ , 于是 $P(|\xi_n| > M) \leq P(\xi > M) < \varepsilon, \forall n \in \mathbb{N}$ 。

### 最小二乘估计和Y-W估计的比较

- 注意对于 $1 \leq k \leq j \leq p$ , 矩阵 $(1/N)\mathbf{X}^T \mathbf{X}$ 的第 $(j, k)$ 元素是

$$\begin{aligned} \tilde{\gamma}_{jk} &= \frac{1}{N} \sum_{t=1}^{N-p} y_{t+p-j} y_{t+p-k} = \frac{1}{N} \sum_{t=p-j+1}^{N-j} y_t y_{t+j-k} \\ &= \frac{1}{N} \sum_{t=1}^{N-(j-k)} y_t y_{t+j-k} - \frac{1}{N} \sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N} \sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\ &= \hat{\gamma}_{j-k} - \frac{1}{N} \sum_{t=1}^{p-j} y_t y_{t+j-k} - \frac{1}{N} \sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}. \end{aligned}$$

这里 $\hat{\gamma}_{j-k}$  由(1.2) 定义.

- 完全类似地得到 $(1/N)\mathbf{X}^T\mathbf{Y}$ 的第 $j$ 个元素

$$\begin{aligned}\tilde{\gamma}_j &= \frac{1}{N} \sum_{t=1}^{N-p} y_{t+p-j} y_{t+p} = \frac{1}{N} \sum_{t=p-j+1}^{N-j} y_t y_{t+j} \\ &= \frac{1}{N} \sum_{t=1}^{N-j} y_t y_{t+j} - \frac{1}{N} \sum_{t=1}^{p-j} y_t y_{t+j} \\ &= \hat{\gamma}_j - \frac{1}{N} \sum_{t=1}^{p-j} y_t y_{t+j}.\end{aligned}$$

- $\sum_{t=1}^{p-j} y_t y_{t+j-k}$  和  $\sum_{t=1}^{p-j} y_t y_{t+j}$  是固定的随机变量所以是 $O_p(1)$ 的,  $\frac{1}{N} \sum_{t=1}^{p-j} y_t y_{t+j-k} = O_p(\frac{1}{N})$ ,  $\frac{1}{N} \sum_{t=1}^{p-j} y_t y_{t+j} = O_p(\frac{1}{N})$ 。
- 同时, 固定随机变量的 $\frac{1}{N}$ 是a.s.趋于零的。
- 由 $\{\varepsilon_t\}$ 独立同分布知AR( $p$ )序列 $\{Y_t\}$ 严平稳遍历, 所以 $\sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k}$ 也是严平稳遍历的, 且一阶矩存在, 由引理3知其为 $O_p(1)$ 的。
- 另一方面, 由 $\{Y_t\}$ 的严平稳遍历性则

$$\begin{aligned}& \frac{1}{N} \sum_{t=N-j+1}^{N-(j-k)} y_t y_{t+j-k} \\ &= \frac{1}{N} \sum_{j=1}^{N-(j-k)} y_t y_{t+j-k} - \frac{1}{N} \sum_{j=1}^{N-j} y_t y_{t+j-k} \\ &= \frac{N-(j-k)}{N} \frac{1}{N-(j-k)} \sum_{j=1}^{N-(j-k)} y_t y_{t+j-k} \\ & \quad - \frac{N-j}{N} \frac{1}{N-j} \sum_{j=1}^{N-j} y_t y_{t+j-k} \\ & \rightarrow 1 \cdot \gamma_{j-k} - 1 \cdot \gamma_{j-k} = 0 \quad \text{a.s.}\end{aligned}$$

- 于是, 对 $1 \leq k \leq j \leq p$ 有

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} = O_p(\frac{1}{N}), \quad \tilde{\gamma}_j - \hat{\gamma}_j = O_p(\frac{1}{N}),$$

- 且

$$\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k} \rightarrow 0, \quad \tilde{\gamma}_j \rightarrow \hat{\gamma}_j, \quad \text{a.s.},$$

- 由 $\hat{\gamma}_j$ 的强相合性知 $\tilde{\gamma}_{jk} \rightarrow \gamma_{j-k}$ ,  $\tilde{\gamma}_j \rightarrow \gamma_j$ , a.s.
- 这样, 当 $N \rightarrow \infty$

$$\begin{aligned}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} &= \left( \frac{1}{N} \mathbf{X}^T \mathbf{X} \right)^{-1} \left( \frac{1}{N} \mathbf{X}^T \mathbf{Y} \right) \\ &\rightarrow \Gamma_p^{-1} \boldsymbol{\gamma}_p = \mathbf{a}, \quad \text{a.s.}\end{aligned}$$



即最小二乘估计也是强相合的。

- 另一方面, 利用 $(\tilde{\gamma}_{jk} - \hat{\gamma}_{j-k}) = O_p(1/N)$  和 $(\tilde{\gamma}_j - \hat{\gamma}_j) = O_p(1/N)$ ,
- 用 $\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p)^T$  和 $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_p)^T$  分别表示Yule-Walker估计和最小二乘估计. 记 $\hat{\gamma}_p = (\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_p)^T$ , 就有

$$\begin{aligned}\tilde{a} - \hat{a} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} - \hat{\Gamma}_p^{-1} \hat{\gamma}_p \\ &= \left( \frac{1}{N} \mathbf{X}^T \mathbf{X} \right)^{-1} \left( \frac{1}{N} \mathbf{X}^T \mathbf{Y} - \hat{\gamma}_p \right) + \left[ \left( \frac{1}{N} \mathbf{X}^T \mathbf{X} \right)^{-1} - \hat{\Gamma}_p^{-1} \right] \hat{\gamma}_p \\ &= \left( \frac{1}{N} \mathbf{X}^T \mathbf{X} \right)^{-1} O_p(1/N) + \left( \frac{1}{N} \mathbf{X}^T \mathbf{X} \right)^{-1} \left[ \hat{\Gamma}_p - \frac{1}{N} \mathbf{X}^T \mathbf{X} \right] \hat{\Gamma}_p^{-1} \hat{\gamma}_p \\ &= O_p(1/N), \text{ 当 } N \rightarrow \infty.\end{aligned}\quad (1.14)$$

- 上式说明对 $1 \leq j \leq p$ ,  $\hat{a}_j - \tilde{a}_j = O_p(1/N)$ . 由于 $1/N$ 收敛到零的速度是很快, 所以也可以说对较大的 $N$ , 最小二乘估计和Yule-Walker 估计的差别不大.

### 最小二乘估计的极限分布

- **定理1.2** 设AR(p)模型(1.1)中的白噪声 $\{\varepsilon_t\}$ 是独立同分布的,  $E\varepsilon_t^4 < \infty$ ,  $(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_p)$  是自回归系数 $(a_1, a_2, \dots, a_p)$ 的最小二乘估计. 则当 $N \rightarrow \infty$ 时

$$\sqrt{N}(\tilde{a}_1 - a_1, \tilde{a}_2 - a_2, \dots, \tilde{a}_p - a_p)$$

依分布收敛到 $p$ -维正态分布 $N(0, \sigma^2 \Gamma_p^{-1})$ .

- 即最小二乘估计和Y-W估计有相同的渐近分布。
- 定理1.2 的证明由定理1.1和(1.14)直接得到.

### 例1.1 Y-W估计和LS估计的模拟比较

- $\{\varepsilon_t\}$ 为标准正态白噪声, 模拟产生AR(4)序列

$$\begin{aligned}X_t &= 1.16X_{t-1} - 0.37X_{t-2} - 0.11X_{t-3} \\ &\quad + 0.18X_{t-4} + \varepsilon_t, t = 1, 2, \dots, N\end{aligned}$$

- 分别取 $N = 25, 100, 400, 1600$ , 重复 $M = 10000$ 次产生模拟数据并分别用Y-W方法和最小二乘方法估计参数。
- 结果发现在样本量较小时, 系数估计Y-W方法较好, 预测方差 $\sigma^2$ 估计最小二乘方法较好. 样本量大时二者没有明显差别。

例1.1 Y-W估计和LS估计的模拟比较:  $N = 25, 100$ 

	$N = 25$	$a_1$	$a_2$	$a_3$	$a_4$	$\sigma^2$
True parameter	1.16	-0.37	-0.11	0.18	1	
Y-W avg	0.929	-0.302	-0.052	0.018	1.311	
LS avg	1.012	-0.376	-0.056	0.050	0.751	
Y-W std	0.208	0.251	0.228	0.158	0.518	
LS std	0.252	0.331	0.318	0.226	0.267	

