응용통계연구 제 10권 1호
1997년 3월 pp. 121-136

공간 자기회귀모형의 식별

손 건 태[1) , 백 지 선[2)

요 약

공간자료는 공간 위치의 변화에 따라 관찰되는 자료이다. 본 논문에서는 공간자료를 가지고 행 방향, 열 방향, 대각선 방향으로 나누어 시계열의 모형 식별에서 사용되는 Box-Jenkins 방법과 식별통계량, 형태인식법을 공간 자기회귀모형에 적용하여 모형을 식별해 보고 모의실험을 통하여 식별 방법들을 비교해 보았다.

1. 서 론

공간 위치의 변화에 따라 관찰되는 공간자료(Spatial data)는 기상학, 지질학, 천문학, 생태학 등 여러 분야에서 광범위하게 나타난다. 최근에는 공간자료의 분석에 있어서 시계열에서의 모형과 개념을 적용시키는데 많은 관심을 가져왔다. 시계열에서 모형을 식별하는 방법으로는 Box-Jenkins의 자기상관함수(Autocorrelation Function), 편자기상관함수(Partial Autocorreiation Function)와 최후예측오차방법(Akaike, 1970), 아카이케 정보판단기준방법(Akaike, 1972), 베이지안 정보판단기준방법(Akaike, 1977), 반복대수법칙방법(Hannan and Quinn, 1979) 등의 식별통계량이 있다. 또한 최근에 제시된 R배열, S배열, RS배열, 모퉁이 방법, GAPC 방법, 3-배열, ESACF 방법, SCAN 방법등 형태인식을 이용하는 방법이 있다. 이 형태인식 방법들은 현존하는 식별 방법들 중 가장 뛰어나다고 평가되고 있다. 그러나 시계열에서의 모형과 개념들을 공간자료에 적용시키는 데는 많은 문제점들이 있다. 왜냐하면 시계열은 현재의 값이 과거에 의해 영향을 받는, 즉 시간의 흐름에 의한 1차원의 구조이지만 공간자료는 모든 방향에서 영향을 받는 구조로 1차원, 2차원, 3차원 또는 시간적 흐름을 포함한 4차원까지도 고려하여야 하기 때문이다. 따라서 차수의 증가에 따라 모수의 수가 기하급수적으로 증가하여 기존의 방법을 그대로 적용하기에는 어려움이 많다.

여기서 다룰 공간자료는 단일 변량의 자료이므로 본 논문에서는 이 점에 착안하여 단일 변량 시계열 모형 식별 방법으로 공간 자기회귀과정의 차수를 결정하여 모형 식별을 하고자 한다. 식별에 사용되는 모수의 수를 최소화 하기 위하여 행 방향과 열 방향 그리고 대각선 방향으로의 식별 방법을 제안하였다. 세 방향으로 나누어 식별통계량과 Choi(1988, 1992)에 의하여 제안된 3-배열 형태인식 방법을 각각 적용시켜 공간자료의 모형을 식별하고 모의실험을 통하여 식별 방법들을 비교해 보고자 한다.

---

1) (609-735) 부산시 금정구 장전동, 부산대학교 자연과학대학 통계학과 부교수(컴퓨터 및 정보통신 연구소 연구원)
2) (609-735) 부산시 금정구 장전동, 부산대학교 자연과학대학 통계학과

2. 공간 자료에 대한 모형의 식별

평면상에서 (p_1, p_2)의 차수를 가지는 공간 자기회귀과정(Spatial Autoregressive Process)을 SAR(p_1, p_2)로 나타내며, 모형식은 다음과 같다.

$$Z_{r,c} = \sum_{k_1=0}^{p_1} \sum_{k_2=0}^{p_2} \alpha_{k_1,k_2} Z_{r-k_1,c-k_2} + a_{r,c}, \qquad r, c = 0, \pm 1, \pm 2, \cdots ㄴ$$

여기서 $\alpha_{0,0} = 1$이고 $a_{r,c}$는 평균이 0이고 분산이 $\sigma_a{}^2$인 백색잡음과정이다.

2.1 자기상관함수와 편 자기상관함수의 추정

관측된 자료 $Z_{r,c}$, $r = 1, \cdots, n$, $c = 1, \cdots, m$ 이 주어진 경우 각 방향에 대한 자기 상관 함수를 다음과 같이 추정한다.

$$\widehat{\rho_r}(k) = \frac{\sum_{c=1}^{m} \sum_{t=1}^{n-k} (Z_{t,c} - \overline{Z})(Z_{t+k,c} - \overline{Z})}{\sum_{r=1}^{n} \sum_{c=1}^{m} (Z_{r,c} - \overline{Z})^2}, \qquad \widehat{\rho_c}(k) = \frac{\sum_{r=1}^{n} \sum_{t=1}^{m-k} (Z_{r,t} - \overline{Z})(Z_{r,t+k} - \overline{Z})}{\sum_{r=1}^{n} \sum_{c=1}^{m} (Z_{r,c} - \overline{Z})^2},$$

$$\widehat{\rho_d}(k) = \frac{A(M)}{\sum_{r=2}^{M} \sum_{t=0}^{n-r} (Z_{r+t,1+t} - \overline{Z})^2 + \sum_{c=1}^{M} \sum_{t=0}^{m-c} (Z_{1+t,c+t} - \overline{Z})^2}.$$

