응용통계연구 제 **16**권 1호, 2003년, pp. 141–150

AR(1)모형에서 자기회귀계수의 다중검정을 위한 베이지안방법

김경숙 [1)] 손영숙 [2)]

요 약

본 논문은 베이즈인자(Bayes factor)를 이용하여 정상(stationary) AR(1)모형의 자기회귀계수에 대해 다중검정하는 방법을 제시한다. 모수들에 대한 사전분포로는 무정보 사전분포(noninformative prior distribution)를 가정한다. 이러한 경우에 통상적으로 사용되는 베이즈인자를 근사없이 정확히 계산하여 각 모형에 대한 사후확률(posterior probability)을 얻는다. 최종적으로 모의실험 자료 및 실제 자료에 적용하여 이론의 결과가 잘 부합되는지를 검토한다.

주요용어: AR(1)모형, 자기회귀계수, 베이즈인자, 무정보 사전분포, 사후확률.

1. 서 론

자기회귀과정은 현 시점의 상태를 과거시점의 상태들 즉, 과거의 관측값들과 현 시점의 오차의 함수 형태로 나타내며, 확률과정 $\{Z_t\}$는 $Z_t = f(Z_{t-1}, Z_{t-2}, \ldots) + \epsilon_t$의 관계를 갖는다. 함수 f의 형태로는 선형함수가 많이 쓰이며, 오차항 $\{\epsilon_t\}$은 서로 독립이고 동일하게 $N(0, \sigma^2)$를 따르는 백색잡음과정(white noise process)을 가정한다. 특히, 확률과정 $\{Z_t\}$이 정상 1차-자기회귀과정(stationary first order autoregressive process)을 따를 때 이를 AR(1)과정이라고 부르며, 확률모형은 $Z_t = \delta + \phi Z_{t-1} + \epsilon_t (t = 1, 2, \ldots, n)$와 같이 정의되고, 확률변수 Z_t의 확률분포는 다음과 같다.

$$f(z_t|\phi, \delta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp\Big\{-\frac{1}{2\sigma^2}(z_t - \phi z_{t-1} - \delta)^2\Big\}, \ -1 < \phi < 1, \ -\infty < \delta < \infty, \ \sigma > 0.$$

이 때, 1차-자기회귀계수 ϕ에 대한 가설검정으로서 다음과 같이 K개의 모형(가설)으로 정의된 다중검정을 고려해 보자.

$$M_i : Z_t \sim f(z_t \mid \phi, \delta, \sigma), \ \phi \in \Phi_i, \ i = 1, 2, \ldots, K, \tag{1.1}$$

여기서, $\Phi_i \subset \Phi = \{\phi| -1 < \phi < 1\}$이다.

자기회귀계수 ϕ에 대한 다중검정 문제를 해결하는데 있어서 고전적인 방법은 다음과 같은 문제점이 있다. 첫째, 다중가설을 동시에 검정할 수 없다는 점이다. 예를들어 모수 ϕ를

1) (500-757) 광주광역시 북구 용봉동 300 전남대학교 자연과학대학 통계학과 박사과정
E-mail: ksook620@stat.chonnam.ac.kr
2) (500-757) 광주광역시 북구 용봉동 300 전남대학교 자연과학대학 통계학과 교수
E-mail: ysson@chonnam.ac.kr

모수공간에 포함된 임의의 상수 ϕ_0와 비교하는 다음의 세 가설, $H_0 : \phi < \phi_0$, $H_1 : \phi = \phi_0$, $H_2 : \phi > \phi_0$를 검정한다고 하자. 가장 흔히 접근되는 방법은 먼저 양측검정에 대한 두 가설 $H_0 : \phi = \phi_0$, $H_1 : \phi \neq \phi_0$를 검정하여 귀무가설이 기각되면, 다음으로 단측검정($H_1 : \phi < \phi_0$ 또는 $H_1 : \phi > \phi_0$)을 하는 것이다. 대안적인 방법으로는 귀무가설 H_0가 기각되면 모수에 대한 신뢰구간을 제시하는 것이다. 그러나 이러한 방법은 엄격한 의미의 표본이론에 있어서는 다소 문제점을 안고 있다. 이러한 문제를 해결하기 위해 고전적 접근방법의 측면에서 연구가 계속 진행되고 있으나 이는 상당히 까다로운 과제이다(Berger and Mortera, 1999).

