时间序列分析（TIME SERIES ANALYSIS）

主讲：吴尚
复旦大学管理学院统计与数据科学系

平稳时间序列模型

- 时间序列的平稳性
- MA(q)模型
- AR(p)模型
- ARMA(p, q)模型

一些基本概念

- 设$\{Y_t\}$是任意时间序列（随机变量序列）
- 均值$\mu_t = E(Y_t)$
- 方差$\mathrm{Var}(Y_t) = E(Y_t - \mu_t)^2$
- 自协方差
  $\gamma_{t,s} = \mathrm{Cov}(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)] = E(Y_tY_s) - \mu_t\mu_s$
- **自相关函数（ACF）**

$$\rho_{t,s} = \mathrm{Corr}(Y_t, Y_s) = \frac{\mathrm{Cov}(Y_t, Y_s)}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$$

时间序列的平稳性

- 弱平稳（宽平稳）——二阶矩平稳
- 满足如下条件的序列称为（弱）平稳序列
  - 二阶矩有限 $EY_t^2 < \infty$
  - 均值为常数 $EY_t = \mu$
  - 自协方差函数只取决于两个时刻的时间之差 k，而与时间的起始点无关

$$\gamma_k := \mathrm{Cov}(Y_t, Y_{t-k}) = \gamma_{t,t-k} = \gamma_{k,0}$$

- 根据定义，$\gamma_{-k} = \gamma_{t,t+k} = \gamma_{t+k,t} = \gamma_k$
- 自相关函数（ACF）也有类似性质，且 $\rho_k = \frac{\gamma_k}{\gamma_0}$

白噪声

- 定义：均值为0，方差有限的独立同分布的随机变量序列。
- 有时候也用不相关代替独立性这个条件，但目前我们先用独立白噪声的定义，尽管更多时候我们只用到它的不相关性。
- 一般用$\{e_t\}$或$\{\varepsilon_t\}$来表示白噪声。
- $E(e_t) = 0$
- 定义$\sigma_e^2 \coloneqq \gamma_0 = Var(e_t)$
- 对于$k \neq 0, \gamma_k = 0$

MA(q)模型的定义

- 具有如下结构的模型称为q阶滑动平均模型，简记为$MA(q)$

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} = \Theta(B)\, e_t$$

- 其中， $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$，$\theta_q \neq 0$

常用MA模型的自相关系数

- MA(1)模型
- $\rho_k = \begin{cases} 1 & k=0 \\ \frac{-\theta_1}{1+\theta_1^2} & k=1 \\ 0 & k\geq 2 \end{cases}$

- MA(2)模型
- $\rho_k = \begin{cases} 1 & k=0 \\ \frac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & k=1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & k=2 \\ 0 & k\geq 3 \end{cases}$

AR(p)模型的定义

- 具有如下结构的模型称为 p 阶自回归模型，简记为 $AR(p)$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

即 $\Phi(B)Y_t = e_t$

- 其中，$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \phi_p \neq 0$

AR(1)序列平稳的条件

$$Y_t = \phi Y_{t-1} + e_t$$
$$Y_t = \phi(\phi Y_{t-2} + e_{t-1}) + e_t$$

$$\vdots$$

$$\Downarrow$$

$$Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \phi^3 e_{t-3} + \cdots$$

是否平稳?

- 方差有限?
- 均值为常数?
- 自协方差只与 $t-s$ 有关?

一般线性过程

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots = \sum_{k=0}^{\infty} \psi_k e_{t-k}$$

其中， $\psi_0 = 1.$

结论： 当 $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ 时， Y_t有定义，且$\{Y_t\}$平稳。

$$\mathrm{Var}(Y_t) = \sigma_e^2 \sum_{k=0}^{\infty} \psi_k^2$$

例： 当$\psi_k = \frac{1}{k}(k > 0)$时， $\{Y_t\}$平稳。

证： $\sum_{k=0}^{n} \psi_k^2 = 1 + 1 + \sum_{k=2}^{n} \frac{1}{k^2} < 2 + \sum_{k=2}^{n} \frac{1}{k(k-1)} = 3 - \frac{1}{n} < 3$

所以$\sum_{k=0}^{\infty} \psi_k^2 < \infty$，此时$Y_t$有定义且$\{Y_t\}$平稳。

AR(1)序列平稳的条件

- $Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \phi^3 e_{t-3} + \cdots$
- 当$|\phi| < 1$，$\sum_{k=0}^{\infty} \psi_k^2 = 1 + \phi^2 + \phi^4 + \phi^6 + \cdots = \frac{1}{1-\phi^2} < \infty$,