여기서 $\overline{Z}$는 $n \times m$개의 $Z_{r,c}$들의 평균이며, $A(M) = \sum_{r=2}^{M} \sum_{t=0}^{n-r-k} (Z_{r+t,1+t} - \overline{Z})(Z_{r+t+k,1+t+k} - \overline{Z}) + \sum_{c=1}^{M} \sum_{t=0}^{n-c-k} (Z_{1+t,c+t} - \overline{Z})(Z_{1+t+k,c+t+k} - \overline{Z})$이고, $\widehat{\rho_r}(k)$, $\widehat{\rho_c}(k)$, $\widehat{\rho_d}(k)$는 각각 행 방향, 열 방향, 대각선 방향의 자기상관함수이다. 대각선 방향의 자기상관함수는 $n = m$인 경우에 한하여 구하기로 한다. M는 시차 k를 고려하여 결정하는데, M을 변화시키면서 $\widehat{\rho_d}(k)$의 값의 변화를 보면 그 차이는 무시할 만하다. 추정되어진 자기상관함수는 2.3절의 3-배열의 값을 계산하는데 사용되어 진다.

2.2 식별통계량을 이용한 식별

Akaike(1974)의 AIC 함수의 유도 과정을 따라 SAR 모형의 경우에 대해 AIC를 확장시키면 다음과 같다. $Z_{1,1}, Z_{1,2}, \cdots, Z_{n_1,n_2}$을 확률밀도함수 $g(Z)$의 $n=n_1\times n_2$개의 독립인 관측치라고 하고, $f(Z \mid \boldsymbol{\alpha})$를 모수 $\boldsymbol{\alpha}$를 가지는 Z의 확률밀도함수라고 하자. $Z_{r,c}$가 주어졌을 때, 모수 $\boldsymbol{\alpha}$에 대한 정보의 측도로서 평균 로그우도함수 $\frac{1}{n}\sum_{r=1}^{n_1}\sum_{c=1}^{n_2}\ln f(Z_{r,c} \mid \boldsymbol{\alpha})$를 사용할 수 있다. Kolmogorov의 강대수법칙에 의해서 위의 평균 로그우도함수는 $\ln f(Z\mid \boldsymbol{\alpha})$가 적분 가능한 경우 n이 무한대로 증가함에 따라 거의 모든 점에서(almost everywhere) $S(g;f(\cdot\mid\boldsymbol{\alpha})) = \int g(Z)\ln f(Z\mid\boldsymbol{\alpha})dZ = E(\ln f(Z\mid\boldsymbol{\alpha}))$로 수렴하고, 그 추정치로 위의 평균 로그우도함수를 사용할 수가 있다. 또한 $g(Z)$와 $f(Z\mid\boldsymbol{\alpha})$의 판별에 대한 정보로서 Kullback-Leibler의 정보량 $I(g;f(\cdot\mid\boldsymbol{\alpha})) = S(g;g) - S(g;f(\cdot\mid\boldsymbol{\alpha}))$을 사용한다.

$g(Z)=f(Z\mid\boldsymbol{\alpha}_0)$인 경우에 $S(g;f(\cdot\mid\boldsymbol{\alpha}))\equiv S(\boldsymbol{\alpha}_0;\boldsymbol{\alpha})$, $I(g;f(\cdot\mid\boldsymbol{\alpha}))\equiv I(\boldsymbol{\alpha}_0;\boldsymbol{\alpha})$로 표현하기로 하자. $\boldsymbol{\alpha}$가 $\boldsymbol{\alpha}_0$에 충분히 가까울 때 $\boldsymbol{\alpha}=\boldsymbol{\alpha}_0+\Delta\boldsymbol{\alpha}$로 놓으면

$$I(\boldsymbol{\alpha}_0;\boldsymbol{\alpha}) = I(\boldsymbol{\alpha}_0;\boldsymbol{\alpha}_0+\Delta\boldsymbol{\alpha}) \simeq \frac{1}{2}\|\Delta\boldsymbol{\alpha}\|_J^2$$

이다. 여기서 $\|\Delta\boldsymbol{\alpha}\|_J^2 = \Delta\boldsymbol{\alpha}' J\Delta\boldsymbol{\alpha}$이고, $J_{rc} = E\{\frac{\partial}{\partial\alpha_r}\ln f(Z\mid\boldsymbol{\alpha})\frac{\partial}{\partial\alpha_c}\ln f(Z\mid\boldsymbol{\alpha})\}$는 J의 (r, c)번째 성분이다. $\widehat{\boldsymbol{\alpha}}$를 $\boldsymbol{\alpha}_0$의 MLE라고 하면 n이 무한대로 커질 때 MLE의 정규근사 성질에 의해서 $\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\rightarrow N(0,J^{-1})$이 되고, $n\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_J^2$의 차수가 k인 경우 $\chi^2(k)$가 됨을 알 수 있다. $\boldsymbol{\alpha}$는 $\boldsymbol{\alpha}_0$에 충분히 가까우므로 $n\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\|_J^2$의 분포는 n이 충분히 클 때 $\chi^2(k)$에 근사적으로 따른다고 생각할 수 있다. 따라서

$$E_\infty 2nI(\boldsymbol{\alpha}_0;\widehat{\boldsymbol{\alpha}}) = E_\infty n\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_J^2 \simeq k+n\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|_J^2 \qquad (1)$$

이 성립된다. 여기서 E_∞는 점근분포의 평균을 의미한다.

