시계열자료분석 클린업 2주차

[목차]

+) 2주차 흐름 정리

총 19p.

1. 1주차 복습

클린업 2주차에서는 대표적인 시계열 모형들에 대해 배워보겠습니다. 그 전에, 1주차에 배웠던 내용들을 다시 짚어볼까요?

시계열 자료의 가장 중요한 특징은 관측치들 사이에 **dependency**가 존재하는 것이었습니다. 이 때문에 일반적인 방법과 다른 시계열 분석법을 사용해야 했고, 그때 중요한 역할을 하는 것이 바로 정상성이었습니다. 정상성은 강정상성과 약정상성으로 구분할 수 있으며, 현실과 클린업에서 주로 사용하는 개념은 바로 약정상성이었습니다. 약정상성의 조건은 평균과 분산이 일정하고, 자기공분산 함수인 ACVF가 시차에만 의존하는 것임을 기억해야 합니다!

시계열 자료가 정상성을 만족하는지 여부에 따라 정상 시계열과 비정상 시계열로 나눌 수 있었고, 비정상 시계열의 경우 정상화 과정을 통해 보다 더 정확하고 안정적인 예측과 분석을 할 수 있었습니다. 분산이 일정하지 않은 경우에는 VST를 사용해 분산안정화를 진행하고, 평균이 일정하지 않은 경우에는 시계열 분해를 통해 회귀, 평활, 차분을 거쳐 비정상 부분을 제거할 수 있습니다.

정상화가 끝난 후 남아있는 정상 부분, 즉 오차에 대해 정상성을 만족하는지를 검정해주어야 합니다. 그 과정에서 WN과 IID에 대해 배웠고, 오차가 WN 또는 IID인지의 여부는 ACVF와 ACF를 통해 확인하였습니다.

2주차 내용을 잘 이해하기 위해서는 정상 시계열과 백색잡음의 관계를 명확히 이해해야 하기 때문에 마지막으로 강조하고 넘어가겠습니다! 정상 시계열은 약정상성의 3가지 조건을 모두 만족하는 시계열 자료를 말하며, 백색잡음은 정상 시계열 중에서도 자기상관이 존재하지 않는 시계열 자료를 말합니다. 즉, **모든 백색잡음은 정상 시계열이지만, 모든 정상 시계열이 백색잡음은 아님**을 기억해야 합니다!! (중요중요..)

이렇게 1주차 때 배운 내용을 잘 간직한 채 2주차 클린업으로 들어가보도록 하겠습니다!

2 모형의 식별

2.1 시계열 모형의 필요성

시계열 모형에 대해 배우기에 앞서, 그 필요성을 짚고 넘어가도록 하겠습니다.

오차항 Y_t의 공분산 행렬은 아래와 같습니다.

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

1 주차에서 배운 정상성 조건 중 자기공분산은 시차에만 의존한다는 내용이 있었습니다. 해당 성질을 이용하여 위의 공분산 행렬을 정의하면 아래와 같습니다.

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

만약 Y_t가 백색잡음이라면, 대각요소를 제외한 요소들은 모두 0 이 되어 분산만 추정해 공분산 행렬을 구할 수 있습니다. 그러나, Y_t가 백색잡음이 아니라면 확률변수 사이에 상관관계가 존재해 하삼각요소를 모두 추정해야 합니다. 1 주차에 배운 것처럼 $\gamma(h)$는 SAVCF 를 통해 구할 수 있습니다.

$$\hat{\gamma}_Y(h) = \frac{1}{n}\Sigma_{j=1}^{n-h}(Y_j - \bar{Y})(Y_{j+h} - \bar{Y})$$

하지만 위 식을 통해 확인할 수 있는 것처럼 SAVCF를 통해 추정할 경우 h에 따라 추정의 정확도가 달라질 수 있다는 단점이 존재합니다. SAVCF의 단점을 해결하면서, 이 복잡한 과정을 보다 쉽게 해결해 줄 수 있는 방법이 바로 오늘 배울 시계열 모형입니다!

정리하자면, 정상화를 통해 비정상 시계열을 정상 시계열로 변환하였지만, **남아있는 오차가 IID 나 WN 이 아닐 때** $\gamma(h)\ (h \neq 0)$를 추정하기 위해 시계열 모형이 필요한 것입니다.

시계열 모형에도 여러 종류가 존재하는데요, 어떤 상황에서 어떤 모형을 사용해야 하는지 판단할 수 있는 ACF 와 PACF 에 대해 배워보겠습니다.

2.2 자기상관함수 ACF(Autocorrelation Function)

1 주차에서 배운 ACF 와 동일한 함수입니다. 간단히만 설명하자면, 자기상관함수는 **시차가 $\boldsymbol{h}$인 시계열 간의 상관관계**를 의미하며, 정상성을 만족한다면 시차에만 의존하는 특성이 있었습니다. 대표적인 특징은 아래와 같습니다.

