# STAT 443 Final Exam Review

LaTeXer: W. Kong

## 1 Basic Definitions

**Definition 1.1.** The time series $\{X_t\}$ with $E[X_t^2] < \infty$ is said to be **weakly stationary** if:

1. $\mu_X(t) = E[X_t]$ is independent of $t$
2. $\gamma_X(t, t+h) = Cov(X_t, X_{t+h})$ is independent of $t$ for all $h$; the covariance only depends on the distance $h$ instead of $t$
3. $E[X_t^2] < \infty$ is also one of the conditions for weak stationarity.

**Definition 1.2.** Let $x_1, ..., x_n$ be observations of a time series. The **sample mean** of $x_1, ..., x_n$ is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. The **sample autocovariance function** is

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

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, h \in (-n, n)$$

## 2 Statistical Tests

The **Shapiro-Wilk Test** is as follows:

- $H_0 : Y_1, ..., Y_n$ come from a Gaussian distribution
- Reject $H_0$ if the $p$-value of this test is small
- In R, if the data is stored in the vector $y$, then use the command *shapiro.test(y)*.

The **Difference Sign Test** is as follows:

- Count the number $S$ of values such that $y_i - y_{i-1} > 0$
- For large i.i.d. sequences
$$\mu_S = E[S] = \frac{n-1}{2}, \sigma_S^2 = \frac{n+1}{12}$$
- For large $n$, $S$ is approximately $N(\mu_S, \sigma_S^2)$, therefore,
$$W = \frac{S - \mu_S}{\sqrt{\sigma_S^2}} \sim N(0, 1)$$
- A large positive value of $S - \mu_S$ indicates the presence of increasing (decreasing) trend
- We reject ($H_0 :$ data is random) if $|W| > z_{1-\alpha/2}$ but this may not work for seasonal data

The **Runs Test** is as follows:

- Estimate the median and call it $m$
- Let $n_1$ be the number of observations $> m$ and $n_2$ be the number of observations $< m$
- Let $R$ be the number of consecutive observations which are all smaller (larger) than $m$
- For large i.i.d. sequences

$$\mu_R = E[R] = 1 + \frac{2n_1 n_2}{n_1 + n_2}, \sigma_R^2 = \frac{(\mu_R - 1)(\mu_R - 2)}{n_1 + n_2 - 1}$$

- For large number of observations,

$$\frac{R - \mu_R}{\sigma_R} \sim N(0,1)$$

# 3 Filters and Smoothing

1. (Finite Moving Average Filter) Let $q$ be a non-negative integer and consider the two-sided moving average of the series $X_t$. We have

$$m_t \approx \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j} = \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t-j} + \underbrace{\frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t-j}}_{\approx 0}$$

2. (Exponential Smoothing) For fixed $\alpha \in [0,1]$ define the recursion

$$\hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1}$$

with initial condition $\hat{m}_1 = X_1$. This gives an exponentially decreasing weighted moving average where in the general $t \geq 2$ case,

$$\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1$$

Note that a smaller $\alpha$ creates a smoother plot compared to a larger $\alpha$.

3. (Polynomial Regression) This is just developing a parametric polynomial form of $m_t$ in the form

$$m_t = \sum_{i=0}^{k} \beta_i t^i$$

where $k$ is chosen arbitrarily.

4. We can also eliminate the trend through **differencing** where

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

and $\nabla, B$ are known to be the differencing and backshift operators respectively. Exponentiating these operators is equivalent to function composition. In this case, we are applying differencing to get a stationary process (by eliminating the trend).

Holt-Winters (Special Cases)

- In the case that $\beta = \gamma = 0$ we have no trend or seasonal updates in the H-W algorithm

- Here, we have $L_t = \alpha X_t + (1-\alpha)L_{t-1}$ which is **exactly (simple) exponential smoothing under** $\alpha$
- In the case that $\gamma = 0$ we have no seasonal component and there are two H-W equations for updating $L_t$ and $T_t$
- We call the above case **double exponential smoothing**

# 4 Linear Processes

**Definition 4.1.** A process $\{X_t\}$ is called a **moving average process of order** $q$ if

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta_1, ..., \theta_q$ are constants. Sometimes $Z_t$ is referred to as the **innovation**. Notice that these innovations are uncorrelated, have constant variance and zero mean. Deriving the mean and autocovariance function of $MA(q)$, it is easy to see that this process is stationary.

**Definition 4.2.** We say that a process $\{X_t\}$ is $q-$**dependent** if $X_t$ and are $X_s$ are independent if $|t-s| > q$. That is, they are dependent if they are within $q$ steps of each other. Similarly, we saay that that stationary time series is $q-$**correlated** if $\gamma(h) = 0$ whenever $|h| > q$.

**Example 4.1.** It is easy to show that the $MA(q)$ process is $q-$correlated. The inverse of this statement is also true.

**Proposition 4.1.** *If $\{X_t\}$ is a stationary $q-$correlated time series with mean 0, then it can be represented as the $MA(q)$ process.*

**Definition 4.3.** process $\{X_t\}$ is called a **autoregressive process of order** $p$ if

$$X_t = X_t + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + Z_t$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\phi_1, ..., \phi_p$ are constants.

