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

复习上节课的加法模型

- 问题一：线性趋势项加周期为12的季节项构成的序列，满足什么模型?

- 问题二：二次趋势加周期为7的季节项构成的序列，满足什么模型?

- 思考题：如果有不规则项（随机项），怎么办?

复习上节课的加法模型

- 问题一：线性趋势项加周期为12的季节项构成的序列，满足什么模型?
- 答:
$$(1-B^{12})(1-B)Y_t=0$$
- 问题二：二次趋势加周期为7的季节项构成的序列，满足什么模型?
- 答:
$$(1-B^{7})(1-B)^2Y_t=0$$
- 思考题：如果有不规则项（随机项），怎么办?

数据与模型

- 我们观测到的时间序列，通常是具体的数据值，为了从概率角度进行理性分析，挖掘数据的深层次信息，我们通常认为该序列是某个总体概率模型（随机变量构成的序列）的一次样本实现。
- 总体模型是一种概率模型，定义了随机变量序列$\{Y_t\}$的分布特点、生成方式等。
- 而观测到的序列$\{y_t\}$是该随机变量序列的一次实现。
- 未观测到数据之前，样本看作是随机变量序列$\{Y_t\}$
- 观测到数据之后，样本看作是具体数值构成的序列$\{y_t\}$
- 这就是样本的两重性。

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
  $\gamma_{t,s} = \mathrm{Cov}(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)] = E(Y_t Y_s) - \mu_t \mu_s$
- **自相关函数（ACF）**

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

时间序列分析的数据结构

- 时间序列分析的数据结构有它的特殊性，时间序列在任意时刻的值都是一个随机变量，而且由于时间的不可重复性，任意时刻的变量只能获得唯一的一个样本观察值。
- 由于样本信息太少，在没有其他条件的情况下，通常这种数据结构是没有办法进行统计分析的，而序列平稳性可以有效地解决这个困难。

时间序列的平稳性

- 序列的随机特征不随时间变化而变化：平移不变性
- 严平稳——分布平稳
- 对一切时滞k和时点$t_1, t_2, \cdots, t_n$，都有

$$Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n} \text{ 与 } Y_{t_1-k}, Y_{t_2-k}, \ldots, Y_{t_n-k}$$