1) $\rho(0) = 1 \ \ (\because \ \gamma(0) = var(X_t))$

2) $\rho(-h) = \rho(h)$

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

2.3 부분자기상관함수 PACF(Partial Autocorrelation Function)

PACF 는 이름에서도 알 수 있는 것처럼 ACF 에 "Partial"이 추가된 함수입니다. 간단히 이야기하면, X_t 와 X_{t+k} 의 상관관계를 구할 때, 그 둘 사이에 존재하는 $\{\boldsymbol{X_{t+1}}, \boldsymbol{X_{t+2}}, \dots, \boldsymbol{X_{t+k-1}}\}$ **의 영향을 제외하고 구한 상관계수**입니다. 더 나은 이해를 위해 예시를 통해 부분상관계수(partial correlation coefficient)에 대해 먼저 알아보겠습니다!

X 를 아이스크림 판매량, Y 를 범죄발생건수, Z 를 인구수라고 설정하겠습니다. 만약 인구수 Z 가 시간에 따라 증가한다면 X 와 Y 역시 증가하게 되고, 그 결과 X 와 Y 의 상관관계가 높아질 것입니다. 하지만 이는 두 변수 모두 Z 의 영향을 받아 발생한 결과이기 때문에, Z 의 영향을 제거한 순수한 X 와 Y 의 상관계수를 구해야 합니다. 이때 사용하는 개념이 부분상관계수입니다.

인구수의 영향을 제거한 X 와 Y 의 부분상관계수는 아래 식을 통해 구할 수 있습니다.

$$\rho_{XY,Z} = \frac{E\{[X - E(X|Z)] \cdot [Y - E(Y|Z)]\}}{\sqrt{E[X - E(X|Z)]^2 \cdot E[Y - E(Y|Z)]^2}} = \rho_{X^*,Y^*}$$

조건부 기댓값 $E(X|Z)$: X 가 Z 에 의해 설명되는 부분

조건부 기대값 $E(Y|Z)$: Y 가 Z 에 의해 설명되는 부분

$X^* = X - E(X|Z)$: X 에서 Z 의 영향력을 제거하고 남은 부분

$Y^* = Y - E(Y|Z)$: Y 에서 Z 의 영향력을 제거하고 남은 부분

➔ X와 Y 의 부분상관계수 : $\rho_{XY,Z} = \rho_{X^*,Y^*}$

부분자기상관함수(Partial Autocorrelation Coefficient)는 부분상관계수(Partial Correlation Coefficient)에 "Auto"가 추가되었음을 알 수 있습니다. 즉, 자기 자신과의 부분상관계수입니다.

PACF 는 보통 $\alpha(k)$로 표현하며, 다음과 같이 정의합니다.

- $\alpha(0) = Corr(X_1, X_1) = 1$
- $\alpha(1) = Corr(X_2, X_1) = \rho(1)$
- $\alpha(k) = Corr(X_{k+1} - P_k^* X_{k+1}, X_1 - P_k^* X_1),\ \ k \geq 2$
  - ⇨ $P_k^* X_{k+1} = \ Best\ Linear\ Predictor\ on\ \{1, X_2, \cdots, X_k\}$
  - ⇨ $P_k^* X_1 = \ Best\ Linear\ Predictor\ on\ \{1, X_2, \cdots, X_k\}$

(BLP(Best Linear Predictor)은 뒤에서 더 자세히 다루겠지만, 지금 단계에서는 $\{1, X_2, \cdots, X_k\}$가 X_1과 X_{k+1}에 미치는 영향 정도로 생각해주시면 됩니다!)

주요 아이디어는 "중간값들의 영향력을 선형회귀로 추정하여 제거하자!"입니다. 식을 통해 더 자세히 알아볼까요?

구하고자 하는 X_{k+1}을 $X_1, \dots\ , X_k$를 이용한 회귀식으로 표현하면 다음과 같습니다.

- $X_{k+1} = \phi_{11} X_k + \ \epsilon_{k+1}$
- $X_{k+1} = \ \phi_{21} X_k + \phi_{22} X_{k-1} + \epsilon_{k+1}$

$$\vdots$$

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

위 식을 통해 X_{k+1}의 추정값인 BLP(Best Linear Predictor)을 구하고자 한다면, 오차항을 최소화하는 아래 방법을 이용할 수 있습니다.

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

위 회귀식의 계수인 $\boldsymbol{\phi_{kk}}$는 $\{X_{2,} \dots X_k\}$가 고정되어 있을 때, X_{k+1}과 X_1 간의 **선형적인 상관관계를 나타내는 수치**입니다. 따라서 PACF 는 다음과 같이 정의할 수 있습니다.

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

3. 선형과정(Linear Process)

$\{Z_t\} \sim WN(0, \sigma^2)$**들의 선형결합**으로 표현된 X_t를 선형과정이라고 합니다.

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

(이때 선형결합의 계수는 $\sum_j |\psi_j| < \infty$ (absolutely summable) 조건을 만족해야 합니다.)

선형과정은 아래 식과 같이 1 주차에 배운 후향연산자를 이용하여 표현할 수도 있습니다.

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

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

$$\because \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j B^j Z_t = \left(\sum_{j=-\infty}^{\infty} \psi_j B^j\right) Z_t$$

클린업 2 주차에 배우는 여러 모형들은 모두 선형과정인데요, 왜 시계열을 선형과정으로 표현하는 것일까요? 그 이유로는 4 가지를 들 수 있습니다.

1) 공분산 계산이 간단하다.

$$Cov(a_1Z_1 + a_2Z_2 + a_3Z_3, b_2Z_2 + b_3Z_3) = a_2b_2 + a_3b_3$$

$$Cov(a_1Z_1, b_2Z_2 + b_3Z_3) = 0$$

   위 식과 같이 인덱스가 동일한 경우만 계산해 공분산을 구할 수 있습니다.

