时间序列分析（初级）

非平稳序列分析 (I)

陈垚翰

安徽大学 大数据与统计学院

本章结构

- Cramer 分解定理
- 差分平稳
- ARIMA 模型

Cramer 分解定理

Cramer 分解定理

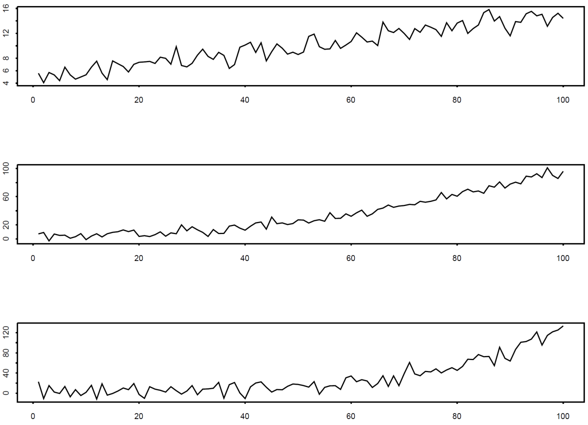
- 任何一个时间序列 $\{x_t\}$ 都可以分解成两部分的叠加

$$x_t = \mu_t + \varepsilon_t \tag{1}$$

  其中

$$\begin{aligned} \mu_t &= f(t) \quad \text{确定性影响} \\ \varepsilon_t &= \Psi(B)\, a_t \quad \text{随机性影响} \end{aligned}$$

- 暂时令 $\mu_t = \sum_{j=0}^{d} \beta_j t^j$



- $f(t) = c_1 + c_2 t$
- $f(t) = Ae^{rt}$

差分平稳

差分运算的实质

- 多项式的差分分解

$$\nabla^j t^d = c, \quad c \text{ 为常数}, \quad \forall j \geqslant d \tag{1}$$

- 任意 d 阶差分的形式

$$\nabla^d x_t = (1-B)^d x_t = \sum_{i=0}^{d} (-1)^i C_d^i x_{t-i}$$

其中

$$C_d^i = \frac{d!}{d!(d-i)!}$$

- d 阶差分实质是 d 阶自回归过程

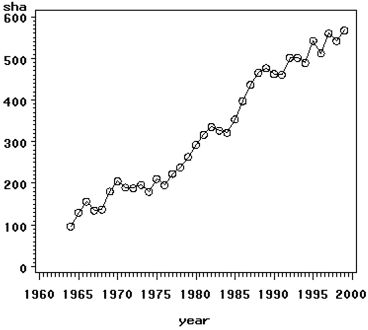
$$\nabla^d x_t = x_t + \sum_{i=1}^{d} (-1)^i C_d^i x_{t-i}$$

$$\Downarrow$$

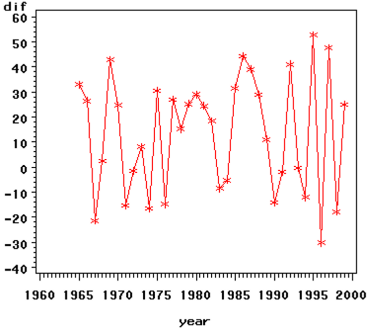
$$x_t = \sum_{i=1}^{d} (-1)^{i+1} C_d^i x_{t-i} + \nabla^d x_t$$