---

	$N = 100$	$a_1$	$a_2$	$a_3$	$a_4$	$\sigma^2$
True parameter	1.16	-0.37	-0.11	0.18	1	
Y-W avg	1.100	-0.343	-0.087	0.133	1.083	
LS avg	1.133	-0.373	-0.092	0.147	0.947	
Y-W std	0.102	0.144	0.139	0.094	0.178	
LS std	0.105	0.156	0.154	0.102	0.141	

例1.1 Y-W估计和LS估计的模拟比较:  $N = 400, 1600$ 

	$N = 400$	$a_1$	$a_2$	$a_3$	$a_4$	$\sigma^2$
True parameter	1.16	-0.37	-0.11	0.18	1	
Y-W avg	1.145	-0.362	-0.104	0.169	1.022	
LS avg	1.154	-0.370	-0.106	0.173	0.988	
Y-W std	0.049	0.074	0.074	0.048	0.075	
LS std	0.050	0.076	0.076	0.049	0.071	

---

	$N = 1600$	$a_1$	$a_2$	$a_3$	$a_4$	$\sigma^2$
True parameter	1.16	-0.37	-0.11	0.18	1	
Y-W avg	1.1563	-0.3683	-0.1082	0.1767	1.0060	
LS avg	1.1586	-0.3705	-0.1086	0.1777	0.9975	
Y-W std	0.0247	0.0377	0.0377	0.0247	0.0355	
LS std	0.0247	0.0379	0.0380	0.0249	0.0350	

## §6.1.3 最大似然估计

## 最大似然估计

- Y-W估计是矩估计，最小二乘估计与Y-W估计渐近相同。矩估计容易计算但可能精度不高。
- 最大似然估计一般精度较高。
- 设AR( $p$ )模型(1.1)的白噪声

$$\varepsilon_t = X_t - \sum_{j=1}^p a_j X_{t-j}$$

服从正态分布, 则 $\varepsilon_{p+1}, \varepsilon_{p+2}, \dots, \varepsilon_N$ 有联合密度函数

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=p+1}^N \varepsilon_t^2\right). \quad (1.15)$$

- 利用(1.15)得到基于 $x_1, x_2, \dots, x_N$ 的似然函数

$$L(\mathbf{a}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=p+1}^N \left(x_t - \sum_{j=1}^p a_j x_{t-j}\right)^2\right).$$

- $L(\mathbf{a}, \sigma^2)$ 的最大值点 $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p, \hat{\sigma}^2)$ 是 $(a_1, a_2, \dots, a_p, \sigma^2)$ 最大似然估计.
- 由于实际计算中首先对数据进行零均值化, 所以对数似然函数应当定义为

$$\begin{aligned} l(\mathbf{a}, \sigma^2) &= \ln L(\mathbf{a}, \sigma^2) \\ &= -\frac{N-p}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=p+1}^N \left[y_t - \sum_{j=1}^p a_j y_{t-j}\right]^2 + c \\ &= -\frac{N-p}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} S(\mathbf{a}) + c, \end{aligned}$$

这里 $c = -[(N-p)/2] \ln(2\pi)$ 是常数,  $S(\mathbf{a}) = S(a_1, a_2, \dots, a_p)$ 按(1.9)定义.

- $l(\mathbf{a}, \sigma^2)$ 关于 $\sigma^2$ 求偏导并令偏导等于零, 得

$$\sigma^2 = \frac{1}{N-p} S(\mathbf{a}) \quad (1.16)$$

- 于是最大值点是 $S(\mathbf{a})$ 的最小值点, 最大似然估计等同于最小二乘估计。

#### §6.1.4 AR(p)模型定阶

##### AR模型定阶

- 对于时间序列观测样本, 假设它来自某个AR(p)模型, 但是 $p$ 一般是未知的。
- 要想办法估计 $p$ 。
- 估计 $p$ 可以用AR(p)的偏相关 $p$ 步截尾性质, 计算样本偏相关系数 $\{\hat{a}_{kk}\}$ , 看 $\{\hat{a}_{kk}\}$ 在何处截尾。
- 估计 $p$ 也可以在拟合优度和模型简单之间进行权衡, AIC和BIC准则就是这样的办法。

### 样本偏相关系数

- 只要观测样本  $x_1, \dots, x_N$  不完全相同则样本自协方差阵  $\hat{\Gamma}_k$  正定 ( $k < N$ )。样本偏相关系数  $\{\hat{a}_{k,k}\}$  可以由样本自协方差唯一决定

$$\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} \hat{a}_{k,1} \\ \hat{a}_{k,2} \\ \vdots \\ \hat{a}_{k,k} \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{pmatrix}$$

可以用Levinson递推公式计算样本偏相关系数。

- 定理1.3** 如果AR( $p$ )模型(1.1)中的白噪声时独立同分布的, 则对任何  $k > p$ , 有

$$\lim_{N \rightarrow \infty} \hat{a}_{k,j} = \begin{cases} a_j, & \text{当 } j \leq p, \\ 0, & \text{当 } j > p. \end{cases} \quad \text{a.s.}$$

### 定理1.3证明

- 由§4.2 的定理2.1 知道样本自协方差函数  $\hat{\gamma}_k$  是  $\gamma_k$  的强相合估计。
- 对任何矩阵  $(c_{j,k}(N))$  定义极限符号

$$\lim_{N \rightarrow \infty} (c_{j,k}(N)) \triangleq \left( \lim_{N \rightarrow \infty} c_{j,k}(N) \right).$$

- 对任何  $k > p$ , 利用  $\Gamma_k$  的正定性得到:

$$\begin{aligned} \lim_{N \rightarrow \infty} \begin{pmatrix} \hat{a}_{k,1} \\ \hat{a}_{k,2} \\ \vdots \\ \hat{a}_{k,k} \end{pmatrix} &= \lim_{N \rightarrow \infty} \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{pmatrix}^{-1} \times \lim_{N \rightarrow \infty} \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{pmatrix} \\ &= \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{pmatrix}^{-1} \times \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix} = \begin{pmatrix} a_{k,1} \\ a_{k,2} \\ \cdots \\ a_{k,k} \end{pmatrix}. \end{aligned}$$

- 从§2.3 的(3.9)知道

$$(a_{k,1}, a_{k,2}, \dots, a_{k,k}) = (a_1, a_2, \dots, a_p, 0, \dots, 0). \quad (1.19)$$

所以(1.18)成立.

## 偏相关系数的检验

- 为检验  $H_0: a_{k,k} = 0$ , 需要知道  $\hat{a}_{k,k}$  的分布。但我们只能得到  $\hat{a}_{k,k}$  的极限分布。
- **定理1.4** 设  $(a_{k,1}, a_{k,2}, \dots, a_{k,k})$  由(1.19)定义, 如果AR(p)模型(1.1)中的白噪声是独立同分布的,  $E\varepsilon_t^4 < \infty$ , 则对确定的  $k > p$ , 当  $N \rightarrow \infty$  时,

$$\sqrt{N}(\hat{a}_{k,1} - a_{k,1}, \hat{a}_{k,2} - a_{k,2}, \dots, \hat{a}_{k,k} - a_{k,k})$$

依分布收敛到  $k$ -维正态分布  $N(0, \sigma^2 \Gamma_k^{-1})$ .

- **推论1.5** 在定理1.4的条件下, 对  $k > p$ ,  $\sqrt{N}\hat{a}_{k,k}$  依分布收敛到标准正态分布  $N(0, 1)$ .

## 推论1.5证明

- 从定理1.4知道只需对  $k > p$  证明  $\sigma^2 \Gamma_k^{-1}$  的第  $(k, k)$  元素是1.
- 用  $A_{(j,j)}$  表示矩阵  $A$  的第  $(j, j)$  元素. 利用  $\Gamma_k^{-1}$  的伴随矩阵表示得到

$$\Gamma_k^{-1} = \frac{1}{\det(\Gamma_k)} \begin{pmatrix} \det(\Gamma_{k-1}) & * & \\ & * & \\ * & & \det(\Gamma_{k-1}) \end{pmatrix}.$$

- 于是

$$(\Gamma_k^{-1})_{(k,k)} = (\Gamma_k^{-1})_{(1,1)} = \frac{\det(\Gamma_{k-1})}{\det(\Gamma_k)}, \quad \forall k \geq 1.$$

- 对  $k \geq p$ , 利用Yule-Walker 方程得到

$$\Gamma_{k+1} \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{从而有} \quad \begin{pmatrix} 1 \\ -\mathbf{a}_k \end{pmatrix} = \Gamma_{k+1}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- 于是得到

$$(\sigma^2 \Gamma_{k+1}^{-1})_{(k+1,k+1)} = (\sigma^2 \Gamma_{k+1}^{-1})_{(1,1)} = 1, \quad k \geq p.$$

