Time series analysis

Contents

0. Remind

1 주차 세미나 하시느라 다들 수고하셨습니다!! 벌써 클린업 중반에 접어들었네요,, 이번주에는 드디어 시계열 모형들에 대해 본격적으로 배우는 시간입니다. 모두 힘을 내세요!!

그 전에 오늘 필요한 몇 가지 통계량 계산 방식을 복습하겠습니다.

- ACF

  $\gamma_k = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$

- ACVF

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

- $\boldsymbol{Cov(X,Y) = E(XY) - E(X)E(Y)}$

- $\boldsymbol{Cov(ax + by + c, dx + fy + e)}$

  $= adVar(x) + afCov(x,y) + bdCov(x,y) + bfVar(y)$

- 무한등비급수

  $\lim_{n\to\infty}(a + ar + ar^2 + ar^3 + \cdots) = \frac{a}{1-r} \quad iff\ |r| < 1$

- 조건부 확률

  $P(A|B) = \frac{P(A\cap B)}{P(B)}$

I. 1주차 복습

1. 정상성

정상성은 시계열의 확률적 성질들이 시간에 의존하지 않는 성질이었습니다.

i)	$E[\|X_t\|]^2 < \infty \ , \forall t \in \mathbb{Z}$
ii)	$E[X_t] = m \ , \forall t \in \mathbb{Z}$
iii)	$\gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

여기서 가장 핵심은 시점"t"에 확률적 성질이 의존하지 않는다! 입니다.

2. 정상화

정상화는, 비정상시계열을 정상시계열로 바꿔주는 과정을 의미했습니다.

1) 분산이 일정하지 않은 경우 ⇨ VST (로그변환, 제곱근변환, Box-Cox 변환)

2) 평균이 일정하지 않은 경우 ⇨ 회귀, 평활, 차분

3. 모형의 필요성

지난주, 오늘 모형에 대해 배울 예정이니 모형의 필요성에 대해 소개하였습니다. 여기서 중요한 개념 또 하나가 바로 백색잡음이었죠!!

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

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

백색잡음인 경우 ⇨ 대각요소인 분산을 제외한 나머지 요소가 모두 0이고,

백색잡음이 아닌 경우 ⇨ 분산 이외의 공분산을 추정해주어야 하는데 이때, 모형을 활용해 추정한다고 했습니다.

II. 모형의 식별

다시 정리하고 넘어가자면, 회귀/평활/차분 등의 방법을 통하여 비정상 시계열 ⇨ 정상시계열로 변환하여 주었지만, 여전히 IID 나 WN 은 아닐 때, $\gamma(h)$ $(h \neq 0)$들을 추정해 주어야 하고, 이때 모델을 통하여 추정하는 것입니다. 오늘 배울 AR, MA, ARMA 를 다음의 ACF 와 PACF 의 패턴을 통해 구분할 수 있습니다.

1. 자기상관함수 ACF(Autocorrelation Function)

지난주에 배워 이제는 익숙해졌겠죠? 그러니 오늘은 ACF 의 의미에서 더 나아가 ACF 의 특징에 대해 주로 소개하겠습니다. $\rho(h)$는 시차가 h인 시계열 간의 상관관계를 의미하며, 정상성을 만족할 경우 "시점"이 변해도 이 값은 변하지 않으며 오직 "시차"에만 의존해야 한다! 기억하시죠? ACF 는 다음과 같은 특징을 가진 함수입니다.

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

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

$\because$

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

⇨

2. 부분자기상관함수 PACF(Partial Autocorrelation Function)

Covariance 에 “auto”가 붙은 건 어떤 의미인지 지난 주에 배웠는데, 이번주에는 “partial”이 한 개 더 붙었네요! 그렇다면 다 이유가 있어서 붙었겠죠? “partial”의 의미를 알아보기 위해 부분상관계수(partial correlation coefficient)에 대해 알아보겠습니다.

Partial Correlation Coefficient

> 시계마을에는 매년 새로운 시계들이 탄생한다고 해봅시다. 이 시계마을에서 시계들이 힘을 얻고 싶을 때 건전지를 산다고 합시다. 이때, X를 건전지의 판매량, Y를 시계들의 범죄발생건수라고 해봅시다. X와 Y의 상관계수를 구해 보면 매우 상관이 높은 것으로 나타날 것입니다. 왜 그럴까요? 시간이 지남에 따라 새로운 시계들은 계속 태어나고 이에 따라 당연히 건전지의 판매량과 시계들의 범죄발생건수도 증가했기 때문입니다. 따라서 X와 Y의 순수한 상관관계를 구하기 위해서는 “시간”의 효과를 제거한 후 상관계수를 구해줘야 합니다. 이때, 이를 부분상관계수(partial correlation coefficient)라고 합니다. 예시가 왜 이래 푸핫

시간을 Z 라고 할 때, Z 의 효과를 배제한 X 와 Y 의 부분상관계수는 다음과 같습니다.

