# 시계열자료분석팀 2주차

**[ 목차 ]**



# 1 1 주차 복습

시계열 자료는 보통 **dependency** 를 가집니다. 시계열 자료의 확률 분포를 보다 쉽게 추정하기 위해, 확률적 성질(공분산)이 시간의 흐름에 의존하지 않고 시차(lag)에만 의존한다는 **정상성 가정**을 합니다. 대부분의 정상성은 **약정상성**을 의미하는데, **평균과 분산이 일정하고, 자기공분산(ACVF)이 시차에만 의존**해야 합니다. 대부분의 시계열은 의존성이 있는 비정상 시계열인데, 이를 nonstationary part (추세, 계절성)과 stationary part (정상성을 만족하는 오차)로 나누어 추세와 계절성을 추정하여 제거해주는 **정상화**를 거쳐야 합니다. 정상화 방법으로는 **회귀, 평활, 차분**이 있습니다. 비정상 부분을 제거한 다음에는 남은 오차항이 **WN sequence** 인지 확인하는 **정상성 검정**을 해야 합니다. 특히 자기상관을 확인하기 위해 correlogram (=ACF 그래프)를 확인합니다.

정상화 후 오차 $Y_t$에 대하여 정상성 검정을 한 결과, 오차항이 정상성을 만족하지 않을 수도 있습니다. 즉, 오차항이 백색잡음이 아닌 경우, 오차항에 대해 추가적인 모델링이 필요한데, 이 역할을 하는 것이 바로 오늘 배우게 될 시계열 모형입니다.

# 2 모형의 식별

## 2.1 시계열 모형의 필요성

오차항이 백색잡음이 아닌 경우 왜 추가적인 모델링이 필요한 걸까요? 분산-공분산 행렬 $\Gamma$에서 그 이유를 찾을 수 있습니다.

$$\Gamma = \begin{pmatrix} \mathrm{Cov}(Y_1, Y_1) & \mathrm{Cov}(Y_1, Y_2) & \dots & \mathrm{Cov}(Y_1, Y_n) \\ \mathrm{Cov}(Y_2, Y_1) & \mathrm{Cov}(Y_2, Y_2) & \dots & \mathrm{Cov}(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(Y_n, Y_1) & \mathrm{Cov}(Y_n, Y_2) & \dots & \mathrm{Cov}(Y_n, Y_n) \end{pmatrix}$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}$$

자기공분산함수 ACVF는 $\gamma_x(h) = Cov(X_t, X_{t+h})$ 와 같이 표현됩니다. 즉, 우리가 최종적으로 추정해야 하는 분산-공분산 행렬은 위와 같습니다.

오차항이 백색잡음인 경우, $Cov(X_t, X_s) = 0, if\ t \neq s$ 을 만족하므로 인덱스가 다른 확률변수 간에 상관관계가 없습니다. 즉, 대각요소를 제외한 상삼각과 하삼각 요소가 전부 0이기 때문에, 우리가 추정해야 하는 것은 결과적으로 $\sigma^2$ 뿐입니다.

하지만, 오차항이 백색잡음이 아니라면 어떨까요? 확률변수 간에 상관관계가 남아있기 때문에 분산-공분산 행렬 전체의 원소를 추정해야 합니다. (물론 $\gamma(0), \dots, \gamma(n-1)$만 찾으면 됩니다.)이 때 사용되는 모형이 오늘 배울 시계열 모형들입니다.

2.2와 2.3에서는 ACF와 PACF를 통해 어떤 경우에 어떤 모형을 사용해야 할지 판별하는 방법에 대해 알아보겠습니다.

### 2.2 자기상관함수 ACF (auto-correlation function)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t+h}) = \frac{Cov(X_t\ , X_{t+h})}{\sqrt{var(X_t)}\sqrt{var(X_{t+h})}}$$

클린업 1 주차에서 간단히 짚고 넘어갔던 ACF 의 특성을 좀 더 알아보겠습니다. ACF $\rho(h)$ 는 시차가 h 인 확률변수 간의 상관관계를 의미하며, 정상 시계열인 경우 시점이 변해도 동일한 시차에 대해서는 같은 값을 가져야 합니다. (ex. $Cov(X_1, X_2) = \gamma(1) = Cov(X_5, X_6)$) ACF 는 다음과 같은 특성을 갖습니다.

1) $\rho(0) = 1$

2) $\rho(-h) = \rho(h) \quad (even\ function)$

3) $|\gamma(h)| \leq \gamma(0)\ for\ all\ h \in \mathbb{Z} \quad \rightarrow \quad |\rho(h)| \leq 1$

### 2.3 부분자기상관함수 PACF (partial auto-correlation function)


부분자기상관이란 $X_1$과 $X_{k+1}$ 사이에 $X_2, \dots, X_k$의 영향을 제거했을 때의 조건부 상관관계 (conditional correlation)을 의미합니다.

