응용통계연구 제 **20**권 *1*호, *2007*년, *pp. 103–115*

안장점근사를 이용한 자기회귀계수에 대한 소표본 점근추론*

나종화[1)] 김정숙[2)]

요 약

본 논문에서는 1차 자기회귀모형에서 자기회귀계수에 대한 여러 가지 추정량들의 분포함수에 대한 근사 방법에 대해 연구하였다. 자기회귀계수의 여러 추정량들을 이차형식의 관점에서 이해하고, Na와 Kim(2005)에 의한 안장점근사의 결과를 이용한 새로운 근사법을 제시하였다. 이 방법은 정규근사를 비롯한 기존의 근사법과는 달리 추정량에 대한 근사분포의 유도과정이 불필요하며, 소표본은 물론 통계적 추론의 주요 관심영역에서의 근사정도가 매우 뛰어난 장점을 가지고 있다. 모의실험을 통해 Edgeworth근사를 비롯한 기존의 여러 근사법보다 효율이 뛰어남을 확인하였다.

주요용어: 자기회귀계수, Edgeworth 근사, 안장점근사, 이차형식.

1. 서론

다음과 같이 정의되는 1차 자기회귀모형(first-order autoregressive model)을 생각하자.

$$X_t = \alpha X_{t-1} + e_t, \quad t = 1, \ldots, n. \tag{1.1}$$

여기서 모든 t에 대해 $|\alpha| < 1$이고, e_t 는 평균이 0 이고 분산이 σ^2이며 서로 독립인 정규분포를 따른다. 또한 X_0는 평균이 0이고 분산이 $\sigma^2/(1-\alpha^2)$인 정규분포를 가정한다. 식 (1.1)의 자기회귀계수 α에 대한 최소제곱추정량은 다음과 같다.

$$\widehat{\alpha} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_t^2}. \tag{1.2}$$

보통 자기회귀계수 α에 대한 추론은 $\hat{\alpha}$의 점근분포가 충분히 큰 n에 대해, 근사적으로 평균이 α이고, 분산이 $(1-\alpha^2)/n$인 정규분포를 따른다는 사실에 기초하고 있다. 그러나 정규분포에 기초한 추론은 매우 큰 표본의 크기를 요구하고 있어, 실제 중 · 소표본의 자료에 대해 큰 제한으로 작용한다. 이를 개선하기 위한 많은 연구가 진행되어 왔으며, 특히

* 이 논문은 2006학년도 충북대학교 학술연구지원사업의 연구비지원에 의하여 연구되었음.
1) (361-763) 청주시 흥덕구 개신동 12, 충북대학교 정보통계학과 & 기초과학연구소, 교수
E-mail: cherin@cbnu.ac.kr
2) (121-749) 서울시 서초구 서초3동 1586-9, 건강보험심사평가원 정보통신실, 책임연구원(Ph.D.)
E-mail: chastity@hiramail.net

Sargan(1976)과 Phillips(1977, 1978)는 $\hat{\alpha}$ 의 분포함수에 대한 Edgeworth전개를 통한 근사법을 제안하여, 정규근사의 결과를 부분적으로 개선하였으나 자기회귀계수의 값이 큰 경우에는 근사의 정확도가 여전히 만족스럽지 못한 결과를 보여주고 있다. 한편 자기회귀계수와 관련하여 Daniels(1956)는 다음의 변형된 추정량

$$\hat{\alpha}^* = \frac{X_1X_2 + \cdots + X_{n-1}X_n}{\frac{1}{2}X_1^2 + X_2^2 + \cdots + X_{n-1}^2 + \frac{1}{2}X_n^2} \tag{1.3}$$

의 밀도함수(density function)에 대한 효과적인 근사로 안장점근사(saddlepoint approximation)를 적용한 바 있다. 이 근사는 대단히 정확한 결과를 제공하며, 식 (1.2) 의 최소제곱추정량에도 적용이 가능하다. 그러나 자기회귀계수의 추론에 필요한 꼬리확률에 대한 결과를 얻기 위해서는 밀도함수에 대한 근사식을 다시 수치적분 해야 하는 어려움이 남아있어, 통계적 추론의 목적에는 그다지 효율적인 방법이라 말할 수 없다. 또한 이 방법은 밀도함수에 대한 근사식의 유도 과정에 Fourier 역변환을 통한 고전적인 안장점 근사의 유도과정이 요구되며, 이 과정을 변형된 여러가지 추정량들에 적용하기에는 매우 번거로운 전개를 필요로 한다. 본 연구에서는 식 (1.2)와 (1.3) 등의 여러 추정량들을 이차형식 통계량의 구조로 재표현하여 이해함으로서, 이차형식에 대한 안장점근사와 관련된 Na와 Kim(2005)의 결과를 이용하는 새로운 근사법을 제시하고자 한다.