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

- 조건부 기댓값 $E(X|Z)$ = X 를 Z 에 회귀시킨 최적선형 예측값으로 X 가 Z 에 의해 설명되는 부분

  $X^* = X - E(X|Z)$: X 를 Z 에 회귀시킨 후의 잔차 ⇨ Z 의 영향력을 제거

- 조건부 기댓값 $E(Y|Z)$ = Y를 Z에 회귀시킨 최적선형 예측값으로 Y 가 Z에 의해 설명되는 부분

  $Y^* = Y - E(Y|Z)$: Y 를 Z 에 회귀시킨 후의 잔차 ⇨ Z 의 영향력을 제거

  ⇨ X 와 Y 의 부분상관계수 $\rho_{XY,Z}$ = ρ_{X^*,Y^*}

여기까지가 “auto”가 붙지 않은 부분상관계수였습니다. 그렇다면 부분자기상관함수는 자기자신과의 부분상관계수 이겠지요?

Partial Autocorrelation Coefficient

즉, X_t와 X_{t+k}의 상관관계를 구할 때, 이 둘 사이에 존재하는 $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$의 영향을 제거한 후 구한 상관 계수입니다.

“Conditional correlation of X_1 *and* X_{k+1} *given intermediate values* $X_2, X_3, \dots, X_k$*”*

부분자기상관함수(partial autocovariance function ; PACF)를 구하는 식은 다음과 같습니다. PACF 를 $\alpha(k) = Cov(X_1, X_{1+k})$ 라고 할 때, $X_2, \dots, X_k$까지의 영향을 제거하여야 한다고 했습니다.

⇨ 그래서 저희는 두 시계열 사이에 존재하는 시계열들의 영향력을 어떻게 제거하여 주어야 할지 먼저 알아보겠습니다.

기본적인 흐름은 "X_1에서 $X_2, \dots, X_k$의 영향을 선형회귀로 추정해서 제거해주면 순수한 X_1만 남겠지?"입니다. 식을 통해 더 구체적으로 알아보겠습니다.

X_{k+1}을 $X_1, \dots, x_k$로 회귀식을 표현하면 다음과 같이 표현할 수 있습니다.

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

이때, ϕ_{11}는 $\rho(1)$라고 생각할 수 있습니다. 왜냐하면! X_{k+1}과 X_k의 선형적인 연관성을 의미하기 때문입니다.

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

$$\vdots$$

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

이때, $\{X_1, \dots, X_k\}$ 를 이용하여 추정한 X_{k+1}의 BLP(Best Linear Predictor)는 오차항을 최소화하는 추정치로 다음과 같이 표현할 수 있습니다.

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

위 회귀식의 계수 ϕ_{kk}는 $\{X_{2,} \dots X_k\}$가 고정되어 있을 때, X_{k+1}과 X_1간의 선형적인 상관관계를 나타내는 수치이므로 PACF 는 다음과 같이 표현할 수 있습니다.

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

이 부분 회귀계수의 해석이 어렵다면 회귀 1주차 복습 gogosing~

III. 선형과정 (Linear Process)

선형과정은 한 마디로 하면 $\{Z_t\} \sim WN(0, \sigma^2)$들의 선형결합입니다. 이를 식으로 표현하면 다음과 같습니다.

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

이때, 선형결합의 계수는 $\sum_j |\psi_j| < \infty$의 조건을 만족해야 합니다. 선형과정의 1)공분산 계산은 매우 간단하며, 2)해석과 추정이 잘 발달되어 있으며 3)**주어진 정상 확률 과정의 선형 결합은 또 다시 정상 확률 과정**이라는 성질이 있어 매우 유용합니다. 또한 4) Wold Decomposition에 따르면, 모든 약한 의미의 정상확률과정은 선형과정과 결정적 과정으로 나타낼 수 있습니다. 오늘 배울 AR과 MA 또한 선형과정의 모형 중 하나입니다.

IV. AR(Auto Regressive Model): 자기회귀 모형

1. 정의

현 시점의 관측값을 과거 관측값과 현 시점의 오차의 함수 형태로 나타내는 모형입니다. 이렇게 관측값이 자기 자신의 과거에 회귀시킨다는 의미에서 자기회귀라는 표현을 사용하는 모델입니다. $Z_t \sim WN(0,\sigma^2)$일 때, 다음과 같이 표현할 수 있습니다.

AR(1)모형

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

AR(p)모형

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

2. 특성방정식

식에서도 알 수 있듯이 AR(p)의 의미는 p 시점 전까지의 관측값의 선형결합으로 표현함을 의미합니다. 1 주차에 배웠던 후향연산자(Backshift Operator)를 기억하시나요? 후향연산자를 활용하여 AR(p) 모형을 간단하게 표현할 수 있습니다.

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

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

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

여기서, $(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$를 **특성방정식(characteristic equation)** $\phi(B)$라고 합니다. 이 특성방정식을 사용하여 AR모델을 다음과 같이 표현할 수 있습니다.

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

3. AR 모형의 조건

AR 모형은 다음과 같은 두 가지 특성을 만족해야 합니다.