## 用样本偏相关系数定阶

- $N$ 较大时若  $k > p$ , 则  $\sqrt{N}\hat{a}_{k,k}$  近似服从  $N(0,1)$ ,  $\hat{a}_{k,k}$  约以95%概率落入区间

$$(-1.96/\sqrt{N}, 1.96/\sqrt{N}).$$

- 对某固定的 $K$ , 以

$$\hat{p} = \max\{j : |\hat{a}_{j,j}| > 1.96/\sqrt{N}, 1 \leq j \leq K\}$$

作为 $p$ 的估计是合理的。

- 实际问题中可以取 $K$ 为允许的阶数的一个上界。
- 这种办法只要画出样本偏相关系数图就可以定阶。(R中用pacf(x)作样本偏相关系数图)。

### AIC和BIC定阶

- 可以通过对拟合优度的要求加上对参数个数的惩罚指定一个准则来定阶。
- AIC准则:

$$\text{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k = 0, 1, \dots, P_0$$

( $P_0$ 是可取的阶的上限)

- BIC定阶:

$$\text{BIC}(k) = \ln \hat{\sigma}_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \dots, P_0$$

- 取AIC或BIC的最小值点作为 $p$ 的估计。
- AIC定阶不相合, 倾向于高估; BIC定阶在 $N \rightarrow \infty$ 时是强相合的, 但对较小的 $N$ 倾向于低估。

### §6.1.5 模型拟合检验

#### 模型拟合检验

- 对观测数据 $x_1, x_2, \dots, x_N$ , 如果得到了 $AR(p)$ 的阶 $p$ 和自回归系数 $(a_1, a_2, \dots, a_p)$ 的估计量 $\hat{p}$  和 $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{\hat{p}})$ , 定义残差

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \quad t = \hat{p} + 1, \hat{p} + 2, \dots, N. \quad (1.21)$$

- 这里的 $y_t = x_t - \bar{x}_N$ 是 $x_t$ 零均值化。
- 对上述的残差序列进行白噪声的检验。
- 如果能够判定(1.21)是白噪声, 就认为建立的模型是合理的。否则可以改动 $\hat{p}$  的值后重新计算, 或改用MA( $q$ ) 或ARMA( $p, q$ ) 模型。

## §6.1.6 AR谱密度估计

## AR谱密度估计

- 满足 $AR(p)$ 模型(1.1)的 $AR(p)$ 序列有谱密度函数

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 - \sum_{j=1}^p a_j e^{ij\lambda} \right|^{-2}. \quad (1.22)$$

- 把 $\sigma^2$ ,  $p$  和 $(a_1, a_2, \dots, a_p)$ 的估计量 $\hat{\sigma}^2$ ,  $\hat{p}$  和 $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p)$  代入(1.22)后, 得到 $f(\lambda)$  的估计

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} \left| 1 - \sum_{j=1}^{\hat{p}} \hat{a}_j e^{ij\lambda} \right|^{-2}. \quad (1.23)$$

- 通常称 $\hat{f}(\lambda)$ 为AR谱估计或极大熵谱估计.
- 对于AIC或BIC定阶 $\hat{p}$ , 如果 $\{\varepsilon_t\}$ 是独立同分布的 $WN(0, \sigma^2)$ , 可以证明 $\hat{f}(\lambda)$ 一致收敛到 $f(\lambda)$ .

## §6.2 MA( $q$ )模型的参数估计

### MA(1)参数估计

- 考虑MA(1)

$$X_t = \varepsilon_t + b\varepsilon_{t-1}, \quad t = 1, 2, \dots$$

其中 $|b| < 1$ 。

- $\gamma_0 = EX_t^2 = \sigma^2(1 + b^2)$ ,  $\gamma_1 = \sigma^2 b$ .
- $\rho_1 = \frac{b}{1+b^2}$ , 关于 $b$ 的方程为

$$\rho_1 b^2 - b + \rho_1 = 0$$

- 当 $|\rho_1| \leq 0.5$ 时 $b$ 才有实解:

$$b = \frac{1 - \sqrt{1 - 4\rho_1^2}}{2\rho_1}$$

(另一个解使 $|b| > 1$ , 抛弃)

- 用 $\hat{\rho}_1 = \hat{\gamma}_1/\hat{\gamma}_0$ 代替理论值可得 $b$ 的矩估计

$$\hat{b} = \frac{1 - \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}$$

- 如果 $\{\varepsilon_t\}$ 是独立同分布的白噪声, 由§4.2 定理2.1 知道 $\hat{\rho}_1$ 是 $\rho_1$ 的强相合估计:  $\lim_{N \rightarrow \infty} \hat{\rho}_1 = \rho_1$ , a.s., 于是

•

$$\lim_{N \rightarrow \infty} \hat{b} = \frac{1 - \sqrt{1 - 4\rho_1^2}}{2\rho_1} = b, \text{ a.s.,}$$

- 所以 $\hat{b}$ 是 $b$ 的强相合估计.
- 实际上利用§4.2 定理2.2 还可以证明, 当 $N \rightarrow \infty$ 时,  $\sqrt{N}(\hat{b} - b)$  依分布收敛到正态分布([27]).

$$N \left( 0, \frac{1 + b^2 + b^4 + b^6 + b^8}{(1 - b^2)^2} \right).$$

### §6.2.1 MA( $q$ )模型的矩估计及其计算

#### 矩估计



- 考虑可逆MA( $q$ )模型:

$$X_t = \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (2.2)$$

$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$ , 系数满足可逆条件:

$$B(z) = 1 + \sum_{j=1}^q b_j z^j \neq 0, \quad |z| \leq 1, \quad (2.3)$$

- 假定 $q$ 已知, 估计 $\mathbf{b} = (b_1, \dots, b_q)^T$ 和 $\sigma^2$ 。

- 参数与自协方差函数关系为

$$\gamma_k = \sigma^2(b_0 b_k + b_1 b_{k+1} + \dots + b_{q-k} b_q), \quad 0 \leq k \leq q. \quad (2.4)$$

- 可以用解非线性方程组的方法求解 $\mathbf{b}$ 和 $\sigma^2$ , 但不能保证解唯一, 也不能保证可逆性条件。
- 求解方法包括线性迭代方法和Newton-Raphson迭代方法。
- 还可以根据§3.1 C计算矩估计。由MA( $q$ )的 $q$ 步截尾性, 可定义

$$\tilde{\gamma}_k = \begin{cases} \hat{\gamma}_k, & 0 \leq k \leq q, \\ 0, & k > q, \end{cases}$$

用 $\{\tilde{\gamma}_k\}$ 作为 $\{\gamma_k\}$ 的估计代入§3.1C 的计算公式。

- **定理2.1** 如果模型(2.2)中的 $\{\varepsilon_t\}$ 是独立同分布的 $\text{WN}(0, \sigma^2)$ , 则几乎必然地当 $N$ 充分大后由(2.6)计算的 $\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q$ 满足可逆条件(2.3)。
- **证明** 由于当 $N \rightarrow \infty, \hat{\gamma}_k \rightarrow \gamma_k$ , a.s., 所以

$$\begin{aligned} \hat{f}(\lambda) &= \frac{1}{2\pi} \sum_{k=-q}^q \hat{\gamma}_k e^{-ik\lambda} \\ &\rightarrow f(\lambda) = \frac{1}{2\pi} \sum_{k=-q}^q \gamma_k e^{-ik\lambda}, \quad \text{a.s., 当 } N \rightarrow \infty \end{aligned}$$

在 $[-\pi, \pi]$ 上一致成立. 于是当 $N$ 充分大后 $\hat{f}(\lambda)$ 恒正. 利用§3.1的引理1.2知道有惟一的 $(b'_1, b'_2, \dots, b'_q)$ 满足可逆性条件(2.3), 并且使得

$$\hat{f}(\lambda) = \frac{\sigma_0^2}{2\pi} \left| 1 + \sum_{j=1}^q b'_j e^{-ij\lambda} \right|^2$$

是

$$Y_t = e_t + \sum_{j=1}^q b'_j e_{t-j}, \quad \{e_t\} \sim \text{WN}(0, \sigma_0^2)$$

的谱密度. 这时,  $\tilde{\gamma}_k = E(Y_t Y_{t+k})$ . 再利用§3.1的C知道

$$(\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q) = (b'_1, b'_2, \dots, b'_q).$$

- 从定理2.1的证明知道这样得到的 $(\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q)$ 满足

$$1 + \sum_{j=1}^q \hat{b}_j z^j \neq 0, \quad |z| \leq 1$$

的充分条件是

$$\sum_{k=-q}^q \hat{\gamma}_k e^{-ik\lambda} > 0, \quad \lambda \in [-\pi, \pi].$$

### §6.2.2 MA(q)模型的逆相关函数法

逆相关函数

- 因为AR的Yule-Walker估计能保证最小相位性所以想到把MA变成一个AR再估计。
- 定义: 设平稳序列 $\{X_t\}$ 有恒正的谱密度 $f(\lambda)$ . 通常称

$$f_y(\lambda) = \frac{1}{4\pi^2 f(\lambda)}. \quad (2.7)$$

为 $\{X_t\}$ 的逆谱密度. 称

$$\gamma_y(k) \triangleq \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda) d\lambda$$

为 $\{X_t\}$ 的逆相关函数或逆自相关函数.

