ARMA 모형의 상태공간에 의한 최대 우도 적합에 관한 연구

Maximum Likelihood Fitting of ARMA Models With State-Space Approach

박 종 구*
오 대 호**
신 호 중***

ABSTRACT

The state-space model has been extensively applied to modelling data from economics, medicine, engineering, etc. The model also can be used to the calculation of likelihood function of a stationary autoregressive and moving average(ARMA) time series when a time series data is fitted by maximum likelihood. Here rewriting of ARMA model into state-space model, the form of the likelihood function and Kalman filter, Kalman smoothing techniques are reviewed. Computation of initial state covariance matrix is important when Kalman recursion is applied and its achievement is shown by based on impulse response. The model parameter estimation by maximum likelihood fitting leads to difficult nonlinear optimization technique and here recursive estimation algorithm by EM(expectation maximization) particulary useful in state-space involving unobserved components or irregularly observed data is dicussed. And its example is represented.

1. 서 론

상태공간 모형은 Kalman과 Bucy에 의해 도입되었다. 주로 항공관련 분야에 사용하기 위해 도입 되었기 때문에 공학과 제어 분야의 연구가 많이 이루어졌다. 그러나 최근 통계학자와 계량경제 학자의 많은 연구에 의해 경제, 경영 그리고 의학, 지질 등 많은 분야에서 연구, 적용되어졌다. Kailath(1970)의 innovation 접근법은 여러 분야의 응용에서의 발전에 근간을 이룬 기여를 하였다.

거의 모든 기존 시계열 모형은 상태공간 모형의 특수한 경우로 생각 할 수 있다. 시계열 자료에 대해 어떤 ARMA 모형이 가능하다면 그것은 상태공간 모형으로 쉽게 변형될수 있다. 그러나 자료로 부터 직접적으로 상태공간 모형을 구축하는것이 더욱 의미있다고 생각하는 학자들도 있다. Aoki(1987)는 모형

* 원광대 컴퓨터 공학과 교수
** 동국대 통계학과 박사 과정 수료
*** 재정 경제원 예산실 사무관

의 구조 식별과 모수 추정에 있어서 제어 이론의 최소 실현(minimal realization) 알고리즘을 사용했는데 이는 자료로 부터 직접 모형을 구축한 것 이었다.

상태공간 모형은 두 개의 방정식으로 구성된다. 시점 t 에서의 체계의 상태를 표현하는 상태 방정식과 관측 되지않은 상태 벡터와 관측 오차의 함수인 관측 방정식이 그것이다. 시계열 자료를 모형화 하기위해 상태공간 방법론을 이용 하는데는 모형 식별과 모수 추정 문제가 요구 된다. 본 논문에서는 모수 추정 문제를 다루기로 한다.

Harvey & Phillips(1979)은 ARMA 과정에서의 우도 함수 계산에 있어 Kalman 반복법을 이용 할 수 있음을 보였다. 이에 따르면, 다른 방법론에 비해 계산적 장점은 첫째, 미래값의 예측과 이에 대한 조건부 평균 자승 오차(conditional mean square error)들이 매우 쉽게 얻어 질 수 있으며 둘째, 모형이 정확히 규정 되었고 ARMA 계수를 알면, 평균이 영이고 상수 분산의 정규 분포를 따르는 예측 오차들을 생성 한다는 것이다.

시계열 모형에 있어서 최우 추정치의 계산 문제는 비선형 최적화 기법을 이용 해야 하는데 본 논문에서는 Kalman 반복에 의한 EM 알고리즘의 적용을 논의 한다. EM 알고리즘은 비관측된 요소 혹은 불규칙적 관측 자료를 포함한 시계열 응용에 적절하다(Dempsteretal (1977)). 불완전 자료 y는 비 관측된 신호 과정 Z와 비 관측된 잡음 과정 v의 기지(known) 함수로서 관측 된다 가정 하면 이 때, 불완전-자료에 근거한 로그 우도 함수는 $\ln L(y,\lambda)$, 완전-자료에 근거한 우도 함수는 $\ln L(Z,v,\theta)$으로 정의 할 수 있다. 이 때 완전-자료에 근거한 $\ln L(Z,v,\theta)$을 최대화 하는것이 용이하며 주어진 관측 자료 y 의 완전-자료 우도의 조건부 기대값을 최대화 하는 반복적 절차가 요구 된다. 즉 EM 알고리즘의 E-단계인 $Q(\lambda|\lambda_k)=E_k[\ln L(Z,v,\lambda)|y]$을 최대화하여 모수들을 갱신 한다. λ_k는 단계 k에서의 모수값이며, E_k는 λ_k 하에서의 기대값이다. 본 논문에서는 안정적 ARMA 모형의 우도 함수를 계산하기 위해 상태 공간 모형으로의 재표현과 예측 오차 분할에 의해 우도 함수가 간단히 표현됨을 논의하며, 모수 추정에 있어서 Kalman 반복 추정이 검토되며, 이 때의 초기값 문제와 EM 알고리즘에 의한 비선형 최적화 기법이 논의 되어 진다.

