Time Series

Guanxu Gleason WANG

University of Glasgow

版本：Past Paper of STAT5030 Time Series, UofG

更新：May 6, 2025

目录

定义 0.1 (Purely random process) A **purely random process** (also known as a **white noise process**) is a time series process $\{X_t \mid t \in T\}$ defined by

$$\mathbb{E}[X_t] = \mu$$
$$\mathrm{Var}[X_t] = \sigma^2.$$

where each X_t is independent.

定义 0.2 (Random walk process) A **random walk process** is a time series process $\{X_t \mid t \in T\}$ defined by

$$X_t = X_{t-1} + Z_t$$

where Z_t is a purely random process with mean μ and variance σ^2. The process is started at $X_0 = 0$, so that $X_1 = Z_1$, $X_2 = X_1 + Z_2$, etc.

定理 0.1 (Testing for independence) Given a time series $(x_1, x_2, \cdots, x_n)$, the hypotheses for this test are:

$$H_0 : \text{the time series is independent} \quad \text{v.s.} \quad H_1 : \text{the time series is not independent}$$

Two competing test statistics have been developed for this test, namely

$$\textbf{Box-Pierce statistic} \quad Q_{\mathrm{BP}} = n \sum_{\tau=1}^{h} \widehat{\rho}_\tau^2 \qquad \textbf{Ljung-Box statistic} \quad Q_{\mathrm{LB}} = n(n+2) \sum_{\tau=1}^{h} \frac{\widehat{\rho}_\tau^2}{n-\tau}$$

Under H_0 both these statistics have a χ_h^2 distribution, so we reject H_0 at the 5% level if the test statistic is greater than a $\chi_{h,0.95}^2$. 除此之外，既然是独立性检验，就应该能检验出是否存在 short-term correlation.

1 自回归 (Autoregressive, AR) 过程

1.1 定义

The notation $\mathrm{AR}(p)$ indicates an autoregressive model of order p. The $\mathrm{AR}(p)$ model is defined as

$$\begin{aligned} X_t &= \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t \\ &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t, \end{aligned}$$

where we assume $X_0 = X_{-1} = \cdots = X_{1-p} = 0$.

The parameters $\phi_1, \cdots, \phi_p$ are the parameters of the model, where ϕ_i is called the lag i coefficient, and Z_t is a purely random process with mean zero and variance σ_z^2.

This can be equivalently written using the backshift operator (滞后算子) B as

$$\begin{aligned} X_t = \sum_{i=1}^{p} \phi_i B^i X_t + Z_t &= \left(\sum_{i=1}^{p} \phi_i B^i \right) X_t + Z_t \\ &= (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) X_t + Z_t, \end{aligned}$$

where $X_{t-i} = B^i X_t$, so that, moving the summation term to the left side and using polynomial notation, we have

$$\phi(B) X_t = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = Z_t.$$

不难发现 AR(p) 等价于 MA(0), 因此 AR(p) **天然可逆**.

1.2 平稳性 (Stationarity)

1.2.1 宽平稳 (Weak Stationarity)

时间序列 $\{X_t\}$ 被称为宽平稳 (也称二阶平稳)，若

1. 均值恒定（Constant Mean）：
$$\mathbb{E}\left[X_t\right] = \mu, \quad \forall t.$$
2. 方差恒定（Constant Variance）：
$$\operatorname{Var}\left[X_t\right] = \mathbb{E}\left[(X_t - \mu)^2\right] = \sigma^2, \quad \forall t.$$
3. 自协方差仅依赖于滞后阶数 τ（Autocovariance Depends only on Lag）：
$$\gamma_X(\tau) = \operatorname{Cov}\left[X_t, X_{t+\tau}\right] = \mathbb{E}\left[(X_t - \mu)(X_{t+\tau} - \mu)\right]$$
仅依赖于滞后阶数 τ 而与 t 无关。

例 1.1 Prove that a sum of two mutually independent weakly stationary processes with corresponding means μ_x, μ_y, variances σ_x^2, σ_y^2 and autocovariance functions $\gamma_\tau^x, \gamma_\tau^y$ is weakly stationary.

证明. Define a new process $Z_t = X_t + Y_t$, then

1. 均值恒定:

$$\mathbb{E}\left[Z_t\right] = \mathbb{E}\left[X_t + Y_t\right] = \mathbb{E}\left[X_t\right] + \mathbb{E}\left[Y_t\right] = \mu_x + \mu_y$$

   is constant.

2. 方差恒定:

$$\mathrm{Var}\left[Z_t\right] = \mathrm{Var}\left[X_t + Y_t\right] = \mathrm{Var}\left[X_t\right] + \mathrm{Var}\left[Y_t\right] + 2\mathrm{Cov}\left[X_t, Y_t\right] = \sigma_x^2 + \sigma_y^2$$

   is constant.

3. 自协方差:

$$\begin{aligned}\gamma_\tau^z &= \mathrm{Cov}\left[Z_t, Z_{t+\tau}\right] = \mathrm{Cov}\left[X_t + Y_t, X_{t+\tau} + Y_{t+\tau}\right] \\ &= \mathrm{Cov}\left[X_t, X_{t+\tau}\right] + \mathrm{Cov}\left[X_t, Y_{t+\tau}\right] + \mathrm{Cov}\left[Y_t, X_{t+\tau}\right] + \mathrm{Cov}\left[Y_t, Y_{t+\tau}\right] = \gamma_\tau^x + \gamma_\tau^y\end{aligned}$$

   only depends on τ.

Some parameter constraints are necessary for the model to remain weak-sense stationary. For example, processes in the $\mathrm{AR}(1)$ model with $|\phi_1| \geqslant 1$ are not stationary. More generally, for an $\mathrm{AR}(p)$ model to be weak-sense stationary, the roots of the **characteristic polynomial (特征多项式)** $\Phi(z) := 1 - \sum_{i=1}^{p} \phi_i z^i$ must lie outside the unit circle, i.e., each (complex) root z_i must satisfy $|z_i| > 1$.

不难发现, 其特征函数 $\Phi(z)$ 的形式等同于 $\phi(B)$.

证明. 给定 AR(1) 过程

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

得到特征方程

$$1 - \phi_1 B = 0 \quad \Rightarrow \quad B = \frac{1}{\phi_1},$$

若使 $|B| > 1$, 则需使 $|\phi_1| < 1$, 此时 AR(1) 宽平稳; 当 $|\phi_1| \geqslant 1$ 时, AR(1) 不平稳.

平稳的 AR(p) 且 $\phi_1 + \phi_2 + \cdots + \phi_p \neq 1$ 的期望 $\mu = 0$.

