WEEK2

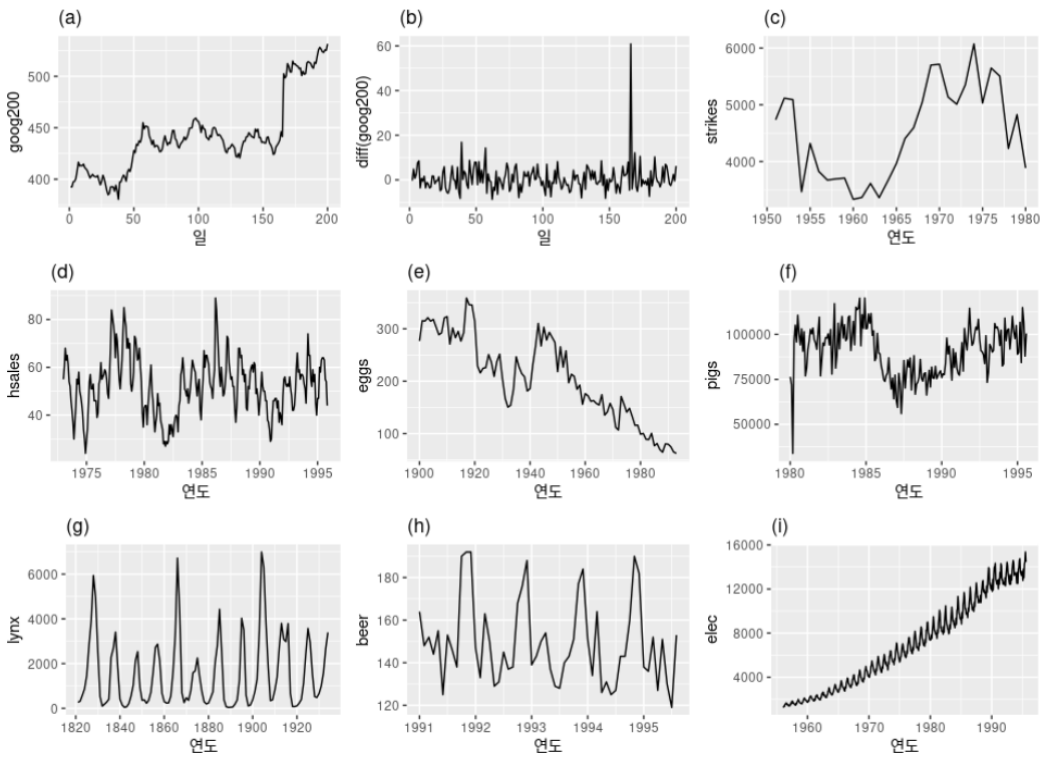
- 정상적 ARMA 모형: AR모형, MA모형, ARMA모형
- 비정상적 모형: ARIMA 모형, 계절성 ARIMA 모형
- 다변량 시계열: VAR

1. 자기상관함수
- 시계열 자료의 큰 특징 = 시간의 흐름에 따라 독립적이지 않다는 것 ⇒ 자기상관관계를 가짐
- 따라서, 시간에 따른 상관 정도를 나타내야 함 ⇒ ‘자기상관함수(AutoCovariance Function)’

- 정상성이란?
  1) 일정한 평균
  - 모든 시점에 대해 평균이 일정해야 함: if not, 차분을 통해 정상화 필요
    ⇒ 차분 = (현 시점의 자료 값) - (전 시점의 자료 값)
  2) 일정한 분산
  - 모든 시점에 대해 분산이 동일해야 함: if not, 변환을 통해 정상화 필요
    ⇒ 변환: 자료 값에 지수 혹은 로그를 취함 → 시간에 따라 변하는 분산 크기 안정화
  3) 시차에만 의존하는 공분산
  - 시차 의존, 시점 의존 X
  - t는 시점, s는 시차라고 했을 때, ‘t시점과 t+s시점의 공분산’과 ‘t시점과 t-s시점의 공분산’은 서로 같음
  - 다만, 시차에 따라 공분산 값이 다를 수는 있음

  + 추세나 계절성 등의 시계열 자료는 정상성을 띠지 않는다고 보아야 함
    But, 주기성 행동을 가지지만(추세나 계절성은 없는)시계열은 정상성을 나타내는 시계열임

