시계열 분석 기법과 응용

Week 3. ARMA 모형의 식별과 예측
3-2. ARMA 모형의 추정

전치혁 교수
(포항공과대학교 산업경영공학과)

ARMA 모형의 식별

- 모형의 식별, 추정 및 검증 과정
  - 단계 1: 시계열 데이터를 그래프로 나타내어 시각적으로 관찰하고 정상성 여부를 검토한다. 비정상적인 경우에는 추세제거, 계절성제거, 분산안정화 등을 통하여 정상적 시계열로 변환한다 (이 부분은 비정상적 시계열에서 설명).
  - 단계 2. 시계열 데이터에 대한 표본 ACF 및 표본 PACF를 산출하고 정상성 여부를 확인한다. 비정상적인 경우 단계 1을 반복한다.
  - 단계 3. (모형의 식별) 표본 ACF 및 표본 PACF를 다양한 ARMA모형들의 이론적 ACF 및 PACF와 비교하여 ARMA모형의 차수인 p와 q를 구한다.
  - 단계 4. (모형의 추정) 단계 3에서 얻은 모형에 대한 계수들을 추정하고 잔차(residual)를 구한다.
  - 단계 5. (모형의 검증) 잔차가 백색잡음을 따르는지 검정한다. 잔차가 백색잡음을 따르면 단계 3의 모형이 제대로 식별되었으며, 아니면 다른 차수 p와 q를 구한 후 과정을 반복한다.

시계열 모형 추정방법

- 최소자승법 (least squares method): AR모형의 경우 가능
- 비선형 최소자승법 (nonlinear least squares method): ARMA모형에 적용
- 최우 추정법 (maximum likelihood estimation)
  - 오차항이 서로 독립인 정규분포를 따르므로 우도함수 (likelihood function)을 유도하여 이를 최대로 하는 모형 계수들을 추정. ARMA 모형의 경우 정확한 우도함수 도출이 어렵고 초기치에 대한 가정이 필요.
  - 정확한 우도함수
  - 조건있는 우도함수: 임의로 초기치 가정하여 사용
  - 조건없는 우도함수: 과거의 초기치를 후방예측 (backcasting)하여 사용

최우추정법

- 정확한 우도함수

  - 평균 μ, 분산 σ^2의 정규분포를 따르는 확률변수 X에 대한 관측치 $(x_1, \dots, x_n)$가 있다하자.
  - 이 때 관측치는 서로 독립이므로 n개 관측치의 결합확률분포는 정규확률밀도함수의 곱으로 표현됨

$$f(x_1, \dots, x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

  - 관측 데이터를 대입하면 모수 μ와 σ^2의 함수로 간주할 수 있으며 이를 우도함수라 함.

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

  - 이 우도함수를 최대로 하는 모수 μ와 σ^2를 추정하는 방법을 최우추정법이라 함.
  - 로그함수를 취한 로그우도함수를 최대로 하는 것이 편함

$$logL(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

최우추정법

- ARMA모형의 경우 관측치 $(Z_1, Z_2, \dots, Z_n)$들은 서로 독립이 아니므로 이에 대한 우도함수의 구성이 어렵다.
- 대신 백색잡음 $(a_1, a_2, \dots, a_n)$이 서로 독립임을 활용하여 이에 대한 우도함수를 구성한다.

$$L(\theta; a_1, \dots, a_n) = (2\pi\sigma_a^2)^{-n/2} exp\left(-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right)$$

- (예 - AR(1) 모형) $Z_t = \phi_1 Z_{t-1} + a_t$
  - $a_t = Z_t - \phi_1 Z_{t-1}$
  - $a_1 = Z_1 - \phi_1 Z_0$ (즉, a_1을 구하기 위해 초기치 Z_0 필요)
- 초기치 처리 방법
  - 적절한 가정으로 초기치 결정 ⇒ 조건있는 우도함수 (conditional likelihood function)
  - 초기치를 예측하여 사용 ⇒ 조건없는 우도함수 (unconditional likelihood function)