둘째, 자기회귀계수 ϕ에 대한 가설검정을 위해 Durbin과 Watson (1951)의 D통계량이나 Marr와 Quesenberry (1991)의 NU통계량 등이 가장 널리 사용되는데, 이러한 검정통계량은 귀무가설 $H_0 : \phi = 0$이 참이라는 가정 하에서 도출된 통계량이라는 제한점이 있다. 따라서 일반적인 귀무가설인 $H_0 : \phi = \phi_0$(ϕ_0 : 임의의 상수)를 검정하고자 할 때에는 이러한 통계량을 사용할 수 없다는 문제가 발생한다.

반면에, 베이지안 방법은 두 개 이상의 검정하고자 하는 모든 모형(가설)을 동시에 검정할 수 있을 뿐만 아니라, 고전적인 방법으로는 검정결과에 따라 가설을 채택하거나 기각하는데 그치지만, 베이지안 방법을 적용하면 각 모형에 대한 사후확률 (posterior probability)을 계산함으로서 모형의 지지 정도를 수량화 할 수 있는 장점이 있다.

본 논문에서는 AR(1)모형의 자기회귀계수인 ϕ에 대한 다중검정, 즉, 임의의 상수 $\phi_0 \in (-1, 1)$에 대하여 $M_1 : \phi < \phi_0$, $M_2 : \phi = \phi_0$, $M_3 : \phi > \phi_0$, 을 위해 베이지안 방법을 사용하고자 한다. 이에 관련된 선행연구를 살펴보면, 김혜중과 한성실 (1998)에 의해 1차 자기회귀과정을 가지는 단순선형회귀모형, $y_t = \beta x_t + u_t$, $u_t = \phi u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 1/\tau)$, 에 대해 두 가설 $M_1 : \phi = \phi_0$, $M_2 : \phi \neq \phi_0$(ϕ_0: 임의의 상수)에 대한 검정방법이 연구되었다. ϕ와 (β, τ)에 대한 사전분포로는 각각 균일분포와 정규-감마분포를 가정하였다. 또한, 한성실과 김혜중 (1999)에 의해 1차 자기회귀과정을 가지는 중회귀모형, $y_t = \underline{x_t}'\underline{\beta} + u_t$, $u_t = \phi u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 1/\tau)$, 에 대해 다중가설 $M_1 : \phi = \phi_0$, $M_2 : \phi < \phi_0$, $M_3 : \phi > \phi_0$, 에 대한 검정방법이 연구되었다. ϕ와 (β, τ)에 대한 사전분포로는 각각 절단된 정규분포(truncated normal distribution)와 무정보 부적절 사전분포(noninformative improper prior distribution)를 가정하였다. 위의 두 논문 모두에서 계산이 힘든 복잡한 형태의 베이즈인자(Bayes factor)는 일반화 Savage-Dickey 밀도비의 형태로 변환시킨 후, Gibbs 샘플링을 이용하여 근사적으로 추정하였다.

본 논문에서는 모수 ϕ와 (δ, σ)에 대해 각각 균일분포와 무정보 부적절 사전분포의 가정하에 보통 사용되는 베이즈인자를 근사없이 정확하게 계산하여 사후확률을 얻었다.

본 논문의 2절에서는 무정보 사전분포의 가정 하에 베이즈인자를 이용하여 다중검정하는 방법에 대해 소개하였고, 3절에서는 이러한 방법론을 적용하여 정상 AR(1)모형의 자기회귀계수인 ϕ에 대해 다중검정하는 방법을 논의하였다. 마지막으로 4절에서는 모의실험 및 실제자료를 적용하여 이론을 뒷받침하는 근거로 제시하였다.