证明. 给定 AR(p) 过程

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t,$$

取期望, 得

$$\begin{aligned}
\mathbb{E}[X_t] &= \phi_1 \mathbb{E}[X_{t-1}] + \phi_2 \mathbb{E}[X_{t-2}] + \cdots + \phi_p \mathbb{E}[X_{t-p}] + \mathbb{E}[Z_t] \\
\mu &= \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu + 0 \\
(1 - \phi_1 - \phi_2 - \cdots - \phi_p)\mu &= 0 \\
\mu &= 0 \quad (\phi_1 + \phi_2 + \cdots + \phi_p \neq 1)
\end{aligned}$$

若使 $|B| > 1$, 则需使 $|\phi_1| < 1$, 此时 AR(1) 宽平稳; 当 $|\phi_1| \geqslant 1$ 时, AR(1) 不平稳.

1.2.2 严平稳 (Strict Stationarity)

时间序列 $\{X_t\}$ 被称为严平稳, 若 $\{X_t\}$ 只与 n（时间长度）有关，而与时间起始点、h（时间差）无关, 即

$$\big((X(t_1+h),\cdots,X(t_n+h)\big)\overset{d}{=}\big((X(t_1),\cdots,X(t_n)\big),$$

即相同联合分布. **宽平稳**仅要求均值、方差和自协方差不变，而**严平稳**要求整个概率分布保持不变。

1.3 Yule-Walker Equation

假设 AR(1) 过程平稳, 即 $|\phi_1|<1$ 时, 有

$$\begin{aligned}X_t &= \phi_1 X_{t-1}+Z_t\\ X_t X_{t-\tau} &= \phi_1 X_{t-1}X_{t-\tau}+Z_t X_{t-\tau}\\ \mathbb{E}\left[X_t X_{t-\tau}\right] &= \phi_1\mathbb{E}\left[X_{t-1}X_{t-\tau}\right]+\mathbb{E}\left[Z_t X_{t-\tau}\right]\end{aligned}$$

其中, 由于 AR(1) 过程平稳, 且 $\phi_1\neq 1$, 则 $\mathbb{E}\left[X_i\right]=0$. 且白噪声序列 $\{Z_t\}$ 期望为 0, 那么有

$$\gamma_\tau=\mathrm{Cov}\left[X_t,X_{t-\tau}\right]=\mathbb{E}\left[X_tX_{t-\tau}\right]-\mathbb{E}\left[X_t\right]\mathbb{E}\left[X_{t-\tau}\right]=\mathbb{E}\left[X_tX_{t-\tau}\right].$$

因此,

$$\gamma_\tau=\phi_1\gamma_{\tau-1}.$$

当 $\tau = 1$ 时,

$$\gamma_1 = \phi_1 \gamma_0 \quad \Rightarrow \quad \phi_1 = \frac{\gamma_1}{\gamma_0} = \rho_1.$$

由

$$\rho_\tau = \mathrm{Corr}[X_t, X_{t-\tau}] = \frac{\mathrm{Cov}\left[X_t, X_{t-\tau}\right]}{\sqrt{\mathrm{Var}\left[X_t\right]}\sqrt{\mathrm{Var}\left[X_{t-\tau}\right]}} = \frac{\mathrm{Cov}\left[X_t, X_{t-\tau}\right]}{\mathrm{Var}\left[X_t\right]} = \frac{\gamma_\tau}{\gamma_0},$$

当 $\tau = 0$ 时, $\rho_0 = \dfrac{\gamma_0}{\gamma_0} = 1$. 因此

$$\rho_\tau = \phi_1 \rho_{\tau-1},$$

当 $\tau = 1$ 时,

$$\rho_1 = \phi_1 \rho_0 = \phi_1,$$

例 1.2 Consider a time series process

$$X_t = \frac{40X_{t-1} - 25X_{t-2} + 17Z_t}{32}.$$

解. 显然是个 AR(2), MA(0) 过程. 并且得到特征方程

$$\begin{aligned}
32 - 40B + 25B^2 = 0 \\
B &= \frac{40 \pm \sqrt{40^2 - 4 \times 25 \times 32}}{2 \times 25} \\
&= \frac{40 \pm \sqrt{-1600}}{50} = \frac{40 \pm 40\sqrt{-1}}{50} = \frac{4}{5}(1 \pm i)
\end{aligned}$$

因此得到两根分别为 $\frac{4}{5}(1+i)$ 和 $\frac{4}{5}(1-i)$, 得到模为 $\frac{4}{5}\sqrt{2}>1$, 因此平稳, 且天然可逆 (MA(0) 可逆).

推导 Yule-Walker 方程:

$$\begin{aligned}
X_t &= \frac{40X_{t-1}-25X_{t-2}+17Z_t}{32} \\
X_tX_{t-\tau} &= \frac{40X_{t-1}X_{t-\tau}-25X_{t-2}X_{t-\tau}+17Z_tX_{t-\tau}}{32} \\
\mathbb{E}\left[X_tX_{t-\tau}\right] &= \frac{40\mathbb{E}\left[X_{t-1}X_{t-\tau}\right]-25\mathbb{E}\left[X_{t-2}X_{t-\tau}\right]+17\mathbb{E}\left[Z_tX_{t-\tau}\right]}{32} \\
\gamma_\tau &= \frac{40\gamma_{\tau-1}-25\gamma_{\tau-2}}{32} \\
\rho_\tau &= \frac{40\rho_{\tau-1}-25\rho_{\tau-2}}{32}
\end{aligned}$$

当 $\tau=1$ 时,

$$\rho_1=\frac{40\rho_0-25\rho_{-1}}{32} \quad\Rightarrow\quad \rho_1=\frac{40\rho_0-25\rho_1}{32} \quad\Rightarrow\quad \rho_1=\frac{40}{32+25}\rho_0=\frac{40}{57}\approx 0.7018.$$

当 $\tau=2$ 时,

$$\rho_2=\frac{40\rho_1-25\rho_0}{32}=\frac{40\times\frac{40}{57}-25}{32}=\frac{175}{1824}\approx 0.09594.$$

例 1.3 Consider $\mathrm{AR}(2)$ process. Assuming $\rho_1=0.454$, $\rho_2=-0.448$, and $\rho_3=-0.743$, estimate the parameters of the AR model you selected.

