目录

一、 时间序列

1.1 时间序列的定义

按时间次序排列的随机变量序列

$$X_1, X_2, \cdots \tag{1.1}$$

称为时间序列, 记为 $X(t)$ 或 X_t. 如果用 $x_1, x_2, \cdots, x_N$ 分别表示随机变量 $X_1, X_2, \cdots, X_N$ 的观测值, 则称其为时间序列 (1.1) 的 N 个观测样本. 如果用 $x_1, x_2, \cdots$ 表示 $X_1, X_2, \cdots$ 的依次观测值, 则称其为时间序列 (1.1) 的一次实现或一条轨道.

1.2 时间序列的分解

(i) 趋势项 + 随机项

模型:

$$X_t = T_t + \varepsilon_t, \quad t = 1, 2, \cdots, n \tag{1.2}$$

常用方法:

a) 趋势拟合法.

基本步骤: 作数据图; 判断趋势类型; 估计趋势项; 残差序列.

b) 平滑法.

基本步骤: 作数据图; 选择方法; 计算移动平均或滑动平均并作为估计; 残差序列.

1. 移动平均法:

n 期移动平均: $\widetilde{x}_t = \dfrac{1}{n}(x_t + x_{t-1} + \cdots + x_{t-n+1})$;

n 期中心移动平均:

$$\widetilde{x}_t = \begin{cases} \dfrac{1}{n}(x_{t-\frac{n-1}{2}} + x_{t-\frac{n-1}{2}+1} + \cdots + x_t + \cdots + x_{t+\frac{n-1}{2}-1} + x_{t+\frac{n-1}{2}}), & n \text{ 为奇数}, \\ \dfrac{1}{n}\left(\dfrac{1}{2}x_{t-\frac{n}{2}} + x_{t-\frac{n}{2}+1} + \cdots + x_t + \cdots + x_{t+\frac{n}{2}-1} + \dfrac{1}{2}x_{t+\frac{n}{2}}\right), & n \text{ 为偶数}. \end{cases}$$

注: 以周期长度作为移动平均的间隔长度, 可以消除周期效应的影响; 移动平均的期数越多 (少), 拟合趋势越平滑 (敏感).

2. 指数平滑法:

简单指数平滑: $\widetilde{x}_t = \alpha x_t + (1-\alpha)\widetilde{x}_{t-1} = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \cdots$

两参数指数平滑 (含趋势项):

$$\begin{aligned} &\widetilde{x}_t = \alpha x_t + (1-\alpha)(\widetilde{x}_{t-1} + T_{t-1}),\ T_t = \beta(\widetilde{x}_t - \widetilde{x}_{t-1}) + (1-\beta)T_{t-1}, \\ &\widehat{x}_t(h) = \widetilde{x}_t + hT_t, \quad h = 1, 2, \cdots \end{aligned}$$

注: 对于变化缓慢 (迅速) 的序列, α 常取较小 (大) 的值; 经验表明 α 的值介于 0.05 至 0.3 之间, 修匀效果比较好.

(ii) 季节项 + 随机项

模型:

$$X_t = S_t + \varepsilon_t, \quad t = 1, 2, \cdots, n \tag{1.3}$$

设模型的周期为 m, 计算周期内各期平均数 $\overline{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ik}$, $k=1,2,\cdots,m$ 即为季节项估计. 随机项序列则为 $\varepsilon_{ij} = x_{ij} - \overline{x}_j$.

(iii) 综合分析模型

加法模型: $X_t = T_t + S_t + \varepsilon_t, \quad t = 1,2,\cdots,n;$

乘法模型: $X_t = T_t \cdot S_t \cdot \varepsilon_t, \quad t = 1,2,\cdots,n;$

混合模型: $X_t = T_t \cdot S_t + \varepsilon_t, \quad t = 1,2,\cdots,n.$

1.3 平稳序列

(i) 平稳序列及自协方差函数

如果时间序列 $\{X_t\} = \{X_t : t \in \mathbb{N}\}$ 满足

(1) 对任何 $t \in \mathbb{N}$, $\mathrm{E}X_t^2 < \infty$;

(2) 对任何 $t \in \mathbb{N}$, $\mathrm{E}X_t = \mu$;

(3) 对任何 $t, s \in \mathbb{N}$, $\mathrm{E}[(X_t - \mu)(X_s - \mu)] = \gamma_{t-s}$,

就称 $\{X_t\}$ 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.

自协方差函数的三条基本性质:

(1) 对称性: $\gamma_k = \gamma_{-k}$ 对所有 $k \in \mathbb{Z}$ 成立;

(2) 非负定性: 对任何 $n \in \mathbb{N}_+$, n 阶自协方差矩阵

$$\Gamma_n = (\gamma_{k-j})_{k,j=1}^n = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} \tag{1.4}$$

是非负定矩阵;

(3) 有界性: $|\gamma_k| \leq \gamma_0$ 对所有 $k \in \mathbb{Z}$ 成立.

注: 任何满足上述三个性质的实数列都被称为非负定序列; 每个非负定序列都可以是一个平稳序列的自协方差函数.

(ii) 自相关系数

性质 设平稳序列 $\{X_t\}$ 的期望为 μ, 自协方差函数为 $\gamma(t)$, 考虑 X_t 的线性变换 $Y_t = aX_t + b, t \in \mathbb{Z}$, 则有

$$\mathrm{E}Y_t = a\mathrm{E}X_t + b = a\mu + b, \quad \mathrm{Cov}(Y_s, Y_{s+t}) = a^2\mathrm{Cov}(X_s, X_{s+t}) = a^2\gamma(t),$$

则 $\{Y_t\}$ 也是一个平稳序列.

特别地, 取

$$Y_t = \frac{X_t - \mu}{\sqrt{\gamma_0}}, \quad t \in \mathbb{Z}, \tag{1.5}$$

则 $\mathrm{E}Y_t = 0, \mathrm{Var}(Y_t) = 1$, 称 $\{Y_t\}$ 为 $\{X_t\}$ 的标准化序列.

平稳序列 $\{X_t\}$ 的标准化序列的自协方差函数 $\rho_k = \gamma_k/\gamma_0, k \in \mathbb{Z}$ 称为平稳序列 $\{X_t\}$ 的自相关系数. 它也是一个非负定序列.

(iii) 白噪声

设 $\{\varepsilon_t\}$ 是一个平稳序列. 如果对任何 $s, t \in \mathbb{N}$,

$$\mathrm{E}\varepsilon_t = 0, \quad \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \sigma^2\delta_{t-s} = \begin{cases} \sigma^2, & t = s, \\ 0, & t \neq s, \end{cases} \tag{1.6}$$

则称 $\{\varepsilon_t\}$ 是一个白噪声, 记作 $\mathrm{WN}(\mu, \sigma^2)$.

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;

当 $\mu = 0, \sigma^2 = 1$ 时, 称 $\{\varepsilon_t\}$ 为标准白噪声;

对独立白噪声, 当 ε_t 服从正态分布时, 称 $\{\varepsilon_t\}$ 为正态白噪声.

(iv) 正交平稳序列

对于平稳序列 $\{X_t\}$ 和 $\{Y_t\}$,

(1) 如果对任何 $s, t \in \mathbb{Z}$, $\mathrm{E}(X_tY_s) = 0$, 则称 $\{X_t\}$ 和 $\{Y_t\}$ 是正交的;

(2) 如果对任何 $s, t \in \mathbb{Z}$, $\mathrm{Cov}(X_t, Y_s) = 0$, 则称 $\{X_t\}$ 和 $\{Y_t\}$ 是不相关的.

对零均值的平稳序列, 正交性和不相关性等价.

定理 设 $\gamma_X(k)$ 和 $\gamma_Y(k)$ 分别是平稳序列 $\{X_t\}$ 和 $\{Y_t\}$ 的自协方差函数. 记 $\mu_X = \mathrm{E}X_t$ 和 $\mu_Y = \mathrm{E}Y_t$. 定义

$$Z_t = X_t + Y_t, \quad t \in \mathbb{Z}. \tag{1.7}$$

(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

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k) - 2\mu_X\mu_Y, \quad k = 0, 1, 2, \cdots; \tag{1.8}$$

(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

$$\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k), \quad k = 0, 1, 2, \cdots. \tag{1.9}$$

1.4 线性平稳序列与线性滤波

(i) 有限运动平均

设 $\{\varepsilon_t\} = \{\varepsilon_t : t \in \mathbb{Z}\}$ 是 $\mathrm{WN}(0, \sigma^2)$. 对于非负整数 q 和常数 $a_0, a_1, \cdots, a_q$, 我们称

$$X_t = \sum_{j=0}^{q} a_j\varepsilon_{t-j} = a_0\varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_q\varepsilon t - q, \quad t \in \mathbb{Z} \tag{1.10}$$

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.

通过简单计算可知

$$\mathrm{E}X_t = 0, \quad \gamma_k = \mathrm{E}X_kX_0 = \begin{cases} \sigma^2\sum_{j=0}^{q-k} a_ja_{j+k}, & 0 \leq k \leq q, \\ 0, & k > q. \end{cases}$$

当 $k > q$ 时, $\gamma_k = 0$, 即时间间隔大于 q 步的随机干扰是不相关的, 称这样的序列是 q 相关的.

(ii) 线性平稳序列

为了把有限运动平均推广到无限的场合, 需要借助如下两个定理.

定理 (单调收敛定理) 如果非负随机变量序列单调不减: $0 \le \xi_1 \le \xi_2 \le \cdots$, 则当 $\xi_n \to \xi$ a.s. 时, 有 $\mathrm{E}\xi = \lim\limits_{n\to\infty} \mathrm{E}\xi_n$.

对于任何时间序列 $\{Y_t\}$, 利用单调收敛定理得到:

$$\mathrm{E}\Big[\sum_{t=-\infty}^{\infty} |Y_t|\Big] = \lim_{n\to\infty} \mathrm{E}\Big[\sum_{t=-n}^{n} |Y_t|\Big] = \sum_{t=-\infty}^{\infty} \mathrm{E}|Y_t|. \tag{1.11}$$

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

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, 记作 $\{a_j\} \in l_1$. 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

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{1.12}$$

则 $\{X_t\}$ 是平稳序列.

由 (1.11) 式和 Schwarz 不等式有

$$\mathrm{E}\Big[\sum_{j=-\infty}^{\infty} |a_j\varepsilon_{t-j}|\Big] = \sum_{j=-\infty}^{\infty} |a_j| \mathrm{E}|\varepsilon_{t-j}| \le \sigma \sum_{j=-\infty}^{\infty} |a_j| < \infty.$$

于是, (1.12) 右边的无穷级数是 a.s. 绝对收敛的, 从而是 a.s. 收敛的. 再利用控制收敛定理得

$$\Big|\sum_{j=-n}^{n} a_j\varepsilon_{t-j}\Big| \le \sum_{j=-\infty}^{\infty} |a_j\varepsilon_{t-j}| \implies \mathrm{E}X_t = \lim_{n\to\infty} \mathrm{E}\Big[\sum_{j=-n}^{n} a_j\varepsilon_{t-j}\Big] = 0.$$

类似地, 可以计算得到自协方差函数

$$\gamma(t-s) = \mathrm{E}(X_tX_s) = \lim_{n\to\infty} \sum_{j=-n}^{n}\sum_{k=-n}^{n} a_ja_k\mathrm{E}(\varepsilon_{t-j}\varepsilon_{t-k}) = \sigma^2 \sum_{j=-\infty}^{\infty} a_ja_{j+(t-s)}.$$

注: 一般的线性平稳序列只要求 $\{a_j\}$ 平方可和, 由 (1.12) 定义的时间序列 $\{X_t\}$ 仍然是平稳序列, 有零均值和自协方差函数

$$\gamma(k) = \sigma^2 \sum_{j=-\infty}^{\infty} a_ja_{j+k}, \quad k \in \mathbb{Z}. \tag{1.13}$$

此时可以证明 (1.12) 右边的无穷级数是均方收敛的.

定理 设 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0,\sigma^2)$, 实数列 $\{a_j\}$ 平方可和, 线性平稳序列 $\{X_t\}$ 由 (1.12) 定义, 则自协方差函数 $\lim\limits_{k\to\infty} \gamma_k = 0$.

(iii) 时间序列的线性滤波

在线性滤波问题中, 绝对可和的实数列 $H=\{h_j\}$ 被称为一个保时线性滤波器. 信号 $\{X_t\}$ 通过滤波器 H 后得到输出

$$Y_t=\sum_{j=-\infty}^{\infty}h_jX_{t-j},\quad t\in\mathbb{Z}. \tag{1.14}$$

如果 $\{X_t\}$ 是平稳信号, 有数学期望 $\mathrm{E}X_t=\mu$ 和自协方差函数 $\{\gamma_k\}$, 则 (1.14) 是平稳信号, 且有

$$\mu_Y=\mathrm{E}Y_t=\sum_{j=-\infty}^{\infty}h_jEX_{t-j}=\mu\sum_{j=-\infty}^{\infty}h_j,$$

$$\gamma_Y(n)=\mathrm{Cov}(Y_{n+1},Y_1)=\sum_{j,k=-\infty}^{\infty}h_jh_k\mathrm{E}[(X_{n+1-j}-\mu)(X_{1-k}-\mu)]=\sum_{j,k=-\infty}^{\infty}h_jh_k\gamma_{n+k-j}.$$

1.5 正态平稳序列与严平稳序列

(i) 正态平稳序列

对于时间序列 $\{X_t\}$, 如果对任何 $n\geq 1$ 和 $t_1,t_2,\cdots,t_n\in\mathbb{N}$, $(X(t_1),X(t_2),\cdots,X(t_n))$ 服从多元正态分布, 就称 $\{X_t\}$ 是正态时间序列. 特别地, 当 $\{X_t\}$ 还是平稳序列时, 又称为正态平稳序列.

定理 如果正态时间序列 $\{\xi_n,n\in\mathbb{N}_+\}$ 依分布收敛到随机变量 ξ, 则 $\xi\sim N(\mathrm{E}\xi,\mathrm{Var}(\xi))$, 并且

$$\mathrm{E}\xi_n\to\mathrm{E}\xi,\quad \mathrm{Var}(\xi_n)\to\mathrm{Var}(\xi).$$

定理 如果 $\{\varepsilon_t\}$ 是正态 $\mathrm{WN}(0,\sigma^2)$, 实数列 $\{a_j\}$ 绝对可和, 则由 (1.12) 定义的平稳序列是零均值正态序列, 自协方差函数由 (1.13) 给出.

注: 定理中可将 $\{a_j\}$ 绝对可和减弱为平方可和.

(ii) 严平稳序列

设 $\{X_t:t\in\mathbb{N}\}$ 是时间序列. 如果对任何正整数 n 和 $k\in\mathbb{N}$, 随机向量 $(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 和 $(X_{1+k},X_{2+k},\cdots,X_{n+k})^{\mathrm{T}}$ 同分布, 就称 $\{X_t:t\in\mathbb{N}\}$ 是严平稳序列.

如果严平稳序列是遍历的, 那么从它的一次实现 $x_1,x_2,\cdots$ 就可以推断出这个严平稳序列的所有有限维分布:

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

称有遍历性的严平稳序列为严平稳遍历序列.

定理 (遍历定理) 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 $\mathrm{E}|X_1|<\infty$, 则 $\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{t=1}^{n}X_t=\mathrm{E}X_1$, a.s..

(2) 对任何多元函数 $\varphi(x_1,x_2,\cdots,x_m)$, $Y_t=\varphi(X_{t+1},X_{t+2},\cdots,X_{t+m})$ 是严平稳遍历序列.

定理 如果 $\{\varepsilon_t\}$ 是正态 $\mathrm{WN}(0,\sigma^2)$, 实数列 $\{a_j\}$ 平方可和, 则由 (1.12) 定义的平稳序列是严平稳遍历的.

1.6 Hilbert 空间中的平稳序列

(i) Hilbert 空间

通常考虑用历史资料 $X_1, X_2, \cdots, X_n$ 的线性组合来对未来的 X_{n+1} 进行预测. 用 $L^2(X)$ 表示平稳序列 $\{X_t\}$ 中随机变量有限线性组合的全体:

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

则对任何 $X, Y, Z \in L^2(X), a, b \in \mathbb{R}$, 有以下结论成立:

(1) $X + Y = Y + X \in L^2(X), \quad (X + Y) + Z = X + (Y + Z)$;

(2) $0 \in L^2(X), \quad X + 0 = X, \quad X + (-X) = 0 \in L^2(X)$;

(3) $a(X + Y) = aX + aY \in L^2(X), \quad (a + b)X = aX + bX, \quad a(bX) = (ab)X$.

所以 $L^2(X)$ 是一个线性空间. 进一步, 在 $L^2(X)$ 上定义内积 $\langle X, Y\rangle = \mathrm{E}(XY)$, 则有:

(4) $\langle X, Y\rangle = \langle X, Y\rangle, \quad \langle aX + bY, Z\rangle = a\langle X, Z\rangle + b\langle Y, X\rangle$;

(5) $\langle X, X\rangle \geq 0$, 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, 有 Schwarz 不等式成立. 更进一步, 在 $L^2(X)$ 上引入距离 $\|X - Y\| = \left(\langle X - Y, X - Y\rangle\right)^{1/2}$, 则有:

(6) $\|X - Y\| = \|Y - X\| \geq 0$, 并且 $\|X - Y\| = 0$ 当且仅当 $X = Y$, a.s..

(7) (三角不等式) $\|X - Y\| \leq \|X - Z\| + \|Z - Y\|$.

这样, $L^2(X)$ 又成为距离空间.

如果用 L^2 表示二阶矩有限的随机变量的全体, 即 $L^2 = \{X : \mathrm{E}X^2 < \infty\}$, 则 L^2 也是内积空间和距离空间, 并且 $L^2(X)$ 是 L^2 的子空间.

下面给出两个基本概念. 对 $\xi_n \in L^2, \xi_0 \in L^2$,

(1) 如果 $\lim\limits_{n\to\infty} \|\xi_n - \xi_0\| = 0$, 就称 ξ_n 在 L^2 中 (均方) 收敛到 ξ_0, 记作 $\xi_n \xrightarrow{\text{m.s.}} \xi_0$;

(2) 如果当 $n, m \to \infty$ 时, $\|\xi_n - \xi_m\| \to 0$, 就称 $\{\xi_n\}$ 是 L^2 中的基本列或 Cauchy 列.

定理 如果 $\{\xi_n\}$ 是 L^2 中的基本列, 则 (在 a.s. 的意义下) 有唯一的 $\xi \in L^2$, 使得 $\xi_n \xrightarrow{\text{m.s.}} \xi$.

注: 这与数列极限中的 Cauchy 收敛准则类似.

如果内积空间中的每个基本列都有极限, 而且这个极限也在内积空间中, 就称这个内积空间是完备的. 完备的内积空间又称为 Hilbert 空间. 上面的定理告诉我们, L^2 是 Hilbert 空间.

(ii) 内积的连续性

定理 在内积空间中, 如果 $\|\xi_n - \xi\| \to 0, \|\eta_n - \eta\| \to 0$, 则有:

(1) $\|\xi_n\| \to \|\xi\|, \quad \|\eta_n\| \to \|\eta\|$;

(2) $\langle \xi_n, \eta_n\rangle \to \langle \xi, \eta\rangle$.

如果随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 是零均值平稳序列 $\{X_t\}$ 一段, 则它的线性组合全体构成的内积空间

$$L_n \xlongequal{\text{记为}} \mathrm{sp}(X_1, X_2, \cdots, X_n) \xlongequal{\text{def}} \{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} : \boldsymbol{a} \in \mathbb{R}^n\}$$

是 Hilbert 空间, 称为由 X 生成的 Hilbert 空间.

(iii) 复值时间序列

如果 X 和 Y 是随机变量, 就称 $Z = X + \mathrm{i}Y$ 是复随机变量. 如果 $\mathrm{E}X$ 和 $\mathrm{E}Y$ 都存在, 就称 Z 的数学期望存在, 并且定义 $\mathrm{E}Z = \mathrm{E}X + \mathrm{i}\mathrm{E}Y$. 如果 $\mathrm{E}|Z|^2 = \mathrm{E}X^2 + \mathrm{E}Y^2 < \infty$, 就称 Z 是二阶矩有限的复随机变量.

用 H 表示二阶矩有限的复随机变量的全体, $\overline{Y}$ 表示 Y 的共轭. 在 H 上定义内积 $\langle X, Y\rangle = \mathrm{E}(X\overline{Y})$, 可以证明 H 是复 Hilbert 空间.

按时间次序排列的复随机变量 $\{Z_n\}$ 的序列称为复时间序列. 如果复时间序列 $\{Z_n\}$ 满足

$$\mathrm{E}Z_n = \mu, \quad \gamma_{n-m} = \mathrm{E}[(Z_n - \mu)(\overline{Z_m - \mu})], \quad n, m \in \mathbb{Z},$$

就称 $\{Z_n\}$ 是一个复值平稳序列, $\{\gamma_k\}$ 是 $\{Z_n\}$ 的自协方差函数.

特别当

$$\mathrm{E}Z_n = 0, \quad \gamma_{n-m} = \mathrm{E}[Z_n\overline{Z_m}] = \sigma^2\delta_{n-m}, \quad n, m \in \mathbb{Z}$$

时, 称 $\{Z_n\}$ 是一个复值零均值白噪声, 这里 δ_{n-m} 是 Kronecker 函数.

1.7 平稳序列的谱函数

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

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

$$\gamma_k = \int_{-\pi}^{\pi} e^{\mathrm{i}k\lambda}\mathrm{d}F(\lambda), \quad F(-\pi) = 0, \quad k \in \mathbb{Z}, \tag{1.15}$$

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

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

$$\gamma_k = \int_{-\pi}^{\pi} f(\lambda)e^{\mathrm{i}k\lambda}\mathrm{d}\lambda, \quad k \in \mathbb{Z}, \tag{1.16}$$

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

如果 $\{X_t\}$ 有谱密度 $f(\lambda)$, 则变上限积分 $F(\lambda) = \displaystyle\int_{-\pi}^{\lambda} f(s)\mathrm{d}s$ 就是 $\{X_t\}$ 的谱函数. 当谱函数 $F(\lambda)$ 绝对连续, 它的几乎处处导函数 $F'(\lambda)$ 就是谱密度.

定理 (Herglotz 定理) 平稳序列的谱函数是惟一存在的.

定理 如果 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0, \sigma^2)$, 实数列 $\{a_j\}$ 平方可和, 则由 (1.12) 定义的平稳序列有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi}\Big|\sum_{j=-\infty}^{\infty} a_j e^{\mathrm{i}j\lambda}\Big|^2. \tag{1.17}$$

定理 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 $f_Y(\lambda)$, 则平稳序列 $\{Z_t\}$ 有谱密度 $f_Z(\lambda) = f_X(\lambda) + f_Y(\lambda)$.

下面考虑从线性滤波器输出的平稳序列的谱函数和谱密度的变化. 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. 当输入过程是 $\{X_t\}$ 时, 输出过程 (1.14), 则有

$$\begin{aligned}\gamma_Y(k)=\sum_{l,j=-\infty}^{\infty}h_lh_j\gamma_{k+l-j}&=\sum_{l,j=-\infty}^{\infty}h_lh_j\int_{-\pi}^{\pi}\exp(\mathrm{i}(k+l-j)\lambda)\mathrm{d}F_X(\lambda)\\&=\int_{-\pi}^{\pi}\sum_{l,j=-\infty}^{\infty}h_lh_j\exp(\mathrm{i}(l-j)\lambda)e^{\mathrm{i}k\lambda}\mathrm{d}F_X(\lambda)\\&=\int_{-\pi}^{\pi}|H(e^{-\mathrm{i}\lambda})|^2e^{\mathrm{i}k\lambda}\mathrm{d}F_X(\lambda)\end{aligned}\tag{1.18}$$

其中 $H(z)=\sum\limits_{j=-\infty}^{\infty}h_jz^j,\ |z|\le 1.$

于是 $\{Y_t\}$ 的谱函数为

$$F_Y(\lambda)=\int_{-\pi}^{\lambda}|H(e^{-\mathrm{i}s})|^2\mathrm{d}F_X(s).\tag{1.19}$$

特别地, 当 $\{X_t\}$ 有谱密度 $f_X(\lambda)$ 时, $\{Y_t\}$ 有谱密度

$$f_Y(\lambda)=|H(e^{-\mathrm{i}\lambda})|^2f_X(\lambda).\tag{1.20}$$

二、 自回归模型

2.1 推移算子和常系数差分方程

(i) 推移算子

对任何时间序列 $\{X_t\}$ 和无穷级数 $\psi(z)=\sum\limits_{j=-\infty}^{\infty}b_jz^j$, 只要级数 $\sum\limits_{j=-\infty}^{\infty}b_jX_{t-j}$ 在某种意义下收敛, 就定义

$$\psi(\mathscr{B})=\sum_{j=-\infty}^{\infty}b_j\mathscr{B}^j,\quad \psi(\mathscr{B})X_t=\sum_{j=-\infty}^{\infty}b_j\mathscr{B}^jX_t=\sum_{j=-\infty}^{\infty}b_jX_{t-j},\tag{2.1}$$

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

推移算子又称为时滞算子或延迟算子, 且具有以下性质:

(1) 对和 t 无关的随机变量 Y, 有 $\mathscr{B}Y=Y$.

(2) 对整数 n, 常数 a, 有 $\mathscr{B}^n(aX_t)=a\mathscr{B}^nX_t=aX_{t-n}$.

(3) 对整数 n,m, 有 $\mathscr{B}^{n+m}X_t=\mathscr{B}^n(\mathscr{B}^m)X_t=X_{t-n-m}$.

(4) 对多项式 $\psi(z)=\sum\limits_{j=0}^{p}c_jz^j$, 有 $\psi(\mathscr{B})X_t=\sum\limits_{j=0}^{p}c_jX_{t-j}$.