  (ex) 정상성을 띠는 시계열 찾기



- 이 중, (b),(g)만이 정상성을 나타내는 시계열 후보임. 특히, (g)의 경우는 일정한 주기를 띤다고 착각할 수 있지만, 해당 주기는 불규칙적임(→ (g): 캐나다 북서부의 맥킨지 강 지역에서 연간 포획된 스라소니의 전체 수). 먹이를 구하기 힘들만큼 살쾡이 개체 수가 너무 많이 늘어나 번식을 멈춰서, 개체수가 작은 숫자로 줄어들고, 그 다음 먹이를 구할 수 있게 되어 개체수가 다시 늘어나는 식임. 장기적으로 볼 때, 이러한 주기의 시작이나 끝은 예측할 수 없으므로 정상성을 띤다고 보아야 함

- 자기 공분산: 시계열의 시간에 따른 연관 패턴을 나타냄
  - 시차 k의 자기공분산

$\gamma(k) = Cov[Z_t, Z_{t-k}] = E[(Z_t - \mu_Z)(Z_{t-k} - \mu_Z)]$ (약 정상성 가정)
$\quad = E[Z_t\ Z_{t-k}]$ ($\mu_Z = 0$ 가정)
$\gamma(0) = Var[Z_t]$
$\gamma(k) = \gamma(-k)$
$\gamma(k)$ 를 k의 함수 ($k = 0, 1, 2, \ldots$)로 볼 때, 자기공분산 함수 (autocovariance function)라 함

- 자기상관 함수(ACF)
  - 일반적으로 ACF 값은 시차와 크기와 반비례

- 비교적 단순한 형태의 정상적 시계열 모형을 주로 다루며, ACF로 모형을 식별할 수 있음

시차 k의 자기상관계수:

$$\rho(k) = Corr[Z_t, Z_{t-k}] = \frac{Cov[Z_t, Z_{t-k}]}{Var[Z_t]} = \frac{\gamma(k)}{\gamma(0)}$$

✓ $\rho(0) = 1$

✓ $\rho(k) = \rho(-k)$

공분산을 분산으로 나눈 형태

2. **편자기상관함수**

   - 정상적 시계열의 형태를 식별하기 위해 ACF외에 PACF의 정보를 활용함
   - PACF는 시차 k 사이의 값들은 conditional이라고 가정. 따라서 조건부 상관계수를 고려함
   - P(k)는 다음의 회귀모형에서의 kk번째 회귀계수와 같다고 알려져 있음
   - ACF와의 차이점은, 다른 시점의 확률변수 영향력은 통제하고 '상관관계만' 보여준다는 것
   (ex) 교회수와 범죄발생률이 비례한다고 가정 → 교회가 많다는 것은 인구수 또한 많다는 것 → '인구' 요소를 제외하고, 오로지 교회와 범죄발생률만을 놓고 상관계수를 분석

$$- Z_t = \phi_{k1} Z_{t-1} + \phi_{k2} Z_{t-2} + \cdots + \phi_{kk} Z_{t-k} + b_t$$

- 시계열 표현방식
  - 자기회귀(AR) 방식: 시점 t의 값을 과거시점의 값들을 이용한 회귀식으로 표현. 과거시점의 값들과 백색잡음(a_t)이 포함된 식임
  - 이동평균(MA) 방식: 시점 t의 값을 현재와 과거시점의 백색잡음만으로 표현
- 후향연산자 표현

$$- Z_{t-1} = BZ_t$$
$$- Z_{t-2} = BZ_{t-1} = B^2 Z_t$$
$$- Z_{t-k} = B^k Z_t, k = 1,2, \ldots$$

-B가 곱해져 있는 만큼, 데이터를 한 시점 뒤로 옮기는 효과를 나타냄

AR과정의 표현

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \cdots + a_t \Rightarrow Z_t = (\pi_1 B + \pi_2 B^2 + \cdots)Z_t + a_t$$
$$\Rightarrow (1 - \pi_1 B - \pi_2 B^2 - \cdots)Z_t = a_t$$
$$\Rightarrow Z_t = (1 - \pi_1 B - \pi_2 B^2 - \ldots)^{-1} a_t$$

MA과정의 표현

$$Z_t = a_t - \psi_1 \underbrace{a_{t-1}}_{Ba_t} - \psi_2 \underbrace{a_{t-2}}_{B^2 a_t} - \cdots \Rightarrow Z_t = (1 - \psi_1 B - \psi_2 B^2 - \cdots)a_t$$
$$\Rightarrow (1 - \psi_1 B - \psi_2 B^2 - \ldots)^{-1} Z_t = a_t$$

(예 2-4) 다음의 시계열을 AR표현방식과 MA표현방식으로 나타내라.