### MA(q)的逆相关函数

- 设 $\{X_t\}$ 为可逆MA(q)序列. 其谱密度为

$$f(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2$$

- 逆谱密度为

$$f_y(\lambda) = \frac{\sigma^{-2}}{2\pi} |B(e^{i\lambda})|^{-2} \quad (2.8)$$

$f_y$ 为如下AR(q)序列的谱密度

$$Y_t = - \sum_{j=1}^q b_j Y_{t-j} + e_t, \quad t \in \mathbb{Z} \quad (2.9)$$

其中 $\{e_t\} \sim \text{WN}(0, \sigma^{-2})$ 。

- $\{Y_t\}$ 的自协方差函数

$$\gamma_y(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda) d\lambda, \quad k = 0, 1, 2, \dots \quad (2.10)$$

是 $\{X_t\}$ 的逆相关函数。

- 只要能估计 $\{\gamma_y(k)\}$ , 就可以用Y-W方法估计 $\{Y_t\}$ 的参数 $b_1, \dots, b_q$ 和 $\sigma^{-2}$ , 从而得到MA(q) 序列 $\{X_t\}$ 的参数估计, 且系数满足可逆性条件。
- **引理2.2** 如果 $(a_1, a_2, \dots, a_p)$  和 $\sigma^2$ 分别是AR(p)模型

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z},$$

的自回归系数和白噪声 $\{\varepsilon_t\}$ 的方差, 则 $\{X_t\}$ 的逆相关函数为

$$\gamma_y(k) = \begin{cases} \frac{1}{\sigma^2} \sum_{j=0}^{p-k} a_j a_{j+k}, & 0 \leq k \leq p, \quad a_0 \triangleq -1, \\ 0, & k > p. \end{cases}$$

- **引理2.2证明**  $\{X_t\}$ 有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi} |A(e^{i\lambda})|^{-2}, \quad \text{其中 } A(z) = 1 - \sum_{j=1}^p a_j z^j,$$

和逆谱密度

$$f_y(\lambda) = \frac{1}{4\pi^2 f(\lambda)} = \frac{1}{2\pi\sigma^2} |A(e^{i\lambda})|^2,$$

于是有逆相关函数

$$\gamma_y(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda) d\lambda = \begin{cases} \frac{1}{\sigma^2} \sum_{j=0}^{p-k} a_j a_{j+k}, & 0 \leq k \leq p, \\ 0, & k > p. \end{cases}$$

### MA(q)的逆相关函数参数估计法

- 满足(2.2)的可逆MA(q)序列 $\{X_t\}$ 可以写成无穷阶自回归的形式

$$X_t - \sum_{j=1}^{\infty} a_j X_{t-j} = \varepsilon_t, \quad t = 1, 2, \dots, \quad (2.13)$$

这里的回归系数 $\{a_j\}$  由 $1/B(z)$  在单位圆内的Taylor级数决定:

$$\frac{1}{B(z)} = 1 - \sum_{j=1}^{\infty} a_j z^j, \quad |z| \leq 1.$$

- 由于当  $j \rightarrow \infty$ ,  $a_j$  是以负指数阶收敛到0的, 所以对较大的正整数  $p$  可以将无穷阶自回归模型(2.13)写成近似的长阶( $p$ 阶)自回归的形式

$$X_t \approx \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (2.14)$$

- 利用引理2.2 知道  $\{X_t\}$  的逆相关函数  $\gamma_y(k)$  满足

$$\gamma_y(k) \approx \frac{1}{\sigma^2} \sum_{j=0}^{p-k} a_j a_{j+k}, \quad 0 \leq k \leq p, \quad a_0 \triangleq -1. \quad (2.15)$$

- 用样本  $x_1, x_2, \dots, x_N$  计算逆相关函数  $\hat{\gamma}_y(k)$  和  $\hat{b}_1, \dots, \hat{b}_p, \hat{\sigma}^2$  的方法如下。

1. 首先利用  $\{x_t\}$  的样本自协方差函数  $\hat{\gamma}_k$  建立一个  $AR(p_N)$  模型, 这里  $p_N$  可以是  $AR$  模型的 AIC 定阶, 也可以取作  $K \ln(N)$  的整数部分,  $K$  是一个正数.
2. 对  $p \equiv p_N$  解样本 Yule-Walker 方程(1.3), (1.4), 得到样本 Yule-Walker 系数

$$(\hat{a}_{p,1}, \hat{a}_{p,2}, \dots, \hat{a}_{p,p}) \quad \text{和} \quad \hat{\sigma}_p^2.$$

3. 计算样本逆相关函数

$$\hat{\gamma}_y(k) = \frac{1}{\hat{\sigma}_p^2} \sum_{j=0}^{p-k} \hat{a}_{p,j} \hat{a}_{p,j+k}, \quad k = 0, 1, 2, \dots, p, \quad \hat{a}_{p,0} \triangleq -1.$$

4. 利用样本 Yule-Walker 方程(2.11) 和(2.12) 计算出  $MA(q)$  系数的估计  $\hat{\mathbf{b}}_q = (\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q)^T$  和  $\hat{\sigma}^2$ .

- 这种办法是用一个长阶自回归来近似  $\{X_t\}$ , 自回归序列的逆自相关函数很容易计算, 这样得到逆相关函数。

### §6.2.3 MA( $q$ )模型的新息估计方法

#### MA参数的新息估计

- 用新息预报公式可以计算  $\{X_t\}$  的样本新息。

$$\begin{aligned} \hat{\varepsilon}_{t+1} &= X_{t+1} - \hat{X}_{t+1} \triangleq X_{t+1} - L(X_{t+1} | X_1, \dots, X_t) \\ &= X_{t+1} - L(X_{t+1} | \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_t) \\ &= X_{t+1} - \sum_{j=1}^t \theta_{t,j} \hat{\varepsilon}_{t+1-j}, \quad t = 1, 2, \dots \end{aligned}$$

- 其中  $\hat{X}_1 \triangleq 0$ ,  $\{\theta_{t,j}\}$  可递推计算, 预报误差  $\nu_t = E\hat{\varepsilon}_{t+1}^2$  可递推计算。

- 对MA(q)序列 $\{X_t\}$ 上述新息预报公式中 $\theta_{t,j} = 0$ , 对 $t \geq q, j > q$ , 新息预报公式变成

$$\hat{X}_{t+1} = \sum_{j=1}^q \theta_{t,j} \hat{\varepsilon}_{t+1-j}, \quad t \geq q \quad (2.16)$$

- 对MA(q)序列 $\{X_t\}$ 其白噪声项 $\{\varepsilon_t\}$ 是新息:

$$\begin{aligned} \varepsilon_{t+1} &= X_{t+1} - L(X_{t+1}|X_t, X_{t-1}, X_{t-2}, \dots) \\ &\approx X_{t+1} - L(X_{t+1}|X_t, X_{t-1}, X_{t-2}, \dots, X_1) \\ &= \hat{\varepsilon}_{t+1}, \quad t \rightarrow \infty \end{aligned}$$

(由§5.2(P161)定理2.4, 无穷长历史最佳线性预测是有限历史预测的极限)

- 于是 $t$ 较大时

$$\begin{aligned} X_t &= \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q} \\ &\approx \hat{\varepsilon}_t + b_1 \hat{\varepsilon}_{t-1} + \dots + b_q \hat{\varepsilon}_{t-q} \\ &= X_t - \hat{X}_t + b_1 \hat{\varepsilon}_{t-1} + \dots + b_q \hat{\varepsilon}_{t-q} \end{aligned}$$

说明

$$\hat{X}_t \approx b_1 \hat{\varepsilon}_{t-1} + \dots + b_q \hat{\varepsilon}_{t-q}$$

- 上式与(2.16)的新息预报公式比较可知当 $t$ 较大时可以用 $\hat{b}_j = \theta_{t,j}$ 估计 $b_j$ , 用 $\hat{\nu}_t$ 估计 $\sigma^2$ 。这种估计称为**新息估计**。

### MA参数的新息估计算法

- 给定观测数据 $x_1, x_2, \dots, x_N$ , 取 $m = o(N^{1/3})$ .
- 计算样本自协方差函数 $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_m$ .
- $b$  和 $\sigma^2$ 的新息估计

$$(\hat{b}_1, \hat{b}_2, \dots, \hat{b}_q) = (\hat{\theta}_{m,1}, \hat{\theta}_{m,2}, \dots, \hat{\theta}_{m,q}), \quad \hat{\sigma}^2 = \hat{\nu}_m. \quad (2.17)$$

由下面的新息预报递推公式得到.

$$\begin{cases} \hat{\nu}_0 = \hat{\gamma}_0, \\ \hat{\theta}_{n,n-k} = \hat{\nu}_k^{-1} [\hat{\gamma}_{n-k} - \sum_{j=0}^{k-1} \hat{\theta}_{k,k-j} \hat{\theta}_{n,n-j} \hat{\nu}_j], & 0 \leq k \leq n-1, \\ \hat{\nu}_n = \hat{\gamma}_0 - \sum_{j=0}^{n-1} \hat{\theta}_{n,n-j}^2 \hat{\nu}_j, & 1 \leq n \leq m, \end{cases} \quad (2.18)$$

其中 $\sum_{j=0}^{-1}(\cdot) \triangleq 0$ .