2. 베이즈인자를 이용한 다중검정 방법

K개의 모형인 식(1.1)의 모형선택을 위해 베이지안 통계에서 이용되는 가장 기본적인 도구는 베이즈인자이다. $\boldsymbol{Z} = \{Z_1, Z_2, \ldots, Z_n\}$의 관측값이 $\boldsymbol{z} = \{z_1, z_2, \ldots, z_n\}$로 주어졌을 때, 임의의 모형 M_i에 대한 모형 M_j의 베이즈인자 $B_{ji}(\boldsymbol{z})$는 다음과 같이 정의된다.

$$B_{ji}(\boldsymbol{z}) = m_j(\boldsymbol{z}) \; / \; m_i(\boldsymbol{z}), \; i, j = 1, 2, \ldots, K, \; i \neq j, \tag{2.1}$$

여기서, $m_i(\boldsymbol{z}) = \int_{\phi \in \Phi_i} \int_{-\infty}^{\infty} \int_0^{\infty} \pi_i(\phi, \delta, \sigma) L(\phi, \delta, \sigma | \boldsymbol{z}) d\sigma d\delta d\phi$이고, $\pi_i(\phi, \delta, \sigma)$는 모형 M_i 하에서 모수 (ϕ, δ, σ)에 대해 가정된 사전분포이며, $L(\phi, \delta, \sigma | \boldsymbol{z}) = \prod_{k=1}^{n} f(x_k | \phi, \delta, \sigma)$는 우도함수(likelihood function)를 나타낸다. 또한 $m_i(\boldsymbol{z})$는 모형 M_i하에서 표본 $\boldsymbol{z}$에 대한 주변확률밀도함수(marginal density) 또는 예측확률밀도함수(predictive density)라고도 불린다. 식(2.1)의 의미를 살펴보면, 모형 M_i에 속하는 표본들에 대한 모형 M_j에 속하는 표본들의 오즈(odds)로서 해석할 수 있으며, 또한 모형 M_i에 대한 모수의 사전분포인 $\pi_i(\phi, \delta, \sigma)$를 그 모형에 대한 가중치로서 적용하여 표본에 대한 확률을 가중평균 낸 후 표본에 의해 지지되는 두 모형 M_i, M_j의 비(ratio)를 산출한 것으로 해석할 수도 있다.

모수에 대한 사전정보가 없는 베이지안 통계 실험의 초기 단계에서는 보통 모수에 대한 사전분포로서 무정보 사전분포를 가정하며, 이러한 무정보 사전분포는 흔히 부적절(improper)한 경우가 많다. 부적절 사전분포를 사용하는 경우에는 베이즈인자에 임의의 상수항 c_j/c_i이 내포되어 있어서 이를 정확히 계산할 수가 없다. 이러한 문제는 Spiegelhalter와 Smith (1982)가 제안한 가상의 상수(imaginary constant)를 이용한 가상적 트레이닝 표본법(imaginary training sample method), Berger와 Pericchi (1996)의 내재적 베이즈인자(intrinsic Bayes factor, IBF), 또는 O'Hagan (1995)의 부분 베이즈인자(fractional Bayes factor, FBF) 등을 이용하여 해결할 수 있다. 한편, Jeffreys (1961)는 모수에 대한 사전분포로서 각 모형에 공통으로 포함되어 있는 모수에 대해서만 무정보 부적절 사전분포를 가정한 경우에는 베이즈인자를 계산하는 과정에서 임의의 상수항 c_j/c_i가 상쇄됨을 지적하였다.

각 모형에 대한 사후확률은 위에서 정의된 베이즈인자, $B_{ji}(\boldsymbol{z})$, 를 이용하여 다음과 같이 정의된다.

$$P(M_i|\boldsymbol{z}) = \left\{ \sum_{j=1}^{K} \frac{p_j}{p_i} B_{ji}(\boldsymbol{z}) \right\}^{-1}, \; i = 1, 2, \ldots, K, \tag{2.2}$$

여기서 p_i는 각 모형에 대한 사전확률을 의미하며, 일반적으로 각 모형에 대한 사전정보가 거의 없는 경우에는 동일한 값을 부여한다. 마지막으로 각 모형에 대한 사후확률을 계산하여 가장 큰 확률을 갖는 모형을 선택함으로써 다중검정의 결론을 얻는다.