$$Z_t = a_t - 0.5a_{t-1} + 0.3a_{t-2}$$

(풀이)

- 이미 MA방식으로 표현되어있다. 즉,

$$\psi_1 = 0.5, \psi_2 = -0.3, \psi_k = 0, k = 3,4, \ldots$$

- AR방식으로 표현하기 위해 후방연산자를 사용하면

$$Z_t = (1 - 0.5B + 0.3B^2 - \cdots)a_t \Rightarrow (1 - 0.5B + 0.3B^2 - \ldots)^{-1} Z_t = a_t$$
$$\Rightarrow (1 + 0.5B - 0.05B^2 - 0.175B^3 - \cdots)Z_t = a_t$$
$$\Rightarrow Z_t = -0.5Z_{t-1} + 0.05Z_{t-2} + 0.175Z_{t-3} + \cdots + a_t$$

$$\frac{1}{1 - 0.5B + 0.3B^2 + \cdots}$$

따라서

$$\pi_1 = -0.5, \pi_2 = 0.05, \pi_3 = 0.175, \cdots$$

3. **AR 모형과 MA 모형**

- AR모형: 변수의 과거 값의 선형 결합을 통해서 예측을 수행. 자기 회귀(Auto-Regressive)라는 단어 속에 자기 자신에 대한 변수의 회귀라는 의미가 담겨있음. 따라서 차수 p의 자기 회귀 모모형은 다음과 같이 쓸 수 있음

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

e_t는 백색잡음(~N(0, $sigma^2$)). p 자기 회귀 모형인 AR(p) 모델이라고 부름

- AR(1) / AR(2) 모형
  * AR(1) 모형 :

$$Z_t = \phi_1 Z_{t-1} + a_t$$

-정상성을 갖기 위한 조건: -1 < ϕ < 1
-Z_t는 일반적으로 평균을 0으로 가정함

*AR(2) 모형 :

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$$

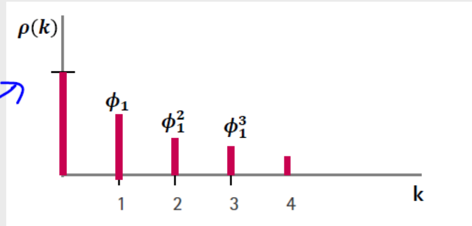
-정상성을 갖기 위한 조건:

$$-1 < \phi_2 < 1, \phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1$$

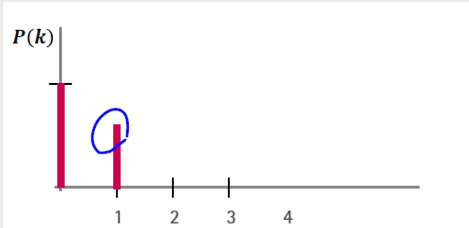
- AR 모형의 ACF/PACF

- AR(1) 모형의 ACF
  - $Z_t = \phi_1 Z_{t-1} + a_t$ $-1<\phi_1<1$
  - $Var[Z_t] = \gamma(0) = \frac{\sigma_a^2}{1-\phi_1^2}$
  - $\gamma(k) = \phi_1^k \gamma(0),\ k = 0, 1, 2, \ldots$
  - $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k,\ k = 0, 1, 2, \ldots$: 지수적으로 감소하는 패턴