2) 해석이 용이하며, 추정 방법이 잘 발달되어 있다. (regression/linear algebra)

3) 정상 확률 과정의 선형 결합은 다시 정상 확률 과정 조건을 만족한다.

   (증명 부록 참고!)

4) 모든 약한 의미의 정상 확률 과정은 선형 과정의 합과 결정적 과정으로 표현될 수 있다. (Wold decomposition)

그럼 지금부터 선형과정 모형인 AR, MA 그리고 ARMA 에 대해 배워보도록 하겠습니다.

4. AR(Auto Regressive Model) : 자기회귀 모형

4.1 정의

AR 모형이란 현 시점의 관측값을 과거 시점의 관측값과 현 시점의 오차의 선형결합 형태로 표현하는 모형입니다. $\boldsymbol{Z_t \sim WN(0, \sigma^2)}$ 이며, 아래의 식과 같이 표현할 수 있습니다.

AR(1) 모형

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

AR(p) 모형

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

위 식들은 우리가 흔하게 알고 있는 회귀식과 유사합니다. 이처럼 AR 모형은 관측값을 자기 자신의 과거로 회귀시킨다는 의미에서 '**자기회귀 모형**'이라고 부릅니다.

4.2 특성방정식

AR(p)에서 p는 몇 시점 전까지의 관측값을 사용했는지를 나타내는 모수입니다. P 시점 전까지의 관측값으로 나타낸 선형결합은 지난 시간에 배웠던 후향연산자를 사용해 표현할 수 있습니다.

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t \\ &= \phi_1 B X_t + \phi_2 B^2 X_t + \cdots + \phi_p B^P X_t + Z_t \end{aligned}$$

후향연산자를 이용해서 나타낸 식을 Z_t에 대해 다시 정리하면 아래와 같습니다.

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

$\boldsymbol{\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)}$ 를 AR(p)의 **특성방정식**(characteristic equation)이라고 부르며, $\boldsymbol{\phi(B)}$ 로 표현합니다. 특성방정식을 이용해 AR모형을 표현한 식은 아래와 같습니다.

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

AR(p)를 나타낸 식들을 정리하면 아래와 같습니다.

1) 정의를 이용해 나타낸 식

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

2) 후향연산자를 이용해 나타낸 식

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

3) 특성방정식을 이용해 나타낸 식

⇨ $Z_t = \phi(B) X_t$

4.3 AR 모형의 조건

AR 모형은 두 가지 조건을 만족해야 합니다.

1. **정상성(Stationarity)** : 시계열의 확률적 특성이 시점에 의존하지 않는 특성
2. **인과성(Causality)** : t 시점의 관측값이 과거시점의 오차항으로 설명되는 특성

$$\psi_j = 0\,, \forall j < 0 \;\leftrightarrow\; X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

선형과정 $\{X_t\}$가 위 조건을 만족할 때, $\{X_t\}$는 인과성을 가집니다.

AR 모형이 이 특성들을 어떻게 만족하는지 AR(1) 모형을 통해 확인해볼까요?

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + Z_t \\ &= \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = \phi_1^2 X_{t-2} + \phi_1 Z_{t-1} + Z_t \\ &= \phi_1^2(\phi_1 X_{t-3} + Z_{t-2}) + \phi_1 Z_{t-1} + Z_t = \phi_1^3 X_{t-3} + \phi_1^2 Z_{t-2} + \phi_1 Z_{t-1} + Z_t \\ &\vdots \\ &= \phi_1^{M+1} X_{t-M-1} + \Sigma_{j=0}^{M} \phi_1^j Z_{t-j} \end{aligned}$$

AR(1) 식은 위와 같이 과거 시점의 관측값과 오차들의 선형결합으로 정리할 수 있습니다. 인과성을 만족하기 위해서는 오차항만으로 관측값을 설명해야 합니다. 해당 조건의 만족 여부는 ϕ_1 범위에 따라 아래와 같이 달라집니다. 식을 통해 확인해볼까요?

1) $|\phi_1| < 1$

$X_t = \phi_1^{M+1} X_{t-M-1} + \Sigma_{j=0}^{M} \phi_1^j Z_{t-j}$ 에서 $M \rightarrow \infty$이면

$\phi_1^{M+1} X_{t-M-1} \rightarrow 0$ 으로 수렴, 나머지 부분은 $\Sigma_{j=0}^{\infty} \phi_1^j Z_{t-j}$ 이 됩니다.

위 결과와 같이 **$|\phi_1| < 1$일 때는 오차항의 선형결합(Z_t)**만 남아 인과성을 만족합니다. 또한 정상 시계열의 선형결합은 여전히 정상 시계열임을 확인하였기 때문에 정상성 역시 만족함을 알 수 있습니다.