조건있는 우도함수

- $(a_1, a_2, \dots, a_n)$의 로그 우도함수

$$logL(\theta|a_1,\dots,a_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}log\sigma_a^2 - \frac{\sum_{t=1}^{n} a_t^2}{2\sigma_a^2}$$

- ARMA (p,q)모형에서 $(a_1, a_2, \dots, a_n)$ 산출

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p}$$
$$a_1 = \theta_1 a_0 + \cdots + \theta_q a_{1-q} + Z_1 - \phi_1 Z_0 - \cdots - \phi_p Z_{1-p}$$

다음의 초기치 필요

$$a_* = (a_{1-q}, \dots, a_{-1}, a_0) = (0, \dots, 0,0)$$
$$z_* = (z_{1-p}, \dots, z_{-1}, z_0) = (z, \dots, z, z)$$

초기 조건하에서 로그 우도함수

$$logL_*(\theta|a_1,\dots,a_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}log\sigma_a^2 - \frac{S_*(\theta)}{2\sigma_a^2}$$

조건제곱합

$S_*(\theta) = \sum_{t=1}^{n} a^2{}_t\,(\theta|a_*, z_*, \mathrm{z}_1, \dots, \mathrm{z_n}) \Rightarrow$ 최소화 $\Rightarrow \hat{\theta} \Rightarrow \hat{\sigma}_a^2 = \frac{S_*(\hat{\theta})}{df}$ $(df = n - 2p - q - 1)$

조건있는 우도함수

(예- AR(1) 모형) 다음 10개 시계열 데이터가 AR(1)모형을 따른다.

-0.17, 0.24, -0.51, 1.10, 0.08, 0.32, -0.44, -1.16, -0.58, -0.28

$\phi_1 = 0.2$와 $\phi_1 = 0.4$ 일때 조건제곱합을 각각 구하라.

(풀이)

$\phi_1 = 0.2$: $a_t = Z_t - 0.2Z_{t-1}$

$\phi_1 = 0.4$: $a_t = Z_t - 0.4Z_{t-1}$

$z_0 = 0$ 가정

$$S_*(\phi_1 = 0.2) = 3.5234$$

$$S_*(\phi_1 = 0.4) = 3.7106$$

Z_t	$\phi_1 = 0.2$		$\phi_1 = 0.4$	
	a_t	a_t^2	a_t	a_t^2
-0.17	-0.17	0.0289	-0.17	0.0289
0.24	0.274	0.0751	0.308	0.0949
-0.51	-0.558	0.3114	-0.606	0.3672
1.1	1.202	1.4448	1.304	1.7004
0.08	-0.14	0.0196	-0.36	0.1296
0.32	0.304	0.0924	0.288	0.0829
-0.44	-0.504	0.2540	-0.568	0.3226
-1.16	-1.072	1.1492	-0.984	0.9683
-0.58	-0.348	0.1211	-0.116	0.0135
-0.28	-0.164	0.0269	-0.048	0.0023

조건없는 우도함수

- 조건없는 우도함수

$$logL(\theta|z_1, \dots, z_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}log\sigma_a^2 - \frac{S(\theta)}{2\sigma_a^2}$$

- 조건없는 제곱합

$S(\theta) = \sum_{t=-M}^{n} E^2[a_t|\theta; z_1, \dots, z_n]$ (M: 과거치 예측수)
백색잡음의 기대치로 오차항 예측 - 이 때 과거 관측치 예측 필요

- 과거값 예측에 후방예측 (backcasting) 사용 - 시간축을 반대방향으로 생각

$$Z_t = \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$
$$\Rightarrow Z_t = \phi_1 Z_{t+1} - \cdots - \phi_p Z_{t+p} + b_t - \theta_1 b_{t+1} - \cdots - \theta_q b_{t+q}$$
⇒ 미래값 (관측치)으로 과거값 예측

- 예 –AR(1)모형 $Z_t = \phi_1 Z_{t-1} + a_t \Leftrightarrow Z_t = \phi_1 Z_{t+1} + b_t$

$$E[a_0] = E[Z_0] - \phi_1 E[Z_{-1}]$$
$$E[Z_0] = \phi_1 E[Z_1] + E[b_0] = \phi_1 z_1$$
$$E[Z_{-1}] = \phi_1 E[Z_0] + E[b_{-1}] = \phi_1 \phi_1 z_1$$

조건없는 우도함수

(예- AR(1))다음 10개 시계열 데이터가 AR(1)모형을 따른다.

