

第 2 章 一元时间序列的基本概念和 ARIMA 模型

本章介绍时间序列的概念、模型及一些方法,体现了人们用数学语言来描述现实世界的努力.任何时间序列的模型都试图近似地描述一些真实的时间序列.当然它们都不等同于实际的序列.本章要介绍的一些模型反映了人们所掌握的数学手段,但并不一定反映人们可能面对的现实世界.这些模型的定义都含有各种在实践中无法验证的假定.实际上,这些模型仅仅是人们所发明的各种数学模型的一部分.读者肯定明白,数学模型的复杂性不一定与其实用性成正比.

2.1 时间序列的平稳性及相关性度量

在经典的回归分析教科书中,多数数据都是所谓的横截面数据,即每个对象只取一次观测值.比如一个金融机构记录的各个客户企业的总资产、货币资产、净资产、净债务、经营活动现金流、信用等级等就组成了一个横截面数据.但如果这个金融机构对其客户每年都做同样内容的记录,那么多年的记录就形成了多个变量(多元)的时间序列,其中任何一个变量的序列就是单变量或一元时间序列.由于时间序列是一种随机过程,我们也经常使用术语“过程”来表示时间序列.

在经典数理统计基础教科书中,经常会考虑一个变量的独立同分布的样本,也就是说观测值 $\{Y_1, \dots, Y_n\}$ 都是来自一个总体的独立观测值.但如果对于一个对象在不同时间进行观测,所得到的观测值就不一定独立同分布了.比如一个地点每小时所进行的气温记录 $\{X_1, \dots, X_n\}$ 就显然不是独立同分布的了,每个时间的温度都和前后的温度相关(称为自相关,即温度变量自身各个观测值之间的相关).因此,在独立同分布条件下的统计方法就不适用了,需要引进时间序列分析的各种方法.

观测值之间最重要的关系度量就是相关性的度量,本节就介绍时间序列的相关性

度量, 以及在时间序列分析中很重要的平稳性概念和差分的方法.

一个有 N 个顺序观测值的时间序列 $\{X_1, X_2, \dots, X_N\}$ 可以看成随机过程 $\{X_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ 的一个部分.^①

2.1.1 平稳、自协方差函数和自相关函数

平稳性

对于时间序列 X_1, X_2, \dots , 如果称之为**严平稳的** (strictly stationary), 则它必须满足以下条件: 对于任何 t_1, \dots, t_k , 滞后期 τ 和 k , X_{t_1}, \dots, X_{t_k} 的联合分布与 $X_{t_1+\tau}, \dots, X_{t_k+\tau}$ 的联合分布相同. 如果 $k = 1$, 那么 X_t 的分布对所有的 t 都相同, 而且均值和方差都不随 t 而变.

对于时间序列 X_1, X_2, \dots , 如果称之为**弱平稳的** (weak stationary), 或者**二阶平稳的** (second-order stationary), 或者**协方差平稳的** (covariance stationary), 则它必须满足下面条件: X_t 的均值 (数学期望) 不随时间而改变, 即对于任何 t , $E(X_t) = \mu$ (这里 μ 为一个常数), 而且, 对于任何滞后期 τ , X_t 与 $X_{t+\tau}$ 的相关系数 $\text{Cov}(X_t, X_{t+\tau}) = \gamma_\tau$, 即该相关系数仅仅依赖于 τ , 与时间 t 无关. 显然, 平稳时间序列的方差也是一个常数: $\text{Var}(X_t) = \gamma_0$.

注意, 除非在某些假定条件下, 严平稳和弱平稳没有一个包含另一个的关系, 一般教科书都着重讨论弱平稳 (即二阶平稳), 本书后面所涉及的平稳都是弱平稳. 弱平稳时间序列有不随时间改变的一阶和二阶矩.

一类很重要的过程是正态过程, 即序列 X_{t_1}, \dots, X_{t_k} 对于所有的 t_1, \dots, t_k 都是多元正态分布的. 而多元正态分布则完全被该分布的一阶矩 $\mu(t)$ 和二阶矩 $\gamma(t_1, t_2)$ 所决定, 因此, 对于正态过程, 弱平稳或二阶平稳就意味着严平稳. 但如果过程非常“不正态”, 仅靠 μ 和 γ_τ 就不能很好地描述平稳过程了.

自协方差函数和自相关函数

假定 X_t 和 X_s 的均值分别为 $\mu_t = E(X_t)$, $\mu_s = E(X_s)$, 那么 X_t 和 X_s 的**自协方差函数** (auto-covariance function, acvf) 定义为

$$\gamma(t, s) = \text{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)].$$

^① 人们通常用小写字母来记该时间序列的观测值 (随机序列的具体实现): x_1, x_2, \dots, x_N , 但在某些不会混淆的情况下, 随机序列及其实现都用小写字母表示. 此外, 为简单计, 一个时间序列 X_1, X_2, \dots, X_N 往往用花括号 $\{X_t\}_1^N$ 表示, 类似地, 对无穷序列有 $\{X_t\}_1^\infty$ 或 $\{X_t\}_-\infty^\infty$ 之类的表示, 在不会混淆时, 均用 $\{X_t\}$ 表示.

对于平稳时间序列, 如果 $\tau = s - t$ 及 $\mu = E(X_t) = E(X_s)$, 则有自协方差函数^①

$$\begin{aligned}\gamma(h) &\equiv \gamma_\tau \equiv \gamma(t, s) \\ &= E[(X_t - \mu_t)(X_s - \mu_s)] \\ &= E[(X_t - \mu)(X_{t+\tau} - \mu)].\end{aligned}$$

也就是说, 平稳时间序列的自协方差函数 γ_τ 仅仅依赖于时间差 τ , 与绝对时间无关.

平稳时间序列的自相关函数 (auto-correlation function, acf) 定义为^②

$$\rho(\tau) \equiv \rho_\tau \equiv \frac{\gamma_\tau}{\gamma_0} = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\text{Var}(X_t)}.$$

对于平稳序列, 有

$$\gamma_\tau = \text{Cov}(X_t, X_{t+\tau}) = \text{Cov}(X_{t-\tau}, X_t) = \gamma_{-\tau}.$$

自相关函数的性质如下:

若记方差 $\sigma^2 = \text{Var}(X_t) = \gamma_0$, 由 $\rho_\tau = \gamma_\tau/\gamma_0 = \gamma_\tau/\sigma^2$, 可得

$$\rho_\tau = \rho_{-\tau}.$$

此外, 关于 ρ_τ , 还有

$$\|\rho_\tau\| \leq 1.$$

这是因为, 根据方差的性质, 对于任意的常数 λ_1, λ_2 , 有

$$\begin{aligned}0 &\leq \text{Var}(\lambda_1 X_t + \lambda_2 X_{t+\tau}) \\ &= \lambda_1^2 \text{Var}(X_t) + \lambda_2^2 \text{Var}(X_{t+\tau}) + 2\lambda_1 \lambda_2 \text{Cov}(X_t, X_{t+\tau}) \\ &= (\lambda_1^2 + \lambda_2^2) \sigma^2 + 2\lambda_1 \lambda_2 \gamma_\tau,\end{aligned}$$

所以得到

$$\text{当 } \lambda_1 = \lambda_2 = 1 \text{ 时: } \gamma_\tau \geq -\sigma^2 \Rightarrow \rho_\tau \geq -1;$$

^① 为方便计, 我们既用 $\gamma(h)$ 也用 γ_h 表示平稳序列的自协方差函数.

^② 为方便计, 我们既用 $\rho(h)$ 也用 ρ_h 表示平稳序列的自相关函数.

当 $\lambda_1 = 1, \lambda_2 = -1$ 时: $\gamma_\tau \leq \sigma^2 \Rightarrow \rho_\tau \leq +1$.

即得 $\|\rho_\tau\| \leq 1$.

注意, 自相关函数 acf 并不唯一识别背景模型. 虽然一个给定的随机过程都有唯一的 acf, 但是一个给定的自相关函数并不一定对应唯一的随机过程. 即使是不同的正态过程, 也可能有同样的自相关函数. 这时, 就需要所谓的**可逆性条件** (invertibility condition) 来保证唯一性. 这样就可以在可逆性条件下通过自相关函数去识别一个随机过程. 关于可逆性条件, 请参见后面章节中的介绍.

如果 $\{x_t\}_{t=1}^T$ (代表 x_1, \dots, x_T) 为时间序列 $\{X_t\}_{t=1}^T$ (代表 X_1, \dots, X_T) 的样本, 滞后期 τ 的样本自协方差函数 (sample auto-covariance function, sacvf) 及滞后期 τ 的样本自相关函数 (sample auto-correlation function, sacf) 分别定义为

$$\hat{\gamma}_\tau = \frac{1}{T} \sum_{t=\tau+1}^T (x_t - \bar{x})(x_{t-\tau} - \bar{x})$$

和

$$\hat{\rho}_\tau = \frac{\hat{\gamma}_\tau}{\hat{\gamma}_0},$$

这里 $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ 为样本均值. 在不会发生误解的情况下, 上面两个样本函数的“样本”二字往往省略, 它们的英文缩写也用 acvf 和 acf.

2.1.2 差分算子和后移算子

差分算子

人们总是希望把非平稳时间序列变换成平稳序列, 以便于用数学方法来处理, 而最常用的变换是差分变换.