일반적으로 안장점근사는 관심 통계량에 대한 정규근사나 Edgeworth근사 보다 더욱 정확한 근사 결과를 제공한다. 특히 안장점근사는 소표본의 경우에도 근사의 정도를 잃지 않으며, 통계적 추론에서의 주요 관심 영역인 꼬리부분(tail area)에서도 근사의 정확도가 매우 뛰어난 장점을 가지고 있어 정밀한 추론이 요구되는 많은 응용문제에 기존의 여러 근사법에 대한 보다 개선된 대안으로 활용되고 있다. 다만 안장점근사의 적용을 위해서는 제시된 통계량의 누율생성함수의 유도과정이 필요하며, 수치해를 비롯한 계산상의 어려움 등으로 인해 iid 표본평균을 비롯한 매우 제한된 형태의 통계량을 중심으로 발전되어 왔다. 특히 본 논문에서 다루게 되는 자기회귀계수의 경우와 같이 non−iid 표본에 기초한 통계량에 대해서는 안장점근사와 관련된 연구가 매우 미비하다. 본 논문에서는 non−iid 확률변수들의 함수로 표현되는 자기회귀계수에 대한 여러 추정량들을 이차형식의 통계량 구조로 이해하고, Na와 Kim(2005)에 의한 안장점근사의 결과를 이용하여, 이들 추정량들의 분포함수에 대한 근사를 수행하고자 한다. 2절에서는 이차형식에 대한 안장점근사의 결과를 소개하고, 3절에서는 자기회귀계수 추정량들의 분포함수에 대한 안장점근사의 적용과정을 구체적으로 다루었으며, 4절에서는 모의실험을 통해 이 근사법의 정확도 및 기존의 방법과의 비교분석을 수행하였다.

2. 이차형식에 대한 안장점근사

$X = (X_1, \ldots, X_n)'$를 평균이 $\mu = (\mu_1, \ldots, \mu_n)'$이고 공분산행렬이 Σ인 다변량 정규분포를 따르는 확률벡터라 하자. 이 때, 이차형식 $Q(X)$는 다음과 같이 정의된다.

$$Q(X) = \sum_{1 \leqq i \leqq j \leqq n} a_{ij} X_i X_j = X'AX. \tag{2.1}$$

위 식에서 $A = (a_{ij})_{n \times n}$ 는 일반성을 잃지 않고 대칭(symmetric)으로 가정할 수 있다. 이차형식 $Q(X)$의 누율생성함수를 $K(t)$라 할 때, Lugannani와 Rice(1980)의 분포함수에 대한 안장점근사식은 다음과 같다(Daniels, 1987).

$$\Pr\{Q(X) \leqq q\} \simeq \begin{cases} \Phi(w) + \phi(w)\{\dfrac{1}{w} - \dfrac{1}{\zeta}\}, & q \neq E(Q(X)) \\ \dfrac{1}{2} - \dfrac{K^{(3)}(0)}{6\sqrt{2\pi(K''(0))^3}}, & q = E(Q(X)). \end{cases} \tag{2.2}$$

위 식에서 $\Phi(\cdot)$와 $\phi(\cdot)$는 표준정규분포의 분포함수와 밀도함수를 나타내며, ω와ζ는 다음과 같이 정의된다.

$$\omega = \mathrm{sgn}(t_0)[2n\{t_0 q - K(t_0)\}]^{1/2}, \quad \zeta = t_0\{nK''(t_0)\}^{1/2}.$$

여기서 안장점(saddlepoint)으로 불리는 t_0는 다음의 안장점방정식(saddlepoint equation)에 대한 수치해를 통해 구해진다.

$$K'(t_0) = q.$$

다음으로 분포함수에 대한 Jensen(1992, 1995)의 안장점근사식은 다음과 같다. 이 근사식은 적절한 가정하에서 식 (2.2)와 근사의 정도(precision)가 동일한 것으로 알려져 있다. 또한 근사식 (2.2)는 분포함수의 근사값으로 음의 값을 취할 수도 있는 구조를 가진 반면, 아래의 근사식은 항상 비음(non-negative)의 값으로 분포함수를 추정하는 보다 바람직한 성질을 가지고 있다.

$$Pr\{Q(X) \leqq q\} \simeq \Phi\left\{\omega + \frac{1}{\omega}\log\left(\frac{\zeta}{\omega}\right)\right\}. \tag{2.3}$$

위의 근사식을 적용하기 위해서는 이차형식 $Q(X)$의 누율생성함수 및 이의 1, 2차 미분식의 형태가 요구된다. 다음의 정리 2.1과 보조정리 2.1은 이와 관련된 결과들을 정리한 것이다.