식 (1)의 두번째 항은 $n\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|_J^2 \simeq 2\{\sum_{r=1}^{n_1}\sum_{c=1}^{n_2}\ln f(Z_{r,c}\mid\boldsymbol{\alpha}_0) - \sum_{r=1}^{n_1}\sum_{c=1}^{n_2}\ln f(Z_{r,c}\mid\widehat{\boldsymbol{\alpha}})\}+k$로 근사시킬 수 있다. 그러므로 위의 근사식을 식 (1)에 대입하면 다음과 같이 된다.

$$E_{\infty}2n\,I(\boldsymbol{\alpha}_0;\ \widehat{\boldsymbol{\alpha}}) \simeq 2\{\sum_{r=1}^{n_1}\sum_{c=1}^{n_2}\ln f(Z_{r,c}|\,\boldsymbol{\alpha}_0) - \sum_{r=1}^{n_1}\sum_{c=1}^{n_2}\ln f(Z_{r,c}|\,\widehat{\boldsymbol{\alpha}})\} + 2k.$$

모형의 식별에 있어서는 여러가지 모형식을 비교하는 것이 목적이므로 위의 식에서 차수와 무관하게 공통부분이 되는 첫번째 항을 제외시키면 SAR (p_1, p_2)모형을 적합시킨 경우 AIC 함수는 다음과 같이 정의된다.

$$\mathrm{AIC}\,(p_1, p_2) = -2\sum_{r=1}^{n_1}\sum_{c=1}^{n_2}\ln f(Z_{r,c}|\,\widehat{\boldsymbol{\alpha}}) + 2\{(p_1+1)(p_2+1)-1\}$$

$$= -2\ln\{\text{ maximum likehood}\} + 2\{(p_1+1)(p_2+1)-1\}.$$

$S(\boldsymbol{\alpha}) = \sum_{r=-\infty}^{n_1}\sum_{c=-\infty}^{n_2}[E(a_{r,c}|\,\boldsymbol{\alpha}, Z)]^2$라 하면 SAR 모형에서 로그우도함수를 최대화시키는 분산의 추정치는 $\hat{\sigma}_a^{\,2} = \dfrac{S(\widehat{\boldsymbol{\alpha}})}{n}$이므로 로그우도함수는 다음과 같다.

$$\ln L = -\frac{n}{2}\ln\ \hat{\sigma}_a^{\,2} - \frac{n}{2}(1+\ln 2\pi).$$

따라서 AIC 함수는 다음과 같이 수정되어 진다.

$$\mathrm{AIC}\,(p_1, p_2) = \ln\ \hat{\sigma}_a^{\,2}(p_1, p_2) + \frac{2[(p_1+1)(p_2+1)-1]}{n}.$$

여기서 $\hat{\sigma}_a^{\,2}(p_1, p_2)$는 SAR (p_1, p_2)모형을 적합시킨 경우 백색잡음 과정의 분산에 대한 최대 우도 추정량을 의미한다. 이 함수를 최소화하는 (p_1, p_2)를 최소 AIC 추정량이라고 한다.

SAR 모형에 대해 BIC 통계량은 다음과 같이 정의되어 진다.

$$\mathrm{BIC}\,(p_1, p_2) = \ln\ \widehat{\sigma_a^2}(p_1, p_2) + \frac{[(p_1+1)(p_2+1)-1]\ln n}{n}.$$

AIC, BIC 함수를 계산하기 위해서는 $\sigma_a(p_1, p_2)$의 추정이 필요하고, $\sigma_a(p_1, p_2)$을 추정하기 위해서는 각각의 SAR 모형에서 모수의 추정이 필요하다(Ha, 1992). 모형식에서 추정되어질 모수의 개수는 $(p_1+1)(p_2+1)-1$개이다. 자기공분산함수를 다음과 같이 정의한다.

$$R(s, t) = E(Z_{r,c}Z_{r+s,c+t})$$

1차원 율-워커 방정식(Yule-Walker equation)과 유사하게 공간 율-워커 방정식을 다음과 같이 쓸 수 있다(Ha, 1992).

$$R(s, t) = \sum_{k,l\in S_p}\alpha_{kl}R(s-k, t-l), \quad s, t\geq 0,\ (s, t)\neq 0.$$

여기서 $S_p = \{(r, c)\,|\,0\leq r\leq p_1,\ \ 0\leq c\leq p_2\}$ 이다.