对于序列 $\{X_1, \dots, X_N\}$, 其一阶 (向后) 差分 (first-order (backward) difference) 定义为 $\{y_2, \dots, y_N\}$, 这里 $y_t = \nabla x_t \equiv x_t - x_{t-1}$ ^①.

二阶差分 (second-order difference) 定义为

$$\begin{aligned} \nabla^2 x_t &= \nabla(\nabla x_t) \\ &= \nabla x_t - \nabla x_{t-1} \end{aligned}$$

^① 一些文献用 Δ 标记差分算子.

$$\begin{aligned}
 &= (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) \\
 &= x_t - 2x_{t-1} + x_{t-2},
 \end{aligned}$$

类似地, p 阶差分为

$$\nabla^p x_t = \nabla^{p-1}(\nabla x_t) = \nabla^{p-1} x_t - \nabla^{p-1} x_{t-1},$$

s 期滞后差分为

$$\nabla_s x_t = x_t - x_{t-s}.$$

显然, $\nabla \nabla_s \equiv \nabla_s \nabla$; 对于常数 c , $\nabla c = 0$.

后移算子

后移 (滞后) 算子 (back-shift (lag) operator) B 定义为 $By_t \equiv y_{t-1}$,^① 显然, $\nabla \equiv 1 - B$. 后移算子有下面的性质:

- $B^s y_t = y_{t-s}$
- $Bc = c$
- $Bcy_t = cy_{t-1} = cBy_t$
- $(a_1 B^i + a_2 B^j)y_t = a_1 B^i y_t + a_2 B^j y_t = a_1 y_{t-i} + a_2 y_{t-j}$
- $B^i B^j y_t = B^{i+j} y_t = y_{t-i-j}$
- $\frac{1}{(1-aB)} y_t = \{1 + aB + a^2 B^2 + \cdots\} y_t = y_t + ay_{t-1} + a^2 y_{t-2} + \cdots$, 假定 $|a| < 1$
- $B^0 = 1$
- $B(cx_t) = cBx_t = cx_{t-1}$
- $B(x_t + y_t) = x_{t-1} + y_{t-1}$
- $(1 - B)^n = \sum_{i=0}^n (-1)^i \binom{n}{i} B^i$
- p 阶差分: $\nabla^p x_t = (1 - B)^p x_t = \sum_{i=0}^p (-1)^i \binom{p}{i} x_{t-i}$
- k 期滞后差分: $\nabla_k x_t = x_t - x_{t-k} = (1 - B^k)x_t$

2.2 白噪声

如果序列 $\{w_t\}$ 的所有观测值都是独立同分布的, 而且其均值 μ 及方差 σ^2 均为有穷的常数 (通常定义均值 $\mu = 0$), 记为 $w_t \sim \text{iid}(\mu, \sigma^2)$, 则称序列 $\{w_t\}$ 为白噪声过程

^① 一些文献用 L 标记后移算子.

(white noise process) 或纯随机过程 (purely random process). 白噪声的定义通常用符号表示为

$$w_t \sim wn(\mu, \sigma^2).$$

如果白噪声的分布是均值为 0 的正态分布, 即 $w_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$, 则 $\{w_t\}$ 也称为**高斯白噪声** (Gaussian white noise). 更一般的白噪声的定义为: 对于任何平稳过程 $\{w_t\}$, 如果其自协方差函数 γ_h 满足条件

$$\gamma_h = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

则可以称该过程为白噪声.^① 这里没有要求常数均值, 但通常都定义白噪声的均值为零, 以体现其为纯粹的“噪声”.

根据上面的定义可以知道白噪声序列 $\{w_t\}$ 的自协方差函数为

$$\gamma_k = \text{Cov}(w_t, w_{t+k}) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k = \pm 1, \pm 2, \dots \end{cases}$$

因而, 自相关函数为

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k = \pm 1, \pm 2, \dots \end{cases}$$

白噪声过程是二阶平稳的, 独立性假定保证了白噪声也是严平稳的.

图2.1为模拟的高斯白噪声序列图 (左) 及其样本自相关函数图 (右).

产生图2.1的代码为:

```
set.seed(10);x=rnorm(150)
par(mfrow=c(1,2))
ts.plot(x);acf(x)
```

白噪声过程可以成为许多复杂过程的基本组件, 比如后面要介绍的移动平均过程. 一些文献在白噪声定义中把独立性减弱成观测值互不相关, 这对于线性正态过程来说是等价的, 但对于非线性过程则需要独立性假定了.

^① 这是因为其谱密度 $f(\nu) = \sum_{h=-\infty}^{\infty} \gamma_h e^{2\pi i \nu h} = \gamma_0 = \sigma^2$, 为常数, 说明所有频率强度相同, 如同白光一样, 故称其为白噪声. 参见后面有关谱分析的章节.

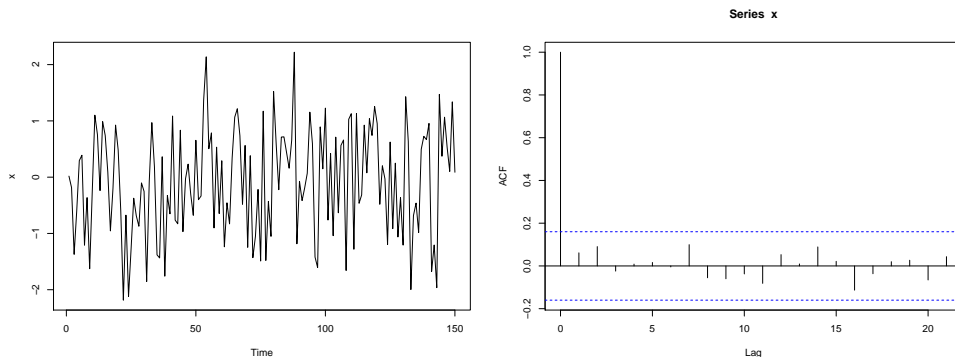


图 2.1 模拟的高斯白噪声序列图 (左) 及其样本自相关函数图 (右)

2.3 随机游走

假定 $\{w_t\}$ 为白噪声, 均值为 0, 方差为 σ^2 , 如果

$$X_t = \mu + X_{t-1} + w_t,$$

则过程 $\{X_t\}$ 称为带有漂移 (drift) μ 的随机游走 (random walk), 亦称随机徘徊. 如果 X_0 固定, 那么有

$$X_t = \mu t + X_0 + \sum_{i=1}^t w_i.$$

因此

$$E(X_t) = \mu t + X_0, \text{Var}(X_t) = t\sigma^2.$$

随 t 而变的均值及方差意味着该过程是非平稳的. 然而, 随机游走的一阶差分

$$\nabla X_t = X_t - X_{t-1} = \mu + w_t$$

是一个纯随机过程, 它是平稳的.

图2.2为模拟的带漂移 ($\mu = 0.005$) 的随机游走 (左下) 及其样本自相关函数图 (右上), 下面左右两图分别为原序列的差分及差分的样本自相关函数图.

产生图2.2的代码为:

```
set.seed(1010)
rwd=cumsum(rnorm(n=100, mean=(1:100)*.005))
```

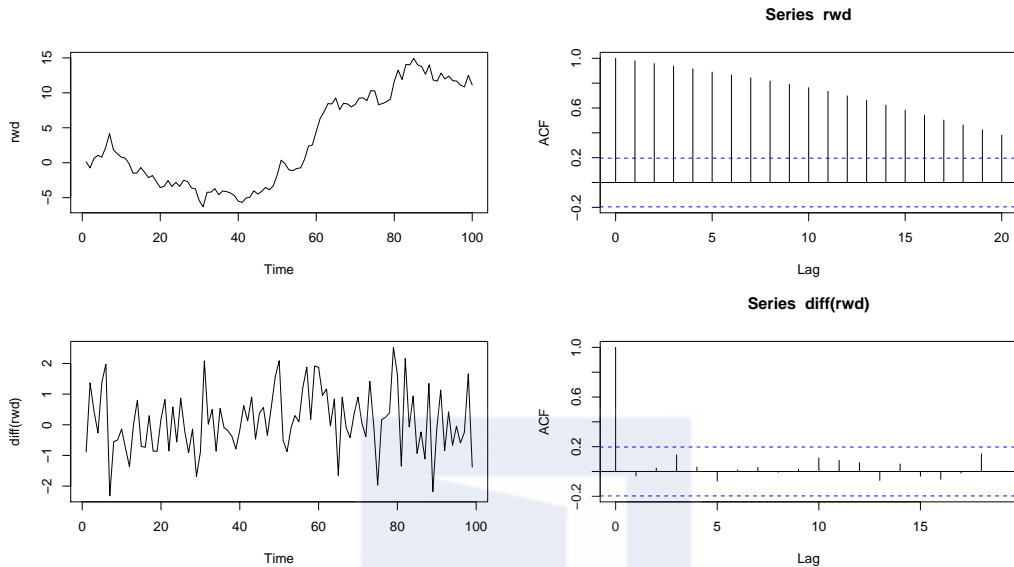



图 2.2 模拟的随机游走 (左上) 及其差分 (左下) 以及它们的样本自相关函数图 (右上和右下)

```
par(mfrow=c(2,2));ts.plot(rwd);acf(rwd)
ts.plot(diff(rwd));acf(diff(rwd))
```

在应用中, 一些人假定股票价格服从随机游走模型, 也就是说, 今天的股票价格等于昨天的股票价格加上一个随机误差. 这种假定在数学上很方便, 可以推导出很漂亮的数学结果, 但在实际生活中可能是非常不合理的, 特别是在金融危机、政治危机或者其他因素影响下, 用随机游走模型来描述股票价格就很荒谬了.