먼저 콜레스키 분해를 통해 양정치 공분산행렬(Σ)을 $\Sigma = \Gamma\Gamma'$ 으로 표현하자. 여기서 Γ는 비정칙 하삼각행렬이다. 또한 $B = \Gamma' A \Gamma$이라 하고, 행렬 B의 고유치 $\lambda_i (i = 1, \ldots, n)$들로 구성된 대각행렬을 $\Lambda = \mathrm{diag}(\lambda_i)$라고 하면, $B = P\Lambda P'$ 으로 표현된다. 여기서 P는 $P'P = I$를 만족하는 고유벡터들로 구성된 직교행렬이다.

정리 2.1 이차형식 $Q(X) = X'AX$의 누율생성함수는 다음과 같이 주어진다.

$$\begin{aligned} K_Q(t) &= -\frac{1}{2}\sum_{i=1}^{n}\log(1-2t\lambda_i) + \frac{1}{2}\sum_{i=1}^{n}\frac{\delta_i^2\lambda_i t}{(1-2t\lambda_i)} \\ &= -\frac{1}{2}\sum_{i=1}^{n}\log(1-2t\lambda_i) + \frac{1}{2}\sum_{i=1}^{n}\frac{\delta_i^2}{(1-2t\lambda_i)} - \frac{1}{2}\sum_{i=1}^{n}\delta_i^2. \end{aligned}$$

여기서 $\delta = P \cdot \Gamma^{-1} \cdot \mu$ 으로 정의된다.

증명: 일반화 이차형식 $Q_L(X) = X'AX + b'X + c$ 의 누율생성함수는 다음과 같이 주어진다(Na와 Kim, 2005).

$$\begin{aligned} K_{Q_L}(t) &= -\frac{1}{2}\log|I-2tB| + \frac{1}{2}t^2\nu'(I-2tB)^{-1}\nu + tu, \\ B &= \Gamma' A,\ \nu = \Gamma'(b+2A\mu),\quad \mu = \mu' A\mu + b'\mu + c. \end{aligned}$$

위 식에서 첫째항은

$$\begin{aligned} &-\frac{1}{2}\log|I-2tB| = -\frac{1}{2}\log|P(I-2t\Lambda)P'| = -\frac{1}{2}\log|I-2t\Lambda| \\ &= -\frac{1}{2}\log[\prod_{i=1}^{n}(1-2t\lambda_i)] = -\frac{1}{2}\sum_{i=1}^{n}\log(1-2t\lambda_i) \end{aligned}$$

이고, 두 번째 항은

$$\begin{aligned} &\frac{1}{2}t^2 v'(I-2tB)^{-1}v \\ &= \frac{1}{2}t^2(2\cdot\Gamma A\mu)'(I-2tB)^{-1}(2\cdot\Gamma A\mu) \\ &= \frac{1}{2}(2\mu'\Gamma^{-1}B)(I-2tB)^{-1}(2B\Gamma^{-1}\mu), \qquad B = \Gamma' A\Gamma \\ &= \frac{1}{2}(\mu'\Gamma^{-1}(2tB))(I-2tB)^{-1}((2tB)\Gamma^{-1}\mu) \qquad (A) \\ &= \frac{1}{2}\mu'\Gamma^{-1}[(I-2tB)^{-1} - 2tB - I]\Gamma^{-1}\mu \qquad (B) \\ &= \frac{1}{2}\mu'\Gamma^{-1}(I-2tB)^{-1}\Gamma^{-1}\mu - \mu'\Gamma^{-1}(tB)\Gamma^{-1}\mu - \frac{1}{2}\mu'\Gamma^{-1}\Gamma^{-1}\mu \\ &= \frac{1}{2}\mu'\Gamma^{-1}P\ \mathrm{diag}\left(\frac{1}{1-2t\lambda_i}\right)P'\Gamma^{-1}\mu - t\mu' A\mu - \frac{1}{2}\mu'\Gamma^{-1}P'P\Gamma^{-1}\mu \\ &= \frac{1}{2}\sum_{i=1}^{n}\frac{\delta_i^2}{1-2t\lambda_i} - \frac{1}{2}\sum_{i=1}^{n}\delta_i^2 - t\mu' A\mu, \quad \delta = P\cdot\Gamma^{-1}\cdot\mu. \end{aligned}$$

으로 표현된다. 위 식의 전개과정에서 식 (A)에서 식 (B)로 표현되는 과정은 다음의 결과가 성립하기 때문이다.

$$\begin{aligned} (I-2tB)^{-1} &= I + (2tB) + (2tB)^2 + \cdots, \\ 2tB(I-2tB)^{-1} &= 2tB + (2tB)^2 + 2tB^3 + \cdots \\ &= (I-2tB)^{-1} - I, \\ (2tB)(I-2tB)^{-1}(2tB) &= (I-2tB)^{-1} - 2tB - I. \end{aligned}$$