1) 정상성 : 시계열의 확률적 특성이 시점에 의존하지 않아야 하는 특성

2) 인과성(Causality) : t시점의 관측값이 과거시점의 오차항으로 설명될 수 있다는 특성

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

AR모형이 이 두 가지 특성을 어떻게 만족하는지 가장 간단한 AR(1) 모형을 통해 알아보겠습니다. 다들 심호흡 한번 하시고,,

[AR(1) 모델]

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

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 → ∞ 이면

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

이때, 정상시계열의 선형결합은 여전히 정상 시계열이므로 **정상성**을 만족하고, 동시에 과거의 값에만 의존하고 있다는 점에서 **인과성**도 만족함을 알 수 있습니다.

2) $|\phi_1| = 1$

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

이는 대표적인 비정상 확률 과정 중 하나인 확률보행과정(random walk process)입니다.

[참고] ε_t를 어떤 사람이 임의로 움직이는 보폭이라고 하면 원점($X_0 = 0$)에서 출발하여 t 시점에서의 현 위치를 의미합니다. 이 확률 과정의 경우 평균은 0 으로 시간에 의존하진 않지만 자기공분산이 시간에 의존하여 비정상 확률과정입니다.

3) $|\phi_1| > 1$

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

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

$$
\begin{aligned}
X_t &= \frac{1}{\phi_1} X_{t+1} - \frac{1}{\phi_1} Z_t = \frac{1}{\phi_1}\left(\frac{1}{\phi_1} X_{t+2} - \frac{1}{\phi_1} Z_{t+2}\right) - \frac{1}{\phi_1} Z_{t+1} \\
&= \frac{1}{\phi_1}\left(\frac{1}{\phi_1} X_{t+3} - \frac{1}{\phi_1} Z_{t+3}\right) \frac{1}{\phi_1^2} Z_{t+2} - \frac{1}{\phi_1} Z_{t+1} \\
&\vdots \\
&= \left(\frac{1}{\phi_1}\right)^m X_{t+j} - \sum_{j=1}^{m} \left(\frac{1}{\phi_1}\right)^j Z_{t+j}
\end{aligned}
$$

이때, $m \to \infty$ 일 때 $\left(\frac{1}{\phi_1}\right)^m X_{t+j}$은 0 으로 수렴하고 $\Sigma_{j=1}^{m} \left(\frac{1}{\phi_1}\right)^j Z_{t+j}$은 미래의 오차항들의 결합으로 과거시점의 오차항으로 설명되어야 한다는 인과성을 만족하지 못합니다.

⇨ AR 모형은 $|\phi| < 1$ 일 때 **인과성**과 **정상성**을 만족합니다. 또한, 이는 '$\phi(B) = 0$의 근이 절대값이 1 보다 커야한다' 와 동치입니다.

4. ACF

AR(1)모형의 ACF 를 계산해보겠습니다. 이때, 계산의 편의를 위해 $E(X_t) = 0$임을 가정합니다.

Key idea : 양변에 X_{t-h}를 곱해 기댓값을 취하자!!

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

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

이제 여기에 양변에 기댓값을 취해보겠습니다.

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

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

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

⇨ 정상성을 만족시키는 경우 AR 모형은 $|\phi| < 1$ 이기 때문에

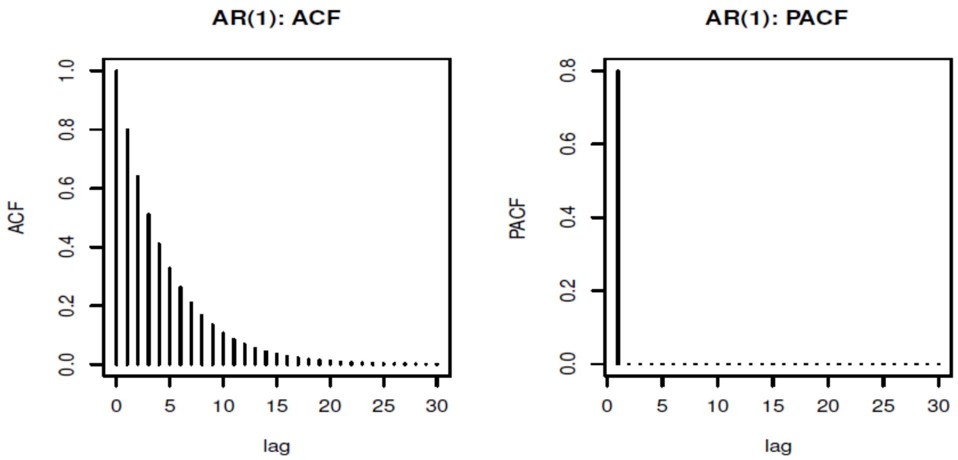
h ↑ → ACF 는 지수적으로 감소합니다.

5. PACF

AR(1)모형의 PACF 에 대하여 알아보겠습니다. AR(p)모형의 경우 X_{k+1} 을 p 시점 이전의 값인 $X_{k+1-p}, \dots, X_k$로만 표현하는 모델입니다.