-0.17, 0.24, -0.51, 1.10, 0.08, 0.32, -0.44, -1.16, -0.58, -0.28

$\phi_1 = 0.2$와 $\phi_1 = 0.4$ 일때 조건없는 제곱합을 각각 구하라.

$S(\theta) = \sum_{t=-M}^{n} E^2[a_t|\theta; z_1, \ldots, z_n]$

- $Z_t = \phi_1 Z_{t-1} + a_t \Leftrightarrow Z_t = \phi_1 Z_{t+1} + b_t$
- $E[a_t] = Z_t - \phi_1 Z_{t-1},\ t = 2,3, \ldots$
- $E[a_1] = E[Z_1] - \phi_1 E[Z_0] = (1 - \phi_1^2) z_1$
- $E[a_{-j}] = \phi_1^{j+1}(1 - \phi_1^2) z_1$, j= 0,1,2, …
- $S(\phi_1 = 0.2) = 3.5236$
- $S(\phi_1 = 0.4) = 3.7060$

Z_t		$\phi_1 = 0.2$		$\phi_1 = 0.4$	
		a_t	a_t^2	a_t	a_t^2
-3	-	-0.0040	0.0000	-0.00366	0.0000
-2	-	-0.0100	0.0001	-0.00914	0.0001
-1	-	-0.0250	0.0006	-0.02285	0.0005
0	-	-0.0626	0.0039	-0.05712	0.0033
1	-0.17	-0.1564	0.0245	-0.1428	0.0204
2	0.24	0.2740	0.0751	0.308	0.0949
3	-0.51	-0.5580	0.3114	-0.606	0.3672
4	1.1	1.2020	1.4448	1.304	1.7004
5	0.08	-0.1400	0.0196	-0.36	0.1296
6	0.32	0.3040	0.0924	0.288	0.0829
7	-0.44	-0.5040	0.2540	-0.568	0.3226
8	-1.16	-1.0720	1.1492	-0.984	0.9683
9	-0.58	-0.3480	0.1211	-0.116	0.0135
10	-0.28	-0.1640	0.0269	-0.048	0.0023

ARMA 모형의 추정

- 예 (나일강 유량) 본 시계열이 AR(2)모형을 따른다고 식별되므로 다음의 모형 추정 $Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$ 추정 결과는 다음과 같다.

	추정치	표준오차	t-값	p-값
상수 (c)	357.5	14.55	24.56	0.000
AR1 (ϕ_1)	0.4039	0.1000	4.04	0.000
AR2 (ϕ_2)	0.2064	0.1006	2.05	0.043

#1

- 두 계수 ϕ_1, ϕ_2의 추정치는 유의수준 5%에서 유의
- 추정식은 다음과 같다. $\hat{Z}_t = 357.5 + 0.4039 Z_{t-1} + 0.2064 Z_{t-2}$