→ 자기상관계수

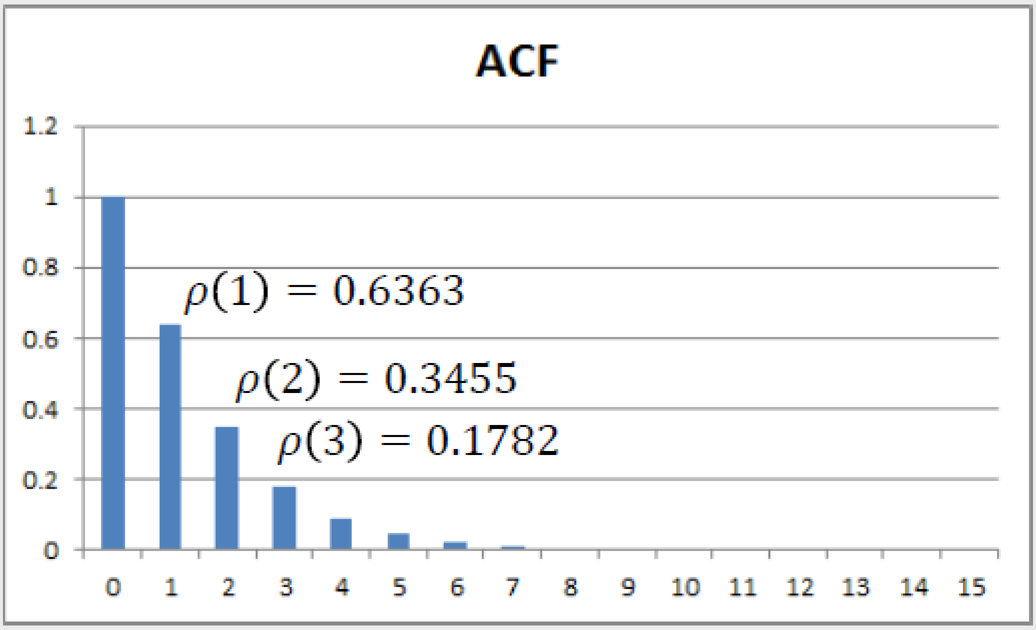


- AR(1) 모형의 PACF
  - $P(1) = \rho(1) = \phi_1$
  - $P(k) = 0,\ k = 2, 3, \cdots$: 시차 1 이후 절단 패턴



- AR(1) 모형의 경우, ACF는 지수적으로 감소. $|\phi| < 1$ 이므로, ϕ의 exponential은 값이 계속 줄어드는 것임

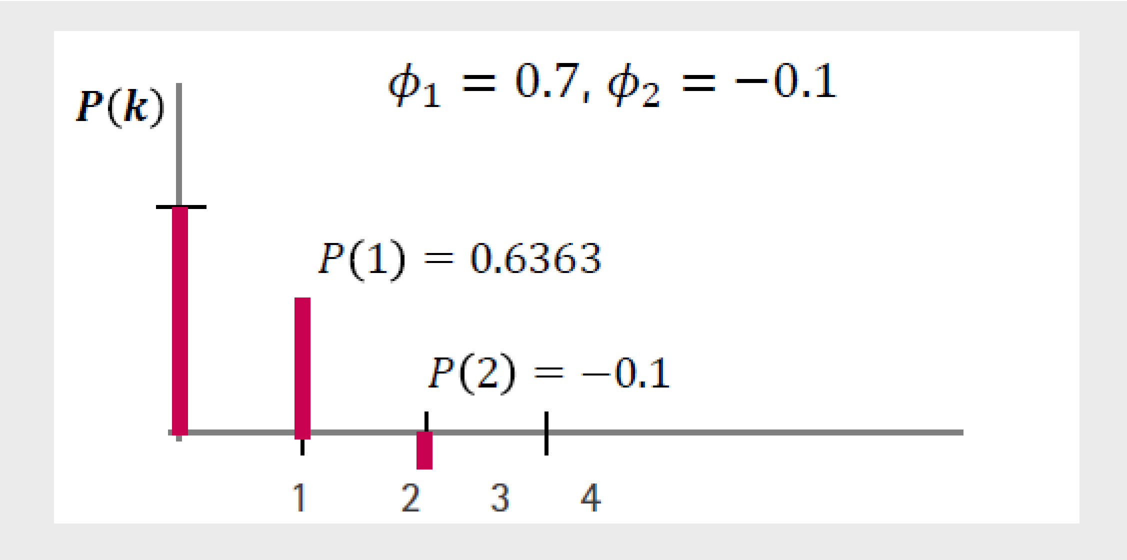
- PACF의 경우는, 1일 때의 값만 존재하고, 그 이후로는 P(k)=0으로 계산됨(절단 패턴)



$$\phi_1 = 0.7,\ \phi_2 = -0.1$$

- AR(2) 모형의 경우, 역시 ACF는 지수적으로 감소하는 패턴을 띔

- PACF는 P(2) 이후로 절단 패턴을 띔



- ACF/PACF의 기능 → 만약, ACF가 지수적으로 감소하고, PACF는 k=2까지만 값을 갖고 3 이후로는 0의 값을 갖는다면 우리가 가진 데이터는 AR(2)를 따르는 데이터라는 것을 유추할 수 있음. PACF에서 절단되는 지점을 통해 AR 모형의 차수 p를 결정할 수 있고, ACF의 형태를 통해 AR 모형인지 아닌지를 판단할 수 있음