따라서 이차형식 $Q(X)$에 대한 누율생성함수는 위 식에서 $b = c = 0$인 경우에 해당하며 다음의 식으로 주어진다.

$$K_Q(t) = -\frac{1}{2}\sum_{i=1}^{n}\log(1-2t\lambda_i) + \frac{1}{2}\sum_{i=1}^{n}\frac{\delta_i^2}{(1-2t\lambda_i)} - \frac{1}{2}\sum_{i=1}^{n}\delta_i^2.$$

□

위의 증명은 보다 단순한 과정을 통해 보여질 수도 있으나(Kuonen, 1999), 본 논문에서는 일반화 이차형식에 대한 Na와 Kim(2005)의 결과에서 $b = c = 0$인 특별한 경우가 Kuonen(1999)의 결과가 됨을 보이고 있다.

보조정리 2.1 이차형식 $Q(X)$의 1, 2차 미분식은 다음과 같이 주어진다.

$$\begin{aligned} K_Q'(t) &= \frac{d}{dt}K_Q(t) = \sum_{i=1}^{n}\frac{\delta_i^2\lambda_i}{(1-2t\lambda_i)^2} + \sum_{i=1}^{n}\frac{\lambda_i}{1-2t\lambda_i}. \\ K_Q''(t) &= \frac{d^2}{dt^2}K_Q(t) = \sum_{i=1}^{n}\frac{4\delta_i^2\lambda_i^2}{(1-2t\lambda_i)^3} + \sum_{i=1}^{n}\frac{2\lambda_i^2}{(1-2t\lambda_i)^2}. \end{aligned}$$

증명: 정리 2.1로부터 쉽게 유도될 수 있다. □

3. 자기회귀계수에의 응용

3.1. 평균이 알려진 경우

식 (1.1)로 부터 $X = (X_1, \ldots, X_n)'$는 평균이 $\mu = (0, \ldots, 0)'$이고 공분산행렬이

$$\Sigma = \frac{\sigma^2}{1-a^2}\begin{pmatrix} 1 & a & a^2 & \cdots & a^{n-1} \\ a & 1 & a & \cdots & a^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a^{n-1} & \cdots & \cdots & \cdots & 1 \end{pmatrix} \tag{3.1}$$

인 다변량 정규분포를 따른다. 또한 본 논문의 주된 관심사인 식 (1.2) 통계량의 분포 함수는

$$\begin{aligned} \Pr\ (\widehat{a} \le q) &= \Pr\ (V_n \le 0), \\ V_n &= \sum_{t=2}^{n} X_t X_{t-1} - q\sum_{t=1}^{n} X_t^2 \end{aligned} \tag{3.2}$$

으로 표현되며, 위 식에서 통계량 V_n은 다음과 같이 이차형식의 통계량으로 재표현될 수 있다.

$$V_n = \Big(X_1, \ldots, X_n\Big) \begin{pmatrix} -q & 1/2 & 0 & 0 & 0 & \cdots & 0 \\ 1/2 & -q & 1/2 & 0 & 0 & \cdots & 0 \\ 0 & 1/2 & -q & 1/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1/2 & -q & 1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1/2 & -q \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \tag{3.3}$$

$$= X'AX.$$

이제 식 (3.2)에 대해 앞 절에서 소개한 이차형식에 대한 안장점근사의 결과를 적용할 수 있다. 이 과정은 식 (1.3) 및 다음과 같이 표현되는 Durbin(1980)의 추정량

$$\widehat{\alpha}' = \frac{\frac{1}{2}X_1^2 + X_1X_2 + \cdots + X_{n-1}X_n + \frac{1}{2}X_n^2}{X_1^2 + \cdots + X_n^2} \tag{3.4}$$

에 대해서도 동일하게 적용될 수 있다. 예를 들어, 식 (1.3)의 변형된 추정량 $\hat{\alpha}^*$의 분포함수는

$$\begin{aligned} Pr(\hat{\alpha}^* \le q) &= Pr(V_n^* \le 0), \\ V_n^* &= \sum_{t=2}^{n} X_tX_{t-1} - y\left(\frac{1}{2}X_1^2 + X_2^2 + \cdots + X_{n-1}^2 + \frac{1}{2}X_n^2\right) \end{aligned}$$

이며, V_n^*는 다시 다음의 이차형식

$$V_n^* = \Big(X_1, \ldots, X_n\Big) \begin{pmatrix} -q/2 & 1/2 & 0 & 0 & 0 & \cdots & 0 \\ 1/2 & -q & 1/2 & 0 & 0 & \cdots & 0 \\ 0 & 1/2 & -q & 1/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1/2 & -q & 1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1/2 & -q/2 \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$= X'BX$$