$$\alpha(0) = Corr(X_1, X_1) = 1$$

$$\alpha(1) = Corr(X_2, X_1) = \rho(1)$$

$$\boldsymbol{\alpha(k) = Corr(X_{k+1} - P_k^* X_{k+1}, X_1 - P_k^* X_1)}, \ \ k \geq 2$$

$$\text{where } \ P_k^* X_{k+1} = \ Best\ Linear\ Predictor\ on\ \{1, X_2, \cdots, X_k\}$$

$$P_k^* X_1 = \ Best\ Linear\ Predictor\ on\ \{1, X_2, \cdots, X_k\}$$

$P_k^* X_{k+1}$ 는 중간 시점인 $\{1, X_2, \cdots, X_k\}$ 만으로 $X_{k+1}$ 을 설명한 것입니다. $P_k^* X_1$ 은 중간 시점만으로 $X_1$ 을 설명한 것입니다. 각각을 빼서 $X_1$ 과 $X_{k+1}$ 사이에 중간 시점의 영향을 제거한 뒤 구한 것을 부분자기상관함수 PACF 라 합니다.

BLP 를 구하는 방법을 좀 더 자세하게 살펴보겠습니다.

$$X_{k+1} = \ \phi_{11} X_k + \ \epsilon_{k+1}$$

$$X_{k+1} = \ \phi_{21} X_k + \phi_{22} X_{k-1} + \epsilon_{k+1}$$

$$\vdots$$

$$X_{k+1} = \ \phi_{k1} X_k + \phi_{k2} X_{k-1} + \cdots + \phi_{kk} X_1 + \epsilon_{k+1}$$

순서대로 lag=1, lag=2, ⋯, lag=k 인 경우에 대해 $X_{k+1}$을 회귀식으로 표현한 것입니다. $\{X_1, \dots, X_k\}$로 $X_{k+1}$을 설명한 회귀식의 오차항을 최소화하는 추정량은 다음과 같습니다.

$$\hat{X}_{k+1} = arg\min_{\phi} E(X_{k+1} - \ \phi_{k1} X_k - \phi_{k2} X_{k-1} - \cdots - \ \phi_{kk} X_1)^2$$

이 때, **계수 $\boldsymbol{\phi_{kk}}$는 $\{\boldsymbol{X_2, \dots X_k}\}$이 설명하는 부분을 제거한 뒤 남은 부분에 대하여 $\boldsymbol{X_1}$의 $\boldsymbol{X_{k+1}}$에 대한 영향, 즉 순수한 $\boldsymbol{X_1}$의 $\boldsymbol{X_{k+1}}$ 사이의 상관관계**를 나타냅니다.

$$\boldsymbol{\alpha(k) = \ \phi_{kk}}, \qquad k \geq 1$$

## 3 선형과정 (Linear Process)

**선형과정이란 $\{\boldsymbol{Z_t}\} \sim \boldsymbol{WN(0, \sigma^2)}$ 들의 선형결합입니다**. 이 때, 상수들의 합이 발산하지 않도록 $\sum_j |\psi_j| < \infty$ (absolutely summable) 조건을 만족해야 합니다. 선형과정에서 $Z_t$를 innovation 이라고 부릅니다.

$$\boldsymbol{X_t} = \sum_{\boldsymbol{j=-\infty}}^{\infty} \boldsymbol{\psi_j Z_{t-j}} \coloneqq \ \psi(B) Z_t$$

$$\text{where } \psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$$

(example) $X_t = Z_t + 0.2\, Z_{t-1} = (1 + 0.2B)Z_t$ , $\psi(B) = (1 + 0.2B)$

선형과정은 time dependence 를 설명하기 위해 사용되는 구조로, 곧 공부할 AR, MA, ARMA 모형 등이 모두 선형과정의 일종입니다. 선형과정을 사용하는 이유는 1) 공분산 계산이 쉬워지며, 2) 해석과 추정에 용이하며, 3) 정상 선형과정의 선형 결합은 정상 선형과정을 다시 만족하며, 4) Wold decomposition 에 의해 약정상성을 만족하는 모든 정상 선형과정은 선형과정의 선형결합으로 표현 가능하기 때문입니다.

3)의 증명은 생략하지만 (궁금한 분께는 자료를 보내드리겠습니다), 그 과정에서 함께 증명되는 $\{Y_t\} \sim WN(0, \sigma^2)$ 인 경우의 ACVF 공식은 짚고 넘어가겠습니다. 즉, 아래 식은 **WN sequence(백색잡음)을 innovation 으로 갖는 선형과정의 ACVF** 입니다.

$$X_t = \sum_j \psi_j Y_{t-j}, \ \sum_j |\psi_j| < \infty$$

$$\gamma_X(h) = \sum_j \psi_j \psi_{j+h} \sigma^2$$

# 4 AR 모형 (Auto-Regressive)

## 4.1 AR 모형의 정의

**AR 모형은 $X_t$ 가 과거 관측값과 현 시점에서의 오차의 함수로 표현되는 모형입니다**. 현재를 자신의 과거로 설명(회귀)한다는 점에서 '자기회귀'라는 표현을 사용합니다. $Z_t \sim WN(0, \sigma^2)$일 때 다음과 같이 표현할 수 있습니다.

**AR(1) 모형**

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

**AR(p) 모형**

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

## 4.2 AR 모형의 특성방정식

위의 AR(p) 모형을 후향 연산자를 활용하여 간단하게 표현할 수 있습니다.