的联合分布相同，则称序列（过程）$\{Y_t\}$ 是严平稳的。

- 不难发现，（二阶矩有限的）严平稳序列具有相同的均值和方差，且

$$\gamma_{t,s} = \gamma_{0,t-s}$$

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

$$\gamma_{t,s} = \gamma_{0,t-s} \text{ 或 } \gamma_{t,t-k} = \gamma_{0,k}$$

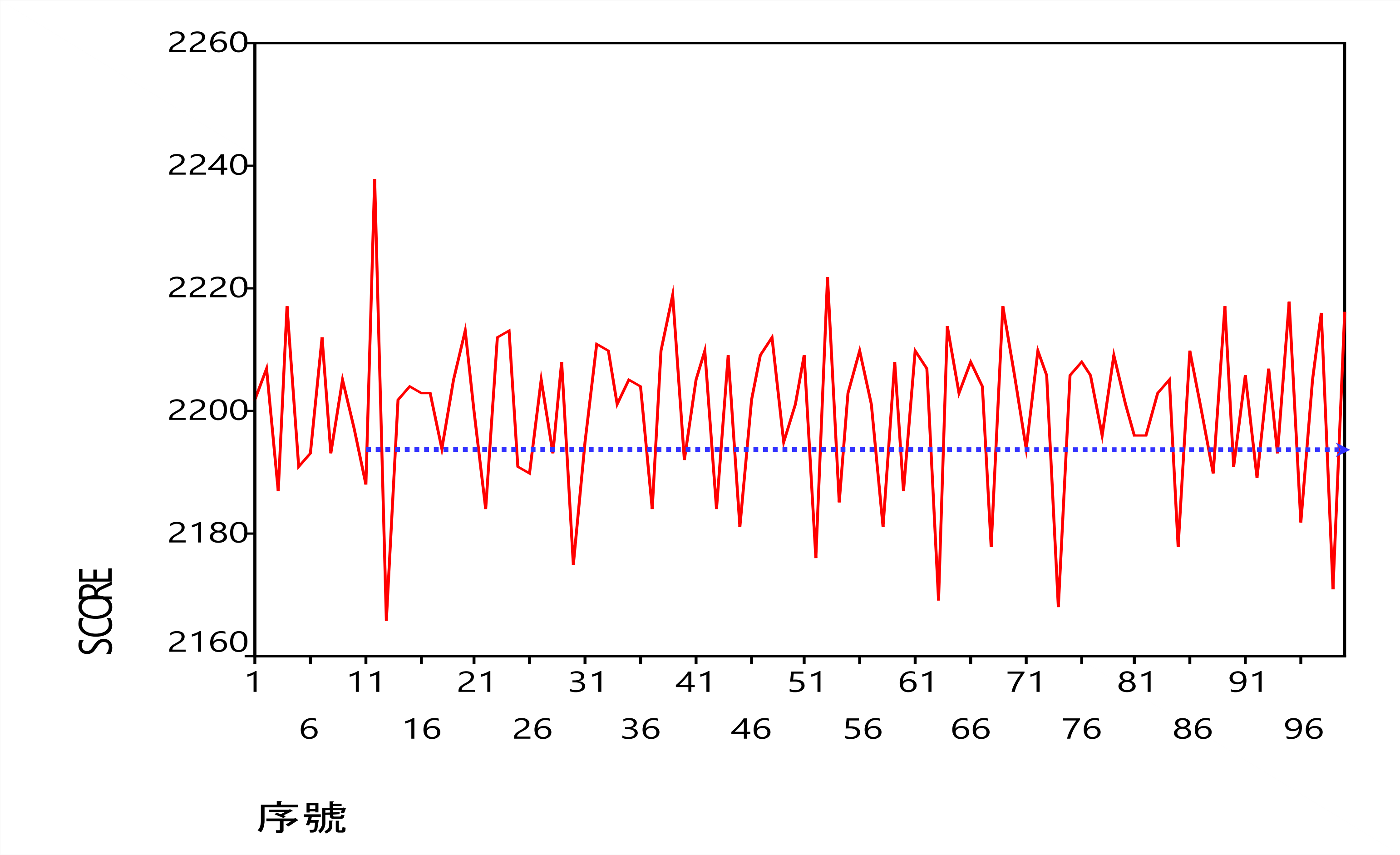
$$\text{定义} \gamma_k \coloneqq \gamma_{0,k}$$