- MA 모형: 회귀식에 변수의 과거 값을 이용하는 대신에, 과거 예측 오차를 이용함

$$y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

e_t는 백색잡음

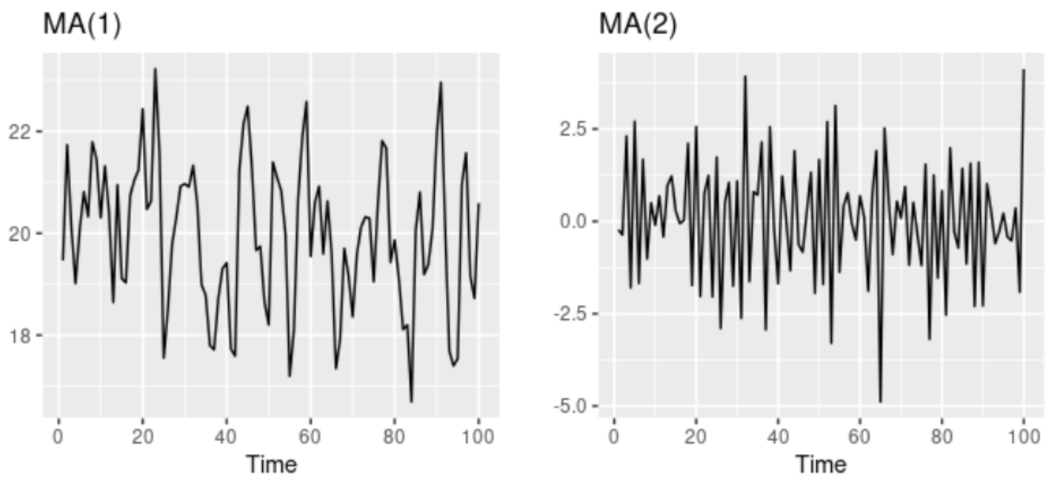
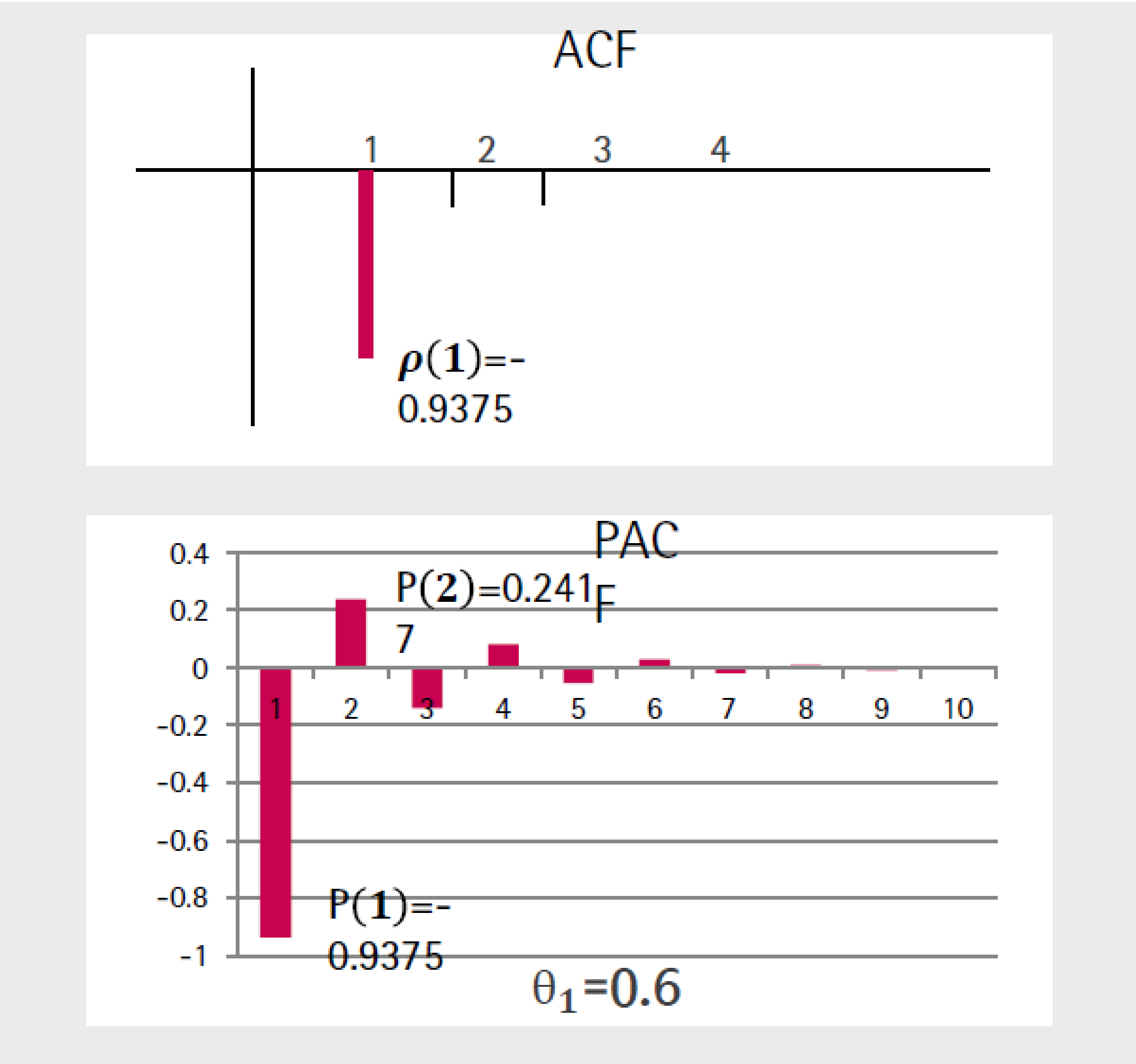