$Z_{r,c}$가 주어졌을 때, 자기공분산함수는 다음과 같이 추정한다.

$$\hat{R}(s,t)=\frac{1}{n}\sum_{r=1}^{n_1-s}\sum_{c=1}^{n_2-t}Z_{r,c}Z_{r+s,c+t}, \qquad s,t\geq 0$$

$$\hat{R}(s,-t)=\frac{1}{n}\sum_{r=1}^{n_1-s}\sum_{c=t+1}^{n_2}Z_{r,c}Z_{r+s,c-t}, \qquad s,t\geq 0$$

$$\hat{R}(s,t) = \hat{R}(-s,-t), \qquad \hat{R}(s,-t)=\hat{R}(-s,t).$$

이 추정량을 율-워커 추정량이라고 한다. 그러나 이 추정량은 근사적으로 편의를 가진다. 그러므로, 근사적으로 편의를 가지지 않는 아래의 최소제곱추정량 $\tilde{R}(s,t)$을 사용한다.

$$\tilde{R}(s,t) = \frac{n}{(n_1-s)(n_2-t)}\hat{R}(s,t), \qquad s,t\geq 0$$

$$\tilde{R}(-s,-t) = \tilde{R}(s,t), \qquad \tilde{R}(s,-t)=\tilde{R}(-s,t).$$

이 최소제곱추정량을 율-워커 방정식에 적용하여 모수를 추정하고, 추정되어진 모수로써 각 모형의 분산을 추정하면 다음과 같다.

$$S_*(\alpha)=\sum_{r=p_1+1}^{n_1}\sum_{c=p_2+1}^{n_2}a_{r,c}^2(\alpha), \qquad a_{r,c}=(Z_{r,c}-\hat{Z}_{r,c})$$

$$\widehat{\sigma_a^2}=\frac{S_*(\hat{\alpha})}{\text{d.f.}}, \qquad \text{d.f.}=(n_1-p_1)(n_2-p_2)-((p_1+1)(p_2+1)-1).$$

2.3 3-배열을 이용한 공간 자기회귀모형 식별

Choi(1988, 1992)는 일반화 레빈슨-더빈 알고리듬을 사용하여 세개의 배열, 즉 θ-배열, λ-배열, η-배열을 제안하였다. 일반적으로 3-배열은 자기회귀이동평균모형의 식별에 사용되어지며 식별하는 방법을 간략하게 소개하면 다음과 같다.

ARMA(p,q)의 모형식은 다음과 같이 표현된다.

$$\phi(B)Z_t=\theta(B)a_t.$$

여기서 $\phi(B)=-\phi_0-\phi_1B-\cdots-\phi_pB^p$, $\theta(B)=-\theta_0-\theta_1B-\cdots-\theta_qB^q$, $\phi_0=\theta_0=-1$, B는 $B^kZ_t=Z_{t-k}$인 후향 연산자이며, a_t는 평균이 0이고 분산이 σ^2인 백색잡음과정이다.

ARMA(p,q)의 확률과정은 다음과 같은 확장된 율-워커 방정식(Extended Yule-Walker equation ; EYW)을 만족한다.(Box and Jenkins, 1976)

$$\rho_j = \phi_1\rho_{j-1}+\phi_2\rho_{j-2}+\cdots+\phi_p\rho_{j-p}, \qquad j=q+1,q+2,\cdots$$

먼저 다음과 같은 행렬과 벡터를 정의한다.

$$B(k,i)=\begin{pmatrix} \rho_i & \rho_{i-1} & \cdots & \rho_{i-k+1} \\ \rho_{i+1} & \rho_i & \cdots & \rho_{i-k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i+k-1} & \rho_{i+k-2} & \cdots & \rho_i \end{pmatrix}$$

$$\rho(k,i) = (\rho_{i+1}, \cdots, \rho_{i+k})^t, \quad r(k,i) = (\rho_{i-k}, \cdots, \rho_{i-1})^t$$

또한 임의의 벡터 $x = (x_1, x_2, \cdots, x_n)^t$에 대하여 $\tilde{x} = (x_n, x_{n-1}, \cdots, x_1)^t$로 표기하기로 하자.

$B(k,i)$가 정칙행렬인 경우, $\phi(k,i) = B(k,i)^{-1}\rho(k,i)$, $\tilde{\pi}(k,i) = B(k,i)^{-1} r(k,i)$를 정의하고, 다음과 같은 세가지 형태의 함수를 정의한다.

$$\theta(k,i) = \rho_{i+k-1} - \tilde{\phi}(k,i)^t \rho(k,i)$$

$$\eta(k,i) = \rho_{i-k-1} - \pi(k,i)^t r(k,i)$$

$$\lambda(k,i) = \rho_i - \pi(k,i)^t \rho(k,i)$$

여기서, $\theta(0,i) = \rho_{i+1}$, $\eta(0,i) = \rho_{i-1}$, $\lambda(0,i) = \rho_i$ 이다.

$k > p$이고 $i > q$이면, $B(k,i)$가 비정칙행렬이 되므로 이 경우에는 $B(k,i)$의 일반화 역행렬을 사용하여 세가지 함수를 확장시키고, 각각에 대해 일반화된 θ-함수, η-함수, λ-함수로 명칭한다. ARMA (p,q)의 확률과정을 따르는 경우 위의 세가지 함수는 다음의 표의 형태를 나타내며, 각각의 형태를 이용하여 모형을 식별하게 된다.