- 自相关函数（ACF）也有类似性质，且 $\rho_k = \frac{\gamma_k}{\gamma_0}$

严平稳与弱平稳的关系

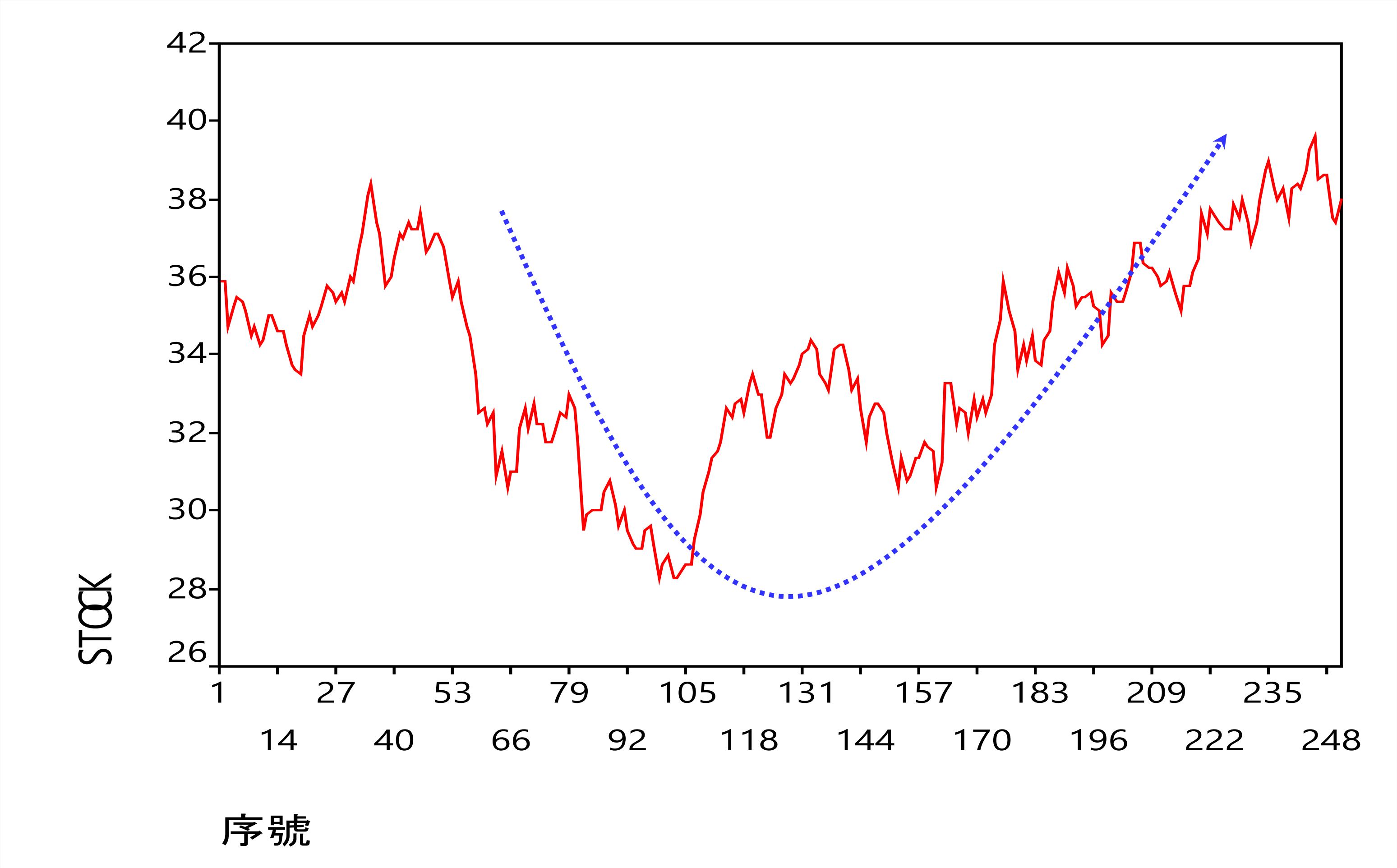
- 弱平稳序列未必是严平稳序列。
- 严平稳序列未必是弱平稳序列。
- 二阶矩有限的严平稳序列必为弱平稳序列。
- 对正态序列：严平稳与弱平稳等价。

平稳时间序列的例子



2260
2240
2220
2200
2180
2160
SCORE
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
序號

非平稳时间序列的例子



42
40
38
36
34
32
30
28
26
STOCK
1
27
53
79
105
131
157
183
209
235
14
40
66
92
118
144
170
196
222
248
序號

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

MA模型的性质

- 自协方差函数

$$\gamma_k = \begin{cases} (1+\theta_1^2+\cdots+\theta_q^2)\sigma_e^2 & k=0 \\ \left(-\theta_k+\sum_{i=1}^{q-k}\theta_i\theta_{k+i}\right)\sigma_e^2 & 1 \le k \le q \\ 0 & k>q \end{cases}$$

- 自相关系数（自相关函数）

$$\rho_k = \begin{cases} 1 & k=0 \\ \frac{-\theta_k+\sum_{i=1}^{q-k}\theta_i\theta_{k+i}}{1+\theta_1^2+\cdots+\theta_q^2} & 1 \le k \le q \\ 0 & k>q \end{cases}$$