따라서, $\hat{X}_{k+1} = \phi_1 X_k + \phi_2 X_{k-1} + \cdots + \phi_p X_{k+1-p} + 0X_{k-p} + \cdots + 0X_1$ 이며 $\alpha(p) = \phi_p$ 이며 p 이후로는 PACF 가 0 이 된다는 사실, 이해가시나요?

↓ PACF 를 구하는 방식을 다시 떠올려봅시다.



AR(1)모형의 ACF 및 PACF입니다. 위에서 확인해본 바와 같이 ACF는 지수적으로 감소하고 PACF는 p이후, 즉 여기선 1 이후에 절단된 양상임을 알 수 있습니다.

V. MA(Moving Average): 이동평균모형

1. 정의

AR 모형이 과거의 관측값과 현재의 오차를 통해 현재의 관측값을 설명하는 모형이라면, 이동평균 모형은 과거시점의 오차항을 이용하여 관측값을 설명하는 모델입니다. 표현은 다음과 같습니다. AR 모형과 마찬가지로 $Z_t \sim WN(0, \sigma^2)$입니다.

MA(1)모형

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

MA(q)모형

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

2. 특성방정식

MA 모형은 후향연산자 B 를 사용하여 다음과 같이 표현 가능합니다.

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

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

$$= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) Z_t$$

자, 그렇다면 이제 MA(q)의 특성방정식은? ⇨ $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

어렵지 않죠? 따라서 $X_t = \theta(B) Z_t$와 같이 표현할 수 있습니다.

3. 모형의 조건

MA 모형은 백색잡음이라고 가정하는 오차들의 선형결합, 그 중에서도 t 시점 이전의 오차항들로 표현하는 모델이기 때문에 **정상성**과 **인과성**을 만족합니다. 그럼 끝인걸까요? 아닙니다!! **가역성**이라는 조건도 만족되어야 합니다. 가역성은 처음 등장하죠?

가역성(Invertibility)

: t 시점의 **오차항**을 **과거 시점의 관측값**으로 표현할 수 있는 성질입니다. 이러한 성질은 예측에 유용합니다. 다음의 식을 만족하면 확률 과정이 invertible 하다고 할 수 있습니다.

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad for\ all\ t \quad (\text{단}, \sum_{j=0}^{\infty} |\pi_j| < \infty)$$

그렇다면 가역성은 어떤 조건에서 만족하는지 MA(1)모형을 통해 알아보겠습니다.

$X_t = \theta(B)Z_t$ ⇨ $\theta(B)^{-1}X_t = Z_t$ 이렇게 표현 가능한 거 다 이해하시나요? 그렇다면 이 식이 MA(1) 모형의 경우엔 다음과 같습니다.

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

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

⇨ 이러한 전개는 $|\theta| < 1$일 때만 성립합니다. 따라서, **MA 모형은 $|\theta| < 1$일 때 가역성을 만족한다**는 사실을 알 수 있습니다. 또한 이 말은 **$\theta(B)$의 근의 절댓값이 1 보다 크다**는 것을 의미합니다.

여기서 잠깐!

가역성은 왜 필요할까? AR모형은 가역성을 만족하지 않아도 되는걸까?

먼저 가역성이 왜 필요한 조건인지, 가역성의 의미가 와닿지 않을 것 같아 설명하겠습니다. 두 모형의 ACF가 동일할 때 ACF만을 갖고는 어떤 모형을 적합시킬지에 대한 판단이 어려울 수 있습니다. 이러한 경우, ACF와 모형의 일대일 관계를 성립시켜주기 위하여 제약 조건을 가해 준 것인데 이 조건을 **"가역성(invertibility)"**이라고 합니다. 일반적으로 MA(q)모형의 경우 2^q개의 모형이 동일한 ACF형태를 가질 수 있으며, 이 중에서 하나의 모형만을 선택하기 위해서는 $\theta(B) = 0$의 근의 절댓값이 1보다 크다는 조건을 부여해 주면 됩니다.

그렇다면 두번째 물음! AR모형에서는 왜 가역성을 만족하는지 확인하지 않고 넘어갔을까요?

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

AR모형에서는 별도의 조건이 없어도 우변에서 Z_t 즉, 오차항을 제외한 나머지 항들을 좌변으로 이항시키면 그 자체로 가역성이 만족되기 때문입니다.

4. ACF

앞서 AR 모델에서 ACF 구했던거 다들 기억하시나요? 핵심 아이디어는 같습니다! X_{t-h}를 곱하고, 기댓값을 취하여 ACF 를 구해주겠습니다.

[MA(1) 모형]

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

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

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

- $h = 0$ 인 경우

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

- $h = 1$ 일 경우

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

- $h \geq 2$ 일 경우

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

  ⇨ MA(1) 모형의 ACF 는 다음과 같이 정리할 수 있습니다.

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

  ⇨ MA(q) 모형의 ACF 는 시차 q 이후 절단된 모양입니다.