**Definition 4.4.** $\{X_t\}$ is called a **Gaussian time series** if all its joint distributions are multivariate normal. That is for any set $i_1, ..., i_m$ with each $n \in \mathbb{N}$, the random vector $(X_{i_1}, ..., X_{i_m})$ follows a multivariate normal distribution.

**Example 4.2.** Consider the stationary Gaussian time series $\{X_t\}$. Suppose $X_n$ has been observed and we want to forecast $X_{t+h}$ using $m(X_n)$, a function of $X_n$. Let us measure the quality of the forecast by

$$MSE = E\left( [X_{n+h} - m(X_n)]^2 \,|X_n \right)$$

It can be shown that $m(\cdot)$ which minimizes MSE in a general case is $m(X_n) = E(X_{n+h}|X_n)$.

**Example 4.3.** We now consider the problem of predicting $X_{n+h}, h > 0$ for a stationary time series with known mean $\mu$ and ACVF $\gamma(\cdot)$ based on previous values $\{X_n, ..., X_1\}$ showing the linear predictor of $X_{n+h}$ by $P_n X_{n+h}$. We are interested in

$$P_n X_{n+h} = a_0 + a_1 X_n + a_2 X_{n-1} + ... + a_n X_1$$

which minimizes

$$S(a_0, ..., a_n) = E\left[ (X_{n+h} - P_n X_{n+h})^2 \right]$$

To get $a_0, a_1, ..., a_n$ we need to solve the system $\frac{\partial S}{\partial a_j} = 0$ for $j = 0, 1, ..., n$. Doing so, we get

$$a_0 = \mu \left( 1 - \sum_{i=1}^{n} a_i \right), \Gamma_n a_n = \gamma_n(h)$$

where

$$a_n = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}, \gamma_n(h) = \begin{pmatrix} \gamma(h) \\ \gamma(h+1) \\ \vdots \\ \gamma(n+h-1) \end{pmatrix}$$

Here, $a_n = \Gamma_n^{-1}\gamma_n(h) \implies a_n = \frac{\Gamma_n^{-1}}{\gamma(0)} \cdot \rho_n(h)$ where $\rho_n(h) = \frac{\gamma_n(h)}{\gamma(0)}$.

*Note* 1. Here are some properties from the above:

- $P_n X_{n+h}$ is defined by $\mu, \gamma(h)$
- $P_n X_{n+1} = \mu + \sum_{i=1}^{n} a_i(X_{n+1-i} - \mu)$
- It can be shown that $E\left[(X_{n+h} - P_n X_{n+h})^2\right] = \gamma(0) - a_n^T \gamma_n(h)$
- $E(X_{n+h} - P_n X_{n+h}) = 0$
- $E\left[(X_{n+h} - P_n X_{n+h})X_j\right] = 0$ for $j = 1, 2, ..., n$

**Example 4.4.** Derive the one-step prediction for the $AR(1)$ model. (Here, $h = 1$)

To find the linear predictor, we need to solve

$$\begin{aligned} \Gamma_n a_n = \gamma_n(h) \quad &\implies \quad \frac{\Gamma_n a_n}{\gamma(0)} = \frac{\gamma_n(h)}{\gamma(0)} \\ &\implies \quad \begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{pmatrix} \end{aligned}$$

An obvious solution is

$$\begin{aligned} a_n = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad &\implies \quad P_n X_{n+1} = \mu + \sum_{i=1}^{n} a_i(X_{n+1-i} - \mu) \\ &\implies \quad P_n X_{n+1} = \sum_{i=1}^{n} a_i X_{n+1-i} = a_1 X_n + 0 = \phi X_n \end{aligned}$$

You can use the formula of MSE to get

$$\begin{aligned} MSE &= \gamma(0) - a_n^T \gamma_n(h) \\ &= \gamma(0) - \phi\gamma(1) \\ &= \gamma(0) - \phi^2\gamma(0) \\ &= \gamma(0)[1 - \phi^2] = \sigma^2 \end{aligned}$$

# 5 Causal and Invertible Processes

**Definition 5.1.** The time series $\{X_t\}$ is a **linear process** if $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ for all $t$ where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\psi_j$ is a sequence of constants such that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

**Example 5.1.** Show that $AR(1)$ with $|\phi| < 1$ is a **linear process**. We know that

$$X_t = \phi X_{t-1} + \underbrace{Z_t}_{\sim WN(0,\sigma^2)}$$

and we showed before that $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$. Since $|\phi| < 1$ then if $\psi_j = \phi^j$ then $\sum_{j=-\infty}^{\infty} |\psi_j|$ and therefore all assumptions in the definition above are satisfied. So $AR(1)$ is a linear process.

**Definition 5.2.** A linear process $\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ is **causal** or **future independent** if $\psi_j = 0$ for any $j < 0$.