- 当$k>q$时，自相关系数为0，这种性质称为（q期）截尾。

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

对比如下MA模型的ACF

- $Y_t = e_t - 2e_{t-1}$ 和 $Y_t = e_t - 0.5e_{t-1}$

- $Y_t = e_t - \frac{4}{5}e_{t-1} + \frac{16}{25}e_{t-2}$

和 $Y_t = e_t - \frac{5}{4}e_{t-1} + \frac{25}{16}e_{t-2}$

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（在满足$EX^2 < \infty$的随机变量空间L^2中）（以下内容仅供拓展，不做要求*）

证：令$S_n = \sum_{k=0}^{n} \psi_k^2, Y_{t,n} = \sum_{k=0}^{n} \psi_k e_{t-k}$，由于$S_n \to \sum_{k=0}^{\infty} \psi_k^2 < \infty$，所以$\{S_n\}$为柯西列：对于任意小的$\varepsilon > 0$, 存在$N$, 当$n > m > N$时，

$S_n - S_m = \sum_{k=m+1}^{n} \psi_k^2 < \frac{\varepsilon}{\sigma_e^2}$, 则 $E\left(Y_{t,n} - Y_{t,m}\right)^2 = \sigma_e^2 \sum_{k=m+1}^{n} \psi_k^2 < \varepsilon$，即对于给定的$t, \{Y_{t,n}\}$是柯西列（在$L^2$中），由于$L^2$空间的完备性，存在一个随机变量$Y_t$，使得 $E\left(Y_{t,n} - Y_t\right)^2 \to 0$

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（教材4.1）：当$\sum_{k=0}^{\infty} \psi_k^2 < \infty$时，$\{Y_t\}$平稳。

例：$\psi_k = 0.5^k + 0.3^k$，证明$\{Y_t\}$平稳。

证：只需要证明$\sum_{k=0}^{\infty} \psi_k^2 < \infty$，经计算可得

$\sum_{k=0}^{\infty} \psi_k^2 = \sum_{k=0}^{\infty} (0.5^k + 0.3^k)^2 < \sum_{k=0}^{\infty} (2 \times 0.5^k)^2 = 4 \sum_{k=0}^{\infty} 0.25^k < \infty$

思考：$\psi_k = C_1 \lambda_1^k + C_2 \lambda_2^k + \cdots + C_p \lambda_p^k$, $|\lambda_j| < 1, j = 1,2,\ldots,p$, 证明$\{Y_t\}$平稳。

AR(1)序列平稳的条件

- $Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \phi^3 e_{t-3} + \cdots$
- $E(Y_t) = 0$
- $\mathrm{Var}(Y_t) = \sigma_e^2(1 + \phi^2 + \phi^4 + \phi^6 + \cdots)$
- 当$|\phi| < 1$, $\mathrm{Var}(Y_t) = \frac{\sigma_e^2}{1-\phi^2}$
- $\gamma_{t,t+k} = \mathrm{Cov}(Y_t, Y_{t+k}) = E(Y_t Y_{t+k})$

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

- 结论：$|\phi| < 1$时，一阶自回归序列平稳。

AR模型平稳性判别方法

- $AR(p)$ 模型平稳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 $\Phi(x)=(1-\lambda_1 x)\cdots(1-\lambda_p x)$, 则该条件等价于 $|\lambda_j|<1, j=1,\ldots,p$.
- 比如$AR(1)$模型，$\Phi(x)=1-\phi x$, 则平稳条件为$|\phi|<1$.
- 为什么（*）？
- 必要条件：$\phi_1+\cdots+\phi_p<1, \quad |\phi_p|<1$
- 另外，$\lambda_1,\ldots,\lambda_p$是$y^p-\phi_1 y^{p-1}-\cdots\phi_{p-1}y-\phi_p=0$的根