解. Consider $\mathrm{AR}(2)$ process

$$X_t=\phi_1X_{t-1}+\phi_2X_{t-2}+Z_t.$$

Derive the Yule-Walker equation for this process

$$\rho_\tau = \phi_1 \rho_{\tau-1} + \phi_2 \rho_{\tau-2}.$$

Let $\tau = 1$ and $\tau = 2$,

$$\begin{cases} \rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} \\ \rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 \end{cases} \Rightarrow \begin{cases} 0.454 = \phi_1 + 0.454\phi_2 \\ -0.448 = 0.454\phi_1 + \phi_2 \end{cases} \Rightarrow \begin{cases} \phi_1 =? \\ \phi_2 =? \end{cases}$$

2 移动平均 (Moving Average, MA) 过程

2.1 定义

The notation $\mathrm{MA}(q)$ refers to the moving average model of order q:

$$\begin{aligned} X_t &= \mu + Z_t + \sum_{i=1}^{q} \theta_i Z_{t-i} \\ &= \mu + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \end{aligned}$$

where μ is the mean of the series, the $\theta_1, \cdots, \theta_q$ are the coefficients of the model and $Z_t, Z_{t-1}, \cdots, Z_{t-q}$ are the error terms. The value of q is called the order of the MA model.

This can be equivalently written in terms of the backshift operator B as

$$\begin{aligned} X_t &= \mu + (1 + \theta_1 B + \cdots + \theta_q B^q) Z_t \\ &= \mu + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) Z_t = \mu + \theta(B) Z_t. \end{aligned}$$

Thus, a moving-average model is conceptually a linear regression of the current value of the series against current and previous (observed) white noise error terms or random shocks. The random shocks at each point are assumed to be mutually independent and to come from the same distribution, typically a normal distribution, with location at zero and constant scale, 即 $\mathrm{MA}(q)$ 天然平稳.

2.2 可逆性（Invertibility）

时间序列模型的可逆性通常针对移动平均（MA）过程来定义。如果一个 MA(q) 过程：

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q},$$

可以等价地表示为一个收敛的自回归过程：

$$Z_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots,$$

并且该自回归过程是平稳的，则称该 MA(q) 过程是可逆的。不难发现 MA(q) 等价于 AR(∞).

定义 2.1 (Invertibility) An MA(q) process is invertible if it can be written as an autoregressive process of infinite order, i.e. as

$$X_t = \sum_{i=1}^{n} b_i X_{t-i} + Z_t,$$

where the coefficients b_j form a convergent (收敛) sum, i.e. $\sum_{i=1}^{\infty} b_i < \infty$.

In other words, a time series is invertible **if it can be expressed as a sum of its own past values**. Non-invertible processes may be expressed as a sum of future values. For this reason, invertible processes are more interpretable.

证明. 考虑 MA(1) 过程：

$$X_t = Z_t + \theta Z_{t-1}$$

我们尝试通过迭代展开来表示:

$$\begin{aligned} Z_t &= X_t - \theta Z_{t-1} \\ &= X_t - \theta(X_{t-1} - \theta Z_{t-2}) \\ &= X_t - \theta X_{t-1} + \theta^2 Z_{t-2} \\ &= \cdots \\ &= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \cdots \end{aligned}$$

若级数收敛, 即 $|\theta| < 1$，则该过程是可逆的。

More generally, for an $\mathrm{MA}(q)$ model to be invertibility, the roots of the polynomial $\Theta(z) := 1 + \sum_{i=1}^{q} \theta_i z^i$ must lie outside the unit circle, i.e., each (complex) root z_i must satisfy $|z_i| > 1$.

不难发现, 其特征函数 $\Theta(z)$ 的形式等同于 $\theta(B)$.

2.3 Autocorrelation Function

It is easy to know that

$$\mathbb{E}\left[X_t\right] = 0, \quad \mathrm{Var}\left[X_t\right] = \left(1 + \sum_{i=1}^{q} \theta_i^2\right) \sigma_z^2.$$

The autocovariance function is calculated as:

$$\begin{aligned}
\mathrm{Cov}\left[X_t, X_{t-\tau}\right] &= \mathbb{E}\left[X_t X_{t-\tau}\right] - \mathbb{E}\left[X_t\right]\mathbb{E}\left[X_{t-\tau}\right] \\
&= \mathbb{E}\left[X_t X_{t-\tau}\right] \\
&= \mathbb{E}\left[\left(\sum_{i=0}^{q}\theta_i Z_{t-i}\right) \times \left(\sum_{j=0}^{q}\theta_j Z_{t-j-\tau}\right)\right] \\
&= \sum_{i=0}^{q}\sum_{j=0}^{q}\theta_i\theta_j \mathbb{E}\left[Z_{t-i}Z_{t-j-\tau}\right]
\end{aligned}$$

The expectations $\mathbb{E}\left[Z_r Z_s\right]$ can be split into two cases:

- If $r \neq s$ we have $\mathbb{E}\left[Z_r Z_s\right] = \mathbb{E}\left[Z_r\right]\mathbb{E}\left[Z_s\right] = 0$ because the Z_t are independent.
- If $r = s$ we have $\mathbb{E}\left[Z_r^2\right] = \mathrm{Var}\left[Z_r\right] + \mathbb{E}\left[Z_r\right]^2 = \sigma_z^2$.

Therefore the autocovariance function is given by

$$\gamma_\tau = \mathrm{Cov}\left[X_t, X_{t-\tau}\right] = \begin{cases} \sigma_z^2 \displaystyle\sum_{i=0}^{q-\tau}\theta_i\theta_{i+\tau}, & \tau = 0, 1, \cdots, q, \\ 0, & \tau > q, \end{cases}$$

and the autocorrelation function is given by

$$\rho_\tau = \mathrm{Corr}\left[X_t, X_{t-\tau}\right] = \begin{cases} 1, & \tau = 0, \\ \dfrac{\sigma_z^2 \sum_{i=0}^{q-\tau} \theta_i \theta_{i+\tau}}{\sigma_z^2 \sum_{i=0}^{q} \theta_i^2}, & \tau = 1, \cdots, q, \\ 0, & \tau > q, \end{cases}$$

where again $\theta_0 = 1$.

3 定阶

3.1 自相关函数 (Autocorrelation Function, ACF)

ACF 衡量的是不同时间滞后（lag）下的自相关性，即:

$$\rho_k = \frac{\gamma_k}{\gamma_0},$$

其中:

- $\gamma_k = \mathrm{Cov}\left[X_t, X_{t-k}\right]$ 是滞后 k 的自协方差;
- $\gamma_0 = \mathrm{Var}\left[X_t\right]$ 是时间序列的方差.

ACF 显示的是总体的相关性，包括直接和间接的影响, 可以显示出**季节性**。

定阶

1. **截尾:** 对于 MA(q) 过程，ACF 在滞后 q 之后迅速截断（即 $> q$ 阶趋近于 0）, 即 q 阶截尾。
2. **拖尾:** 如果 ACF 逐渐衰减（指数或正弦波衰减）, 即拖尾，可能是 AR 过程.

3.2 偏自相关函数 (Partial Autocorrelation Function, PACF)

通过 Yule-Walker 方程计算可以得出

$$\rho_k = \phi_{1,k}\rho_{k-1} + \phi_{2,k}\rho_{k-2} + \cdots + \phi_{k,k}\rho_0,$$

其中, 最后一项系数 $\phi_{k,k}$ 即为 PACF.