(5) 对于多项式 $\psi(z)=\sum\limits_{j=0}^{p}c_jz^j$ 和 $\varphi(z)=\sum\limits_{j=0}^{q}d_jz^j$ 的乘积 $A(z)=\psi(z)\varphi(z)$, 有

$$A(\mathscr{B})X_t=\psi(\mathscr{B})[\varphi(\mathscr{B})X_t]=\varphi(\mathscr{B})[\psi(\mathscr{B})X_t].$$

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

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

(ii) 常系数齐次线性差分方程

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

$$X_t-[a_1X_{t-1}+a_2X_{t-2}+\cdots+a_pX_{t-p}]=0,\quad t\in\mathbb{Z} \tag{2.2}$$

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 (2.2) 的实值 (或复值) 时间序列 $\{X_t\}$ 称为 (2.2) 的解.

利用推移算子 $\mathscr{B}$, 可以把 (2.2) 写成等价的形式 $A(\mathscr{B})X_t=0,\ t\in\mathbb{Z}$, 其中

$$A(z)=1-\sum_{j=1}^{p}a_jz^j \tag{2.3}$$

称为 (2.2) 的特征多项式.

定理 设 $A(z)$ 有 k 个互不相同的零点 $z_1,z_2,\cdots,z_k$, 其中 z_j 是 $r(j)$ 重零点, 则 $\{z_j^{-t}t^l\},\ l=0,1,\cdots,r(j)-1,\ j=1,2,\cdots,k$ 是 (2.2) 的 p 个解; 而且, (2.2) 的任何解 $\{X_t\}$ 都可以写成这 p 个解的线性组合:

$$X_t=\sum_{j=1}^{k}\sum_{l=0}^{r(j)-1}U_{l,j}t^lz_j^{-t},\quad t\in\mathbb{Z}, \tag{2.4}$$

其中的随机变量 $U_{l,j}$ 可以由 $\{X_t\}$ 的初值 $X_0,X_1,\cdots,X_{p-1}$ 惟一决定.

将 $U_{l,j}$ 和 z_j 写成指数形式 $U_{l,j}=V_{l,j}e^{\mathrm{i}\theta_{l,j}},\ z_j=\rho_je^{\mathrm{i}\lambda_j},\ \mathrm{i}=\sqrt{-1}$, 则 (2.4) 中 $U_{l,j}t^lz_j^{-t}=V_{l,j}t^l\rho_j^{-t}e^{-\mathrm{i}(\lambda_jt-\theta_{l,j})}$ 的实部是 $V_{l,j}t^l\rho_j^{-t}\cos(\lambda_jt-\theta_{l,j})$. 所以, 满足差分方程 (2.2) 的任何实值时间序列可以表示成

$$X_t=\sum_{j=1}^{k}\sum_{l=0}^{r(j)-1}V_{l,j}t^l\rho_j^{-t}\cos(\lambda_jt-\theta_{l,j}),\quad t\in\mathbb{Z}. \tag{2.5}$$

作为 $U_{l,j}$ 的模和辐角, $V_{l,j}$ 和 $-\theta_{l,j}$ 也由 $X_0,X_1,\cdots,X_{p-1}$ 惟一决定.

如果特征多项式 $A(z)$ 的根都在单位圆外: $|z_j|>1$, 则有正数 α 使得 $1<\alpha<\min\{|z_j|:1\leq j\leq k\}$. 利用当 $t\to\infty$ 时, $t^l|z_j|^{-t}=t^l(\alpha/|z_j|)^t\alpha^{-t}=o(\alpha^{-t})$ 得到 (2.2) 的任何解 $\{X_t\}$ 满足

$$|X_t|\leq\sum_{j=1}^{k}\sum_{l=0}^{r(j)-1}|U_{l,j}|t^l|z_j|^{-t}=o(\alpha^{-t}),\quad t\to\infty. \tag{2.6}$$

这时, 我们称 $\{X_t\}$ 以负指数阶收敛到零.

如果多项式 $A(z)$ 在单位圆上有根 $z_j=e^{\mathrm{i}\lambda_j}$, 则 (2.2) 有一个实值解 $X_t=\cos(\lambda_jt),t\in\mathbb{Z}$. 这是一个周期解.

如果多项式 $A(z)$ 在单位圆内有根 $z_j=\rho_je^{\mathrm{i}\lambda_j}$, 则 (2.2) 有一个实值解 $X_t=\rho_j^{-t}\cos(\lambda_jt),t\in\mathbb{Z}$. 这个解具有快速发散到性质.

(iii) 非齐次线性差分方程

设 $A(z)$ 由 (2.3) 定义, 非齐次线性差分方程由

$$A(\mathscr{B})X_t = Y_t, \quad t \in \mathbb{Z} \tag{2.7}$$

定义, 其中 $\{Y_t\}$ 是实值时间序列.

如果 $\{X_t^{(0)}\}$ 是 (2.7) 的解 (称为特定解), 利用通解 (2.4) 可得非齐次线性差分方程 (2.7) 的通解

$$X_t = X_t^{(0)} + \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} U_{l,j} t^l z_j^{-t}, \quad t \in \mathbb{Z}. \tag{2.8}$$

2.2 自回归模型及其平稳性

如果 $\{\varepsilon_t\}$ 是白噪声 $\mathrm{WN}(0, \sigma^2)$, 实数 $a_1, a_2, \cdots, a_p (a_p \neq 0)$ 使得多项式 $A(z)$ 的零点都在单位圆外:

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j \neq 0, \quad |z| \leq 1, \tag{2.9}$$

就称 p 阶差分方程

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z} \tag{2.10}$$

是一个 p 阶自回归模型, 简称为 AR(p) 模型. 满足 AR(p) 模型 (2.10) 的平稳时间序列 $\{X_t\}$ 称为平稳解或 AR(p) 序列. 称 $\boldsymbol{a} = (a_1, a_2, \cdots, a_p)^{\mathrm{T}}$ 是 AR(p) 模型的自回归系数. 称条件 (2.9) 是稳定性条件或最小相位条件.

利用时间 t 的向后推移算子将 (2.10) 改写为 $A(\mathscr{B})X_t = \varepsilon_t$, $t \in \mathbb{Z}$. 设多项式 $A(z)$ 的互异根是 $z_1, z_2, \cdots, z_k$, 则对 $1 < \rho < \min\{|z_j|\}$, $A^{-1}(z) = 1/A(z)$ 是 $\{z : |z| \leq \rho\}$ 内的解析函数, 从而有 Taylor 级数

$$A^{-1}(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq \rho. \tag{2.11}$$

由级数 (2.11) 在 $z = \rho$ 的绝对收敛性可知

$$\psi_j = o(\rho^{-j}), \quad j \to \infty \tag{2.12}$$

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

定义 $A^{-1}(\mathscr{B}) = \sum_{j=0}^{\infty} \psi_j \mathscr{B}^j$, 于是有

$$X_t = A^{-1}(\mathscr{B})A(\mathscr{B})X_t = A^{-1}(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \tag{2.13}$$

其中, $\{\psi_j\}$ 称为平稳序列 $\{X_t\}$ 的 Wold 系数.

除了利用 $A^{-1}(z)=1/A(z)$ 的 Taylor 级数展开求 Wold 系数, 还可以采用递推公式

$$\psi_0=1,\quad \psi_k=\sum_{j=1}^{p}a_j\psi_{k-j},\quad k\geq 1. \tag{2.14}$$

其中对 $k<0$ 定义 $\psi_k=0$.

定理 设多项式 $A(z)$ 的 k 个互异根是 $z_1,z_2,\cdots,z_k$, 其中 z_j 是 $r(j)$ 重零点.

(1) 由 (2.12) 定义的时间序列 $\{X_t\}$ 是 AR(p) 模型 (2.10) 的惟一平稳解.

(2) AR(p) 模型的通解有如下的形式

$$Y_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}+\sum_{j=1}^{k}\sum_{l=0}^{r(j)-1}U_{l,j}t^l z_j^{-t},\quad t\in\mathbb{Z}, \tag{2.15}$$

其中 $\{U_{l,j}\}$ 是随机变量.

2.3 AR(p) 序列的谱密度和 Yule-Walker 方程

(i) AR(p) 序列的自协方差

设 $\{X_t\}$ 是 AR(p) 模型 (2.10) 的平稳解, 由 (2.13) 定义, 利用线性平稳序列的性质知 $\{X_t\}$ 为零均值, 自协方差函数为

$$\gamma_k=\mathrm{E}(X_{t+k}X_t)=\sigma^2\sum_{j=0}^{\infty}\psi_j\psi_{j+k},\quad k=0,1,\cdots. \tag{2.16}$$

设正数 $\rho\in(1,\min\{|z_j|\})$, 利用 Cauchy 不等式和 Wold 系数的负指数阶收敛性 (2.12) 可得

$$|\gamma_k|\leq\sigma^2\left[\sum_{j=0}^{\infty}\psi_j^2\sum_{j=0}^{\infty}\psi_{j+k}^2\right]^{1/2}\leq c_0\left[\sum_{j=k}^{\infty}\rho^{-2j}\right]^{1/2}\leq c_1\rho^{-k}, \tag{2.17}$$

其中 c_0,c_1 是正常数. 所以, 和 Wold 系数一样, AR(p) 序列的自协方差函数 $\{\gamma_k\}$ 也是以负指数阶收敛到 0 的. 而且 $\min\{|z_j|\}$ 越大, γ_k 收敛得越快.

由以上分析可知, $\{X_t\}$ 序列前后的相关性减少得很快, 这种现象称为时间序列的短记忆性.

(ii) AR(p) 序列的谱密度

由线性平稳序列的谱密度公式 (1.17) 可得 AR(p) 序列平稳解 (2.13) 的谱密度函数

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty}\psi_j e^{\mathrm{i}j\lambda}\right|^2=\frac{\sigma^2}{2\pi|A(e^{\mathrm{i}\lambda})|^2}, \tag{2.18}$$

其中 $A(z)$ 是 AR(p) 模型 (2.10) 的特征多项式.

如果 $A(z)$ 有靠近单位圆的复根 $z_j=\rho_j e^{\mathrm{i}\lambda_j}$, 则 $|A(e^{\mathrm{i}\lambda_j})|$ 会接近零, 造成谱密度在 $\lambda=\lambda_j$ 处有一个峰值, 说明平稳序列的能量在 λ_j 处比较集中, 且当 $\rho_j\to 1$ 时, $f(\lambda_j)\to\infty$, 平稳性将遭到破坏.

下面定理表明平稳序列的谱密度可以由它的自协方差函数表示出来.

定理 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

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-\mathrm{i}k\lambda}. \tag{2.19}$$

由于谱密度是实值函数, 所以 (2.19) 还可以写成

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

推论 由 (2.13) 定义的 AR(p) 序列 $\{X_t\}$ 有谱密度

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

(iii) Yule-Walker 方程

设 AR(p) 序列 $\{X_t\}$ 是 (2.10) 的解, 由 (2.13) 定义. 对任何 $k \geq 1$, 利用控制收敛定理可知

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

这说明模型 (2.10) 和解 (2.13) 的合理性: t 时刻的随机现象 X_t 和 t 时刻以后的随机干扰无关.

对 $n \geq p$, 可以将模型 (2.10) 改写为向量的形式:

$$\begin{bmatrix} X_t \\ X_{t+1} \\ \vdots \\ X_{t+n-1} \end{bmatrix} = \begin{bmatrix} X_{t-1} & X_{t-2} & \cdots & X_{t-n} \\ X_t & X_{t-1} & \cdots & X_{t+1-n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{t+n-2} & X_{t+n-3} & \cdots & X_{t-1} \end{bmatrix}\boldsymbol{a}_n + \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t+1} \\ \vdots \\ \varepsilon_{t+n-1} \end{bmatrix}, \tag{2.20}$$

其中 $\boldsymbol{a}_n = (a_{n,1}, a_{n,2}, \cdots, a_{n,n})^{\mathrm{T}} \equiv (a_1, a_2, \cdots, a_p, 0, \cdots, 0)^{\mathrm{T}}$.

记 Γ_n 为 (1.4) 式, $\boldsymbol{\gamma}_n = (\gamma_1, \gamma_2, \cdots, \gamma_n)^{\mathrm{T}}$, 在 (2.20) 两边同时乘上 X_{t-1} 后取数学期望, 可得

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

对于 $k \geq 1$, AR(p) 序列的自协方差函数满足和 AR(p) 模型 (2.10) 相应的齐次差分方程:

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

而

$$\gamma_0 = \mathrm{E}X_t^2 = \mathrm{E}\Big(\sum_{j=1}^{p}a_jX_{t-j}\Big)^2 + \mathrm{E}\varepsilon_t^2 = \boldsymbol{a}_n^{\mathrm{T}}\Gamma_n\boldsymbol{a}_n + \sigma^2 = \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{\gamma}_n + \sigma^2, \quad n \geq p.$$

定理 (Yule-Walker 方程) AR(p) 序列的自协方差函数满足

$$\boldsymbol{\gamma}_n = \Gamma_n\boldsymbol{a}_n, \quad \gamma_0 = \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{\gamma}_n + \sigma^2, \quad n \geq p, \tag{2.21}$$

其中 $\boldsymbol{a}_n = (a_{n,1}, a_{n,2}, \cdots, a_{n,n})^{\mathrm{T}} \equiv (a_1, a_2, \cdots, a_p, 0, \cdots, 0)^{\mathrm{T}}$.

(iv) 自协方差函数的周期性

对 $k<0$, 定义 $\psi_k=0$, 可以把 Yule-Walker 方程写成更一般的形式.

推论 AR(p) 序列的自协方差函数 $\{\gamma_k\}$ 满足和 AR(p) 模型 (2.10) 相应的差分方程:

$$\gamma_k-(a_1\gamma_{k-1}+a_2\gamma_{k-2}+\cdots+a_p\gamma_{k-p})=\sigma^2\psi_{-k},\quad k\in\mathbb{Z}. \tag{2.22}$$

设特征多项式 $A(z)$ 的所有根 $z_j=\rho_j e^{\mathrm{i}\lambda_j}$ 互异, 则 $A(z)$ 有因式分解

$$A(z)=(1-z/z_1)(1-z/z_2)\cdots(1-z/z_p).$$

则存在非零实数 $c_1,c_2,\ldots,c_p$, 使得

$$A(z)^{-1}=\frac{c_1}{1-z/z_1}+\frac{c_2}{1-z/z_2}+\cdots+\frac{c_p}{1-z/z_p},\quad |z|\le 1, \tag{2.23}$$

其中 $c_j=\dfrac{\prod_{i=1,i\ne j}^p z_i}{\prod_{i=1,i\ne j}^p(z_i-z_j)}$. 于是可以从差分方程 (2.22) 得到

$$\begin{aligned}\gamma_k&=A^{-1}(\mathscr{B})\sigma^2\psi_{-k}=\sigma^2\sum_{j=1}^p c_j(1-z_j^{-1}\mathscr{B})^{-1}\psi_{-k}=\sigma^2\sum_{j=1}^p c_j\sum_{l=0}^\infty z_j^{-l}\mathscr{B}^l\psi_{-k}\\&=\sigma^2\sum_{j=1}^p c_j\sum_{l=0}^\infty z_j^{-l}\psi_{-k+l}=\sigma^2\sum_{j=1}^p c_j\sum_{l=k}^\infty \psi_{l-k}z_j^{-l+k}z_j^{-k}\\&=\sigma^2\sum_{j=1}^p c_jA^{-1}(z_j^{-1})z_j^{-k}=\sigma^2\sum_{j=1}^p A_j\rho_j^{-k}\cos(\lambda_jk+\theta_j),\quad k\ge 0,\end{aligned} \tag{2.24}$$

其中 $A_j\cos(\lambda_jk+\theta_j)$ 是 $c_jA^{-1}(z_j^{-1})e^{-\mathrm{i}\lambda_jk}$ 的实部.

从式 (2.24) 可以看出, 当 $A(z)$ 有接近单位圆的复根 z_j 时, $\{\gamma_k\}$ 的频率特性开始增强. 如果 $\rho_j=\min\{\rho_k\}$, 则在 $\{X_t\}$ 中, 角频率 λ_j 的作用最重要, 此时 $\{X_t\}$ 的能量在 λ_j 处最集中.

(v) 自协方差函数的正定性

如果协方差矩阵 Γ_p 是正定的, 在 Yule-Walker 方程 (2.21) 中取 $n=p$, 可以惟一解出 AR(p) 模型的自回归系数和白噪声的方差如下

$$\boldsymbol{a}_p=\Gamma_p^{-1}\boldsymbol{\gamma}_p,\quad \sigma^2=\gamma_0-\boldsymbol{\gamma}_p^{\mathrm{T}}\Gamma_p^{-1}\boldsymbol{\gamma}_p. \tag{2.25}$$

下面的定理表明, 在实际问题中, 许多平稳序列的自协方差矩阵是正定的. 特别地, AR(p) 序列的自协方差矩阵总是正定的.

定理 设 Γ_n 是平稳序列 $\{X_t\}$ 的 n 阶自协方差矩阵.

(1) 如果 $\{X_t\}$ 的谱密度 $f(\lambda)$ 存在, 则对任何 $n\ge 1$, Γ_n 正定;

(2) 如果当 $k\to\infty$ 时, $\gamma_k\to 0$, 则对任何 $n\ge 1$, Γ_n 正定.

推论 (系数平方可和的) 线性平稳序列的自协方差矩阵总是正定的.

(vi) 时间序列的可完全预测性

对于方差有限的随机变量 $Y_1, Y_2, \cdots, Y_n$, 如果有不全为零的常数 $b_0, b_1, \cdots, b_n$, 使得

$$\mathrm{E}\Big(\sum_{j=1}^{n} b_j Y_j - b_0\Big)^2 = 0,$$

就称随机变量 $Y_1, Y_2, \cdots, Y_n$ 是线性相关的, 否则称为线性无关的. 线性相关且当 $b_n \neq 0$ 时, Y_n 可以由 $Y_1, Y_2, \cdots, Y_{n-1}$ 线性表示, 此时称 Y_n 可以由 $Y_1, Y_2, \cdots, Y_{n-1}$ 完全线性预测.

对于平稳序列 $\{X_t\}$ 的 n 阶自协方差矩阵 Γ_n 和 n 维向量 $\boldsymbol{b} = (b_1, b_2, \cdots, b_n)^{\mathrm{T}}$, 总有

$$\mathrm{E}\Big(\sum_{j=1}^{n} b_j X_{t-j} - b_0\Big)^2 \geq \mathrm{E}\Big[\sum_{j=1}^{n} b_j (X_{t-j} - \mathrm{E}X_t)\Big]^2 = \boldsymbol{b}^{\mathrm{T}}\Gamma_n \boldsymbol{b} \geq 0.$$

于是 $X_{t-1}, X_{t-2}, \cdots, X_{t-n}$ 线性无关的充分必要条件是 Γ_n 正定. 所以线性平稳序列不能完全线性预测.

2.4 平稳序列的偏相关系数和 Levinson 递推公式

(i) 最优线性预测

设 $X_1, X_2, \cdots, X_n, Y$ 为随机变量. 考虑估计问题

$$L(Y|X_1, \cdots, X_n) = \underset{\widehat{Y} = a_0 + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}}{\operatorname{argmin}} \ \mathrm{E}(Y - \widehat{Y})^2,$$

其中 $\boldsymbol{a} = (a_n, \cdots, a_1)^{\mathrm{T}}, \boldsymbol{X} = (X_n, \cdots, X_1)^{\mathrm{T}}$. 称 $L(Y|X_1, \cdots, X_n)$ 为 Y 关于 $X_1, \cdots, X_n$ 的最优线性估计.

设 $\Sigma_{\boldsymbol{X}} = \mathrm{Var}(\boldsymbol{X})$ 正定, 记 $\Sigma_{\boldsymbol{X},Y} = \mathrm{Cov}(\boldsymbol{X}, Y)$, 不妨设 $\mathrm{E}\boldsymbol{X} = \boldsymbol{0}, \mathrm{E}Y = 0$, 可求得估计问题的最小值点及误差

$$\boldsymbol{a} = \Sigma_{\boldsymbol{X}}^{-1}\Sigma_{\boldsymbol{X},Y}, \quad \mathrm{E}(Y - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X})^2 = \mathrm{Var}(Y) - \Sigma_{\boldsymbol{X},Y}^{\mathrm{T}}\Sigma_{\boldsymbol{X}}^{-1}\Sigma_{\boldsymbol{X},Y}. \tag{2.26}$$

(ii) 平稳序列的最优线性预测

设 $\{\gamma_k\}$ 和 Γ_n 分别是平稳序列 $\{X_t\}$ 的自协方差函数和 n 阶自协方差矩阵, $\boldsymbol{\gamma}_n = (\gamma_1, \gamma_2, \cdots, \gamma_n)^{\mathrm{T}}$. 方程组 $\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n$ 称为 $\{\gamma_k\}$ 的 n 阶 Yule-Walker 方程, 其中的 $\boldsymbol{a}_n = (a_{n,1}, a_{n,2}, \cdots, a_{n,n})^{\mathrm{T}}$ 称为 $\{\gamma_k\}$ 的 n 阶 Yule-Walker 系数.

设 $\{X_t\}$ 为零均值平稳序列. 考虑用 $X_1, \cdots, X_n$ 的线性组合预测 X_{n+1}. 设 $\Gamma_n > 0$, 且记 $\boldsymbol{X}_n = (X_n, \cdots, X_1)^{\mathrm{T}}$, 则由 (2.26) 的结果可知

$$L(X_{n+1}|X_1, \cdots, X_n) = (\Sigma_{\boldsymbol{X}_n}^{-1}\Sigma_{\boldsymbol{X}_n,Y})^{\mathrm{T}}\boldsymbol{X}_n = (\Gamma_n^{-1}\boldsymbol{\gamma}_n)^{\mathrm{T}}\boldsymbol{X}_n = \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{X}_n \tag{2.27}$$

所以 n 阶 Yule-Walker 系数 $\boldsymbol{a}_n$ 是用 X_t 的前面 n 个的线性组合进行预测的最优线性组合系数. 因而称 $\boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{X}_n$ 是 X_{n+1} 的最佳线性预测, 预测的均方误差为

$$\sigma_n^2 = \mathrm{E}(X_{n+1} - \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{X}_n)^2 = \gamma_0 - a_{n,1}\gamma_1 - \cdots - a_{n,n}\gamma_n. \tag{2.28}$$

如果 $\{\gamma_k\}$ 是 AR(p) 序列的自协方差函数, 则 p 阶 Yule-Walker 系数 $\boldsymbol{a}_p=(a_1,a_2,\cdots,a_p)^{\mathrm{T}}=\Gamma_p^{-1}\boldsymbol{\gamma}_p$ 就是 AR(p) 模型的自回归系数, 满足最小相位条件. 对一般的平稳序列, 有如下定理:

定理 如果实数列 $\gamma_k, k=0,1,\cdots,n$ 使得

$$\Gamma_{n+1}=\begin{bmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_n\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-1}\\ \vdots & \vdots & \ddots & \vdots\\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_0\end{bmatrix}>0,$$

则由方程组 $\Gamma_n\boldsymbol{a}_n=\boldsymbol{\gamma}_n$ 解出的 n 阶 Yule-Walker 系数 $\boldsymbol{a}_n$ 满足最小相位条件:

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

线性平稳序列的自协方差矩阵正定, 所以其任意 n 阶 Yule-Walker 系数都满足最小相位条件.

(iii) Levinson 递推公式

为了更快地计算 Yule-Walker 系数, 通常采用下面的递推公式.

定理 (Levinson 递推公式) 如果 Γ_{n+1} 正定, 对 $1\leq k\leq n$, 有

$$\begin{cases}\sigma_0^2=\gamma_0,\\ a_{1,1}=\gamma_1/\gamma_0,\\ \sigma_k^2=\sigma_{k-1}^2(1-a_{k,k}^2),\\ a_{k+1,k+1}=\dfrac{\gamma_{k+1}-a_{k,1}\gamma_k-a_{k,2}\gamma_{k-1}-\cdots-a_{k,k}\gamma_1}{\gamma_0-a_{k,1}\gamma_1-a_{k,2}\gamma_2-\cdots-a_{k,k}\gamma_k},\\ a_{k+1,j}=a_{k,j}-a_{k+1,k+1}a_{k,k+1-j},\quad 1\leq j\leq k,\end{cases}\tag{2.29}$$

其中 σ_k^2 为用 $\boldsymbol{X}_k$ 预测 X_{k+1} 时的均方误差 (2.28). 下式更加清晰地体现了 $a_{k+1,j}$ 的递推公式:

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

(iv) 偏相关系数

如果 Γ_n 正定, 称 $a_{n,n}$ 为 $\{X_t\}$ 或 $\{\gamma_k\}$ 的 n 阶偏相关系数.

设 $\{X_t\}$ 是 AR(p) 序列, 其自协方差矩阵正定. 由 Yule-Walker 方程可知, 它的 Yule-Walker 系数为

$$\boldsymbol{a}_n=(a_1,\cdots,a_p,0,\cdots,0)^{\mathrm{T}}=(a_{n,1},a_{n,2},\cdots,a_{n,n})^{\mathrm{T}},\quad n\geq p.\tag{2.30}$$

所以其偏相关系数满足

$$a_{n,n}=\begin{cases}a_p\neq 0, & n=p,\\ 0, & n>p.\end{cases}\tag{2.31}$$

此时称偏相关系数 $a_{n,n}$ 是 p 步 (后) 截尾的.

注: 偏相关截尾隐含要求自协方差矩阵正定.

反之, 下面定理告诉我们, 如果一个零均值平稳序列偏相关系数 p 步 (后) 截尾的, 则它必是 AR(p) 序列.

定理 零均值平稳序列 $\{X_t\}$ 是 AR(p) 序列的充分必要条件是, 它的偏相关系数 $a_{n,n}$ 在 p 步 (后) 截尾.