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

$$= \phi_1 B X_t + \phi_2 B^2 X_t + \cdots + \phi_p B^P X_t + Z_t$$

위 식을 $Z_t$에 대한 식으로 바꾸면 다음과 같습니다.

$$\boldsymbol{Z_t} = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) X_t = \boldsymbol{\phi(B) X_t}$$

이 때, $(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$을 AR(p)의 특성방정식 $\phi(B)$라 합니다.

### 4.3 AR 모형의 조건

AR 모형은 정상성과 인과성 두 가지의 조건을 만족해야 합니다.

1) **정상성** (stationarity) : 시계열의 확률적 특성이 시점이 아닌 시차에만 의존하는 특성

2) **인과성** (causality) : t 시점의 관측값이 **과거 시점의 오차항으로 설명**되는 특성

$$[Causality] \quad \psi_j = 0\,, \forall j < 0 \leftrightarrow X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

AR(1) 모형을 통해 정상성과 인과성을 어떻게 만족하는지 확인해보겠습니다.

$$X_t = \phi_1 X_{t-1} + Z_t$$
$$= \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = \phi_1^2 X_{t-2} + \phi_1 Z_{t-1} + Z_t$$
$$\vdots$$
$$= \phi_1^{M+1} X_{t-M-1} + \Sigma_{j=0}^{M} \phi_1^j Z_{t-j}$$

이 때, 선형과정이기 위해서는 innovation 들의 계수의 합이 유한해야 하는 absolutely summable 조건을 만족해야 합니다. 즉, $M \to \infty$일 때 $\Sigma_{j=0}^{M} \phi_1^j$ 가 유한하려면 기하급수의 수렴 조건(공비의 절댓값이 1 보다 작을 것)대로 $\boldsymbol{|\phi_1| < 1}$ 이어야 합니다.

$|\phi_1| < 1$ 조건에서, AR(1) 모형은 선형 과정의 성질에 따라 정상 시계열의 선형결합은 또 다시 정상 시계열이므로 정상성을 만족하고, $X_t$ 가 과거의 값들로 설명되고 있다는 점에서 인과성도 만족합니다. 따라서 AR 모형의 계수의 절댓값은 1 보다 작아야 합니다. 이는 특성방정식 $\phi(B) = 0$의 근의 절댓값이 1 보다 커야 한다는 조건과 동치입니다.

2 절에서 ACF 와 PACF 를 통해서 어떤 모형을 사용해야 하는지 판별할 수 있다고 했습니다. 그렇다면 각 모형의 ACF 와 PACF 가 어떤 특성을 갖는지 알고 있어야겠죠? 지금 다루고 있는 AR 모형의 ACF 와 PACF 에 대해 알아보겠습니다.

### 4.4 AR 모형의 ACF

자기상관함수 ACF (auto-correlation function)가 나온 김에 다시 한 번 복습하겠습니다.

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t+h}) = \frac{Cov(X_t, X_{t+h})}{\sqrt{var(X_t)}\sqrt{var(X_{t+h})}}$$

$$(\text{자기공분산함수 } ACVF) \quad \gamma_x(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

AR(1) 모형의 ACF 를 구해보겠습니다. 계산의 편의를 위해 $E(X_t) = 0$ 임을 가정합니다. 양 변에 $X_{t-h}$를 곱하고, 기댓값을 취하겠습니다.

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

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

$$E[X_t X_{t-h}] = \phi_1 E[X_{t-1} X_{t-h}] + E[Z_t X_{t-h}]$$

$$\gamma(h) = \phi_1 \gamma(h-1) + Cov(Z_t, X_{t-h}) = \phi_1 \gamma(h-1)$$

$$\gamma(h) = \phi_1\big(\phi_1 \gamma(h-2)\big) = \cdots = \phi_1^h \gamma(0)$$

$$\frac{\gamma(h)}{\gamma(0)} = \phi_1^h = \rho(h)$$

따라서 AR 모형은 **시차(lag) h 가 커질수록** 계수 $\phi_1$의 절댓값이 1 보다 작기 때문에 **ACF 가 지수적으로 감소**하게 됩니다.

### 4.5 AR 모형의 PACF

2.3 절에서 공부한 부분자기상관함수 PACF (partial auto-correlation function)은 중간 시점의 영향을 BLP 를 활용하여 제거한 순수한 상관관계를 의미했습니다. PACF 의 공식을 복습하자면 다음과 같습니다.

$$\alpha(k) = Corr(X_{k+1} - P_k^* X_{k+1}, X_1 - P_k^* X_1), \ \ k \geq 2$$

$X_{k+1}$을 AR(p) 모형을 통해 추정하면 다음과 같은 식을 구할 수 있습니다.

$$\hat{X}_{k+1} = \phi_1 X_k + \phi_2 X_{k-1} + \cdots + \phi_p X_{k+1-p} + 0X_{k-p} + \cdots + 0X_1$$

2.3 절에서 확인한 것처럼, $\alpha(0)$와 $\alpha(1)$의 경우 중간 시점이 없기 때문에 다음과 같이 정의됩니다.