AR(2)模型平稳条件

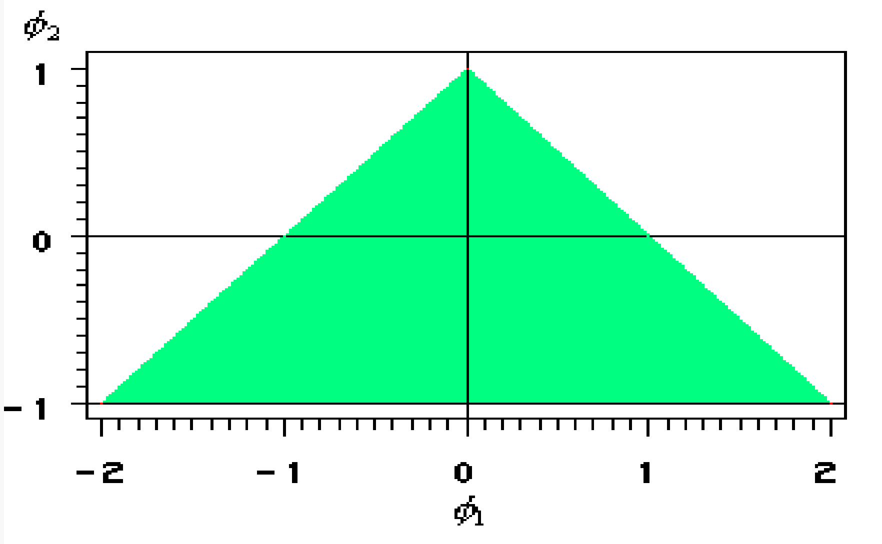
- 特征根的倒数

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

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

特征根判别： $|\lambda_j| < 1, j = 1,2$

- 平稳域



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

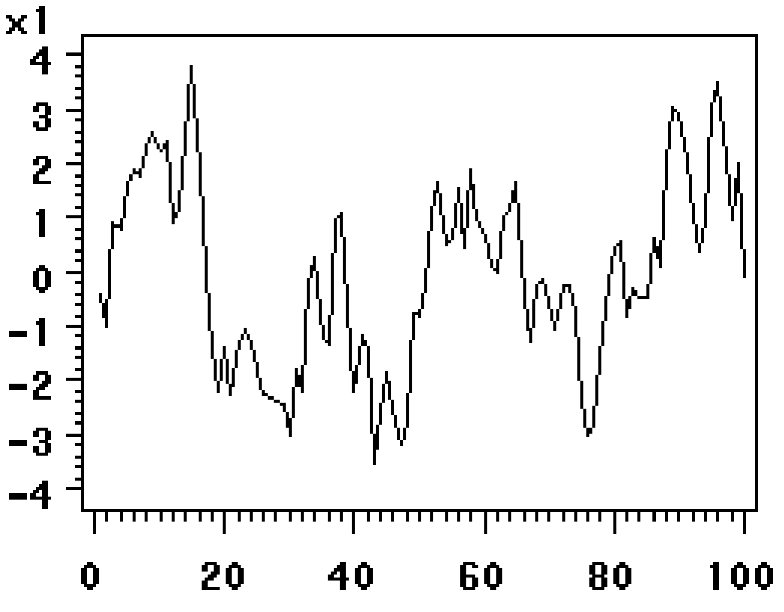
考察如下四个模型的平稳性

- $(1)Y_t = 0.8Y_{t-1} + e_t$
- $(2)Y_t = -1.1Y_{t-1} + e_t$
- $(3)Y_t = Y_{t-1} - 0.5Y_{t-2} + e_t$
- $(4)Y_t = Y_{t-1} + 0.5Y_{t-2} + e_t$

