시계열 분석 기법과 응용

Week 3. ARMA 모형의 식별과 예측
3-3. ARMA모형의 예측

전치혁 교수
(포항공과대학교 산업경영공학과)

시계열 예측

최소 평균제곱오차 예측치

- 관측치: $Z_1, \dots, Z_n$
  - $f_{n,k}$ (k =1,2,⋯): 시점 n에서 k시점 이후 (즉, $n+k$ 시점) 시계열 예측치
  - 과거 시계열 관측치의 선형결합으로 예측

$$f_{n,k} = b_k Z_n + b_{k+1} Z_{n-1} +$$

$$\Downarrow\ Z_t = a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \cdots$$

$$f_{n,k} = c_k a_n + c_{k+1} a_{n-1} + \cdots$$

- 평균제곱오차 (mean squared error)

$$Q = E\left[\left(Z_{n+k} - f_{n,k}\right)^2\right] = E\left[\left(\sum_{j=0}^{\infty} \psi_j a_{n+k-j} - \sum_{j=k}^{\infty} c_j a_{n+k-j}\right)^2\right]$$

$$= \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} \left(\psi_{j+k} - c_{j+k}\right)^2$$

시계열 예측

최소 평균제곱오차 예측치

- Q를 최소로 하는 c_{j+k}의 추정치

  $\hat{c}_{j+k} = \psi_{j+k},\ k = 1,2,\ldots$

  $f_{n,k} = \psi_k a_n + \psi_{k+1} a_{n-1} + \cdots$

  $f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \ldots]$: k시점 이후 예측치는 조건부 기대치와 동일

- 예측오차

  $e_{n,k} = Z_{n+k} - f_{n,k} = a_{n+k} + \psi_1 a_{n+k-1} + \cdots + \psi_{k-1} a_{n+1},\ k = 1,2,\ldots$

  예측오차 분산

  $v_{n,k} = Var[e_{n,k}] = \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2$

  $v_{n,k} = Var[Z_{n+k} | Z_n, Z_{n-1}, \ldots]$: k시점 이후 예측오차분산은 조건부 분산과 동일

ARMA모형 예측

- 예측식: 시점 n까지의 과거 데이터로 부터 미래 시점 예측은 조건부 기대치 사용
  - 한단계 예측식: $f\_(n,1)=E[Z\_(n+1) \quad |Z\_n, Z\_(n-1),\cdots]$
  - k-단계 예측식: $f\_(n,k)=E[Z\_(n+k) \quad |Z\_n, Z\_(n-1),\cdots], \ k=1, 2,\cdots$
  - 계수는 추정값 이용, 과거 오차항은 잔차를 이용하여 추정

- 예측오차 분산: 예측구간이 필요한 경우 사용
  - 한단계 예측오차: $e\_(n,1)=Z\_(n+1)-f\_(n,1)=a\_(n+1)$
  - k-단계 예측오차: $e\_(n,k)=Z\_(n+k)-f\_(n,k)$
  - 한단계 예측오차 분산: $\mathrm{Var}[e\_(n,1)\ ]=Var[a\_(n+1)\ ] \ = \sigma\_a\hat{}2$
  - k-단계 예측오차 분산: $\mathrm{Var}[e\_(n,k)\ ]= Var[Z\_(n+k) \quad |Z\_n, Z\_(n-1),\cdots], \ k=1, 2,\cdots$

ARMA모형 예측

AR모형의 예측

- AR(2) 모형: $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$
  - 한단계 예측식: $f_{n,1} = E[Z_{n+1}|Z_n, Z_{n-1}, \dots] = \phi_1 Z_n + \phi_2 Z_{n-1}$
  - 2-단계 예측식: $f_{n,2} = E[Z_{n+2}|Z_n, Z_{n-1}, \dots] = \phi_1 f_{n,1} + \phi_2 Z_n$
  - 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
  - 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2}|Z_n, Z_{n-1}, \dots] = (1+\phi_1^2)\sigma_a^2$

- AR(4) 모형: $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_4 Z_{t-4} + a_t$
  - 한단계 예측식: $f_{n,1} = E[Z_{n+1}|Z_n, Z_{n-1}, \dots] = \phi_1 Z_n + \phi_2 Z_{n-1} + \cdots + \phi_4 Z_{n-3}$
  - 2-단계 예측식: $f_{n,2} = E[Z_{n+2}|Z_n, Z_{n-1}, \dots] = \phi_1 f_{n,1} + \phi_2 Z_n + \phi_3 Z_{n-1} + \phi_4 Z_{n-2}$
  - 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
  - 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2}|Z_n, Z_{n-1}, \dots] = (1+\phi_1^2)\sigma_a^2$

ARMA모형 예측

(예- AR(2) 모형) 시점 1에서 20까지의 시계열데이터가 다음과 같다.

시점 1-10: -1.356, -1.567, -0.994, -0.417, 0.840, -0.991, 0.166, 0.889, 0.514, -0.491
시점 11-20: -0.766, -1.936, -2.223, -1.395, -1.512, -0.582, 1.204, 1.706, -0.768, -0.313

- 과거데이터를 바탕으로 시계열모형은 다음의 AR(2)를 따르는 것으로 추정되었다.

$$Z_t = 0.7Z_{t-1} - 0.2Z_{t-2} + a_t \ (\sigma_a^2 = 1.0)$$

- 시점 10에서부터 한단계이후 예측치를 구하고 다음 시점의 실제값과 비교하여 예측오차를 산출하라.