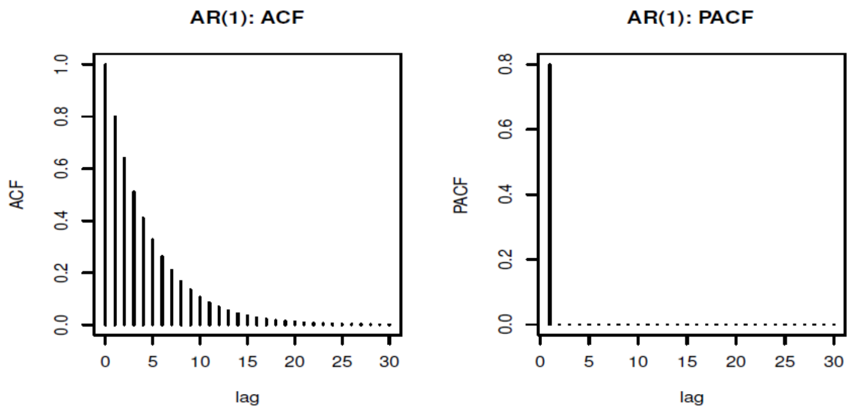
$$\alpha(0) = Corr(X_1, X_1) = 1$$

$$\alpha(1) = Corr(X_2, X_1) = \rho(1)$$

또한, 2.3 절에서 BLP 를 구하는 회귀식에 OLS 를 적용했을 때, $\alpha(k) = \phi_{kk}$ 였던 것과 같이 $\alpha(p)$와 $\alpha(k), k > p$ 는 다음과 같이 정의됩니다.

$$\alpha(p) = \phi_p \ \ \& \ \ \alpha(k) = 0 \ \ if \ k > p$$

정리하자면, **AR 모형의 PACF 는 시차(lag) p 까지만 존재**하며, **시차가 p 를 넘어가면 PACF=0** 이 됩니다. 이를 "AR(p) 모형의 PACF는 시차 p 이후에 절단된다"고 표현합니다.




AR(1) 모형의 ACF, PACF 그래프입니다. **시차가 커질수록 ACF 가 지수적으로 감소**하는 AR 모형의 ACF 의 특징이 잘 드러납니다. 또한, **PACF 그래프가 1 이후로 절단**된 것을 보아 AR(1) 모형임을 잘 보여줍니다.

# 5 MA 모형 (Moving Average)

## 5.1 MA 모형의 정의

**MA 모형은 $X_t$ 가 과거 시점의 오차항들의 함수로 표현되는 모형**입니다. AR 모형과 마찬가지로 $Z_t \sim WN(0, \sigma^2)$를 따르는 innovations 으로 다음과 같이 표현합니다.

**MA(1) 모형**

$$\boldsymbol{X_t = Z_t + \theta_1 Z_{t-1}}$$

**MA(p) 모형**

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

### 5.2 MA 모형의 특성방정식

위의 MA(p) 모형을 후항 연산자를 활용하여 다음과 같이 표현할 수 있습니다.

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

$$= Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_t - \cdots + \theta_q B^q Z_t$$

$$= \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) Z_t$$

$$\boldsymbol{X_t = \theta(B) Z_t}$$

이 때, $1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ 를 MA(p)의 특성 방정식 $\theta(B)$라 합니다.

### 5.3 MA 모형의 조건

MA 모형은 정상성, 인과성, 가역성 세 가지의 조건을 만족해야 합니다.

1) **정상성** (stationarity) : 시계열의 확률적 특성이 시점이 아닌 시차에만 의존하는 특성

2) **인과성** (causality) : **t 시점의 관측값이 과거 시점의 오차항으로 설명**되는 특성

3) **가역성** (invertibility) : **t 시점의 오차항이 과거 시점의 관측값으로 설명**되는 특성

$$[Invertibility] \quad Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \ \ for\ all\ t, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty$$

MA 모형은 백색잡음을 따르는 오차의 선형결합이므로 정상성을 만족하며, t 시점 이전의 오차항들로 $X_t$ 를 설명하기 때문에 인과성 역시 만족합니다. 그렇다면 가역성은 어떤 조건에서 만족하는지 MA(1) 모형을 통해 알아보겠습니다.

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

$$(1 + \theta B)^{-1} X_t = Z_t$$

$$(1 + \theta B)^{-1} = \frac{1}{1 - (-\theta B)} = 1 - \theta B + (\theta B)^2 - (\theta B)^3 + \cdots$$

위와 같이 $Z_t$ 를 $X_t$ 에 대한 함수로 표현한 결과, 무한등비급수의 합 형태가 됩니다. 선형과정이기 위해서는 계수가 absolutely summable 해야 하기 때문에, $\boldsymbol{|\theta| < 1}$ 일 때만 이러한 조건을 만족합니다. 이는 특성방정식 $\theta(B) = 0$ 의 근의 절댓값이 1 보다 커야 한다는 조건과 동치입니다.