- 递推次序是

$$\hat{\nu}_0; \hat{\theta}_{1,1}, \hat{\nu}_1; \hat{\theta}_{2,2}, \hat{\theta}_{2,1}, \hat{\nu}_2; \hat{\theta}_{3,3}, \hat{\theta}_{3,2}, \hat{\theta}_{3,1}, \hat{\nu}_3; \dots$$

## 新息估计的相合性

**定理2.3**([18]) 设 $\{X_t\}$ 是可逆ARMA( $p, q$ )序列, 满足

$$A(\mathcal{B})X_t = B(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}.$$

$\{\psi_j\}$ 是 $B(z)/A(z)$ 的Taylor级数系数. 如果 $\{\varepsilon_t\}$ 是4阶矩有限的独立同分布 $WN(0, \sigma^2)$ , 正整数列 $m = m(N) < N$  满足当 $N \rightarrow \infty$  时,  $m \rightarrow \infty$  和 $m = o(N^{1/3})$ . 则对任何正整数 $q$ , 当 $N \rightarrow \infty$

$$\sqrt{N}(\hat{\theta}_{m,1} - \psi_1, \hat{\theta}_{m,2} - \psi_2, \dots, \hat{\theta}_{m,q} - \psi_q)$$

依分布收敛到 $q$ 维正态分布 $N(0, A)$ , 其中 $q \times q$ 矩阵

$$A = (a_{i,j}), \quad a_{i,j} = \sum_{k=1}^{\min(i,j)} \psi_{i-k}\psi_{j-k}.$$

并且 $\hat{\nu}_m$  依概率收敛到 $\sigma^2$ .

- **推论** 对于MA( $q$ )序列 $\{X_t\}$ , 当模型(2.2)中的白噪声是4阶矩有限的独立同分布序列时, 新息估计(2.17)是相合估计, 当 $N$ 趋于无穷时,

$$\sqrt{N}(\hat{b}_1 - b_1, \hat{b}_2 - b_2, \dots, \hat{b}_q - b_q)$$

依分布收敛到 $q$ 维正态分布 $N(0, A)$ , 其中 $q \times q$ 矩阵

$$A = (a_{i,j}), \quad a_{i,j} = \sum_{k=1}^{\min(i,j)} b_{i-k}b_{j-k}, \quad b_0 \triangleq 1.$$

并且 $\hat{\nu}_m$  依概率收敛到 $\sigma^2$ .

- 注意需要 $m \rightarrow \infty$ 时新息方法得到的参数估计才是相合的, 不能用 $\theta_{q,1}, \dots, \theta_{q,q}$ 作为 $b_1, \dots, b_1$ 的估计.
- 实际中,  $m$ 也不能取得太大, 因为新息预测的递推公式中用到 $\hat{\gamma}_{m-1}$ .

§6.2.4 MA( $q$ )模型的定阶方法

## MA序列的定阶方法

- 由于MA( $q$ )序列的特征是自相关系数 $q$ 后截尾, 所以当样本自相关系数 $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$  从某一点 $\hat{q}$ 后变得很小时, 可以 $\hat{q}$ 作为 $q$ 的估计.
- 从§4.2 的定理2.2及例2.1 知道, 对于 $m > q$ ,  $\sqrt{N}\hat{\rho}_m$ 依分布收敛到期望为0, 方差为

$$1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_q^2$$

的正态分布. 这样由样本自相关系数 $\hat{\rho}_k, k = 1, 2, \dots, m$ 的图形可以大致得到 $q$ 的估计, 同时也可以判断采用MA( $q$ )模型的合理与否.

- 还可以用AIC定阶方法: 如果根据问题的背景或数据的特性能够判定MA(q)模型阶数 $q$ 的上界是 $Q_0$ . 对于 $m = 0, 1, 2, \dots, Q_0$ 按前述的方法逐个拟合MA(m)模型. 白噪声方差 $\sigma^2$ 的估计量记做 $\hat{\sigma}_m^2$ . 定义AIC函数

$$\text{AIC}(m) = \ln(\hat{\sigma}_m^2) + 2m/N, \quad m = 0, 1, \dots, Q_0.$$

这里 $N$ 是样本个数. AIC(m)的最小值点 $\hat{q}$  (如不惟一, 应取小的) 称为MA(q)模型的AIC定阶.

### §6.2.5 MA(q)模型的拟合检验

#### MA模型估计的拟合检验

- 从观测数据 $x_1, x_2, \dots, x_N$  得到模型的参数估计 $\hat{q}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{\hat{q}}$  和 $\hat{\sigma}^2$  后, 取

$$\begin{aligned} \hat{\varepsilon}_{1-\hat{q}} &= \hat{\varepsilon}_{2-\hat{q}} = \dots = \hat{\varepsilon}_0 = 0, \\ y_t &= x_t - \bar{x}_N, \\ \hat{\varepsilon}_t &= y_t - \sum_{j=1}^{\hat{q}} \hat{b}_j \hat{\varepsilon}_{t-j}, \quad t = 1, 2, \dots, N. \end{aligned}$$

- 对 $L = O(N^{1/3})$ , 如果 $\{\hat{\varepsilon}_t : t = L, L+1, \dots, N\}$  能够通过白噪声检验, 就认为模型的选择合适. 否则改变 $\hat{q}$ 的取值, 拟合新的MA模型或改用其他的模型, 例如改用ARMA模型等.

### §6.2.6 MA谱密度估计

#### MA(q)序列的谱密度估计

- 如果从数据得到了MA 模型的参数估计, 模型的检验也已经通过, 可以利用

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} \left| 1 + \sum_{j=1}^{\hat{q}} \hat{b}_j e^{ij\lambda} \right|^2$$

作为所关心的平稳序列的谱密度的估计.

- 这是因为如果观测数据确实是MA(q)序列(2.2)时, 它的谱密度是

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 + \sum_{j=1}^q b_j e^{ij\lambda} \right|^2.$$

- 不难看出, 如果 $\hat{q}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{\hat{q}}$ 和 $\hat{\sigma}^2$ 分别是 $q, b_1, b_2, \dots, b_q$  和 $\sigma^2$ 的相合估计, 则 $\hat{f}(\lambda)$  是 $f(\lambda)$ 的相合估计.

### §6.3 ARMA( $p, q$ )模型的参数估计

#### ARMA模型

- 对于零均值化后的平稳观测数据 $x_1, x_2, \dots, x_N$ , 如果拟合AR( $p$ )和MA( $q$ )模型的效果都不理想, 就要考虑ARMA( $p, q$ )模型的拟合.
- 这时可以假设 $x_1, x_2, \dots, x_N$ 满足如下的可逆ARMA( $p, q$ )模型

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}, t = 1, 2, \dots \quad (3.1)$$

- 其中 $\{\varepsilon_t\}$ 是 $WN(0, \sigma^2)$ , 未知参数 $\mathbf{a} = (a_1, a_2, \dots, a_p)^T$ 和 $\mathbf{b} = (b_1, b_2, \dots, b_q)^T$ 使得多项式

$$A(z) = 1 - \sum_{j=1}^p a_j z^j, \quad B(z) = 1 + \sum_{j=1}^q b_j z^j \quad (3.2)$$

互素, 并且满足

$$A(z)B(z) \neq 0, |z| \leq 1. \quad (3.3)$$

- 以下仍然用 $\hat{\gamma}_k$ 表示由(1.2)定义的样本自协方差函数.
- 我们先设 $p, q$ 是已知的.