在实际应用中, 有了观测样本 $x_1, x_2, \cdots, x_n$ 后, 可以得到样本自协方差函数的点估计

$$\widehat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x}).$$

利用 Levinson 递推公式得到偏相关系数的估计 $\{\widehat{a}_{k,k}\}$. 如果其表现出 $\widehat{p}$ 步 (后) 截尾, 它即为 p 的估计, 进而可以得到 AR($\widehat{p}$) 模型回归系数 $\boldsymbol{a}_{\widehat{p}}$ 的矩估计, 对模型进行拟合. 事实上, 只要样本自协方差函数使得 $\widehat{\Gamma}_{p+1} = (\widehat{\gamma}_{k-j})_{(\widehat{p}+1)\times(\widehat{p}+1)} > 0$, 就可以得到具有稳定性的 AR($\widehat{p}$) 模型.

2.5 AR(p) 序列举例

(i) AR(1) 模型

对 $|a| < 1$, AR(1) 模型

$$X_t = aX_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2) \tag{2.32}$$

有平稳解 $X_t = \sum\limits_{j=0}^{\infty} a^j \varepsilon_{t-j}$, 自协方差函数

$$\gamma_0 = \sigma^2 \sum_{j=0}^{\infty} a^{2j} = \frac{\sigma^2}{1-a^2}, \quad \gamma_k = a\gamma_{k-1} = \cdots = a^k\gamma_0,$$

自相关系数 $\rho_k = \gamma_k/\gamma_0 = a^k$ 和谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi|1 - a\exp(\mathrm{i}\lambda)|^2} = \frac{\sigma^2}{2\pi(1 + a^2 - 2a\cos\lambda)}, \quad \lambda \in [-\pi, \pi].$$

(ii) AR(2) 模型

AR(2) 模型

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2) \tag{2.33}$$

的特征多项式 $A(z) = 1 - a_1z - a_2z^2$ 的根都在单位圆外的充分必要条件是 $a_2 \pm a_1 < 1, |a_2| < 1$, 称之为稳定性条件.

从 Yule-Walker 方程可知, 平稳解的自相关系数 $\{\rho_k\}$ 满足

$$\rho_0 = 1, \quad \rho_1 = \frac{a_1}{1-a_2}, \quad \rho_k = a_1\rho_{k-1} + a_2\rho_{k-2}, \quad k \geq 1.$$

更进一步, 当 (a_1, a_2) 在

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

内变化时, (ρ_1, ρ_2) 在

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

中取值. 称 $\mathscr{C}$ 是 AR(2) 的允许域, 称 $\mathscr{A}$ 是 AR(2) 的稳定域.

AR(2) 序列的谱密度是

$$f(\lambda) = \frac{\sigma^2}{2\pi|1 - a_1 e^{\mathrm{i}\lambda} - a_2 e^{2\mathrm{i}\lambda}|^2}.$$

当 $z_1 = \rho e^{\mathrm{i}\lambda_0}, z_2 = \rho e^{-\mathrm{i}\lambda_0}$, ρ 接近 1 时, 谱密度在 λ_0 附近有一个峰值, 此时 AR(2) 序列的角频率大约是 λ_0, 周期大约在 $2\pi/\lambda_0$ 附近.

注: 这块具体看课本 P81-86 和 PPT.

三、 滑动平均模型与自回归滑动平均模型

3.1 滑动平均模型

(i) MA(q) 模型和 MA(q) 序列

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

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

就称

$$X_t = \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z} \tag{3.1}$$

是 q 阶滑动平均模型, 简称为 MA(q) 模型, 由 (3.1) 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 $|z| \leq 1$, 就称 (3.1) 是可逆的 MA(q) 模型, 称相应的平稳序列是可逆的 MA(q) 序列.

利用时间的向后推移算子 $\mathscr{B}$ 可将 MA(q) 模型改写成 $X_t = B(\mathscr{B})\varepsilon_t$, $t \in \mathbb{Z}$. 对于可逆的 MA(q) 模型, $B^{-1}(z)$ 有 Taylor 展式

$$B^{-1}(z) = \sum_{j=0}^{\infty} \varphi_j z^j, \quad |z| \leq 1. \tag{3.2}$$

于是

$$\varepsilon_t = B^{-1}(\mathscr{B})X_t = \sum_{j=0}^{\infty} \varphi_j X_{t-j}, \quad t \in \mathbb{Z}. \tag{3.3}$$

引入 $b_0 = 1$, 对 MA(q) 序列容易计算出 $\mathrm{E}X_t = 0$,

$$\gamma_k = \mathrm{E}(X_t X_{t+k}) = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} b_j b_{j+k}, & 0 \leq k \leq q, \\ 0, & k > q. \end{cases} \tag{3.4}$$

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

$$\gamma_q = \sigma^2 b_q \neq 0, \quad \gamma_k = 0, \quad |k| > q. \tag{3.5}$$

并且有谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi}|B(e^{\mathrm{i}\lambda})|^2 = \frac{1}{2\pi}\sum_{k=-q}^{q} \gamma_k e^{-\mathrm{i}k\lambda}, \quad \lambda \in [-\pi, \pi]. \tag{3.6}$$

定理 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自协方差函数 q 后截尾, 即

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

利用如下引理就可以证明上面判定 MA(q) 序列的定理.

引理 设实常数 $\{c_j\}$ 使得 $c_q \neq 0$ 和 $g(\lambda) = \dfrac{1}{2\pi}\sum_{j=-q}^{q} c_j e^{-\mathrm{i}j\lambda} \geq 0,\ \lambda \in [-\pi, \pi]$, 则有惟一的实系数多项式

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

使得 $g(\lambda) = (\sigma^2/2\pi)|B(e^{\mathrm{i}\lambda})|^2$. 这里 σ^2 为某个正常数.

(ii) MA(q) 系数的计算

根据引理可知, 系数 $\boldsymbol{b}_q = (b_1, b_2, \cdots, b_q)^{\mathrm{T}}$ 可以由 $\gamma_0, \gamma_1, \cdots, \gamma_q$ 惟一决定. 定义

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{q\times q}, \qquad \Omega_k = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_k \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_q & \gamma_{q+1} & \cdots & \gamma_{q+k-1} \end{bmatrix},$$

$$C = (1, 0, \cdots, 0)^{\mathrm{T}}_{1\times q}, \qquad \boldsymbol{\gamma}_q = (\gamma_1, \gamma_2, \cdots, \gamma_q)^{\mathrm{T}}, \tag{3.7}$$

则有

$$\boldsymbol{b}_q = \frac{1}{\sigma^2}(\boldsymbol{\gamma}_q - A\Pi C), \quad \sigma^2 = \gamma_0 - C^{\mathrm{T}}\Pi C, \tag{3.8}$$

其中 $\Pi = \lim_{k\to\infty} \Omega_k \Gamma_k^{-1} \Omega_k^{\mathrm{T}}$.

(iii) MA(q) 模型举例

对可逆 MA(1) 序列

$$X_t = \varepsilon_t + b\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathrm{WN}(0, \sigma^2), \quad |b| < 1, \tag{3.9}$$

容易计算 $\gamma_0 = \sigma^2(1+b^2), \gamma_1 = \sigma^2 b$ 和 $\gamma_k = 0, |k| \geq 2$; 自相关系数

$$\rho_k = \begin{cases} b/(1+b^2), & |k| = 1, \\ 0, & |k| > 1; \end{cases}$$

谱密度

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

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

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

对可逆 MA(2) 序列

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}, \quad \varepsilon_t \sim \mathrm{WN}(0,\sigma^2), \quad |b| < 1, \tag{3.10}$$

其可逆域为

$$\{(b_1,b_2): B(z) \neq 0, |z| \le 1\} = \{(b_1,b_2): b_2 \pm b_1 > -1, |b_2| < 1\};$$

容易计算 $\gamma_0 = \sigma^2(1+b_1^2+b_2^2), \gamma_1 = \sigma^2(b_1+b_1b_2), \gamma_2 = \sigma^2 b_2$ 和 $\gamma_k = 0, |k| \ge 3$; 自相关系数

$$\rho_k = \begin{cases} (b_1+b_1b_2)/(1+b_1^2+b_2^2), & |k| = 1, \\ b_2/(1+b_1^2+b_2^2), & |k| = 2, \\ 0, & |k| > 2; \end{cases}$$

谱密度

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 + b_1 e^{\mathrm{i}\lambda} + b_2 e^{\mathrm{i}2\lambda}|^2, \quad \lambda \in [-\pi,\pi].$$

3.2 自回归滑动平均 (ARMA) 模型

(i) ARMA(p,q) 模型及其平稳解

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

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j \neq 0, \quad |z| \le 1, \qquad B(z) = \sum_{j=0}^{q} b_j z^j \neq 0, \quad |z| < 1, \tag{3.11}$$

则称

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z} \tag{3.12}$$

是一个自回归滑动平均模型, 简称为 ARMA(p,q) 模型. 称满足 (3.12) 的平稳序列 $\{X_t\}$ 为平稳解或 ARMA(p,q) 序列.

利用推移算子可以将 (3.12) 改写为 $A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t$, $t \in \mathbb{Z}$. 由于 $A(z)$ 满足最小相位条件, 所以有 $\rho > 1$, 使得在 $\{z : |z| \le \rho\}$ 内, $A^{-1}(z)B(z)$ 解析, 从而有 Taylor 展式

$$\Psi(z) = A^{-1}(z)B(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \le \rho. \tag{3.13}$$

易知 $\psi_j = o(\rho^{-j}), j \to \infty$, 进一步可以得到 ARMA$(p,q)$ 模型惟一平稳解

$$X_t = A^{-1}(\mathscr{B})B(\mathscr{B})\varepsilon_t = \Psi(\mathscr{B})\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}. \tag{3.14}$$

称 (3.14) 中的 $\{\psi_k\}$ 为 $\{X_t\}$ 的 Wold 系数.

模型 (3.12) 的任意解可以写成

$$Y_t = X_t + \sum_{j=1}^{k} \sum_{l=0}^{r(j)-1} V_{l,j} t^l \rho_j^{-t} \cos(\lambda_j t - \theta_{l,j}), \quad t \in \mathbb{Z}, \tag{3.15}$$

其中 X_t 是平稳解 (3.14), $z_1, z_2, \cdots, z_k$ 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 $V_{l,j}, \theta_{l,j}$ 由 $Y_0 - X_0, Y_1 - X_1, \cdots, Y_{p-1} - X_{p-1}$ 惟一决定.

对满足 (3.15) 的时间序列 $\{Y_t\}$, 有

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

所以在 t 充分大后 $\{Y_t\}$ 就和平稳解相当. 由此可以模拟产生 ARMA(p,q) 序列.

(ii) ARMA(p,q) 序列的自协方差函数

由 (3.14) 可知 ARMA(p,q) 序列 $\{X_t\}$ 的自协方差函数可由 Wold 系数表示:

$$\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \quad k \ge 0. \tag{3.16}$$

由于 $\psi = o(\rho^{-j}), j \to \infty$, 所以 $\gamma_k = o(\rho^{-j}), j \to \infty$.

记 $b_0 = 1, b_j = 0,\ j > q;\ \psi_j = 0,\ j < 0$. 由 ARMA$(p,q)$ 模型参数 $\boldsymbol{a}_p = (a_1, \cdots, a_p)^{\mathrm{T}}, \boldsymbol{b}_q = (b_1, \cdots, b_q)^{\mathrm{T}}$ 可以得到 Wold 系数的递推公式

$$\psi_j = \begin{cases} 1, & j = 0, \\ b_j + \sum\limits_{j=1}^{p} a_k \psi_{j-k}, & j = 1, 2, \cdots. \end{cases} \tag{3.17}$$

(iii) ARMA(p,q) 模型的可识别性

ARMA(p,q) 模型中要求 $A(z)$ 和 $B(z)$ 没有公因子, 这个条件保证了模型的可识别性, 即可以由平稳解的自协方差函数惟一决定模型的参数 $(\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}}, \sigma^2) = (a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_q, \sigma^2)$.

引理 设 $\{X_t\}$ 是 (3.12) 的平稳解. 如果又有白噪声 $\{\eta_t\}$ 和实系数多项式 $C(\mathscr{B}), D(\mathscr{B})$ 使得

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

成立, 则 $C(z)$ 的阶数 $\ge p$, $D(z)$ 的阶数 $\ge q$.

下面导出 ARMA(p,q) 模型的 Yule-Walker 方程. 首先注意到有 $\mathrm{E}(\varepsilon_{t+k} X_t) = 0, k > 0$. 这表明 ARMA$(p,q)$ 序列 $\{X_t\}$ 的合理性: X_t 不受 t 之后的噪声干扰. 另一方面,

$$\begin{aligned} \gamma_k = \mathrm{E}(X_t X_{t-k}) &= \mathrm{E}\Big[\Big(\sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}\Big) X_{t-k}\Big] \\ &= \sum_{j=1}^{p} a_j \gamma_{k-j} + \mathrm{E}\Big(\sum_{j=0}^{q} b_j \varepsilon_{t-j} \sum_{l=0}^{\infty} \psi_l \varepsilon_{t-k-l}\Big) = \sum_{j=1}^{p} a_j \gamma_{k-j} + \sigma^2 \sum_{j=0}^{q} b_j \psi_{j-k}, \quad k \in \mathbb{Z} \end{aligned}$$

所以 ARMA(p,q) 序列的自协方差函数 $\{\gamma_k\}$ 满足差分方程

$$\gamma_k-\sum_{j=1}^{p}a_j\gamma_{k-j}=\begin{cases}\sigma^2\sum\limits_{j=0}^{q}b_j\psi_{j-k}, & k\le q,\\ 0, & k>q,\end{cases}\tag{3.18}$$

其中对 $k<0$, 补充定义 $\psi_k=0$, 并有 $b_0=1$.

对 $k>q$, 可以得出延伸的 Yule-Walker 方程:

$$\begin{bmatrix}\gamma_{q+1}\\ \gamma_{q+2}\\ \vdots\\ \gamma_{q+p}\end{bmatrix}=\begin{bmatrix}\gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1}\\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2}\\ \vdots & \vdots & \ddots & \vdots\\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q\end{bmatrix}\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_p\end{bmatrix}.\tag{3.19}$$

如果 (3.19) 的系数矩阵 $\Gamma_{p,q}=(\gamma_{|q+i-j|})_{i,j=1,2,\cdots,p}$ 可逆, 则已知 p,q 时, 参数 $a_1,a_2,\cdots,a_p$ 可以由 $\gamma_0,\gamma_1,\cdots,\gamma_{q+p}$ 惟一决定. 更进一步, 此时

$$Y_t=X_t-\sum_{j=1}^{p}a_jX_{t-j}=A(\mathscr{B})X_t=B(\mathscr{B})\varepsilon_t,\quad t\in\mathbb{Z}$$

是一个 MA(q) 序列, 它的自协方差函数 $\{\gamma_y(k)\}$ 是 q(步) 后截尾的, 且

$$\begin{aligned}\gamma_y(k)=\mathrm{E}(Y_tY_{t-k})&=\mathrm{E}\left(\sum_{j=0}^{p}a_jX_{t-j}\sum_{l=0}^{p}a_lX_{t-k-l}\right)=\sum_{j=0}^{p}\sum_{l=0}^{p}a_ja_l\gamma_{k+l-j},\\ &=\boldsymbol{a}^{\mathrm{T}}\begin{bmatrix}\gamma_k & \gamma_{k+1} & \gamma_{k+2} & \cdots & \gamma_{k+p}\\ \gamma_{k-1} & \gamma_k & \gamma_{k+1} & \cdots & \gamma_{k+p-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \gamma_{k-p} & \gamma_{k-p+1} & \gamma_{k-p+2} & \cdots & \gamma_k\end{bmatrix}\boldsymbol{a},\quad |k|\le q,\end{aligned}\tag{3.20}$$

其中 $\boldsymbol{a}^{\mathrm{T}}=(a_0,a_1,\cdots,a_p),a_0=-1$.

根据 3.1(ii) 提供的 MA(q) 系数的计算方法可以从 $\gamma_y(0),\gamma_y(1),\cdots,\gamma_y(q)$ 惟一解出参数 $b_1,b_2,\cdots,b_q$ 和 σ^2. 所以只要 $\Gamma_{p,q}$ 可逆, ARMA(p,q) 序列的自协方差函数和 ARMA(p,q) 模型的参数 $(\boldsymbol{a}_q^{\mathrm{T}},\boldsymbol{b}_q^{\mathrm{T}},\sigma^2)$ 相互惟一决定.

下面定理说明了 $\Gamma_{p,q}$ 可逆的条件.

定理 设 $\{\gamma_k\}$ 为 ARMA(p,q) 序列 $\{X_t\}$ 的自协方差函数列, 则 $m\ge p$ 时,

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

可逆.

证明: 如果 $\Gamma_{m,q}(m\times m$ 矩阵) 不满秩, 则存在 $\boldsymbol{\beta}=(\beta_0,\beta_1,\cdots,\beta_{m-1})^{\mathrm T}\neq\mathbf{0}$, 使得 $\Gamma_{m,q}\boldsymbol{\beta}=\mathbf{0}$, 即

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

再利用延伸的 Yule-Walker 方程 (3.18) 可得

$$\sum_{l=0}^{m-1}\beta_l\gamma_{q+m-l}=\sum_{k=1}^{p}a_k\sum_{l=0}^{m-1}\beta_l\gamma_{q+m-l-k}=0.$$

依次类推可以得到对 $k\geq 0$, $\sum_{l=0}^{m-1}\beta_l\gamma_{q+k-l}=0$. 令 $Y_t=\sum_{l=0}^{m-1}\beta_lX_{t-l}$, 则 $\{Y_t\}$ 是零均值平稳列. 利用

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

可得 $\mathrm{E}(Y_tY_{t-q-k})=0, k\geq 0$. 所以 $\gamma_y(k)=\mathrm{E}(Y_0Y_k)$ 是 $(q-1)$ (步) 后截尾的, $\{Y_t\}$ 是 MA$(q-1)$ 序列. 所以存在 $\alpha_0,\alpha_1,\cdots,\alpha_{q-1}$ 使得

$$\sum_{l=0}^{m-1}\beta_lX_{t-l}=\sum_{j=0}^{q-1}\alpha_j\varepsilon_{t-j},\quad \{\varepsilon_t\}\sim\mathrm{WN}(0,\sigma^2),$$

与本节引理矛盾.

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

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

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

(iv) ARMA 序列的谱密度和可逆性

由于 ARMA 序列的自协方差函数 $\{\gamma_k\}$ 绝对可和, 所以 ARMA(p,q) 序列 (3.14) 有谱密度

$$f(\lambda)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_ke^{-\mathrm{i}k\lambda}=\frac{\sigma^2}{2\pi}\Big|\sum_{j=0}^{\infty}\psi_je^{\mathrm{i}j\lambda}\Big|^2=\frac{\sigma^2}{2\pi}\Big|\frac{B(e^{\mathrm{i}\lambda})}{A(e^{\mathrm{i}\lambda})}\Big|^2. \tag{3.21}$$

形如 (3.21) 的谱密度被称为有理谱密度.

在前面 ARMA(p,q) 模型的定义中, 如果进一步要求 $B(z)$ 在单位圆上无根:

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

就称 ARMA(p,q) 模型 (3.12) 为可逆 ARMA 模型, 称相应的平稳解为可逆 ARMA(p,q) 序列. 可逆 ARMA(p,q) 序列和它的噪声序列可以相互线性表示:

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

3.3 广义 ARMA 模型和 ARIMA(p,d,q) 模型

(i) 广义 ARMA 模型

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 $a_p b_q \neq 0$ 和

$$A(z)=1-\sum_{j=1}^{p}a_j z^j,\qquad B(z)=1+\sum_{j=1}^{q}b_j z^j,$$

如果不对 $A(z),B(z)$ 的根做任何其他限制, 则称差分方程

$$A(\mathscr{B})X_t=B(\mathscr{B})\varepsilon_t,\quad t\in\mathbb{Z} \tag{3.22}$$

为广义 ARMA(p,q) 模型. 称满足 (3.22) 的时间序列 $\{X_t\}$ 为广义 ARMA(p,q) 序列.

若 $A(z)$ 在单位圆上有根, 可以证明 (3.22) 没有平稳解.

若 $A(z)$ 在单位圆上没有根, 则从 (3.22) 可以得到惟一的平稳解

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

若 $A(z)$ 在单位圆内有根, 则这个平稳解是白噪声的双边无穷滑动和, 从实际意义上看不合理: t 时的观测受到 t 以后的干扰的影响. 此时称为爆炸模型.

(ii) 求和 ARIMA(p,d,q) 模型

采用求和 ARIMA(p,d,q) 模型拟合数据的过程, 实质上是先对观测数据进行 d 次差分处理, 然后再拟合 ARMA(p,q) 模型. 具体来说, 设 d 是一个正整数, 如果

$$Y_t=\nabla^d X_t=(1-\mathscr{B})^d X_t=\sum_{k=0}^{d}\mathrm{C}_d^k(-1)^k X_{t-k},\quad t\in\mathbb{Z} \tag{3.23}$$

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 ARIMA(p,d,q) 序列. 所以 ARIMA(p,d,q) 序列满足的模型为

$$A(\mathscr{B})(1-\mathscr{B})^d X_t=B(\mathscr{B})\varepsilon_t,\quad t\in\mathbb{Z}, \tag{3.24}$$

其中 $A(z),B(z),\{\varepsilon_t\}$ 满足 ARMA(p,q) 模型的定义.

注: (1) 通常 $d=1$ 或 $d=2$; (2) $\{X_t\}$ 是非平稳序列.

(iii) 特殊 ARIMA(p,d,q) 模型

(1) ARI$(1,1)$ 模型: 对 $|a|<1$, 模型

$$X_t-X_{t-1}=a(X_{t-1}-X_{t-2})+\varepsilon_t \iff X_t=(1+a)X_{t-1}-aX_{t-2}+\varepsilon_t \tag{3.25}$$

的非平稳解为 $X_t=\sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k}$, 其中 $\Psi(z)=A^{-1}(z)=\sum_{k=0}^{\infty}\psi_k z^k$.

为了求权系数 ψ_k, 利用 $A(z)\Psi(z)=1$ 可以得到如下方程

$$[1-(1+a)z+az^2]\sum_{k=0}^{\infty}\psi_k z^k=1$$

比较系数可得 $\psi_k=(1-a^{k+1})/(1-a), k\geq 0$.

(2) IMA(1,1) 模型: 在经济和商业中产生的序列, 常具有该模型性质.

$$X_t=X_{t-1}+\varepsilon_t-b\varepsilon_{t-1} \tag{3.26}$$

通常假设序列的首次观测时间为 $-m$, 则当 $t<-m$ 时, $X_t=0$, 所以

$$\begin{aligned}X_t&=X_{t-1}+\varepsilon_t-b\varepsilon_{t-1}\\&=X_{t-2}+\varepsilon_t+(1-b)\varepsilon_{t-1}-b\varepsilon_{t-2}\\&=\cdots=\varepsilon_t+(1-b)\varepsilon_{t-1}+\cdots+(1-b)\varepsilon_{-m}-b\varepsilon_{-m-1}\end{aligned}$$

所以 $\mathrm{Var}(X_t)=[1+b^2+(1-b)^2(t+m)]\sigma^2$, 且对较大的 m 和中等大小的 k, 有

$$\mathrm{Corr}(X_t,X_{t+k})\approx\sqrt{\frac{t+m}{t+m+k}}\approx 1,$$

所以当 t 增加时, $\mathrm{Var}(X_t)$ 会随着增加并无限增大, 同时, 对多个滞后期数 k, X_t 和 X_{t+k} 呈现高度正相关.

3.4 季节 ARIMA 模型

(i) 季节 MA 模型

定义季节周期为 s 的 Q 阶季节 MA(Q) 模型如下:

$$X_t=\varepsilon_t+\Theta_1\varepsilon_{t-s}+\Theta_2\varepsilon_{t-2s}+\cdots+\Theta_Q\varepsilon_{t-Qs} \tag{3.27}$$

其中特征多项式满足

$$B(z)=1+\Theta_1 z^s+\Theta_2 z^{2s}+\cdots+\Theta_Q z^{Qs}\neq 0,\quad |z|\leq 1. \tag{3.28}$$

模型得到的序列 $\{X_t\}$ 是平稳序列, 且其自相关函数只在 $s,2s,\cdots,Qs$ 等季节滞后上非零, 且有

$$\rho_{ks}=\frac{\Theta_k+\Theta_1\Theta_{k+1}+\cdots+\Theta_{Q-k}\Theta_Q}{1+\Theta_1^2+\Theta_2^2+\cdots+\Theta_Q^2},\quad k=1,2,\cdots,Q.$$

注: 模型可逆的条件是特征方程 $B(z)=0$ 的所有根都在单位圆外; 季节 MA(Q) 模型可以看作阶数 $q=Qs$ 的非季节性 MA 模型的特例.

(ii) 季节 AR 模型

定义季节周期为 s 的 P 阶季节 AR(P) 模型如下:

$$X_t = \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \cdots + \Phi_P X_{t-Ps} + \varepsilon_t \tag{3.29}$$

其中特征多项式满足

$$A(z) = 1 - \Phi_1 z^s - \Phi_2 z^{2s} - \cdots - \Phi_P z^{Ps} \neq 0, \quad |z| \leq 1. \tag{3.30}$$

该模型有平稳解, 且其自相关函数只在 $s, 2s, \cdots, Qs$ 等季节滞后上非零, 并按指数衰减函数趋向于零.

注: 季节 AR(P) 模型可以看作阶数 $p = Ps$ 的非季节性 AR 模型的特例.

(iii) 季节 ARMA 模型

定义季节周期为 s 的季节 ARMA$(p,q) \times (P,Q)_s$ 模型如下:

$$A_0(\mathscr{B})A(\mathscr{B}^s)X_t = B_0(\mathscr{B})B(\mathscr{B}^s)\varepsilon_t, \tag{3.31}$$

其中 p, q 分别是多项式 $A_0(z), B_0(z)$ 的阶数, P, Q 分别是多项式 $A(z), B(z)$ 的阶数. 一般 p, q 不超过 3.