2.4 趋势平稳过程

趋势平稳过程 (trend stationary process) 定义为

$$X_t = f(t) + y_t, \quad (2.1)$$

这里的 $f(t)$ 是 t 的任意实值函数, 而 $\{y_t\}$ 为一个平稳过程. 它可以看成围绕着 $f(t)$ 周围的平稳过程. 前面的带漂移的随机游走是一个最简单的趋势平稳过程, 它可以写成下面形式:

$$X_t = \mu + \beta t + w_t, \quad w_t \sim \text{iid}(0, \sigma^2).$$

式 (2.1) 中的 $f(t)$ 在这里是 $E(X_t) = \mu + \beta t$, 而 $y_t = w_t$, 方差 $\text{Var}(X_t) = \sigma^2$, 因均值随时间而变, 它不是平稳过程. 在取一阶差分后, 可以得到

$$\nabla X_t = \beta + w_t - w_{t-1}.$$

显然, ∇X_t 是一个平稳的 MA(1) 过程 (后面要介绍), 均值为 β , 方差为 $2\sigma^2$.

图2.3为模拟的趋势平稳过程 (左上) 和样本自相关函数图 (右上), 以及其一阶差分序列 (左下) 和样本自相关函数图 (右下).

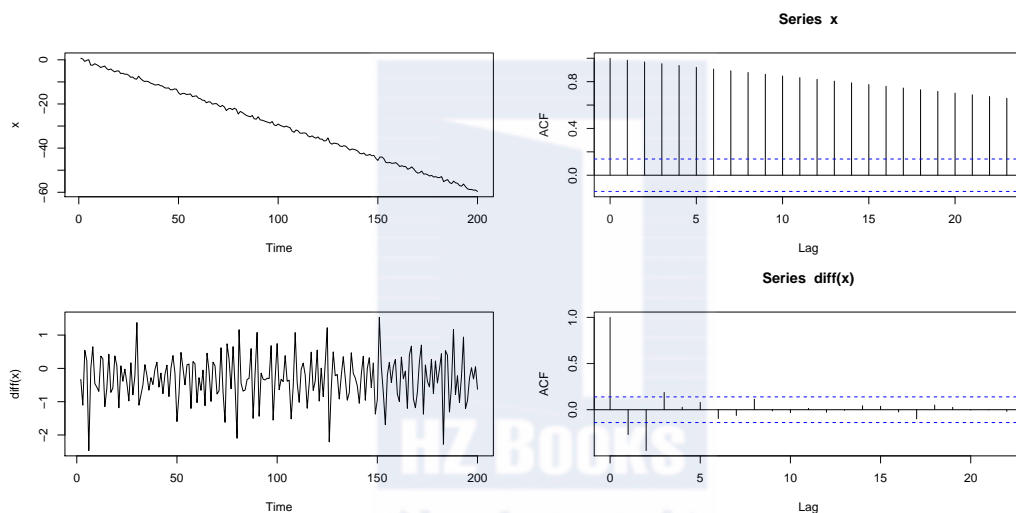


图 2.3 模拟的趋势平稳过程 (左上) 和样本自相关函数图 (右上), 以及其一阶差分序列 (左下) 和样本自相关函数图 (右下)

产生图2.3的代码为:

```
set.seed(1010)
y=arima.sim(n = 200, list(ma = c(0.2, -0.4)),sd = sqrt(0.2))
x=0.5-0.3*(1:200)+y
par(mfrow=c(2,2));ts.plot(x);acf(x)
ts.plot(diff(x));acf(diff(x))
```

这里模拟的序列为 $x_t = 0.5 - 0.3t + y_t$, 其中的 y_t 属于后面要介绍的平稳的移动平均过程 MA(2).

2.5 联合平稳性和互相关函数

考虑两个时间序列 X_t 和 Y_t , 如果它们皆为平稳的, 而且它们的互协方差函数 (cross-covariance function)

$$\gamma_{XY}(h) = \text{Cov}(X_{t+h}, Y_t) = E[(X_{t+h} - \mu_X)(Y_t - \mu_Y)]$$

仅仅是滞后 h 的函数, 则称 X_t 和 Y_t 为联合平稳的 (jointly stationary).

联合平稳时间序列 X_t 和 Y_t 的互相关函数 (cross-correlation function, ccf) 定义为

$$\rho_{XY}(h) = \frac{\gamma_{XY}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}.$$

显然互相关函数有下面性质: $-1 \leq \rho_{XY}(h) \leq 1$, $\rho_{XY}(h) \neq \rho_{XY}(-h)$ 以及 $\rho_{XY}(h) = \rho_{YX}(-h)$.

样本 x_t 和 y_t 的样本互协方差函数 (sample cross-covariance function) 定义为

$$\hat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}),$$

而 $\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{xy}(h)$ 确定负滞后时的函数值. 样本互相关函数 (sample cross-correlation function) 定义为

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}.$$

2.6 一般线性模型

假定 $\{w_t\}$ 是均值为 0 的白噪声序列 (假定其方差为 σ^2), 那么该序列可以通过下面的线性滤波 (linear filter) 变换成序列 $\{X_t\}$:

$$\begin{aligned} X_t &= \mu + \sum_{j=0}^{\infty} \psi_j w_{t-j} \\ &= \mu + (1 + \psi_1 B + \psi_2 B^2 + \cdots) w_t \\ &= \mu + \psi(B) w_t, \end{aligned} \tag{2.2}$$

式(2.2)中, $\psi_0 = 1$, 而算子 $\psi(B) \equiv 1 + \psi_1 B + \psi_2 B^2 + \dots$ 为 $\{w_t\}$ 变换到 $\{X_t\}$ 的算子, 这里的序列 $\{X_t\}$ 是后面会介绍的无穷阶移动平均过程 $MA(\infty)$.

显然, 只有在系数满足 $\sum_{j=0}^{\infty} |\psi_j| < \infty$ 时, 序列 $\{X_t\}$ 才是平稳的, 并且以 μ 作为均值. 由于均值不会改变协方差的性质, 往往假定其为 0, 这时, 一般线性模型为 $X_t = \psi(B)w_t$. 在满足 $\sum_{j=0}^{\infty} |\psi_j| < \infty$ (这时称序列为绝对可加的 (absolutely summable)) 或满足 $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ (这时称序列为二次可加的 (square summable) 或 ℓ_2 可加的)^①时, 式 (2.2) 的序列称为因果的 (causal). 这个因果 (causality) 概念源于目前的观测值是日前和过去的白噪声项的函数. 对于预测来说, 只有因果平稳过程才有意义, 因此在中所提到的平稳都是因果平稳, 也就是依历史的平稳, 而不是非因果 (依未来) 的平稳.

根据 **Wold 分解定理** (Wold's decomposition theorem): 任何 (弱) 平稳时间序列都可以表示成上面的一般线性模型形式, 因此式 (2.2) 也称为 **Wold 形式** (Wold form) 或者 **Wold 表示** (Wold representation). 除了前面提到的均值 $E[X_t] = \mu$ 之外, 我们还有

$$\begin{aligned}\gamma_0 &= \text{Var}(X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2, \\ \gamma_k &= \text{Cov}(X_t, X_{t-k}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \\ \rho_k &= \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}.\end{aligned}$$

由于 $\partial X_{t+j} / \partial w_t = \psi_j, j = 1, 2, \dots$, Wold 形式中的权重 ψ_j 又称为脉冲响应 (impulse response).

式 (2.2) 如同一个回归, 而其所有的解释函数 (即 $\{X_t\}$) 都是不可观测的误差 (即 $\{w_t\}$) 的函数. 所谓的可逆性条件 (invertibility condition) 意味着能够把这些误差表示成目前及以前的观测值的加权和:

$$w_t = \sum_{j=1}^{\infty} \pi_j X_{t-j}.$$

在绝对可加 ($\sum_{j=0}^{\infty} |\pi_j| < \infty$) 条件下, 这样的序列 $\{X_t\}$ 称为可逆的 (invertible), 或者更严格地说 $\{X_t\}$ 是 $\{w_t\}$ 的可逆函数 (invertible function). 从预测的观点, 可逆性表示了现在和过去的误差所提供的信息等价于现在和过去观测值所提供的信息. 或者, 对

^① 注意, 可以证明: $\sum_{j=0}^{\infty} |\psi_j| < \infty \Rightarrow \sum_{j=0}^{\infty} \psi_j^2 < \infty$, 以及 $\sum_{j=0}^{\infty} \psi_j^2 < \infty \Rightarrow \sum_{j=0}^{\infty} |\psi_j| < \infty$.

于任意随机变量 Y , 有

$$E[Y|X_t, X_{t-1}, \dots] = E[Y|w_t, w_{t-1}, \dots].$$

目前这个关系依赖于无穷多个过去的值. 后面我们将继续讨论可逆性是否可能, 什么平稳序列满足可逆性条件, 是否能够用有穷项来表示等问题.