2. 마아코프 표현

마아코프 표현은 현 상태벡터의 최적추정에 있어서 미래를 예측하는데 필요한 과거와 현재의 관측치들에서의 정보를 함축한다. 다음 시점에서의 관측치와 현 상태의 최적 추정치로부터의 예측간의 차이는 이전 관측치들에 대해 직교이며(Kalman, 1960) 가우스 과정(Gaussian process)에 대해 이것은 우도함수 계산에 대해 간단한 반복법을 제공한다.

차수 p, q의 ARMA 모형은 다음과 같이 쓸 수 있다. 즉

$$x_t = \sum_{k=1}^{p} \phi_k x_{t-k} + \varepsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k}$$

단, ε_t 은 무상관의 확률잡음(random noise, or shock)으로 $E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma^2$ 이다. 안정성(stationarity)과 가역성(invertibility) 조건을 만족하기 위해,

$1-\sum_{k=1}^{p} \phi_k B^k = 0$ 와 $1+\sum_{k=1}^{q} \theta_k B^k = 0$ 의

근들이 단위원 외부(outside the unit circle)에 존재한다고 가정하자.

상태공간 표현을 위해 위 과정의 상태를 길이 $m = Max(p, q+1)$의 열벡터로 정의한다

$$Z(t) = \begin{bmatrix} x(t|t) \\ x(t+1|t) \\ \vdots \\ x(t+m-1|t) \end{bmatrix}$$

단, $x(t+j|t) = E\{x(t+j)|x(s), s \le t\}$으로 j-단계 예측(j-step prediction)이다.

이 때 $x(t|t) = x_t$ 이다.

1-단계 예측은

$$x(t+1|t) = \sum_{k=1}^{p} \phi_k x_{t+1-k} + \sum_{k=1}^{q} \theta_k \varepsilon_{t+1-k}. \quad (2.1)$$

j-단계 예측은

$$x(t+j|t) = \sum_{k=1}^{j-1} \phi_k x(t+j-k|t)$$

$$+ \sum_{k=j}^{p} \phi_k x_{t+j-k} + \sum_{k=j}^{q} \theta_k \varepsilon_{t+j-k} \text{ 이며 } (2.2)$$

$$x(t+j|t+1) = \sum_{k=1}^{j-2} \phi_k x(t+j-k|t+1)$$

$$+ \sum_{k=j-1}^{p} \phi_k x_{t+j-k} + \sum_{\theta_k} \varepsilon_{t+j-k} \text{ 이다 } (2.3)$$

식 (2.2)와 (2.3)으로부터

$$x(t+j|t+1) - x(t+j|t) = \sum_{k=1}^{j-1} \phi_k \{x(t+j-k|t+1) - x(t+j-k|t)\}$$

$$+ \theta_{j-1} \varepsilon_{t+1} \text{ 이다.} \quad (2.4)$$

$x(t+j|t+1) - x(t+j|t) = g_j \theta_{t+1}$ 이라 하자. 그러면

$$x(t+j|t+1) - x(t+j|t)$$

$$= \{\theta_{j-1} + \sum_{k=1}^{j-1} \phi_k g_{j-k}\} \times \varepsilon_{t+1}$$

$$= g_j \varepsilon_{t+1} \text{ 이다.} \quad (2.5)$$

단,

$$g_1 = 1, g_j = \theta_{j-1} + \sum_{k=1}^{j-1} \phi_k g_{j-k}, \quad \theta_j = 0, j > q$$

이다. (2.6)

(m-1)-단계 예측은,

$$x(t+m|t+1) = \sum_{k=1}^{p} \phi_k x(t+m-k|t) + g_m \varepsilon_{t+1}$$

이다. (2.7)

위 방정식들은 다음의 상태방정식으로 표현될 수 있다.

$$\begin{bmatrix} x(t+1|t+1) \\ x(t+2|t+1) \\ \vdots \\ x(t+m|t+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots \\ \phi_m & \cdots & & \phi_2 & \phi_1 \end{bmatrix}$$

$$\times \begin{bmatrix} x(t|t) \\ x(t+1|t) \\ \vdots \\ x(t+m-1|t) \end{bmatrix} + \begin{bmatrix} 1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} \varepsilon_{t+1}$$

이 때 $\phi_i = 0,\ i > p$ 이다. 행렬식으로 쓰면

$$Z(t+1) = FZ(t) + G\varepsilon(t+1) \qquad (2.8)$$

이다.