(iv) 非平稳季节 ARIMA 模型

如果 $Y_t = \nabla^d \nabla_s^D X_t$ 满足季节周期为 s 的季节 ARMA$(p,q) \times (P,Q)_s$ 模型, 则称过程 $\{X_t\}$ 为季节周期为 s、非季节阶数为 p, d 和 q、季节阶数为 P, D 和 Q 的乘法季节 ARIMA 模型, 或季节周期为 s 的 ARIMA$(p,d,q) \times (P,D,Q)_s$ 模型.

实际问题中 d 和 D 都很小, 一般是 $D = 0$ 或 1.

四、 均值和自协方差函数的估计

4.1 均值的估计

设 $x_1, x_2, \cdots, x_N$ 是平稳序列 $\{X_t\}$ 的观测, 则均值 $\mu = \mathrm{E}X_t$ 的点估计为

$$\widehat{\mu} = \overline{x}_N = \frac{1}{N}\sum_{k=1}^{N} x_k. \tag{4.1}$$

在研究估计量的统计性质时, 用大写的 X_t 代替 x_t, 用大写的 $\overline{X}_N$ 代替 $\overline{x}_N$.

(i) 相合性

对于 (4.1) 定义的 $\overline{X}_N$, 有

$$\mathrm{E}\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N}\mathrm{E}X_k = \frac{1}{N}\sum_{k=1}^{N}\mu = \mu,$$

所以 $\overline{X}_N$ 是均值 μ 的无偏估计.

另一方面, 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

$$\begin{aligned}\mathrm{E}(\overline{X}_N-\mu)^2&=\mathrm{E}\Big[\frac{1}{N}\sum_{k=1}^{N}(X_k-\mu)\Big]^2=\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\gamma_{k-j}=\frac{1}{N^2}\sum_{j=1}^{N}\sum_{m=1-j}^{N-j}\gamma_m\\&=\frac{1}{N^2}\sum_{m=-N+1}^{N-1}\sum_{j=(1-m)\vee 1}^{(N-m)\wedge N}\gamma_m=\frac{1}{N^2}\sum_{m=-N+1}^{N-1}(N-|m|)\gamma_m\\&\leq\frac{1}{N}\sum_{m=-N}^{N}|\gamma_m|\to 0,\quad N\to\infty,\end{aligned}$$

所以 $\overline{X}_N$ 均方收敛到 μ. 再利用 Chebyshev 不等式可知 $\overline{X}_N$ 依概率收敛到 μ, 即 $\overline{X}_N$ 是 μ 的相合估计. 上述论述总结为如下定理.

定理 设平稳序列 $\{X_t\}$ 有均值 μ 和自协方差函数 $\{\gamma_k\}$, 则

(1) $\overline{X}_N$ 是 μ 的无偏估计;

(2) 如果 $\gamma_k\to 0$, 则 $\overline{X}_N$ 是 μ 的相合估计;

(3) 如果 $\{X_t\}$ 是严平稳遍历序列, 则 $\overline{X}_N$ 是 μ 的强相合估计.

(ii) 中心极限定理

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

$$X_t=\mu+\sum_{k=-\infty}^{\infty}\psi_k\varepsilon_{t-k},\quad t\in\mathbb{Z},\tag{4.2}$$

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

$$f(\lambda)=\frac{\sigma^2}{2\pi}\Big|\sum_{k=-\infty}^{\infty}\psi_k e^{-\mathrm{i}k\lambda}\Big|^2$$

在 $\lambda=0$ 处连续, 并且 $f(0)\neq 0$, 则当 $n\to\infty$ 时,

$$\sqrt{N}(\overline{X}_N-\mu)\xrightarrow{d}N(0,2\pi f(0)).\tag{4.3}$$

当 $\{\psi_k\}$ 绝对可和时, $f(\lambda)$ 连续. 可以得到如下推论.

推论 如果 $\sum\limits_{k=-\infty}^{\infty}|\psi_k|<\infty$ 和 $\sum\limits_{k=-\infty}^{\infty}\psi_k\neq 0$ 成立, 则当 $n\to\infty$ 时, (4.3) 式成立, 并且

$$2\pi f(0)=\gamma_0+2\sum_{j=1}^{\infty}\gamma_j.\tag{4.4}$$

由此可以给出均值 $\mu=\mathrm{E}X_1$ 的渐近置信区间和关于假设 $\mu=\mu_0$ 的检验.

(iii) 收敛速度

收敛速度的描述方法之一是所谓的重对数律. 我们把 $\overline{X}_N$ 服从的重对数律叙述为如下定理.

定理 设 $\{\varepsilon_t\}$ 是独立同分布的 $\mathrm{WN}(0,\sigma^2)$. 线性平稳序列 $\{X_t\}$ 由 (4.2) 定义, 谱密度 $f(0)\neq 0$. 当以下的条件之一成立时:

(1) 当 $k\to\pm\infty$, $\psi_{|k|}$ 以负指数阶收敛于 0;

(2) 谱密度 $f(\lambda)$ 在 $\lambda=0$ 处连续, 且 $\mathrm{E}|\varepsilon_t|^r<\infty$ 对某个 $r>2$ 成立, 则有重对数律

$$\begin{aligned}\limsup_{N\to\infty}\sqrt{\frac{N}{2\ln\ln N}}(\overline{X}_N-\mu)&=\sqrt{2\pi f(0)},\quad \text{a.s.},\\ \liminf_{N\to\infty}\sqrt{\frac{N}{2\ln\ln N}}(\overline{X}_N-\mu)&=-\sqrt{2\pi f(0)},\quad \text{a.s..}\end{aligned}\tag{4.5}$$

重对数律成立时, 得到的收敛速度的阶数一般是 $O\Big(\sqrt{\dfrac{2\ln\ln N}{N}}\Big)$. 除了个别的情况, 这个阶数一般不能再被改进.

4.2 自协方差函数的估计

(i) γ_k 的点估计和样本自协方差矩阵的正定性

给定平稳序列 $\{X_t\}$ 的 N 个样本 $x_1,x_2,\cdots,x_N$, 样本自协方差函数

$$\widehat{\gamma}_k=\frac{1}{N}\sum_{j=1}^{N-k}(x_j-\overline{x}_N)(x_{j+k}-\overline{x}_N),\ 0\le k\le N-1;\quad \widehat{\gamma}_{-k}=\widehat{\gamma}_k,\ 1\le k\le N-1\tag{4.6}$$

是 $\{X_t\}$ 的自协方差函数 $\gamma_k=\mathrm{Cov}(X_1,X_{k+1})$ 的点估计. 样本自相关系数

$$\widehat{\rho}_k=\widehat{\gamma}_k/\widehat{\gamma}_0,\quad |k|\le N-1\tag{4.7}$$

是 $\{X_t\}$ 的自相关系数 $\rho_k=\gamma_k/\gamma_0$ 的点估计.

样本自协方差函数点估计的分母取 N 的一个重要原因是为了保证样本自协方差矩阵正定. 事实上, 只要 $x_1,x_2,\cdots,x_N$ 不全相同, 则 $y_i=x_i-\overline{x}_N, i=1,2,\cdots,N$ 不全为零, 于是 $N\times(2N-1)$ 矩阵

$$\boldsymbol{A}=\begin{bmatrix}0&\cdots&0&y_1&y_2&\cdots&y_{N-1}&y_N\\0&\cdots&y_1&y_2&y_3&\cdots&y_N&0\\\vdots&\ddots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\y_1&\cdots&y_{N-1}&y_N&0&\cdots&0&0\end{bmatrix}\tag{4.8}$$

是满秩的, 则样本自协方差矩阵 $\widehat{\Gamma}_N=\dfrac{1}{N}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 正定. 主子式 $\widehat{\Gamma}_n(1\le n\le N)$ 也是正定的.

(ii) $\widehat{\gamma}_k$ 的相合性

定理 设平稳序列的样本自协方差函数 $\widehat{\gamma}_k$ 由 (4.6) 定义.

(1) 若当 $k\to\infty$ 时 $\gamma_k\to 0$, 则对每个确定的 k, $\widehat{\gamma}_k$ 是 γ_k 的渐近无偏估计:

$$\lim_{N\to\infty}\mathrm{E}\widehat{\gamma}_k=\gamma_k.$$

(2) 若 $\{X_t\}$ 是严平稳遍历序列, 则对每个确定的 k, $\widehat{\gamma}_k$ 和 $\widehat{\rho}_k$ 分别是 γ_k 和 ρ_k 的强相合估计:

$$\lim_{N\to\infty}\widehat{\gamma}_k=\gamma_k,\ \text{a.s.},\qquad \lim_{N\to\infty}\widehat{\rho}_k=\rho_k,\ \text{a.s..}$$

(iii) $\widehat{\gamma}_k$ 的渐近分布

设 $\{\varepsilon_t\}$ 是四阶矩有限, 独立同分布的 $\mathrm{WN}(0,\sigma^2)$, 实数列 $\{\psi_k\}$ 平方可和. 线性平稳序列

$$X_t=\sum_{j=-\infty}^{\infty}\psi_j\varepsilon_{t-j},\quad t\in\mathbb{Z} \tag{4.9}$$

有自协方差函数 $\gamma_k=\sigma^2\sum\limits_{j=-\infty}^{\infty}\psi_j\psi_{j+k}$ 和谱密度函数 $f(\lambda)=\dfrac{\sigma^2}{2\pi}\Big|\sum\limits_{j=-\infty}^{\infty}\psi_je^{\mathrm{i}j\lambda}\Big|^2$.

当自协方差函数 $\{\gamma_k\}$ 平方可和时, 对于独立同分布的正态标准白噪声 $\{W_t\}$, 定义正态时间序列

$$\xi_j=(M_0\gamma_j)W_0+\sum_{t=1}^{\infty}(\gamma_{t+j}+\gamma_{t-j})W_t,\quad j\geq 0, \tag{4.10}$$

$$R_j=\sum_{t=1}^{\infty}(\rho_{t+j}+\rho_{t-j}-2\rho_t\rho_j)W_t,\quad j\geq 1, \tag{4.11}$$

其中 $M_0=(\mu_4-\sigma^4)^{1/2}/\sigma^2,\mu_4=\mathrm{E}\varepsilon_1^4$. 则有如下中心极限定理.

定理 设 $\{\varepsilon_t\}$ 是独立同分布的 $\mathrm{WN}(0,\sigma^2)$, 满足 $\mu_4=\mathrm{E}\varepsilon_1^4<\infty$. 如果线性平稳序列 (4.9) 的谱密度平方可积, 即

$$\int_{-\pi}^{\pi}f(\lambda)\mathrm{d}\lambda=\int_{-\pi}^{\pi}\frac{\sigma^2}{2\pi}\Big|\sum_{j=-\infty}^{\infty}\psi_je^{\mathrm{i}j\lambda}\Big|^2\mathrm{d}\lambda<\infty,$$

则对任何正整数 h, 当 $N\to\infty$ 时, 有以下结果:

(1) $\sqrt{N}(\widehat{\gamma}_0-\gamma_0,\widehat{\gamma}_1-\gamma_1,\cdots,\widehat{\gamma}_h-\gamma_h)$ 依分布收敛到 $(\xi_0,\xi_1,\cdots,\xi_h)$.

(2) $\sqrt{N}(\widehat{\rho}_0-\rho_0,\widehat{\rho}_1-\rho_1,\cdots,\widehat{\rho}_h-\rho_h)$ 依分布收敛到 $(R_0,R_1,\cdots,R_h)$.

实际问题中谱密度平方可积的条件很难验证, 下面定理解决了这一问题.

定理 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函数平方可积.

为了方便白噪声的检验, 下面中心极限定理不要求噪声项的四阶矩有限.

定理 设 $\{\varepsilon_t\}$ 是独立同分布的 $\mathrm{WN}(0,\sigma^2)$, 平稳序列 $\{X_t\}$ 由 (4.9) 定义. 如果自协方差函数 $\{\gamma_k\}$ 平方可和, 并且对某个常数 $\alpha>0.5$,

$$m^\alpha\sum_{|k|\geq m}\psi_k^2\to 0,\quad m\to\infty, \tag{4.12}$$

则对任何正整数 h, 当 $N\to\infty$ 时, $\sqrt{N}(\widehat{\rho}_0-\rho_0,\widehat{\rho}_1-\rho_1,\cdots,\widehat{\rho}_h-\rho_h)$ 依分布收敛到 $(R_0,R_1,\cdots,R_h)$.

推论 如果 $\{X_t\}$ 是独立同分布的白噪声,

$$\widehat{\rho}_k=\frac{\sum_{t=1}^{N-k}(x_t-\overline{x}_N)(x_{t+k}-\overline{x}_N)}{\sum_{t=1}^{N}(x_t-\overline{x}_N)^2}$$

是样本自相关系数, 则对任何正整数 h,

(1) $\sqrt{N}(\widehat{\rho}_1,\widehat{\rho}_2,\cdots,\widehat{\rho}_h)$ 依分布收敛到多元正态分布 $N(0,\boldsymbol{I}_h)$, 其中 $\boldsymbol{I}_h$ 是 $h\times h$ 的单位矩阵.

(2) 如果 $\mu_4=\mathrm{E}\varepsilon_t^4<\infty$, 则 $\sqrt{N}(\widehat{\gamma}_0-\sigma^2,\widehat{\gamma}_1,\cdots,\widehat{\gamma}_h)$ 依分布收敛到 $\sigma^2(M_0W_0,W_1,\cdots,W_h)$.

4.3 白噪声检验

(i) 白噪声的 χ^2 检验

根据 4.2(iii) 的推论, 对于充分大的 N, 构造检验统计量

$$\widehat{\chi}^2(m) = N(\widehat{\rho}_1^2 + \widehat{\rho}_2^2 + \cdots + \widehat{\rho}_m^2) \overset{approx}{\sim} \chi^2(m) \tag{4.13}$$

这里要求 $m \leq \sqrt{N}$. 由此可以做出假设检验:

$$H_0 : \{X_t\} \text{ 是独立白噪声 } \quad \text{vs} \quad H_1 : \{X_t\} \text{ 是相关序列.}$$

由于在原假设 H_0 下, $\rho_1^2 + \rho_2^2 + \cdots + \rho_m^2 = 0$, 所以当检验统计量取值较大时拒绝原假设.

(ii) 白噪声的正态分布检验

检验 $\{X_t\}$ 是白噪声的另一个简单方法是计算

$$Q(m) = \frac{1}{m}\sharp\{j \big| \sqrt{N}|\widehat{\rho}_j| \geq 1.96,\ 1 \leq j \leq m\}, \tag{4.14}$$

其中 $\sharp A$ 表示集合 A 中的元素个数.

根据 4.2(iii) 的推论, 在原假设 H_0 下, 对较大的 N, 应当有 95% 的 $\sqrt{N}|\widehat{\rho}_j| \leq 1.96$, 所以当 $Q(m)$ 取值 ≥ 0.05 时, 应当拒绝原假设, 认为 $\{X_t\}$ 是相关序列.

五、 时间序列的预报

5.1 最佳线性预测的基本性质

(i) 最佳线性预测

设 $X_1, X_2, \cdots, X_n$ 为某时间序列的一段, Y 是一个随机变量. 考虑使用 $X_1, X_2, \cdots, X_n$ 的线性组合对 Y 进行预测. 记

$$\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}, \quad \boldsymbol{a} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}} \in \mathbb{R}^n,$$

则 Y 的线性预测有以下形式 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}$. 现要找一个 $\boldsymbol{a}$ 使得 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}$ 离 Y 最近.

定义 设 Y 和 $X_j (1 \leq j \leq n)$ 是均值为零, 方差有限的随机变量. 如果 $\boldsymbol{a} \in \mathbb{R}^n$, 使得对任何的 $\boldsymbol{b} \in \mathbb{R}^n$, 有

$$\mathrm{E}(Y - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X})^2 \leq \mathrm{E}(Y - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X})^2,$$

则称 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}$ 是用 $X_1, X_2, \cdots, X_n$ 对 Y 进行预测的最佳线性预测, 记作 $L(Y|\boldsymbol{X})$ 或 $\widehat{Y}$. 于是

$$\widehat{Y} = L(Y|\boldsymbol{X}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}. \tag{5.1}$$

称 $Y - \widehat{Y} = Y - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}$ 是预测误差, $\mathrm{E}(Y - \widehat{Y})^2 = \mathrm{E}(Y - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X})^2$ 是预测的均方误差.

更进一步, 如果 $\mathrm{E}Y = b, \mathrm{E}\boldsymbol{X} = \boldsymbol{\mu}$, 定义

$$L(Y|\boldsymbol{X}) = L(Y - b|\boldsymbol{X} - \boldsymbol{\mu}) + b, \tag{5.2}$$

并称 $L(Y|\boldsymbol{X})$ 是用 $X_1, X_2, \cdots, X_n$ 对 Y 进行预测时的最佳线性预测. 当 $b = 0, \boldsymbol{\mu} = \boldsymbol{0}$ 时, 两种定义 (5.1) 与 (5.2) 一致.

用 $\Gamma = \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})$ 表示 $\boldsymbol{X}$ 的协方差矩阵, 用 $\Sigma_{\boldsymbol{X}Y} = \mathrm{E}(\boldsymbol{X}Y)$ 表示 $\boldsymbol{X}$ 和 Y 的协方差向量. 以下总假设所述随机变量的均值为零. 最佳线性预测有以下的基本性质.

性质 1 如果 $\boldsymbol{a} \in \mathbb{R}^n$, 使得

$$\Gamma\boldsymbol{a} = \mathrm{E}(\boldsymbol{X}Y) = \Sigma_{\boldsymbol{X}Y}, \tag{5.3}$$

则 $L(Y|\boldsymbol{X}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}$, 并且有

$$\mathrm{E}(Y - L(Y|\boldsymbol{X}))^2 = \mathrm{E}Y^2 - \mathrm{E}[L(Y|\boldsymbol{X})]^2 = \mathrm{E}Y^2 - \boldsymbol{a}^{\mathrm{T}}\Gamma\boldsymbol{a}. \tag{5.4}$$

注: 如果 Γ 和 $\Sigma_{\boldsymbol{X}Y}$ 已知, 以 $\boldsymbol{a}$ 为未知数的线性方程组 (5.3) 被称为预测方程. 此时有 $Y - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}$ 与 $\boldsymbol{X}$ 正交, 式 (5.4) 可被理解为空间中的勾股定理.

性质 2 (1) 如果 $\Gamma = \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})$ 可逆, 则 $\boldsymbol{a} = \Gamma^{-1}\Sigma_{\boldsymbol{X}Y}$;

(2) 预测方程 $\Gamma\boldsymbol{a} = \Sigma_{\boldsymbol{X}Y}$ 总有解;

(3) 如果 $\det(\Gamma) = 0$, 取正交矩阵 A, 使得

$$A\Gamma A = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \cdots, 0), \quad \lambda_j \neq 0, j = 1, 2, \cdots, r.$$

定义 $\boldsymbol{Z} = A\boldsymbol{X} = (Z_1, Z_2, \cdots, Z_n)^{\mathrm{T}} = (Z_1, Z_2, \cdots, Z_r, 0, \cdots, 0)^{\mathrm{T}}$ 和 $\boldsymbol{\xi} = (Z_1, Z_2, \cdots, Z_r)^{\mathrm{T}}$, 则 $\mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}})$ 正定, 并且当取

$$\boldsymbol{\alpha} = [\mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}})]^{-1}\mathrm{E}(\boldsymbol{\xi}Y) \tag{5.5}$$

时, $L(Y|\boldsymbol{X}) = L(Y|\boldsymbol{\xi}) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\xi}$.

注: 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; 第三条说明当第一条不成立时, $L(Y|\boldsymbol{X})$ 可以通过 $\boldsymbol{X}$ 的基表示.

性质 3 尽管 $\boldsymbol{a}$ 由 $\Gamma\boldsymbol{a} = \Sigma_{\boldsymbol{X}Y}$ 决定时可以不惟一, 但 $L(Y|\boldsymbol{X})$ 总是 (a.s.) 惟一的.

性质 4 (1) 如果 $\mathrm{E}(\boldsymbol{X}Y) = 0$, 则 $L(Y|\boldsymbol{X}) = 0$;

(2) 如果 $Y = \sum_{j=1}^{m} b_j X_j$, 则 $L(Y|\boldsymbol{X}) = Y$.

性质 5 设 $Y_1, Y_2, \cdots, Y_n$ 是随机变量, b_j 是常数. 如果 $Y = \sum_{j=1}^{m} b_j Y_j$, 则

$$L(Y|\boldsymbol{X}) = \sum_{j=1}^{m} b_j L(Y_j|\boldsymbol{X}).$$

这一性质告诉我们, 求最佳线性预测的运算 $L(\cdot|\boldsymbol{X})$ 是一种线性运算.

性质 6 设 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$, $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_m)^{\mathrm{T}}$. 如果 $\mathrm{E}(\boldsymbol{X}\boldsymbol{Z}^{\mathrm{T}}) = 0$, 则有

$$L(Y|\boldsymbol{X}, \boldsymbol{Z}) = L(Y|\boldsymbol{X}) + L(Y|\boldsymbol{Z}). \tag{5.6}$$

其中 $L(Y|\boldsymbol{X}, \boldsymbol{Z})$ 是用 $X_1, X_2, \cdots, X_n, Z_1, Z_2, \cdots, Z_m$ 预测 Y 时的最佳线性预测.

性质 7 设 $\widetilde{Y} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}$ 是 $\boldsymbol{X}$ 的线性组合, 则 $\widetilde{Y} = L(Y|\boldsymbol{X})$ 的充分必要条件是

$$\mathrm{E}\big(X_j(Y - \widetilde{Y})\big) = 0, \quad 1 \leq j \leq n. \tag{5.7}$$

即 $\mathrm{E}\big(\boldsymbol{X}(Y - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X})\big) = 0$.

性质 8 如果 $\widehat{Y}=L(Y|X_1,X_2,\cdots,X_n)$, $\widetilde{Y}=L(Y|X_1,X_2,\cdots,X_{n-1})$, 则有

$$L(\widehat{Y}|X_1,X_2,\cdots,X_{n-1})=\widetilde{Y}, \tag{5.8}$$

并且有 $\mathrm{E}(Y-\widehat{Y})^2\le\mathrm{E}(Y-\widetilde{Y})^2$.

性质 9 如果 $\mathrm{E}Y=b,\mathrm{E}\boldsymbol{X}=\boldsymbol{\mu}$, 则对任何 $c_0\in\mathbb{R},\boldsymbol{c}\in\mathbb{R}^n$, 有

$$\mathrm{E}[Y-L(Y|\boldsymbol{X})]^2\le\mathrm{E}[Y-(c_0+\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X})]^2. \tag{5.9}$$

性质 10 设 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 分别是 m 和 n 维向量, 如果有实矩阵 A,B, 使得 $\boldsymbol{X}=A\boldsymbol{Y},\boldsymbol{Y}=B\boldsymbol{X}$, 则

$$L(Z|\boldsymbol{X})=L(Z|\boldsymbol{Y}) \tag{5.10}$$

(ii) Hilbert 空间中的投影

最佳线性预测实际上是 Hilbert 空间中的投影. 用 L^2 表示全体二阶矩有限的随机变量构成的 Hilbert 空间, 设 H 是 L^2 的闭子空间, $Y\in L^2$, 可以证明在 H 中存在惟一的 $\widehat{Y}$ 使得

$$\mathrm{E}(Y-\widehat{Y})^2=\inf_{\xi\in H}\mathrm{E}(Y-\xi)^2. \tag{5.11}$$

定义: 如果 H 是 L^2 的闭子空间, $Y\in L^2$, $\widehat{Y}\in H$ 使得 (5.11) 式成立, 就称 $\widehat{Y}$ 是 Y 在 H 上的投影, 记作 $P_H(Y)$, 并且称 P_H 是投影算子.

设 $Y\in L^2$, 如果对 H 中的任何 ξ, $\mathrm{E}(Y\xi)=0$, 就称 Y 垂直于 H.

定理 设 $Y\in L^2,\widehat{Y}\in H$, 则 $\widehat{Y}=P_H(Y)$ 的充分必要条件是 $(Y-\widehat{Y})$ 垂直于 H.

用 $L^2(\boldsymbol{X})$ 表示 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的元素和常数 1 生成的 Hilbert 空间. 设 $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^{\mathrm{T}}=\mathrm{E}\boldsymbol{X}$. 对任何二阶矩有限的随机变量 Y, 设 $\mathrm{E}Y=b,\widehat{Y}=L(Y|\boldsymbol{X})$ 由 (5.2) 式定义, 则有

$$Y-\widehat{Y}=(Y-b)-L(Y-b|\boldsymbol{X}-\boldsymbol{\mu}).$$

利用性质 7 可得 $Y-\widehat{Y}$ 垂直于 $H\xlongequal{\text{def}}L^2(\boldsymbol{X})$. 由定理可知 $L(Y|\boldsymbol{X})=P_H(Y)$.

基于上述分析, 当 H 是 $\{X_j:j\in T\}$ 和常数 1 生成的 Hilbert 空间时, 也用

$$L(Y|1,X_j,j\in T)\text{ 或 }L(Y|H)$$

表示 $P_H(Y)$, 其中 T 是一个可列的指标集.

对于 L^2 的闭子空间 H, 其投影算子 P_H 具有如下性质.