5. PACF

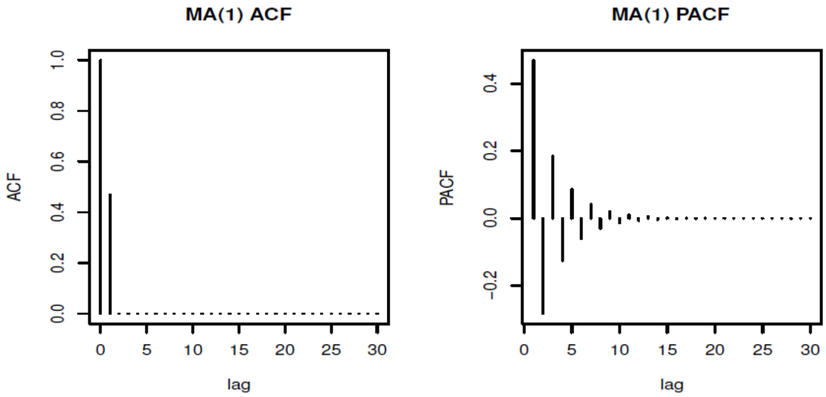
다음은 MA 모형에서 PACF 는 어떤 패턴을 보이는지 알아보겠습니다.

[MA(1) 모형]

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

⇨ MA 모형에서는 PACF 가 지수적으로 감소함을 알 수 있습니다.

위의 MA 모델의 PACF 는 Crammer 공식을 사용하여 얻어진 결과입니다. 그러나 우리는 계산 과정을 자세히 다루지는 않으려고 합니다.



위 그림은 MA(1)의 ACF 및 PACF입니다. 앞서 구한 것 과 같이 ACF는 p시점 이후 절단, PACF는 지수적으로 감소함을 알 수 있습니다.

VI. AR, MA의 쌍대성(Duality)

1. 의미

1)유한차수의 AR ⇨ 무한차수의 MA 로 표현 가능하고, 2)유한차수의 MA ⇨ 무한차수의 AR 로 표현가능합니다.

1) AR(1) ⇨ MA(∞)

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

$$= \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t$$

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

$$= \phi_1^2(\phi_1 X_{t-3} + Z_{t-2}) + \phi_1 Z_{t-1} + Z_t = \phi_1^3 X_{t-3} + \phi_1^2 Z_{t-2} + \phi_1 Z_{t-1} + Z_t$$

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

이때, AR(1)모형의 ϕ 의 절댓값이 1 보다 작고 $M \to \infty$ 일때 $\phi_1^{M+1} X_{t-M-1} \to 0$ 인거 기억나시죠!

남은 항을 정리해보면 $X_t = \Sigma_{j=0}^{\infty} \phi_1^j Z_{t-j}$ ⇨ 즉, 과거시점의 오차항들의 선형결합인 무한 차수의 MA 모형이 되는 것을 확인할 수 있습니다. 후향연산자 B 를 이용해 다시 표현해 보겠습니다.

$$X_t = \phi_1 B X_t + Z_t$$

$$X_t - \phi_1 B X_t = Z_t$$

$$(1 - \phi_1 B) X_t = Z_t$$

$$X_t = \frac{1}{(1 - \phi_1 B)} Z_t$$

$$X_t = (1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots) Z_t$$

2) MA(1) ⇨ AR(∞)

$$X_t = -\theta_1 B Z_t + Z_t$$

$$X_t = (1-\theta_1 B)Z_t, \rightarrow \ Z_t = \frac{1}{1-\theta_1 B}X_t$$

$$(1+\theta_1 B + \theta_1^2 B^2 + \cdots)X_t = Z_t$$

$$X_t + \theta_1 B X_t + \theta_1^2 B^2 X_t + \cdots = Z_t$$

$$X_t = -\theta_1 B X_t - \theta_1^2 B^2 X_t - \cdots - Z_t$$

2. ACF/PACF

유한차수의 AR과정의 ACF와 유한차수의 MA과정의 PACF는 지수적으로 감소하는 형태를, 유한차수의 AR과정의 PACF와 유한차수의 MA과정의 ACF는 절단 형태를 보입니다.

3. 모형의 조건

유한자수의 AR과정에 대해서 가역성 조건은 필요하지 않지만 **정상성**을 위해서는 $\phi(B) = 0$의 근들이 단위원 밖에 존재해야 한다는 조건이 필요합니다. 반면, 유한차수의 MA과정에 대해선 정상성 조건은 필요하지 않으나 **가역성**을 위해서는 $\theta(B) = 0$의 근들의 절댓값이 1보다 커야 한다는 조건이 필요합니다!

VII. ARMA(p,q)

ARMA모형은 이름 그대로 AR 모형과 MA 모형을 동시에 포함하는 모형입니다. 그렇다면 왜 두 모형을 통합한 모형이 또 필요한 것일까요? 이는 "**모수의 절약**(parsimony)"와 관련이 있습니다. 시계열자료를 순수한 자기회귀모형이나 이동평균모형으로만 설명하려면 p나 q의 값이 너무 커질 가능성이 있습니다. 이때, 추정해야 할 모수의 개수가 많아지면 일반적으로 추정의 효율성이 떨어지고 해석도 쉽지 않으므로 적합한 모형을 선택하는 기준의 하나가 바로 모수의 개수가 적은가? 입니다.