定阶

1. **截尾:** 对于 AR(p) 过程，PACF 在滞后 p 之后迅速截断（即 $> p$ 阶趋近于 0），即 p 阶截尾。
2. **拖尾:** 如果 PACF 逐渐衰减（指数或正弦波衰减），即拖尾，可能是 MA 过程.
3. **均趋于 0:** 可能为白噪声过程.

例 3.1 给定统一数据的不同 AR(p) 过程:

$$\begin{aligned}
X_t &= \phi_{1,1}X_{t-1} + Z_t \\
X_t &= \phi_{2,1}X_{t-1} + \phi_{2,2}X_{t-1} + Z_t \\
X_t &= \phi_{3,1}X_{t-1} + \phi_{3,2}X_{t-1} + \phi_{3,3}X_{t-1} + Z_t \\
X_t &= \phi_{4,1}X_{t-1} + \phi_{4,2}X_{t-1} + \phi_{4,3}X_{t-1} + \phi_{4,4}X_{t-1} + Z_t
\end{aligned}$$

Thus the first four estimated partial autocorrelation coefficients for these data

$$\widehat{\pi}_1 = \phi_{1,1} \quad \widehat{\pi}_2 = \phi_{2,2} \quad \widehat{\pi}_3 = \phi_{3,3} \quad \widehat{\pi}_4 = \phi_{4,4}.$$

95% confidence interval $\pm 1.96\sqrt{n}$, where n is the number of the observations.

3.3 定阶步骤

说白了 ACF 和 PACF 如果在 95% CI (虚线) 以外, 那就是**显著 (significant)**, 否则就不显著. 因此,

- ACF 图的最后一个显著点是 q, 那就是 $\mathrm{MA}(q)$ (has only q significant lags);
- PACF 图的最后一个显著点是 p, 那就是 $\mathrm{AR}(p)$ (has only p significant lags);
- 指数下降就是**拖尾** (exponentially decreasing with a large number of significant lags);
- 如果 ACF 和 PACF **都显著 (没有骤降)**, 那就是 ARMA. 只能通过先拟合简单的模型, 然后再慢慢增加模型的复杂度, 直到残差体现出没有短期相关性 (short-term correlation), 即白噪声序列, 或称为纯随机序列 (purely random process).

规范说法

1. Verify that the time series is (weakly) stationary, i.e. that it has constant mean and variance. 确定宽平稳.
2. Examine the autocorrelation function and partial autocorrelation function. 检查 ACF 和 PACF.
   - If an $\mathrm{MA}(q)$ is appropriate,
     - the ACF will have significant values for lags 1 to q, and no significant values for lags greater than q; ACF 在滞后 1 到 q 处应有显著值, 在 $q+1$ 之后应不再显著;
     - the PACF will not indicate that a simpler AR model would be appropriate, i.e. have significant

values at lags 1 to p where $p < q$.

PACF 不应呈现明显的 $\mathrm{AR}(p)$ 模型特征（即不应在前 p 阶显著且后面迅速为 0）

- If an $\mathrm{AR}(p)$ is appropriate,
  - the PACF will have significant values for lags 1 to p, and no significant values for lags greater than p;
  - the ACF will not indicate that a simpler MA model would be appropriate, i.e. have significant values at lags 1 to q where $q < p$.

3. Estimate the chosen $\mathrm{MA}(q)$ or $\mathrm{AR}(p)$ models and examine its residuals to verify that they resemble a purely random process, that is, they contain no short-term correlation.
   估计所选的 $\mathrm{MA}(q)$ 或 $\mathrm{AR}(p)$ 模型，并检查其残差以验证它们类似于一个纯粹随机过程，也就是说，它们不包含短期相关性。

4 自相关差分移动平均 (Autoregressive Integrated Moving Average, ARIMA) 过程

4.1 定义

The time series X_t is an Autoregressive Integrated Moving Average process of order (p, d, q), denoted $\mathrm{ARIMA}(p, d, q)$ if the d-th order differenced process

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

is an $\mathrm{ARMA}(p, q)$ process. An $\mathrm{ARIMA}(p, d, q)$ process can be written most easily in terms of characteristic polynomials. If we write the $\mathrm{ARMA}(p, q)$ process for Y_t as

$$\phi(B) Y_t = \theta(B) Z_t$$

then as $Y_t = (1 - B)^d X_t$, an $\mathrm{ARIMA}(p, d, q)$ process can be written as

$$\phi(B)(1 - B)^d X_t = \theta(B) Z_t \quad \Leftrightarrow \quad \left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1 - B)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) Z_t$$

显而易见 $\mathrm{ARIMA}(p, 0, q) = \mathrm{ARMA}(p, q)$.

例 4.1 For an $\mathrm{ARIMA}(1,1,1)$ process the characteristic polynomials are given by $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$ and $\theta(B) = 1 + \theta B$, meaning that the full model is given by

$$
\begin{aligned}
(1 - \phi_1 B - \phi_2 B^2)(1 - B)X_t &= (1 + \theta B)Z_t \\
(1 - \phi_1 B - \phi_2 B^2 - B + \phi_1 B^2 + \phi_2 B^3)X_t &= (1 + \theta B)Z_t \\
X_t - (1 + \phi_1)BX_t + (\phi_1 - \phi_2)B^2 X_t + \phi_2 B^3 X_t &= Z_t + \theta B Z_t \\
X_t &= (1 + \phi_1)X_{t-1} - (\phi_1 - \phi_2)X_{t-2} - \phi_2 X_{t-3} + Z_t + \theta Z_{t-1}
\end{aligned}
$$

So an $\mathrm{ARIMA}(2,1,1)$ model is essentially a non-stationary $\mathrm{ARMA}(3,1)$ model.

Similarly, an $\mathrm{ARIMA}(p,d,q)$ model is essentially a non-stationary $\mathrm{ARMA}(p+d,q)$ model.

4.2 平稳性

由于 ARIMA 模型的 AR 部分的特征多项式 (characteristic polynomial) 为

$$\phi^*(B) = \phi(B)(1 - B)^d$$

有 d 个根为 1, 因此 $\mathrm{ARIMA}(p,d,q)$ 不可能平稳, 除非 $d = 0$.

5 预测

5.1 指数平滑法 (Exponential smoothing)

Given data $(x_1, \cdots, x_n)$, the one step ahead forecast using exponential smoothing is given by

$$\widehat{x}_n(1) = \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + \cdots + \alpha(1-\alpha)^{n-1}x_1$$

where $\alpha \in [0, 1]$ is a smoothing parameter.