이 때, $Z(t)$ 는 벡터값의 마아코프 과정이며, F 는 $m \times m$ 상태 추이 행렬으로 마지막 행에는 AR 계수들을 포함하고 있다. G 벡터의 원소들은 ARMA 모수들로부터 계산된다.

상태공간 표현에서의 두 번째 방정식은 관측 방정식(observational equation)으로 다음과 같다

$$y(t) = [1\ 0\ \cdots\ 0] \begin{bmatrix} x(t|t) \\ x(t+1|t) \\ \vdots \\ x(t+m-1|t) \end{bmatrix} + v(t). \qquad (2.9)$$

혹은 $y(t) = HZ(t) + v(t)$, 단 $v(t)$는 시점에 관해 무상관의 관측 오차(observational error)인 확률변수 이며 θ 와는 독립이고, 평균이 영이며 분산은

$$R = E(v(t))^2 \text{ 이다.}$$

3. 예측오차의 분할 (prediction error decomposition)

$x = (x(1), x(2), \cdots, x(n))'$은 ARMA 모형,

$\phi(B)x(t) = \theta(B)\varepsilon(t)$

단,

$\phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ 이며,

$\theta(B) = 1 + \theta_1(B) + \cdots + \theta_q B^q$의 실현이라하자.

그리고 $\{\varepsilon(t)\}$ 은 서로 독립인 정규분포, $N(0, \sigma^2)$를 따른다고 하자. 이때 우도함수는 다음과 같이 쓸수 있다.

$$L(x(1), x(2), \cdots, x(n)) = L(x(1)).$$
$$\times \prod_{t=2}^{n} L(x(t)|x(t-1), \cdots, x(1)) \qquad (3.1)$$

이 형태를 우도(likelihoo)의 "예측오차 분할"이라 한다.

$t = 2, \cdots, n$에 대해 다음을 정의하자.

$$\hat{x}(t|t-1) = E(x(t)|x(t-1), \cdots, x(1))$$

$$\sigma^2(t|t-1) = var(x(t)|x(t-1), \cdots, x(1))/\sigma^2$$
$$= E(x(t) - \hat{x}(t|t-1))^2/\sigma^2.$$

그리고,

$$\hat{x}(1|0) = E(x(1)) = 0,$$

$\sigma^2(1|0) = (1/\sigma^2)var(x(1))$이다.

위의 정의에 의해 우도함수(3.1)은 다음과 같이 쓸 수 있다.

$$(2\pi\sigma^2)^{-n/2}\left(\prod_{t=1}^{n} \sigma^2(t|t-1)\right)^{-1/2} \times$$

$$\exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{n}((x(t)-\hat{x}(t|t-1))/\sigma(t|t-1))^2\right)$$

$e(t), t=1, \cdots, n$은 표준화된 1-단계 예측 오차들, $e(t)=(x(t)-\hat{x}(t|t-1))/\sigma(t|t-1)$이라 하고,

$e=(e(1), \cdots, e(n))'$라 하자. 그러면 로그-우도함수(log-likelihood function)은 다음과 같다.

$$l=\frac{-1}{2}\{nln(\sigma^2)+\ln(\prod_{t=1}^{n}\sigma^2(t|t-1))+e'e/\sigma^2\}$$

이 때 σ^2의 m.l.e 는 $\hat{\sigma}^2=(1/n)e'e$ 이며 상수항을 제외한 로그-우도함수는

$$l^*=\frac{-1}{2}\{nln(e'e)+\ln(\prod_{t=1}^{n}\sigma^2(t|t-1))\}$$ 이다.

따라서 ARMA 모형의 우도 함수는 예측오차들로 표현될수 있으며 Kalman filter 는 이 값을 구하는데 이용될수 있다.

4. 상태공간 표현과 Kalman 반복 (recursion)

식(2.8)과 (2.9)로부터 다음 상태공간 모형을 고려하자.

$$Z(t+1)=FZ(t)+G\varepsilon(t+1) \qquad (4.1)$$

$$y(t)=HZ(t)+v(t) \qquad (4.2)$$

단, $Z(t)$ 와 $y(t)$ 은 결합안정 가우스 과정이며, ε_t 와 $v(t)$은 결합안정 가우스의 평균이 영인 백색소음과정이고,

$E[\varepsilon(t)v(s)']=0\ \forall s, t$이다. 또한,

$E[Z(t)]=0\ \forall\ t,$

$E(x(t)\varepsilon(s)')=0\ \forall\ s\geq t,$

$E(Y(t)\varepsilon(s)')=0\ \forall\ s\geq t$

(Priestley, Subba Rao(1975)) 이다.

상태공간 모형에서 $Z(t)$를 추정함에 있어서 Kalman 반복 절차를 이용하기 위해 Kalman 필터추정치와 Kalman 평활 추정치를 고려하기로 한다.