1. 정의

ARMA(p,q)모형

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

2. 특성방정식

ARMA(p,q)모형의 특성방정식은 다음과 같습니다.

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

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

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

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

3. 모형의 조건

ARMA 모형은 AR 모형과 MA 모형의 조건을 모두 만족시겨야 합니다. 따라서 AR 의 정상성 및 MA 의 가역성을 둘 다 만족해야 합니다.

ex)ARMA(1,2)

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

이때, **정상성**을 만족시키기 위해 $\boldsymbol{\phi(B) = 0}$ 의 근의 절댓값이 1 보다 커야하고,

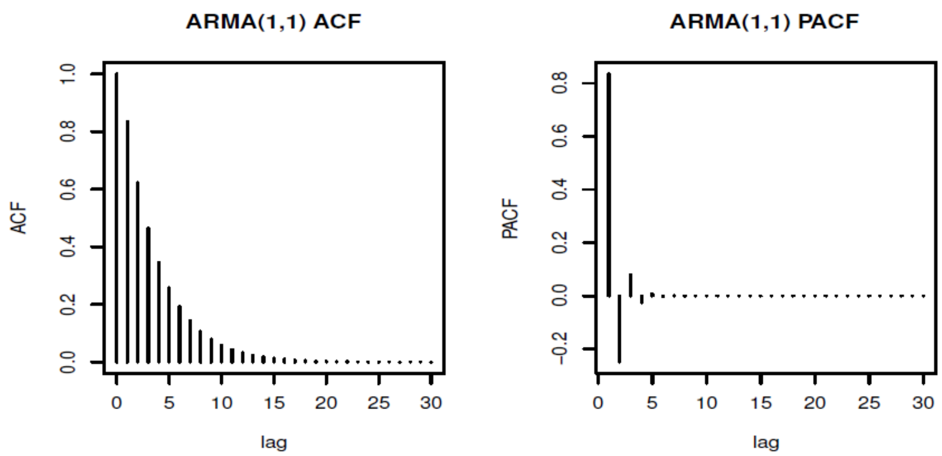
가역성을 만족시키기 위해 $\boldsymbol{\theta(B) = 0}$ 의 근의 절댓값이 1 보다 커야합니다.

4. ACF

지수적으로 감소하거나 싸인함수 형태로 소멸하는 모양입니다.

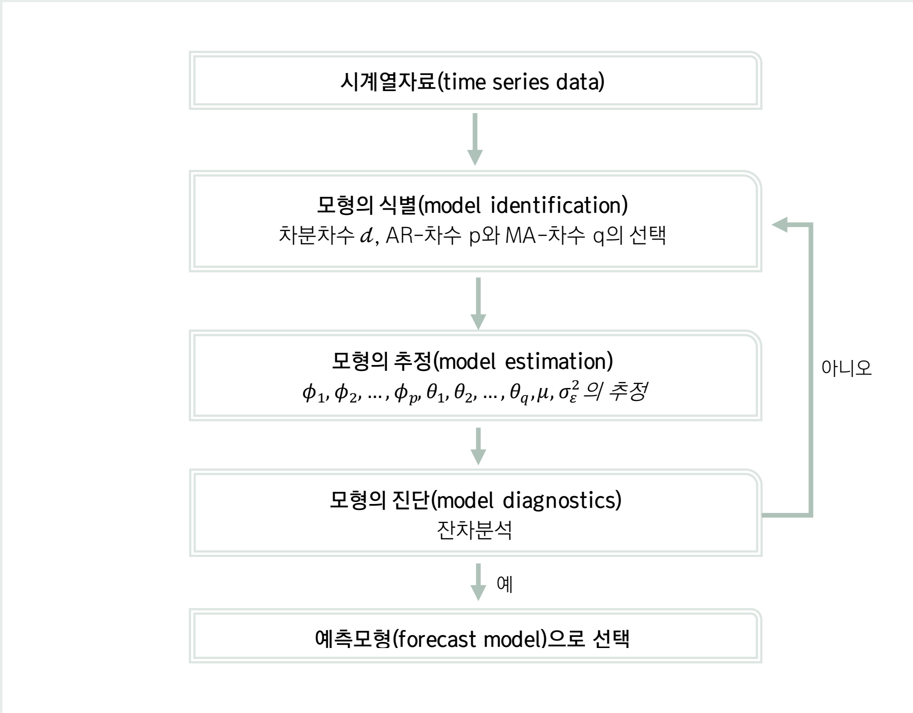
5. PACF

PACF 또한 지수적으로 감소하거나 싸인함수 형태로 소멸하는 모양입니다.