  该时间序列平稳，此时，
- $E(Y_t) = 0$，$\mathrm{Var}(Y_t) = \sigma_e^2(1 + \phi^2 + \phi^4 + \phi^6 + \cdots) = \frac{\sigma_e^2}{1-\phi^2}$
- $\gamma_{t,t+k} = \mathrm{Cov}(Y_t, Y_{t+k}) = E(Y_t Y_{t+k})$

$$= \sigma_e^2 \phi^k (1 + \phi^2 + \phi^4 + \phi^6 + \cdots) = \frac{\sigma_e^2 \phi^k}{1 - \phi^2}$$

AR模型平稳性判别方法

- 假设 e_t 独立于 $Y_{t-1}, Y_{t-2}, \ldots$, 则 $AR(p)$ 模型存在平稳解的充要条件是该模型的 AR 特征多项式 $\Phi(x)=0$ 的根（称为特征根）都在单位圆外（模大于1）
- 假设 $\Phi(x)=(1-\lambda_1 x)\cdots(1-\lambda_p x)$, 则该平稳条件等价于 $|\lambda_j|<1, j=1,\ldots,p$.
- $\lambda_1,\ldots,\lambda_p$ 是 $y^p-\phi_1 y^{p-1}-\cdots\phi_{p-1}y-\phi_p=0$ 的根。

- 比如 $AR(1)$ 模型，$\Phi(x)=1-\phi x$, 则平稳条件为 $|\phi|<1$.

AR(2)模型平稳条件

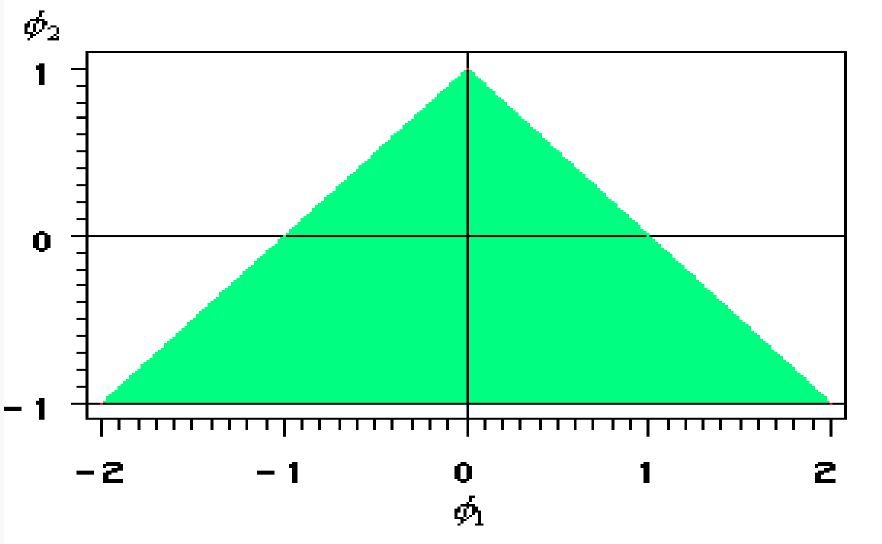
- λ_1, λ_2是$y^2 - \phi_1 y - \phi_2 = 0$的根

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

$$\lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

平稳条件： $|\lambda_j| < 1, j = 1,2$

- 平稳域



等价平稳条件：

$\{\phi_1, \phi_2 \mid |\phi_2| < 1$，且$\phi_2 \pm \phi_1 < 1\}$

例：AR(2)模型平稳条件

- $Y_t = Y_{t-1} - 0.5Y_{t-2} + e_t$
- $\phi_1 = 1$，$\phi_2 = -0.5$
- $y^2 - y + 0.5 = 0$的根为$\lambda_1 = \frac{1+i}{2}$，$\lambda_2 = \frac{1-i}{2}$
- $|\lambda_1| = |\lambda_2| = \frac{\sqrt{2}}{2} < 1$，所以平稳。
- 或者直接验证$|\phi_2| < 1$，$\phi_2 + \phi_1 = 0.5 < 1$，$\phi_2 - \phi_1 = -1.5 < 1$，所以平稳

协方差函数

- 在平稳$AR(p)$模型两边同乘Y_{t-k}，再求期望，可得

$$E(Y_tY_{t-k}) = \phi_1 E(Y_{t-1}Y_{t-k}) + \cdots + \phi_p E\left(Y_{t-p}Y_{t-k}\right) + E(e_tY_{t-k})$$