으로 표현되어 안장점근사의 결과를 쉽게 적용할 수 있게 된다. 동일한 방법을 식 (3.4)의 통계량에 적용하면, 대응하는 이차형식은 다음과 같이 주어진다.

$$V_n' = \Big(X_1, \ldots, X_n\Big) \begin{pmatrix} 1/2-q & 1/2 & 0 & 0 & 0 & \cdots & 0 \\ 1/2 & -q & 1/2 & 0 & 0 & \cdots & 0 \\ 0 & 1/2 & -q & 1/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1/2 & -q & 1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1/2 & 1/2-q \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$= X'CX.$$

위의 적용 과정에서 알 수 있듯이 자기회귀계수에 대한 안장점근사는 다양한 추정량의 형태에 따라 이차형식의 행렬 형태에만 약간의 변형이 일어나며, 추정량의 형태에 관계없이 동일한 근사과정을 통해 수행됨을 알 수 있다. 이는 정규근사를 비롯한 기존의 근사법들이 서로 다른 여러 추정량 각각에 대한 근사분포를 요구하는 것에 비해 매우 단순한 절차를 통해 수행됨을 의미한다. 실제로 변형된 형태의 여러 추정량들은 통계적으로 개선된 의미를 가지고 있으나, 기존의 방법을 적용할 경우에는 근사분포의 유도과정에 큰 어려움이 존재한다.

3.2. 평균이 미지인 경우

이 절에서는 모형 (1.1)의 확장으로 $\{X_t\}$의 평균이 미지($\mu : unknown$)인 다음의 자기회귀모형을 생각하자.

$$X_t - \mu = \alpha(X_{t-1} - \mu) + e_t, \quad t = 1, \ldots, n. \tag{3.5}$$

여기서 모든 t에 대해 $|a| < 1$이고 e_t 는 평균이 0이고 분산이 σ^2 이며 서로 독립인 정규분포를 따르며, $E(X_t) = \mu,\ t = 1, \ldots, n$임을 가정하자. 또한 자기회귀계수 α에 대한 다음의 추정량을 생각하자.

$$\widehat{\alpha} = \frac{\sum_{t=2}^{n}(X_t - \widehat{\mu})(X_{t-1} - \widehat{\mu})}{\sum_{t=1}^{n}(X_t - \widehat{\mu})^2}. \tag{3.6}$$

위 식에서 $\widehat{\mu}$은 μ의 최우추정량 등이 고려될 수 있으나, 본 논문에서는 사용하기 간편한 추정량으로 $\widehat{\mu} = \overline{X}$ 를 고려하고, 이때의 추정량을 $\widehat{\alpha}^u$ 으로 표기하자. 이 경우 n이 충분히 크면, $\widehat{\alpha}^u$의 점근분포는 근사적으로 평균이 α이고, 분산이 $(1-\alpha^2)/n$ 인 정규분포를 따르며, 이 점근분포에 기초하여 자기회귀계수 α에 대한 추론이 수행되어 왔다.

이제 식 (3.6)의 분포함수에 대한 안장점근사의 과정을 알아보자. 먼저 추정량 $\widehat{\alpha}^u$의 분포함수는

$$Pr(\widehat{\alpha}^u \le q) = \Pr(V_n^u \le 0),$$

$$V_n^u = \sum_{t=2}^{n}(X_t - \overline{X})(X_{t-1} - \overline{X}) - q\sum_{t=1}^{n}(X_t - \overline{X})^2$$

으로 표현되며, 통계량 V_n^u은 다시 다음과 같은 이차형식으로 나타낼 수 있다.

$$V_n^u = \left(X_1 - \overline{X}, \ldots, X_n - \overline{X}\right)\begin{pmatrix} -q & 1/2 & 0 & 0 & 0 & \cdots & 0 \\ 1/2 & -q & 1/2 & 0 & 0 & \cdots & 0 \\ 0 & 1/2 & -q & 1/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1/2 & -q & 1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1/2 & -q \end{pmatrix}\begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix}$$

$$= Y'DY\ ,\ Y = (X_1 - \overline{X}, \ldots, X_n - \overline{X})'.$$

위 식에서 Y의 평균은 0이며, 공분산행렬은 다음의 정리 3.1과 같이 주어진다.

정리 3.1 $Y = (X_1 - \overline{X}, \ldots, X_n - \overline{X})'$ 의 공분산 행렬을

$$\Sigma = (\sigma_{ij})_{n\times n}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n$$

으로 표현할 때, 공분산행렬의 각 원소는 다음과 같이 주어진다.