3. AR(1)모형의 자기회귀계수 ϕ에 대한 다중검정

정상 AR(1)과정에서 자기회귀계수인 ϕ에 대해 다음과 같은 다중검정을 고려해 보자.

$$M_1 : \phi < \phi_0, \qquad M_2 : \phi = \phi_0, \qquad M_3 : \phi > \phi_0.$$

모수에 대한 사전분포로서 모수 ϕ는 균일분포(uniform distribution)를, 모수 $(\delta,\ \sigma)$는 무정보 부적절 확률분포를, 그리고 $(\delta,\ \sigma)$는 ϕ와 서로 독립임을 가정하였다. 이에 따라 다음과 같은 사전분포를 구성한다.

$$\pi_i(\phi,\delta,\sigma)=\pi_i(\delta,\sigma)\ \pi_i(\phi),\ i=1,2,3,$$

여기서,

$$\pi_i(\delta,\sigma)=\frac{c_i}{\sigma},\quad -\infty<\delta<\infty,\ \sigma>0,$$

c_i는 정의되지 않은 정규화상수(*normalizing constant*),

$$\begin{cases}\pi_1(\phi)=\dfrac{1}{\phi_0+1}, & \phi\in\Phi_1=\{\phi\mid -1<\phi<\phi_0\},\\ \pi_2(\phi)=1, & \phi\in\Phi_2=\{\phi\mid \phi=\phi_0\},\\ \pi_3(\phi)=\dfrac{1}{1-\phi_0}, & \phi\in\Phi_3=\{\phi\mid \phi_0<\phi<1\}.\end{cases}$$

크기 n인 시계열 관측값 $\boldsymbol{z}=\{z_1,z_2,\ldots,z_n\}$이 주어졌을때 우도함수는 다음과 같다.

$$L(\phi,\delta,\sigma|\boldsymbol{z})=(2\pi\sigma^2)^{-n/2}\exp\Big(-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(z_t-\phi z_{t-1}-\delta)^2\Big).$$

각 모형에 대한 주변확률밀도함수인 $m_i(\boldsymbol{z})$는 모수에 대한 사전분포와 우도함수를 이용하여 다음과 같이 정의된다.

$$m_i(\boldsymbol{z})=\begin{cases}\int_{\phi\in\Phi_i}\int_{-\infty}^{\infty}\int_0^{\infty}\pi_i(\phi,\delta,\sigma)L(\phi,\delta,\sigma|\boldsymbol{z})d\sigma d\delta d\phi, & i=1,3,\\ \int_{-\infty}^{\infty}\int_0^{\infty}\pi_i(\phi_0,\delta,\sigma)L(\phi_0,\delta,\sigma|\boldsymbol{z})d\sigma d\delta, & i=2.\end{cases}$$

위의 계산에서 σ와 (δ,ϕ)에 대한 적분은 각각 역감마-확률밀도함수와 t-확률밀도함수의 모수공간에서의 적분을 이용하여 다음과 같은 계산결과를 얻는다.

$$\begin{aligned}m_1(\boldsymbol{z})&=\frac{C_1}{1+\phi_0}\cdot\Gamma\Big(\frac{n-2}{2}\Big)\cdot g_1(\boldsymbol{z})^{-1/2}\cdot g_2(\boldsymbol{z})^{1-n/2}\cdot\Big[F_\phi(g_5(\boldsymbol{z}):\nu)-F_\phi(g_4(\boldsymbol{z}):\nu)\Big],\\ m_2(\boldsymbol{z})&=C_2\cdot\pi^{-1/2}\cdot\Gamma\Big(\frac{n-1}{2}\Big)\cdot g_3(\boldsymbol{z})^{(1-n)/2},\\ m_3(\boldsymbol{z})&=\frac{C_3}{1-\phi_0}\cdot\Gamma\Big(\frac{n-2}{2}\Big)\cdot g_1(\boldsymbol{z})^{-1/2}\cdot g_2(\boldsymbol{z})^{1-n/2}\cdot\Big[F_\phi(g_6(\boldsymbol{z}):\nu)-F_\phi(g_5(\boldsymbol{z}):\nu)\Big],\end{aligned}$$