$\{X_t, t \in T\}$ is an $ARMA(p,q)$ process if

1) $\{X_t, t \in T\}$ is stationary

2) $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$ where $\{Z_t\} \sim WN(0, \sigma^2)$

3) Polynomials $(1 - \phi_1 z - ... - \phi_p z^p)$ and $(1 + \theta_1 z + ... + \theta_q z^q)$ have no common factors/roots (IMPORTANT FOR THE FINAL!)

**Definition 5.3.** An $ARMA(p,q)$ process $\phi(B)X_t = \theta(B)Z_t$ where $Z_t \sim WN(0, \sigma^2)$ is **causal** if there exists constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ for any $t$. This condition is equivalent to

$$\phi(z) = 1 - \phi_1 z_1 - \phi_2 z^2 - ... - \phi_p z^p \neq 0$$

for any $z \in \mathbb{C}$ such that $|z| \leq 1$.

*Remark* 5.1. f the condition above holds true, then

$$\begin{aligned} \frac{\theta(z)}{\phi(z)} = \psi(z) \quad &\implies \quad \theta(z) = \phi(z) \cdot \psi(z) \\ &\implies \quad 1 + \theta_1 z + ... + \theta_q z^q = (1 - \phi_1 z - ... - \phi_p z^p)(\psi_0 + \psi_1 z + ...) \end{aligned}$$

and we have

$$\begin{aligned} 1 &= \psi_0 \\ \theta_1 &= \psi_1 - \phi_1 \psi_0 \\ &\vdots \end{aligned}$$

**Definition 5.4.** An $ARMA(p,q)$ process $\{X_t\}$ is **invertible** if there exists constants $\{\Pi_j\}$ such that $\sum_{j=0}^{\infty} |\Pi_j| < \infty$ and $Z_t = \sum_{j=0}^{\infty} \Pi_j X_{t-j}$ for all $t$. Invertibility is equivalent to the condition

$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q \neq 0$$

for any $z \in \mathbb{C}$ such that $|z| \leq 1$. Using the same methods above, one can get that

$$\begin{aligned} \Pi_0 &= 1 \\ -\phi_1 &= \Pi_0 \theta_1 + \Pi_1 \\ &\vdots \end{aligned}$$

**Example 5.2.** Consider $\{X_t, t \in T\}$ satisfying $X_t - 0.5X_{t-1} = Z_t + 0.4Z_{t-1}$ where $\{Z_t\} \sim WN(0, \sigma^2)$. Investigate the causality and invertibility of $X_t$. If the series is causal (invertible) then provide the causal (invertible) solutions. These are called the $MA(\infty)$ and $AR(\infty)$ representations.

[Causality] We have $\phi(z) = 1 - 0.5z \implies z = 2 \implies |z| > 1$. Since this is outside the unit circle, $X_t$ is causal. We then have

$$\begin{aligned} 1 + 0.4z = (1 - 0.5z)(\psi_0 + \psi_1 z + ...) \quad &\implies \quad \psi_0 = 1, \psi_1 - 0.5\psi_0 = 0.4, \psi_2 - 0.5\psi_1 = 0, ... \\ &\implies \quad \psi_0 = 1, \psi_1 = 0.9, \psi_2 = 0.9(0.5), \psi_3 = 0.9(0.5)^2, ... \end{aligned}$$

We can kind of see the pattern (and prove using induction)

$$\psi_j = \begin{cases} \psi_j = 1 & j = 0 \\ \psi_j = 0.9(0.5)^{j-1} & j \neq 0 \end{cases} \implies X_t = Z_t + 0.9\sum_{j=1}^{\infty}(0.5)^{j-1}Z_{t-j}$$

[Invertibility] We have $\theta(z) = 1 + 0.4z = 0 \implies z = -10/4 \implies |z| > 1$. Since this is outside the unit circle, $X_t$ is invertible. We then have, like above,

$$\begin{aligned} 1 - 0.5z = (1 + 0.4z)(\Pi_0 + \Pi_1 z + ...) &\implies \Pi_0 = 1, \Pi_1 + 0.4\Pi_0 = -0.5, \Pi_2 + 0.4\Pi_1 = 0, ... \\ &\implies \Pi_0 = 1, \Pi_0 = -0.9, \psi_2 = -0.9(-0.4), \psi_3 = -0.9(-0.4)^2, ... \end{aligned}$$

We can kind of see the pattern (and prove using induction)

$$\psi_j = \begin{cases} \psi_j = 1 & j = 0 \\ \psi_j = -0.9(-0.4)^{j-1} & j \neq 0 \end{cases} \implies X_t = Z_t - 0.9\sum_{j=1}^{\infty}(-0.4)^{j-1}Z_{t-j}$$

*Remark* 5.2. (ACVF of ARMA processes) Consider a causal, stationary process $\phi(B)X_t = \theta(B)Z_t$ with $Z_t \sim WN(0, \sigma^2)$. The $MA(\infty)$ representation of $X_t$ is $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ where $E[X_t] = 0$. We have

$$\begin{aligned} \gamma(h) &= E[X_t X_{t+h}] - \underbrace{E[X_t]E[X_{t+h}]}_{=0} \\ &= E\left[\left(\sum_{j=0}^{\infty}\psi_j Z_{t-j}\right)\left(\sum_{j=0}^{\infty}\psi_j Z_{t+h-j}\right)\right] \end{aligned}$$