2.7 MA 模型

假定 $w_t \sim wn(0, \sigma_w^2)$, 如果序列 $\{X_t\}$ 满足

$$X_t = \mu + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q},$$

则称其为 q 阶 **MA 过程**, 即 q 阶移动平均过程 (moving average process), 记为 $MA(q)$. 令移动平均算子 $\theta(B) = 1 + \theta_1 B^1 + \theta_2 B^2 + \dots + \theta_q B^q$, 则该模型为

$$X_t = \mu + \theta(B)w_t.$$

显然

$$E(X_t) = \mu, \text{Var}(X_t) = \sigma^2 \left(1 + \sum_{i=1}^q \theta_i^2 \right).$$

考虑 B 为一个复变量, $\theta(B) = 0$ 称为该模型的**特征方程** (characteristic equation). 根据 $\{w_t\}$ 作为白噪声的性质, 可以导出 $MA(q)$ 过程的各种矩为

$$\begin{aligned} E[X_t] &= \mu, \\ \gamma_0 &= \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2), \\ \gamma_j &= \begin{cases} \sigma^2 \left(\theta_j + \sum_{i=1}^{q-j} \theta_i \theta_{i+j} \right), & j = 1, 2, \dots, q \\ 0, & j > q \\ \gamma_{-j}, & j < 0 \end{cases} \end{aligned}$$

由此, 可以得出 $MA(q)$ 的自相关函数 (acf):

$$\gamma_0 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2),$$

$$\rho_j = \begin{cases} \frac{\theta_j + \sum_{i=1}^{q-j} \theta_i \theta_{i+j}}{1 + \theta_1^2 + \dots + \theta_q^2}, & j = 1, 2, \dots, q \\ 0, & j > q \\ \rho_{-j}, & j < 0 \end{cases}$$

由这个表达式可以看出, MA(q) 的自相关函数 ρ_j 在 $j > q$ 时为 0, 因此, 可用 acf 来大致判断 MA 过程的阶数 (见图 2.4).

图 2.4 为模拟的不同参数的 MA(2) 过程的 acf 图, 三组过程的 θ 参数分别为 (0.5, 0.3), (0.5, -0.3), (-0.75, -0.3). 注意三个图在滞后期为 0 的 acf 均为 1.

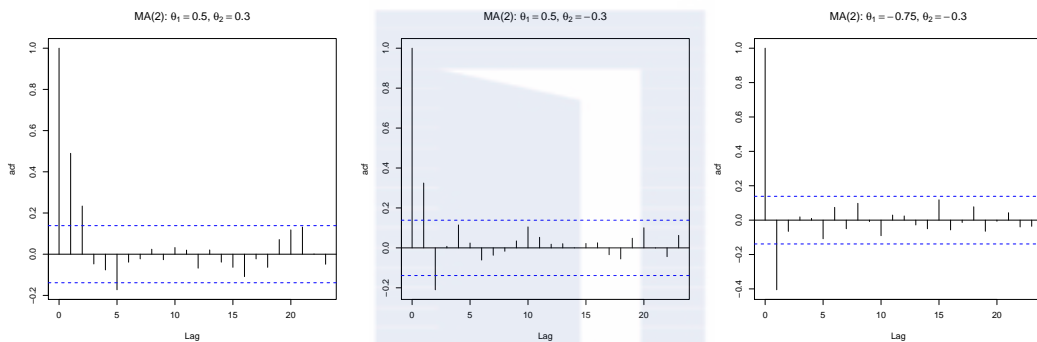


图 2.4 模拟的不同参数的 MA(2) 过程的 acf 图

产生图 2.4 的代码为:

```
set.seed(63010);par(mfrow=c(1,3))
plot(acf(arima.sim(n=200,list(ma=c(.5,.3))),plot=F),ylab="acf",
type="h", main=expression(paste("MA(2): ",
theta[1]==0.5," ", theta[2]==0.3)))
plot(acf(arima.sim(n=200,list(ma=c(.5,-.3))),plot=F),ylab="acf",
type="h", main=expression(paste("MA(2): ",
theta[1]==0.5," ", theta[2]==-0.3)))
plot(acf(arima.sim(n=200,list(ma=c(-.75,-.3))),plot=F),ylab="acf",
type="h", main=expression(paste("MA(2): ",
theta[1]==-0.75," ", theta[2]==-0.3)))
```

显然, 无论那些系数 $\{\theta_i\}$ 是多少, MA(q) 过程总是平稳的, 在 MA(∞) 时需要系数的绝对可加条件.

关于可逆性, 请先看下面例子.

例 2.1 具有同样 acf 的序列不一定都可逆. 假定有两个 MA(1) 过程:

$$A: X_t = w_t + \theta w_{t-1},$$

$$B: X_t = w_t + \theta^{-1} w_{t-1}.$$

容易验证, 它们有完全一样的 acf. 因此不能从给定的 acf 来唯一地确定 MA 过程. 现在试图用 $\{X_t\}$ 来表示 $\{w_t\}$, 得到

$$A: w_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots,$$

$$B: w_t = X_t - \theta^{-1} X_{t-1} + \theta^{-2} X_{t-2} - \cdots.$$

如果 $|\theta| < 1$, 则模型 A 的 X_{t-j} 的系数序列收敛, 而模型 B 的 X_{t-j} 的系数序列不收敛. 因此模型 B 不是可逆的.

对于一般的 MA(q) 来说,

$$X_t - \mu = \theta(B)w_t,$$

即

$$w_t = \theta^{-1}(B)(X_t - \mu).$$

如果

$$\theta(B) = \prod_{i=1}^q (1 - \lambda_i B),$$

做部分分式展开, 得到

$$\pi(B) = \theta^{-1}(B) = \sum_{i=1}^q \frac{m_i}{1 - \lambda_i B}.$$

只有在 $|\lambda_i| < 1$ ($i = 1, \dots, q$), 权重 $\pi_j = -\sum_{i=1}^q m_i \lambda_i^j$ 是绝对可加时, $\pi(B)$ 收敛. 由于特征方程 $\theta(B) = 0$ 的根为 λ_i^{-1} , 所以, 如果 MA(q) 过程的特征方程的根都在单位圆外, 则该过程是可逆的. 实际上, 这里说的可逆性是可以把序列 $\{X_t\}$ 用目前及过去的信息来表示, 也可以说是依历史可逆 (invertible in the past), 而 MA 过程的特征方程的根只要不在单位圆上, 就可以把该过程表示成可逆 (依历史可逆) 或者不可逆 (依未来可逆) 的形式. 为此, 请看下面例子.

例 2.2 MA(1) 过程的两种转换. 考虑 $\mu = 0$ 的 MA(1) 过程

$$X_t = w_t + \theta w_{t-1}.$$

它可以写成 $w_t = X_t - \theta w_{t-1}$, 通过迭代, 得到

$$\begin{aligned} w_t &= X_t - \theta(X_{t-1} - \theta w_{t-2}) = X_t - \theta X_{t-1} + \theta^2 w_{t-2} = \cdots \\ &= X_t - \theta X_{t-1} + \cdots + (-\theta)^p X_{t-p} + (-\theta)^{p+1} w_{t-(p+1)} = \cdots \\ &= X_t + \sum_{j=1}^{\infty} (-\theta)^j X_{t-j} \end{aligned}$$

或者

$$X_t = w_t - \sum_{j=1}^{\infty} (-\theta)^j X_{t-j}. \quad (2.3)$$

完全类似地, 根据 $X_{t+1} = w_{t+1} + \theta w_t$, 或者 $w_t = \theta^{-1} X_{t+1} - \theta^{-1} w_{t+1}$, 类似于 (2.3) 式的推导, 该 MA(1) 过程也可以写成

$$w_t = - \sum_{j=1}^{\infty} (-\theta^{-1})^j X_{t+j} \quad \text{或者} \quad X_t = \theta w_{t-1} - \sum_{j=1}^{\infty} (-\theta^{-1})^j X_{t+j}. \quad (2.4)$$

当 $|\theta| < 1$ 时, (2.3) 中的序列是收敛的, 因此该 MA(1) 过程是 (依历史) 可逆的, 而当 $|\theta| > 1$ 时, (2.4) 中的序列是收敛的, 但却是“依未来可逆”, 它对于预测没有意义, 不符合我们关于可逆性的定义.

2.8 AR 模型

假定 $\{w_t\}$ 是均值为 0、方差为 σ^2 的白噪声序列, 如果均值为 0 的序列 $\{X_t\}$ 满足

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t,$$

则称它为 p 阶 AR 过程, 即 p 阶自回归过程 (autoregressive process) AR(p). 如果 X_t 的均值 μ 不等于 0, 则上式等价于

$$X_t = \alpha + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t,$$

式中的 $\alpha = (1 - \phi_1 - \cdots - \phi_p)\mu$. 令 $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, 则该模型可写成

$$\phi(B)X_t = \alpha + w_t.$$

考虑 B 为一个复变量, 则

$$\phi(B) = 0$$

称为该 AR 模型的特征方程.

根据前面的内容可知, 如果在绝对可加的条件下, AR(p) 有 Wold 表示 (2.2), 则该 AR 过程称为因果的. 可以证明, $\phi(B)X_t = \alpha + w_t$ 有唯一的平稳解的充分必要条件是特征方程的解不在单位圆上. 如果 AR(p) 的特征方程的根都在单位圆外, 则该过程为因果平稳的.