#### §6.3.1 ARMA( $p, q$ )模型的矩估计方法

##### ARMA模型的矩估计方法

- 利用§3.2的(2.15)知道ARMA( $p, q$ )序列的自协方差函数满足延伸的Yule-Walker方程

$$\begin{pmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \dots \\ \gamma_{p+q} \end{pmatrix} = \begin{pmatrix} \gamma_q & \gamma_{q-1} & \dots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \dots & \gamma_{q-p+2} \\ \dots & \dots & \dots & \dots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \dots & \gamma_q \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{pmatrix} \quad (3.4)$$

- 这是参数 $\mathbf{a}$ 的估计方程, 从它得到 $\mathbf{a}$ 的矩估计

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \dots \\ \hat{a}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_q & \hat{\gamma}_{q-1} & \dots & \hat{\gamma}_{q-p+1} \\ \hat{\gamma}_{q+1} & \hat{\gamma}_q & \dots & \hat{\gamma}_{q-p+2} \\ \dots & \dots & \dots & \dots \\ \hat{\gamma}_{q+p-1} & \hat{\gamma}_{q+p-2} & \dots & \hat{\gamma}_q \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}_{q+1} \\ \hat{\gamma}_{q+2} \\ \dots \\ \hat{\gamma}_{q+p} \end{pmatrix}. \quad (3.5)$$



- 利用§3.2的定理2.3知道(3.4)中的 $p \times p$ 矩阵 $\Gamma_{p,q}$ 是可逆的. 用 $\hat{\Gamma}_{p,q}$ 表示(3.5)中的 $p \times p$ 矩阵. 当ARMA(p, q)模型中的白噪声 $\{\varepsilon_t\}$ 独立同分布时,  $\hat{\gamma}_k$  a.s.收敛到 $\gamma_k$ .
- 于是当 $N \rightarrow \infty$ ,

$$\det(\hat{\Gamma}_{p,q}) \rightarrow \det(\Gamma_{p,q}) \neq 0.$$

- 所以当 $N$ 充分大后, (3.5)中的矩阵 $\hat{\Gamma}_{p,q}$ 也是可逆的. 这时矩估计是惟一的. 还可以看出, 在上述的条件下, 矩估计(3.5)是强相合的:

$$\lim_{N \rightarrow \infty} \hat{a}_j = a_j, \text{ a.s., } 1 \leq j \leq p. \quad (3.6)$$

- 下面估计MA(q)部分的参数.
- 由于

$$z_t \triangleq x_t - \sum_{j=1}^p a_j x_{t-j}, \quad t = p+1, p+2, \dots, N$$

满足MA(q)模型

$$z_t = B(\mathcal{B})\varepsilon_t,$$

- 所以得到 $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ 后,

$$y_t = x_t - \sum_{j=1}^p \hat{a}_j x_{t-j}, \quad t = p+1, p+2, \dots, N. \quad (3.7)$$

是一个MA(q)序列的近似观测数据. 它的样本自协方差函数由

$$\hat{\gamma}_y(k) = \sum_{j=0}^p \sum_{l=0}^p \hat{a}_j \hat{a}_l \hat{\gamma}_{k+j-l}, \quad k = 0, 1, \dots, q \quad (3.8)$$

定义, 其中 $\hat{a}_0 = -1$ .

- 现在将(3.8)看成一个MA(q)序列的样本自协方差函数, 利用§6.2的方法就可以估计出MA(q)部分的参数 $\mathbf{b}$ 和 $\sigma^2$ .

### §6.3.2 ARMA(p, q)模型的回归逼近法

#### ARMA模型的回归逼近法

- 如果ARMA模型中已知 $\varepsilon_t$ 则 $a_1, \dots, a_p, b_1, \dots, b_q$ 可以看成是回归系数.
- $\varepsilon_t$ 作为一步预报误差, 可以用样本新息估计. 但是样本新息直接计算困难, 所以可以拟合长阶自回归模型, 用自回归模型的残差作为一步预报误差的估计.

- 首先为数据建立AR模型. 取自回归阶数的上界  $P_0 = [\sqrt{N}]$ . 这里  $[a]$  表示  $a$  的整数部分. 采用AIC定阶方法得到AR模型的阶数估计  $\hat{p}$  和自回归系数的估计

$$(\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_{\hat{p}}).$$

计算残差

$$\hat{\varepsilon}_t = x_t - \sum_{j=1}^{\hat{p}} \hat{\phi}_j x_{t-j}, \quad t = \hat{p} + 1, \hat{p} + 2, \dots, N.$$

- 然后写出近似的ARMA( $p, q$ )模型

$$x_t = \sum_{j=1}^p a_j x_{t-j} + \hat{\varepsilon}_t + \sum_{j=1}^q b_j \hat{\varepsilon}_{t-j}, \quad t = L + 1, L + 2, \dots, N.$$

这里  $L = \max(\hat{p}, p, q)$ ,  $a_j, b_k$  是待定参数.

- 最后对目标函数

$$Q(\mathbf{a}, \mathbf{b}) = \sum_{t=L+1}^N \left( x_t - \sum_{j=1}^p a_j x_{t-j} - \sum_{j=1}^q b_j \hat{\varepsilon}_{t-j} \right)^2 \quad (3.9)$$

极小化, 得到最小二乘估计  $(\hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q)$ .  $\sigma^2$  的最小二乘估计由下式定义.

$$\hat{\sigma}^2 = \frac{1}{N-L} Q(\hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q).$$

### §6.3.3 正态时间序列的似然函数

#### 正态时间序列的似然函数

- 设  $\{X_t\}$  是零均值正态时间序列, 对  $n \geq 1$ ,  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)^T$  的协方差矩阵  $\Gamma_n$  正定.
- 采用§5.3的记号可以得到最佳线性预测

$$\hat{X}_n \triangleq L(X_n | \mathbf{X}_{n-1}) = \sum_{j=1}^{n-1} \theta_{n-1, n-j} Z_j,$$

其中  $Z_j = X_j - \hat{X}_j$ ,  $Z_1 = X_1$ .

- 于是

$$\begin{aligned} X_n &= \hat{X}_n + Z_n = \sum_{j=1}^{n-1} \theta_{n-1, n-j} Z_j + Z_n \\ &= \sum_{j=1}^n \theta_{n-1, n-j} Z_j \quad (\theta_{n-1, 0} \triangleq 1) \\ &= (\theta_{n-1, n-1}, \theta_{n-1, n-2}, \dots, \theta_{n-1, 1}, 1) \cdot \mathbf{Z}_n, \end{aligned}$$

其中  $\mathbf{Z}_n = (Z_1, Z_2, \dots, Z_n)^T$ .

- 为了方便引入下三角矩阵

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{pmatrix}$$

- 则有

$$\mathbf{X}_n = C\mathbf{Z}_n.$$

- 由于

$$r_{k-1} \triangleq EZ_k^2, \quad k = 1, 2, \dots, n-1$$

是预测的均方误差(原来新息预报中 $r_{k-1}$ 用 $\nu_{k-1}$ 表示, 但是下面 $\nu_{k-1}$ 要表示对 $X_t$ 作变换后的序列 $Y_t$ 的样本新息方差), 所以用 $Z_1, Z_2, \dots, Z_n$ 的正交性得到

$$D \triangleq E(\mathbf{Z}_n \mathbf{Z}_n^T) = \text{diag}(r_0, r_1, \dots, r_{n-1}).$$

- 由此得到 $\mathbf{X}_n$ 的协方差矩阵:

$$\Gamma_n = E(C\mathbf{Z}_n \mathbf{Z}_n^T C^T) = CDC^T,$$

$$\det(\Gamma_n) = \det(D) = r_0 r_1 \cdots r_{n-1},$$

$$\mathbf{X}_n^T \Gamma_n^{-1} \mathbf{X}_n = \mathbf{Z}_n^T C^T (CDC^T)^{-1} C \mathbf{Z}_n = \sum_{j=1}^n Z_j^2 / r_{j-1}.$$

- 由于 $\mathbf{X}_n$ 的分布由 $\Gamma_n$ 决定, 而 $r_j, \theta_{k,j}$ 都是 $\Gamma_n$ 的函数, 所以可得基于 $\mathbf{X}_n$ 的似然函数

$$\begin{aligned} L(\Gamma_n) &= \frac{1}{(2\pi)^{n/2} [\det(\Gamma_n)]^{1/2}} \exp \left( -\frac{1}{2} \mathbf{X}_n^T \Gamma_n^{-1} \mathbf{X}_n \right) \\ &= \frac{1}{(2\pi)^{n/2} (r_0 r_1 \cdots r_{n-1})^{1/2}} \exp \left( -\frac{1}{2} \sum_{j=1}^n Z_j^2 / r_{j-1} \right). \end{aligned} \quad (3.10)$$

### §6.3.4 ARMA(p, q)模型的最大似然估计

#### ARMA(p, q)模型的最大似然估计

- 设 $\{X_t\}$ 是满足ARMA模型(3.1)的平稳序列. 采用§5.4的C中符号, 利用§5.4的(4.11)得到逐步预报的递推公式

$$\hat{X}_{k+1} = \begin{cases} \sum_{j=1}^k \theta_{k,j} Z_{k+1-j}, & 1 \leq k < m, \\ \sum_{j=1}^p a_j X_{k+1-j} + \sum_{j=1}^q \theta_{k,j} Z_{k+1-j}, & k \geq m. \end{cases} \quad (3.11)$$

这里  $m = \max(p, q)$ ,

$$Z_k = X_k - \hat{X}_k = X_k - L(X_k | X_{k-1}, X_{k-2}, \dots, X_1), \quad (3.12)$$

$$r_{k-1} = E Z_k^2 = E(X_k - \hat{X}_k)^2 = \sigma^2 \nu_{k-1}. \quad (3.13)$$

- $\theta_{k,j}, \nu_k$  可以通过第五章的(4.9)和(3.6)递推计算, 而(4.9)中的  $\sigma^{-2}\gamma_k$  由第三章的(2.10), (2.11)计算. 它们都是和  $\sigma^2$  无关的量, 由ARMA模型(3.1)的参数

$$\beta = (a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q)^T \quad (3.14)$$

惟一决定.