## 5.4 MA 모형의 ACF

$$\gamma_X(h) = \sum_j \psi_j \psi_{j+h} \sigma^2$$

3 장에서 언급했듯이, 백색잡음 WN(0, $\sigma^2$)를 innovation 으로 갖는 선형과정의 ACVF 는 위와 같이 정의됩니다. 이 때, 백색잡음은 확률변수 간에 상관관계가 없어야 하므로, 인덱스가 다른 경우에는 공분산이 0 입니다. 따라서 인덱스가 겹치는 innovations 에 대해서만 계수의 합을 구하면 됩니다.

MA 모형은 위 ACVF 의 조건을 만족하기 때문에 이 공식을 이용하여 MA(1) 모형의 ACVF 를 계산해보겠습니다. ACF 는 $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$ 이므로, 아래의 ACVF 를 $\gamma(0)$으로 나눠주면 ACF 를 구할 수 있습니다.

1) $h = 0$인 경우

$$\gamma(0) = Cov(Z_t + \theta_1 Z_{t-1}, Z_t + \theta_1 Z_{t-1}) = \sigma^2 + \theta_1^2\sigma^2 = (1+\theta_1^2)\sigma^2$$

$$\rho(0) = 1$$

2) $h = 1$인 경우

$$\gamma(1) = Cov(Z_{t-1} + \theta_1 Z_{t-2}, Z_t + \theta_1 Z_{t-1}) = \theta_1\sigma^2$$

$$\rho(1) = \frac{\theta}{1+\theta^2}$$

3) $h \geq 2$인 경우

$$\gamma(h) = Cov(Z_{t-h} - \theta_1 Z_{t-h-1}, Z_t - \theta_1 Z_{t-1}) = 0$$

$$\rho(2) = 0$$

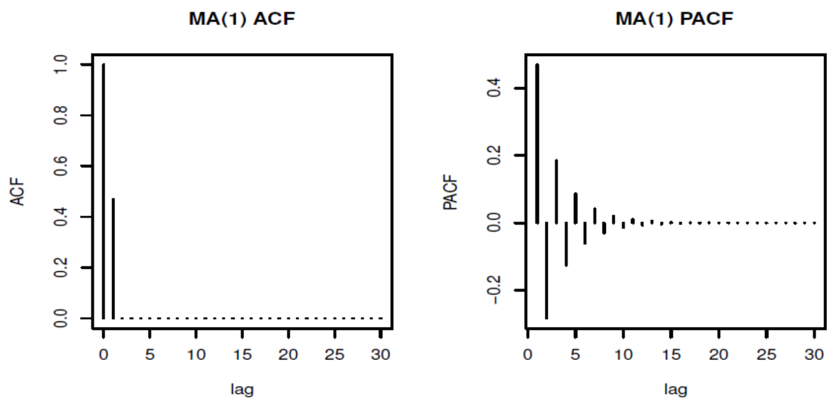
위 식에 따르면, 시차 h 가 1 보다 커지면 인덱스가 겹치는 innovation 쌍이 없기 때문에, 남아있는 공분산이 없게 됩니다. 정리하자면, **MA(q) 모형은 시차 h 가 q 보다 커지면 ACF 가 절단**되는 특성을 가집니다.

## 5.5 MA 모형의 PACF

MA 모형의 PACF 는 Crammer 공식을 사용하여 계산할 수 있습니다. 풀이 과정을 자세히 다루지는 않고 PACF 식과 특징에 대해서만 알아보겠습니다. ACF 와 PACF 를 다루는 이유는 결국 모형 선택을 잘 하기 위해서, 각 모형이 어떤 특징을 갖는지 판별하기 위함이니까요!

$$\alpha(k) = \phi_{kk} = \frac{-(-\theta)^k}{(1+\theta^2+\cdots+\theta^{2k})}, k \geq 1$$

위 식은 MA(1) 모형의 PACF 를 계산한 결과입니다. **시차 k 가 커질수록 PACF 는 점점 0 으로 수렴**하는 특징을 갖습니다.




MA(1) 모형의 ACF, PACF 그래프입니다. ACF 그래프의 경우, **시차(lag) 1 까지만 ACF 가 존재하고 2 이상부터는 절단**된 것을 확인할 수 있습니다. PACF 그래프의 경우, **시차가 커질수록 PACF 가 점점 0 으로 수렴**하는 모습을 확인할 수 있습니다.

# 6 ARMA 모형

## 6.1 ARMA 모형의 정의

ARMA 모형은 AR 과 MA 를 동시에 포함하는 모형으로, $Z_t \sim WN(0,\sigma^2)$일 때 다음과 같이 표현합니다.

**ARMA(1, 1) 모형**

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

**ARMA(p, q) 모형**

$$\boldsymbol{X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}}$$

ARMA(p, q)에서 p 와 q 는 각각 AR 의 차수, MA 의 차수를 나타냅니다.

AR 과 MA 를 통합한 모형이 필요한 이유는 '모수의 절약(parsimony)'과 관련이 있습니다. 시계열 자료를 AR 모형이나 MA 모형만 가지고 설명하려면 모수인 p, q 의 값이 너무

커질 가능성이 있습니다. 이 때, 추정해야 할 모수의 개수가 많아지면 일반적으로 추정의 효율성이 떨어지고, 해석이 어려우며, 적합 시간도 오래 걸리는 문제가 있습니다. 이 경우, ARMA 모형을 사용한다면 모수를 절약할 수 있다는 장점이 있습니다.