1. $i = j$ 인 경우

   (a) $i = 1$ 인 경우

$$\frac{\sigma^2}{1-\alpha^2}\left\{1 - \frac{2(1-\alpha^n)}{n(1-\alpha)} + \frac{1}{n} + \frac{2\alpha}{n(1-\alpha)} - \frac{2\alpha(1-\alpha^{n-1})}{n^2(1-\alpha)^2}\right\}.$$

   (b) $i > 1$ 인 경우

$$\frac{\sigma^2}{1-\alpha^2}\left\{1 - \frac{2\alpha(1-\alpha^{i-1})}{n(1-\alpha)} - \frac{2(1-\alpha^{n-i+1})}{n(1-\alpha)} + \frac{1}{n} + \frac{2\alpha}{n(1-\alpha)} - \frac{2\alpha(1-\alpha^{n-1})}{n^2(1-\alpha)^2}\right\}.$$

2. $i \ne j$ 인 경우

   (a) $i = 1$ 인 경우

$$\frac{\sigma^2}{1-\alpha^2}\left\{\alpha^{j-i} - \frac{1-\alpha^{n-i+1}}{n(1-\alpha)} - \frac{\alpha(1-\alpha^{j-1})}{n(1-\alpha)} - \frac{1-\alpha^{n-j+1}}{n(1-\alpha)} + \frac{1}{n} + \frac{2\alpha}{n(1-\alpha)} - \frac{2\alpha(1-\alpha^{n-1})}{n^2(1-\alpha)^2}\right\}.$$

(b) $i > 1$ 인 경우

$$\frac{\sigma^2}{1-\alpha^2}\left\{\alpha^{j-i} - \frac{\alpha(1-\alpha^{i-1})}{n(1-\alpha)} - \frac{1-\alpha^{n-i+1}}{n(1-\alpha)} - \frac{\alpha(1-\alpha^{j-1})}{n(1-\alpha)} - \frac{1-\alpha^{n-j+1}}{n(1-\alpha)} + \frac{1}{n} + \frac{2\alpha}{n(1-\alpha)} - \frac{2\alpha(1-\alpha^{n-1})}{n^2(1-\alpha)^2}\right\}.$$

증명: 공분산의 정의와 단순계산을 반복하여 구할 수 있다. □

4. 모의실험

여기서는 앞 절에서 소개한 자기회귀계수에 대한 여러가지 유형의 추정량들 가운데, 가장 보편적으로 사용되는 식 (1.2)의 최소제곱추정량에 대한 안장점근사의 결과를 제시하고, 기존의 여러 근사법과의 비교를 수행한다. 표5.1~ 표5.4는 식 (3.1)에서 $\mu = 0$, $\sigma^2 = 1$ 이고, 자기회귀계수가 $\rho = 0.2, 0.8$ 이며, 표본의 크기 $n = 5, 30$인 경우에 대한 여러 가지 근사 결과를 제시하고 있다. 각 표에서 Exact는 정확한 값으로 10만번의 모의실험을 통해 구해진 값이며, Normal은 정규근사로 아래 Edgeworth근사식(Phillips, 1978)의 첫째항을 이용한 것이며, Edge1과 Edge2는 아래의 근사식에서 $O(n^{-1/2})$와 $O(n^{-1})$항까지의 근사를 나타낸다.

$$\begin{aligned}
&Pr\{\sqrt{n}(\widehat{\alpha}-\alpha)/(1-\alpha^2)^{1/2} \le x\} \\
&= \Phi(x) + \Phi(x)\big[n^{-1/2}\alpha(1-\alpha^2)^{-1/2}(x^2+1) \\
&\quad + \frac{1}{4n}(1-\alpha^2)^{-1}\{(1-\alpha^2)x + (1+\alpha^2)x^3 - 2\alpha^2x^5\}\big] + O(n^{-3/2}).
\end{aligned}$$

또한 Saddle1과 Saddle2는 각각 식 (2.3)과 (2.2)의 안장점근사의 결과를 나타낸다. 또한 본 논문에서의 모든 모의실험은 S−Plus를 통해 수행되었다.

각 표에서 알 수 있듯이 안장점근사(Saddle1, Saddle2)의 정도(precision)는 매우 뛰어난 결과를 제공하는데 비해, 기존의 정규근사 및 Edgeworth근사는 통계적 추론의 관심 영역인 꼬리부분에서 그 정도가 매우 떨어지는 것을 확인할 수 있다. 좀 더 구체적으로 살펴보면 자기회귀계수가 비교적 작은 경우($\alpha = 0.2$)에는 기존의 근사들도 어느 정도 정확도를 유지하나, 큰 값($\alpha = 0.8$)의 경우에는 매우 불안정한 근사를 보이고 있다. 그러나 안장점근사의 경우에는 α의 값이 큰 경우에도 매우 정확한 근사값을 제공할 뿐 이니라, 매우 작은 소표본($n = 5$)의 경우에도 기존의 근사법에 비해 상당히 정확한 값을 제공하고 있음을 알 수 있다.