여기서,

$$\begin{aligned}&\nu=n-2,\quad C_i=(c_i/2)n^{-1/2}\pi^{1-n/2},\\ &g_1(\boldsymbol{z})=\Big(1-\frac{1}{n}\Big)\textstyle\sum_{t=1}^{n}z_{t-1}^2-\frac{2}{n}\sum_{t=1}^{n-1}\sum_{p=t+1}^{n}z_{t-1}z_{p-1},\\ &g_2(\boldsymbol{z})=\Big(1-\frac{1}{n}\Big)\textstyle\sum_{t=1}^{n}z_t^2-\frac{2}{n}\sum_{t=1}^{n-1}\sum_{p=t+1}^{n}z_tz_p-\{g_8(\boldsymbol{z})\}^2,\\ &g_3(\boldsymbol{z})=\textstyle\sum_{t=1}^{n}(z_t-\phi_0z_{t-1})^2-\frac{1}{n}\Big\{\sum_{t=1}^{n}(z_t-\phi_0z_{t-1})\Big\}^2,\\ &g_4(\boldsymbol{z})=g_7(\boldsymbol{z})\big(-1-g_8(\boldsymbol{z})\big),\quad g_5(\boldsymbol{z})=g_7(\boldsymbol{z})\big(\phi_0-g_8(\boldsymbol{z})\big),\\ &g_6(\boldsymbol{z})=g_7(\boldsymbol{z})\big(1-g_8(\boldsymbol{z})\big),\quad g_7(\boldsymbol{z})=\sqrt{(n-2)g_1(\boldsymbol{z})/g_2(\boldsymbol{z})}\\ &g_8(\boldsymbol{z})=\frac{1}{g_1(\boldsymbol{z})}\Big\{\Big(1-\frac{1}{n}\Big)\textstyle\sum_{t=1}^{n}z_tz_{t-1}-\frac{1}{n}\sum_{t=1}^{n-1}\sum_{p=t+1}^{n}(z_{t-1}z_p+z_tz_{p-1})\Big\},\end{aligned}$$

이고, $F_\phi(\cdot)$는 t-확률분포의 누적분포함수를 의미하며, 확률변수 ϕ가 자유도 ν인 t-확률분포를 따른다고 할 때 $F_\phi(\phi_0 : \nu) = P(\phi \leq \phi_0)$로 정의된다.

본 논문에서는 각 모형에서 공통의 모수에 대해서만 무정보 부적절 사전분포를 가정하였으므로, 베이즈인자를 계산하는 과정에서 발생하는 임의의 상수항 c_j/c_i는 $c_i = c_j (i \neq j)$이므로 상쇄되어 베이즈인자를 정확히 계산할 수 있다. 따라서 식 (2.1)에 정의된 베이즈인자를 이용하여 본 논문에서 검정하고자 하는 문제를 충분히 해결할 수 있다.

마지막으로, 식(2.2)를 이용하여 각 모형 $M_i(i = 1, 2, 3)$의 사후확률을 계산하며, 이 가운데 가장 큰 확률을 갖는 모형을 선택함으로써 다중검정의 결론을 얻을 수 있다.

4. 모의실험 및 실제자료 분석

이제 앞 절에서 계산한 베이즈인자에 의한 사후확률의 유용성을 평가하기 위해 수치분석을 수행하기로 한다. 모든 계산은 MATLAB(The MATH WORKS Inc., 1998)을 사용하였으며, 특히 t-누적분포함수값의 계산은 tcdf함수를 이용하였다.

AR(1)모형에 포함되어 있는 모수인 상수항 δ와 오차항의 분산 σ^2은 일반성을 잃지 않고서 각각 0과 1로 고정하였고, 자기회귀계수 ϕ의 값은 각각 $(-0.9, -0.7, -0.5, -0.3, 0, 0.3, 0.5, 0.7, 0.9)$ 일 때에 대해 표본크기 $n = 30$인 자료를 300회 반복추출하였다. 각 경우에 대해 모수 ϕ에 대한 임의의 상수 ϕ_0값을 $(-0.9, -0.5, 0, 0.5, 0.9)$로 설정하여 다중검정을 시행하였다. 그 결과로서 각 모형 M_i에 대해 300회 반복검정하여 얻은 사후확률의 평균과 표준편차를 [표 4.1]에 제시하였다.

[표 4.1]은 대부분의 경우에 모의실험결과가 이론에 부합됨을 보였다. 일반적으로 M_2모형(다중검정에 대한 참모형(true model)이 하나의 값에 대해 검정하는 정확한(precise) 모형)인 경우에 M_1모형이나 M_3모형에 비해 사후확률의 표준편차가 더 컸으며, 모수 ϕ와 임의의 상수 ϕ_0의 차이가 클수록 표준편차는 점점 더 작아지는 결과를 보였다.