Notice that $E[Z_t Z_s] = 0$ when $t \neq s$. We then have

$$\gamma(h) = \begin{cases} \sum_{j=0}^{\infty}\psi_j\psi_{j+h}E[Z_j^2] & h \geq 0 \\ \sum_{j=0}^{\infty}\psi_j\psi_{j-h}E[Z_j^2] & h < 0 \end{cases} = \sigma^2\sum_{j=0}^{\infty}\psi_j\psi_{j+|h|}$$

**Example 5.3.** Derive the ACVF for the following $ARMA(1, 1)$ process

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

where $Z_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$. Note that $\phi(z)$ is causal because $1 - \phi z = 0 \implies z = 1/\phi > 1$. It can be shown, with similar methods above, that

$$\psi_j = \begin{cases} \psi_j = \phi(\phi + \theta) & j = 0 \\ \psi_j = \phi^{j-1}(\phi + \theta) & j \neq 0 \end{cases}$$

Now if $h = 0$ then

$$\begin{aligned} \gamma(0) = \sigma^2\sum_{j=0}^{\infty}\psi_j^2 &= \sigma^2\left[1 + \sum_{j=1}^{\infty}\psi_j^2\right] \\ &= \sigma^2\left[1 + (\theta + \phi)^2\sum_{j=1}^{\infty}\phi^{2(j-1)}\right] \\ &= \sigma^2\left[1 + (\theta + \phi)^2\sum_{i=0}^{\infty}\phi^{2i}\right] \\ &= \sigma^2\left[1 + \frac{(\theta + \phi)^2\phi}{1 - \phi^2}\right] \end{aligned}$$

If $h \neq 0$ then

$$\begin{aligned}
\gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} &= \sigma^2 \left[ \psi_0 \psi_{|h|} + \sum_{j=1}^{\infty} \psi_j \psi_{j+|h|} \right] \\
&= \sigma^2 \left[ \phi^{|h|-1}(\theta + \phi) + (\theta + \phi)^2 \sum_{j=1}^{\infty} \phi^{j-1} \phi^{j+|h|} \right] \\
&= \sigma^2 \left[ \phi^{|h|-1}(\theta + \phi) + (\theta + \phi)^2 \phi^{|h|-1} \sum_{j=1}^{\infty} \phi^{2j} \right] \\
&= \sigma^2 \left[ \phi^{|h|-1}(\theta + \phi) + \frac{(\theta + \phi)^2 \phi^{|h|+1}}{1 - \phi^4} \right]
\end{aligned}$$

*Summary* 1. For ACF and PACF, we have the following summary:

| | ACF | PACF |
|---|---|---|
| $MA(q)$ | Zero after lag $q$ | Decays exponentially |
| $AR(p)$ | Decays exponentially | Zero after lag $p$ |

In the general case of ARMA processes, the PACF is defined as $\alpha(0) = 1$ and $\alpha(h) = \Phi_{hh}$ for $h \geq 1$ where $\Phi_{hh}$ is the last component of the vector $\Phi_h = \Gamma_h^{-1} \gamma_h$ in which

$$\Gamma_h = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(h-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \cdots & \gamma(0) \end{pmatrix}, \gamma_h = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(h) \end{pmatrix}$$

**Example 5.4.** Calculate $\alpha(2)$ for an $MA(1)$ process

$$X_t = Z_t + \theta Z_{t-1}, \{Z_t\} \sim WN(0, \sigma^2)$$

We have shown before that

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2 & h = 0 \\ \theta\sigma^2 & h = 1 \\ 0 & h \geq 2 \end{cases}$$

We have $\Phi = \Gamma_h^{-1} \gamma_h$. So $\alpha(h)$ is the last element of $\Phi_h$ and

$$\begin{aligned}
h = 1 \quad &\Longrightarrow \quad \Phi_{11} = (\gamma(0))^{-1} \gamma(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1 + \theta^2} \\
h = 2 \quad &\Longrightarrow \quad \begin{pmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} \theta\sigma^2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\theta(1+\theta^2)\sigma^4}{(1+\theta^2)^2\sigma^4 - \theta^2\sigma^4} \\ \frac{-\theta\sigma^2}{(1+\theta^2)^2\sigma^4 - \theta^2\sigma^4} \end{pmatrix}
\end{aligned}$$

Where the last element of the case of $h = 2$, in reduced form, is

$$\alpha(2) = \Phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4}$$

It can be shown, in general, that

$$\alpha(h) = \Phi_{hh} = \frac{-(-\theta)^h}{\sum_{i=0}^{h} \theta^{2h}}$$

# 6 ARIMA/SARIMA Models

**Definition 6.1.** Let $d$ be a non-negative integer. $\{X_t, t \in T\}$ is an $ARIMA(p,d,q)$ process if $Y_t = (1-B)^d X_t$ is a causal $ARMA(p,q)$ process. The definition above means that $\{X_t, t \in T\}$ satisfies an equation of the form

$$\phi^*(B)X_t \equiv \phi(B)(1-B)^d X_t = \theta(B)Z_t, \{Z_t\} \sim WN(0,\sigma^2)$$

**Note that** $\phi^*(1) = 0 \implies X_t$ **is not stationary unless** $d = 0$. Therefore, $\{X_t\}$ is stationary iff $d = 0$ in which case it is reduced to an $ARMA(p,q)$ process in the previous case.