Figure 8.6: 매개변수를 다르게 설정한 이동 평균 모델로부터 얻은 데이터의 두 가지 예. 왼쪽: $y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$인 MA(1). 오른쪽: $y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}$인 MA(2). 두 가지 경우 모두, ε_t은 평균이 0이고 분산이 1인 정규 분포를 따르는 백색잡음입니다.

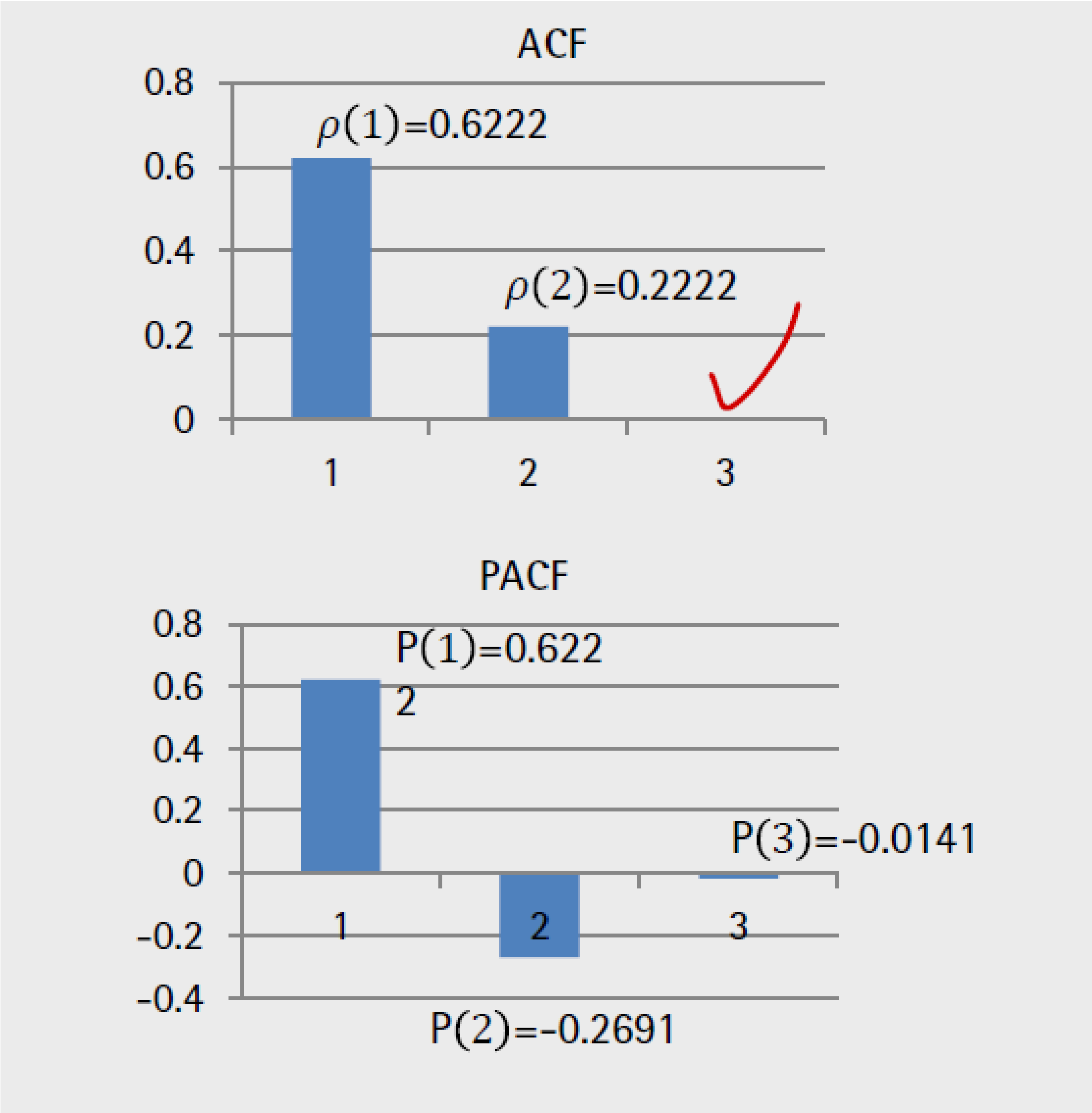
- MA 모형의 가역성 조건

$$\text{MA(1) 모델의 경우: } -1 < \theta_1 < 1.$$
$$\text{MA(2) 모델의 경우: } -1 < \theta_2 < 1,\ \ \theta_2 + \theta_1 > -1,\ \ \theta_1 - \theta_2 < 1$$

- MA 모형의 ACF와 PACF



- MA(1) 모형의 경우, ACF는 K=1 이후에서 절단 패턴을 띔
- PACF의 경우는 지수적으로 감소하는 패턴



- MA(2) 모형에서는 ACF는 2 이후에서 절단 패턴, PACF는 지수적으로 감소하는 패턴을 띰

- AR 모형과 MA 모형의 관계
  - 정상성을 나타내는 어떤 AR(p) 모형은 MA(∞) 모형으로 쓸 수 있음.
  예를 들어, AR(1) 모형의 경우 다음과 같은 식으로 변형할 수 있음
  (정상성 조건에 의해, k가 커질수록 ϕ^k의 값은 작아질 것임) ⇒ MA(∞) 모형이 유도됨

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \varepsilon_t \\
&= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&\text{etc.}
\end{aligned}
$$

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots$$

뿐만 아니라, 어떤 가역적인 MA(q) 모형은 AR(∞) 모형으로 쓸 수 있음

4. **ARMA 모형**
   - ARMA(1,1) 모형: AR(1) 모형과 MA(1) 모형의 복합 형태
   - 시점 t의 값은 시점 t-1의 값, 시점 t와 t-1의 오차항(백색잡음)으로 생성됨
   - 정상성 조건: -1 < ϕ < 1
   - 가역성 조건: -1 < $theta_1$ < 1