定理 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) $\mathrm{E}Y^2=\mathrm{E}[L(Y|H)]^2+\mathrm{E}[Y-L(Y|H)]^2$;

(3) $\mathrm{E}[L(Y|H)]^2\le\mathrm{E}Y^2$;

(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 $P_HP_M=P_MP_H=P_H$, 并且对 $Y\in L^2$,

$$E[Y-L(Y|M)]^2\le E[Y-L(Y|H)]^2.$$

(iii) 最佳预测

最佳线性预测只是考虑用 $\boldsymbol{X}$ 和常数 1 的线性组合对 Y 进行预测, 并没有考虑用 $\boldsymbol{X}$ 的任意可测函数

$$g(\boldsymbol{X}) : \mathrm{E}g^2(\boldsymbol{X}) < \infty \tag{5.12}$$

对 Y 进行预测. 明显, 考虑在所有形如 (5.12) 的非线性函数中寻找 Y 的最佳预测是有意义的, 并且会得到更好的预测结果.

用 M 表示全体形如 (5.12) 的随机变量生成的 Hilbert 空间:

$$M = \overline{\mathrm{sp}}\{g(\boldsymbol{X}) : \mathrm{E}g^2(\boldsymbol{X}) < \infty,\ g(\cdot) \text{ 是可测函数}\}. \tag{5.13}$$

用 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 对 Y 进行预测时, 称 $L(Y|M) = P_M(Y)$ 为 Y 的最佳预测.

由于 $L^2(\boldsymbol{X})$ 是 M 的子空间, 则在方差最小的意义下, 最佳预测确实比最佳线性预测好. 但由于 M 比 $L^2(\boldsymbol{X})$ 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. 但是对于正态序列来讲, 最佳预测和最佳线性预测是一致的.

定理 如果 $(X_1, X_2, \cdots, X_n, Y)^{\mathrm{T}}$ 服从联合正态分布 $N(\boldsymbol{\mu}, \Sigma)$, M 由 (5.13) 定义, 则

$$L(Y|M) = L(Y|X_1, X_2, \cdots, X_n).$$

5.2 非决定性平稳序列及其 Wold 表示

对平稳序列, 考虑用所有的历史 $\{X_t, t \le n\}$ 对 X_{n+1} 进行线性预测. 当预测误差为零时, X_{n+1} 的信息完全含在历史资料中, 这样的平稳序列称为决定性的, 否则称为非决定性的.

(i) 非决定性平稳序列

设 $\{X_n : n \in \mathbb{Z}\}$ 是零均值平稳序列, 记 $\boldsymbol{X}_{n,m} = (X_n, X_{n-1}, \cdots, X_{n+1-m})^{\mathrm{T}}$, 这里 n 表示向量的第一个脚标, m 表示向量的维数. 定义

$$\widehat{X}_{n+1,m} = L(X_{n+1}|\boldsymbol{X}_{n,m}).$$

由性质 8 可知, $\sigma_{1,m}^2 = \mathrm{E}(X_{n+1} - \widehat{X}_{n+1,m})^2$ 是 m 的单调递减函数, 所以定义 $\sigma_1^2 = \lim\limits_{m\to\infty} \sigma_{1,m}^2 < \infty$.

定理 $\sigma_1^2 = \lim\limits_{m\to\infty} \sigma_{1,m}^2$ 与 n 无关.

证明: 设 $\boldsymbol{a} = (a_1, a_2, \cdots, a_m)^{\mathrm{T}}$ 是预测方程 (5.3) 的解, 则 $\boldsymbol{a}$ 与 n 无关. 另一方面, 由于

$$Y_n \xlongequal{\text{def}} X_{n+1} - \sum_{j=1}^{m} a_j X_{n+1-j} = X_{n+1} - \widehat{X}_{n+1,m}, \quad n \in \mathbb{Z}$$

是平稳序列, 所以 $\sigma_{1,m}^2 = \mathrm{E}Y_n^2 = \mathrm{E}Y_0^2$ 与 n 无关. 最后, $\sigma_1^2 = \lim\limits_{m\to\infty} \sigma_{1,m}^2$ 与 n 无关.

这个定理告诉我们, 当 $m \to \infty$ 时, $\sigma_{1,m}^2 \to 0$ 表明 X_{n+1} 可以由所有历史 $X_n, X_{n-1}, \cdots$ 进行完全预测; 当 $\sigma_1^2 > 0$ 时, 表明 X_{n+1} 不可以由所有历史 $X_n, X_{n-1}, \cdots$ 进行完全预测.

进一步给出如下定义: 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, 并称 $\sigma_1^2 = \lim\limits_{m\to\infty} \sigma_{1,m}^2$ 为 $\{X_t\}$ 的一步预测的均方误差.

如果 $\{X_t\}$ 不是零均值的, 有 $\mathrm{E}X_t=\mu$, 引入 $\{Y_t\}=\{X_t-\mu\}$ 和 m 维向量 $\boldsymbol{\mu}_m=(\mu,\cdots,\mu)^{\mathrm{T}}$. 于是

$$\widehat{X}_{n+1,m}=\mu+L(X_{n+1}-\mu|\boldsymbol{X}_{n,m}-\boldsymbol{\mu}_m)=\mu+\widehat{Y}_{n+1,m},$$

所以 $\mathrm{E}(Y_{n+1}-\widehat{Y}_{n+1,m})^2=\mathrm{E}(X_{n+1}-\widehat{X}_{n+1,m})^2$. 因此, 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.

常见的决定性平稳序列:

(1) $\{X_t=\xi\}$, ξ 为随机变量;

(2) $n+1$ 阶自协方差矩阵退化的 $\{X_t\}$;

(3) 离散谱序列 $Z_j(t)=\xi_j\cos(t\lambda_j)+\eta_j\sin(t\lambda_j),\ t\in\mathbb{Z}$.

(ii) 纯非决定性平稳序列

与前面类似地, 可以定义用 $\boldsymbol{X}_{n,m}$ 预测 X_{n+k} 时的均方误差:

$$\sigma_{k,m}^2=\mathrm{E}\big[X_{n+k}-L(X_{n+k}|X_n,X_{n-1},\cdots,X_{n-m+1})\big]^2.$$

这个 $\sigma_{k,m}^2$ 也是 m 的单调递减函数, 并且也与 n 无关. 于是 $\sigma_k^2=\lim\limits_{m\to\infty}\sigma_{k,m}^2$ 与 n 无关.

更进一步, 由性质 8 可知

$$\begin{aligned}\sigma_k^2&=\lim_{m\to\infty}\mathrm{E}\big[X_{n+k}-L(X_{n+k}|X_n,X_{n-1},\cdots,X_{n-m})\big]^2\\&=\lim_{m\to\infty}\mathrm{E}\big[X_{n+k-1}-L(X_{n+k-1}|X_{n-1},\cdots,X_{n-m-1})\big]^2\\&\le\lim_{m\to\infty}\mathrm{E}\big[X_{n+k-1}-L(X_{n+k-1}|X_n,X_{n-1},\cdots,X_{n-m-1})\big]^2=\sigma_{k-1}^2,\end{aligned}$$

所以 σ_k^2 是 k 的单调不减函数, 且 $\sigma_k^2\le\mathrm{E}(X_{n+k}-0)^2=\gamma_0$. 所以当 $k\to\infty$ 时, 如果 $\sigma_k^2\to\gamma_0$, 说明充分多的历史对遥远的将来进行预测和用 0 进行预测的效果差不多. 于是引入下面的定义:

设 $\{X_t\}$ 是非决定性平稳序列, 如果 $\lim\limits_{k\to\infty}\sigma_k^2=\gamma_0$, 则称 $\{X_t\}$ 是纯非决定性的.

对于纯非决定性平稳序列, 做长期或超长期预测不太合适, 即

$$\lim_{k\to\infty}\lim_{m\to\infty}\mathrm{E}\big[L(X_{n+k}|X_n,X_{n-1},\cdots,X_{n-m+1})\big]^2=0.$$

(iii) Wold 表示定理及其证明

对于零均值平稳序列 $\{X_t\}$, 用

$$H_n\xlongequal{\text{def}}\overline{\mathrm{sp}}\{X_n,X_{n-1},X_{n-2},\cdots\}$$

表示由 $\{X_n,X_{n-1},X_{n-2},\cdots\}$ 生成的闭子空间. X_{n+1} 在 H_n 中的投影

$$\widehat{X}_{n+1}=L(X_{n+1}|H_n)=L(X_{n+1}|X_{n-j},j=0,1,\cdots)$$

是用全部历史资料对 X_{n+1} 的最佳线性预测.

引理 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, 则对任何 $\xi\in H_A$, 必存在 $\xi_n\in L_A,\ n=1,2,\cdots$ 使得

$$\|\xi_n-\xi\|\to0,\quad n\to\infty.$$

特别地, 当 $A=\{X_n,X_{n-1},\cdots\}$ 时, 则有 $H_A=H_n$ 为 A 所张成的子 Hilbert 空间.

定理 设 $Y \in L^2, \xi \in H_n$, 则 $\xi = L(Y|H_n)$ 的充分必要条件是

$$\mathrm{E}\big[X_j(Y-\xi)\big] = 0 \;\Leftrightarrow\; Y - \xi \perp X_j, \quad j = n, n-1, n-2, \cdots.$$

下面定理说明对 $Y \in L^2$, $L(Y|H_n)$ 是有穷维最佳线性预测的极限.

定理 设 $\boldsymbol{X}_{n,m} = (X_n, X_{n-1}, \cdots, X_{n-m+1})^{\mathrm{T}}$, 当 $m \to \infty$ 时,

$$L(Y|\boldsymbol{X}_{n,m}) \xrightarrow{\text{m.s.}} \widehat{Y} = L(Y|H_n). \tag{5.14}$$

取 $Y = X_{n+1}$, 则 $\widehat{X}_{n+1} = L(X_{n+1}|H_n)$ 是 $\widehat{X}_{n+1,m} = L(X_{n+1}|\boldsymbol{X}_{n,m})$ 的均方极限. 利用内积的连续性可知

$$\sigma_1^2 = \lim_{m\to\infty} \mathrm{E}(X_{n+1} - \widehat{X}_{n+1,m})^2 = \mathrm{E}\big[X_{n+1} - L(X_{n+1}|H_n)\big]^2 = \mathrm{E}\big[X_1 - L(X_1|H_0)\big]^2.$$

所以 $\sigma_1^2 = 0 \Leftrightarrow X_1 = L(X_1|H_0) \Leftrightarrow X_1 \in H_0$.

类似地, 取 $Y = X_{n+k}$, 可以得到

$$\sigma_k^2 = \lim_{m\to\infty} \mathrm{E}(X_{n+k} - \widehat{X}_{n+k,m})^2 = \mathrm{E}\big[X_{n+k} - L(X_{n+k}|H_n)\big]^2 = \mathrm{E}\big[X_k - L(X_k|H_0)\big]^2.$$

于是得到下面的定理.

定理 设 $\{X_t\}$ 是零均值平稳序列.

(1) $\{X_t\}$ 是决定性的当且仅当对某个 n, 有

$$X_{n+1} \in H_n. \tag{5.15}$$

如果 (5.15) 对某个 n 成立, 则对一切 n 成立. 此时 $H_n = H_{n-1}$ 对一切 n 成立.

(2) $\{X_t\}$ 是纯非决定性的当且仅当对某个 n, 有

$$\sigma_k^2 \xlongequal{\text{def}} \mathrm{E}\big[X_{n+k} - L(X_{n+k}|H_n)\big]^2 \to \gamma_0, \quad k \to \infty. \tag{5.16}$$

如果 (5.16) 对某个 n 成立, 则对一切 n 成立.

定理 (Wold 表示定理) 任一非决定性的零均值平稳序列 $\{X_t\}$ 可以表示成

$$X_t = U_t + V_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} + V_t, \quad t \in \mathbb{Z}, \tag{5.17}$$

其中 (1) $\varepsilon_t = X_t - L(X_t|X_{t-1}, X_{t-2}, \cdots)$ 是零均值白噪声, 满足

$$\mathrm{E}\varepsilon_t^2 = \sigma^2 > 0, \quad a_0 = 1, \quad a_j = \frac{\mathrm{E}(X_t \varepsilon_{t-j})}{\sigma^2}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty;$$

(2) $\{U_t\}$ 是和 $\{V_t\}$ 正交的平稳序列;

(3) 定义 $H_\varepsilon(t) = \overline{\mathrm{sp}}\{\varepsilon_j : j \le t\}, H_U(t) = \overline{\mathrm{sp}}\{U_j : j \le t\}$. 对任何 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 $f(\lambda) = \dfrac{\sigma^2}{2\pi}\Big|\displaystyle\sum_{j=0}^{\infty} a_j e^{\mathrm{i}j\lambda}\Big|^2$;

(5) $\{V_t\}$ 是决定性的平稳序列. 对任何 $t, k \in \mathbb{Z}$, $V_t \in H_{t-k}$.

证明: 取 $\varepsilon_t = X_t - L(X_t|H_{t-1}) \in H_t$, 且由 H_t 的单调性可知 $\varepsilon_t \in H_s, \forall t < s$, 因此 $H_\varepsilon(t) \subset H_t$.

(1) 下面证明 $\{\varepsilon_t\}$ 是白噪声. 由定理可知

$$L(X_t|H_{t-1}) = \lim_{m\to\infty} L(X_t|X_{t-1}, \cdots, X_{t-m}) = \lim_{m\to\infty} \boldsymbol{a}_m^{\mathrm{T}} \boldsymbol{X}_{t-1,m},$$

其中 $\boldsymbol{a}_m$ 是 m 阶 Yule-Walker 系数 (预测方程的解), 不依赖于 t.
由内积的连续性, $\mathrm{E}\varepsilon_t = \mathrm{E}X_t - \lim_{m\to\infty} \mathrm{E}\boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{X}_{t-1,m} = 0$, 及

$$\begin{aligned}\mathrm{E}\varepsilon_t^2 &= \lim_{m\to\infty} \mathrm{E}\big[X_t - L(X_t|X_{t-1}, \cdots, X_{t-m})\big]^2 \\ &= \lim_{m\to\infty} (\gamma_0 - \boldsymbol{a}_m^{\mathrm{T}}\Gamma_m \boldsymbol{a}_m) = \lim_{m\to\infty} \sigma_{1,m}^2 = \sigma^2 > 0\end{aligned}$$

对 $s > t$, 有 $\varepsilon_s \perp H_{s-1} \supset H_t \supset H_\varepsilon(t)$, 所以 $\varepsilon_s \perp \varepsilon_t$. 于是 $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$.

(2) 下面证明 $\{\varepsilon_t\}$ 和 $\{V_t\}$ 正交. 定义

$$V_t = X_t - L(X_t|H_\varepsilon(t)) \in H_\varepsilon(t) \subset H_t,$$

所以 $V_t \perp \varepsilon_s$, $\forall s \le t$. 当 $s > t$ 时, $\varepsilon_s \perp H_{s-1} \supset H_t$, $V_t \in H_t$, 所以 $\varepsilon_s \perp V_t$, $\forall s > t$. 即 $\{\varepsilon_t\}$ 和 $\{V_t\}$ 正交.

(3) 下面证明 $\{\varepsilon_t\}$ 和 $\{X_t\}$ 平稳相关. 当 $s > t$ 时, $\varepsilon_s \perp H_t$, 所以 $\mathrm{E}(\varepsilon_s X_t) = 0$; 当 $s \le t$ 时,

$$L(X_t|H_{t-1}) = \lim_{m\to\infty} L(X_t|\boldsymbol{X}_{t-1,m}) = \lim_{m\to\infty} \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{X}_{t-1,m}$$

由内积的连续性

$$\mathrm{E}(\varepsilon_s X_t) = \lim_{m\to\infty} \mathrm{E}\big[X_t(X_s - \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{X}_{t-1,m})\big] = \gamma_{t-s} - \lim_{m\to\infty}(a_{m1}\gamma_{t-s+1} + \cdots + a_{mm}\gamma_{t-s+m})$$

只依赖于 $t - s$. 于是 $\{\varepsilon_t\}$ 和 $\{X_t\}$ 平稳相关. 所以可以定义

$$a_j = \frac{\mathrm{E}(X_t\varepsilon_{t-j})}{\sigma^2} \text{ 与 } t \text{ 无关, 且 } a_0 = \frac{\mathrm{E}(X_t\varepsilon_t)}{\sigma^2} = \frac{\mathrm{E}\varepsilon_t^2}{\sigma^2} = 1.$$

(4) 下面证明 U_t 是相应系数 a_j 的单边滑动和. 令 $U_t = L(X_t|H_\varepsilon(t))$, 定义

$$U_{t,n} = L(X_t|\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-n}) \xlongequal{\text{def}} \sum_{j=0}^n b_j\varepsilon_{t-j},$$

则对 $j = 0, 1, \cdots, n$, 有

$$\sigma^2 a_j = \mathrm{E}(X_t\varepsilon_{t-j}) = \mathrm{E}\big[(U_{t,n} + X_t - U_{t,n})\varepsilon_{t-j}\big] = \mathrm{E}(U_{t,n}\varepsilon_{t-j}) = \sigma^2 b_j,$$

所以 $b_j = a_j$, 又注意到 $\mathrm{E}U_{t,n}^2 \le \mathrm{E}X_t^2 = \gamma_0$, 所以

$$\mathrm{E}U_{t,n}^2 = \sigma^2\sum_{j=0}^n a_j^2 \le \gamma_0 < \infty.$$

因此 $\{a_j\}$ 平方可和, 且有 $U_{t,n}$ 均方收敛到 U_t, 即

$$U_t = \sum_{j=0}^\infty a_j\varepsilon_{t-j}.$$

(5) 下面证明 $\{V_t\}$ 和 $\{U_t\}$ 正交.

由于 $\{\varepsilon_t\}$ 和 $\{V_t\}$ 正交, 所以 $V_s \perp H_\varepsilon(t)$, 而 $U_t \in H_\varepsilon(t)$, 所以 $V_s \perp U_t$, 即 $\{V_t\}$ 和 $\{U_t\}$ 正交. 进一步结合 (4) 的证明结果可知, $\{V_t\}$ 和 $\{U_t\}$ 都是平稳列.

(6) 下面证明 $H_U(t)=H_\varepsilon(t)$. 显然 $U_t\in H_\varepsilon(t)\Rightarrow H_U(t)\subset H_\varepsilon(t)$. 注意到 $\varepsilon_t\in H_t\subset\overline{\mathrm{sp}}\{U_s,V_s:s\le t\}$, 因此存在 $\xi_m\in L_{\{V_t,V_{t-1},\cdots,V_{t-m}\}},\eta_m\in L_{\{U_t,U_{t-1},\cdots,U_{t-m}\}}$, 使得

$$\|\xi_m+\eta_m-\varepsilon_t\|^2\to 0$$

结合正交关系可知

$$\|\xi_m+\eta_m-\varepsilon_t\|^2=\|\xi_m\|^2+\|\eta_m-\varepsilon_t\|^2\ge\|\eta_m-\varepsilon_t\|^2,$$

所以 $\varepsilon_t\in H_U(t)\Rightarrow H_\varepsilon(t)\subset H_U(t)$, 进一步有 $H_U(t)=H_\varepsilon(t)$.

(7) 下面证明 $\{U_t\}$ 是纯非决定性的平稳序列. 注意到

$$L(U_{t+k}|H_U(t))=L(U_{t+k}|H_\varepsilon(t))=\sum_{j=k}^{\infty}a_j\varepsilon_{t+k-j},$$

于是

$$\mathrm{E}\big[U_{t+k}-L(U_{t+k}|H_U(t))\big]^2=\mathrm{E}\Big[\sum_{j=0}^{k-1}a_j\varepsilon_{t+k-j}\Big]^2=\sigma^2\sum_{j=0}^{k-1}a_j^2\to\sigma^2\sum_{j=0}^{\infty}a_j^2=\mathrm{E}U_t^2.$$

所以 $\{U_t\}$ 是纯非决定性的.

(8) 最后证明 $\{V_t\}$ 是决定性的平稳序列. 注意 $\varepsilon_{t-j}\in H_{t-j}$, 且

$$V_t=X_t-U_t=X_t-\varepsilon_t-\sum_{j=1}^{\infty}a_j\varepsilon_{t-j}=L(X_t|H_{t-1})-\sum_{j=1}^{\infty}a_j\varepsilon_{t-j}\in H_{t-1},$$

而 $H_{t-1}\subset\overline{\mathrm{sp}}\{U_s,V_s:s\le t-1\}$, 因此存在 $\xi_m\in L_{\{V_{t-1},\cdots,V_{t-m}\}},\eta_m\in L_{\{U_{t-1},\cdots,U_{t-m}\}}$, 使得

$$\|\xi_m+\eta_m-V_t\|^2\to 0$$

结合正交关系可知

$$\|\xi_m+\eta_m-V_t\|^2=\|\xi_m-V_t\|^2+\|\eta_m\|^2\ge\|\xi_m-V_t\|^2,$$

所以 $V_t\in H_V(t-1)$, 即 $\{V_t\}$ 为决定性的, 且 $H_V(t)=H_V(t-k),k\in\mathbb{Z}$, 所以

$$V_t\in H_V(t-k)\subset H_{t-k},t,k\in\mathbb{Z}.$$

在上述 Wold 表示定理中, 称 (5.17) 是 $\{X_t\}$ 的 Wold 表示, $\{U_t\}$ 是 $\{X_t\}$ 的纯非决定性部分, $\{V_t\}$ 是 $\{X_t\}$ 的决定性部分; 称 $\{a_j\}$ 是 $\{X_t\}$ 的 Wold 系数; 称一步预测误差

$$\varepsilon_t=X_t-L(X_t|X_{t-1},X_{t-2},\cdots)$$

是 $\{X_t\}$ 的 (线性) 新息序列, $\sigma^2=\mathrm{E}\varepsilon_t^2$ 为一步预测的均方误差.

Wold 表示定理告诉我们, 任何纯非决定性平稳序列可以表达为新息的单边滑动和. 我们还可以证明, 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平稳序列.

注: “新息”是新信息的简称, 意思是完全不能被历史线性预测, $\varepsilon_t\notin H_{t-1}$.

例: ARMA(p,q) 序列 $\{X_t\}$ 是纯非决定性的平稳序列, 其 Wold 表示为

$$X_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j},\quad t\in\mathbb{Z},\tag{5.18}$$

其中 $\{\varepsilon_t\}$ 是 $\{X_t\}$ 的新息序列, $\{\psi_j\}$ 是 $\{X_t\}$ 的 Wold 系数.

(iv) 多步预测的误差

设 $\{X_t\}$ 是非决定性的平稳序列, 在 Wold 表示 (5.17) 中, 对任何 $n>0$, $V_{t+n}\in H_t$, 于是

$$L(X_{t+n}|H_t)=L(U_{t+n}|H_t)+L(V_{t+n}|H_t)=\sum_{j=n}^{\infty}a_j\varepsilon_{t+n-j}+V_{t+n}.$$

称 $L(X_{t+n}|H_n)$ 是 X_{t+n} 的 n 步预报. 其预报误差为

$$X_{t+n}-L(X_{t+n}|H_t)=\sum_{j=0}^{n-1}a_j\varepsilon_{t+n-j},$$

预报的均方误差是

$$\sigma^2(n)=\sigma^2\sum_{j=0}^{n-1}a_j^2\to \mathrm{E}U_t^2,\ n\to\infty.$$

(v) 最佳预测和最佳线性预测相等的条件

设 $\{X_t\}$ 是平稳序列, 用 $\mathscr{F}_t=\sigma\{X_t,X_{t-1},\cdots\}$ 表示由 $X_t,X_{t-1},\cdots$ 生成的 σ-代数. 称条件数学期望

$$\mathrm{E}(X_{t+k}|\mathscr{F}_t)$$

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. 可以证明, 在方差最小的意义下, 条件数学期望 (最佳预测) 是最好的, 还是惟一的. 对任意方差有限的随机变量 $\xi=g(X_t,X_{t-1},\cdots)$, 利用条件数学期望的性质可得

$$\begin{aligned}\mathrm{E}(X_{t+k}-\xi)^2&=\mathrm{E}\big[X_{t+k}-\mathrm{E}(X_{t+k}|\mathscr{F}_t)+\mathrm{E}(X_{t+k}|\mathscr{F}_t)-\xi\big]^2\\&=\mathrm{E}\big[X_{t+k}-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big]^2+\mathrm{E}\big[\xi-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big]^2+\\&\qquad 2\mathrm{E}\big[\big(X_{t+k}-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big)\big(\xi-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big)\big]\\&=\mathrm{E}\big[X_{t+k}-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big]^2+\mathrm{E}\big[\xi-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big]^2\\&\geq\mathrm{E}\big[X_{t+k}-\mathrm{E}(X_{t+k}|\mathscr{F}_t)\big]^2.\end{aligned}$$

取等号时当且仅当 $\xi=\mathrm{E}(X_{t+k}|\mathscr{F}_t)$, a.s..

下面定理告诉我们, 对纯非决定性平稳序列来讲, 如果它的新息序列是独立序列, 则最佳预测和最佳线性预测等价.

定理 设平稳序列 $\{X_t\}$ 有 Wold 表示 $X_t=\sum_{j=0}^{\infty}a_j\varepsilon_{t-j},\ t\in\mathbb{Z}$, 则

$$L(X_{t+n}|H_t)=\mathrm{E}(X_{t+n}|\mathscr{F}_t),\quad n\geq 1,\quad t\in\mathbb{Z}\tag{5.19}$$

成立的充分必要条件是

$$\mathrm{E}(\varepsilon_{t+1}|\varepsilon_t,\varepsilon_{t-1},\cdots)=0,\quad t\in\mathbb{Z}.\tag{5.20}$$

注: 满足 (5.20) 的白噪声通常被称为鞅差白噪声.