관측 자료가 $y(1), y(2), \cdots, y(n)$ 일때 다음의 조건부 기대값에 대해,

$$Z(t|s)=E[Z(t)|y(s), \cdots, y(1)],$$

이때 s=t-1 일때를 Kalman 필터 추정치라 하며, s=n 일때 Kalman 평활 추정치라 한다.

$e(t+1)=y(t+1)-y(t+1|t)$ 이라 하자.

이는 $y(t+1)$ 의 1-단계 예측 오차이다. 식(2.3.2)의 양변에 조건부 기대값을 취하면

$$\begin{aligned} y(t+1|t) &= E(y(t+1)|y(t), \cdots) \\ &= E(HZ(t+1)+v(t+1)|y(t), \cdots) \\ &= HZ(t+1|t). \end{aligned} \qquad (4.3)$$

식 (4.2)와 (4.3)으로 부터,

$$\begin{aligned} e(t+1) &= H[Z(t+1)-Z(t+1|t)]+v(t+1) \\ &= Hf(t+1)+v(t+1) \end{aligned}$$

단, $f(t+1)=Z(t+1)-Z(t+1|t). \qquad (4.4)$

이 때, $E[f(t+1)]=0$이므로

$$cov[f(t+1)] = E[f(t+1)f(t+1)'] = P(t+1|t).$$

모형(4.4)는 인자분석(factor analysis)모형과 같다. 따라서 $f(t+1)$ 의 추정치를 구하기 위해 회귀적 방법(regression approach)을 통한 인자점수(factor score)를 추정하는것을 이용하면 된다.(R.A. Johnson, D.W. Wichern).

인자분석 모형에 (4.4) 을 적용하면

$e(t+1) \sim N(0,\ HP(t+1|t)H' + R)$이 된다.

그리고 $cov[e(t+1),\ f(t+1)] = HP(t+1|t)$,

또한 $f(t+1)$의 추정치는 다음과 같다.

$$\begin{aligned}\hat{f}(t+1) &= E[f(t+1)|y(t+1), y(t), \cdots] \\ &= E[f(t+1)|e(t+1)] \\ &= E[Z(t+1)-Z(t+1|t)|e(t+1)] \\ &= Z(t+1|t+1) - Z(t+1|t) \qquad (4.7)\end{aligned}$$

e(t+1) 와 f(t+1)은 평균이 영이고 공분산행렬이 Σ^*인 결합정규분포를 따른다 단,

$$\Sigma^* = \begin{bmatrix} HP(t+1|t)H' & HP(t+1|t) \\ P(t+1|t)'H' & P(t+1|t) \end{bmatrix}$$

$$E[f(t+1)|e(t+1)] =$$

$$P(t+1|t)'H'(HP(t+1|t)H'+R)^{-1}e(t+1)$$

$$cov[f(t+1)|e(t+1)] = P(t+1|t)$$

$$-P(t+1|t)'H'(HP(t+1|t)H'+R)^{-1}$$

$$\times HP(t+1|t)$$

따라서, $\hat{f}(t+1) = P(t+1|t)'H'\Sigma_e^{-1}$

$$\times[y(t+1) - y(t+1|t)],$$

$\Sigma_e = E[e(t)e(t)']$이다. (4.8)

(4.7)과 (4.8)로 부터,

$$Z(t+1|t+1) = Z(t+1|t) + K\{y(t+1) - HZ(t+1|t)\} \qquad (4.9)$$

단, $K = P(t+1|t)'H'\Sigma_e^{-1}$ 이며 Kalman 이득(gain)이라 한다. 또한,

$$\begin{aligned}Z(t+1|t) &= E[Z(t+1)|y(t), \cdots] \\ &= FE[Z(t)|y(t), \cdots] \\ &= FZ(t|t). \qquad (4.10)\end{aligned}$$

(4.4)와 (4.10)으로부터,

$Z(t+1|t+1) = Z(t+1)$이므로

$$\begin{aligned}f(t+1) &= Z(t+1) - Z(t+1|t) \\ &= F[Z(t) - Z(t|t)] + G\varepsilon(t+1).\end{aligned}$$

따라서, $f(t+1)$의 공분산행렬 $P(t+1|t)$는

$$\begin{aligned}cov[f(t+1)] &= P(t+1|t) \\ &= FP(t|t)F' + \sigma^2 GG',\end{aligned}$$

$$P(t|t) = E[(Z(t) - Z(t|t))(Z(t) - Z(t|t))']. \quad (4.11)$$

그리고 (4.9)와 (4.10)으로 부터,

$Z(t+1|t+1) = FZ(t|t) + Ke(t+1)$ 이다.