2) $|\phi_1| = 1$

$$X_t = X_{t-1} + \varepsilon_t \quad or \quad X_t = -X_{t-1} + \varepsilon_t$$

이는 대표적인 비정상 확률 과정 중 하나인 확률보행과정(random walk process)입니다. *(확률보행과정은 부록 참고!)*

3) $|\phi_1| > 1$

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

위 식을 통해 알 수 있는 것처럼, $|\phi_1| > 1$일 경우 $\phi_1^{M+1} X_{t-M-1}$ 부분을 제거하지 못해 인과성을 만족하지 못합니다. 따라서 AR 모형 조건을 만족하지는 못하지만, 아래와 같은 흥미로운 내용을 이끌어낼 수 있습니다!

$$X_{t+1} = \phi_1 X_t + Z_{t+1}$$

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

$$X_t = \frac{1}{\phi_1} X_{t+1} - \frac{1}{\phi_1} Z_{t+1} = \frac{1}{\phi_1}\left(\frac{1}{\phi_1} X_{t+2} - \frac{1}{\phi_1} Z_{t+2}\right) - \frac{1}{\phi_1} Z_{t+1}$$

$$= \frac{1}{\phi_1^2} X_{t+2} - \frac{1}{\phi_1^2} Z_{t+2} - \frac{1}{\phi_1} Z_{t+1}$$

$$\vdots$$

$$= \left(\frac{1}{\phi_1}\right)^k X_{t+j} - \sum_{j=1}^{k} \left(\frac{1}{\phi_1}\right)^j Z_{t+j}$$

위 식과 같이 과거의 값이 아닌 미래 관측값들의 선형결합으로 표현할 수 있습니다. 하지만 이는 AR 모형으로는 사용할 수 없음을 기억해야 합니다.

위 내용을 정리하자면, AR모형은 정상성과 인과성이 모두 만족되는 $|\phi_1| < 1$가 성립할 때만 사용할 수 있습니다. 이는 '$\phi(B) = 0$의 근이 절대값이 1보다 커야한다'와 동치입니다.

→ 결론만 강조해도 ok!

4.4 AR 모형의 ACF

4.4에서는 마찬가지로 AR(1) 모형을 이용해 ACF를 계산해보도록 하겠습니다. 계산의 편의를 위해 $E(X_t) = 0$를 가정하겠습니다.

[1] ACF식을 유도하기 위해 AR(1) 식 양변에 X_{t-h} 를 곱해줍니다.

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

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

[2] 위 식에 기댓값을 취해 ACF 식을 구합니다.

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

$$(\because Cov(Z_t, X_{t-h}) = Cov\left(Z_t, \phi_1^{M+1} X_{t-h-M-1} + \Sigma_{j=0}^{M} \phi_1^j Z_{t-h-j}\right) = 0)$$

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

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

4.3에서 정상성을 만족하기 위해서 $|\phi_1| < 1$가 되어야 함을 확인했기 때문에, h가 커짐에 따라 **AR모형의 ACF는 지수적으로 감소함**을 알 수 있습니다.

4.5 AR 모형의 PACF

4.5에서는 AR(p) 모형식을 이용해 AR 모형의 PACF에 대해 알아보겠습니다.

AR(p)는 X_{k+1}을 p 시점 이전 값들로만 표현하는 것이며, 식은 아래와 같습니다.

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

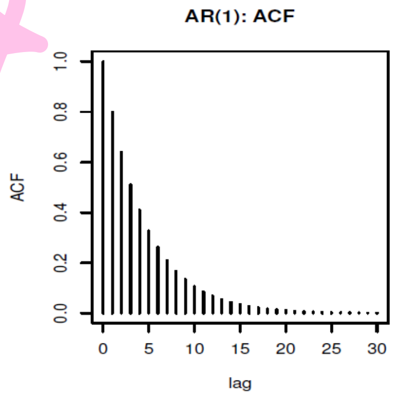
PACF를 유도하기 위해 위의 식을 조금 더 자세히 풀어서 써보면

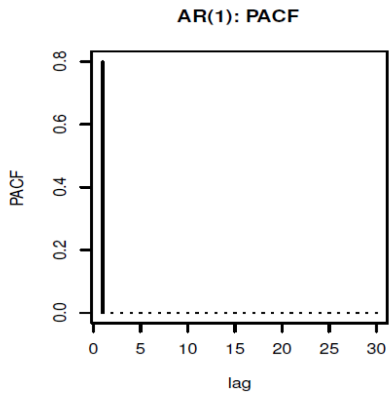
$$\hat{X}_{k+1} = \phi_1 X_k + \phi_2 X_{k-1} + \cdots + \phi_p X_{k+1-p} + 0X_{k-p} + \cdots + 0X_1$$

위와 같은 식을 완성할 수 있습니다. 이 식과 PACF의 성질을 이용해 AR 모형의 PACF를 정리하면 아래와 같습니다.

$$\alpha(0) \coloneqq 1,\ \ \alpha(p) = \phi_p \ \ \&\ \ \alpha(k) = 0\ \ if\ k > p$$

즉, AR 모형의 PACF는 p 시차 전까지만 존재하며, p 이후로는 모두 0이 됩니다. 이를 "**AR(p) 모형의 PACF는 시차 p 이후에 절단된다**"라고 표현합니다.





그림을 통해 AR 모형의 ACF와 PACF를 살펴보면, 위에서 확인한 바와 같이 **ACF는 지수적으로 감소, PACF는 p 이후(위 그림에서 p=1)에 절단**된 양상을 보이고 있습니다.