VIII. 모형의 적합절차

주어진 자료를 가장 잘 설명해주는 모형은 어떻게 선택하는 것일까요? 이렇게 여러 모델을 배웠고, 또 배울 예정인데 이 중 적절한 모델을 잘 선택해 주어야 의미있는 것이겠죠? 예측을 위한 모형은 다음과 같은 4단계를 거쳐서 선택됩니다.



1. 모형의 식별

모형의 식별 단계에서는 시계열 그림, 자기상관함수, 부분자기상관함수등을 이용하여 차분의 필요 여부와 모형의 차수(p,q)를 잠정적으로 결정합니다. 이 단계에서 유의할 점은 p와 q를 크게 잡으면 추정해야 할 모수의 개수가 증가하므로 비효율적이라는 점입니다. 예측 모형이 복잡해질 뿐만 아니라 추정의 효율성도 떨어지므로 될 수 있으면 간단한 모형을 선호하며 이러한 원칙을 "간결의 원칙"(principle of parsimony)이라고 합니다. 또한 모형 선택의 기준으로 사용되는 AIC, SBC 등을 모형 식별에 사용하기도 합니다.

AIC = $n\ln\hat{\sigma}^2{}_{\varepsilon} + 2(p+q)$ ⇨ 가장 작은 값을 가진 AIC모형을 선택합니다.

SBC = $n\ln\hat{\sigma}^2{}_{\varepsilon} + (p+q)\ln n$ ⇨ 가장 작은 값을 가진 SBC모형을 선택합니다.

2. 모수 추정

모수의 추정 단계에서는 적률추정법, 조건부 최소제곱추정법, 비조건부 최소제곱추정법, 최대가능도추정법 등을 이용하여 모수들을 추정합니다. 추정해야 할 모수는 $\phi = (\phi_1, \phi_2, \dots, \phi_p)$,$\theta = (\theta_1, \dots, \theta_p), \mu, \sigma^2$ 총 (p+q+2)개 입니다. 이때, 모형의 식별과 모수의 추정 단계를 거쳐서 얻은 특정한 모형을 잠정모형이라고 합니다.

- 적률추정법: 모집단의 적률을 상응하는 표본의 적률로 대체한 후 방정식을 풀어 추정량을 구하는 방법
- 조건부 최소제곱법, 비조건부 최소제곱법: 오차의 제곱합이 가장 작게 되도록 하는 모수들의 추정량을 구하는 방법을 최소제곱추정법이라고 하며, 오차제곱합을 계산하는 과정에서 어떠한 조건을 주는가에 따라 조건부/비조건부 최조제곱추정법이 있습니다.
- 최대가능도추정법: 관측된 시계열의 결합확률밀도함수인 모수의 가능도함수를 최대화시켜 모수를 구하는 방법

3. 모형의 진단

모형의 적합성 진단 단계에서는 잔차의 시계열그림, 잔차의 자기상관함수와 부분자기상관함수, Portmanteau 통계량을 이용한 잔차 분석과 과대적합에 의하여 잠정모형의 적합 정도를 진단합니다. 즉, 모형의 검토과정을 통해서 잠정모형의 타당성이 입증되면 모형이 최종적으로 적합된 것으로 판단하며, 적합된 모형은 예측모형으로 사용합니다.

- 포트맨토검정(Portmanteau test) : 잔차의 표본부분자기상관계수를 이용하여 적합성 검정을 시행합니다. H0: 모형이 적합하다. (잔차들 사이에 유의한 상관관계가 존재하는지)

$$H_0 : \rho_1(e) = \rho_2(e) = \cdots = \rho_k(e) = 0$$

4. 예측

- 최종선정된 모델을 이용하여 예측합니다.
- 새로운 자료가 관측될 때마다 이에 대한 예측값과 비교해 안정성을 조사합니다. 안정성 조사에서는 예측값과 실제 관측값과의 차이가 큰 경우에 모형 또는 모수가 예측기간 동안에 변하지 않았는지를 알아봅니다. 만약 변화가 있었다면, 예측모형 자체를 새로 구하거나 기존의 모형을 갱신하여 사용합니다.
- 새로운 관측값과 이전에 구한 예측값을 이용하여 미래의 예측값을 갱신합니다.

IX. 정리

오늘 배운 내용들이 복잡한 감이 있기에 한번 정리를 해볼께요!!

(1) 정상성과 가역성

	AR(p)	MA(q)	ARMA(p,q)
정상성	조건필요	자체만족	조건필요
가역성	자체만족	조건필요	조건필요

(2) 모형의 ACF와 PACF패턴

	AR(p)	MA(q)	ARMA(p,q)
ACF	지수적으로 감소	q+1차부터 절단	지수적으로 감소
PACF	p+1차부터 절단	지수적으로 감소	지수적으로 감소

[부록]

이번주 부록은,,,, 아무래도 수식도 너무 많고 계산도 너무 많고 통계적 개념들이 바탕이 되어야 이해하고 계산 가능한 것들이 많아서 다 넣기엔 개인차가 있고 또 시간의 제약도 있고! 그리고 우린 할 일이 정말 많잖아요? 그래서 모수의 추정은 어떤 식으로 이루어지는지 궁금하신 분들이 계실까봐 여러 방법이 있지만 느낌만 보시라고 이렇게 부록으로 남겨두었습니다.