표 2.1 θ-배열

$k \backslash i$	$\cdots$	$q-1$	q	$q+1$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p-1$	$\cdots$	$\theta(p-1, q-1)$	$\theta(p-1, q)$	$\theta(p-1, q+2)$	$\cdots$
p	$\cdots$	$\theta(p, q-1)$	0	0	$\cdots$
$p+1$	$\cdots$	$\theta(p+1, q-1)$	0	0	$\cdots$
$p+2$	$\cdots$	$\theta(p+2, q-1)$	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

표 2.2 η-배열

$k\backslash i$	$\cdots$	$q+1$	$q+2$	$q+3$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p-1$	$\cdots$	$\eta(p-1, q+1)$	$\eta(p-1, q+2)$	$\eta(p-1, q+3)$	$\cdots$
p	$\cdots$	$\eta(p, q+1)$	0	0	$\cdots$
$p+1$	$\cdots$	∞	$\eta(p, q+1)$	0	$\cdots$
$p+2$	$\cdots$	∞	∞	$\eta(p, q+1)$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

표 2.3 λ-배열

$k\backslash i$	$\cdots$	q	$q+1$	$q+2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p-1$	$\cdots$	$\lambda(p-1, q)$	$\lambda(p-1, q+1)$	$\lambda(p-1, q+2)$	$\cdots$
p	$\cdots$	$\lambda(p, q)$	0	0	$\cdots$
$p+1$	$\cdots$	$\lambda(p, q)$	X	0	$\cdots$
$p+2$	$\cdots$	$\lambda(p, q)$	X	X	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

(X는 0이 아님을 나타냄)

이제 공간자료 SAR(p_1, p_2)에 대하여 3-배열을 적용해 보면 다음과 같다. 각각의 방향에 대해 3-배열은 자기상관함수에 따라 값이 변할 뿐, 방향에 따른 3-배열을 구하는 방법에는 차이가 없다. 또한 η-배열의 경우 대각선 방향으로 형태가 나타나므로 본 논문의 경우에 적용할 수 가 없다. 그러므로 본 논문에서는 행 방향의 자기상관함수를 가지고 θ-배열과 λ-배열에 적용하여 p_1의 차수를 결정하는 방법을 소개한다. 이동평균모형을 고려하지 않은 $AR(p_1)$의 확률과정은 다음과 같은 율-워커 방정식을 만족한다.(Box and Jenkins, 1976).

$$\sigma(k) = \phi_1 \sigma(k-1) + \cdots + \phi_p \sigma(k-p), \qquad k > 0.$$

3-배열을 구하기 위한 자기상관함수 행렬과 열 벡터들은 다음과 같다.

$$B(k,0)=\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{-k+1} \\ \rho_{-1} & \rho_0 & \cdots & \rho_{-k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_0 \end{pmatrix}$$

$$\rho(k,0)=(\rho_1,\cdots,\rho_k)^t\ ,\quad \gamma(k,0)=(\rho_{-k},\cdots,\rho_{-1})^t$$
$$\phi(k,0)=B(k,0)^{-1}\rho(k,0),\quad \pi(k,0)=B(k,0)^{-1}\gamma(k,0).$$

그러므로 θ-배열은

$$\theta(k,0)=\rho_{k+1}-\tilde{\phi}(k,0)^t\rho(k,0)\text{이며,}$$

λ-배열은

$$\lambda(k,0)=\rho_0-\pi(k,0)^t\rho(k,0)\text{이 된다.}$$

위의 θ-배열과 λ-배열은 다음의 표와 같은 형태를 나타낸다.

표 2.4 공간자료에서 나타나는 θ-배열과 λ-배열

k	$\theta(k,0)$	$\lambda(k,0)$
$\vdots$	$\vdots$	$\vdots$
p_1-1	$\theta(p_1-1,0)$	$\lambda(p_1-1,0)$
p_1	0	$\lambda(p_1,0)$
p_1+1	0	$\lambda(p_1,0)$
$\vdots$	$\vdots$	$\vdots$

3. 모의실험

식별통계량과 3-배열 형태인식법을 활용하여 본 논문에서 제안한 방법의 식별력을 조사하기 위하여, 주어진 구조에 의해 모의자료를 생성하고 제안한 방법들이 올바로 식별하는 지를 알아보고자 한다. 주어진 각각의 모형에 따라 10×10, 20×20, 30×30개의 모의자료를 생성하여

모형식별에 이용하기로 한다. 이 때 생성되는 모의자료는 초기치의 분포가 $U(0,1)$이며, 백색잡음 $a_{r,c}$의 분포는 $N(0,1)$이다.