- 从而  $\hat{X}_k$  也是和  $\sigma^2$  无关的量, 仅由观测数据  $X_1, X_2, \dots, X_{k-1}$  和  $\beta$  决定. 将(3.11)和(3.13)代入(3.10)就得到基于观测数据  $X_1, X_2, \dots, X_N$  的似然函数

$$L(\beta, \sigma^2) = \frac{\exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^N Z_k^2 / \nu_{k-1}\right)}{(2\pi)^{N/2} (\sigma^2 \nu_0 \nu_1 \dots \nu_{N-1})^{1/2}} \quad (3.15)$$

- 引入

$$S(\beta) = \sum_{k=1}^N Z_k^2 / \nu_{k-1}. \quad (3.16)$$

- 忽略常数项后, 可以得到对数似然函数

$$\ln L(\beta, \sigma^2) = -\frac{N}{2} \ln \sigma^2 - \frac{1}{2} \ln(\nu_0 \nu_1 \dots \nu_{N-1}) - \frac{1}{2\sigma^2} S(\beta). \quad (3.17)$$

- 利用

$$\frac{\partial}{\partial \sigma^2} \ln L(\beta, \sigma^2) = -\frac{N}{2\sigma^2} + \frac{S(\beta)}{2\sigma^4} = 0$$

- 得到

$$\sigma^2 = \frac{1}{N} S(\beta).$$

- 将上式代入(3.17)得到

$$\begin{aligned} l(\beta) &\triangleq -\frac{2}{N} \ln L(\beta, S(\beta)/N) - 1 \\ &= \frac{1}{N} \ln(\nu_0 \nu_1 \dots \nu_{N-1}) + \ln[S(\beta)/N]. \end{aligned} \quad (3.18)$$

- 通常称  $l(\beta)$  是约化似然函数. 可以看出,  $l(\beta)$  的最小值点

$$\hat{\beta} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_q)^T \quad (3.19)$$

是  $\beta$  的最大似然估计, 而

$$\hat{\sigma}^2 = \frac{1}{N} S(\hat{\beta}) \quad (3.20)$$

是  $\sigma^2$  的最大似然估计.

## 最大似然估计的计算

- 在计算 $l(\beta)$ 的最小值点时, 可采用一般的最优化方法.
- 要计算 $l(\beta)$ 函数值, 通过第五章的(4.9)和(3.6)递推计算出 $\theta_{k,j}$ ,  $\nu_j$  和  $Z_k = X_k - \hat{X}_k$ , 然后计算出 $l(\beta)$ .
- 为了加快搜索的速度和提高估计的精度, 初始值 $\beta^{(0)}$  应当选择在 $l(\beta)$ 的最小值附近.
- 实际计算中, 初始值应当选成在A中定义的矩估计或B中定义的回归逼近估计. 为了得的估计模型的合理性和可逆性, 还应当选择初始值 $\beta^{(0)}$  使得 $A(z)$ ,  $B(z)$ 在闭单位圆内没有零点.

## 最大似然估计的与最小二乘估计

- 由于当 $k \rightarrow \infty$ , 用§5.2 定理2.4(或习题4.4)得到

$$\begin{aligned} \nu_k &= E(X_{k+1} - \hat{X}_{k+1})^2 / \sigma^2 = E[X_n - L(X_n | X_{n-1}, X_{n-2}, \dots, X_{n-k})]^2 / \sigma^2 \\ &\rightarrow E[X_n - L(X_n | X_{n-1}, X_{n-2}, \dots)]^2 / \sigma^2 = E\varepsilon_n^2 / \sigma^2 = 1. \end{aligned}$$

- 所以 $N \rightarrow \infty$  时,

$$\frac{1}{N} \ln(\nu_0 \nu_1 \cdots \nu_{N-1}) \rightarrow 0.$$

- 于是, 对较大的 $N$ ,  $l(\beta)$ 和 $S(\beta)$ 的最小值点近似相等. 于是也可以用 $S(\beta)$ 的最小值点 $\tilde{\beta}$ 作为 $\beta$ 的估计. 通常也称 $\tilde{\beta}$ 是 $\beta$ 的最小二乘估计, 相应的白噪声方差 $\sigma^2$ 的估计定义成

$$\hat{\sigma}^2 = \frac{1}{N - p - q} S(\tilde{\beta}). \quad (3.21)$$

## 最大似然估计的极限分布

- 可以证明最小二乘估计 $\tilde{\beta}$ 和最大似然估计 $\hat{\beta}$  有相同的极限分布,  $\hat{\sigma}^2$ 和 $\hat{\sigma}^2$ 有相同的极限分布.
- **定理3.1** 如果 $\{X_t\}$ 是平稳可逆的ARMA(p, q)序列, 白噪声是独立同分布序列,  $E\varepsilon_t^4 < \infty$ . 则当 $N \rightarrow \infty$  时,

$$\sqrt{N}(\hat{\beta} - \beta)$$

依分布收敛到正态分布 $N(0, V(\beta))$ , 其中

$$V(\beta) = \begin{cases} \sigma^2 \begin{pmatrix} E(UU^T) & E(UV^T) \\ E(VU^T) & E(VV^T) \end{pmatrix}^{-1}, & p > 0, q > 0, \\ \sigma^2 (E(UU^T))^{-1}, & q = 0, p > 0, \\ \sigma^2 (E(VV^T))^{-1}, & p = 0, q > 0. \end{cases} \quad (3.22)$$

$\mathbf{U} = (U_p, U_{p-1}, \dots, U_1)^T$ ,  $\mathbf{V} = (V_q, V_{q-1}, \dots, V_1)^T$ ,  $\{U_t\}$ 和 $\{V_t\}$ 分别是AR( $p$ )和AR( $q$ )序列, 分别满足 $A(\mathcal{B})U_t = \varepsilon_t$ 和 $B(\mathcal{B})V_t = \varepsilon_t$ .

### 极限分布的利用

- 用 $v_{jj}$ 表示 $V(\beta)$ 的第 $(j, j)$ 的元素, 则

$$\sqrt{N}(\hat{\beta}_j - \beta_j)$$

依分布收敛到正态分布 $N(0, v_{jj})$ .

- 在实际问题中, 真值 $V(\beta)$ 是未知的, 通常用估计量 $V(\hat{\beta})$ 代替.
- 于是可以用 $V(\hat{\beta})$ 的第 $(j, j)$ 的元 $\hat{v}_{jj}$ 作为 $v_{jj}$ 的近似.
- 这样, 当 $N$ 较大时, 利用

$$P[\sqrt{N}|\hat{\beta}_j - \beta_j|/\sqrt{\hat{v}_{jj}} \leq 1.96] \approx 0.95$$

就可以在置信水平0.95下得到 $\beta_j$ 的近似置信区间

$$[\hat{\beta}_j - 1.96\sqrt{\hat{v}_{jj}/N}, \quad \hat{\beta}_j + 1.96\sqrt{\hat{v}_{jj}/N}].$$

### 例3.1: ARMA(4,2)的估计与模拟分析

对于ARMA(4,2)模型

$$\begin{aligned} X_t + 0.9X_{t-1} + 1.4X_{t-2} + 0.7X_{t-3} + 0.6X_{t-4} \\ = \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2}, \quad t \in \mathbb{Z}, \end{aligned} \quad (3.23)$$

的模拟计算, 其中 $\{\varepsilon_t\}$ 是标准正态WN(0, 1).

利用§3.2的方法产生模型(3.23)的300个数据 $\{x_t\}$ .