Recall that if $\{X_t\}$ exhibits a polynomial trend of the form $m(t) = \alpha_0 + \alpha_1 t + ... + \alpha_d t^d$ then $(1-B)^d X_t$ will not have that trend any more. Therefore, ARIMA models (when $d \neq 0$) are appropriate when the trend in the data is well approximated by a polynomial degree $d$.

Recall the operator $B$ where $B^k X_t = X_{t-k}$. Clearly $(1-B^k)$ and $(1-B)^k$ are different filters. The latter is performing $k$ times differencing, but the former is differencing once in lag $k$. In R, we will write

$$\begin{aligned} \text{diff(x,difference=k)} &\equiv (1-B)^k X_t \\ \text{diff(x,lag=k)} &\equiv (1-B^k) X_t \end{aligned}$$

**Definition 6.2.** If $d, D$ are non-negative integers, then $\{X_t t \in T\}$ is a seasonal $ARIMA(p,d,q) \times (P,D,Q)_S$ process with period $S$ if the differenced series

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

is a causal ARMA process defined by

$$\phi(B)\Phi(B^S)Y_t = \theta(B)\Theta(B^S)Z_t, Z_t \sim WN(0,\sigma^2)$$

*Remark* 6.1. Notice that the process $\{X_t, t \in T\}$ is causal iff $\phi(z) \neq 0 \wedge \Phi(z) \neq 0$ for all $\forall z : |z| < 1$.

**Example 6.1.** Derive the ACF of $SARIMA(0,0,1)_{12} = SARIMA(0,0,0) \times (0,0,1)_{12}$. This gives us the general form

$$X_t = Z_t + \Theta_1 Z_{t-12}, Z_t \sim WN(0,\sigma^2)$$

Show, as an exercise, that

$$\gamma(h) = Cov(X_t, X_{t+h}) = \begin{cases} (1+\Theta_1^2)\sigma^2 & h = 0 \\ \Theta_1 \sigma^2 & h = 12 \\ 0 & \text{otherwise} \end{cases}$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1 & h = 0 \\ \frac{\theta}{1+\theta^2} & h = 12 \\ 0 & \text{otherwise} \end{cases}$$

**Definition 6.3.** Consider a causal $AR(p)$ model

$$(1)\ X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t$$

with causal solution $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ where $\{Z_t\} \sim WN(0,\sigma^2)$. Multiply both sides of (1) by $X_{t-j}$ with $j = 0,1,2,...,p$ and taking expectations will give us

$$E[X_t X_{t-j}] - \phi_1 E[X_{t-1}X_{t-j}] - ... - \phi_p E[X_{t-p}X_{t-j}] = E[Z_t X_{t-j}]$$

$$\implies \gamma(j) - \phi_1 \gamma(j-1) - ... - \phi_p \gamma(j-p) = E[Z_t X_{t-j}]$$

We then have

$$\begin{cases} E[Z_tX_{t-j}] = E[Z_tX_t] = E\left[Z_t \sum_{j=0}^{\infty} \psi_j Z_{t-j}\right] = E[Z_t^2] = \sigma^2 & j = 0 \\ E[Z_tX_{t-j}] = 0 & j > 0 \end{cases}$$

So the original equation reduces to

$$\begin{cases} \gamma(0) - \phi_1\gamma(1) - ... - \phi_p\gamma(p) = \sigma^2 & j = 0 \\ \gamma(j) - \phi_1\gamma(|j-1|) - .... - \gamma(|j-p|) = 0 & j \neq 0 \end{cases}$$

These are called the **Yule-Walker equations**. This can be easily generalized to a matrix form $\Gamma_p\phi = \gamma_p$. Based on a sample $\{x_1, x_2, ..., x_n\}$ the parameters $\phi$ and $\sigma^2$ can be estimated by

$$\hat{\phi} = \hat{\Gamma}_p^{-1}\hat{\gamma}_p$$

where the matrices are defined in a similar fashion as the best linear predictor section. The system above is called the **sample Yule-Walker equations**. We can write Yule-Walker equations in terms of ACF too.