例1. 1964 年–1999 年中国纱年产量

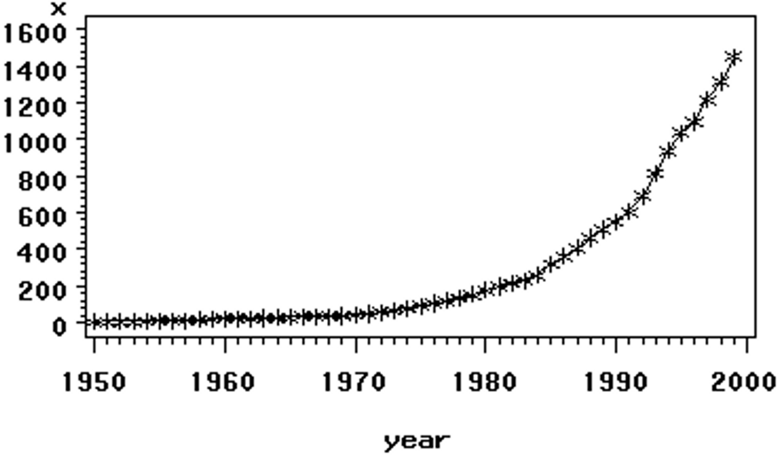
- 原序列时序图



- 差分后序列时序图

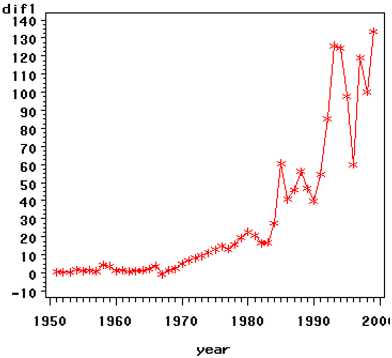


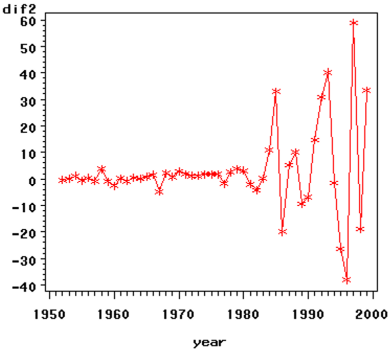
例2. 1950 年–1999 用车量拥有量



例2. 1950 年–1999 用车量拥有量：差分后序列时序图

- 一阶差分
- 二阶差分





过差分

- 过度的差分会造成有用信息的浪费
- 比较序列 $x_t = \beta_0 + \beta_1 t + a_t$ 的一阶差分和二阶差分
- 一阶差分
  - 平稳

$$\begin{aligned}\nabla x_t &= x_t - x_{t-1} \\ &= \beta_1 + a_t - a_{t-1}\end{aligned}$$

  - 方差小

$$\mathrm{Var}\left(\nabla x_t\right) = \mathrm{Var}\left(a_t - a_{t-1}\right) = 2\sigma^2$$

- 二阶差分（过差分）
  - 平稳

$$\begin{aligned}\nabla^2 x_t &= \nabla x_t - \nabla x_{t-1} \\ &= a_t - 2a_{t-1} + a_{t-2}\end{aligned}$$

  - 方差大

$$\mathrm{Var}\left(\nabla^2 x_t\right) = \mathrm{Var}\left(a_t - 2a_{t-1} + a_{t-2}\right) = 6\sigma^2$$

ARIMA 模型

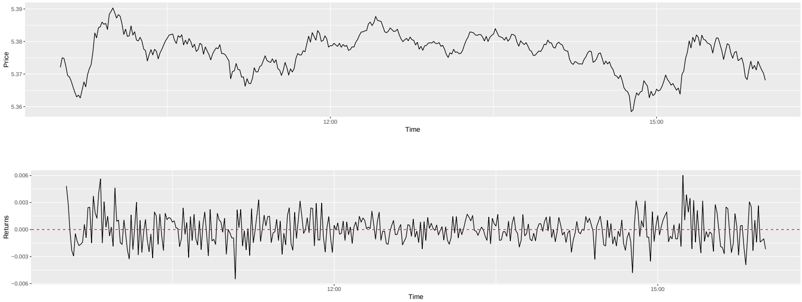
- 模型结构

$$
\begin{cases}
\Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t & (1) \\
\mathbb{E}(\varepsilon_t) = 0, \quad \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2, \quad \mathbb{E}(\varepsilon_t\varepsilon_s) = 0, \quad s \neq t & (2) \\
\mathbb{E}(x_s\varepsilon_t) = 0, \quad \forall s < t & (3)
\end{cases}
$$

- $d = 0$
$$\text{ARIMA}(p, d, q) = \text{ARMA}(p, q)$$
- $p = 0$
$$\text{ARIMA}(P, d, q) = \text{IMA}(d, q)$$
- $q = 0$
$$\text{ARIMA}(P, d, q) = \text{ARI}(p, d)$$
- $d = 1, p = q = 0$

$$\text{ARIMA}(p, d, q) = \text{随机游走模型 (random walk model)}$$

Apple Inc. 2018 年 10 月 11 日股票价格



- 假设 x_t 服从 ARIMA(p, d, q) 模型

$$\Phi\left(B\right)\nabla^{d} x_t = \Theta\left(B\right)\varepsilon_t$$

其中

$$\begin{aligned}\nabla^{d} &= (1-B)^{d}\\ \Phi(B) &= 1-\phi_1 B-\cdots-\phi_p B^{p}\\ \Theta(B) &= 1-\theta_1 B-\cdots-\theta_q B^{q}\end{aligned}$$