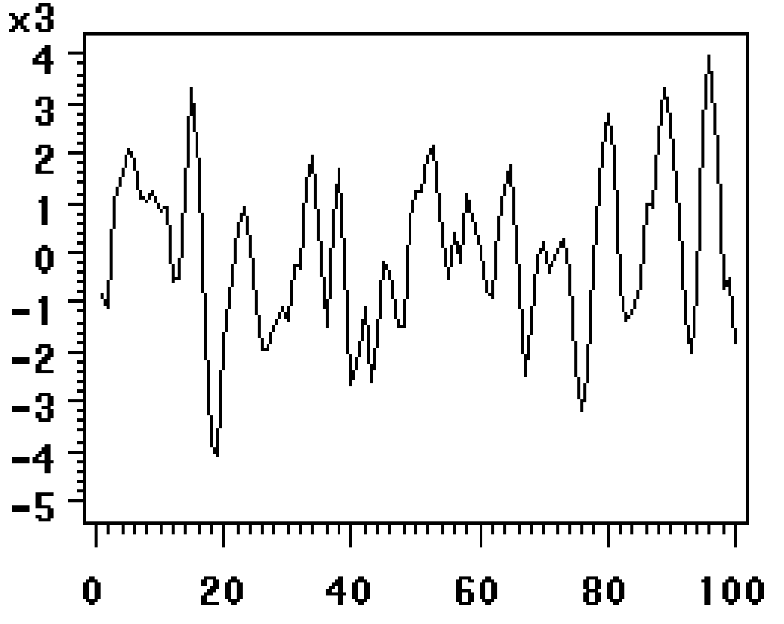
平稳性判别

模型	特征根判别	平稳域判别	结论
(1)	$\lambda_1 = 0.8$	$\phi = 0.8$	平稳
(2)	$\lambda_1 = -1.1$	$\phi = -1.1$	非平稳
(3)	$\lambda_1 = \frac{1+i}{2}$ $\lambda_2 = \frac{1-i}{2}$	$\lvert\phi_2\rvert = 0.5, \phi_2 + \phi_1 = 0.5, \phi_2 - \phi_1 = -1.5$	平稳
(4)	$\lambda_1 = \frac{1+\sqrt{3}}{2}$ $\lambda_2 = \frac{1-\sqrt{3}}{2}$	$\lvert\phi_2\rvert = 0.5, \phi_2 + \phi_1 = 1.5, \phi_2 - \phi_1 = -0.5$	非平稳