例 2.3 AR 过程的两种转换. 考虑 AR(1) 过程

$$X_t = \phi_1 X_{t-1} + w_t,$$

有

$$\begin{aligned} X_t &= \phi(\phi X_{t-2} + w_{t-1}) + w_t \\ &= \phi^2(\phi X_{t-3} + w_{t-2}) + \phi w_{t-1} + w_t = \cdots \\ &= \sum_{j=0}^{\infty} \phi^j w_{t-j}. \end{aligned} \quad (2.5)$$

当 $|\phi_1| < 1$ 时, (2.5) 的过程收敛, 因此该过程是因果平稳的. 我们还可以实行下面的迭代:

$$X_{t+1} = \phi_1 X_t + w_{t+1} \quad \text{即} \quad X_t = \phi_1^{-1} X_{t+1} - \phi_1^{-1} w_{t+1}.$$

类似地, 有 $X_{t+1} = \phi_1^{-1} X_{t+2} - \phi_1^{-1} w_{t+2}$, 代入上式, 得到

$$X_t = \phi_1^{-2} X_{t+2} - \phi_1^{-2} w_{t+2} - \phi_1^{-1} w_{t+1}.$$

如此下去, 得到

$$X_t = \phi_1^{-k} X_{t+k} - \sum_{j=1}^k \phi_1^{-j} w_{t+j},$$

最终得到

$$X_t = - \sum_{j=1}^{\infty} \phi_1^{-j} w_{t+j}. \quad (2.6)$$

序列 (2.6) 在 $|\phi| > 1$ 时是平稳的, 但是依赖于未知的未来, 因此不是因果平稳的. 所以, 只要 AR 过程的特征方程的根不在单位圆上, 它就是因果平稳的或者是 (依未来) 平稳的, 但后者对于预测没有意义.

AR(p) 过程总是可逆的, 在 AR(∞) 时需要系数的绝对可加条件.

AR(p) 过程的 acf 满足所谓的 **Yule-Walker** 方程组^①:

$$\begin{aligned} \rho_0 &= 1, \\ \rho_j &= \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p}, \end{aligned}$$

从上面第二式, 可以得到

$$\begin{aligned} \rho_1 &= \phi_1 \rho_0 + \phi_2 \rho_1 + \phi_3 \rho_2 + \cdots + \phi_p \rho_{p-1}, \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \rho_0 + \phi_3 \rho_1 + \cdots + \phi_p \rho_{p-2}, \\ &\vdots \\ \rho_{p-1} &= \phi_1 \rho_{p-2} + \phi_2 \rho_{p-3} + \phi_3 \rho_{p-4} + \cdots + \phi_p \rho_1, \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \cdots + \phi_p \rho_0. \end{aligned}$$

令 $\boldsymbol{\rho}_p = (\rho_1, \cdots, \rho_p)^\top$, $\boldsymbol{\phi}_p = (\phi_1, \cdots, \phi_p)^\top$ 以及 (下面矩阵的对角线上 $\rho_0 = 1$)

$$\mathbf{P}_p = \begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & \rho_0 \end{pmatrix}$$

这样, Yule-Walker 方程组的矩阵形式为

$$\boldsymbol{\rho}_p = \mathbf{P}_p \boldsymbol{\phi}_p \text{ 或 } \boldsymbol{\phi}_p = \mathbf{P}_p^{-1} \boldsymbol{\rho}_p,$$

^① 实际上, Yule-Walker 方程对于自协方差函数 $\{\gamma_k\}$ 也适用, 仅仅把 ρ_k 换成 γ_k , 把 $\rho_k = 1$ 换成 $\gamma_0 = \sigma^2$ 即可.

$$\sigma^2 = \rho_0 - \boldsymbol{\rho}^\top \mathbf{P}_p^{-1} \boldsymbol{\rho}_p.$$

Yule-Walker 方程是一组差分方程, 其一般解为

$$\rho_k = A_1 \pi_1^{|k|} + \cdots + A_p \pi_p^{|k|},$$

这里的 $\{\pi_i\}$ 是 Yule-Walker 方程 $\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}$ 的所谓辅助方程 (auxiliary equation)

$$y^p - \phi_1 y^{p-1} - \cdots - \phi_p = 0$$

的根, 而常数 $\{A_i\}$ 满足依赖于 $\rho_0 = 1$ 的初始条件, 这意味着 $\sum A_i = 1$. 前 $p-1$ 个 Yule-Walker 方程利用 $\rho_0 = 1$ 和 $\rho_k = \rho_{-k}$ 给出了对 $\{A_i\}$ 的进一步约束.

显然, 只要对于所有的 i , $|\pi_i| < 1$, 则当 k 趋于无穷时, ρ_k 趋于 0, 这就是过程为 (因果) 平稳的充分必要条件. 这等价于前面提到的特征方程的根落在单位圆之外意味着过程为 (因果) 平稳的条件.

和 AR 模型有关的一个概念是偏自相关函数 (partial auto-correlation function, pacf), 后面会用样本 pacf 条形图来判断 AR 的阶数. pacf 是基于下面一系列 AR 模型定义的:

$$\begin{aligned} x_t - \mu &= \phi_{11} x_{t-1} + w_{1t}, \\ x_t - \mu &= \phi_{21} x_{t-1} + \phi_{22} x_{t-2} + w_{2t}, \\ &\vdots \\ x_t - \mu &= \phi_{p1} x_{t-1} + \phi_{p2} x_{t-2} + \cdots + \phi_{pp} x_{t-p} + w_{pt}, \end{aligned}$$

这里的系数 $\{\phi_{jj}\}$, 即每个 AR 模型的最后一个系数, 称为偏自相关函数. 样本偏自相关函数 $\hat{\phi}_{jj}$ 是根据上面 p 个 AR 模型用最小二乘法解出来的. 也可以用上面的 Yule-Walker 方程组来计算 pacf, 利用各阶的样本 acf 矩阵 $\hat{\mathbf{P}}_i$ 及样本 acf $\hat{\rho}_i$ ($1 \leq i \leq p$), 得到

$$\hat{\boldsymbol{\phi}}^{(i)} = \hat{\mathbf{P}}_i^{-1} \hat{\boldsymbol{\rho}}_i = (\hat{\phi}_1^{(i)}, \dots, \hat{\phi}_i^{(i)})^\top$$

只留下向量 $\boldsymbol{\phi}^{(i)}$ 的最后一项 $\hat{\phi}_i^{(i)}$, 这就是 pacf 的第 i 个值 $\text{pacf}(i)$.

注意, 在正态假定下, 偏自相关函数为

$$\phi_k = \text{Cov}(X_t, X_{t+k} | X_{t+1}, X_{t+2}, \dots, X_{t+k-1}).$$

可以表明, 对于 $\text{AR}(p)$ 过程, 当 $j > p$ 时, 偏自相关系数 $\phi_{jj} = 0$, 因此可以用 `pacf` 来大致判断 AR 模型的阶数.

图 2.5 为模拟的不同参数的 $\text{AR}(2)$ 过程的 `pacf` 图, 三组过程的 ϕ 参数分别为 $(0.5, 0.3)$, $(0.5, -0.5)$, $(-0.75, -0.3)$.

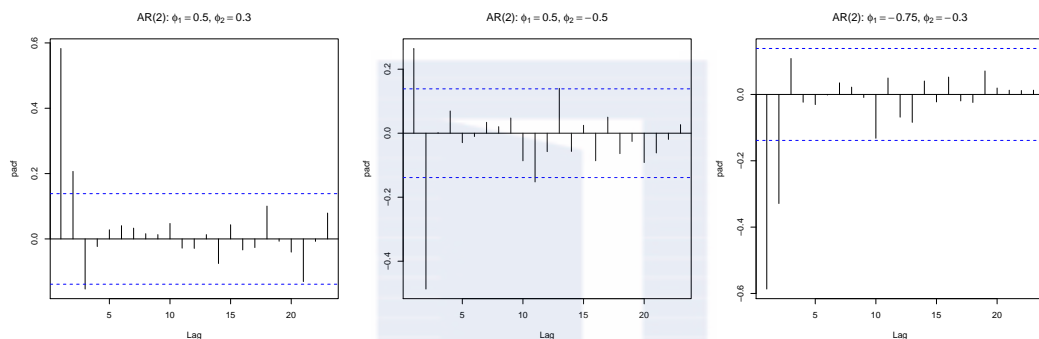


图 2.5 模拟的不同参数的 $\text{AR}(2)$ 过程的 `pacf` 图

产生图 2.5 的代码为:

```
set.seed(6301);par(mfrow=c(1,3))
plot(pacf(arima.sim(n=200,list(ar=c(.5,.3))),plot=F),ylab="pacf",
type="h",main=expression(paste("AR(2): ",
phi[1]==0.5," ", phi[2]==0.3)))
plot(pacf(arima.sim(n=200,list(ar=c(.5,-.5))),plot=F),ylab="pacf",
type="h", main=expression(paste("AR(2): ",
phi[1]==0.5," ", phi[2]==-0.5)))
plot(pacf(arima.sim(n=200,list(ar=c(-.75,-.3))),plot=F),ylab="pacf",
type="h", main=expression(paste("AR(2): ",
phi[1]==-0.75," ", phi[2]==-0.3)))
```

例 2.4 判断一个 $\text{AR}(2)$ 过程的平稳性. 考虑 $\text{AR}(2)$ 过程

$$X_t = X_{t-1} - 0.5X_{t-2} + w_t$$

是否 (因果) 平稳的问题. 其特征方程为

$$\phi(B) = 1 - B + 0.5B^2 = 0.$$

这个特征方程可以用 R 代码 `polyroot(c(1, -1, .5))` 来解, 并用函数 `Mod()` 求其模. 得到它的两个根为 $1 \pm i$, 两个根的模均为 $\sqrt{2} = 1.414\ 214$, 显然在单位圆外, 因此过程是 (因果) 平稳的. 而 Yule-Walker 方程

$$\rho_k = \rho_{k-1} - 0.5\rho_{k-2} \quad \text{或} \quad \rho_k - \rho_{k-1} + 0.5\rho_{k-2} = 0$$

的辅助方程为

$$y^2 - y + 0.5 = 0.$$

其解为 $0.5 \pm 0.5i$, 而模为 $1/\sqrt{2} = 0.707\ 106\ 8$, 在单位圆内, 这也从另一方面验证了该过程为 (因果) 平稳的.