Explicitly, if we divide $\hat{\gamma}_p$ by $\gamma(0)$ and multiply it in $\hat{\Gamma}_p$ then

$$\hat{\phi} = \hat{R}_p^{-1}\hat{\rho}_p$$

$$\hat{R}_p = \frac{\hat{\Gamma}_p}{\hat{\gamma}(0)} \implies \hat{R}_p^{-1} = \hat{\Gamma}_p^{-1} \cdot \hat{\gamma}(0)$$

$$\hat{\rho}_p = \hat{\gamma}_p/\hat{\gamma}(0)$$

where $\hat{\sigma}^2 = \hat{\gamma}(0)\left[1 - \hat{\phi} \cdot \hat{\rho}_p\right]$. Notice that $\hat{\gamma}(0)$ is the sample variance of $\{x_1, ..., x_n\}$. Based on a sample $\{x_1, ..., x_n\}$, the above equations will provide the parameter estimates. Using advanced probability theory, it can be shown that

$$\tilde{\phi} = \begin{bmatrix} \tilde{\phi}_1 \\ \vdots \\ \tilde{\phi}_p \end{bmatrix} \sim MVN\left(\phi = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}, \frac{\sigma^2}{n}\Gamma_p^{-1}\right)$$

for large $n$. If we replace $\sigma^2$ and $\Gamma_p$ by their sample estimates $\hat{\sigma}^2$ and $\hat{\Gamma}_p$ we can use this result for large-sample confidence intervals for the parameters $\phi_1, ..., \phi_p$.

**Example 6.2.** Based on the following sample ACF and PACF, an $AR(2)$ has been proposed for the data. Provide the Yule-Walker estimates of the parameters as well as 95% confidence intervals for the parameters in $\phi$. The data was collected over a window of 200 points with sample variance 3.69 with the following table:

| $h$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $\hat{f}(h)$ | 1 | 0.821 | 0.764 | 0.644 | 0.586 | 0.49 | 0.411 | 0.354 |
| $\hat{\alpha}(h)$ | 1 | 0.821 | 0.277 | -0.121 | 0.052 | -0.06 | -0.072 | - |

We want to estimate $\phi_1$ and $\phi_2$ in

$$X_t = \phi_1X_{t-1} + \phi_2X_{t-2} + Z_t, \{Z_t\} \sim N(0, \sigma^2)$$

The system is

$$\hat{\phi} = \begin{bmatrix} 1 & 0.821 \\ 0.821 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.821 \\ 0.764 \end{bmatrix} = \begin{bmatrix} 0.594 \\ 0.276 \end{bmatrix}$$

Similarly,

$$\hat{\sigma}^2 = \underbrace{\hat{\gamma}(0)}_{3.69}\left[1 - \hat{\phi}\begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \end{bmatrix}\right] = 1.112$$

Therefore the estimated model is

$$X_t = 0.594X_{t-1} + 0.276X_{t-1} + Z_t, \{Z_t\} \sim WN(0, 1.112)$$

Now

$$\begin{aligned}
\tilde{\phi} \sim N\left(\phi, \frac{\sigma^2}{n}\Gamma_2^{-1}\right) &= N\left(\begin{bmatrix} 0.594 \\ 0.276 \end{bmatrix}, \frac{1.112}{200}\begin{bmatrix} 0.831 & -0.683 \\ -0.683 & 0.831 \end{bmatrix}\right) \\
&= N\left(\begin{bmatrix} 0.594 \\ 0.276 \end{bmatrix}, \begin{bmatrix} 0.005 & -0.004 \\ -0.004 & 0.005 \end{bmatrix}\right)
\end{aligned}$$

So the 95% C.I.'s for $\phi_1, \phi_2$ are

$$\begin{aligned}
\hat{\phi}_1 \pm 1.96\sqrt{\hat{Var}(\tilde{\phi})} &= 0.594 \pm 1.96\sqrt{0.005} = & (0.455, 0.733) \\
\hat{\phi}_2 \pm 1.96\sqrt{\hat{Var}(\tilde{\phi})} &= 0.276 \pm 1.96\sqrt{0.005} = & (0.137, 0.415)
\end{aligned}$$

# 7 Forecasting

We discuss how forecasting works under our studied processes.

## 7.1 Forecasting AR(p)

Let $X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + Z_t, Z_t \sim WN\{0, \sigma^2\}$ be a causal $AR(p)$ process. We have

$$\begin{aligned}
\hat{X}_{n+h} &= E[X_{n+h}|X_1, ..., X_n], h > 0 \\
&= E\left[\sum_{j=1}^{h-1} \phi_j X_{n+h-j} + \sum_{j=h}^{p} \phi_j X_{n+h-j}|X_1, ..., X_n\right] + \underbrace{E\left[Z_{n+h}|X_1, ..., X_n\right]}_{=0} \\
&= E\left[\sum_{j=1}^{h-1} \phi_j X_{n+h-j}|X_1, ..., X_n\right] + E\left[\sum_{j=h}^{p} \phi_j X_{n+h-j}|X_1, ..., X_n\right]
\end{aligned}$$

due to the uncorrelatedness of $Z_{n+h}$ with respect to $X_k$. If $h = 1$, then the above equation becomes

$$\hat{X}_{n+1} = \sum_{j=1}^{p} \phi_j X_{n+1-j}$$

If $h = 2, 3, ..., p$ then remark that

$$\begin{aligned}
j < h &\implies & n + h - j > n \\
j \geq h &\implies & n + h - j \leq n
\end{aligned}$$

and so

$$\begin{aligned}
\hat{X}_{n+h} &= \sum_{j=h}^{p} \phi_j X_{n+h-j} + \sum_{j=1}^{h-1} \phi_j E\left(X_{n+h-j}|X_1, ..., X_n\right) \\
&= \sum_{j=1}^{h-1} \phi_j \hat{X}_{n+h-j} + \sum_{j=h}^{p} \phi_j X_{n+h-j}
\end{aligned}$$