(풀이) 한단계이후 예측치는 다음과 같다.

$$f_{n,1} = 0.7Z_n - 0.2Z_{n-1}$$
$$f_{11,1} = 0.7Z_{11} - 0.2Z_{10}$$

n	Z_n	예측치	오차
10	-0.491	-	-
11	-0.766	-0.4465	-0.3195
12	-1.936	-0.4380	-1.4980
13	-2.223	-1.2020	-1.0210
14	-1.395	-1.1689	-0.2261
15	-1.512	-0.5319	-0.9801
16	-0.582	-0.7794	0.1974
17	1.204	-0.1050	1.3090
18	1.706	0.9592	0.7468
19	-0.768	0.9534	-1.7214
20	-0.313	-0.8788	0.5658

시계열 예측

MA모형의 예측

- MA(2) 모형: $Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

- 한단계 예측식: $f_{n,1} = E[Z_{n+1}|Z_n, Z_{n-1}, \dots] = -\theta_1 a_n - \theta_2 a_{n-1}$
- 2-단계 예측식: $f_{n,2} = E[Z_{n+2}|Z_n, Z_{n-1}, \dots] = -\theta_2 a_n$
- 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
- 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2}|Z_n, Z_{n-1}, \dots] = (1 + \theta_1^2)\sigma_a^2$

- MA(q) 모형: $Z_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$

- 한단계 예측식: $f_{n,1} = E[Z_{n+1}|Z_n, Z_{n-1}, \dots] = -\theta_1 a_n - \theta_2 a_{n-1} - \cdots - \theta_q a_{n+1-q}$
- 2-단계 예측식: $f_{n,2} = E[Z_{n+2}|Z_n, Z_{n-1}, \dots] = -\theta_2 a_n - \theta_3 a_{n-1} - \cdots - \theta_q a_{n+2-q}$
- 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
- 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2}|Z_n, Z_{n-1}, \dots] = (1 + \theta_1^2)\sigma_a^2$
- 실제 예측치 산출을 위해서는 추정치 사용, , 이 때 a_n의 추정치로는 해당 시점의 잔차를 사용

ARMA모형 예측

(예- MA(2) 모형) 시점 1에서 20까지의 시계열데이터가 다음과 같다.

시점 1-10: 1.377, 1.856, -0.655, -0.587, -0.188, 1.414, 0.731, -1.628, -0.511, -0.294

시점 11-20: 0.499, -0.442, 1.019, -1.705, -0.139, 0.219, 1.131, -0.508, 0.541, -0.809

- 과거데이터를 바탕으로 시계열모형은 다음의 MA(2)를 따르는 것으로 추정되었다.

$$Z_t = a_t - 0.6a_{t-1} - 0.2a_{t-2} \ (\sigma_a^2 = 1.0)$$

- 시점 11에서부터 한단계이후 예측치를 구하고 다음 시점의 실제값과 비교하여 예측오차를 산출하라.

(풀이) 한단계이후 예측치는 다음과 같다.

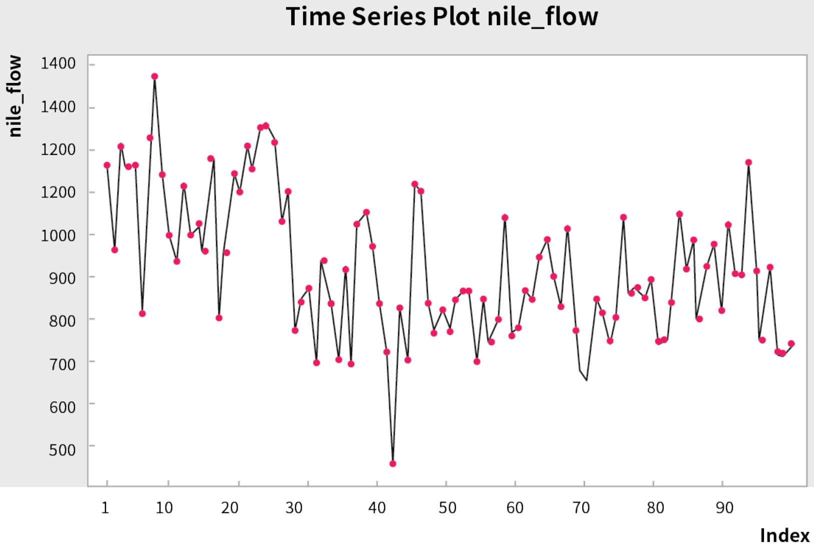
$$f_{n,1} = -0.6a_n - 0.2a_{n-1}$$

여기서 $a_t = (1 - 0.6B - 0.2B^2)^{-1} \approx (1 + 0.6B + 0.56B^2 + 0.456B^3 + 0.3856B^4)Z_t$

n	Z_n	a_n	예측치	오차
10	-0.294	-0.6337	-	-
11	0.499	-0.4240	0.4814	0.0176
12	-0.442	-1.1680	0.3812	-0.8232
13	1.019	0.7021	0.7856	0.2334
14	-1.705	-1.2269	-0.1877	-1.5173
15	-0.139	-0.6005	0.5957	-0.7347
16	0.219	-0.5250	0.6057	-0.3867
17	1.131	0.8000	0.4351	0.6959
18	-0.508	-0.4276	-0.3750	-0.1330
19	0.541	0.9158	0.0965	0.4444
20	-0.809	-0.1687	-0.4640	-0.3450

시계열 예측

- 예 (나일강 유량) 1872-1970년 사이 나일강 유량 시계열의 추정 모형은 AR(2)이며 추정식은 다음과 같다.
  - $\hat{Z}_t = 357.5 + 0.4039 Z_{t-1} + 0.2064 Z_{t-2}$. 1967년 부터 5년간 한단계 예측값은 표와 같다.



년도	실제값	예측값	잔차
1967	919	847.0	72.0
1968	718	882.7	-164.7
1969	714	837.2	-123.2
1970	740	794.1	-54.1
1971	-	803.8	

#1

Reference

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