단위근(Unit root)

데이터가 정상성을 만족하는지 만족하지 않는지를 검정하는 방법 중 단위근검정이 있습니다. 단위근이란 무엇이고 단위근 검정이란 무엇인지 알아보겠습니다. 오늘 배운 내용까지 알아야 온전히 이해할 수 있는 내용이어서 오늘 부록에서 배워볼께요! 다음과 같은 AR(p)과정을 생각해봅시다.

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

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

위의 모형이 정상성을 만족하기 위해서는 $\phi(B) = 0$의 모든 근의 절댓값이 1 보다 커야 했습니다. 만약 절댓값이 1 보다 크지 않은 근이 존재하는 경우 비정상 확률과정이라고 하며, 근의 크기가 1 인 근을 단위근(unit root)라과 합니다. AR(1)과정에서 $(1 - \phi B) = 0$의 근이 1 이 되는 경우 공분산은 $t\sigma^2$가 되어 시간에 의존하는 비정상 시계열이 됨을 알 수 있습니다. 따라서 다음과 같은 가설을 검정하는 것을 단위근 검정이라고 합니다.

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

단위근검정을 위한 검정통계량으로는 LSE 를 통해 추정된 값이 많이 활용됩니다.

Yule-Walker 방정식 – 적률추정법

AR(p) 모형을 고려하자.

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

양변에 X_{t-k}를 곱하고 기댓값을 취하면

$$E(X_t X_{t-k}) = \phi_1 E(X_{t-1} X_{t-k}) + \phi_2 E(X_{t-2} X_{t-k}) + \cdots + \phi_p E\left(X_{t-p} X_{t-k}\right) + E(Z_t X_{t-k})$$

$$\Rightarrow \gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \cdots + \phi_p \gamma(k-p) + E(Z_t X_{t-k})$$

이다. 그런데 $E(Z_t X_{t-k}) = \begin{cases} \sigma^2, k = 0 \\ 0, \quad k \geq 1 \end{cases}$ 이므로

$k = 0$: $\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma^2$ (식 1)

$k \geq 1$: $\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \cdots + \phi_p \gamma(k-p)$ (식 2)

의 관계를 가진다. 그리고 (식 2)의 양변을 $\gamma(0)$로 나누면

$$\rho(k) = \phi_1 (k-1) + \phi_2 (k-2) + \cdots + \phi_p \rho(k-p), k \geq 1$$

와 같이 나타낼 수 있다. 이제 이 방정식을 행렬로 표현하면

$$\begin{bmatrix}\rho_1\\ \rho_2\\ \vdots\\ \rho_k\end{bmatrix}=\begin{bmatrix}1 & \rho_1 & \rho_2 & \cdots\rho_{p-1}\\ \rho_1 & \rho_2 & \rho_1 & \cdots\rho_{p-2}\\ \vdots & \vdots & \ddots & \vdots\ \ \vdots\\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots\ \ 1\end{bmatrix}\begin{bmatrix}\phi_1\\ \phi_2\\ \vdots\\ \phi_p\end{bmatrix}$$

와 같이 나타낼 수 있다. 여기서 $\rho_k = \rho(k)$를 나타낸다. 이것을 AR(p) 모형의 Yule-Walker 방정식이라 부른다. 특히 $\mathbb{P} = Corr(Z_{t-1}, Z_{t-k}) = \begin{bmatrix}1 & \rho_1 & \rho_2 & \cdots\rho_{p-1}\\ \rho_1 & \rho_2 & \rho_1 & \cdots\rho_{p-2}\\ \vdots & \vdots & \ddots & \vdots\ \ \vdots\\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots\ \ 1\end{bmatrix}$는 AR(p) 모형에서 X_{t-1}과 X_{t-k}의 상관구조를 설명한다.

행렬 $\mathbf{P}$는 양정치 행렬(positive definite matrix)이기 때문에 역행렬이 존재한다. 따라서 Yule-Walker 방정식으로부터

$$\phi = \mathbb{P}^{-1}\rho$$

를 얻는다. 그리고 (식 1)은

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \cdots + \gamma_p\gamma(p) + \sigma^2 = \phi^T\begin{bmatrix}\gamma(1)\\ \gamma(2)\\ \vdots\\ \gamma(p)\end{bmatrix} + \sigma^2 = \boldsymbol{\phi}^T\boldsymbol{\gamma} + \sigma^2$$

이므로 $\sigma^2 = \gamma(0) - \boldsymbol{\phi}^T\boldsymbol{\gamma} = \gamma(0) - \rho^T\mathbb{P}^{-1}\boldsymbol{\gamma}$

로서 표현된다.

결국 AR(p) 모형에서는 $\gamma(0), \gamma(1), \ldots\ \gamma(p)$를 알면 계수 $\phi_1, \phi_2, \ldots, \phi_p$와 σ^2이 결정된다.