5. MA(Moving Average) : 이동평균모형

5.1 정의

MA 모형은 관측값을 과거시점의 오차항만을 이용해 설명하는 모형입니다. AR 모형과 마찬가지로 MA 모형의 Z_t 는 $\boldsymbol{Z_t \sim WN(0, \sigma^2)}$ 를 만족하는 백색잡음입니다. 식은 다음과 같습니다.

MA(1) 모형

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

MA(p) 모형

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

5.2 특성방정식

MA 모형 역시 AR 모형과 같이 후향연산자를 이용해 표현할 수 있습니다.

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$
$$= Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_t - \cdots + \theta_q B^q Z_t$$
$$= \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) Z_t$$

위 식에서 $\left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right)$ 를 MA(p)의 특성 방정식이라고 합니다. 따라서 MA(p)를 특성 방정식을 이용해 표현하면 다음과 같습니다.

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

지금까지 배운 MA(p)를 표현하는 식들을 정리하면 다음과 같습니다.

1) 정의를 이용해 나타낸 식

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

2) 후향연산자를 이용해 나타낸 식

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

3) 특성방정식을 이용해 나타낸 식

⇨ $X_t = \theta(B) Z_t$

5.3 MA 모형의 조건

MA모형은 AR모형의 조건에 **가역성(Invertibility)** 조건이 추가되어 총 3가지의 조건을 만족해야 합니다.

(cf. 가역성 조건이 추가된 이유는 상황에 맞는 모형을 적합시키기 위함입니다. 6장에서 더 자세히 배우겠지만, 우리는 ACF 또는 PACF를 통해 적절한 모형을 선정합니다. 이때 두 모형의 ACF가 같아 모형을 선정하기 어려운 경우를 대비해 조건을 추가했다고 이해하시면 됩니다!)

1) **정상성(Stationarity)** : 시계열의 확률적 특성이 시점에 의존하지 않는 특성

2) **인과성(Causality)** : t 시점의 관측값이 과거시점의 오차항으로 설명되는 특성

3) **가역성(Invertibility)** : t 시점의 오차항이 과거시점의 관측값으로 설명되는 특성

$\sum_{j=0}^{\infty} |\pi_j| < \infty$ 인 $\{\pi_j\}$가 존재하며, 아래 식을 만족할 때 가역성을 가진다.

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad for\ all\ t$$

(즉, Z_t에 대한 식을 X_t에 대한 식으로 표현할 수 있는지 여부를 따지는 조건이라고 이해하시면 됩니다!)

정의에서 확인한 것처럼 MA모형은 백색잡음을 만족하는 오차들의 선형결합이기 때문에 자연스럽게 인과성과 정상성을 만족합니다. 따라서 가역성 조건의 만족 여부만 확인해주면 됩니다!

가역성 조건 역시 가장 간단한 MA(1)을 통해 확인하도록 하겠습니다.

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

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

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

위와 같이 MA(1)식을 X_t에 대해 정리한 후 식을 풀어보면 무한등비급수의 합 형태로 표현할 수 있습니다. 따라서 가역성 조건 역시 $\boldsymbol{|\theta| < 1}$일 때만 성립하며, 이는 특성방정식 $\theta(B) = 0$의 근의 절댓값이 1 보다 커야 한다는 조건과 동치입니다.

5.4 MA 모형의 ACF

AR 모형에서와 같이, 양변에 X_{t-h}를 곱하고, 기댓값을 취하여 ACF를 구해주겠습니다.

[MA(1) 모형]

$$X_{t-h}X_t = X_{t-h}(Z_t + \theta_1 Z_{t-1}) = X_{t-h}Z_t + \theta_1 X_{t-h}Z_{t-1}$$

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

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

생략 가능

1) $h = 0$인 경우

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

2) $h = 1$인 경우

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

3) $h \geq 2$인 경우

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

이를 더 간단히 정리하면 다음과 같이 표현할 수 있습니다.

$$\rho(k) = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{\theta}{(1+\theta^2)}, & k = 1 \\ 0 \quad , & k \geq 2 \end{cases}$$

따라서 MA(q) 모형의 ACF는 **시차 q 이후 절단된다**는 것을 알 수 있습니다.

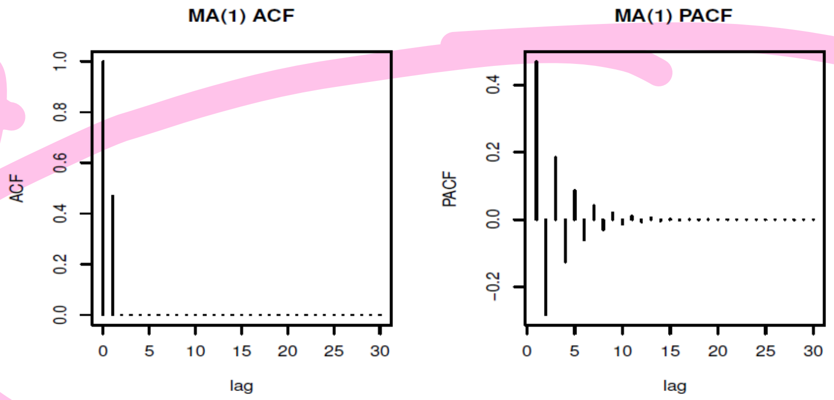
5.5 MA 모형의 PACF

MA 모형의 PACF는 Crammer 공식을 사용해 구할 수 있습니다.

(자세한 유도 과정은 부록 참고!)