3.1 모의실험을 위한 모형구조식

(1) SAR(1,1) : $Z_{r,c}= 0.2Z_{r-1,c}+0.5Z_{r,c-1}+0.3Z_{r-1,c-1}+a_{r,c}$

(2) SAR(2,3) :

$$\begin{aligned} Z_{r,c}= &\ 0.3Z_{r-1,c}-0.51_{Z}r-2,c+0.06Z_{r,c-1}+0.01Z_{r-1,c-1} \\ &+0.31Z_{r-2,c-1}-0.34Z_{r,c-2}+0.13Z_{r-1,c-2}-0.023Z_{r-2,c-2} \\ &+0.33Z_{r,c-3}-0.11Z_{r-1,c-3}+0.21Z_{r-2,c-3}+a_{r,c} \end{aligned}$$

(3) SAR(3,3) :

$$\begin{aligned} Z_{r,c}= &\ 0.3Z_{r-1,c}-0.51Z_{r-2,c}+0.2Z_{r-3,c}+0.06Z_{r,c-1} \\ &+0.01Z_{r-1,c-1}+0.31Z_{r-2,c-1}-0.05Z_{r-3,c-1}-0.34Z_{r,c-2} \\ &+0.13Z_{r-1,c-2}-0.023Z_{r-2,c-2}-0.001Z_{r-3,c-2}+0.33Z_{r,c-3} \\ &-0.11Z_{r-1,c-3}+0.21Z_{r-2,c-3}+0.04Z_{r-3,c-3}+a_{r,c} \end{aligned}$$

3.2 모형식별

10×10개의 모의실험에 대한 결과는 표 3.1에서 표 3.3에 걸쳐 수록되어져 있으며, 20×20개인 경우는 표 3.4에서 표 3.6에, 그리고 30×30개의 경우는 표 3.7에서 표 3.9에 수록되어 있다. 먼저 10×10개의 모의실험 결과를 살펴보면, 식별통계량의 경우 SAR(1,1) 모형에 대해서는 모형을 잘 식별하고 있으나 SAR(2,3) 모형과 SAR(3,3) 모형에 대해서는 과소식별을 하는 경향을 보이고 있다. 3-배열을 이용한 경우에도 SAR(2,3) 모형에서는 제대로 식별하고 있으나 나머지 모형에 대해서는 과대식별 경향을 보이고 있다. 이는 모형식별을 하는데 필요한 자료의 개수가 너무 적은데 원인이 있는 것으로 판단된다. 시계열의 경우에도 모형을 식별하기 위해서는 어느 정도의 자료의 개수가 필요하다. 20×20개의 경우에는 식별통계량은 SAR(3,3) 모형에서만 SAR(2,3) 모형으로 과소식별을 하고 있다. 3-배열을 경우 SAR(3,3) 모형에서 모호한 점이 있지만 모수절약의 원칙에서 식별을 한다면 SAR(3,3)으로 판단할 수 있을 것이다. 30×30개의 경우에는 각 모형에 대해서 식별을 잘 하고 있는 것으로 판단되어 진다.

이 외에도 백색잡음 $a_{r,c}$의 분포를 $N(0,10^2)$으로 하여 식별통계량을 가지고 모의실험을 수행하여 보았다. 본 논문에 그 결과를 수록하지는 않았지만 제안한 방법들은 모형을 잘 식별하는 것으로 나타났다.

표 3.1 SAR(1,1)모형의 식별 (10×10)

시차	형태 인식법						식별통계량		
	행방향		열방향		대각선방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$	$\theta(p_1=p_2)$	$\lambda(p_1=p_2)$			
1	-.12232	.92288	.01895	.97003	.12249	.95074	(1,1)	.00876	.08692
2	-.22745	.90667	-.10790	.96966	-.08923	.93496	(1,2)	.01327	.14352
3	-.04838	.84962	-.04654	.95765	.03468	.92644	(1,3)	.08979	.27216
4	.05095	.84686	.08465	.95539	.08508	.92514	(2,1)	.16325	.29351
5	-.09535	.84380	-.02772	.94789	-.04794	.91732	(2,2)	.19223	.40064
							(2,3)	.42104	.70760
							(3,1)	.20701	.38937
							(3,2)	.31841	.60498
							(3,3)	1.43541	1.82618

표 3.2 SAR(2,3)모형의 식별 (10×10)

시차	형태 인식법				식별통계량		
	행방향		열방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$			
1	-.29917	.88688	-.21002	.99691	(1,1)	.01586	.09402
2	.08150	.78596	.28619	.95266	(1,2)	.13853	.26879
3	.03129	.77751	.02678	.86669	(1,3)	.24429	.42666
4	.00963	.77625	.08430	.86586	(2,1)	-.02511	.10515
5	-.00027	.77613	-.05053	.85765	(2,2)	.15531	.36373
					(2,3)	.34302	.62959
					(3,1)	.05469	.23705
					(3,2)	.11094	.39751
					(3,3)	.37845	.76922

표 3.3 SAR(3,3)모형의 식별 (10×10)