Note that $c_i = \alpha(1-\alpha)^i$, and

$$\begin{aligned}
\sum_{i=0}^{\infty} c_i &= \sum_{i=0}^{\infty} \alpha(1-\alpha)^i \\
&= \alpha \sum_{i=1}^{\infty} (1-\alpha)^i \\
&= \alpha \times \frac{1 - \lim\limits_{n\to\infty}(1-\alpha)^n}{1-(1-\alpha)} \quad = 1
\end{aligned}$$

For finite n, the sum of the coefficients is approximately 1 because for large enough n, $c_n = \alpha(1-\alpha)^n \approx 0$.

The one step ahead forecast can be written recursively as follows

$$\begin{aligned}
\widehat{x}_n(1) &= \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + \cdots + \alpha(1-\alpha)^{n-1}x_1 \\
&= \alpha x_n + (1-\alpha)\big(\alpha x_{n-1} + \alpha(1-\alpha)x_{n-2} + \cdots + \alpha(1-\alpha)^{n-2}x_1\big) \\
&= \alpha x_n + (1-\alpha)\widehat{x}_{n-1}(1),
\end{aligned}$$

making it straightforward computationally to update the forecasts in light of new data. To start the process, we set $\widehat{x}_1(1) = x_2$.

For exponential smoothing it has been shown that an approximate 95% prediction interval for $x_n(1)$ is given by

$$\widehat{x}_n(1) \pm 1.96\sqrt{\mathrm{Var}\left[e_n(1)\right]},$$

where $\mathrm{Var}\left[e_n(1)\right]$ can be approximated as the variance of the forecast errors $e_1(1), e_2(1), \cdots, e_{n-1}(1)$, i.e.

$$\mathrm{Var}\left[e_n(1)\right] = \frac{1}{n-2}\sum_{i=1}^{n-1}\left(e_i(1) - \overline{e}\right),$$

with $\overline{e} = \dfrac{1}{n-1}\sum_{i=1}^{n-1} e_i(1)$.

例 5.1 Given the time series data $x = (1, 2, 4, 4, 3)$, perform one step ahead forecasting using the exponential smoothing method with $\alpha = 0.5$, and produce a point forecast $\widehat{x}_5(1)$.

解. Apparently, $x = (x_1, x_2, x_3, x_4, x_5) = (1, 2, 4, 4, 3)$

$\widehat{x}_5(1) = 0.5 \times 3 + 0.5(1-0.5) \times 4 + 0.5(1-0.5)^2 \times 4 + 0.5(1-0.5)^3 \times 2 + 0.5(1-0.5)^4 \times 1 = 3.15625.$

5.2 Forecasting from AR(p) Models

For an $\mathrm{AR}(p)$ process $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$, the one step ahead forecast is given by

$$\begin{aligned}
\widehat{x}_n(1) &= \mathbb{E}\left[X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[\phi_1 X_n + \phi_2 X_{n-1} + \cdots + \phi_p X_{n-p+1} + Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \phi_1 \mathbb{E}\left[X_n \mid \mathcal{F}_n\right] + \phi_2 \mathbb{E}\left[X_{n-1} \mid \mathcal{F}_n\right] + \cdots + \phi_p \mathbb{E}\left[X_{n-p+1} \mid \mathcal{F}_n\right] + \mathbb{E}\left[Z_{n+1} \mid \mathcal{F}_n\right] \\
&= \phi_1 x_n + \phi_2 x_{n-1} + \cdots + \phi_p x_{n-p+1},
\end{aligned}$$

where x_n is the observed value of the series at time n, and $\mathbb{E}\left[Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] = 0$.

Then for any k, the k steps ahead forecast is given by

$$\widehat{x}_n(k) = \mathbb{E}\left[X_{n+k} \mid X_n, X_{n-1}, \cdots, X_1\right].$$

注 1. If X_{n+kj} has been observed, then $\mathbb{E}\left[X_{n+k-j} \mid X_n, X_{n-1}, \cdots, X_1\right]$ is equal to its observed value, x_{n+kj}, 即已观测时间的值直接用观测值.

2. If X_{n+kj} is a future value, then $\mathbb{E}\left[X_{n+k-j} \mid X_n, X_{n-1}, \cdots, X_1\right]$ has already been forecast as one of $\widehat{x}_n(1), \widehat{x}_n(2), \cdots, \widehat{x}_n(k-1)$, 即未来时间的值需要用预测值.

例 5.2 (以 $\mathrm{AR}(2)$ 为例) Suppose the time series $(x_1, \cdots, x_n)$ is represented by an $\mathrm{AR}(2)$ process $X_t =$

$\phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$. Then the one step and two steps ahead forecast are given by

$$\begin{aligned}
\widehat{x}_n(1) &= \mathbb{E}\left[X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[\phi_1 X_n + \phi_2 X_{n-1} + Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \phi_1 \mathbb{E}\left[X_n \mid X_n, X_{n-1}, \cdots, X_1\right] + \phi_2 \mathbb{E}\left[X_{n-1} \mid X_n, X_{n-1}, \cdots, X_1\right] + \mathbb{E}\left[Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \phi_1 x_n + \phi_2 x_{n-1}, \\
\widehat{x}_n(2) &= \mathbb{E}\left[X_{n+2} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[\phi_1 X_{n+1} + \phi_2 X_n + Z_{n+2} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \phi_1 \mathbb{E}\left[X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] + \phi_2 \mathbb{E}\left[X_n \mid X_n, X_{n-1}, \cdots, X_1\right] + \mathbb{E}\left[Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \phi_1 \widehat{x}_n(1) + \phi_2 x_n \quad = \phi_1(\phi_1 x_n + \phi_2 x_{n-1}) + \phi_2 x_n \quad = (\phi_1^2 + \phi_2) x_n + \phi_1 \phi_2 x_{n-1}.
\end{aligned}$$

The forecast error variance at one step and two steps ahead are given by

$$\begin{aligned}
\operatorname{Var}\left[e_n(1)\right] &= \operatorname{Var}\left[X_{n+1} - \widehat{x}_n(1)\right] \\
&= \operatorname{Var}\left[\phi_1 X_n + \phi_2 X_{n-1} + Z_{n+1} - (\phi_1 X_n + \phi_2 X_{n-1})\right] \\
&= \operatorname{Var}\left[Z_{n+1}\right] \\
&= \sigma_z^2.
\end{aligned}$$

$$
\begin{aligned}
\operatorname{Var}\left[e_n(2)\right] &= \operatorname{Var}\left[X_{n+2} - \widehat{x}_n(2)\right] \\
&= \operatorname{Var}\left[\phi_1 X_{n+1} + \phi_2 X_n + Z_{n+2} - (\phi_1 \widehat{x}_n(1) + \phi_2 X_n)\right] \\
&= \operatorname{Var}\left[\phi_1 (X_{n+1} - \widehat{x}_n(1)) + Z_{n+2}\right] \\
&= \phi_1^2 \operatorname{Var}\left[X_{n+1} - \widehat{x}_n(1)\right] + \operatorname{Var}\left[Z_{n+2}\right] \\
&= \phi_1^2 \operatorname{Var}\left[e_n(1)\right] + \sigma_z^2 \\
&= (\phi_1^2 + 1)\sigma_z^2.
\end{aligned}
$$