- 当$k > 0$时，e_t和Y_{t-k}独立，可得自协方差函数的***递推公式***

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}$$

- 该关系满足常系数齐次线性差分方程。

- 由于$E(e_tY_t) = E\left(e_t(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t)\right) = \sigma_e^2$，所以

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \cdots + \phi_p\gamma_{-p} + \sigma_e^2$$

平稳AR(1)模型的协方差

- 由于$\gamma_1 = \phi\gamma_0$，$\gamma_0 = \phi\gamma_1 + \sigma_e^2$，解得
$$\gamma_0 = \frac{\sigma_e^2}{1-\phi^2}$$
- 递推公式$\gamma_k = \phi\gamma_{k-1} = \phi^k\gamma_0 = \phi^k \frac{\sigma_e^2}{1-\phi^2}$
- $\rho_k = \gamma_k/\gamma_0 = \phi^k$

自相关系数(ACF)递推公式

- 自相关系数 $\rho_k = \frac{\gamma_k}{\gamma_0}$
- 当$k > 0$时，$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}$
- 两边同时除以γ_0，可得自相关系数递推公式

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}$$

$$\rho_k = \rho_{-k}, \rho_0 = 1$$

- 令$k = 1,2, \ldots, p$, 可得尤尔−沃克(Yule−Walker)方程组

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\ \rho_2 = \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\ \quad \vdots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p \end{cases}$$

常用AR模型的ACF

- $AR(1)$，$\rho_k = \phi^k$

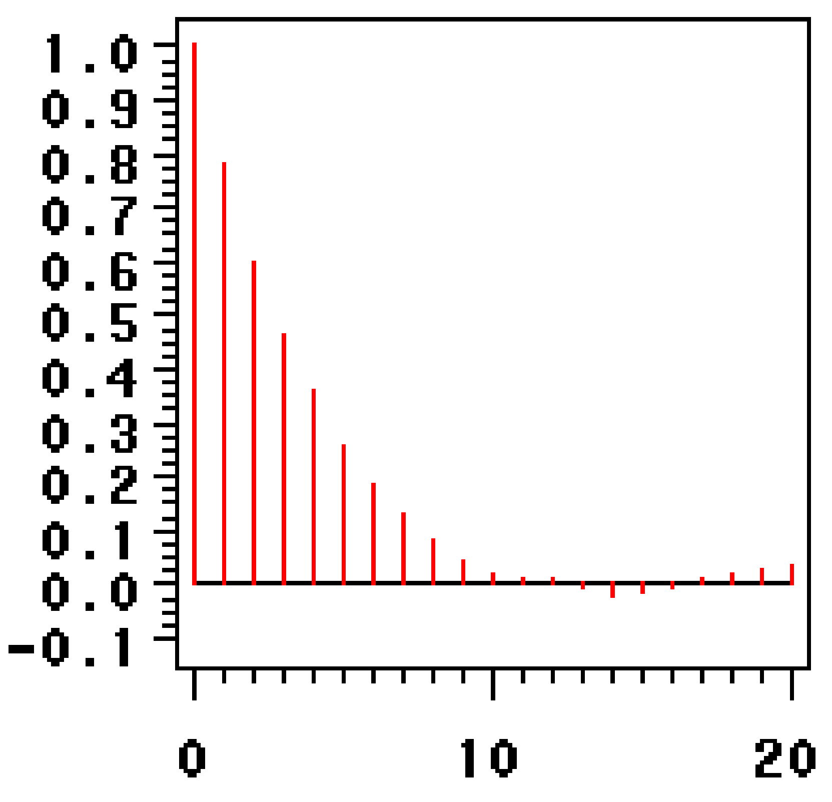
- $AR(2)$

$$\rho_k = \begin{cases} 1 & k = 0 \\ \dfrac{\phi_1}{1 - \phi_2} & k = 1 \\ \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} & k \geq 2 \end{cases}$$