推论 设 ARMA(p,q) 序列 $\{X_t\}$ 由 (5.18) 定义, 如果 $\{\varepsilon_t\}$ 是独立白噪声, 则用全体历史 $\{X_t,X_{t-1},\cdots\}$ 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

(i) 时间序列的递推预测

设 $\{Y_t\}$ 是方差有限的零均值时间序列, 对任何正整数 n, 用 $L_n = \overline{\mathrm{sp}}\{Y_1, Y_2, \cdots, Y_n\}$ 表示 $Y_1, Y_2, \cdots, Y_n$ 的线性组合的全体. 定义 $\boldsymbol{Y}_n = (Y_1, Y_2, \cdots, Y_n)^{\mathrm{T}}$,

$$\widehat{Y}_1 = 0, \quad \widehat{Y}_n = L(Y_n|\boldsymbol{Y}_{n-1}), \quad n = 1, 2, \cdots.$$

引入预测误差 W_n 及其方差 ν_{n-1} 如下:

$$W_n = Y_n - \widehat{Y}_n, \quad \nu_{n-1} = \mathrm{E}W_n^2, \quad n = 1, 2, \cdots.$$

可以证明 $\{W_n\}$ 是一个正交序列, 用 $M_n = \overline{\mathrm{sp}}\{W_1, W_2, \cdots, W_n\}$ 表示 $W_1, W_2, \cdots, W_n$ 的线性组合的全体, 可以进一步证明 $M_n = L_n$. 因此用 $\boldsymbol{W}_n = (W_1, W_2, \cdots, W_n)^{\mathrm{T}}$ 对 Y_{n+1} 进行预测和用 $\boldsymbol{Y}_n = (Y_1, Y_2, \cdots, Y_n)^{\mathrm{T}}$ 对 Y_{n+1} 进行预测是等价的.

定理 设 $\{Y_t\}$ 是零均值时间序列. 如果 $(Y_1, Y_2, \cdots, Y_{m+1})^{\mathrm{T}}$ 的协方差矩阵 $\big(\mathrm{E}(Y_sY_t)\big)_{1\le s,t\le m+1}$ 正定, 则最佳线性预测

$$\widehat{Y}_{n+1} \equiv L(Y_{n+1}|\boldsymbol{Y}_n) = \sum_{j=1}^{n} \theta_{n,j} W_{n+1-j}, \quad n = 1, 2, \cdots, m$$

中的系数 $\{\theta_{n,j}\}$ 和预测的均方误差 $\nu_n = \mathrm{E}W_{n+1}^2$ 满足如下的递推公式:

$$\begin{cases} \nu_0 = \mathrm{E}Y_1^2, \\ \theta_{n,n-k} = \left[\mathrm{E}(Y_{n+1}Y_{k+1}) - \displaystyle\sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j\right] \Big/ \nu_k, \quad 0 \le k \le n-1, \\ \nu_n = \mathrm{E}Y_{n+1}^2 - \displaystyle\sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \end{cases} \tag{5.21}$$

其中 $\displaystyle\sum_{j=0}^{-1}(\cdot) \xlongequal{\mathrm{def}} 0$, 递推的顺序是 $\nu_0; \theta_{1,1}, \nu_1; \theta_{2,2}, \theta_{2,1}, \nu_2; \cdots$.

(ii) 多步预报的递推公式

下面考虑用 $\{Y_1, Y_2, \cdots, Y_n\}$ 预测 Y_{n+k+1} 的问题. 设 $(Y_1, Y_2, \cdots, Y_{n+k+1})^{\mathrm{T}}$ 的自协方差矩阵正定, 记 $\widehat{Y}_{n+k+1} = L(Y_{n+k+1}|\boldsymbol{Y}_{n+k})$, $W_j = Y_j - L(Y_j|\boldsymbol{Y}_{j-1})$, 则有

$$\widehat{Y}_{n+k+1} = \sum_{j=1}^{n+k} \theta_{n+k,j} W_{n+k+1-j}.$$

注意到, 对 $j \ge 0$, W_{n+j+1} 垂直于 L_n, $W_{n-j} \in L_n$, 则

$$L(Y_{n+k+1}|\boldsymbol{Y}_n) = L(\widehat{Y}_{n+k+1}|\boldsymbol{Y}_n) = \sum_{j=k+1}^{n+k} \theta_{n+k,j} W_{n+k+1-j},$$

预测的均方误差为

$$\mathrm{E}\big[Y_{n+k+1} - L(Y_{n+k+1}|\boldsymbol{Y}_n)\big]^2 = \mathrm{E}Y_{n+k+1}^2 - \sum_{j=k+1}^{n+k} \theta_{n+k,j}^2 \nu_{n+k-j},$$

其中系数 $\theta_{n+k,j}$ 和预测的均方误差 ν_{n+k-j} 可用递推公式 (5.21) 计算.

(iii) 正态时间序列的区间预测

如果 $\{Y_t\}$ 是正态时间序列, 则 $\widehat{Y}_{n+1}$ 也是最佳预测. $W_{n+1}=Y_{n+1}-\widehat{Y}_{n+1}$ 作为 $Y_1,Y_2,\cdots,Y_{n+1}$ 的线性组合是服从正态分布 $N(0,\nu_n)$ 的. 于是可以得到 Y_{n+1} 的置信度为 $1-\alpha$ 的置信区间

$$\left[\widehat{Y}_{n+1}-u_{\alpha/2}\sqrt{\nu_n},\widehat{Y}_{n+1}+u_{\alpha/2}\sqrt{\nu_n}\right].$$

(iv) 平稳序列的递推预测

设 $\gamma_k=\mathrm{E}(X_{t+k}X_t)$ 是零均值平稳序列 $\{X_t\}$ 的自协方差函数, Γ_n 是 $\{X_t\}$ 的 n 阶自协方差矩阵. 设

$$\boldsymbol{X}_n=(X_1,X_2,\cdots,X_n)^{\mathrm{T}},\quad Z_n=X_n-L(X_n|\boldsymbol{X}_{n-1}),$$

可将 (i) 中递推预测的定理改述如下.

推论 设 $\{X_t\}$ 是零均值平稳序列, 对任何 $n\in\mathbb{N}_+$, 自协方差矩阵 Γ_n 正定, 则最佳线性预测

$$\widehat{X}_{n+1}\xlongequal{\text{def}}L(X_{n+1}|\boldsymbol{X}_n)=\sum_{j=1}^{n}\theta_{n,j}Z_{n+1-j},\quad n=1,2,\cdots,$$

这里的系数 $\{\theta_{n,j}\}$ 和预测的均方误差 $\nu_n=\mathrm{E}Z_{n+1}^2$ 满足如下的递推公式:

$$\begin{cases}\nu_0=\gamma_0,\\ \theta_{n,n-k}=\left[\gamma_{n-k}-\displaystyle\sum_{j=0}^{k-1}\theta_{k,k-j}\theta_{n,n-j}\nu_j\right]\Big/\nu_k,\quad 0\le k\le n-1,\\ \nu_n=\gamma_0-\displaystyle\sum_{j=0}^{n-1}\theta_{n,n-j}^2\nu_j,\end{cases}\tag{5.22}$$

其中 $\displaystyle\sum_{j=0}^{-1}(\cdot)\xlongequal{\text{def}}0$, 递推的顺序是 $\nu_0;\theta_{1,1},\nu_1;\theta_{2,2},\theta_{2,1},\nu_2;\cdots$.

由于预测误差 $Z_n=X_n-L(X_n|\boldsymbol{X}_{n-1})$ 和 $\boldsymbol{X}_{n-1}$ 正交, 所以是不被 $\boldsymbol{X}_{n-1}$ 包含的信息. 基于这个原因, 人们又称 Z_n 是样本新息. 且当 $n\to\infty$ 时,

$$\nu_n=\mathrm{E}\left[X_1-L(X_1|X_0,X_{-1},\cdots,X_{-n+1})\right]^2\to\sigma^2.$$

这里 σ^2 是用全体历史 $X_t,X_{t-1},\cdots$ 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

(i) AR(p) 序列的预测

设 $\{X_t\}$ 满足 AR(p) 模型

$$X_t=\sum_{j=1}^{p}a_jX_{t-j}+\varepsilon_t,\quad t\in\mathbb{Z},\tag{5.23}$$

其中 $\{\varepsilon_t\}$ 是白噪声, 特征多项式

$$A(z)=1-\sum_{j=1}^{p}a_jz^j\neq0,\quad |z|\le1.$$

考虑用 $\boldsymbol{X}_n=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 预测 X_{n+1} 的问题. 设 $\{\gamma_n\}$ 是 $\{X_t\}$ 的自协方差函数.

对于 $1 \le n \le p-1$, 由最佳线性预测的性质 1 可知

$$\widehat{X}_{n+1} = L(X_{n+1}|\boldsymbol{X}_n) = \boldsymbol{\gamma}_n^{\mathrm{T}}\Gamma_n^{-1}\boldsymbol{X}_n,$$

其中 Γ_n 是 $\{X_t\}$ 的 n 阶自协方差矩阵, $\boldsymbol{\gamma}_n = (\gamma_n, \gamma_{n-1}, \cdots, \gamma_1)^{\mathrm{T}}$. 预测的均方误差为

$$\mathrm{E}(X_{n+1} - \widehat{X}_{n+1})^2 = \gamma_0 - \boldsymbol{\gamma}_n\Gamma_n^{-1}\boldsymbol{\gamma}_n.$$

对于 $n \ge p$, 由于当 $k \ge 1$ 时, $X_t = \sum\limits_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ 与 ε_{t+k} 正交, 所以

$$\widehat{X}_{n+1} = L\left(\sum_{j=1}^{p} a_j X_{n+1-j}\middle|\boldsymbol{X}_n\right) = \sum_{j=1}^{p} a_j X_{n+1-j}.$$

容易看出对 $n \ge p$, $\widehat{X}_{n+1} = L(X_{n+1}|X_n, X_{n-1}, \cdots, X_{n-p+1})$.

下面考虑用 $\boldsymbol{X}_n$ 预测 X_{n+k} 的多步预报问题. 对于 $n \ge p$, 用归纳法可以证明

$$L(X_{n+k}|\boldsymbol{X}_n) = L(X_{n+k}|X_n, X_{n-1}, \cdots, X_{n-p+1}),$$

即用 p 个数据 $X_n, X_{n-1}, \cdots, X_{n-p+1}$ 做预测和用 $\boldsymbol{X}_n$ 做预测的结果相同. 于是有

$$\begin{aligned} L(X_{n+k}|\boldsymbol{X}_n) &= L\left(\sum_{j=1}^{p} a_j X_{n+k-j} + \varepsilon_{n+k}\middle|\boldsymbol{X}_n\right) = L\left(\sum_{j=1}^{p} a_j X_{n+k-j}\middle|\boldsymbol{X}_n\right) \\ &= \sum_{j=1}^{p} a_j L(X_{n+k-j}|X_n, X_{n-1}, \cdots, X_{n-p+1}). \end{aligned}$$

(ii) MA(q) 序列的预测

设 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0, \sigma^2)$, 实系数多项式 $B(z) = 1 + \sum\limits_{j=1}^{q} b_j z^j$ 在单位圆内无根: $B(z) \ne 0, |z| < 1$. 满足 MA(q) 模型

$$X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z} \tag{5.24}$$

的 MA(q) 序列 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$ 是 q 后截尾的. 假设 $\sigma^2, b_1, b_2, \cdots, b_q$ 已知, 考虑用 $X_1, X_2, \cdots, X_n$ 预测 X_{n+k} 的问题.

对 $n \ge 1$, 有

$$L_n = \overline{\mathrm{sp}}\{X_1, X_2, \cdots, X_n\} = \overline{\mathrm{sp}}\{\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \cdots, \widehat{\varepsilon}_n\},$$

这里 $\widehat{\varepsilon}_n = X_n - L(X_n|\boldsymbol{X}_{n-1})$ 是 $\{X_t\}$ 的逐步预测误差, $\boldsymbol{X}_n = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$.

以下假定 $n \ge q$. 由于 $\{\widehat{\varepsilon}_t\}$ 是正交序列, X_{n+1} 和 $\widehat{\varepsilon}_j, 1 \le j \le n-q$ 正交. 根据最佳线性预测的性质 6、性质 4、性质 10 可得 (新息预测公式)

$$L(X_{n+1}|\boldsymbol{X}_n) = L(X_{n+1}|\widehat{\varepsilon}_n, \cdots, \widehat{\varepsilon}_1) = L(X_{n+1}|\widehat{\varepsilon}_n, \cdots, \widehat{\varepsilon}_{n-q+1}) = \sum_{j=1}^{q}\theta_{n,j}\widehat{\varepsilon}_{n+1-j}.$$

预测的均方误差为

$$\nu_n = \mathrm{E}\widehat{\varepsilon}_{n+1}^2 = \gamma_0 - \sum_{j=1}^{q}\theta_{n,j}^2\nu_{n-j},$$

这里的 $\theta_{n,j}, \nu_n$ 可以利用 (5.22) 进行递推计算.

用 $\boldsymbol{X}_n$ 预测 X_{n+k+1} 进行多步预报时, 对 $n \geq q$, 有

$$L(X_{n+k+1}|\boldsymbol{X}_n) = \begin{cases} \sum\limits_{j=k+1}^{q} \theta_{n+k,j}\widehat{\varepsilon}_{n+k+1-j}, & 1 \leq k \leq q-1, \\ 0, & k \geq q. \end{cases}$$

此时预测的均方误差为

$$\mathrm{E}\big[X_{n+k+1} - L(X_{n+k+1}|\boldsymbol{X}_n)\big]^2 = \gamma_0 - \sum_{j=k+1}^{q} \theta_{n+k,j}^2 \nu_{n+k-j}, \quad 1 \leq k \leq q-1,$$

其中的 $\theta_{n+k,j}, \nu_{n+k-j}$ 可以由 (5.22) 递推得到.

(iii) ARMA(p,q) 序列的预测

设实系数多项式 $A(z) = 1 - \sum\limits_{j=1}^{p} a_j z^j$ 满足最小相位条件, $B(z) = 1 + \sum\limits_{j=1}^{q} b_j z^j$ 在单位圆内无零点, $\{\varepsilon_t\}$ 是 WN$(0,\sigma^2)$. 对于满足 ARMA(p,q) 模型

$$A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, \quad t \in \mathbb{Z} \tag{5.25}$$

的 ARMA(p,q) 序列 $\{X_t\}$, 定义 $m = \max(p,q)$ 和

$$Y_t = \begin{cases} X_t/\sigma, & t = 1,2,\cdots,m, \\ A(\mathscr{B})X_t/\sigma, & t = m+1, m+2, \cdots, \end{cases}$$

则 $\{Y_t\}$ 由 ARMA(p,q) 模型 (5.25) 的参数 $\boldsymbol{\beta} = (a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_q)^{\mathrm{T}}$ 和标准白噪声 $\{\varepsilon_t/\sigma\}$ 决定, 从而不依赖 σ. 假设模型 (5.25) 中的参数已知, 考虑用 $X_1, X_2, \cdots, X_n$ 进行逐步预测的问题.

从 Y_t 的定义可知, 对 $t \geq 1$, $Y_t \in L_t = \overline{\mathrm{sp}}\{X_1, X_2, \cdots, X_t\}$, 并且 $X_1, X_2, \cdots, X_m \in \overline{\mathrm{sp}}\{Y_1, Y_2, \cdots, Y_m\}$. 对 $t > m$, 有

$$X_t = \sigma Y_t + \sum_{j=1}^{p} a_j X_{t-j} \in \overline{\mathrm{sp}}\{Y_1, Y_2, \cdots, Y_t\}.$$

进一步可得

$$L_n = \overline{\mathrm{sp}}\{X_1, X_2, \cdots, X_n\} = \overline{\mathrm{sp}}\{Y_1, Y_2, \cdots, Y_n\} = \overline{\mathrm{sp}}\{W_1, W_2, \cdots, W_n\},$$

其中 $W_1 = Y_1, W_t = Y_t - L(Y_t|\boldsymbol{Y}_{t-1})$ 是 $\{Y_t\}$ 的样本新息.

用 γ_k 表示 $\{X_t\}$ 的自协方差函数, 取 $b_0 = 1, b_j = 0, j > q$, 可以计算出

$$\mathrm{E}(Y_s Y_t) = \begin{cases} \sigma^{-2}\gamma_{t-s}, & 1 \leq s \leq t \leq m, \\ \sigma^{-2}\Big[\gamma_{t-s} - \sum\limits_{j=1}^{p} a_j \gamma_{t-s-j}\Big], & 1 \leq s \leq m < t, \\ \sum\limits_{j=0}^{q} b_j b_{j+t-s}, & t \geq s > m, \end{cases} \tag{5.26}$$

其中的 γ_k 可用 (3.16) 和 (3.17) 计算.

定义 $\{X_t\}$ 的逐步预测误差

$$Z_t = X_t - L(X_t|\boldsymbol{X}_{t-1}), \quad Z_1 = X_1,$$

则对 $1 \leq t \leq m$,

$$W_t = X_t/\sigma - L(X_t/\sigma|\boldsymbol{X}_t) = \sigma^{-1}[X_t - L(X_t|\boldsymbol{X}_{t-1})];$$

对 $t \geq m+1$,

$$W_t = \sigma^{-1}\Big[X_t - \sum_{j=1}^{p} a_j X_{t-j} - L\Big(X_t - \sum_{j=1}^{p} a_j X_{t-j}\Big|\boldsymbol{X}_{t-1}\Big)\Big] = \sigma^{-1}[X_t - L(X_t|\boldsymbol{X}_{t-1})].$$

所以 $\{X_t\}$ 和 $\{Y_t\}$ 的预测误差之间总有以下关系:

$$Z_t = \sigma W_t, \quad \mathrm{E}Z_t^2 = \sigma^2 \mathrm{E}W_t^2 = \sigma^2 \nu_{t-1}, \quad t = 1, 2, \cdots.$$

对于 $1 \leq n < m = \max(p, q)$, 从逐步预测公式可得

$$\widehat{Y}_{n+1} = L(Y_{n+1}|\boldsymbol{Y}_n) = \sum_{j=1}^{n} \theta_{n,j} W_{n+1-j},$$

所以

$$\widehat{X}_{n+1} = L(X_{n+1}|\boldsymbol{X}_n) = L(\sigma Y_{n+1}|\boldsymbol{Y}_n) = \sigma \widehat{Y}_{n+1} = \sum_{j=1}^{n} \theta_{n,j} Z_{n+1-j}.$$

对于 $n \geq m$, 利用 $\mathrm{E}(X_t \varepsilon_{t+k}) = 0, k = 1, 2, \cdots$, 得到

$$Y_{n+1} = \sigma^{-1} B(\mathscr{B}) \varepsilon_{n+1} = \sigma^{-1} \sum_{j=0}^{q} b_j \varepsilon_{n+1-j}$$

与 $\{X_j : 1 \leq j \leq n-q\}$ 正交, 从而与

$$\overline{\mathrm{sp}}\{Y_j : 1 \leq j \leq n-q\} = \overline{\mathrm{sp}}\{W_j : 1 \leq j \leq n-q\}$$

中的任何随机变量正交. 利用最佳线性预测的性质 6 和性质 4 可得

$$L(Y_{n+1}|\boldsymbol{Y}_n) = \sum_{j=1}^{q} \theta_{n,j} W_{n+1-j}, \quad n \geq m = \max(p, q).$$

所以

$$\widehat{X}_{n+1} = L\Big(\sum_{j=1}^{p} a_j X_{n+1-j} + \sigma Y_{n+1}\Big|\boldsymbol{X}_n\Big) = \sum_{j=1}^{p} a_j X_{n+1-j} + \sum_{j=1}^{q} \theta_{n,j} Z_{n+1-j}.$$

总结上述推导可以得到

$$\widehat{X}_{n+1} = \begin{cases} \displaystyle\sum_{j=1}^{n} \theta_{n,j} Z_{n+1-j}, & 1 \leq n < m, \\ \displaystyle\sum_{j=1}^{p} a_j X_{n+1-j} + \sum_{j=1}^{q} \theta_{n,j} Z_{n+1-j}, & n \geq m. \end{cases} \tag{5.27}$$

预测的均方误差仍然是 $\mathrm{E}Z_{n+1}^2 = \sigma^2 \nu_n$, 这里的 $\theta_{n,j}, \nu_n = \mathrm{E}W_{n+1}^2$ 可以利用 (5.21) 和 (5.26) 进行递推计算.

六、 ARMA 模型的参数估计

6.1 AR(p) 模型的参数估计

考虑 AR(p) 模型

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t, \quad t \geq 0. \tag{6.1}$$

假设自回归阶数 p 已知, 需要考虑回归系数 $\boldsymbol{a} = (a_1, a_2, \cdots, a_p)^{\mathrm{T}}$ 和零均值白噪声 $\{\varepsilon_t\}$ 的方差 $\sigma^2 = \mathrm{E}\varepsilon_t^2$ 的估计问题.

实际计算时, 首先要对观测数据 $x_1, x_2, \cdots, x_N$ 进行零均值化预处理:

$$y_t = x_t - \overline{x}_N,\ t = 1, 2, \cdots, N, \quad \overline{x}_N = \frac{1}{N}\sum_{j=1}^{N} x_j,$$

然后为数据 $\{y_t\}$ 建立一个 AR(p) 模型. 以下假设数据 $x_1, x_2, \cdots, x_N$ 满足模型 (6.1).

(i) AR(p) 模型的 Yule-Walker 估计

从 AR(p) 模型的基本知识知道, 其自回归系数 $\boldsymbol{a}$ 由 AR(p) 序列的自协方差函数 $\gamma_0, \gamma_1, \cdots, \gamma_p$ 通过 Yule-Walker 方程

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

惟一决定. 白噪声的方差 σ^2 由 $\sigma^2 = \gamma_0 - (a_1\gamma_1 + a_2\gamma_2 + \cdots + a_p\gamma_p)$ 决定. 从观测样本 $x_1, x_2, \cdots, x_N$ 可以构造出样本自协方差函数的估计:

$$\widehat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k}(x_j - \overline{x}_N)(x_{j+k} - \overline{x}_N),\ k = 0, 1, \cdots, p, \quad \widehat{\gamma}_{-k} = \widehat{\gamma}_k,$$

所以 AR(p) 的自回归系数和白噪声方差的矩估计 $\widehat{\boldsymbol{a}} = (\widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_p)^{\mathrm{T}}$ 和 $\widehat{\sigma}^2$ 就由样本 Yule-Walker 方程

$$\begin{bmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \\ \vdots \\ \widehat{\gamma}_p \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_0 & \widehat{\gamma}_1 & \cdots & \widehat{\gamma}_{p-1} \\ \widehat{\gamma}_1 & \widehat{\gamma}_0 & \cdots & \widehat{\gamma}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{p-1} & \widehat{\gamma}_{p-2} & \cdots & \widehat{\gamma}_0 \end{bmatrix} \begin{bmatrix} \widehat{a}_1 \\ \widehat{a}_2 \\ \vdots \\ \widehat{a}_p \end{bmatrix} \tag{6.2}$$

和 $\widehat{\sigma}^2 = \widehat{\gamma}_0 - (\widehat{a}_1\widehat{\gamma}_1 + \widehat{a}_2\widehat{\gamma}_2 + \cdots + \widehat{a}_p\widehat{\gamma}_p)$ 决定.

只要 $N > p$, $x_1, x_2, \cdots, x_N$ 不全相同, p 阶样本自协方差矩阵 $\widehat{\Gamma}_p = (\widehat{\gamma}_{k-j})$ 就是正定的. 在此假设下, $\boldsymbol{a}$ 和 σ^2 的矩估计由样本 Yule-Walker 方程和方差公式惟一确定.

上述矩估计由 Yule-Walker 方程得到, 所以又被称为 Yule-Walker 估计. 它的优点是计算简便, 且样本自回归系数 $\widehat{\boldsymbol{a}} = (\widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_p)^{\mathrm{T}}$ 满足最小相位条件

$$\widehat{A}(z) = 1 - \sum_{j=1}^{p} \widehat{a}_j z^j \neq 0, \quad 当\ |z| \leq 1.$$

在实际工作中, 对于较大的 p 为了加快计算速度还可以采用如下的 Levinson 递推方法:

$$\begin{cases} \widehat{\sigma}_0^2=\widehat{\gamma}_0, \\ \widehat{a}_{1,1}=\widehat{\gamma}_1/\widehat{\gamma}_0, \\ \widehat{\sigma}_k^2=\widehat{\sigma}_{k-1}^2(1-\widehat{a}_{k,k}^2), \\ \widehat{a}_{k+1,k+1}=\dfrac{\widehat{\gamma}_{k+1}-\widehat{a}_{k,1}\widehat{\gamma}_k-\widehat{a}_{k,2}\widehat{\gamma}_{k-1}-\cdots-\widehat{a}_{k,k}\widehat{\gamma}_1}{\widehat{\gamma}_0-\widehat{a}_{k,1}\widehat{\gamma}_1-\widehat{a}_{k,2}\widehat{\gamma}_2-\cdots-\widehat{a}_{k,k}\widehat{\gamma}_k}, \\ \widehat{a}_{k+1,j}=\widehat{a}_{k,j}-\widehat{a}_{k+1,k+1}\widehat{a}_{k,k+1-j}, \quad 1\le j\le k,\ k\le p. \end{cases} \tag{6.3}$$

递推的最后得到矩估计

$$(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p)=(\widehat{a}_{p,1},\widehat{a}_{p,2},\cdots,\widehat{a}_{p,p}), \quad \widehat{\sigma}^2=\widehat{\sigma}_p^2.$$

只要样本自协方差函数是强相合估计, Yule-Walker 估计作为样本自协方差函数的连续函数, 也是强相合估计. 为了保证样本自协方差函数是强相合估计, 只需要求白噪声 $\{\varepsilon_t\}$ 是独立同分布的. 事实上, 这一要求还可以进一步放松.