5. 결론

본 논문에서는 자기회귀계수의 여러 추정량들을 이차형식의 구조로 이해하고, 이차형식의 분포함수에 대한 안장점근사(Na와 Kim, 2005)의 결과를 이들 추정량에 적용하였다. 이 방법은 기존의 정규근사를 크게 개선할 수 있으며, 변형된 형태의 다양한 추정량들에 대

해서도 근사분포의 유도가 불필요하다는 장점을 가진다. 특히, 자기회귀계수의 값이 비교적 큰 값을 가지는 경우, 기존의 여러 근사법들은 매우 불안정하며, 대표본하에서도 매우 부정확한 결과를 제공하는데 비해, 안장점근사는 소표본의 경우에도 매우 뛰어난 근사를 제공함을 모의실험을 통해 확인하였다. 따라서 본 논문에서 제시한 추정량들에 대한 효과적인 근사를 통해 자기회귀계수에 대한 매우 정확한 근사적 추론이 가능하다.

표 5.1: 자기회귀계수의 분포함수 근사($n = 5, \alpha = 0.2$)

		$\Pr\{\hat{a} \le q\}$에 대한 근사				
q	Exact	Normal	Edge1	Edge2	Saddle1	Saddle2
−1.9	0.0133	0.0287	0.0563	0.0325	0.0147	0.0147
−1.7	0.0288	0.0446	0.0780	0.0505	0.0308	0.0308
−1.5	0.0535	0.0668	0.1052	0.0759	0.0556	0.0556
−1.3	0.0899	0.0968	0.1389	0.1100	0.0908	0.0908
−1.1	0.1365	0.1357	0.1796	0.1534	0.1379	0.1379
−0.9	0.1955	0.1841	0.2280	0.2062	0.1973	0.1973
−0.5	0.3465	0.3085	0.3487	0.3376	0.3495	0.3495
0.5	0.7956	0.6915	0.7316	0.7428	0.7974	0.7975
0.7	0.8736	0.7580	0.8005	0.8170	0.8667	0.8668
0.9	0.9278	0.8159	0.8599	0.8817	0.9222	0.9222
1.1	0.9660	0.8643	0.9083	0.9345	0.9633	0.9633
1.3	0.9901	0.9032	0.9453	0.9742	0.9895	0.9895
1.5	0.9999	0.9332	0.9716	1.0009	0.9999	0.9999

표 5.2: 자기회귀계수의 분포함수 근사($n = 5, \alpha = 0.8$)

		$\Pr\{\hat{a} \le q\}$에 대한 근사				
q	Exact	Normal	Edge1	Edge2	Saddle1	Saddle2
−4.1	0.0287	0.0000	0.0010	0.0179	0.0205	0.0205
−3.7	0.0482	0.0001	0.0038	0.0512	0.0362	0.0363
−3.3	0.0760	0.0005	0.0127	0.1182	0.0599	0.0601
−2.9	0.1182	0.0019	0.0353	0.2184	0.0941	0.0944
−2.5	0.1708	0.0062	0.0820	0.3217	0.1416	0.1422
−2.3	0.2051	0.0107	0.1170	0.3593	0.1714	0.1722
−2.1	0.2417	0.0179	0.1598	0.3817	0.2057	0.2068
−1.9	0.2796	0.0287	0.2091	0.3892	0.2447	0.2462
−1.7	0.3287	0.0446	0.2627	0.3869	0.2885	0.2903
−1.5	0.3802	0.0668	0.3178	0.3834	0.3367	0.3389
−1.1	0.5116	0.1357	0.4227	0.4071	0.4464	0.4489
−0.7	0.6491	0.2420	0.5194	0.4934	0.5785	0.5805
−0.3	0.8073	0.3821	0.6300	0.6221	0.7503	0.7512
0.1	0.9730	0.5398	0.7789	0.7810	0.9615	0.9615

표 5.3: 자기회귀계수의 분포함수 근사($n = 30, \alpha = 0.2$)