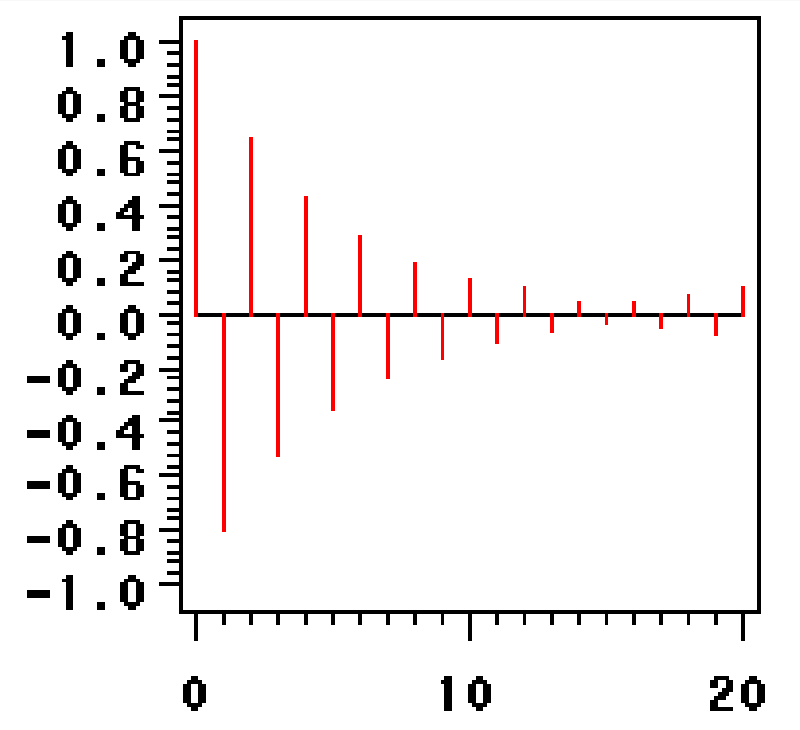
考察如下模型的自相关系数

- $(1) Y_t = 0.8Y_{t-1} + e_t$
- $(2) Y_t = -0.8Y_{t-1} + e_t$
- $(3) Y_t = Y_{t-1} - 0.5Y_{t-2} + e_t$

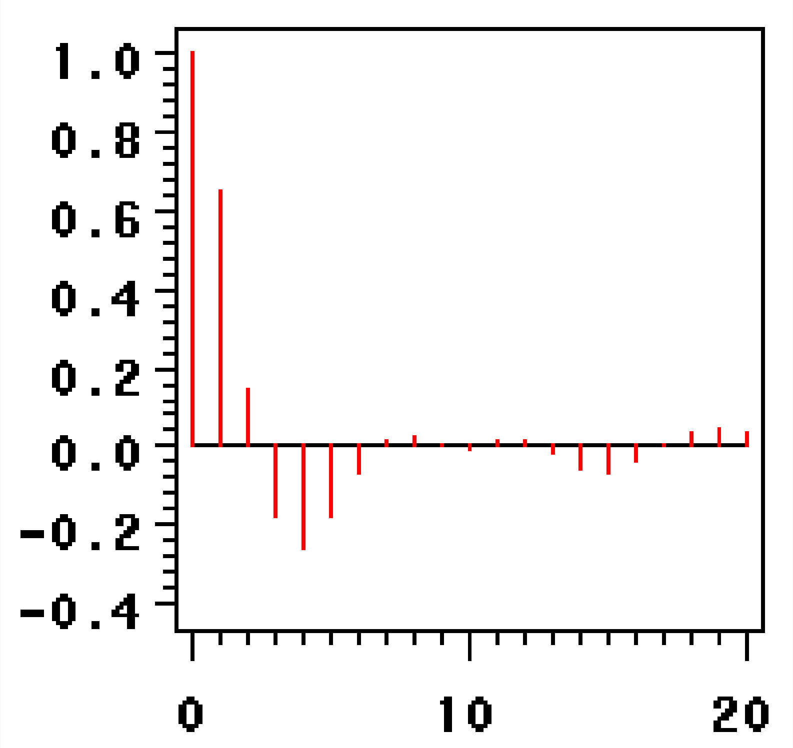
- $(1) Y_t = 0.8Y_{t-1} + e_t$，自相关系数指数收敛到0



- (2)$Y_t = -0.8Y_{t-1} + e_t$，自相关系数正负交替指数收敛到0



- (3)$Y_t = Y_{t-1} - 0.5Y_{t-2} + e_t$，自相关系数阻尼正弦波动



关于阻尼因子和频率

- 对于$AR(2)$模型，$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$，$k \geq 2$
- 设$\Phi(x) = (1-\lambda_1 x)\,(1-\lambda_2 x)$, 其实$\lambda_1, \lambda_2$是$y^2 - \phi_1 y - \phi_2 = 0$的根，平稳条件为$|\lambda_j| < 1, j = 1, \ldots, p$.
- 当λ_j均为实数，ρ_k指数变化；若λ_1, λ_2为共轭复数，我们有（见书53页）

$$\rho_k = R^k \frac{\sin(\Theta k + \Phi)}{\sin(\Phi)}$$

- 其中，$R = \sqrt{-\phi_2}$，$\cos(\Theta) = \frac{\phi_1}{2R}$
- 此时，R称为阻尼因子，Θ称为频率，Φ称为相位