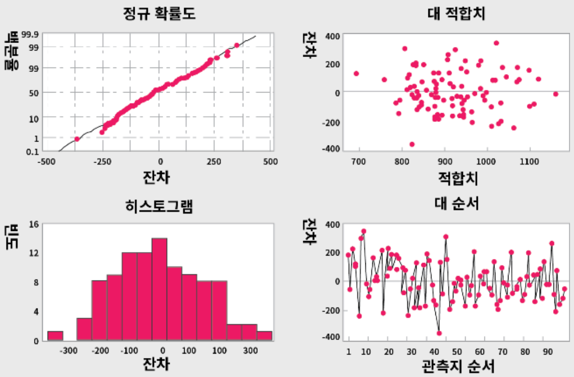
$\widehat{\sigma_a^2} = MSE = 20943$

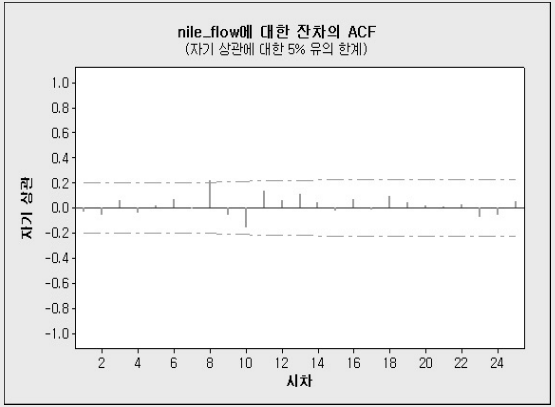
ARMA 모형의 검증

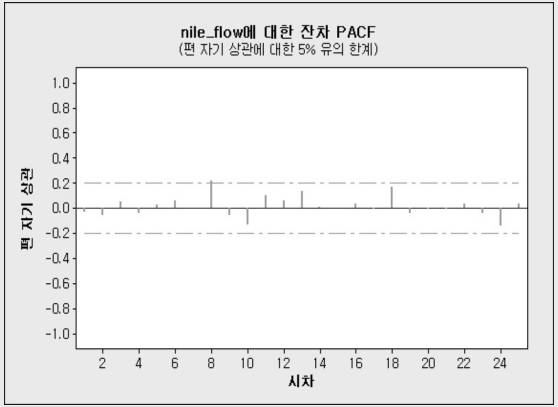
- 모형의 검증
  - 모형의 오차항이 평균 0, 분산 σ_a^2의 정규분포를 따르는 백색잡음이라 가정하고 있기 때문에 이에 대한 검증이 필요
  - 잔차에 대하여 다음을 확인
  - 정규성: 정규확률도표 활용
  - 등분산성 및 패턴 유무: 잔차 산점도 활용
  - 랜덤성:
    - ✓ ACF/PACF로 시차별 상관계수가 모두 0인지 확인
    - ✓ 포트만토 검정 (portmanteau test): 시차별 자기상관계수가 모두 0인지 검정 (Ljung-Box 검정 등 사용)

ARMA 모형의 식별과 추정

- 예 (나일강 유량)