2.9 ARMA 模型

如果 X_t 的期望为 $\mu = 0$, 则自回归移动平均过程 $\text{ARMA}(p, q)$ 定义为

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

或者 $\phi(B)X_t = \theta(B)w_t$, 这里算子

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$$

$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

如果 X_t 的期望 μ 不等于 0, 则上面的定义可写为

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

这里 $\phi_0 = (1 - \phi_1 - \cdots - \phi_p)\mu$.

使得过程 (因果) 平稳的 $\{\phi_i\}$ 值满足 $\phi(B) = 0$ 的根在单位圆外, 而使得过程可逆的 $\{\theta_i\}$ 值满足 $\theta(B) = 0$ 的根在单位圆外. ARMA 比纯粹的 AR 或 MA 有较少的参数.

令算子

$$\psi(B) = \theta(B)/\phi(B) = \sum_{i \geq 1} \psi_i B^i,$$

则 ARMA 可写成纯 MA 过程 $X_t = \psi(B)w_t$, 即

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

令算子

$$\pi(B) = \phi(B)/\theta(B) = 1 + \sum_{i \geq 1} \pi_i B^i,$$

则 ARMA 可写成纯 AR 过程 $\pi(B)X_t = w_t$, 即

$$\sum_{j=0}^{\infty} \pi_j X_{t-j} = w_t.$$

显然, $\pi(B)\psi(B) = 1$, $\pi(B)\theta(B) = \phi(B)$, $\psi(B)\phi(B) = \theta(B)$.

平稳的 ARMA(p, q) 过程的均值为

$$E[X_t] = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p},$$

根据 $\phi(B)X_t = \theta(B)w_t$, $E[\phi(B)X_t] = \theta(B)w_t$. 为了得到 ARMA 模型 $\phi(B)X_t = \theta(B)w_t$ 的关于 $\gamma(h)$ 的方程, 根据

$$E[(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})X_{t-h}] = E[(w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q})X_{t-h}],$$

则有

$$\begin{aligned} & \gamma(h) - \phi_1 \gamma(h-1) - \cdots - \phi_p \gamma(h-p) \\ &= E(\theta_h w_{t-h} X_{t-h} + \cdots + \theta_q w_{t-q} X_{t-h}) \\ &= \sigma_w^2 \sum_{j=0}^{q-h} \theta_{h+j} \psi_j, \end{aligned}$$

这里 $\theta_0 = 1$. 但具体的参数 ψ_j 如何求呢? 根据 $\psi(B)\phi(B) = \theta(B)$, 有

$$\begin{aligned} 1 + \theta_1 B + \cdots + \theta_q B^q &= (\psi_0 + \psi_1 B + \cdots)(1 - \phi_1 B - \cdots - \phi_p B^p), \\ \Leftrightarrow \\ 1 &= \psi_0, \\ \theta_1 &= \psi_1 - \phi_1 \psi_0, \\ \theta_2 &= \psi_2 - \phi_1 \psi_1 - \cdots - \phi_2 \psi_0, \\ &\vdots \end{aligned}$$

这等价于 $\theta_j = \phi(\psi_j)$, 而 $\theta_0 = 1$, 对于 $j < 0$, $j > q$, $\theta_j = 0$.

当然, 也可以直接考虑形式 $X_t = \psi(B)w_t$, 有

$$\gamma(h) = E(X_t X_{t-h}) = \sigma_w^2 (\psi_0 \psi_h + \psi_1 \psi_{h+1} + \psi_2 \psi_{h+2} + \cdots).$$

可以表明: 对于定义为 $\phi(B)X_t = \theta(B)w_t$ 的 ARMA 过程 $\{X_t\}$, 如果多项式 $\theta(z)$ 没有根在单位圆上, 则存在多项式 $\tilde{\phi}$ 和 $\tilde{\theta}$, 以及白噪声序列 $\{\tilde{w}_t\}$, 使得 $\{X_t\}$ 满足 $\tilde{\phi}(B)X_t = \tilde{\theta}(B)\tilde{w}_t$, 而这是一个 (因果) 平稳及可逆的 ARMA 过程. 因此, 对于很一般的 ARMA 过程, 总是可以转换成 (因果) 平稳可逆的 ARMA 过程. 在一些文献和教科书中, 只引入平稳而不专门引入因果平稳这个概念. 在实际应用中, 人们也仅仅考虑因果序列. 本书除了在 2.6 节、2.7 节及 2.8 节中涉及 (因果) 平稳性和 (非因果) 平稳之外, 在其余部分仅仅考虑因果平稳可逆过程, 而且用“平稳”来代替“因果平稳”术语.

根据 acf 和 pacf 图对 ARMA 模型阶数的直观判断

根据平稳序列的性质, 当实际的时间序列均值和方差大致不变时, 有可能用平稳序列来近似, 即可以试试 ARMA 模型来拟合数据. 而如果实际时间序列在一些差分变换之后均值及方差大致不变, 则可以试试 ARIMA 模型, 或者在这些差分后试试 ARMA 模型.

但是, 即使一个序列可以用 ARMA 模型, 如何选择它的阶数也是个问题. 一种直观方法就是用序列的样本 acf 和样本 pacf 条形图来判断其阶数, 虽然比较粗糙, 但总比没有强. 我们通常所说的 acf 图形实际上是样本 acf 图形, 而 pacf 是用最小二乘法估计出来的样本偏自相关函数 $\{\hat{\phi}_{jj}\}$. 后面凡是涉及图形, 均省去“样本”二字, 简称为

偏自相关函数 (pacf).

对于纯随机过程 (ARMA(0,0)), 由于观测值之间独立, acf 和 pacf 的值很小, 它们的条形图没有什么突出的模式. 一个拟合得很好的时间序列的残差的 acf 和 pacf 条形图就应该如同纯随机过程一样, 数值很小.

而对于一般的 ARMA(p, q) 序列阶数的确定, 则有表2.1中介绍的 acf 和 pacf 条形图直观判断法.

表 2.1 如何用 pacf 及 acf 条形图的拖尾和截尾判断 ARMA 模型

| 模型 | AR(p) | MA(q) | ARMA(p, q) |
|---------|-------------|-------------|-------------------|
| pacf 图形 | 第 p 个条后截尾 | 拖尾 | 头 p 个条无规律, 其后拖尾 |
| acf 图形 | 拖尾 | 第 q 个条后截尾 | 头 q 个条无规律, 其后拖尾 |

所谓“拖尾”就是条形图以指数形式或周期形式衰减, 所谓“截尾”就是图形在若干期之后变得很小而且没有什么模式. 如果 acf 和 pacf 的条形图中均没有截尾, 而且至少有一个图没有显示以指数形式或正弦形式衰减, 那么说明该序列不是平稳序列. 上述的图形判别法不太准确, 只能做个参考. 即使是根据软件模拟出来的 ARMA(p, q) 序列的 acf 和 pacf 条形图, 也不一定完全满足表2.1的规律, 但不会差很多.

图2.6展示了 MA(2)、AR(2)、ARMA(2,2) 理论上精确的 pacf 和 acf 图 (上面三个图), 以及模拟的同样参数、同样过程的样本 pacf 图和样本 acf 图 (下面三个图). 从这些图可以大致看出如何利用表2.1的规律来判断 ARMA 的阶数.

产生图2.6的代码为:

```
par(mfcol=c(2,3));set.seed(1010)
plot(ARMAacf(ma=c(.5,-.4),lag.max = 18,pacf=F),type="h",ylab="ACF")
title("Exact ACF for MA(0.5,-0.4)");abline(h=0)
acf(arima.sim(n = 63, list(ma = c(.5, -.4)), sd = sqrt(0.2)),
main="ACF of Simulated MA(0.5,-0.4)")

plot(ARMAacf(ar=c(-.5, .4),lag.max = 18,pacf=T),type="h",ylab="PACF")
title("Exact PACF for AR(-0.5,0.4)");abline(h=0)
pacf(arima.sim(n = 63, list(ar = c(-.5, .4)), sd = sqrt(0.2)),
main="PACF of Simulated AR(-0.5,0.4)")

plot(ARMAacf(ar=c(-.3, .4),ma=c(-0.3,0.25),lag.max = 18,pacf=F),type="h",
ylab="ACF")
title("Exact ACF for ARMA(ar=(-0.3,0.4),ma=(-0.3,0.25))");abline(h=0)
acf(arima.sim(n = 63, list(ar = c(-.3, .4),ma=c(-.3,.25)), sd = sqrt(0.2)),
```