- ARIMA(p, d, q) 的广义自回归多项式

$$\widetilde{\Phi}(B) = \Phi(B)\nabla^{d} = \left[\prod_{i=1}^{p}(1-\lambda_i B)\right](1-B)^{d}$$

- 当 $d \neq 0$ 时，ARIMA(p, d, q) 模型不平稳

- 假设 x_t 在 ARIMA(p,d,q) 下有如下 MA(∞) 表示

$$\begin{aligned} x_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ &= \Psi(B)\varepsilon_t \end{aligned}$$

- 注意到

$$\begin{aligned} \Phi(B)\,(1-B)^d\,x_t &= \Phi(B)\,(1-B)^d\,\Psi(B)\varepsilon_t = \Theta\,(B)\,\varepsilon_t \\ &= \left(1 - \tilde{\phi}_1 B - \tilde{\phi}_2 B^2 - \cdots - \tilde{\phi}_{p+d} B^{p+d}\right)\left(1 + \psi_1 B + \psi_2 B^2 + \cdots\right)\varepsilon_t \\ &= \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)\varepsilon_t \end{aligned}$$

- ARIMA(p, d, q) 的 Green 函数有如下递推公式

$$\begin{cases} \psi_0 = 1 \\ \psi_i = \sum_{j=1}^{i} \tilde{\phi}'_j \psi_{i-j} - \theta'_i \quad i \geqslant 1 \end{cases} \tag{4}$$

其中，

$$\tilde{\phi}'_j = \begin{cases} \tilde{\phi}_j \ , \ j \leqslant p+d \\ 0 \ , \ j > p+d \end{cases}$$

且

$$\theta'_i = \begin{cases} \theta_i \ , \ j \leqslant q \\ 0 \ , \ j > q \end{cases}$$

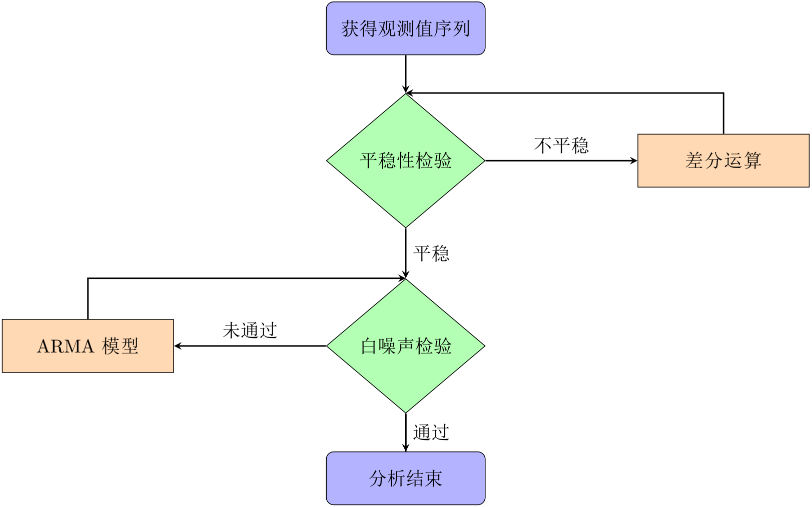
- 预测值

$$x_{t+l} = \underbrace{(\varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \cdots + \psi_{l-1}\varepsilon_{t+1})}_{e_t(l)} + \underbrace{(\psi_l \varepsilon_t + \psi_{l+1}\varepsilon_{t-1} + \cdots)}_{\widehat{x}_t(l)}$$

- 条件均值和方差

$$\begin{aligned}
\mathbb{E}\left(x_{t+1} \mid x_t, x_{t-1}, \cdots\right) &= \widehat{x}_t(l) \\
\operatorname{Var}\left(x_{t+1} \mid x_t, x_{t-1}, \cdots\right) &= \operatorname{Var}\left[e_t(l)\right] \\
&= \sum_{i=0}^{l-1} \psi_i^2 \sigma_\varepsilon^2
\end{aligned}$$

ARIMA 模型建模



- 一个案例