利用公式(3.5)(扩展的Y-W方程)计算出 $\alpha$ 的估计

$$\hat{\alpha} = (-0.8959, -1.3843, -0.6876, -0.5868)^T.$$

再利用公式(3.8)(把 $\hat{A}(\mathcal{B})X_t$ 看成MA序列 $\{Y_t\}$ )计算出 $\hat{\gamma}_y(k)$ ,  $k = 0, 1, \dots, 10$ .

取 $p_N = 10$ , 利用MA模型的逆相关函数方法计算出MA部分的参数估计和白噪声方差的估计如下:

$$\hat{b} = (0.5122, -0.3427)^T, \quad \hat{\sigma}^2 = 1.47.$$

与真实值比较, 自回归部分的参数估计是令人满意的.

下面以 $\hat{a}, \hat{b}, \hat{\sigma}^2$ 为初值计算最大似然估计. 计算采用了在初值附近搜索的方法. 结果为:

$$\begin{aligned} \hat{a} &= (-0.9010 \quad -1.3920 \quad -0.6808 \quad -0.6039)^T, \\ \hat{b} &= (0.5145 \quad -0.3721)^T, \quad \hat{\sigma}^2 = 1.282. \end{aligned}$$

从以上结果看出, 本例中最大似然估计总体上可以改进初估计, 特别是能改进MA参数和 $\sigma^2$ 的估计精度.

将上述的模拟计算重复1000次, 每次重新产生模型(3.23)的300个数据, 然后用上面的初值估计方法得到参数估计(不再进行最大似然估计)。1000次的参数估计的结果综合如下:

$$\begin{aligned} Ave(\hat{a}) &= \begin{matrix} -0.8954 & -1.3880 & -0.6897 & -0.5900 \end{matrix} \\ Std(\hat{a}) &= \begin{matrix} 0.0716 & 0.0873 & 0.0773 & 0.0669 \end{matrix} \\ Ave(\hat{b}) &= \begin{matrix} 0.5258 & -0.3141 \end{matrix} & Ave(\hat{\sigma}^2) &= 1.2791 \\ Std(\hat{b}) &= \begin{matrix} 0.6893 & 0.3188 \end{matrix} & Std(\hat{\sigma}^2) &= 0.1639 \end{aligned}$$

从上述的计算看出, AR部分的估计是令人满意的, MA部分和白噪声方差估计的标准差较大. 采用最大似然估计后通常可以改进估计的精度, 特别是改进MA参数和白噪声方差的估计.

### §6.3.5 ARMA(p, q)模型的检验

#### ARMA模型拟合的检验

- 在得到了ARMA(p, q)模型的参数估计

$$(\hat{a}_1, \dots, \hat{a}_p), (\hat{b}_1, \dots, \hat{b}_q)$$

后, 对模型进行检验是十分必要的.

- 首先要检验模型的平稳性和合理性, 即要检验估计的参数满足(3.3).
- 然后对取定的初值

$$x_0 = x_{-1} = \dots = x_{-p+1} = \hat{\varepsilon}_0 = \dots = \hat{\varepsilon}_{-q+1} = 0$$

递推计算模型的残差

$$\hat{\varepsilon}_t = x_t - \sum_{l=1}^p \hat{a}_l x_{t-l} + \sum_{j=1}^q \hat{b}_j \hat{\varepsilon}_{t-j}, \quad t = 1, 2, \dots \quad (3.24)$$

- 取 $m = O(N^{1/3})$ 和 $m > \max(p, q)$ . 如果残差

$$\hat{\varepsilon}_t, \quad t = m, m+1, \dots, N$$

可以通过白噪声的检验, 就认为模型合适. 否则寻找其他的模型.

- 在实际工作中, 参数(p, q)是未知的. 但是根据数据的性质有时可以知道阶数的大约范围. 可以在这个范围内对每一对(p, q)建立ARMA(p, q)模型, 如果一个模型可以通过检验, 就把这个模型留做备用.

- 如果不能确定阶数的范围, 可以采用从 $p+q=1, p+q=2, \dots$ , 开始由低阶到高阶的依次搜寻的方法. 然后在所有备用的模型中选出 $p+q$ 最小的一个模型.
- 如果 $p+q$ 不能惟一决定 $(p, q)$ , 可以取 $p$ 较大的一个. 也可以在所有的备用模型中, 采用下面的AIC定阶方法, 最后确定一个模型.

### §6.3.6 ARMA( $p, q$ )模型的定阶方法

#### ARMA( $p, q$ )模型的定阶方法

- 和AR模型的定阶方法相似, 给定ARMA( $p, q$ )模型的阶数 $(p, q)$ 的一个估计 $(k, j) = (\hat{p}, \hat{q})$ .
- 无论这个估计是怎样得到的, 按前面的方法都可以估计出ARMA( $k, j$ )模型的参数.
- 用 $\hat{\sigma}^2 = \hat{\sigma}^2(k, j)$ 表示白噪声方差 $\sigma^2$ 的估计.
- 一般来讲, 希望 $\hat{\sigma}^2$ 的取值越小越好. 因为 $\hat{\sigma}^2$ 越小表示模型拟合的越精确.
- 通常较小的残差方差 $\hat{\sigma}^2$ 对应于较大的阶数 $k, j$ .
- 这样, 过多追求拟合的精度, 或说过分追求较小的残差方差 $\hat{\sigma}^2$ 会导致较大的 $\hat{p}$ 和 $\hat{q}$ , 从而导致较多的待估参数.
- 其结果会使建立的模型关于数据过于敏感, 从而降低模型的稳健性.
- 另一方面, 参数过多的模型拟合较好但是对新数据的解释能力差.
- AIC定阶准则就是为了克服参数过多的弊病而提出的.
- ARMA模型的AIC定阶方法和AR模型的AIC定阶方法是相同的.
- 如果已知 $p$ 的上界 $P_0$ 和 $q$ 的上界 $Q_0$ , 对于每一对 $(k, j)$ ,  $0 \leq k \leq P_0, 0 \leq j \leq Q_0$  计算AIC函数

$$AIC(k, j) = \ln(\hat{\sigma}^2(k, j)) + \frac{2(k+j)}{N}. \quad (3.25)$$

- AIC( $k, j$ )的最小值点 $(\hat{p}, \hat{q})$  称为 $(p, q)$ 的AIC定阶.
- 如果最小值不惟一, 应先取 $k+j$ 最小的, 然后取 $j$ 最小的.
- 一般AIC定阶并不是相合的. 也就是说, 如果平稳序列的观测 $\{x_t\}$ 确实是来自一个ARMA( $p, q$ )模型时, AIC定阶 $(\hat{p}, \hat{q})$  并不依概率收敛到真正的 $(p, q)$ . 但是, 和AR( $p$ )模型的AIC定阶相似, 这种不相容性并不能否定AIC定阶的实用性. 因为一方面实际数据并不会真正满足某一个ARMA( $p, q$ )模



型, 所以真正的阶数并不存在, 采用ARMA模型只是对数据的一种处理方法. 另一方面, AIC定阶有时会高估阶数, 但是并不会高估出很多. 这比低估阶数要好, 低估了阶数往往会造成更大的误差和模型的选择不当.

- 将(3.25)中的 $2(k+j)/N$ 改为 $(k+j) \ln N/N$  就得到BIC( $k, j$ )定阶.

### §6.3.7 ARMA( $p, q$ )模型的谱密度估计

#### ARMA谱密度估计

- 如果得到了ARMA( $p, q$ ) 模型的参数估计, 模型的检验也已经通过. 可以利用

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2 |1 + \sum_{j=1}^{\hat{q}} \hat{b}_j e^{ij\lambda}|^2}{2\pi |1 - \sum_{j=1}^{\hat{p}} \hat{a}_j e^{ij\lambda}|^2} \quad (3.26)$$

作为谱密度的估计.

- 这是因为如果观测数据确实是ARMA( $p, q$ )序列(3.1)时, 它的谱密度是

$$f(\lambda) = \frac{\sigma^2 |1 + \sum_{j=1}^q b_j e^{ij\lambda}|^2}{2\pi |1 - \sum_{j=1}^p a_j e^{ij\lambda}|^2}.$$

- 不难看出, 如果 $\hat{p}, \hat{q}, \hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_q$ 和 $\hat{\sigma}^2$ 分别是 $p, q, a_1, \dots, a_p, b_1, b_2, \dots, b_q$ 和 $\sigma^2$ 的相合估计, 则 $\hat{f}(\lambda)$  是 $f(\lambda)$ 的相合估计.
- 通常对谱密度进行估计只是为了解时间序列的频率特性. 后面的第八章将主要介绍谱密度的估计方法.
- 例3.2: 见讲义