The error variance for the k steps ahead forecast has the general form

$$
\operatorname{Var}\left[e_n(k)\right] = \operatorname{Var}\left[X_{n+k} - \widehat{x}_n(k)\right].
$$

Then 95% prediction intervals can be calculated as before using the formula

$$
\widehat{x}_n(k) \pm 1.96 \sqrt{\operatorname{Var}\left[e_n(k)\right]}.
$$

5.3 Forecasting from MA(q) Models

For an $\mathrm{MA}(q)$ process $X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$, the one step ahead forecast is given by

$$
\begin{aligned}
\widehat{x}_n(1) &= \mathbb{E}\left[X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[Z_{n+1} + \theta_1 Z_n + \cdots + \theta_q Z_{n-q+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[Z_{n+1} \mid \mathcal{F}_n\right] + \theta_1 \mathbb{E}\left[Z_n \mid \mathcal{F}_n\right] + \cdots + \theta_q \mathbb{E}\left[Z_{n-q+1} \mid \mathcal{F}_n\right] \\
&= \theta_1 z_n + \theta_2 z_{n-1} + \cdots + \theta_q z_{n-q+1},
\end{aligned}
$$

The last line is true because

- $X_n, X_{n-1}, \cdots, X_1$ do not depend on Z_{n+1}, and hence

$$\mathbb{E}\left[Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] = \mathbb{E}\left[Z_{n+1}\right] = 0.$$

- In contrast, X_n depends on $Z_n, Z_{n-1}, \cdots$ so

$$\mathbb{E}\left[Z_{n-i} \mid X_n, X_{n-1}, \cdots, X_1\right] \neq \mathbb{E}\left[Z_n\right] = 0, \quad i \geqslant 0.$$

Then for any k, the k steps ahead forecast is given by

$$
\begin{aligned}
\widehat{x}_n(k) &= \mathbb{E}\left[Z_{n+k} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \begin{cases} \theta_k z_n + \theta_{k+1} z_{n-1} + \cdots + \theta_q z_{n-q+k}, & k \leqslant q \\ 0, & k > q. \end{cases}
\end{aligned}
$$

例 5.3 (以 $\mathrm{MA}(2)$ 为例) Suppose the time series $(x_1, \cdots, x_n)$ is represented by an $\mathrm{MA}(2)$ process $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$. Then the one step and two steps ahead forecast are given by

$$\begin{aligned}
\widehat{x}_n(1) &= \mathbb{E}\left[X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[Z_{n+1} + \theta_1 Z_n + \theta_2 Z_{n-1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] + \theta_1 \mathbb{E}\left[Z_n \mid X_n, X_{n-1}, \cdots, X_1\right] + \theta_2 \mathbb{E}\left[Z_{n-1} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \theta_1 z_n + \theta_2 z_{n-1}, \\
\widehat{x}_n(2) &= \mathbb{E}\left[X_{n+2} \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[Z_{n+2} + \theta_1 Z_{n+1} + \theta_2 Z_n \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \mathbb{E}\left[Z_{n+2} \mid X_n, X_{n-1}, \cdots, X_1\right] + \theta_1 \mathbb{E}\left[Z_{n+1} \mid X_n, X_{n-1}, \cdots, X_1\right] + \theta_2 \mathbb{E}\left[Z_n \mid X_n, X_{n-1}, \cdots, X_1\right] \\
&= \theta_2 z_n.
\end{aligned}$$

The forecast error variance at one step and two steps ahead are given by

$$\begin{aligned}
\operatorname{Var}\left[e_n(1)\right] &= \operatorname{Var}\left[X_{n+1} - \widehat{x}_n(1)\right] \\
&= \operatorname{Var}\left[Z_{n+1} + \theta_1 Z_n + \theta_2 Z_{n-1} - (\theta_1 Z_n + \theta_2 Z_{n-1})\right] \\
&= \operatorname{Var}\left[Z_{n+1}\right] \\
&= \sigma_z^2.
\end{aligned}$$

$$
\begin{aligned}
\operatorname{Var}[e_n(2)] &= \operatorname{Var}[Z_{n+2} - \widehat{x}_n(2)] \\
&= \operatorname{Var}[Z_{n+2} + \theta_1 Z_{n+1} + \theta_2 Z_n - \theta_2 Z_n] \\
&= \operatorname{Var}[Z_{n+2} + \theta_1 Z_{n+1}] \\
&= (1+\theta_1^2)\sigma_z^2.
\end{aligned}
$$

It is straightforward to show that the k steps ahead error variance is given by

$$
\operatorname{Var}[e_n(k)] = \begin{cases} \left(1 + \sum_{i=1}^{k-1} \theta_i^2\right)\sigma_z^2, & k \leqslant q \\ \left(1 + \sum_{i=1}^{q} \theta_i^2\right)\sigma_z^2, & k > q \end{cases}
$$

Then 95% prediction intervals can be calculated as before using the formula

$$
\widehat{x}_n(k) \pm 1.96\sqrt{\operatorname{Var}[e_n(k)]}.
$$

6 干预模型 (Intervention Models)

有时，在特定时间影响正在建模的过程的外部驱动因素，使得时间序列在外部驱动因素作用前后表现不同。其中外部驱动因素通常被称为**干预 (intervention)**.

Consider the general model specification

$$Y_t = I_t + X_t$$

where

- Y_t is the time series to be modelled.
- I_t is the effect of the intervention.
- X_t is the underlying time series were there no intervention.
- The intervention occurs at time $t = T$. Clearly, the intervention model must satisfy

$$I_t = 0 \quad \text{for } t < T.$$

6.1 阶跃相应干预 (Step-response Interventions)

阶跃响应干预是基于 S_t^T 的一类干预效果模型。它们模拟具有**长期效果**的干预，即随着时间的推移而增加或保持恒定。最简单的是均值移动，其表达式为：

$$I_t = \omega S_t^T = \begin{cases} \omega, & t \geqslant T, \\ 0, & t < T, \end{cases} \quad \text{where } \textbf{step function } S_t^T = \begin{cases} 1, & t \geqslant T, \\ 0, & t < T, \end{cases}$$

which increases or decreases the series at and after time T by a constant ω.