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

위 식은 MA(1) 모형의 PACF를 계산한 결과입니다. 5.3에서 MA 모형은 $|\theta| < 1$일 때 성립함을 배웠으므로, 위 식은 **시차 k가 커질수록 0에 수렴**하게 됩니다.



MA(1) 모형의 ACF와 PACF 그래프입니다. 위에서 공식을 통해 확인한 것과 같이 **ACF는 시차 1까지 존재하며, PACF는 점점 0으로 수렴**하고 있음을 알 수 있습니다.

지금까지 AR 모형과 MA 모형에 대해 배웠습니다. 공부하면서 두 모형식이 굉장히 유사하다는 생각이 들지 않으셨나요? 그런 분들을 위해 두 모형은 서로에 의해 표현될 수 있다는 "쌍대성" 개념을 알아보도록 하겠습니다!

6. AR 모형과 MA 모형의 쌍대성

찍먹~ 가볍게~

위에서 언급한 것처럼 AR 모형과 MA 모형은 서로에 의해 표현될 수 있는 특성이 있습니다. 더 자세히 이야기하면, 유한차수의 AR 모형은 무한차수의 MA 모형으로 표현될 수 있으며, 유한차수의 MA 모형은 무한차수의 AR 모형으로 표현될 수 있습니다. 첫 번째 경우부터 알아보겠습니다.

1) AR(1) ⇨ MA(∞)

$$X_t = \phi_1 X_{t-1} + Z_t$$
$$= \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t$$
$$= \phi_1^2 X_{t-2} + \phi_1 Z_{t-1} + Z_t$$
$$= \phi_1^2(\phi_1 \mathrm{X}_{t-3} + \mathrm{Z}_{t-2}) + \phi_1 \mathrm{Z}_{\mathrm{t}-1} + \mathrm{Z}_{\mathrm{t}} = \phi_1^3 \mathrm{X}_{t-3} + \phi_1^2 \mathrm{Z}_{t-2} + \phi_1 \mathrm{Z}_{t-1} + \mathrm{Z}_t$$
$$= \cdots = \phi_1^{\mathrm{M}+1} \mathrm{X}_{\mathrm{t}-\mathrm{M}-1} + \Sigma_{j=0}^{M} {\phi_1}^j Z_{t-j}$$

AR 모형의 성질 및 조건에 의해 남은 항을 다시 정리해보면 $X_t = \Sigma_{j=0}^{\infty} \phi_1^j Z_{t-j}$ 로, 이는 MA 모형임을 알 수 있습니다.

2) MA(1) ⇨ AR(∞)

$$X_t = \theta_1 B Z_t + Z_t$$
$$X_t = (1 + \theta_1 B) Z_t \rightarrow Z_t = \frac{1}{1 - (-\theta_1 B)} X_t$$
$$(1 - \theta_1 B + \theta_1^2 B^2 - \cdots) X_t = Z_t$$
$$X_t - \theta_1 B X_t + \theta_1^2 B^2 X_t - \cdots = Z_t$$
$$X_t = \theta_1 B X_t - \theta_1^2 B^2 X_t + \cdots + Z_t$$

7 ARMA(p,q)

2주차 클린업의 마지막 모형인 ARMA 모형에 대해 배워보겠습니다! 이름에서 알 수 있는 것처럼 AR(p) 모형과 MA(q) 모형을 모두 포함하는 모형입니다. 두 모형을 함께 사용하는 만큼 ARMA 모형의 장점이 있겠죠? 그건 바로 **"모수의 절약"**입니다. AR 모형만으로 혹은 MA 모형만으로 시계열 자료를 설명할 경우 모수 p나 q가 너무 커질 위험이 있습니다. 추정해야 하는 모수가 많아지면 일반적으로 시간이 오래 걸리거나 효율성이 떨어지고, 해석이 어렵다는 문제가 발생합니다. 이러한 문제를 해결하기 위해 ARMA 모형을 사용하는 것입니다! 그럼 지금부터 ARMA 모형에 대해 자세히 알아볼까요?

7.1 ARMA 모형의 정의

앞서 설명한 바와 같이 ARMA 모형은 AR 모형과 MA 모형을 동시에 포함하는 모형이며, 식은 다음과 같습니다.

[ARMA(p,q) 모형]

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

7.2 ARMA 모형의 특성방정식

ARMA(p,q) 모형의 특성방정식은 다음과 같습니다.

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

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

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

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

(왼쪽 식은 AR 모형의 특성방정식이며, 오른쪽 식은 MA 모형의 특성방정식입니다!)