필터 오차의 공분산행렬 $P(t|t)$ 는

$$\begin{aligned}P(t+1|t+1) &= E[Z(t+1) - Z(t+1|t+1)] \\ &\quad \times [Z(t+1) - Z(t+1|t+1)]' \\ &= E[FZ(t+1) + G\varepsilon(t+1)\end{aligned}$$

$$-FZ(t|t)-Ke(t+1)]$$
$$\times[FZ(t)+G\varepsilon(t+1)$$
$$-FZ(t|t)-Ke(t+1)]'$$
$$=P(t+1|t)-K\Sigma_e K'.$$

따라서 전진 반복(forward recursion)에 의한 Kalman 필터 반복을 기술하면,

$$y(t+1|t)=HFZ(t|t)\ ,$$
$$e(t+1)=y(t+1)-y(t+1|t),$$
$$P(t+1|t)=FP(t|t)F'+\sigma^2 GG',$$
$$\Sigma_e=HP(t+1|t)H'+R,$$
$$K_t=P(t+1|t)H'\Sigma_e^{-1},$$
$$Z(t+1|t+1)=FZ(t|t)+Ke(t+1)$$
$$P(t+1|t+1)=P(t+1|t)-K\Sigma_e K' \text{ 이다.} \tag{4.12}$$

만일 $Z(t)$의 추정치가 모든 관측 자료 $y(1),\cdots,y(n)$에 근거한다면 Kalman 평활 추정치가 필요하며, 이때의 후진 반복(backward recursion)에 의한 Kalman 평활 반복을 요약하면, $t=n,n-1,\cdots,1$ 에 대해,

$$Z(t-1|n)=Z(t-1|t-1)$$
$$+L_{t-1}(Z(t|n)-Z(t|t-1)),$$
$$L_{t-1}=P(t-1|t-1)F'P(t|t-1)^{-1},$$

평활 추정치의 공분산 행렬은

$$P(t-1|n)=P(t-1|t-1)+L_{t-1}(P(t|n)$$
$$-P(t|t-1))L'_{t-1}. \tag{4.13}$$

$t=n,n-1,\cdots,2$ 에 대해서

$$P(t-1,t-2|n)$$
$$=cov(Z(t),Z(t-1)|y(1)\cdots y(n))$$
$$=P(t-1|t-1)L_{t-2}'+L_{t-1}(P(t,t-1|n)$$
$$-FP(t-1|t-1))L_{t-2}'$$

단,

$$P(n,n-1|n)=(I-K_nH)FP(n-1|n-1).$$

5. 우도함수의 계산

4절에서의 Kalman filter 반복 추정에의해 초기 상태 벡터 Z(0|0) 와 공분산 행렬, P(0|0)를 규정함으로써 우도함수를 계산할수 있다.

1-단계 상태 예측

$$Z(t+1|t)=FZ(t|t), \tag{5.1}$$

이 예측의 공분산 행렬

$$P(t+1|t)=FP(t|t)F'+\sigma^2 GG', \tag{5.2}$$

다음 시점의 관측치에 관한 예측치

$$y(t+1|t)=HZ(t+1|t)=x(t+1|t), \tag{5.3}$$

상태 벡터의 갱신

$$Z(t+1|t+1)=Z(t+1|t)$$
$$+K(t+1)[y(t+1)-y(t+1|t)] \tag{5.4}$$

단, $K(t+1)=P(t+1|t)H'$

$$\times[HP(t+1|t)H'+R]^{-1} \tag{5.5}$$

이때, $HP(t+1|t)H'$ 는 $P(t+1|t)$ 의 첫 번째 원소이다.

또한, 갱신된 상태의 공분산 행렬은

$$P(t+1|t+1)=P(t+1|t).$$

$$-K(t+1)HP(t+1|t) \qquad (5.6)$$

관측치의 예측 오차,

$e(t+1)=y(t+1)-y(t+1|t)$ 는 이전 관측치와 직교인 요소이며 분산은,

$V(t+1)=HP(t+1|t)H'+R$ 이다.

이는 시계열과정의 시점 t+1에서의 1-단계 예측 오차 분산과 관측 오차 분산의 합이다. 이 오차들이 가우스 분포를 따르면, n개의 관측치에 대한 우도함수는 3]절의 예측 오차 분할으로부터

$$L=\prod_{i=1}^{n}(2\pi V_i)^{-1/2}\exp(-e_i^2/2V_i)\text{이며,}$$

$l=-2\ln$ 우도함수는

$$l \propto \sum_{i=1}^{n}[\ln V_i+e_i^2/V_i]\text{이다.} \qquad (5.7)$$

6. 초기 상태 공분산 행렬(initial state covariance matrix)의 계산

모수, $\phi_1,\cdots,\phi_p,\theta_1,\cdots,\theta_q$의 예비추정치는 Box & Jenkins(1976, p.201)의 방법으로 부터 구할수 있으며 이 예비 추정치와 상태 벡터 공분산 행렬의 초기값으로부터 Kalman 반복에 의해 우도함수의 값을 계산할 수 있다.