ARMA(p, q)模型的定义

- 具有如下结构的模型称为自回归滑动平均模型，简记为$ARMA(p,q)$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

- 亦可记为$\Phi(B)Y_t = \Theta(B)\,e_t$
- 其中，$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \phi_p \neq 0,$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q,\ \theta_q \neq 0$$

- $ARMA(p,q)$模型存在平稳解的充要条件为该模型的AR特征多项式$\Phi(x) = 0$的根都在单位圆外。（为什么？教材公式（4.4.7））

ARMA(1, 1)模型

- $Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$
- 平稳条件
- 自协方差，自相关系数
- 自相关系数的特点和什么模型类似?
- 见书56页
- 推广到ARMA(p, q)

AR(p)模型的传递形式

- $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$
- 我们要求 e_t 独立于$Y_{t-1}, Y_{t-2}, \dots, e_t$称为新息项。
- 若一个AR(p)模型能够表示成收敛的一般线性过程形式，则称该模型为传递的AR模型。
- 注意：并非所有满足AR(p)模型的序列都平稳。
- $Y_t = \sum_{k=0}^{\infty} \psi_k e_{t-k}$，收敛条件为$\sum_{k=0}^{\infty} \psi_k^2 < \infty$，此时模型平稳，故AR(p)模型"传递"意味着平稳。
- 问题：如何求系数$\{\psi_k\}$? (待定系数法)
- 当$|\lambda_j| < 1, j = 1,2,\dots,p$ 时，$\sum_{k=0}^{\infty} \psi_k^2 < \infty$
- 此时可以把$\Phi(B)Y_t = e_t$写成$Y_t = \frac{1}{\Phi(B)} e_t$

MA(q)模型的可逆性

- 当一个MA(q)模型能够表示成收敛的（无穷阶）AR模型形式，称该模型可逆。（类比传递形式）
- 形式上将$Y_t = \Theta(B)\, e_t$写作$e_t = \frac{1}{\Theta(B)} Y_t$
- 可逆条件是什么?
- 条件：MA多项式$\Theta(x) = 1 - \theta_1 x - \theta_2 x^2 - \cdots - \theta_q x^q = 0$的根都在单位圆外（模长均大于1）
- 等价条件：$y^q - \theta_1 y^{q-1} - \cdots - \theta_q = 0$的根模长小于1

例：MA模型的可逆性

- $Y_t = e_t - 2e_{t-1}$ 和 $Y_t = e_t - 0.5e_{t-1}$
- 前者不可逆，后者可逆
- $Y_t = e_t - \frac{4}{5}e_{t-1} + \frac{16}{25}e_{t-2}$ 和 $Y_t = e_t - \frac{5}{4}e_{t-1} + \frac{25}{16}e_{t-2}$
- 前者可逆，后者不可逆
- 可以证明，在给定自相关函数的情况下，只有唯一的一组参数可以得到可逆的MA模型。
- 何时需要模型可逆？（例：模型预测）

ARMA(p, q)的传递形式和逆转形式

- $\Phi(B)Y_t = \Theta(B)\, e_t$可以形式上记为
- 传递形式： $Y_t = \frac{\Theta(B)}{\Phi(B)} e_t$
- 逆转形式： $e_t = \frac{\Phi(B)}{\Theta(B)} Y_t$
- ARMA模型的平稳性完全由AR部分决定，可逆性完全由MA部分决定，当平稳条件满足时，传递形式成立；当可逆条件满足时，逆转形式成立。两种情况下我们均可以用待定系数法求多项式分式的系数（教材4.4.7)
- 今后，我们最感兴趣的模型都将是平稳可逆的。

练习题

- $Y_t = Y_{t-1} - 0.5Y_{t-2} + e_t - 0.5e_{t-1}$
- 是否平稳？是否可逆?
- 求传递形式的系数、逆转形式的系数（前几个和递推式）
- 求该过程的自协方差函数、自相关系数
- 有几种思路?