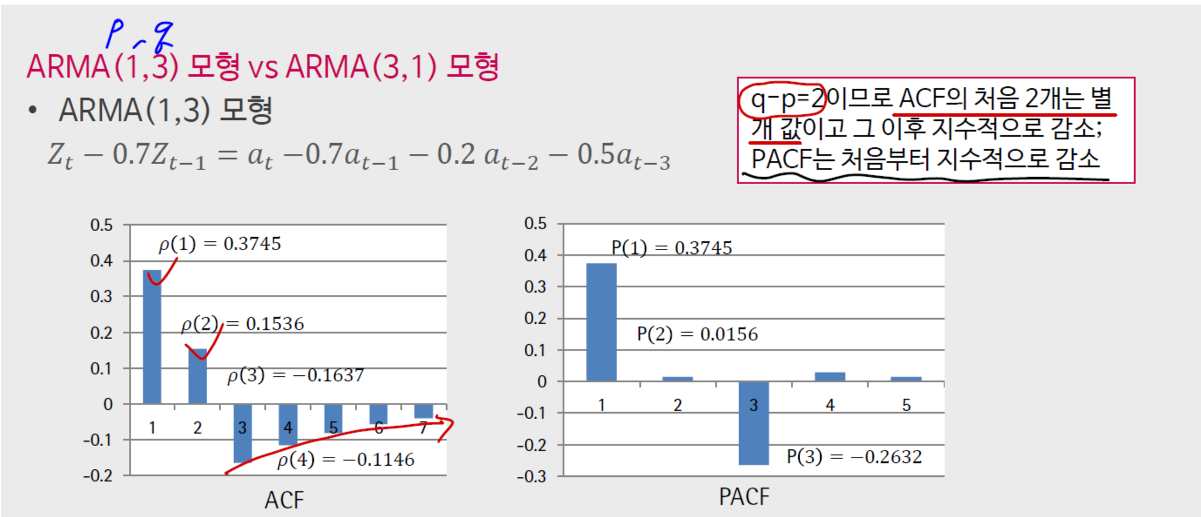
- ARMA(p,q) 모형의 특징

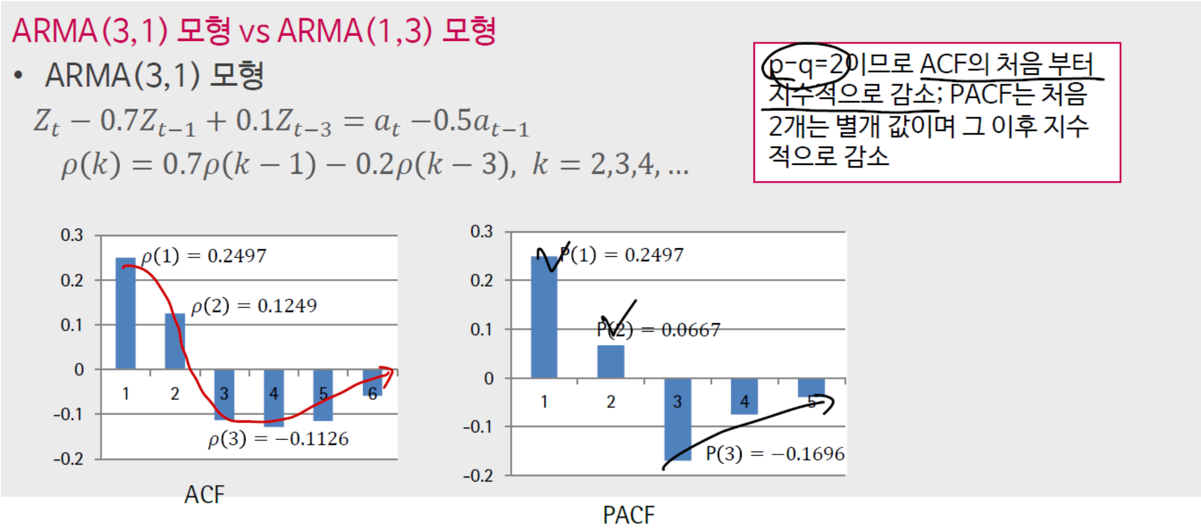
  (1) p≥q+1 일 때, ACF는 AR(p) 모형과 유사하게 0으로 다가감

  (2) p≤q 일 때, q-p개의 값은 특정 값을 가지고 이후로는 0으로 다가감

  (3) p≥q+1 일 때, PACF는 p-q개는 특정 값을 가지고 이후로는 MA(q) 모형과 유사하게 0으로

  (4) p≤q 일 때, 처음부터 MA(q) 모형과 유사하게 0으로 다가감





[ARMA 모형의 표현 요약]

시계열 데이터 X_t가 정상성을 만족하고 $ARMA(p,q)$ 모형을 따른다고 한다면 아래의 관계식을 만족한다.

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t - \sum_{i=1}^{q} \theta_i Z_{t-i} \tag{1.1}$$

여기서 Z_t는 정규분포를 따르는 백색 잡음이다. 즉, Z_t i.i.d. $\sim N(0, \sigma^2)$ 이다.