초기 상태 벡터, $Z(0|0)=[0,0\cdots0]'$ 으로 첫 번째 관측 이전 상태를 표현한다.

공분산 행렬의 초기값은 Akaike(1978)의 방법에 의해서 구할수 있으며 과정(process)과 오차들간의 cross covariance 로 부터 유도된다.

$$\gamma_0=E[x_t\varepsilon_t]=\sigma^2,$$

$$\gamma_k=E[x_{t+k}\varepsilon_t]$$
$$=E[(\sum_{j=1}^{p}\phi_j x_{t+k-j}+\sum_{j=0}^{q}\theta_j\varepsilon_{t+k-j})\varepsilon_t]$$
$$=\sigma^2\theta_k+\sum_{j=1}^{p}\phi_j\gamma_{k-j},\quad k>0$$

((k-1) lagged cross covariance) (6.1)

(6.1)을 σ^2 으로 나누어 정규화(normalization) 하면

$$\gamma_k=\theta_k+\sum_{j=1}^{p}\phi_j\gamma_{k-j},\ \gamma_0=1 \qquad (6.2)$$

이것은 식 (2.6) 에서의 g와 같은 결과이다.

즉, $g_k=\gamma_{k-1}$. (6.3)

따라서 벡터 G 는 impulse response 가 있으며 ARMA 모형을 전달 함수 모형으로 표현하면,

$$x_t=\sum_{k=0}^{\infty}\gamma_k\varepsilon_{t-k}\ \text{단,}\ \gamma_0=1 \qquad (6.4)$$

으로 나타낼수 있다. 따라서 상태 벡터의 원소들은

$$x(t+j|t)=\sum_{k=0}^{\infty}\gamma_{j+k}\varepsilon_{t-k},\ j\geq 0\ \text{이며,} \qquad (6.5)$$

$$x_{t+j}=\sum_{k=-j}^{-1}\gamma_{j+k}\varepsilon_{t-k}+x(t+j|t). \qquad (6.6)$$

ARMA 과정의 자공분산(autocovariance)을 구하면

$$E[x_t x_{t+j}]=E[(\gamma_0\varepsilon_t+\gamma_1\varepsilon_{t-1}+\cdots)$$
$$\times(\gamma_0\varepsilon_{t+j}+\cdots+\gamma_{j-1}\varepsilon_{t+1}+x(t+j|t))]$$
$$=E[x(t|t)x(t+j|t)]\ ,j\geq 0$$

그러므로,

$$C_{j-i}=E[x(t+i)x(t+j)]$$

$$= E[\sum_{k=-i}^{-1} \gamma_{i+k}\varepsilon(t-k) + x(t+i|t)] \times$$

$$[\sum_{k=-j}^{-1} \gamma_{j+k}\varepsilon(t-k) + x(t+j|t)$$

$$= \sigma^2 \sum_{k=0}^{i-1} \gamma_k \gamma_{k+j-i}$$

$$+ E[x(t+i|t)x(t+j|t)]. \qquad (6.7)$$

위 식의 우변은 cross covariance 와 상태백터의 공분산 이며 이를 계산해야 한다.

$$C_k = E[x_{t+k}x_t]$$

$$= E[(\sum_{j=1}^{p} \phi_j x_{t+k-j} + \sum_{j=0}^{q} \theta_j \varepsilon(t+k-j))x(t)]$$

$$+ \sum_{j=0}^{q} \theta_j \varepsilon(t+k-j))x(t)]$$

$$= \sum_{j=1}^{p} \phi_j C_{k-j} + \sum_{j=k}^{q} \theta_j \gamma_{j-k}. \qquad (6.8)$$

(6.2) 으로 부터, cross covariance 들은

$$\gamma_0 = 1$$
$$\gamma_1 = \phi_1\gamma_0 + \theta_1$$
$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 + \theta_2$$
$$\vdots \qquad (6.9)$$

따라서 (6.7)과 (6.8)을 σ^2으로 나누어 상태 공분산행렬의 초기값들을 구할수 있다.

$$P_{0j}(0|0) = C_j,$$

$$P_{ij}(0|0) = C_{j-i} - \sum_{k=0}^{i-1} \gamma_k \gamma_{k+j-i}. \qquad (6.10)$$

7. EM 알고리즘

EM 알고리즘은 비 관측요소 혹은 불규칙적으로 관측된 데이터를 포함한 시계열응용에 적합하다. 시계열에서 불완전 데이터 $y(t)$ 를 주로 관측하며, 이는 비 관측된 신호 과정 Z 와 비 관측 소음과정, v 의 함수이다.