定理 如果 AR(p) 模型 (6.1) 中的 $\{\varepsilon_t\}$ 是独立同分布的 WN$(0,\sigma^2)$, $\mathrm{E}\varepsilon_t^4<\infty$, 则当 $N\to\infty$ 时, (1) $\widehat{\sigma}^2\to\sigma^2$, $\widehat{a}_j\to a_j$, a.s., $1\le j\le p$;

(2) $\sqrt{N}(\widehat{a}_1-a_1,\widehat{a}_2-a_2,\cdots,\widehat{a}_p-a_p)^{\mathrm{T}}$ 依分布收敛到 p 维正态分布 $N(0,\sigma^2\Gamma_p^{-1})$;

(3) $\sqrt{N}\sup\limits_{1\le j\le p}|\widehat{a}_j-a_j|=O(\sqrt{\ln\ln N})$, a.s.,

$\sqrt{N}|\widehat{\sigma}_j^2-\sigma^2|=O(\sqrt{\ln\ln N})$, a.s..

(ii) AR(p) 模型的最小二乘估计

最小二乘估计使残差平方和最小:

$$\min S(a_1,\cdots,a_p)\xlongequal{\text{def}}\sum_{j=p+1}^{N}\left[y_j-(a_1y_{j-1}+a_2y_{j-2}+\cdots+a_py_{t-p})\right]^2$$

最小值点 $\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p$ 称为自回归系数的最小二乘估计.

引入列向量和矩阵

$$\boldsymbol{Y}=\begin{bmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{X}=\begin{bmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} \end{bmatrix}, \quad \boldsymbol{a}=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix},$$

则最小二乘估计的正规方程为 $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\boldsymbol{a}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$. 当 $p\times p$ 对称矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 正定时, 自回归系数当最小二乘估计是惟一的, 由

$$\widehat{\boldsymbol{a}}=(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p)^{\mathrm{T}}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$$

给出. 白噪声方差 σ^2 的最小二乘估计为

$$\widehat{\sigma}^2=\frac{1}{N-p}S(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p).$$

下面对最小二乘估计和 Yule-Walker 估计进行比较, 引入依概率有界的定义.

定义 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

$$\sup_n \mathrm{P}\big(|\xi_n|>M\big)\le\varepsilon,$$

就称时间序列 $\{\xi_n\}$ 是依概率有界的, 记作 $\xi_n=O_p(1)$. 如果 $\{\xi_n/c_n\}=O_p(1)$, 就称 $\xi_n=O_p(c_n)$.

与此相应地, 将随机序列 $\{\xi_n/c_n\}$ 依概率收敛于 0 记作 $\xi_n=o_p(c_n)$.

引理 1 若 $\xi_n=o_p(c_n)$, 则 $\xi_n=O_p(c_n)$.

引理 2 若 $\xi_n=O_p(1)$, 则对任意 $c_n\to\infty$, 有 $\xi_n/c_n=o_p(1)$.

引理 3 若 $\{\xi_n\}$ 同分布, 一阶矩存在, 则 $\xi_n=O_p(1)$.

对于 $1\le k\le j\le p$, 矩阵 $(1/N)\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 的第 (j,k) 元素是

$$\begin{aligned}\widetilde{\gamma}_{jk}&=\frac{1}{N}\sum_{t=1}^{N-p}y_{t+p-j}y_{t+p-k}=\frac{1}{N}\sum_{t=p-j+1}^{N-j}y_ty_{t+j-k}\\&=\widehat{\gamma}_{j-k}-\frac{1}{N}\sum_{t=1}^{p-j}y_ty_{t+j-k}-\frac{1}{N}\sum_{t=N-j+1}^{N-(j-k)}y_ty_{t+j-k},\end{aligned}$$

向量 $(1/N)\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$ 的第 j 个元素是

$$\widetilde{\gamma}_j=\frac{1}{N}\sum_{t=1}^{N-p}y_{t+p-j}y_{t+p}=\frac{1}{N}\sum_{t=p-j+1}^{N-j}y_ty_{t+j}=\widehat{\gamma}_j-\frac{1}{N}\sum_{t=1}^{p-j}y_ty_{t+j}.$$

可以证明, 对 $1\le k\le j\le p$ 有 $\widetilde{\gamma}_{jk}\to\gamma_{j-k}$, $\widetilde{\gamma}_j\to\gamma_j$, a.s., 且

$$\widetilde{\gamma}_{jk}-\widehat{\gamma}_{j-k}=O_p(\frac{1}{N}),\quad \widetilde{\gamma}_j-\widehat{\gamma}_j=O_p(\frac{1}{N}).$$

这样, 当 $N\to\infty$ 时,

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})=\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Big)^{-1}\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}\Big)\to\Gamma_p^{-1}\gamma_p=\boldsymbol{a},\ \text{a.s.},$$

即最小二乘估计也是强相合的.

用 $\widehat{\boldsymbol{a}}=(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p)^{\mathrm{T}}$ 和 $\widetilde{\boldsymbol{a}}=(\widetilde{a}_1,\widetilde{a}_2,\cdots,\widetilde{a}_p)^{\mathrm{T}}$ 分别表示 Yule-Walker 估计和最小二乘估计. 记 $\widehat{\boldsymbol{\gamma}}_p=(\widehat{\gamma}_1,\widehat{\gamma}_2,\cdots,\widehat{\gamma}_p)^{\mathrm{T}}$, 就有

$$\begin{aligned}\widetilde{\boldsymbol{a}}-\widehat{\boldsymbol{a}}&=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})-\widehat{\Gamma}_p^{-1}\widehat{\gamma}_p\\&=\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Big)^{-1}\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}-\widehat{\boldsymbol{\gamma}}_p\Big)+\Big[\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Big)^{-1}-\widehat{\Gamma}_p^{-1}\Big]\widehat{\gamma}_p\\&=\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Big)^{-1}O_p(\frac{1}{N})+\Big(\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Big)^{-1}\Big[\widehat{\Gamma}_p-\frac{1}{N}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Big]\widehat{\Gamma}_p^{-1}\widehat{\boldsymbol{\gamma}}_p=O_p(\frac{1}{N}),\quad N\to\infty.\end{aligned}$$

上式说明, 对 $1\le j\le p$, $\widetilde{\boldsymbol{a}}-\widehat{\boldsymbol{a}}=O_p(1/N)$. 所以说对较大的 N, 最小二乘估计和 Yule-Walker 估计的差别不大.

定理 设 AR(p) 模型 (6.1) 中的白噪声 $\{\varepsilon_t\}$ 是独立同分布的, $\mathrm{E}\varepsilon_t^4<\infty$, $(\widetilde{a}_1,\widetilde{a}_2,\cdots,\widetilde{a}_p)$ 是自回归系数 $(a_1,a_2,\cdots,a_p)$ 的最小二乘估计, 则当 $N\to\infty$ 时,

$$\sqrt{N}(\widetilde{a}_1-a_1,\widetilde{a}_2-a_2,\cdots,\widetilde{a}_p-a_p)$$

依分布收敛到 p 维正态分布 $N(0,\sigma^2\Gamma_p^{-1})$.

(iii) AR(p) 模型的最大似然估计

Yule-Walker 是矩估计, 最小二乘估计与 Yule-Walker 估计接近, 通常精度不高, 而最大似然估计的估计精度比较高. 设 AR(p) 模型 (6.1) 的白噪声 $\{\varepsilon_t\}$ 服从正态分布, 则 $\varepsilon_{p+1}, \varepsilon_{p+2}, \cdots, \varepsilon_N$ 有联合密度函数 (似然函数)

$$L(\boldsymbol{a}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{N}\left(x_t - \sum_{j=1}^{p} a_j x_{t-j}\right)^2\right).$$

通过对对数似然函数求导可以求得 AR(p) 模型的最大似然估计与最小二乘估计相同.

(iv) AR(p) 模型的定阶问题

实际问题中阶数 p 是未知的, 或是根本不存在的. 所以要考虑 p 的估计或选择问题. 由于 AR(p) 序列的特征是偏相关系数 p 后截尾, 所以如果样本偏相关系数 $\{\widehat{a}_{k,k}\}$ 在 $\widehat{p}$ 处截尾: $\widehat{a}_{k,k} \approx 0$, 当 $k > \widehat{p}$, 而 $\widehat{a}_{\widehat{p},\widehat{p}} \neq 0$, 就以 $\widehat{p}$ 作为 p 的估计.

对任何 $k < N$, k 阶样本自协方差矩阵是正定的. 所以样本偏相关系数 $\widehat{a}_{k,k}$ 由样本 Yule-Walker 方程

$$\begin{bmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \\ \vdots \\ \widehat{\gamma}_k \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_0 & \widehat{\gamma}_1 & \cdots & \widehat{\gamma}_{k-1} \\ \widehat{\gamma}_1 & \widehat{\gamma}_0 & \cdots & \widehat{\gamma}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{k-1} & \widehat{\gamma}_{k-2} & \cdots & \widehat{\gamma}_0 \end{bmatrix} \begin{bmatrix} \widehat{a}_{k,1} \\ \widehat{a}_{k,2} \\ \vdots \\ \widehat{a}_{k,k} \end{bmatrix} \tag{6.4}$$

惟一决定. 于是可以利用 Levinson 递推公式 (6.3) 进行递推计算.

定理 如果 AR(p) 模型 (6.1) 中的白噪声是独立同分布的, 则对任何 $k > p$,

$$\lim_{N\to\infty} \widehat{a}_{k,j} = \begin{cases} a_j, & 当\ j \leq p, \\ 0, & 当\ j > p. \end{cases}$$

这个定理保证了上述算法的合理性.

因为无法得到 $\widehat{a}_{k,k}$ 的概率分布, 为解决假设 $H_0 : a_{k,k} = 0$ 的检验问题, 只得求助于 $(\widehat{a}_{k,k} - a_{k,k})$ 的极限分布.

定理 设 $(a_{k,1}, a_{k,2}, \cdots, a_{k,k}) = (a_1, a_2, \cdots, a_p, 0, \cdots, 0)$, 如果 AR($p$) 模型 (6.1) 中的白噪声是独立同分布的, $\mathrm{E}\varepsilon_t^4 < \infty$, 则对确定的 $k > p$, 当 $N \to \infty$ 时,

$$\sqrt{N}(\widehat{a}_{k,1} - a_{k,1}, \widehat{a}_{k,2} - a_{k,2}, \cdots, \widehat{a}_{k,k} - a_{k,k})$$

依分布收敛到 k 维正态分布 $N(0, \sigma^2\Gamma_k^{-1})$.

推论 在定理的条件下, 对 $k > p$, $\sqrt{N}\widehat{a}_{k,k}$ 依分布收敛到标准正态分布 $N(0,1)$.

在实际应用中, AIC 准则和 BIC 准则也是常用的定阶方法. 假定已有阶数 p 的上界 P_0, 在假设 AR 模型的阶数是 k 时, 可以计算出相应的 AR(k) 模型白噪声方差的估计 $\widehat{\sigma}_k^2$. 引入 AIC 函数

$$\mathrm{AIC}(k) = \ln \widehat{\sigma}_k^2 + \frac{2k}{N}, \quad k = 0, 1, \cdots, P_0.$$

AIC(k) 的最小值点 $\widehat{p}$ (如不惟一, 应取小的) 称为 AR(p) 模型的 AIC 定阶.

可以证明 AIC 定阶不是相合的, 而且通常会对阶数略有高估. 当样本量不是很大时, 人们乐于使用 AIC 定阶.

为了克服 AIC 定阶的不相合性, 引入 BIC 函数

$$\mathrm{BIC}(k) = \ln \widehat{\sigma}_k^2 + \frac{k \ln N}{N}, \quad k = 0, 1, \cdots, P_0.$$

$\mathrm{BIC}(k)$ 的最小值点 $\widehat{p}$ (如不惟一, 应取小的) 称为 $\mathrm{AR}(p)$ 模型的 BIC 定阶.

可以证明 BIC 定阶是强相合的, 但当 N 不是很大时, 用 BIC 定阶有时会低估阶数 p, 效果不如 AIC.

(v) $\mathrm{AR}(p)$ 模型的拟合检验

对观测数据 $x_1, x_2, \cdots, x_N$, 如果得到了 $\mathrm{AR}(p)$ 的阶 p 和自回归系数 $(a_1, a_2, \cdots, a_p)$ 的估计量 $\widehat{p}$ 和 $(\widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_p)$, 定义残差

$$\widehat{\varepsilon}_t = y_t - \sum_{j=1}^{\widehat{p}} \widehat{a}_j y_{t-j}, \quad t = \widehat{p}+1, \widehat{p}+2, \cdots, N.$$

对上述残差序列进行白噪声检验即可.

(vi) $\mathrm{AR}(p)$ 序列的谱密度估计

将估计量 $\widehat{\sigma}^2, \widehat{p}$ 和 $(\widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_p)$ 代入 $\mathrm{AR}(p)$ 序列的谱密度函数中, 可得其估计

$$\widehat{f}(\lambda) = \frac{\widehat{\sigma}^2}{2\pi} \left|1 - \sum_{j=1}^{\widehat{p}} \widehat{a}_j e^{\mathrm{i} j \lambda}\right|^{-2}.$$

通常称 $\widehat{f}(\lambda)$ 为 AR 谱估计或极大熵谱估计. 对于 AIC 或 BIC 定阶 $\widehat{p}$, 如果 $\{\varepsilon_t\}$ 是独立同分布的 $\mathrm{WN}(0, \sigma^2)$, 可以证明, $\widehat{f}(\lambda)$ 一致收敛到 $f(\lambda)$.

6.2 MA(q) 模型的参数估计

考虑为数据 $x_1, x_2, \cdots, x_n$ 建立 $\mathrm{MA}(q)$ 模型

$$X_t = \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{6.5}$$

其中 $\{\varepsilon_t\}$ 是 $\mathrm{WN}(0, \sigma^2)$, 待估参数 $\boldsymbol{b} = (b_1, b_2, \cdots, b_q)^{\mathrm{T}}$ 满足可逆条件

$$B(z) = 1 + \sum_{j=1}^{q} b_j z^j \neq 0, \quad |z| \leq 1. \tag{6.6}$$

(i) $\mathrm{MA}(q)$ 模型的矩估计及其计算

先在 q 已知时考虑 $\boldsymbol{b}$ 和 σ^2 的估计. 令 $b_0 = 1$, 可以得到 $\boldsymbol{b}$ 满足方程组

$$\gamma_k = \sigma^2 (b_0 b_k + b_1 b_{k+1} + \cdots + b_{q-k} b_q), \quad 0 \leq k \leq q. \tag{6.7}$$

理论上说, $\boldsymbol{b}$ 和 σ^2 的矩估计可以通过解非线性方程组 (6.7) 得到, 但是实际求解 (6.7) 却不容易, 而且得到的解也不惟一和不满足可逆条件 (6.6).

引入线性迭代方法, 先利用观测数据计算出样本自协方差函数 $\widehat{\gamma}_k, k=0,1,\cdots,q$, 用 $\boldsymbol{b}(j)=\left[b_1(j),b_2(j),\cdots,b_q(j)\right]^{\mathrm{T}}$ 和 $\sigma^2(j)$ 表示第 j 次迭代值, 给定 $\boldsymbol{b}$ 和 σ^2 的初值 $\boldsymbol{b}(0)$ 和 $\sigma^2(0)$ 后, 由 (6.7) 得到迭代公式

$$\begin{cases} \sigma^2(j)=\dfrac{\widehat{\gamma}_0}{1+b_1^2(j-1)+\cdots+b_q^2(j-1)}, \\ b_k(j)=\dfrac{\widehat{\gamma}_k}{\sigma^2(j)}-\left[b_1(j-1)b_{k+1}(j-1)+\cdots+b_{q-k}(j-1)b_q(j-1)\right], \quad 1\le k\le q-1, \\ b_q(j)=\dfrac{\widehat{\gamma}_q}{\sigma^2(j)} \end{cases}$$

对给定的迭代精度 $\delta>0$, 当第 j 次迭代 $\boldsymbol{b}(j)$ 和 $\sigma^2(j)$ 满足

$$\sum_{k=0}^{q}\left|\widehat{\gamma}_k-\sigma^2(j)\sum_{t=0}^{q-k}b_t(j)b_{t+k}(j)\right|<\delta$$

时, 停止迭代, 以第 j 次迭代结果作为 $\boldsymbol{b}$ 和 σ^2 的估计量. 通常还需要验证可逆条件 (6.6) 是否成立, 如果不成立, 需要改变初值重新进行迭代计算.

下面介绍另一种方法. 根据 3.1(ii) 中介绍的方法, 由于 MA(q) 序列的自协方差函数 q 后截尾, 定义

$$\widetilde{\gamma}_k=\begin{cases}\widehat{\gamma}_k, & 0\le k\le q,\\ 0, & k>q,\end{cases} \qquad \widetilde{\Gamma}_k=(\widetilde{\gamma}_{l-j})_{l,j=1,2,\cdots,k}.$$

定义

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0\\ 0 & 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 0 & 1\\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{q\times q}, \qquad \widehat{\Omega}_k=\begin{bmatrix} \widetilde{\gamma}_1 & \widetilde{\gamma}_2 & \cdots & \widetilde{\gamma}_k\\ \widetilde{\gamma}_2 & \widetilde{\gamma}_3 & \cdots & \widetilde{\gamma}_{k+1}\\ \vdots & \vdots & \ddots & \vdots\\ \widetilde{\gamma}_q & \widetilde{\gamma}_{q+1} & \cdots & \widetilde{\gamma}_{q+k-1} \end{bmatrix},$$

$$C=(1,0,\cdots,0)^{\mathrm{T}}_{1\times q}, \qquad \boldsymbol{\gamma}_q=(\widetilde{\gamma}_1,\widetilde{\gamma}_2,\cdots,\widetilde{\gamma}_q)^{\mathrm{T}},$$

则可以定义矩估计

$$(\widehat{b}_1,\widehat{b}_2,\cdots,\widehat{b}_q)^{\mathrm{T}}=\frac{1}{\widehat{\sigma}^2}(\boldsymbol{\gamma}_q-A\widehat{\Pi}C), \quad \widehat{\sigma}^2=\widehat{\gamma}_0-C^{\mathrm{T}}\widehat{\Pi}C, \tag{6.8}$$

其中 $\widehat{\Pi}=\lim\limits_{k\to\infty}\widehat{\Omega}_k\widetilde{\Gamma}_k^{-1}\widehat{\Omega}_k^{\mathrm{T}}$.

定理 如果模型 (6.5) 中的 $\{\varepsilon_t\}$ 是独立同分布的 WN$(0,\sigma^2)$, 则几乎必然地, 当 N 充分大后由 (6.8) 计算的 $\widehat{b}_1,\widehat{b}_2,\cdots,\widehat{b}_q$ 满足可逆条件 (6.6).

(ii) MA(q) 模型的逆相关函数方法

设平稳序列 $\{X_t\}$ 有恒正的谱密度 $f(\lambda)$, 通常称 $f_y(\lambda)=\dfrac{1}{4\pi^2 f(\lambda)}$ 为 $\{X_t\}$ 的逆谱密度; 称 $\gamma_y(k)=\displaystyle\int_{-\pi}^{\pi}e^{\mathrm{i}k\lambda}f_y(\lambda)\mathrm{d}\lambda$ 为 $\{X_t\}$ 的逆相关函数或逆自相关函数.

现设 $\{X_t\}$ 是可逆 MA(q) 序列, 则 $\{X_t\}$ 有谱密度 $f(\lambda)=\dfrac{\sigma^2}{2\pi}|B(e^{\mathrm{i}\lambda})|^2$, 则逆谱密度

$$f_y(\lambda)=\frac{1}{4\pi^2 f(\lambda)}=\frac{\sigma^{-2}}{2\pi}|B(e^{\mathrm{i}\lambda})|^{-2}$$

是 AR(q) 序列

$$Y_t=-\sum_{j=1}^{q} b_j Y_{t-j}+e_t,\quad t\in\mathbb{Z} \tag{6.9}$$

的谱密度, 其中 $\{e_t\}$ 是 WN$(0,\sigma^{-2})$. 于是 $\{Y_t\}$ 的自协方差函数

$$\gamma_y(k)=\int_{-\pi}^{\pi} e^{\mathrm{i}k\lambda} f_y(\lambda)\mathrm{d}\lambda,\quad k=0,1,\cdots \tag{6.10}$$

恰好是 MA(q) 序列 $\{X_t\}$ 的逆相关函数.

如果能够对 $k=0,1,\cdots,q$ 估计 $\{X_t\}$ 的逆相关函数 (6.10), 就可以利用样本 Yule-Walker 方程

$$\begin{bmatrix}\widehat{\gamma}_y(1)\\ \widehat{\gamma}_y(2)\\ \vdots\\ \widehat{\gamma}_y(p)\end{bmatrix}=\begin{bmatrix}\widehat{\gamma}_y(0) & \widehat{\gamma}_y(1) & \cdots & \widehat{\gamma}_y(q-1)\\ \widehat{\gamma}_y(1) & \widehat{\gamma}_y(0) & \cdots & \widehat{\gamma}_y(q-2)\\ \vdots & \vdots & \ddots & \vdots\\ \widehat{\gamma}_y(q-1) & \widehat{\gamma}_y(q-2) & \cdots & \widehat{\gamma}_y(0)\end{bmatrix}\begin{bmatrix}-\widehat{b}_1\\ -\widehat{b}_2\\ \vdots\\ -\widehat{b}_q\end{bmatrix} \tag{6.11}$$

和 $\widehat{\sigma}^{-2}=\widehat{\gamma}_y(0)+\widehat{b}_1\widehat{\gamma}_y(1)+\widehat{b}_2\widehat{\gamma}_y(2)+\cdots+\widehat{b}_q\widehat{\gamma}_y(q)$ 解出 MA(q) 模型 (6.5) 中未知参数 $\boldsymbol{b}$ 和 σ^2 的估计量. 如果 $q+1$ 阶逆相关函数矩阵 $(\widehat{\gamma}_y(k-j))_{k,j=1,2,\cdots,p+1}$ 是正定的, 则得到的 $\widehat{\boldsymbol{b}}$ 满足可逆性条件 (6.6).

如何计算逆相关函数呢? 先介绍一个引理.

引理 如果 $(a_1,a_2,\cdots,a_p)$ 和 σ^2 分别是 AR(p) 模型 (6.1) 的自回归系数和白噪声 $\{\varepsilon_t\}$ 的方差, 则 $\{X_t\}$ 的逆相关函数

$$\gamma_y(k)=\begin{cases}\dfrac{1}{\sigma^2}\displaystyle\sum_{j=0}^{p-k} a_j a_{j+k}, & 0\le k\le p,\ a_0=-1,\\ 0, & k>p.\end{cases}$$

满足 (6.5) 的可逆 MA(q) 序列 $\{X_t\}$ 可以写成无穷阶自回归的形式

$$X_t-\sum_{j=1}^{\infty} a_j X_{t-j}=\varepsilon_t,\quad t=1,2,\cdots,$$

这里的回归系数 $\{a_j\}$ 由 $\dfrac{1}{B(z)}$ 在单位圆内的 Taylor 级数 $\dfrac{1}{B(z)}=1-\displaystyle\sum_{j=1}^{\infty} a_j z^j(|z|\le 1)$ 决定.

由于当 $j\to\infty$ 时, a_j 是以负指数阶收敛到 0 的, 所以对较大的正整数 p 可以将模型写成近似的长阶 (p 阶) 自回归的形式

$$X_t\approx\sum_{j=1}^{p} a_j X_{t-j}+\varepsilon_t,\quad t=1,2,\cdots,$$

利用引理可知, $\{X_t\}$ 的逆相关函数 $\gamma_y(k)$ 满足 $\gamma_y(k)\approx\dfrac{1}{\sigma^2}\displaystyle\sum_{j=0}^{p-k} a_j a_{j+k},\ 0\le k\le p,\ a_0=-1.$

根据上述想法, 可以得出计算逆相关函数 $\widehat{\gamma}_y(k)$ 和 $\widehat{b}_1, \widehat{b}_2, \cdots, \widehat{b}_q, \widehat{\sigma}^2$ 的方法:

(1) 首先利用 $\{x_t\}$ 的样本自协方差函数 $\widehat{\gamma}_k$ 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, 其中 K 是一个正数.

(2) 对 $p \xlongequal{\text{def}} p_N$ 解样本 Yule-Walker 方程 (6.2) 得到样本 Yule-Walker 系数 $(\widehat{a}_{p,1}, \widehat{a}_{p,2}, \cdots, \widehat{a}_{p,p})$ 和 $\widehat{\sigma}_p^2$.

(3) 计算样本逆相关函数

$$\widehat{\gamma}_y(k) \approx \frac{1}{\widehat{\sigma}^2} \sum_{j=0}^{p-k} \widehat{a}_{p,j} \widehat{a}_{p,j+k},\ k = 0, 1, 2, \cdots, q,\ \widehat{a}_{p,0} = -1$$

注意, 当 $\widehat{a}_{p,1}, \widehat{a}_{p,2}, \cdots, \widehat{a}_{p,p}$ 不全相同时, (6.11) 中的 $q \times q$ 系数矩阵就是正定的.

(4) 利用样本 Yule-Walker 方程 (6.11) 计算出 MA(q) 系数的估计量 $\widehat{\boldsymbol{b}}_q$ 和 $\widehat{\sigma}^2$.

(iii) MA(q) 模型的新息估计方法

对于 MA(q) 序列 (6.5), $\widehat{X}_1 = 0$,

$$\widehat{X}_{k+1} = L(X_{k+1} | X_k, X_{k-1}, \cdots, X_1)$$

是用 $X_1, X_2, \cdots, X_k$ 预报 X_{k+1} 时的最佳线性预报, 用

$$\widehat{\varepsilon}_{k+1} = X_{k+1} - L(X_{k+1} | X_k, X_{k-1}, \cdots, X_1)$$

表示样本新息, 用 $\nu_k = \mathrm{E}\widehat{\varepsilon}_{k+1}^2$ 表示预测的均方误差.