형태 인식법							식별통계량		
시차	행방향		열방향		대각선방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$	$\theta(p_1=p_2)$	$\lambda(p_1=p_2)$			
1	-.19072	.97977	-.34949	.99138	-.03692	.91555	(1,1)	-.04944	.02872
2	.12307	94265	.06453	.86817	-.05639	.91406	(1,2)	.06575	.19600
3	-.12829	.92658	.05927	.86337	.04897	.91058	(1,3)	.17184	.35420
4	-.05919	.90882	.08842	.85930	-.04499	.90795	(2,1)	.01725	.14751
5	-.00794	.90496	.07088	.85021	.01572	.90572	(2,2)	.16518	.37360
							(2,3)	.33087	.61744
							(3,1)	.12023	.30259
							(3,2)	.25836	.54492
							(3,3)	.49652	.88729

표 3.4 SAR(1,1)모형의 식별 (20×20)

형태 인식법							식별통계량		
시차	행방향		열방향		대각선방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$	$\theta(p_1=p_2)$	$\lambda(p_1=p_2)$			
1	-.00559	.66714	-.00016	.53133	.06632	.51550	(1,1)	.05067	.08060
2	.00098	.66710	.02427	.53133	-.01466	.50697	(1,2)	.08617	.13606
3	.03254	.66709	-.06515	.53022	-.01092	.50654	(1,3)	.09189	.16174
4	-.02290	.66551	-.00816	.52222	.02758	.50631	(2,1)	.07298	.12287
5	.01197	.66472	.00836	.52209	.00721	.50481	(2,2)	.09916	.17899
6	-.09277	.66450	.00901	.52196	.00400	.50470	(2,3)	.10490	.21466
7	-.04443	.65155	-.03199	.52180	.02194	.50467	(3,1)	.07833	.14818
							(3,2)	.11182	.22159
							(3,3)	.11951	.26919

표 3.5 SAR(2,3)모형의 식별 (20×20)

형태 인식법					식별통계량		
시차	행방향		열방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$			
1	-.66359	.94867	-.56596	.96359	(1,1)	.81235	.84228
2	-.01358	.48450	.14508	.63118	(1,2)	.49289	.54279
3	-.06406	.48412	.06331	.59783	(1,3)	1.59712	1.66697
4	-.01131	.47564	-.01860	.59113	(2,1)	.25547	.30536
5	.07850	.47537	-.00984	.59054	(2,2)	.14767	.22750
6	-.01433	.46241	-.05182	.59038	(2,3)	.08889	.19865
7	.01036	.46197	.01342	.58583	(3,1)	.32202	.39187
					(3,2)	.16118	.27094
					(3,3)	.11579	.26547

표 3.6 SAR(3,3)모형의 식별 (20×20)

형태 인식법							식별통계량		
시차	행방향		열방향		대각선방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$	$\theta(p_1=p_2)$	$\lambda(p_1=p_2)$			
1	-.56211	.99979	-.53279	.98371	.33668	.99199	(1,1)	.74596	.77590
2	.10760	.68376	.17004	.69515	.07467	.87772	(1,2)	.46163	.51152
3	-.04667	.66683	.05386	.65356	-.27874	.87137	(1,3)	.39058	.46043
4	.09130	.66356	-.02099	.64912	.01458	.78221	(2,1)	.28576	.33566
5	.01987	.65100	.00222	.64844	-.01648	.78193	(2,2)	.17627	.25610
6	-.04048	.65039	-.05933	.64843	.08058	.78159	(2,3)	.10161	.21137
7	-.03312	.64787	.01940	.64300	-.03370	.77328	(3,1)	.26017	.33002
							(3,2)	.17518	.28495
							(3,3)	.11818	.26786

표 3.7 SAR(1,1)모형의 식별 (30×30)

형태 인식법							식별통계량		
시차	행방향		열방향		대각선방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$	$\theta(p_1=p_2)$	$\lambda(p_1=p_2)$			
1	.07308	.65213	−.02746	.29243	.00330	.24397	(1,1)	.05938	.07539
2	.03049	.64394	−.02015	.28985	.00167	.24392	(1,2)	.07409	.10077
3	.02061	.64250	−.01684	.28845	−.00160	.24391	(1,3)	.09128	.12863
4	−.02576	.64184	−.01578	.28747	−.00097	.24390	(2,1)	.06478	.09146
5	.01210	.64080	−.01015	.28660	−.00402	.24390	(2,2)	.07994	.12263
6	.04613	.64057	−.00731	.28624	−.00495	.24383	(2,3)	.09897	.15766
7	.03450	.63725	.00004	.28606	.00158	.24373	(3,1)	.06982	.10717
							(3,2)	.08903	.14773
							(3,3)	.11041	.19045

표 3.8 SAR(2,3)모형의 식별 (30×30)

형태 인식법					식별통계량		
시차	행방향		열방향		(p_1, p_2)	AIC	BIC
	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$			
1	−.58225	.97847	−.54395	.97165	(1,1)	.86134	.87735
2	−.01849	.63200	−.21496	.66714	(1,2)	.53784	.56451
3	−.01181	.63146	−.01831	.59788	(1,3)	.44024	.47760
4	.02478	.63124	.02974	.59732	(2,1)	.32463	.35131
5	.02923	.63026	.00348	.59584	(2,2)	.17765	.22034
6	.00492	.62891	−.02647	.59582	(2,3)	.10284	.16153
7	.00964	.62887	.00709	.59464	(3,1)	.32431	.36166
					(3,2)	.18219	.24088
					(3,3)	.11963	.19967