7.3 ARMR 모형의 조건

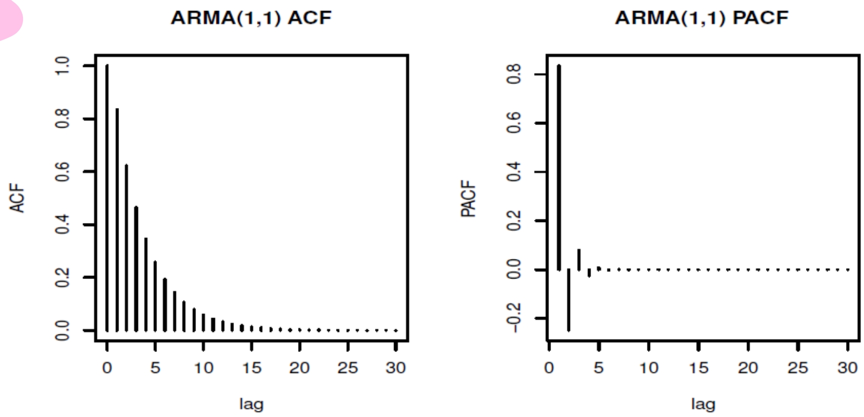
ARMA 모형은 정의에서 유추할 수 있는 것처럼 AR 모형의 조건과 MA 모형의 조건을 모두 만족해야 합니다. 즉, 정상성과 인과성, 가역성을 모두 만족해야 합니다. 여기서 끝나면 정말 좋겠지만.. ARMA 모형은 **식별성(Identifiability)** 조건까지 만족해야 합니다. 식별성이란, 주어진 파라미터 조합에 대해 단 하나의 모형이 대응되는 특성을 의미합니다.

식을 통해 확인해보겠습니다!

$$X_t = Z_t$$

위 식은 $(1-\phi B) = (1+\theta B) \iff \phi = -\theta$를 만족하는 **ARMA(1,1)** 모형입니다. 하지만 동시에 **WN** 이기도 합니다. 이처럼 정상성, 인과성, 가역성만으로는 정확한 모형을 식별할 수 없는 문제가 발생하기 때문에 $\boldsymbol{\phi+\theta \neq 0}$의 식별성 조건이 필요한 것입니다.

7.4 ARMA 모형의 ACF 와 PACF



ARMA 모형의 ACF 그래프와 PACF 그래프는 **모두 지수적으로 감소하거나 싸인함수 형태로 소멸**되는 양상을 보입니다. 여기에서 ARMA 모형의 한계를 발견할 수 있습니다. AR(p) 모형은 PACF가 p+1차부터 절단되었고, MA(q) 모형은 ACF 그래프가 q+1차부터 절단되는 특징이 있었습니다. 하지만 ARMA 모형은 두 그래프 모두 지수적으로 감소하기 때문에 모형 식별을 위한 추가적인 방법이 필요합니다. 해당 방법은 모형 식별 부분에서 다루도록 하겠습니다.

8. 모형의 적합절차

지금까지 WN이 아닌 정상 시계열의 오차를 설명하기 위한 모형인 AR, MA, ARMA에 대해 배웠습니다. 그렇다면 언제, 어떤 모형을 사용해 시계열 자료를 설명해야 할까요? 지금부터 상황에 맞는 모델을 선택하기 위한 모형 적합 절차에 대해 알아보겠습니다! 2주차의 마지막 내용이니 조금만 더 힘을 내봅시다~!

8.1 모형 식별

가장 먼저 사용할 모형과 해당 모형의 차수를 결정해야 합니다. AR모형과 MA 모형은 ACF 혹은 PACF를 확인해 쉽게 차수까지 결정할 수 있습니다. 하지만 앞서 설명했던 것과 같이 ARMA 모형은 두 그래프 모두 지수적으로 감소하기 때문에 모형 식별을 위해 다른 방법이 필요합니다. 이때 AIC, AICC, BIC 등의 **Information Criteria(IC)**를 사용합니다.

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

(iid 가정)

2) 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)

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

여러 모형 중에서 IC를 가장 작게 만드는 모형을 선택해 분석을 진행하면 됩니다!

8.2 모수 추정

모수의 차수를 결정했다면, 파라미터 $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, σ^2 를 추정해야 합니다. 일반적으로 아래의 3가지를 이용해 파라미터를 추정합니다.

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

8.3 모형 진단

모형을 식별하고 모수 추정까지 마쳤다면, 모형이 적합한지 진단해야 합니다. 모형 진단은 모수에 대한 검정과 잔차에 대한 검정으로 진행됩니다.

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 $\neq 0$ 인지 확인

2) 잔차에 대한 검정

- 추세, 계절성, 이상치가 없는지 확인
- WN(0, σ^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

8.4 예측

모형 식별, 파라미터 추정, 모형 진단까지 마쳤다면, 드디어 예측을 진행합니다!

예측에는 과거의 모든 정보를 알고 있다고 가정하는 infinite한 방법과, 알고 있는 자료를 사용해 예측하는 finite한 방법이 있습니다. 클린업에서는 실제로 주로 사용되는 **finite**한 방법을 알아보도록 하겠습니다!

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

위 식은 n개의 자료를 이용해 n+h 시점을 예측했음을 의미합니다. 즉, 가지고 있는 데이터의 **선형결합**을 활용해 미래를 예측하는 방법입니다.

계수 $\{a_0, a_1, \dots, a_n\}$는 MSPE(Mean Squared Prediction Error)를 최소화하는 방향으로 추정합니다.

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

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

계산 방식은 회귀분석에서의 LSE와 동일합니다!