由 5.4(ii) 的内容可知, 用样本新息 $\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \cdots, \widehat{\varepsilon}_m$ 对 X_{m+1} 所做的最佳线性预报

$$\widehat{X}_{m+1} = \sum_{j=1}^{q} \theta_{m,j} \widehat{\varepsilon}_{m+1-j}, \quad m \geq q \tag{6.12}$$

和用 $X_1, X_2, \cdots, X_m$ 预报 X_{m+1} 时的最佳线性预报是一样的. 如果给定了 $\{X_t\}$ 的自协方差函数 $\{\gamma_k\}$, (6.12) 中的 $\theta_{m,j}$ 和预报的均方误差 ν_m 可以按 (5.22) 进行递推计算.

对于 MA(q) 序列 (6.5), 白噪声 $\{\varepsilon_t\}$ 正是这个 MA(q) 序列的新息序列: ε_m 是用所有的历史资料 $X_{m-k}(k = 1, 2, \cdots)$ 预测 X_m 时的预测误差. 所以当 m 取值较大时, $\widehat{\varepsilon}_m = X_m - \widehat{X}_m$ 是 ε_m 的近似. 基于此, 对较大的 t, MA(q) 模型 (6.5) 的近似是

$$\begin{aligned} X_t &\approx \widehat{\varepsilon}_t + b_1 \widehat{\varepsilon}_{t-1} + \cdots + b_q \widehat{\varepsilon}_{t-q} \\ &= X_t - \widehat{X}_t + b_1 \widehat{\varepsilon}_{t-1} + \cdots + b_q \widehat{\varepsilon}_{t-q}, \\ \Rightarrow \widehat{X}_t &= b_1 \widehat{\varepsilon}_{t-1} + \cdots + b_q \widehat{\varepsilon}_{t-q}. \end{aligned}$$

和 (6.12) 比较, 可知取 $\widehat{b}_j = \theta_{m,j}$ 作为 b_j 的估计, 以 ν_m 作为 σ^2 的估计. 这种估计称为新息估计. 具体计算方法如下:

给定观测数据 $x_1, x_2, \cdots, x_N$, 取 $m = o(N^{1/3})$, 计算样本自协方差函数 $\widehat{\gamma}_0, \widehat{\gamma}_1, \cdots, \widehat{\gamma}_m$, $\boldsymbol{b}$ 和 σ^2 的新息估计

$$(\widehat{b}_1, \widehat{b}_2, \cdots, \widehat{b}_q) = (\widehat{\theta}_{m,1}, \widehat{\theta}_{m,2}, \cdots, \widehat{\theta}_{m,q}), \quad \widehat{\sigma}^2 = \widehat{\nu}_m \tag{6.13}$$

由下面的递推公式得到

$$\begin{cases} \widehat{\nu}_0 = \widehat{\gamma}_0, \\ \widehat{\theta}_{n,n-k} = \widehat{\nu}_k^{-1}\Big[\widehat{\gamma}_{n-k} - \sum_{j=0}^{k-1} \widehat{\theta}_{k,k-j}\widehat{\theta}_{n,n-j}\widehat{\nu}_j\Big], \quad 0 \le k \le n-1, \\ \widehat{\nu}_n = \widehat{\gamma}_0 - \sum_{j=0}^{n-1} \widehat{\theta}_{n,n-j}^2 \widehat{\nu}_j, \quad 1 \le n \le m, \end{cases} \tag{6.14}$$

其中 $\sum_{j=0}^{-1}(\cdot) \stackrel{\text{def}}{=\!=} 0$, 递推的顺序是 $\widehat{\nu}_0; \widehat{\theta}_{1,1}, \widehat{\nu}_1; \widehat{\theta}_{2,2}, \widehat{\theta}_{2,1}, \widehat{\nu}_2; \cdots$.

新息估计的合理性可以从下面的定理看出.

定理 设 $\{X_t\}$ 是可逆 ARMA(p,q) 序列, 满足 $A(\mathscr{B})X_t = B(\mathscr{B})\varepsilon_t, t \in \mathbb{Z}$, $\{\psi_j\}$ 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的 WN$(0,\sigma^2)$, 正整数列 $m = m(N) < N$ 满足当 $N \to \infty$ 时, $m \to \infty$ 和 $m = o(N^{1/3})$. 则对任何正整数 q, 当 $N \to \infty$ 时

$$\sqrt{N}(\widehat{\theta}_{m,1} - \psi_1, \widehat{\theta}_{m,2} - \psi_2, \cdots, \widehat{\theta}_{m,q} - \psi_q) \xrightarrow{\text{d}} N_q(0, A),$$

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 $\widehat{\nu}_m \xrightarrow{\text{P}} \sigma^2$.

推论 对于 MA(q) 序列 $\{X_t\}$, 当模型 (6.5) 中的白噪声是 4 阶矩有限的独立同分布序列时, 新息估计 (6.13) 是相合估计; 当 $N \to \infty$ 时,

$$\sqrt{N}(\widehat{b}_1 - b_1, \widehat{b}_2 - b_2, \cdots, \widehat{b}_q - b_q) \xrightarrow{\text{d}} N_q(0, A),$$

其中 $q \times q$ 矩阵

$$A = (a_{i,j}), \quad a_{i,j} = \sum_{k=1}^{\min(i,j)} b_{i-k}b_{j-k}, \quad b_0 \equiv 1,$$

并且 $\widehat{\nu}_m \xrightarrow{\text{P}} \sigma^2$.

(iv) MA(q) 模型的定阶方法

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 $\widehat{\rho}_k = \widehat{\gamma}_k/\widehat{\gamma}_0$ 从某一点 $\widehat{q}$ 后变得很小时, 可以将 $\widehat{q}$ 作为 q 的估计.

另外, 由 4.2(iii) 的定理可知, 对于 $m > q$, $\sqrt{N}\widehat{\rho}_m \xrightarrow{\text{d}} N(0, 1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_q^2)$. 这样由样本自相关系数 $\widehat{\rho}_k (k = 1, 2, \cdots, m)$ 的图形可以大致得到 q 的估计, 同时也可以判断采用 MA(q) 模型到合理性.

下面介绍 AIC 定阶方法: 设 MA(q) 模型阶数 q 的上界是 Q_0, 对于 $m = 0, 1, 2, \cdots, Q_0$ 按前述方法逐个拟合 MA(m) 模型, 得到白噪声方差估计量 $\widehat{\sigma}_m^2$. 定义 AIC 函数

$$\text{AIC}(m) = \ln \widehat{\sigma}_m^2 + 2m/N, \quad m = 0, 1, 2, \cdots, Q_0,$$

其中 N 是样本个数. AIC(m) 的最小值点 $\widehat{q}$ (如不惟一, 应取小的) 称为 MA(q) 模型的 AIC 定阶.

(v) MA(q) 模型的拟合检验

从观测数据 $x_1, x_2, \cdots, x_N$ 得到模型的参数估计 $\widehat{q}, \widehat{b}_1, \widehat{b}_2, \cdots, \widehat{b}_q$ 和 $\widehat{\sigma}^2$ 后, 取

$$\widehat{\varepsilon}_{1-\widehat{q}} = \widehat{\varepsilon}_{2-\widehat{q}} = \cdots = \widehat{\varepsilon}_0 = 0, \quad y_t = x_t - \overline{x}_N$$

和 $\widehat{\varepsilon}_t = y_t - \sum_{j=1}^{\widehat{q}} \widehat{b}_j \widehat{\varepsilon}_{t-j}, t = 1, 2, \cdots, N$. 对 $L = O(N^{1/3})$, 如果 $\{\widehat{\varepsilon}_t : t = L, L+1, \cdots, N\}$ 能够通过白噪声检验, 就认为模型选择合适, 否则改变 $\widehat{q}$ 的值, 拟合新的 MA 模型或改用其它如 ARMA 模型.

(vi) 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, 则其谱密度估计为

$$\widehat{f}(\lambda) = \frac{\widehat{\sigma}^2}{2\pi} \left| 1 + \sum_{j=1}^{\widehat{p}} \widehat{b}_j e^{\mathrm{i}j\lambda} \right|^2.$$

可以证明, $\widehat{f}(\lambda)$ 是 MA(q) 序列 (6.5) 谱密度 $f(\lambda)$ 的相合估计.

6.3 ARMA(p, q) 模型的参数估计

对于零均值化后的平稳观测数据 $x_1, x_2, \cdots, x_N$, 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.

设 $x_1, x_2, \cdots, x_N$ 满足如下的可逆 ARMA(p, q) 模型

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \quad t = 1, 2, \cdots, \tag{6.15}$$

其中 $\{\varepsilon_t\}$ 是 WN$(0, \sigma^2)$, 未知参数 $\boldsymbol{a} = (a_1, a_2, \cdots, a_p)^{\mathrm{T}}$ 和 $\boldsymbol{b} = (b_1, b_2, \cdots, b_q)^{\mathrm{T}}$ 使得多项式

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j, \quad B(z) = 1 + \sum_{j=1}^{q} b_j z^j$$

互素, 并且满足 $A(z)B(z) \neq 0, |z| \leq 1$.

(i) ARMA(p, q) 模型的矩估计方法

利用延伸的 Yule-Walker 方程 (3.19) 可知 ARMA(p, q) 序列的自协方差函数满足

$$\begin{bmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{bmatrix} = \begin{bmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \tag{6.16}$$

设 $\{\widehat{\gamma}_k\}$ 是 ARMA(p, q) 序列的样本自协方差函数, 则由估计方程 (6.16) 可知 $\boldsymbol{a}$ 的矩估计是

$$\begin{bmatrix} \widehat{a}_1 \\ \widehat{a}_2 \\ \vdots \\ \widehat{a}_p \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_q & \widehat{\gamma}_{q-1} & \cdots & \widehat{\gamma}_{q-p+1} \\ \widehat{\gamma}_{q+1} & \widehat{\gamma}_q & \cdots & \widehat{\gamma}_{q-p+2} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{q+p-1} & \widehat{\gamma}_{q+p-2} & \cdots & \widehat{\gamma}_q \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\gamma}_{q+1} \\ \widehat{\gamma}_{q+2} \\ \vdots \\ \widehat{\gamma}_{q+p} \end{bmatrix}. \tag{6.17}$$

利用定理可知 (6.16) 中的 $p\times p$ 矩阵 $\Gamma_{p,q}$ 是可逆的. 用 $\widehat{\Gamma}_{p,q}$ 表示 (6.17) 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, $\widehat{\gamma}_k\to\gamma_k$, a.s.. 于是当 $N\to\infty$ 时,

$$\det(\widehat{\Gamma}_{p,q})\to\det(\Gamma_{p,q})\neq 0.$$

所以当 N 充分大后, (6.17) 中的矩阵 $\widehat{\Gamma}_{p,q}$ 也是可逆的, 由此可知 $\boldsymbol{a}$ 的矩估计是惟一的. 另外, 在上述条件下, 矩估计 (6.17) 是强相合的, 即

$$\lim_{N\to\infty}\widehat{a}_j=a_j,\ \text{a.s.},\quad 1\leq j\leq p.$$

下面估计 MA(q) 部分的参数. 由于

$$y_t\triangleq x_t-\sum_{j=1}^{p}a_jx_{t-j},\quad t=p+1,p+2,\cdots,N$$

满足 MA(q) 模型 $y_t=B(\mathscr{B})\varepsilon_t$, 所以得到 $\widehat{\boldsymbol{a}}=(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p)^{\mathrm{T}}$ 后,

$$y_t=x_t-\sum_{j=1}^{p}\widehat{a}_jx_{t-j},\quad t=p+1,p+2,\cdots,N$$

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

$$\widehat{\gamma}_y(k)=\sum_{j=0}^{p}\sum_{l=0}^{p}\widehat{a}_j\widehat{a}_l\widehat{\gamma}_{k+j-l},\quad k=0,1,\cdots,q \tag{6.18}$$

定义, 其中 $\widehat{a}_0=-1$. 将 (6.18) 看成一个 MA(q) 序列的样本自协方差函数, 利用 6.2 节的方法就可以估计出 MA(q) 部分的参数 $\boldsymbol{b}$ 和 σ^2.

(ii) ARMA(p,q) 模型的自回归逼近法

出发点与想法: 如果 ARMA 模型中已知 ε_t, 则 $a_1,a_2,\cdots,a_p,b_1,b_2,\cdots,b_q$ 可以看成是回归系数. 而 ε_t 作为一步预报误差, 可以用样本新息估计. 但样本新息直接计算困难, 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 $P_0=[\sqrt{N}]$, 其中 $[a]$ 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 $\widehat{p}$ 和自回归系数的估计 $(\widehat{\phi}_1,\widehat{\phi}_2,\cdots,\widehat{\phi}_{\widehat{p}})$, 计算残差

$$\widehat{\varepsilon}_t=x_t-\sum_{j=1}^{\widehat{p}}\widehat{\phi}_jx_{t-j},\quad t=\widehat{p}+1,\widehat{p}+2,\cdots,N.$$

然后写出近似的 ARMA(p,q) 模型

$$x_t=\sum_{j=1}^{p}a_jx_{t-j}+\widehat{\varepsilon}_t+\sum_{j=1}^{q}b_j\widehat{\varepsilon}_{t-j},\quad t=L+1,L+2,\cdots,N,$$

其中 $L=\max(\widehat{p},p,q)$, a_j,b_j 是待定参数.

最后, 由最小二乘估计的想法, 对目标函数

$$Q(\boldsymbol{a},\boldsymbol{b})=\sum_{t=L+1}^{N}\left(x_t-\sum_{j=1}^{p}a_jx_{t-j}-\sum_{j=1}^{q}b_j\widehat{\varepsilon}_{t-j}\right)^2$$

极小化, 可以得到最小二乘估计 $(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p,\widehat{b}_1,\widehat{b}_2,\cdots,\widehat{b}_q)$. σ^2 的最小二乘估计由下式定义

$$\widehat{\sigma}^2=\frac{1}{N-L}Q(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_p,\widehat{b}_1,\widehat{b}_2,\cdots,\widehat{b}_q).$$

(iii) 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 $n \geq 1$, $\boldsymbol{X}_n = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 的协方差矩阵 Γ_n 正定. 采用 5.3 节中的记号可以得到最佳线性预测

$$\widehat{X}_n \xlongequal{\text{def}} L(X_n|X_{n-1}) = \sum_{j=1}^{n-1} \theta_{n-1,n-j} Z_j,$$

其中 $Z_j = X_j - \widehat{X}_j, Z_1 = X_1$. 于是

$$X_n = \widehat{X}_n + Z_n = \sum_{j=1}^{n-1} \theta_{n-1,n-j} Z_j + Z_n = (\theta_{n-1,n-1}, \theta_{n-1,n-2}, \cdots, \theta_{n-1,1}, 1)\boldsymbol{Z}_n,$$

其中 $\boldsymbol{Z}_n = (Z_1, Z_2, \cdots, Z_n)^{\mathrm{T}}$. 为了方便引入下三角矩阵

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix},$$

则有 $\boldsymbol{X}_n = C\boldsymbol{Z}_n$.

由于 $r_{k-1} \xlongequal{\text{def}} \mathrm{E}Z_k^2, k = 1, 2, \cdots, n-1$ 是预测的均方误差, 所以用 $Z_1, Z_2, \cdots, Z_n$ 的正交性得到

$$D \xlongequal{\text{def}} \mathrm{E}(\boldsymbol{Z}_n \boldsymbol{Z}_n^{\mathrm{T}}) = \mathrm{diag}(r_0, r_1, \cdots, r_{n-1}).$$

由此得到 $\boldsymbol{X}_n$ 的协方差矩阵:

$$\begin{aligned} &\Gamma_n = \mathrm{E}(C\boldsymbol{Z}_n \boldsymbol{Z}_n^{\mathrm{T}} C^{\mathrm{T}}) = CDC^{\mathrm{T}}, \\ &\det(\Gamma_n) = \mathrm{def}(D) = r_0 r_1 \cdots r_{n-1}, \\ &\boldsymbol{X}_n^{\mathrm{T}} \Gamma_n^{-1} \boldsymbol{X}_n = \boldsymbol{Z}_n C^{\mathrm{T}} (CDC^{\mathrm{T}})^{-1} C\boldsymbol{Z}_n = \sum_{j=1}^{n} Z_j^2 / r_{j-1}. \end{aligned}$$

由于 $\boldsymbol{X}_n$ 的分布由 Γ_n 决定, 而 $r_j, \theta_{k,j}$ 都是 Γ_n 的函数, 所以可得基于 $\boldsymbol{X}_n$ 的似然函数

$$\begin{aligned} L(\Gamma_n) &= \frac{1}{(2\pi)^{n/2}[\det(\ _{\mathrm{n}})]^{1/2}} \exp\Big(-\frac{1}{2}\boldsymbol{X}_n^{\mathrm{T}} \Gamma_n^{-1} \boldsymbol{X}_n\Big) \\ &= \frac{1}{(2\pi)^{n/2}(r_0 r_1 \cdots r_{n-1})^{1/2}} \exp\Big(-\frac{1}{2}\sum_{j=1}^{n} Z_j^2 / r_{j-1}\Big). \end{aligned} \tag{6.19}$$

(iv) ARMA(p,q) 模型的最大似然估计

设 $\{X_t\}$ 是满足 ARMA 模型 (6.15) 的平稳序列. 采用 5.4(iii) 的逐步预报的递推公式 (5.27) 可知

$$\widehat{X}_{k+1} = \begin{cases} \displaystyle\sum_{j=1}^{k} \theta_{k,j} Z_{k+1-j}, & 1 \leq k < m, \\ \displaystyle\sum_{j=1}^{p} a_j X_{k+1-j} + \sum_{j=1}^{q} \theta_{k,j} Z_{k+1-j}, & k \geq m. \end{cases} \tag{6.20}$$

这里 $m = \max(p,q)$,

$$Z_k = X_k - \widehat{X}_k = X_k - L(X_k|X_{k-1}, X_{k-2}, \cdots, X_1),$$
$$r_{k-1} = \mathrm{E}Z_k^2 = \mathrm{E}(X_k - \widehat{X}_k)^2 = \sigma^2\nu_{k-1}.$$

$\theta_{k,j}, \nu_k$ 可以通过 (5.26) 和 (5.21) 递推计算, 而 (5.26) 中的 $\sigma^{-2}\gamma_k$ 由 (3.16) 和 (3.17) 计算. 它们都是和 σ^2 无关的量, 有 ARMA 模型 (6.15) 的参数

$$\boldsymbol{\beta} = (a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_q)^{\mathrm{T}}$$

惟一决定. 从而 $\widehat{X}_k$ 也是和 σ^2 无关的量, 仅由观测数据 $X_1, X_2, \cdots, X_{k-1}$ 和 $\boldsymbol{\beta}$ 决定. 将以上式子代入 (6.19) 得到似然函数

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{\exp\left(-\dfrac{1}{2\sigma^2}\displaystyle\sum_{k=1}^{N} Z_k^2/\nu_{k-1}\right)}{(2\pi)^{N/2}(\sigma^{2N}\nu_0\nu_1\cdots\nu_{N-1})^{1/2}}. \tag{6.21}$$

记 $S(\beta) = \displaystyle\sum_{k=1}^{N} z_k^2/\nu_{k-1}$, 通过对对数似然函数求导可得 σ^2 的最大似然估计是 $\widehat{\sigma}^2 = \dfrac{1}{N}S(\widehat{\boldsymbol{\beta}})$, 这里 $\widehat{\boldsymbol{\beta}}$ 是约化似然函数

$$l(\boldsymbol{\beta}) \xlongequal{\text{def}} -\frac{2}{N}\ln L(\boldsymbol{\beta}, S(\boldsymbol{\beta})/N) - 1 = \frac{1}{N}\ln(\nu_0\nu_1\cdots\nu_{n-1}) + \ln\frac{S(\boldsymbol{\beta})}{N} \tag{6.22}$$

的最小值点 $\widehat{\boldsymbol{\beta}} = (\widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_p, \widehat{b}_1, \widehat{b}_2, \cdots, \widehat{b}_q)^{\mathrm{T}}$, 是 $\boldsymbol{\beta}$ 的最大似然估计.

在计算 $l(\boldsymbol{\beta})$ 的最小点时, 可采用一般的最优化搜索方法. 给定任何初值 $\boldsymbol{\beta}$, 通过前面得到的 $\theta_{k,j}, \nu_j$ 和 $Z_k = X_k - \widehat{X}_k$, 计算出 $l(\boldsymbol{\beta})$. 为了加快搜索的速度和提高估计的精度, 初值应选择在最小值附近, 实际计算时通常选成矩估计或自回归逼近估计.

下面将最大似然估计与最小二乘估计做对比. 当 $k \to \infty$ 时, 由预报的极限性质

$$\nu_k = \mathrm{E}(X_{k+1} - \widehat{X}_{k+1})^2/\sigma^2 = \mathrm{E}\big[X_n - L(X_n|X_{n-1}, X_{n-2}, \cdots, X_{n-k})\big]^2/\sigma^2 \to \mathrm{E}\varepsilon_n^2/\sigma^2 = 1.$$

所以当 $N \to \infty$ 时, 有 $\dfrac{1}{N}\ln(\nu_0\nu_1\cdots\nu_{N-1}) \to 0$. 于是对较大的 N, $l(\boldsymbol{\beta})$ 和 $S(\boldsymbol{\beta})$ 的最小值点近似相等, 也可以用 $S(\boldsymbol{\beta})$ 的最小值点 $\widetilde{\boldsymbol{\beta}}$ 作为 $\boldsymbol{\beta}$ 的估计, 通常称之为最小二乘估计, 相应的白噪声方差 σ^2 的估计定义为

$$\widetilde{\sigma}^2 = \frac{1}{N-p-q}S(\widetilde{\boldsymbol{\beta}}).$$

可以证明, 最小二乘估计 $\widetilde{\boldsymbol{\beta}}$ 和最大似然估计 $\widehat{\boldsymbol{\beta}}$ 有相同的极限分布, $\widetilde{\sigma}^2$ 与 $\widehat{\sigma}^2$ 有相同的极限分布.

定理 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, $\mathrm{E}\varepsilon_t^4 < \infty$, 则当 $N \to \infty$ 时,

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} N(0, V(\boldsymbol{\beta})),$$

其中

$$V(\boldsymbol{\beta}) = \begin{cases} \sigma^2\begin{bmatrix} \mathrm{E}(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}) & \mathrm{E}(\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}) \\ \mathrm{E}(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}) & \mathrm{E}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}) \end{bmatrix}^{-1}, & p > 0, q > 0, \\ \sigma^2\big[\mathrm{E}(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})\big]^{-1}, & q = 0, p > 0, \\ \sigma^2\big[\mathrm{E}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}})\big]^{-1}, & p = 0, q > 0. \end{cases} \tag{6.23}$$

上式中

$$\boldsymbol{U}=(U_p,U_{p-1},\cdots,U_1)^{\mathrm{T}},\quad \boldsymbol{V}=(V_q,V_{q-1},\cdots,V_1)^{\mathrm{T}}$$

$\{U_t\}$ 和 $\{V_t\}$ 分别是 AR(p) 和 AR(q) 序列, 满足

$$A(\mathscr{B})U_t=\varepsilon_t \text{ 和 } B(\mathscr{B})V_t=\varepsilon_t.$$

注: 在实际问题中, 真值 $V(\boldsymbol{\beta})$ 是未知的, 通常用估计量 $V(\widehat{\boldsymbol{\beta}})$ 代替.

(v) ARMA(p,q) 模型的检验

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 $A(z)B(z)\neq 0,|z|\leq 1$. 然后对取定的初值

$$x_0=x_{-1}=\cdots=x_{-p+1}=\widehat{\varepsilon}_0=\cdots=\widehat{\varepsilon}_{-q+1}=0$$

递推计算模型的残差

$$\widehat{\varepsilon}_t=x_t-\sum_{j=1}^{p}\widehat{a}_jx_{t-j}+\sum_{j=1}^{q}\widehat{b}_j\widehat{\varepsilon}_{t-j},\quad t=1,2,\cdots.$$

取 $m=O(N^{1/3})$ 和 $m>\max(p,q)$, 如果残差 $\widehat{\varepsilon}_t,t=m,m+1,\cdots,N$ 可以通过白噪声检验, 就认为模型合适; 否则寻找其它的模型.

(vi) ARMA(p,q) 模型的定阶方法

在实际计算中, 参数 (p,q) 是未知的, 但根据数据的性质有时可以知道阶数的大致范围, 可以在此范围内对每一对 (p,q) 建立 ARMA(p,q) 模型. 如果不能确定阶数的范围, 可以采用从 $p+q=1,p+q=2,\cdots$, 开始由低阶到高阶依次搜寻, 选出 $p+q$ 最小的一个符合的模型.

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 AIC 定阶方法, 最后确定一个模型.

给定 ARMA(p,q) 模型的阶数 (p,q) 的一个估计 $(k,j)=(\widehat{p},\widehat{q})$, 可以估计出 ARMA($k,j$) 模型的参数, 用 $\widehat{\sigma}^2=\widehat{\sigma}^2(k,j)$ 表示白噪声方差 σ^2 的估计. 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

$$\mathrm{AIC}(k,j)=\ln(\widehat{\sigma}^2(k,j))+\frac{2(k+j)}{N}.$$

AIC(k,j) 的最小值点 $(\widehat{p},\widehat{q})$ 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, 有时会高估阶数.

将 AIC 定阶中的惩罚项修改就可以得到 BIC 定阶

$$\mathrm{BIC}(k,j)=\ln(\widehat{\sigma}^2(k,j))+\frac{(k+j)\ln N}{N}.$$

(vii) ARMA(p,q) 模型的谱密度估计

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

$$\widehat{f}(\lambda)=\frac{\widehat{\sigma}^2}{2\pi}\frac{|1+\sum_{j=1}^{\widehat{q}}\widehat{b}_je^{\mathrm{i}j\lambda}|^2}{|1-\sum_{j=1}^{\widehat{p}}\widehat{a}_je^{\mathrm{i}j\lambda}|^2}$$

作为谱密度的估计. 可以证明 $\widehat{f}(\lambda)$ 是 ARMA(p,q) 序列 (6.15) 谱密度 $f(\lambda)$ 的相合估计.