### 6.2 ARMA 모형의 특성방정식

ARMA(p, q) 모형을 후향 연산자를 활용하여 다음과 같이 나타낼 수 있습니다.

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$

$$X_t - \phi_1 B X_t - \phi_2 B^2 X_t - \cdots - \phi_p B^p X_t = Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_{t-2} + \cdots + \theta_q B^q Z_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t$$

$$\boldsymbol{\phi(B) X_t} = \boldsymbol{\theta(B) Z_t}$$

### 6.3 ARMA 모형의 조건

ARMA 모형 AR 모형의 조건과 MA 모형의 조건을 모두 충족해야 합니다. 즉, **정상성, 인과성, 가역성**을 모두 만족해야 합니다. 예시로 ARMA(2, 2) 모형을 보겠습니다.

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

$$(1 - \phi_1 B - \phi_2 B^2) X_t = (1 + \theta_1 B + \theta_1 B^2) Z_t$$

위 모형은 정상성을 만족하기 위해 $\phi(B) X_t = 0$ 의 근의 절댓값이 1 보다 커야 하고, $\theta(B) Z_t = 0$ 의 근의 절댓값이 1 보다 커야 합니다.

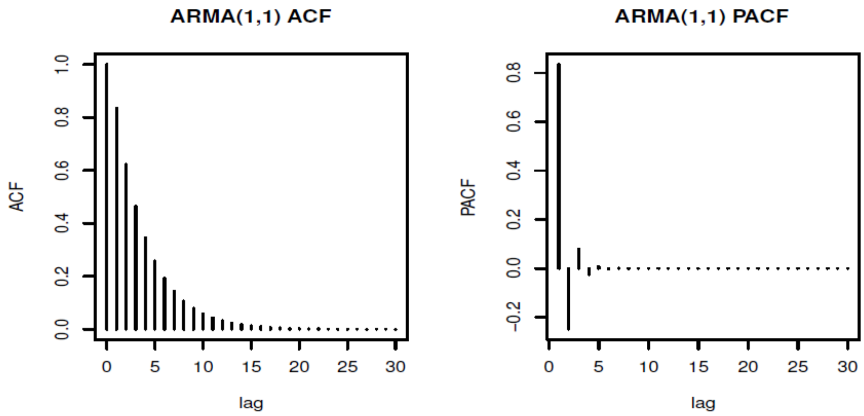
ARMA 모형은 **식별성 (identifiability)** 또한 갖춰야 합니다. 식별성이란, 주어진 파라미터 조합에 대해 하나의 모형만이 대응되는 특성을 의미합니다.

$$(1 - \phi B) = (1 + \theta B) \iff \phi = -\theta$$

만약 위와 같은 경우라면, ARMA(1, 1) 모형의 식은 $X_t = Z_t$가 됩니다. 하지만 이 경우 이 모형이 ARMA(1,1)인지, WN 인지 식별할 수 없는 문제가 생깁니다. 따라서, ARMA 모형의 경우 추가적으로 $\boldsymbol{\phi} + \boldsymbol{\theta} = \mathbf{0}$ 의 식별성에 대한 조건이 필요합니다.

### 6.4 ARMA 모형의 ACF & PACF

**ARMA 모형의 ACF 와 PACF 그래프는 모두 지수적으로 감소하거나 싸인함수 형태로 소멸**하는 양상을 보입니다.




AR(p) 모형은 PACF 가 p+1 차부터 절단되었고, MA(q) 모형은 ACF 가 q+1 차부터 절단되는 특징이 있었습니다. 하지만 ARMA(p, q) 모형의 경우 ACF 와 PACF 모두 지수적으로 감소하기 때문에, ACF 와 PACF 그래프만을 활용하여 모형 식별이 어렵다는 한계가 있습니다.

## 7 모형의 적합 절차

지금까지 AR 모형, MA 모형, ARMA 모형에 대해 알아봤습니다. 복습하자면, 시계열 모형은 오차항이 정상성을 만족하지 않을 때, 오차항에 대해 설명하기 위해 사용되는 모델입니다.

그렇다면 오차항을 모델링하기 위해서 어떤 모형을 선택해야 할까요? 2 주차의 마지막 장에서는 적절한 모델을 선택하기 위한 모형의 적합 절차에 대해 알아보겠습니다.