~~[목록]~~

1 “정상 확률 과정의 선형 결합은 다시 정상 확률 과정 조건을 만족한다.” 증명

Proposition (Introduction to Time Series and Forecasting Ch.2 - 2.2.1)

Let $\{Y_t\}$ be any (weakly) stationary process with mean zero and ACVF γ_Y. Now define new process

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

Is again stationary process with mean zero and

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j)$$

In particular, if $\{Y_t\} \sim WN(0, \sigma^2)$, then

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

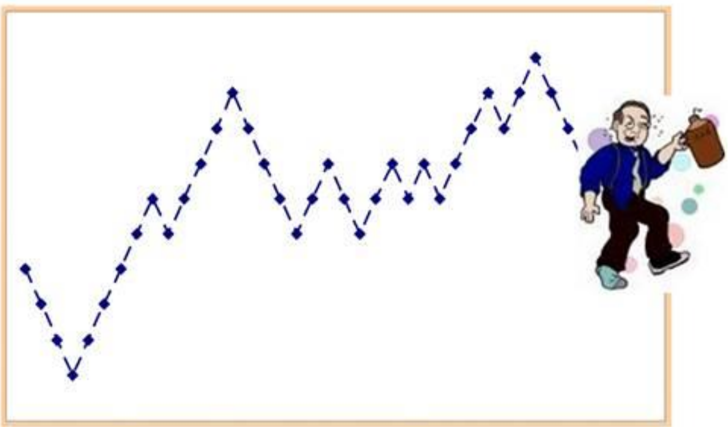
2 확률보행과정(Random Walk Process)

$\varepsilon_t \sim WN(\mu, \sigma^2)$일 때, $X_t = X_{t-1} + \varepsilon_t$ 과 같이 정의되는 과정을 확률보행과정이라고 합니다.

ε_t를 어떤 사람이 임의로 움직이는 보폭, X_t를 t 시점에서의 위치라 생각한다면, X_t는 ε_t들의 누적합으로 표현됩니다.

$$X_0 = \mu, \ X_1 = \varepsilon_1, \ X_2 = X_1 + \varepsilon_2 = \varepsilon_1 + \varepsilon_2, \ \cdots \ , X_t = \sum_{i=1}^{t} \varepsilon_t$$

확률보행과정은 평균은 μ로 시간에 의존하진 않지만, 자기공분산이 시간에 의존하여 비정상 확률과정입니다.



3 Crammer 공식을 통해 유도한 MA 모형의 PACF

우선 Crammer 공식을 간단히 소개하도록 하겠습니다.

연립 일차방정식 Ax=b가 있을 때, 유일한 해는 다음과 같이 나타낼 수 있습니다.

이때 A는 정사각 행렬이며, 행렬식은 0이 아님을 가정합니다.

$$x_j = \frac{\det A_j}{\det A} = \frac{\begin{vmatrix} a_{11} & \cdots & b_1 & \cdots & a_{1n} \\ a_{21} & \cdots & b_2 & \cdots & a_{21} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & b_n & \cdots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{21} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix}} \quad (j = 1,2,\cdots,n)$$

(위 식에서 A_j는 A의 j번째 열을 B로 대신하여 구한 행렬입니다.)

다음과 같은 방법으로 해를 구하는 방법을 Crammer 법칙이라고 합니다. 그럼 이제 Crammer 공식을 이용해 MA(1) 모형의 PACF를 구해보도록 하겠습니다.

1) K=1

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

$$\alpha(1) = \phi_{11} = \rho(1) = \frac{\theta}{(1+\theta^2)}$$

2) K=2

$$\phi_2 = \Gamma_2^{-1}\gamma_2 \ \rightarrow \ \Gamma_2\phi_2 = \gamma_2$$

$$\Gamma_2 = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}, \quad \gamma_2 = [\gamma(1), \gamma(2)]', \qquad \phi_2 = [\phi_{12}, \phi_{22}]'$$

$$\phi_{22} = \frac{\begin{vmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(2) \end{vmatrix}}{\begin{vmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{vmatrix}} = \frac{\begin{vmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 \\ \theta\sigma^2 & 0 \end{vmatrix}}{\begin{vmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 \end{vmatrix}} = \frac{-\theta^2\sigma^4}{\{(1+\theta^2)\sigma^2\}^2 - \{\theta\sigma^2\}^2} = \frac{-\theta^2}{1+\theta^2+\theta^4}$$

3) K=3

$$\phi_3 = \Gamma_3^{-1}\gamma_3 \ \rightarrow \ \Gamma_3\phi_3 = \gamma_3$$

$$\Gamma_3 = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix}, \quad \gamma_3 = [\gamma(1), \gamma(2), \gamma(3)]', \qquad \phi_3 = [\phi_{13}, \phi_{23}, \phi_{33}]'$$

$$\phi_{33} = \frac{\begin{vmatrix} \gamma(0) & \gamma(1) & \gamma(1) \\ \gamma(1) & \gamma(0) & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(3) \end{vmatrix}}{\begin{vmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{vmatrix}} = \frac{\begin{vmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 & \theta\sigma^2 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 & 0 \\ 0 & \theta\sigma^2 & 0 \end{vmatrix}}{\begin{vmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 & 0 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 & \theta\sigma^2 \\ 0 & \theta\sigma^2 & (1+\theta^2)\sigma^2 \end{vmatrix}}$$

$$= \frac{\theta^3\sigma^6}{\{(1+\theta^2)\sigma^2\}^3 - 2(1+\theta^2)\sigma^2\{\theta\sigma^2\}^2}$$

$$= \frac{\theta^3}{1+\theta^2+\theta^4+\theta^6}$$

위와 같은 과정을 반복하여 규칙을 발견하고 정리한 PACF는 다음과 같습니다.

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