If $h > p$, then $n + h - j > n$ and

$$\hat{X}_{n+h} = \sum_{j=1}^{p} \phi_j E\left(X_{n+h-j}|X_1, ..., X_n\right) = \sum_{j=1}^{p} \phi_j \hat{X}_{n+h-j}$$

In summary, for a causal $AR(p)$, the $h-$step predictor is

$$\hat{X}_{n+h} = \begin{cases} \hat{X}_{n+1} = \sum_{j=1}^{p} \phi_j X_{n+1-j} & h = 1 \\ \sum_{j=1}^{h-1} \phi_j \hat{X}_{n+h-j} + \sum_{j=h}^{p} \phi_j X_{n+h-j} & h = 2, 3, ..., p \\ \sum_{j=1}^{p} \phi_j \hat{X}_{n+h-j} & h > p \end{cases}$$

In $AR(p)$, the $h-$step prediction is a linear combination of the previous steps. We either have the previous $p$ steps in $X_1, ..., X_n$ so we substitute the values (like the $h = 1$ case), or we don't have all or some of them, in which case we recursively predict.

Given a dataset, $\phi_j$ can be estimated and $\hat{X}_{n+h}$ will be computed.

**Example 7.1.** Based on the annual sales data of a chain store, an $AR(2)$ model with parameters $\hat{\phi}_1 = 1$ and $\hat{\phi}_2 = -0.21$ has bee fitted. If the total sales of the last 3 years have been 9, 11 and 10 million dollars. Forecast this year's total sales (2013) as well as that of 2015.

We have

$$X_t = X_{t-1} - 0.21X_{t-2} + Z_t, \{Z_t\} \sim WN(0, \sigma^2)$$

Now

$$\begin{aligned} \hat{X}_{2013} &= X_{2012} - 0.21X_{2011} = 6.69 \\ \hat{X}_{2015} &= \hat{X}_{2014} - 0.21\hat{X}_{2013} = \hat{X}_{2014} - 0.21(6.69) \end{aligned}$$

and since

$$\hat{X}_{2014} = \hat{X}_{2013} - 0.21\hat{X}_{2012} = 6.69 - 0.21 \times 9 = 4.8$$

then

$$\hat{X}_{2015} = 4.8 - 0.21(6.69) = 3.4$$

## 7.2 Forecasting MA(q)

MA processes are linear combinations of white noise. To do forecasting in $MA(q)$, we need to estimate $\theta_1, ..., \theta_q$ as well as "approximate" the innovations $Z_t, Z_{t+1}, ....$ First, consider the very simple case of $MA(1)$ where $X_t = Z_t + \theta Z_{t-1}, \{Z_t\} \sim WN(0, \sigma^2)$. We have

$$\begin{aligned} \hat{X}_{n+h} &= E\left[X_{n+h}|X_1, ..., X_n\right] \\ &= E\left[Z_{n+h}|X_1, ..., X_n\right] + \theta E\left[Z_{n+h-1}|X_1, ..., X_n\right] \end{aligned}$$

If $h = 1$, then the above equation is

$$\begin{aligned} \hat{X}_{n+1} &= \underbrace{E\left[Z_{n+1}|X_1, ..., X_n\right]}_{=0} + \theta E\left[Z_n|X_1, ..., X_n\right] \\ &= \theta E\left[Z_n|X_1, ..., X_n\right] \\ &= \theta Z_n \end{aligned}$$

and if $h > 1$ then the equation becomes

$$\hat{X}_{n+1} = E[Z_{n+h}] + \theta E\left[Z_{\underbrace{n+h-1}_{>n}}|X_1, ..., X_n\right] = 0$$

Now we need to plug in a value for $Z_n$. We "approximate" the $Z_i's$ by $U_i's$ as follows. Let $U_0 = 0$ and we estimate

$$\hat{Z}_t = U_t = X_t - \theta U_{t-1}, U_0 = 0$$

from the fact that $Z_t = X_t - \theta Z_{t-1}$. We can then get that

$$\begin{aligned} U_0 &= 0 \\ U_1 &= X_1 \\ U_2 &= X_2 - \theta X_1 \\ U_3 &= X_3 - \theta X_2 + \theta^2 X_1 \\ &\vdots \end{aligned}$$