### 7.1 모형 식별

가장 먼저 ARMA(p, q) 모형의 p, q 의 차수를 결정해야 합니다. AR(p) 또는 MA(q)의 경우, 시계열 자료의 ACF 와 PACF 그래프를 확인해보았을 때, 그래프가 절단되는 특징이 나타나기 때문에 p 또는 q 의 차수를 쉽게 결정할 수 있습니다. 하지만 ARMA 모형의

경우 ACF 와 PACF 모두 지수적으로 감소하기 때문에 ACF 와 PACF 그래프만을 활용하여 모형을 식별하는 데에는 한계가 있습니다. 이 때, AIC, AICC, BIC 등의 **Information Criteria** 를 참고하여 모형을 식별할 수 있습니다.

$$\boldsymbol{Information\,Criteria} = \{\boldsymbol{goodness\,of\,fit} + \boldsymbol{model\,complexity}\}$$

6 장에서 모수의 절약(parsimony) 측면에서 ARMA 모형의 필요성을 설명하였는데, IC 는 모형의 복잡도를 함께 고려하기 때문에 모형 식별을 위한 적절한 지표라고 볼 수 있겠습니다. 가장 작은 IC 값을 가진 모형을 선택하면 됩니다.

- **AIC** (Akaike Information Criteria) : $-2\ln L_n(\hat{\theta}) + 2(p+q+1)$

  (iid 가정)

- **AICC** (AIC bias corrected) : $-2\ln L_n(\hat{\theta}) + \frac{2(p+q+1)n}{n-(p+q+1)+1}$

  (Time dependency 에 대한 correction)

- **BIC** (Bayesian Information Criterion) : $-2\,ln\,L_n\big(\hat{\theta}\big) + (p+q+1)\ln n$

위 식에서 $-2\ln L_n(\hat{\theta})$ 는 SSE (Sum of Squared Error)에 대한 approximation 을 나타냅니다.

### 7.2 모수 추정

모수의 차수를 결정했다면, 그 다음으로는 모수(파라미터) $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, $\sigma^2$를 추정해야 합니다.

1) **최대가능도추정법 (MLE)** : 관측된 시계열의 결합확률밀도함수인 모수의 가능도 함수 (likelihood function)를 최대화하는 모수의 추정량을 구하는 방법

2) **최소제곱법 (LSE)** : 오차의 제곱합이 가장 작게 되도록 하는 모수의 추정량을 구하는 방법

3) **적률추정법 (MME/MoM)** : 모집단의 적률을 상응하는 표본의 적률로 대체한 후, 방정식을 풀어 모수의 추정량을 구하는 방법

보편적으로 MLE 의 성능이 가장 좋다고 알려져 있지만, MLE 를 구하기 위해서는 수학적 최적화가 필요합니다. 효과적인 접근을 위해 LSE 와 MME 를 MLE 에 대한 초기값으로 사용하여 MLE 를 추정하는 전략을 사용할 수 있습니다.

## 7.3 모형 진단

모형을 식별하고 모수를 추정했다면, 모형이 적합한지 진단해야 합니다.

### 1) 모수에 대한 검정

- 모형의 조건을 만족하는지 확인 : 정상성과 가역성 조건 만족 여부 (모수의 절댓값이 1 보다 작은지), 식별성 만족 여부 (모수의 합이 0 이 아닐 것)
- 모수의 유효성 확인 : 모수 $\neq \mathbf{0}$ 인지 확인

### 2) 잔차에 대한 검정

- 추세, 계절성, 이상치가 없는지 확인
- WN(0, $\sigma^2$)을 따르는지 확인 (자기상관성 유무) : 잔차에 대한 ACF, PACF 그래프 / Ljung-Box test / McLeod-Li test / Different sign test
- 정규성을 만족하는지 확인 : 잔차의 QQ plot / Jarque-Bera test

## 7.4 예측

모형을 식별하고 모수(파라미터)를 추정하고 모형에 대한 진단까지 마쳤다면, 선택한 모형을 이용하여 예측을 진행할 수 있습니다.

기본적으로, t+h 시점에 대한 예측 추정값은 $\{1, X_1, X_2, \ldots\}$의 선형결합이라고 가정합니다. 즉, 갖고 있는 데이터의 선형결합을 활용하여 미래에 대해 예측하는 것입니다.

$$\boldsymbol{P_n X_{n+h} \;=\; a_0 \cdot 1 + a_1 \cdot X_n + a_2 \cdot X_{n-1} + \cdots \;+ a_n \cdot X_1}$$

계수 $\{a_0, a_1, \ldots, a_n\}$ 에 대한 추정은 MSPE (Mean Squared Prediction Error)를 최소화하는 값을 사용합니다.

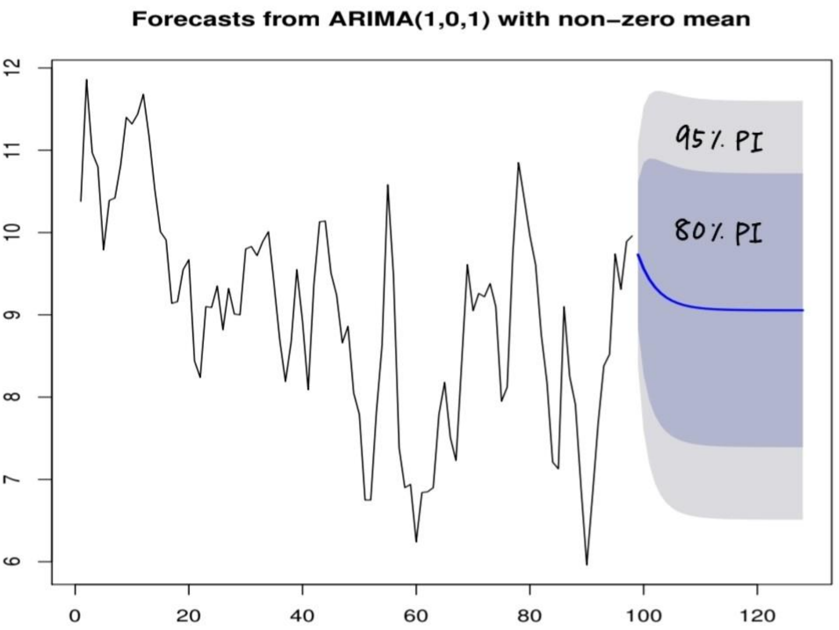
$$\boldsymbol{MSPE \;= E[X_{n+h} - P_n X_{n+h}]^2}$$