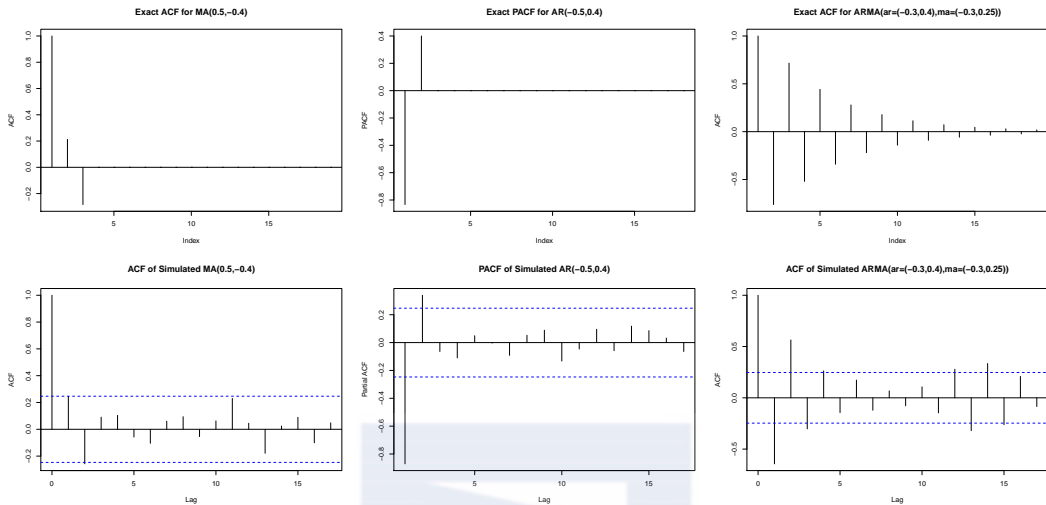


图 2.6 理论的 MA(2)、AR(2)、ARMA(2,2) 及模拟的同样过程的 pacf 图和 acf 图

```
main="ACF of Simulated ARMA(ar=(-0.3,0.4),ma=(-0.3,0.25))"
```

例 2.5 一个 ARMA(1,1) 过程转换成纯 MA 过程和纯 AR 过程. 考虑 ARMA(1,1) 过程

$$X_t = -0.4X_{t-1} + w_t + 0.2w_{t-1}.$$

显然 $\phi(B) = (1 + 0.4B)$, $\theta(B) = (1 + 0.2B)$. 它们的根都在单位圆外, 因此过程是平稳及可逆的. 于是

$$\begin{aligned} \psi(B) &= \theta(B)/\phi(B) \\ &= (1 + 0.2B)/(1 + 0.4B) \\ &= (1 + 0.2B)(1 + (-0.4)B + (-0.4)^2B^2 + (-0.4)^3B^3 + \dots) \\ &= 1 + (0.2 - 0.4)B + (0.2 - 0.4) \times (-0.4)B^2 + (0.2 - 0.4) \times (-0.4)^2B^3 + \dots \end{aligned}$$

因此, 对于 $i = 1, 2, \dots$, 有

$$\psi_i = (0.2 - 0.4) \times (-0.4)^{i-1} = -0.2 \times (-0.4)^{i-1}.$$

类似地, 对于 $i = 1, 2, \dots$, $\pi(B) = \phi(B)/\theta(B)$ 中的系数

$$\pi_i = (0.4 - 0.2) \times (-0.2)^{i-1} = 0.2 \times (-0.2)^{i-1}.$$

显然 $\{\psi_i\}$ 和 $\{\pi_i\}$ 在 i 增加时很快趋于 0.

例 2.6 ARMA(1,1) 过程 ($\phi_1 = 0.9, \theta_1 = 0.5$) 转换成 $MA(\infty)$ 过程的计算. 这可以用 R 计算. 下面是计算代码及 50 个参数 ψ_i 的输出:

```
> ARMAtoMA(ar=.9, ma=.5, 50) # for a list
[1] 1.400000000 1.260000000 1.134000000 1.020600000 0.918540000
[6] 0.826686000 0.744017400 0.669615660 0.602654094 0.542388685
[11] 0.488149816 0.439334835 0.395401351 0.355861216 0.320275094
[16] 0.288247585 0.259422826 0.233480544 0.210132489 0.189119240
[21] 0.170207316 0.153186585 0.137867926 0.124081134 0.111673020
[26] 0.100505718 0.090455146 0.081409632 0.073268669 0.065941802
[31] 0.059347622 0.053412859 0.048071573 0.043264416 0.038937975
[36] 0.035044177 0.031539759 0.028385783 0.025547205 0.022992485
[41] 0.020693236 0.018623913 0.016761521 0.015085369 0.013576832
[46] 0.012219149 0.010997234 0.009897511 0.008907760 0.008016984
> plot(ARMAtoMA(ar=.9, ma=.5, 50),xlib="i",ylab=expression(psi[i]))
> title("ARMA to MA");abline(h=0,lty=2)
```

图 2.7 为这 50 个参数 ψ_i 的点图. 从图上看, 这些参数逐渐地趋于零. 这里所选择的 $\phi_1 = 0.9$ 使得收敛较慢 (当 $\phi_1 = 1$ 时就不平稳了); 如果取 $\phi_1 = 0.2$, 则第 15 个参数就已经很小了: $\psi_{15} = 1.14688 \times 10^{-10}$.

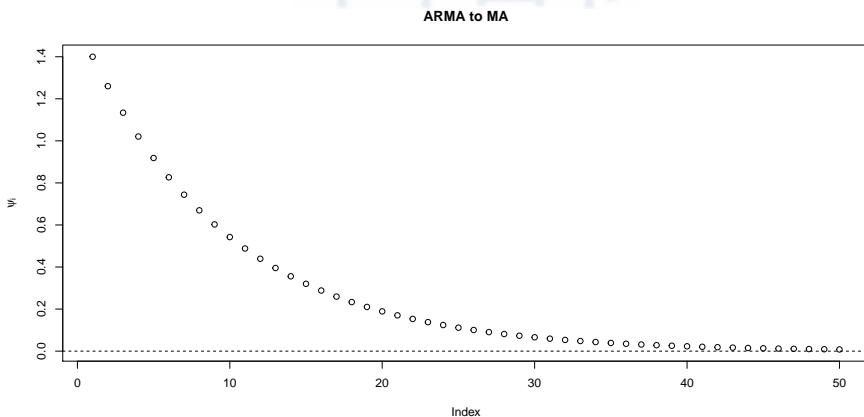


图 2.7 ARAM(0.9,0.5) 过程转换成 $MA(\infty)$ 过程的前 50 个参数 ψ_i

2.10 ARIMA 模型

如果 $\{X_t\}$ 是不平稳的, 而

$$W_t = \nabla^d X_t = (1 - B)^d X_t$$

是平稳的 ARMA(p, q) 过程, 即

$$W_t = \phi_1 W_{t-1} + \cdots + \phi_p W_{t-p} + w_t + \cdots + \theta_q w_{t-q},$$

或者

$$\phi(B)W_t = \theta(B)w_t,$$

则称 $\{X_t\}$ 为整合的 ARMA 模型 ARIMA(p, d, q), 显然,

$$\phi(B)(1 - B)^d X_t = \theta(B)w_t.$$

说明: $\phi(B)(1 - B)^d$ 有 d 个根在单位圆上 (如 $B = 1$), 意味着 $\{X_t\}$ 为非平稳的, 在单位圆上的根叫单位根 (unit root). 作为特例, 随机游走 (随机徘徊) $X_t = X_{t-1} + w_t$, 可以写成 $(1 - B)X_t = w_t$, 为 ARIMA(0,1,0).

如果时间序列 X_t 的 d 阶差分 $\nabla^d X_t$ 是平稳过程, 则称 X_t 为 d 阶单整的, 记为 $I(d)$, 也称序列 X_t 有 d 个单位根. 随机游走有一个单位根.

图2.8为模拟的不同参数的 ARMA(1, 1) 及 ARIMA(1, 1, 1) 序列.

产生图2.8的代码为:

```
set.seed(6301);
par(mfrow=c(2,2))
plot(arima.sim(n=200,list(ar=.4,ma=.5)),ylab="x",type="o",pch=16)
title(expression(paste("ARMA(1,1): ",phi=="0.4,", " ", theta=="0.5")))
plot(arima.sim(n=200,list(ar=-.4,ma=-.5)),ylab="x",type="o",pch=16)
title(expression(paste("ARMA(1,1): ",phi=="-0.4,", " ", theta=="-0.5")))
plot(arima.sim(n=200,list(order=c(1,1,1),ar=.4,d=1,ma=.5)),
ylab="x",type="o",pch=16)
title(expression(paste("ARIMA(1,1,1): ",phi=="0.4,", " ", theta=="0.5")))
plot(arima.sim(n=200,list(order=c(1,1,1),ar=-.4,d=1,ma=-.5)),
ylab="x",type="o",pch=16)
title(expression(paste("ARIMA(1,1,1): ",phi=="-0.4,", " ", theta=="-0.5")))
```