표 3.9 SAR(3,3)모형의 식별 (30×30)

형태 인식법							식별통계량		
	행방향		열방향		대각선방향				
시차	$\theta(p_1)$	$\lambda(p_1)$	$\theta(p_2)$	$\lambda(p_2)$	$\theta(p_1=p_2)$	$\lambda(p_1=p_2)$	(p_1, p_2)	AIC	BIC
1	-.83991	.98924	.75697	.89128	.22237	.79783	(1,1)	1.19425	1.21026
2	.05410	.27612	.06958	.24837	.43641	.73585	(1,2)	.62004	.64672
3	-.02527	.26552	-.00382	.22890	-.16874	.47703	(1,3)	.48513	.52248
4	.02262	.26311	.03658	.22884	-.00267	.41734	(2,1)	.48322	.50990
5	.02433	.26117	-.00629	.22299	-.09703	.41733	(2,2)	.25881	.30150
6	-.01717	.25890	-.01819	.22281	-.04203	.39477	(2,3)	.17208	.23078
7	-.01017	.25776	.01673	.22133	.05092	.39029	(3,1)	.39615	.43351
							(3,2)	.21855	.27725
							(3,3)	.14831	.22835

4. 결 론

공간자료는 시계열과는 다르게 모든 방향에서 영향을 받는다. 그러므로 이전의 시계열의 모형식별 방법을 공간자료의 모형식에 적용시키는 데에는 많은 문제점 들이 있다. 이러한 점들을 고려하여 본 논문에서는 많은 모수를 포함한 모형식을 생각하지 않고 행 방향, 열 방향, 대각선 방향의 세가지 방향으로 형태인식을 통한 모형식별의 한 예인 3-배열을 이용하여 모형을 식별하고, 모든 모수를 포함한 모형식에서 식별통계량을 이용하여 모형을 식별하여 보았다. 모의실험의 결과는 모형의 개수가 적은 경우(10×10인 경우)에는 제안한 방법 모두가 모형식별에 있어서 미흡한 것으로 나타났고, 20×20개와 30×30개의 경우에는 비교적 모형식별을 잘하는 것으로 나타났다. 그러나 시계열을 분석하는 경우 어느 정도의 자료의 개수가 필요하듯이 공간자료의 경우에도 어느 정도의 자료의 개수가 필요하다는 것을 고려한다면 제안한 두가지 방법은 공간자료의 모형식별에도 좋은 통계량이라고 생각되어 진다. 두가지 방법을 비교해 보면, 식별통계량에 의한 식별은 가장 작은 값을 갖는 모형을 선택하면 되므로 식별이 간편한 반면 모든 가능한 차수에 대해 각각의 통계량의 값을 구해야 하므로 복잡하고 많은 시간이 요구된다. 그리고 형태인식에 의한 모형식별은 동시에 모든 시차에 대한 배열값이 구하여 지므로 계산이 간편한 반면 개인에 따라 배열의 형태를 보는 과정에서 다소 주관적인 판단이 개입되어 지는 경향이 있다. 그러므로 모형의 식별에 있어서 한가지 식별 방법만을 사용하는 것보다 식별통계량과 3-배열을 병행하여 사용하는 것이 더 효과적일 것이다.

참 고 문 헌

[1] EunHo Ha (1992). Analysis of Spatial Autoregressive Processes and Rain Rate Estimation, Unpublished, Ph.D. Dissertation, Tesas A & M University.

[2] Akaike, H (1974). A new look at the statistical model indentification, *IEEE Transactions on Automatic Control*, AC-19, 716-723.

[3] ByoungSeon Choi (1988). Three pattern for the order determination of the maximum entropy spectrum, Unpublished.

[4] ByoungSeon Choi (1990). On the asymptotic distribution of the generalized partial autocorrelation function in autoregressive moving-average processes, *Journal of time series analysis*, Vol. 12, No. 3.

[5] ByoungSeon Choi (1992). *ARMA(p,q) Model Identification*, Springer-Verlag, New-York.

[6] ByoungSeon Choi (1993). Two Chi-Square Statistics for Determining the Orders p and q of an ARMA(p,q) Process, *IEEE Transactions on Signal Processing*, Vol. 41, No. 6.

[7] Box, G. E. P. and G. M. jenkins (1976). *Time Series Analysis, Forcasting and Control (Revised Edition)*, Holden-Day, San Francisco.

Model Identification of Spatial Autoregressive Data Analysis

Keon-Tae Sohn[3], Jee-Seon Baik[4]

Abstract

Spatial data is collected on a regular Cartesian lattice. In this paper we consider the model indentification of spatial autoregressive(SAR) models using AIC, BIC, pattern method. The propesed methods are considered as an application of AIC, BIC, 3-patterns for SAR models through three directions; row, column and diagonal directions. Using the Monte Carlo simulation, we test the efficiency of the proposed methods for various SAR models.

3) Associate Professor, Department of Statistics, Pusan National University, Pusan, 609-735, Korea.
4) Department of Statistics, Pusan National University, Pusan, 609-735, Korea.