$$= \; E[X_{n+h} \;-\; (a_0 \cdot 1 + a_1 \cdot X_n + a_2 \cdot X_{n-1} + \cdots \;+ a_n \cdot X_1)]^2$$

계산 방식은 회귀분석에서의 LSE 계산 방법과 동일합니다. MSPE 식을 각각의 계수 $\{a_0, a_1, \ldots, a_n\}$에 대해 미분한 다음, 미분한 식=0 으로 놓고 각 normal equation 을 연립하여 해를 구하면 됩니다. (하지만 n 개의 연립방정식을 직접 계산하여 해를 찾는 것은 사실상 불가능하기 때문에, 컴퓨팅의 힘을 빌리게 됩니다 ㅎㅎ.)

이를 활용하여 PI (Prediction Interval)도 계산할 수 있습니다.

$$P_n X_{n+h} \pm 1.96 \cdot \sqrt{MSPE}$$



고생하셨습니다! (따봉) 시계열자료분석팀 2 주차 클린업이 끝났습니다. 오늘 공부한 내용을 정리하고 마무리하겠습니다.

오차항이 정상성을 만족하지 않을 경우, 시계열 모형을 활용하여 오차항을 모델링할 수 있습니다. 시계열 모형들은 모두 선형과정인데, 선형과정이란 백색잡음을 따르는 innovation $Z_t$들의 선형결합을 말합니다. AR, MA, ARMA 세 가지 모형에 대해 다뤘습니다. AR 모형은 현 시점의 관측값을 과거 관측값들과 현 시점의 오차의 함수로 설명하는 모형입니다. MA 모형은 현 시점의 관측값을 과거의 오차항들로 설명하는 모형입니다. ARMA 모형은 AR 구조와 MA 구조를 모두 포함한 모형입니다. ACF 와 PACF 그래프를 통해 모형을 식별할 수 있습니다. 각 모형의 성립 조건과 ACF, PACF 그래프의 특징을 표로 정리해보겠습니다.

1) 정상성과 가역성

| | AR(p) | MA(q) | ARMA(p,q) |
|---|---|---|---|
| 정상성 | $\|\phi\| < 1$ | 자체 만족 | $\|\phi\| < 1$ |
| 가역성 | 자체 만족 | $\|\theta\| < 1$ | $\|\theta\| < 1$ |

2) ACF, PACF 그래프의 특징

| | AR(p) | MA(q) | ARMA(p,q) |
|---|---|---|---|
| 정상성 | 지수적으로 감소 | q+1 차부터 절단 | 지수적으로 감소 |
| 가역성 | p+1 차부터 절단 | 지수적으로 감소 | 지수적으로 감소 |

모형의 적합 절차는 모형 식별, 모수 추정, 모형 진단, 예측의 순서로 진행됩니다.

But 끝날 때까지 끝난 게 아니다… 부록의 단위근 검정에 대해 살펴보고 마무리하도록 하겠습니다. (찐막) 다들 수고하셨습니다!

**[부록]**

## 단위근 검정이란?

데이터가 정상성을 만족하는지 만족하지 않는지를 검정하는 방법 중 단위근 검정이 있습니다. 클린업 1주차의 4.3절 백색잡음 검정에서 정상성에 대한 검정법으로 살펴본 Kpss, ADF, PP test가 단위근 검정법의 일종입니다. 예시로 AR(p) 과정을 생각해봅시다!

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

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

AR(p) 모형이 정상성을 만족하기 위해서는 $\phi(B) = 0$의 모든 근의 절댓값이 1보다 커야한다고 배웠습니다. 만약 절댓값이 1보다 크지 않은 근이 존재하는 경우 비정상 확률과정이라고 하며, 근의 크기가 1인 근을 단위근(unit root)라고 합니다. 따라서, 다음과 같은 가설을 검정하는 것을 단위근 검정이라고 합니다.

$$H_0 \colon \phi = 1 \; vs \; H_1 : |\phi| < 1$$

대립가설으로 모수의 절댓값이 1 보다 작음을 가정하기 때문에, 귀무가설을 기각하고 대립가설이 채택되었을 때 정상 시계열임을 알 수 있습니다. 7.3 절의 모형의 진단에서 모형의 조건에 대한 만족 여부를 확인하고자 할 때, 단위근 검정을 사용할 수 있습니다.