이때 B를 Back Shift 연산자라고 하자. 즉, $BX_t = X_{t-1}$이고 $B^p X_t = X_{t-p}$이다. 이 연산자를 이용하면 식 (1.1)을 아래와 같이 표현할 수 있다.

$$(1 - \phi_1 B - \cdots \phi_p B^p)X_t = c + (1 - \theta_1 B - \cdots - \theta_q B^q)Z_t \qquad \text{(1-1)}$$

- ARMA 모형 정리
  - 시계열 데이터 설명을 위해 AR/MA 중 하나의 방식으로 표현하는 것보다 모수를 절약하는 효과를 볼 수 있음
  - 차수를 정하기 위해서는, AR은 PACF 함수에서 특정 기준선을 훌쩍 뛰어넘는 막대의 개수, MA는 ACF 함수에서 특정 기준선을 훌쩍 뛰어넘는 막대의 개수에 따라 결정할 수 있음

  (참고)

ARMA 모형의 ACF는 AR과 MA의 특징이 모두 나타난다.

ACF는 지수적으로 감소하는 모습을 보이고 (AR의 특성), PACF에는 음의 상관성도 나타난다. (MA의 특성) 어떤 시계열의 ACF가 이런 모습을 보이고 있으면, 그 시계열은 ARMA 모형에 가까울 가능성이 있다.