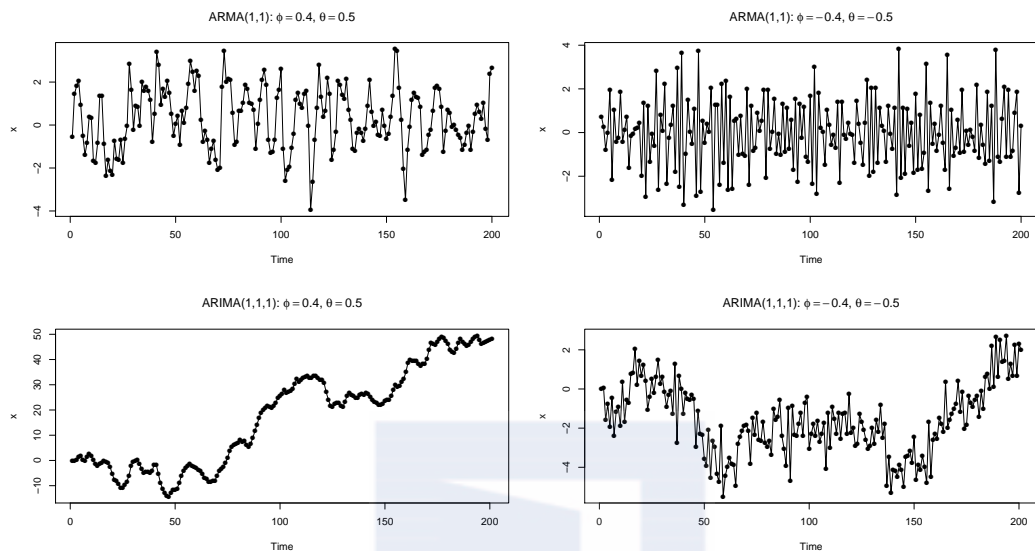


图 2.8 模拟的不同参数的 ARMA(1,1) 及 ARIMA(1,1,1) 序列

2.11 季节模型

为了符号简洁, 本节使用下列算子记号:

$$\begin{aligned}\phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \\ \Phi_P(B) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}, \\ \theta_q(B) &= 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q, \\ \Theta_Q(B) &= 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}.\end{aligned}$$

具有非平稳 (通常的) 阶数 p, d, q , 季节阶数 P, D, Q 及周期 s 的 **Box-Jenkins** 季节 ARIMA 模型 (可乘季节 ARIMA 模型) 记为

$$\text{ARIMA} \underbrace{(p, d, q)}_{\text{非季节}} \times \underbrace{(P, D, Q)}_s$$

其差分序列 $W_t = \nabla^d \nabla_s^D X_t$ 满足具有周期 s 的 $\text{ARMA}(p, q) \times (P, Q)_s$ 模型.

例如 ARIMA(1, 1, 1) × (1, 1, 1)₄ 模型为

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)X_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)w_t.$$

一个一般的可乘季节 ARIMA 模型为

$$\phi_p(B)\Phi_P(B)W_t = \theta_q(B)\Theta_Q(B)w_t.$$

季节 MA(Q) 模型为

$$X_t = w_t + \Theta_1 w_{t-s} + \Theta_2 w_{t-2s} + \cdots + \Theta_Q w_{t-Qs}$$

或

$$X_t = \Theta_Q(B)w_t.$$

季节 AR(P) 模型为

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

或 $\Phi_P(B)X_t = w_t$.

2.12 习题

1. 求下面 X_t 的差分.

(1) 如果 $X_t = \beta_0 + \beta_1 t + Y_t$, 求 ∇X_t .

(2) 如果 $X_t = \sum_{i=0}^k \beta_i t^i + Y_t$, 求 $\nabla^k X_t$.

(3) 如果 $X_t = T_t + S_t + Y_t$, 而 $S_t = S_{t-s}$, 求 $\nabla_s X_t$.

2. 如果 $\{w_t\}$ 是均值为 0、方差为 σ^2 的独立同分布的白噪声, 那么 $X_t = \sum_{i=1}^t w_t$ 为白噪声, 请表明 $\gamma(t+h, t) = t\sigma^2$. 序列 $\{X_t\}$ 是平稳的吗? 模拟并点出序列 $\{X_t\}$ 的图.

3. 对于 MA(1) 过程 $X_t = w_t + \theta w_{t-1}$, 这里 w_t 为白噪声, 表明

$$\gamma(t+h, t) = \begin{cases} \sigma^2(1 + \theta^2), & h = 0 \\ \sigma^2\theta, & h = \pm 1 \\ 0, & \text{其他情况} \end{cases}$$

序列 $\{X_t\}$ 是平稳的吗?

4. 对于 AR(1) 过程 $X_t = \phi X_{t-1} + w_t$, 这里 w_t 为白噪声, 表明: 如果序列 $\{X_t\}$ 是平稳的, 而且 $|\phi| < 1$, 则 $E[X_t] = 0$ 以及

$$E[X_t^2] = \frac{\sigma^2}{1 - \phi^2},$$

$$\gamma_X(h) = \frac{\phi^{|h|} \sigma^2}{1 - \phi^2}.$$

5. 表明: 对于一般线性过程

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

$$\mu_X = E[X_t] = \mu,$$

$$\gamma_X(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{h+j}.$$

6. 对于一般线性过程

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

如果 $E[X_t] = \mu$, 而且在 $j = 0$ 时 $\psi_j = 1$, 在 $j \neq 0$ 时 $\psi_j = 0$, 表明序列 $\{X_t\}$ 是均值为 μ 、方差为 σ_w^2 的白噪声.

7. 假定 $\{Y_t\}$ 为平稳过程, 表明: 对于固定的 n 及任何常数 c_1, \dots, c_n ,

$$X_t = \sum_{j=1}^n c_j Y_{t-j+1}$$

也是平稳的.

8. 假定 $\{Y_t\}$ 为平稳过程, 有自协方差函数 $\gamma_Y(h)$, 表明 ∇Y_t 是平稳的, 并求其均值和自协方差函数; 再表明 $\nabla^2 Y_t$ 也是平稳的.
9. 对于一般线性过程

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

如果 $E[X_t] = 0$, 而且

$$\psi_j = \begin{cases} 1, & j = 0 \\ \theta, & j = 1 \\ 0, & \text{其他情况} \end{cases}$$

表明序列 $\{X_t\}$ 是 MA(1) 过程 $X_t = w_t + \theta w_{t-1}$.

10. 对于一般线性过程

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

如果 $E[X_t] = 0$, 而且

$$\psi_j = \begin{cases} \phi_j, & j \geq 0 \\ 0, & \text{其他情况} \end{cases}$$

表明: 在 $|\phi| < 1$ 时, 序列 $\{X_t\}$ 是 AR(1) 过程 $X_t = \phi X_{t-1} + w_t$.

11. 随机变量 Y 的最好的最小二乘估计为 $E[Y]$, 这是因为 $\min_c (Y - c)^2 = E(Y - E(Y))^2$. 请表明: 给了 X 后, 随机变量 Y 的最好的最小二乘估计为 $E[Y|X]$. 类似地, 表明:

给定 X_n, X_{n+h} 的最好的最小二乘估计为 $f(X_n) = E[X_{n+h}|X_n]$.

12. 对于样本自协方差函数

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n.$$

表明 $\hat{\gamma}(h) = \hat{\gamma}(-h)$ 以及 $\hat{\gamma}$ 为半正定的, 因此有 $\hat{\gamma}(0) \geq 0$ 及 $|\hat{\gamma}(h)| \leq \hat{\gamma}(0)$.

13. 对于样本均值

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n),$$

表明

$$\text{Var}(\bar{X}_n) = \frac{1}{n} \sum_{h=-n}^n \left(1 - \frac{|h|}{n}\right) \gamma(h).$$

14. 对于 MA 过程

$$X_t = \sum_{k=0}^m w_{t-k}/(m+1),$$

表明该过程的 acf 为

$$\rho_k = \begin{cases} (m+1-k)/(m+1), & k=0, 1, \dots, m \\ 0, & k > m \end{cases}$$

15. 考虑 ARMA(1,1) 过程

$$X_t = \phi X_{t-1} + w_t + \theta w_{t-1},$$

这里 $|\phi| < 1, |\theta| < 1$. 表明该过程的 acf 为

$$\begin{aligned} \rho_1 &= (1 + \phi\theta)(\phi + \theta)/(1 + \theta^2 + 2\phi\theta), \\ \rho_k &= \phi\theta(k-1). \end{aligned}$$

16. 考虑过程 $(1-B)(1-0.1B)X_t = (1-0.2B)w_t$.

- (1) 核对该过程为 ARMA(p, d, q) 过程, 找出 p, d, q .
- (2) 核对该过程的平稳性和可逆性.
- (3) 找出把该过程表示成 MA(∞) 模型后的前三个 ψ_i .
- (4) 找出把该过程表示成 AR(∞) 模型后的前三个 π_i .
- (5) 解释上面 (3) 和 (4) 的结果可能说明的意义.

17. 考虑 MA(∞) 过程

$$X_t = w_t + C(w_{t-1} + w_{t-2} + \dots),$$

这里 C 为常数, 表明该过程不是平稳的, 并表明差分 ∇X_t 为平稳的 MA(1) 过程. 求出 ∇X_t 的 acf.

18. 对于 AR(1) 过程

$$X_t = \phi X_{t-1} + w_t,$$

当 $|\phi| < 1$ 时, 表明该过程可以表示成一般线性过程

$$X_t = \sum_{j=1}^{\infty} w_{t-j}.$$

19. 考虑 ARMA(1,1) 过程

$$(1-2.7B)X_t = (1-0.1B)w_t.$$

表明该过程是平稳的及可逆的,但不是因果的.

20. 考虑 ARMA(2,1) 过程

$$(1 + 0.7B^2)X_t = (1 + 2.5B)w_t.$$

表明该过程是平稳的及因果的,但不是可逆的.

21. 表明 AR(2) 过程

$$X_t = X_{t-1} + cX_{t-2} + w_t$$

在 $-1 < c < 0$ 时是平稳的. 求出当 $c = -3/16$ 时的 acf.

22. 表明 AR(3) 过程

$$X_t = X_{t-1} + cX_{t-2} - cX_{t-3} + w_t$$

对于任何 c 都不是平稳的.