A more general class of step-function models is given by

$$I_t = \delta I_{t-1} + \omega S_{t-1}^T$$

with the initial condition that $I_0 = 0$ and the restriction that $0 \leqslant \delta \leqslant 1$.

This model has an $\mathrm{AR}(1)$ type structure, where the intervention at time t depends on the intervention at time $t - 1$. Iterating this equation gives

$$\begin{aligned} I_1 &= 0 \\ &\vdots \\ I_T &= 0 \\ I_{T+1} &= \omega \\ I_{T+2} &= \delta I_{T+1} + \omega S_{T+1}^{T} = \delta\omega + \omega \\ I_{T+3} &= \delta I_{T+2} + \omega S_{T+2}^{T} = \delta(\delta\omega + \omega) + \omega = \delta^2\omega + \delta\omega + \omega \\ &\vdots \\ I_t &= \delta^{t-T-1}\omega + \delta^{t-T-2}\omega + \cdots + \delta\omega + \omega \end{aligned}$$

Yielding

$$I_t = \begin{cases} \omega \dfrac{1 - \delta^{t-T}}{1 - \delta}, & t > T \\ 0, & t \leqslant T. \end{cases}$$

6.2 Pulse-response Interventions

脉冲响应函数是一类基于 P_t^T 的干预效应模型。它们模拟对序列有**短暂影响**的干预措施，即它们具有初始的大效应，但随着时间的推移变得不那么明显。最简单的是单次均值漂移，它由以下给出：

$$I_t = \omega P_t^T = \begin{cases} \omega, & t = T, \\ 0, & t \neq T, \end{cases} \quad \text{where } \textbf{pulse function } P_t^T = \begin{cases} 1, & t = T, \\ 0, & t \neq T, \end{cases}$$

which increases or decreases the series at time T only by a constant ω.

A more general class of pulse-function models is given by

$$I_t = \delta I_{t-1} + \omega P_t^T$$

with the initial condition that $I_0 = 0$ and the restriction that $0 < \delta < 1$.

This model has an $\mathrm{AR}(1)$ type structure, where the intervention at time t depends on the intervention at time $t-1$. Iterating this equation gives

$$\begin{aligned} I_1 &= 0 \\ &\vdots \\ I_T &= \omega \\ I_{T+1} &= \delta I_T = \delta\omega \\ I_{T+2} &= \delta I_{T+1} = \delta^2\omega \\ &\vdots \end{aligned}$$

Yielding

$$I_t = \begin{cases} \delta^{t-T}\omega, & t > T \\ 0, & t \leqslant T. \end{cases}$$

7 Modelling trends and seasonal patterns

我们假设加性时间序列模型适用于我们的数据, 即

$$X_t = m_t + s_t + e_t,$$

如果不是，可以应用对数变换, 即

$$\log(X_t) = \log(m_t \cdot s_t \cdot e_t) = \log(m_t) + \log(s_t) + \log(e_t),$$

where m_t is **trend**, s_t is the **seasonal pattern** and e_t is the stationary **unexplained variation**.

The modelling approach is encompassed in the following two-stage process.

1. Estimate the trend and seasonal variation $\widehat{m}_t + \widehat{s}_t$.
2. Calculate the residual series $e_t^* = X_t - \widehat{m}_t - \widehat{s}_t$, and check that it is stationary. If not, return to (1).

7.1 Method 1: Regression

The idea is to represent the trend and seasonal variation as

$$m_t + s_t = \beta_0 + \beta_1 z_{t1} + \cdots + \beta_p z_{tp} = \boldsymbol{z}_t^\top \boldsymbol{\beta} \quad \Rightarrow \quad \boldsymbol{Z\beta},$$

where

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_n \end{pmatrix} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1p} \\ 1 & z_{21} & \cdots & z_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & z_{n1} & \cdots & z_{np} \end{pmatrix}_{n\times p}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p\times 1}.$$

then estimate the regression parameters using OLS

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{x}.$$

The estimated trend is now

$$\widehat{m}_t + \widehat{s}_t = \widehat{\beta}_0 + \widehat{\beta}_1 z_{t1} + \cdots + \widehat{\beta}_p z_{tp},$$

which can be subtracted from the raw time series to give a stationary residual series

$$e_t^* = X_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 z_{t1} + \cdots + \widehat{\beta}_p z_{tp}\right).$$

To check whether you have adequately modelled the trend and seasonal pattern in the data produce a time plot of the residual series $(e_1^*, e_2^*, \cdots, e_n^*)$, to check it is stationary.

Estimate and 95% Prediction Interval Consider the linear trend model

$$X_t = \beta_0 + \beta_1 z_{t1} + \cdots + \beta_p z_{tp} + \epsilon_t = Z\boldsymbol{\beta} + \epsilon_t,$$

where ϵ_t is a purely random process.

For the observation at $t = (*)$, i.e. $\boldsymbol{z}_* = (1, z_{*1}, \cdots, z_{*p})$, the prediction is given by

$$\mathbb{E}\left[\widehat{x}_*\right] = \boldsymbol{z}_*^\top \widehat{\boldsymbol{\beta}} = \widehat{\beta}_0 + \widehat{\beta}_1 z_{*1} + \cdots + \widehat{\beta}_p z_{*p},$$

and then

$$\begin{aligned}\operatorname{Var}\left[\widehat{x}_*\right] &= \operatorname{Var}\left[\boldsymbol{z}_*^\top \widehat{\boldsymbol{\beta}}\right] + \operatorname{Var}\left[\epsilon_*\right] = \boldsymbol{z}_*^\top \operatorname{Var}\left[\widehat{\boldsymbol{\beta}}\right] \boldsymbol{z}_* + \sigma^2 \\ &= \boldsymbol{z}_*^\top \sigma^2 (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{z}_* + \sigma^2 = \sigma^2\left[1 + \boldsymbol{z}_*^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{z}_*\right].\end{aligned}$$

The general formula for the prediction interval is given by

$$\boldsymbol{z}_*^\top \widehat{\boldsymbol{\beta}} \pm \sqrt{\sigma^2\left[1 + \boldsymbol{z}_*^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{z}_*\right]}.$$

例 7.1 Consider the linear trend model

$$X_t = \beta_0 + \beta_1 t + Z_t,$$

which was applied to 10 data points, $t = 1, \cdots, 10$. When the model was fitted in `R` it produced estimates $\widehat{\beta}_0 = 30$, $\widehat{\beta}_1 = 0.9$ and $\widehat{\sigma}^2 = 4$. So the prediction is given by

$$\boldsymbol{z}_{11}^\top \widehat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 11 \end{pmatrix}^\top \begin{pmatrix} 30 & 0.9 \end{pmatrix} = 39.9, \quad \boldsymbol{z}_{11}^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{z}_{11} = \frac{1}{10 \times 385 - 55^2} \begin{pmatrix} 1 \\ 11 \end{pmatrix} \begin{pmatrix} 385 & -55 \\ -55 & 10 \end{pmatrix} \begin{pmatrix} 1 & 11 \end{pmatrix}$$