平稳序列时序图

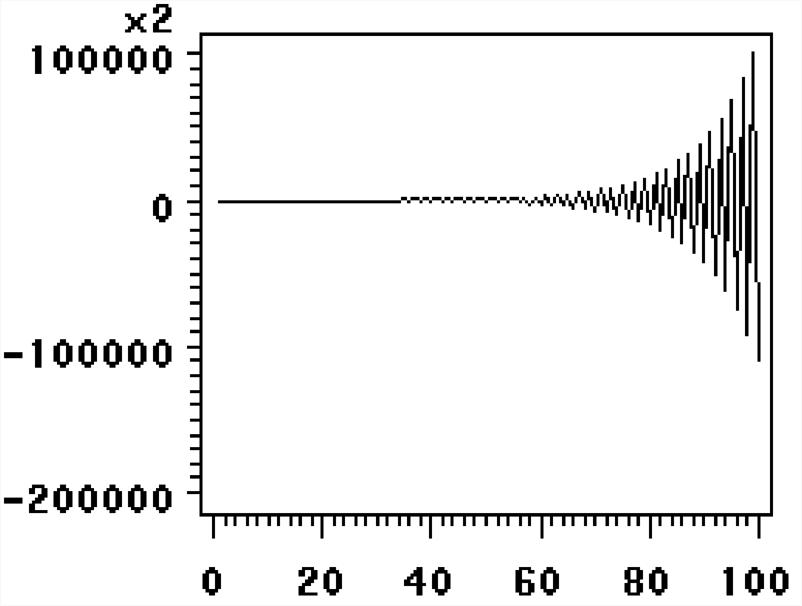


$(1) Y_t = 0.8Y_{t-1} + e_t$

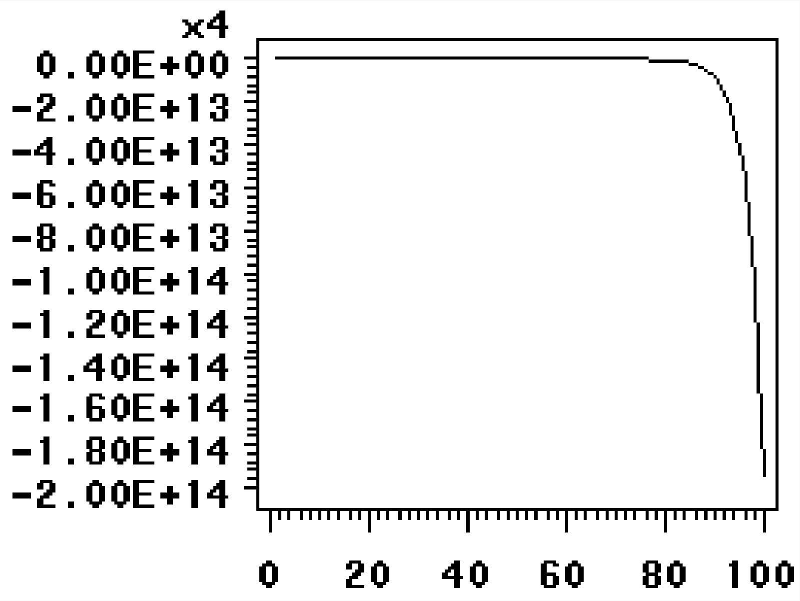


$(3) Y_t = Y_{t-1} - 0.5Y_{t-2} + e_t$

非平稳序列时序图



$(2) Y_t = -1.1Y_{t-1} + e_t$



$(4) Y_t = Y_{t-1} + 0.5Y_{t-2} + e_t$

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

$$E(Y_tY_{t-k})=\phi_1E(Y_{t-1}Y_{t-k})+\cdots+\phi_pE\left(Y_{t-p}Y_{t-k}\right)+E(e_tY_{t-k})$$

- 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$E(e_tY_t)=E\left(e_t(\phi_1Y_{t-1}+\phi_2Y_{t-2}+\cdots+\phi_pY_{t-p}+e_t)\right)=\sigma_e^2$，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