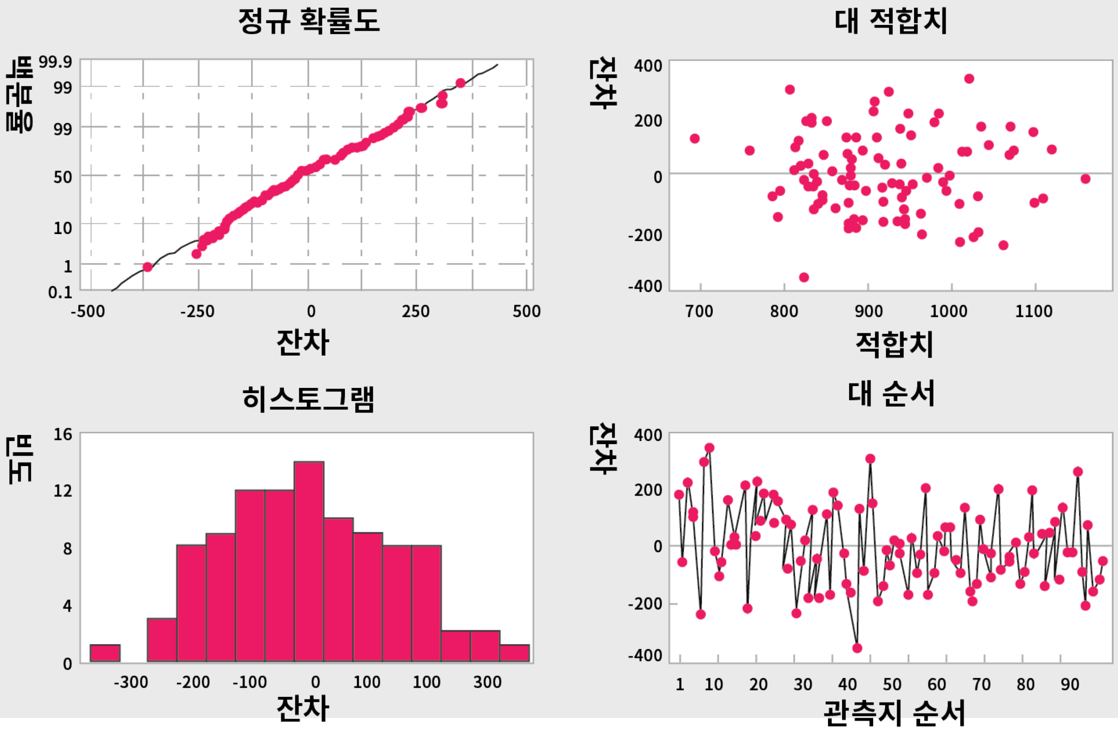
#1

nile_flow의 잔차 그림

ARMA 모형의 식별과 추정

- 예 (나일강 유량)
  - 정규확률도를 볼 때 잔차가 정규분포를 따름
  - 잔차 산점도로 부터 등분산성, 패턴없음이 관찰

nile_flow의 잔차 그림

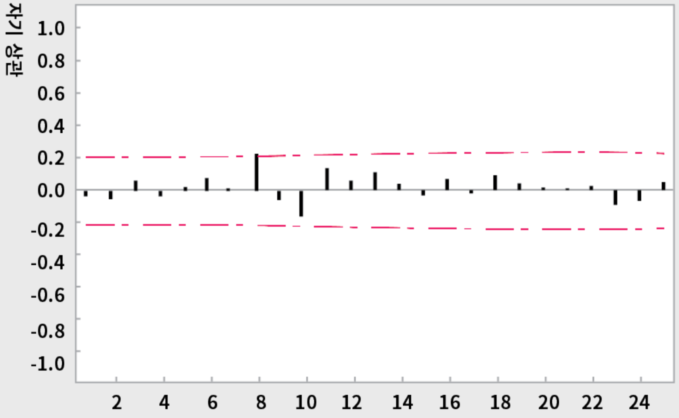


#1

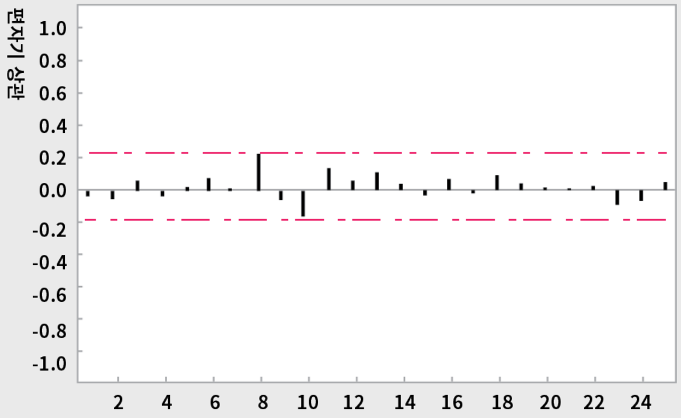
ARMA 모형의 식별과 추정

- 예 (나일강 유량)
  - 잔차에 대한 ACF/PACF로 부터 백색잡음 확인

nile_flow에 대한 잔차의 ACF
(자기 상관에 대한 5% 유의 한계)



nile_flow에 대한 잔차의 PACF
(편 자기 상관에 대한 5% 유의 한계)



#1

ARMA 모형의 식별과 추정

- 포트만토 (Ljung-Box) 검정

시차	12	24	36	48
카이제곱	11.9	16.6	25.0	37.3
자유도	9	21	33	45
p-값	0.222	0.735	0.841	0.787

- 포트만토 검정에서 시차별 자기상관이 없음의 귀무가설을 채택
- 즉, 잔차에 시차별 자기상관이 없다고 판정

Reference

#1. Qlik DataMarket https://datamarket.com/data/set/22w8/mean-annual-nile-flow-1871-1970#!ds=22w8&display=line 2019.12