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 10 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{Z}^\top \boldsymbol{Z} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 10 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 10 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{10} 1 & \sum_{i=1}^{10} i \\ \sum_{i=1}^{10} i & \sum_{i=1}^{10} i^2 \end{pmatrix} = \begin{pmatrix} 10 & 55 \\ 55 & 385 \end{pmatrix}$$

针对不同序列会有不同的合适的模型进行选择, 比如:

Trend - Polynomials

$$m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_q t^q.$$

Seasonality - Harmonics (谐波)

$$s_t = \beta_0 + \sum_{i=1}^{q} \left(\beta_{1i} \sin(\omega_i t) + \beta_{2i} \cos(\omega_i t)\right),$$

where ω can be determined by $\frac{2\pi}{\omega}$. 一般只需要写 $q = 1$ 的情况即可.

因此, 在构造可能的回归模型 (possible regression model) 时, 应该写类似于如下的形式:

$$X_t = \underbrace{\alpha_0 + \alpha_1 t + \alpha_2 t^2}_{\text{trend}} + \underbrace{\beta_0 + \beta_1 \sin(\omega t) + \beta_2 \cos(\omega t)}_{\text{seasonality}} + \underbrace{e_t}_{\text{remaining unexplained variation}}.$$

7.2 Method 2: Moving Average Smoothing

A Moving average smoother estimates the trend and seasonal variation at time t by averaging the current observation and the q either side. In other words, The moving average smother for a **mean function** $\mu_t = m_t + s_t$

is given by

$$\widehat{m}_t + \widehat{s}_t = \frac{1}{2q+1} \sum_{i=-q}^{q} x_{t-i},$$

where q acts as a smoothing parameter, with larger values resulting in smoother estimated trends. The residual series is then calculated by subtraction as with regression, that is

$$e_t^* = x_t - \frac{1}{2q+1} \sum_{i=-q}^{q} x_{t-i},$$

which should have no trend or seasonal variation if an appropriate value of q has been chosen.

Advantage

- 易于实现, 无需为 trend 和 seasonality 制定一个模型;
- 简单灵活, 只有一个参数 (平滑参数 q) 用于选择.

Disadvantage

- trend 和 seasonality 未单独估计;
- 时间序列缩短了 $2q$, mean function 可能无法很好地进行预测.

7.3 Method 3: Differencing

定义 7.1 (Backshift operator (滞后算子)) 记为 B, 有

$$BX_t = X_{t-1}, \quad B^2 X_t = X_{t-2}, \quad \cdots .$$

Removing trends by differencing (通过差分去除趋势)

定义 7.2 (First order difference (一阶差分)) 一阶差分算子记为 ∇, 有

$$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t.$$

同理, 定义 **general order difference operator (一般阶差分算子)** ∇^q 为

$$\nabla^q X_t = \nabla[\nabla^{q-1} X_t] = (1 - B)^q X_t$$

Removing seasonal variation by differencing (通过差分去除季节性变化)

定义 7.3 (Seasonal difference (季节差分)) d 阶季节差分算子 ∇_d 为

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d) X_t.$$

Removing trend and seasonal variation by differencing

例 7.2 $X_t = \beta_0 + \beta_1 t + s_t + e_t$, which has a **linear trend** and **seasonal component** that repeats itself every d time points, that is $s_t = s_{t-d} = s_{t+d}$.

解.

$$\begin{aligned}
\nabla_d X_t &= (1 - B^d)X_t \\
&= (\beta_0 + \beta_1 t + s_t + e_t) - B^d(\beta_0 + \beta_1 t + s_t + e_t) \\
&= (\beta_0 + \beta_1 t + s_t + e_t) - (\beta_0 + \beta_1 (t-d) + s_{t-d} + e_{t-d}) \\
&= d\beta_1 + e_t - e_{t-d}
\end{aligned}$$

例 7.3 $X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + s_t + e_t$, which has a **quadratic trend** and **seasonal component** with period d.

解.

$$\begin{aligned}
\nabla_d X_t &= (1 - B^d)X_t \\
&= (\beta_0 + \beta_1 t + \beta_2 t^2 + s_t + e_t) - B^d(\beta_0 + \beta_1 t + \beta_2 t^2 + s_t + e_t) \\
&= (\beta_0 + \beta_1 t + \beta_2 t^2 + s_t + e_t) - (\beta_0 + \beta_1 (t-d) + \beta_2 (t-d)^2 + s_{t-d} + e_{t-d}) \\
&= d\beta_1 + \beta_2 (2t + d)d + e_t - e_{t-d} \\
&= d\beta_1 + d^2\beta_2 + 2d\beta_2 t + e_t - e_{t-d}
\end{aligned}$$

which **still has a linear trend** in t and is hence not stationary. Applying a **first order difference** to this gives

$$\begin{aligned}\nabla(\nabla_d X_t) &= \nabla\left[d\beta_1 + d^2\beta_2 + 2d\beta_2 t + e_t - e_{t-d}\right] \\ &= (d\beta_1 + d^2\beta_2 + 2d\beta_2 t + e_t - e_{t-d}) - (d\beta_1 + d^2\beta_2 + 2d\beta_2(t-1) + e_{t-1} - e_{t-d-1}) \\ &= 2d\beta_2 + (e_t - e_{t-d}) - (e_{t-1} - e_{t-d-1})\end{aligned}$$

which does not depend on t and is hence stationary.

Advantage

- 简单, 只有一个参数 (差分阶数) 用于选择.

Disadvantage

- 差分不允许估计趋势, 只能去除它. 因此, 如果分析的目标是描述趋势, 则不合适.

7.4 Stabilise the Variance (稳定方差)

上面三节介绍了如何解决 trend (m_t) 和 seasonality (s_t) 的影响, 那么该如何解决残差 (方差) 不恒定的情况, 即 Var $[Z_t] \neq \sigma_Z^2$.

Two of the most common transformations in time series are **natural logarithm** and **square root**, but a more general class of transformations is called the **Box-Cox** transformation, named after two very famous statisticians, George Box and Sir David Cox.

定义 7.4 (Box-Cox transformation) Given an observed time series $\{x_t\}$, the Box-Cox transformation is given by

$$y_t = \begin{cases} \dfrac{x^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(x_t), & \lambda = 0. \end{cases}$$

where the transformation parameter λ is chosen by the time series analyst. Here $\lambda = 0$ corresponds to **natural log** while $\lambda = 0.5$ corresponds to **square root**.