- $AR(1)$，$\rho_k = \phi^k$

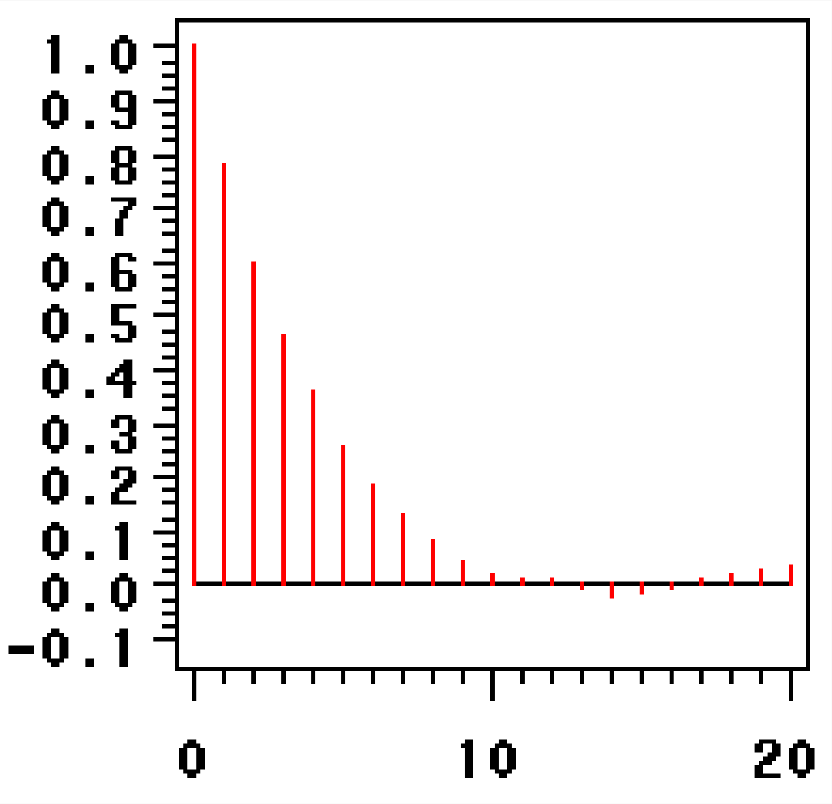
- $AR(2)$

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

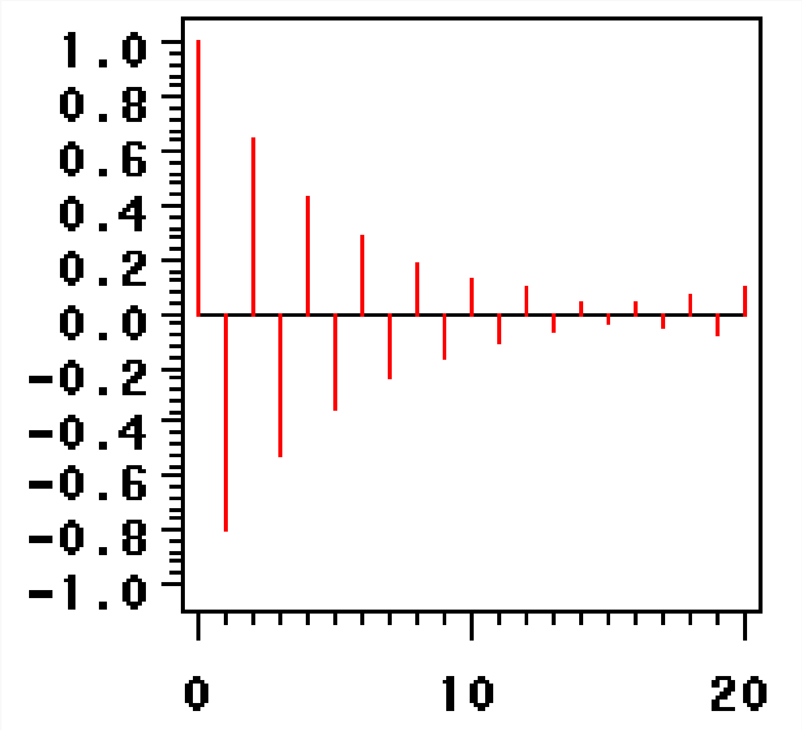
考察如下模型的自相关系数

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

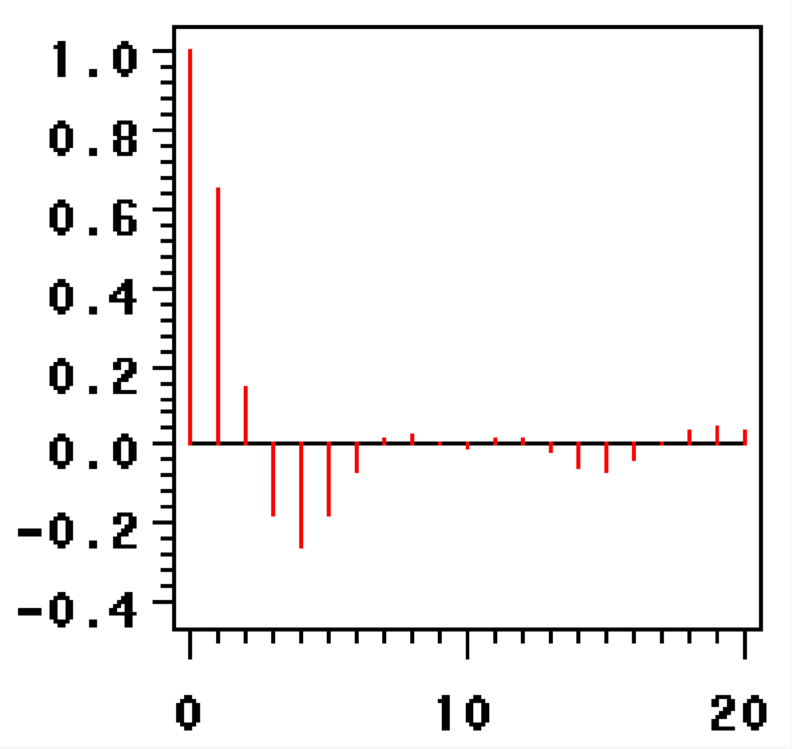
■ (1)$Y_t = 0.8Y_{t-1} + e_t$，自相关系数指数收敛到0



- $(2) Y_t = -0.8Y_{t-1} + e_t$，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



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

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

- $ARMA(p,q)$模型平稳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））