상태공간 모형 (4.1), (4.2)를 다음과 같이 다시 표현하자.

$$Z(t+1) = FZ(t) + W(t+1)$$
$$y(t) = HZ(t) + v(t),$$

단, $W(t+1) = G\varepsilon(t+1)$ (7.1)

$$Q = E[w(t)w(t)'] = \sigma^2 GG'.$$

이때 $Z(0), W(1), \cdots, W(n)$ 과 $v(1), \cdots, v(n)$은 결합정규분포이며 무상관이다. 완전-데이터 (complete-data) 로그 우도함수는 (5.7)로 부터,

$$\ln L(Z, v, \lambda) \propto -\frac{1}{2}\ln|V(0)|$$
$$-\frac{1}{2} Z(0)'V(0)^{-1}Z(0) - \frac{T}{2}\ln|Q|$$
$$-\frac{1}{2}\sum_{t=1}^{n}(Z(t+1) - FZ(t))'$$
$$\times Q^{-1}(Z(t+1) - FZ(t))'$$
$$-\frac{n}{2}\ln R - \frac{1}{2}\sum_{t=1}^{n} R^{-1}$$
$$\times (y(t) - HZ(t))'(y(t) - HZ(t)) \qquad (7.2)$$

이 때, $V(0)$ 는 초기상태, $Z(0)$의 공분산이며 모수 $\lambda = (F, Q, R)$ 이다. 만일 λ의 최근

값이 λ_i 이고 E_i 는 모수 λ_i에서의 기대값 이라면

$$Q(\lambda|\lambda_i)=E_i[\ln L(Z,v,\lambda)|y_1,y_2,\cdots,y_n]$$

$$=-\frac{1}{2}\ln V(0)-\frac{n}{2}\ln|Q|$$

$$-\frac{1}{2}tr\{Q^{-1}[S_t(0)-S_t(1)F'$$

$$-FS_t'(1)+FS_t'(0)F']\}-\frac{n}{2}\ln R$$

$$-\frac{1}{2}tr\{R^{-1}\sum_{t=1}^{n}[(y(t)-HZ(t|n))$$

$$\times(y(t)-HZ(t|n))'+HP(t|n)H']\},$$

단, $S_t(j)=\sum_{t=1}^{n}(P(t,t-j|n)$

$$+Z(t|n)Z(t-j|n)'\ j=0,1 \qquad (7.3)$$

(7.3)를 모수, $\lambda=(F,Q,R)$에 관해서 최대화 하면 EM 알고리즘의 최대화 단계, M-step을 다음과 같이 얻는다

$$F(i+1)=S_t(1)[S_{t-1}(0)]^{-1}, \qquad (7.4)$$

$$Q(i+1)=n^{-1}\{[S_t(0)-S_t(1)[S_{t-1}(0)]^{-1}$$

$$\times S_t'(1)]\}, \qquad (7.5)$$

$$R(i+1)=n^{-1}\sum_{t=1}^{n}[(y(t)-HZ(t|n))$$

$$\times(y(t)-HZ(t|n))'+HP(t|n)H'] \quad (7.6)$$

이제 모수 추정과 우도함수 계산을 위한 전체 절차를 요약하면 다음과 같다.

단계 1) 6)절에서 논한 초기 상태벡터와 초기 공분산 행렬의 계산, 모수 $\lambda_0=(F_0,Q_0,R_0)$을 초기화.

단계 2) Kalman 반복법에 의해 $Z(t|n),P_{tt}^{n},P_{t,t-1}^{n}$ 을 계산.

단계 3) 로그 우도함수 (5.7)를 계산.

단계 4) EM 알고리즘을 이용해서 모수 $\lambda_i=(F_i,Q_i,R_i)$을 갱신.

단계 5) 단계 2)에서 반복.

8. 예 제

ARMA 모형의 최대 우도함수에 의한 적합의 예는, Box 와 Jenkins 의 *Wälfer* sunspot number(Box,Jenkins, 1976, p.530) 의 자료를 사용하였다. 이 자료를 AR(2) 모형으로 적합했을때의 상태 공간 형태는 다음과 같다

$$\begin{bmatrix} x(t+1|t+1) \\ x(t+2|t+1) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ \phi_2 & \phi_1 \end{bmatrix}\begin{bmatrix} x(t|t) \\ x(t+1|t) \end{bmatrix}+\begin{bmatrix} 1 \\ g_2 \end{bmatrix}\varepsilon_{t+1}$$

$$y(t)=[1\ \ 0]\begin{bmatrix} x(t|t) \\ x(t+1|t) \end{bmatrix}+v(t).$$

이 때 AR 계수들의 초기 추정값은 자공분산들로 계산 되어지며, $\widehat{\phi_1}=-0.6345$, $\widehat{\phi_2}=1.3175$ 이며,