		$\Pr\{\hat{a} \le q\}$에 대한 근사				
q	Exact	Normal	Edge1	Edge2	Saddle1	Saddle2
−2.1	0.0228	0.0179	0.0267	0.0235	0.0226	0.0226
−1.7	0.0533	0.0446	0.0582	0.0536	0.0537	0.0537
−1.3	0.1108	0.0968	0.1140	0.1092	0.1115	0.1115
−0.9	0.2045	0.1841	0.2020	0.1984	0.2043	0.2043
−0.5	0.3333	0.3085	0.3249	0.3231	0.3338	0.3338
−0.1	0.4916	0.4602	0.4751	0.4748	0.4895	0.4895
0.1	0.5697	0.5398	0.5548	0.5551	0.5712	0.5712
0.5	0.7285	0.6915	0.7079	0.7097	0.7262	0.7262
0.9	0.8531	0.8159	0.8339	0.8375	0.8511	0.8511
1.3	0.9341	0.9032	0.9204	0.9252	0.9338	0.9339
1.7	0.9766	0.9554	0.9691	0.9736	0.9773	0.9773
2.1	0.9944	0.9821	0.9910	0.9942	0.9944	0.9944
2.5	0.9992	0.9938	0.9985	1.0002	0.9991	0.9991

표 5.4: 자기회귀계수의 분포함수 근사($n = 30, \alpha = 0.8$)

		$\Pr\{\hat{a} \le q\}$에 대한 근사				
q	Exact	Normal	Edge1	Edge2	Saddle1	Saddle2
−3.9	0.0159	0.0000	0.0008	0.0057	0.0143	0.0143
−3.5	0.0264	0.0002	0.0030	0.0152	0.0236	0.0236
−3.1	0.0422	0.0010	0.0094	0.0333	0.0382	0.0382
−2.7	0.0664	0.0035	0.0245	0.0608	0.0605	0.0606
−2.3	0.0984	0.0107	0.0541	0.0945	0.0939	0.0940
−1.9	0.1501	0.0287	0.1024	0.1324	0.1426	0.1428
−1.5	0.2207	0.0668	0.1693	0.1802	0.2115	0.2119
−1.1	0.3133	0.1357	0.2529	0.2503	0.3051	0.3057
−0.7	0.4288	0.2420	0.3552	0.3509	0.4266	0.4273
−0.3	0.5756	0.3821	0.4833	0.4820	0.5955	0.5973
0.1	0.7273	0.5398	0.6374	0.6378	0.7305	0.7314
0.5	0.8699	0.6915	0.7986	0.8014	0.8739	0.8747
0.9	0.9662	0.8159	0.9332	0.9379	0.9678	0.9681
1.3	0.9983	0.9032	1.0154	1.0127	0.9984	0.9985

참고문헌

나종화, 김정숙 (2005). 비동차 이차형식의 분포함수에 대한 안장점근사, <응용통계연구>, **18**, 183–196.

Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients, *Biometrika*, **43**, 169–185.

Daniels, H. E. (1987). Tail probability approximations, *International Statistical Review*, **55**, 37–48.

Durbin, J. (1980). The approximate distribution of partial serial correlation coefficients calculated from residuals from regression on Fourier series, *Biometrika*, **67**, 335–349.

Jensen, J. L. (1992). The modified signed likelihood statistic and saddlepoint approximations, *Biometrika*, **79**, 693–703.

Jensen, J. L. (1995). *Saddlepoint Approximations*, Oxford, U.S.A.

Kuonen, D. (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables, *Biometrika*, **86**, 929–935.

Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables, *Advanced Applied Probability*, **12**, 475–490.

Phillips, P. C. B. (1977). A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators, *Econometrica*, **45**, 1517–1534.

Phillips, P. C. B. (1978). Edgeworth and saddlepoint approximations in the first-order noncircular autoregression, *Biometrika*, **65**, 91–98.

Sargan, J. D. (1976). Economic estimators and the Edgeworth approximation, *Econometrica*, **44**, 421–428.

[2006년 8월 접수, 2006년 10월 채택]

Small Sample Asymptotic Inferences for Autoregressive Coefficients via Saddlepoint Approximation*

Jong-Hwa Na[1)] Jeong-Sook Kim[2)]

ABSTRACT

In this paper we studied the small sample asymptotic inference for the autoregressive coefficient in AR(1) model. Based on saddlepoint approximations to the distribution of quadratic forms, we suggest a new approximation to the distribution of the estimators of the noncircular autoregressive coefficients. Simulation results show that the suggested methods are very accurate even in the small sample sizes and extreme tail area.

Keywords: Autoregressive coefficient, Edgeworth approximation, saddlepoint approximation, quadratic form.

* This work was supported by the research grant of the Chungbuk National University in 2006.

1) Professor, Dept. of Information and Statistics & Institute for Basic Science Research, Chungbuk National University, 12 Gashin-dong, Heungduk-gu, Cheongju, Chungbuk 361-763. Korea
E-mail: cherin@cbnu.ac.kr

2) Principal Researcher(Ph. D.), Information & Communication Dept., Health Insurance B/D 1586-9, Seocho3-Dong, Seocho-Ku, Seoul 137-706. Korea
E-mail: chastity@hiramail.net