Notice that as $i \to \infty$, $U_i$ will need a convergence condition where $|\theta| < 1$ is sufficient. This was the invertibility condition for $MA(1)$. We see that the $U_i's$ are recursively calculable and for an invertible $MA(1)$ process, we have

$$\hat{X}_{n+h} = \begin{cases} \theta U_n & h = 1 \\ 0 & h > 1 \end{cases}, U_t = X_t - \theta U_{t-1}, U_0 = 0$$

Now consider an $MA(q)$ process $X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$. We have

$$\begin{aligned} \hat{X}_{n+h} &= E\left[X_{n+h} | X_1, ..., X_n\right] \\ &= E\left[Z_{n+h} | X_1, ..., X_n\right] + \theta_1 E\left[Z_{n+h-1} | X_1, ..., X_n\right] + ... + \theta_q E\left[Z_{n+h-q} | X_1, ..., X_n\right] \end{aligned}$$

If $h > q$ then the above equation's value is zero since we have $n + h - q > n$. If $0 < h \leq q$ then at least some of the terms in the above are non-zero. In particular,

$$\begin{aligned} \hat{X}_{n+h} &= \sum_{j=1}^{q} \theta_j E\left[Z_{n+h-1} | X_1, ..., X_n\right] \\ &= \sum_{j=h}^{q} \theta_j E\left[Z_{n+h-1} | X_1, ..., X_n\right] \end{aligned}$$

and for $j = h, h+1, ..., q$ we know $E[Z_{n+h-j} | X_1, ..., X_n] = Z_{n+h-j}$ and hence

$$\hat{X}_{n+h} = \sum_{j=h}^{q} \theta_j Z_{n+h-j}$$

Similar to $MA(1)$, we approximate $Z_i's$ by $U_i's$, provided the $MA(q)$ process is invertible. That is, $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q \neq 0$ for all $|z| \leq 1$. Therefore, assuming that

$$U_0 = U_{-1} = U_{-2} = .... = 0$$

then $U_t = X_t - \sum_{j=1}^{q} \theta_j U_{t-j}$ and

$$\begin{aligned} U_0 &= 0 \\ U_1 &= X_1 \\ U_2 &= X_2 - \theta_1 X_1 \\ U_3 &= X_3 - \theta_2 X_2 + \theta_2 \theta_1 X_1 \\ &\vdots \end{aligned}$$

In summary, for an invertible $MA(q)$ process, we have

$$\hat{X}_{n+h} = \begin{cases} \sum_{j=h}^{q} \theta_j U_{n+h-j} & 1 \leq h \leq q \\ 0 & h > q \end{cases}$$

where $U_0 = U_i = .... = 0$, $i < 0$ and $U_t = X_t - \sum_{j=1}^{q} \theta_j U_{t-j}$ for $t = 1, 2, 3, ...$

**Example 7.2.** Consider the $MA(1)$ process $X_t = Z_t + 0.5Z_{t-1}$ where $\{Z_n\} \sim WN(0, \sigma^2)$. If $X_1 = 0.3, X_2 = -0.1, X_3 = 0.1$, predict $X_4, X_5$. Notice that $\hat{X}_5 = \hat{X}_{3+2}$ which is a 2-step prediction based on the history $X_1 = X_2 = X_3$. Since this is an $MA(1)$ model, hence 1-correlated, $\hat{X}_5 = 0$. For $X_4$ we have

$$\hat{X}_4 = \sum_{j=1}^{1} = \theta_j U_{3+1-j} = \theta_1 U_3 = 0.5U_3$$

where

$$\begin{aligned} U_0 &= 0 \\ U_1 &= X_1 - 0.5U_0 = X_1 = 0.3 \\ U_2 &= X_2 - 0.5U_1 = -0.1 - (0.5)(0.3) = 0.25 \\ U_3 &= X_3 - 0.5U_2 = 0.1 - (0.5)(-0.25) = 0.225 \end{aligned}$$

and hence $\hat{X}_4 = 0.5(0.225) = 0.1125$.

**Example 7.3.** Consider the $MA(1)$ process $X_t = Z_t + \theta Z_{t-1}$ with $\{Z_t\} \sim WN(0, \sigma^2)$ and $|\theta| < 1$. Show that the one-step predictor $\hat{X}_{n+1} = \theta U_n$ is equal to the predictor

$$\hat{\hat{X}}_{n+1} = -\sum_{j=1}^{n} (-\theta)^j X_{n-j+1}$$

This is by definition of $U_n$ which we can write the closed form

$$U_n = X_n + \sum_{i=1}^{n-1} (-\theta)^i X_{n-i}, n \geq 2$$

and hence

$$\hat{X}_{n+1} = \theta U_n = \theta X_n - \sum_{i=1}^{n-1} (-\theta)^{i+1} X_{n-i} = -\sum_{i=0}^{n-1} (-\theta)^{i+1} X_{n-i} = -\sum_{j=1}^{n} (-\theta)^j X_{n-j+1} = \hat{\hat{X}}_{n+1}$$

Clearly for $n = 0, 1$ we have $\hat{X}_{n+1} = \hat{\hat{X}}_{n+1}$ as well. This shows that even in the MA process, the predictor may be written as a linear function of the "history".