표 8.1: Kalman 반복과 EM 알고리즘에 의한 모수 추정치

반복	ϕ_1	ϕ_2	q_{11}	q_{12}	q_{22}	$-2\ln L$
1	1.317485	-0.634516	267.629500	352.597823	464.542304	822.329274
2	1.029687	-0.178416	352.416017	449.555611	578.923617	794.893957
3	0.807478	-0.001704	423.608304	512.614002	631.275478	800.074620
4	0.663783	0.096349	487.782852	562.328104	664.082518	803.219856
5	0.567183	0.159731	539.531120	596.534649	679.370149	805.410705
6	0.499647	0.203412	579.280793	617.297160	680.958582	807.053002
7	0.451227	0.234578	609.086164	627.724587	672.937218	808.305351
8	0.415980	0.257283	631.161726	630.700164	658.712028	809.258225
9	0.390104	0.274034	647.420955	628.535798	640.825695	809.978895
10	0.371024	0.286497	659.378414	622.962899	621.062281	810.521392

백색 잡음 ε_t의 분산의 초기 추정치는

$\hat{\sigma}^2 = 267.6296$ 이다.

또한 초기 상태 공분산 행렬 추정치는 식 (6.2)와 (6.9),(6.10)으로 부터

$$P(0|0) = \begin{bmatrix} 4.7786 & 3.8518 \\ 3.8518 & 3.7786 \end{bmatrix} \text{이다.}$$

최우 추정치 계산을 위해서 MATLAB 언어를 이용 하였으며 표 8.1 과 같은 결과를 얻는다.

표 8.1 로 부터 $-2\ln$우도 는 794.894 에서 최소값을 갖으며 이때의 AR 계수들의 값은 $\phi_1 = 1.0297$, $\phi_2 = -0.1784$ 이며 적합된 모형은 $x_t = 1.0297\, x_{t-1} - 0.1784\, x_{t-2} + \varepsilon_t$ 이다.

9. 결 론

본 논문에서는 안정적 ARMA 시계열 모형의 최우 추정치에 의한 적합에 있어서 마아코프 표현에 의한 재 표현과 예측 오차 분할에 의한 우도 함수를 간단히 표현 할 수 있음을 논의 하였으며 모수 추정을 위해 Kalman 반복법을 적용 하기 위해 필터 오차의 공분산 행렬의 초기값 문제를 논의 하였다. 이에 대해서 ARMA 과정과 오차 간의 cross covariance 는 전달 함수 모형의impulse response 가 되며 이로 부터 필터 오차의 공분산 행렬의 초기값을 결정 할수 있음을 논의 하였다.

필터 오차 공분산 행렬의 초기값에 의해 Kalman 필터를 초기화 할 수 있으며 우도 함수를 최대화 하는 모수를 결정함에 있어 계산이 용이한 완전-자료에 근거한 우도함수의 조건부 기대값을 최대화 하는데 Kalman 필터의

결과를 이용하고, k-단계 마다 모수들을 갱신 하여 궁극적으로 불완전-자료에 근거한 우도 함수를 재 평가 하여 그 우도 함수가 최대로 하는 모수 추정 문제를 논의 하였으며 Kalman 반복과 EM 알고리즘에 의한 최우 추정에 의해 쉽게 ARMA 시계열 모형을 적합할 수 있음을 알 수 있다.

참 고 문 헌

[1] ALI, M. M. (1977). Analysis of autoressive-moving average models:Estimation and prediction. Biomerika, 64, 535-545.

[2] AOKI, M. (1987). State Space Modeling of Time Series: Heidelberg: Springer-Verlag.

[3] BOX, G. E. P & JENKINS, G. M. (1976). Time Series Analysis: Forecasting and Control. San Francisco: Holden Day.

[4] CAINES, P. E. & RISSANEN, J. (1974). Maximum likelihood estimation of parameters in multivariate Gaussian stochastic process. IEEE Trans. Info. theo. IT-20, 102-104

[5] DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc., Ser. B. 39, 1-38.

[6] GARDNER, G., HARVEY, A. C. & PHILLIPS, G. D. A. (1979). An algorithm for exact maximum likelihood estimation of autoregressive moving average models by means of Kalman filtering. Appl. Stat., 29, 311-322.

[7] GUPTA, N. K. & MEHRA, R. K. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. IEEE Trans. auto. cont. AC-19, 774-783.

[8] HARVEY, A. C. & PHILLIPS, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika, 66, 49-58.

[9] JOHNSON, R. A. & WICHERN, D. W. (1992). Applied Multivariate Statistical Analysis:Prentice Hall.

[10] PEARLMAN, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. Biomerika, 67, 232-233.

[11] PRIESTLEY, M. D. (1981). Spectral Analysis and Time Series: Academic Press, Inc.