다음으로는 AR(1)모형을 따르는 두 가지 유형의 자료로서, 모수 ϕ가 양(+)의 값을 갖는 경우와 거의 0에 가까운 값을 갖는 경우이다. 첫번째 자료는 모수 ϕ가 양의 값을 갖는 것으로 추정된 경우로서, 어떤 화학 처리에 대한 점성도를 매 시간당 측정한 자료이며, 추정된 모형은 $Z_t = 1.17 + 0.87Z_{t-1} + \epsilon_t$로 알려져 있다(Box와 Jenkins, 1976, p196, p528). 다중검정을 위해 임의의 상수 ϕ_0값은 추정된 자기회귀계수인 $\hat{\phi} = 0.87$을 중심으로 더 크거나 작은 여러 경우를 살펴보았으며, 이에 대한 결과를 [표 4.2]에 제시하였다.

두번째 자료는 모수 ϕ가 거의 0에 가까운 값을 갖는 것으로 추정된 경우로서, 식료품 수급 관련 자료 중에서 일인 당 식료품 소비량과 가처분소득에 대한 일부 자료만을 발췌하였다(Kmenta, 1971). 이들 변인에 대해 회귀분석을 시행하여 얻은 오차항에 대해 Durbin-Watson의 독립성 검정 결과, 검정통계량 $D = 2.297$이고 상한값 $D_U = 1.41$로서 오차항은 독립이라고 결론지을 수 있다. 따라서 추정된 자기회귀계수인 $\hat{\phi} = 0$에 대해 임의의 상수 ϕ_0는 0을 중심으로 더 크거나 작은 여러 경우의 값을 설정하여 다중검정하였고 그 결과를 [표 4.3]에 제시하였다.

위의 두 자료에 대한 결과표를 살펴보면 이는 매우 유사한 형태의 경향을 보였다. 즉, 임

[표 4.1] 자기회귀계수 ϕ인 AR(1)모형을 따르는 모의실험 자료에 대한 다중검정
$M_1 : \phi < \phi_0$, $M_2 : \phi = \phi_0$, $M_3 : \phi > \phi_0$ (n=30, 반복 300회)

ϕ		ϕ_0	true model	Mean $P(M_1)$	Mean $P(M_2)$	Mean $P(M_3)$	S.D. $P(M_1)$	S.D. $P(M_2)$	S.D. $P(M_3)$
-0.9	=	-0.9	M_2	0.4140	0.4913 *	0.0947	0.1160	0.0834	0.1325
-0.9	<	-0.5	M_1	0.9307 *	0.0651	0.0042	0.1477	0.1360	0.0128
-0.9	<	0.0	M_1	0.9983 *	0.0016	0.0001	0.0117	0.0108	0.0009
-0.9	<	0.5	M_1	1.0000 *	0.0000	0.0000	0.0001	0.0001	0.0000
-0.9	<	0.9	M_1	1.0000 *	0.0000	0.0000	0.0000	0.0000	0.0000
-0.7	>	-0.9	M_3	0.2705	0.4299 *	0.2996	0.1005	0.1390	0.2376
-0.7	<	-0.5	M_1	0.6236 *	0.3402	0.0362	0.2549	0.2173	0.0513
-0.7	<	0.0	M_1	0.9658 *	0.0310	0.0032	0.0984	0.0861	0.0142
-0.7	<	0.5	M_1	0.9999 *	0.0001	0.0000	0.0005	0.0005	0.0001
-0.7	<	0.9	M_1	1.0000 *	0.0000	0.0000	0.0000	0.0000	0.0000
-0.5	>	-0.9	M_3	0.1254	0.2144	0.6602 *	0.0987	0.1603	0.2587
-0.5	=	-0.5	M_2	0.2810	0.5679 *	0.1511	0.1550	0.1164	0.1470
-0.5	<	0.0	M_1	0.8516 *	0.1333	0.0150	0.2052	0.1796	0.0271
-0.5	<	0.5	M_1	0.9946 *	0.0046	0.0008	0.0324	0.0273	0.0051
-0.5	<	0.9	M_1	1.0000 *	0.0000	0.0000	0.0003	0.0002	0.0001
-0.3	>	-0.9	M_3	0.0463	0.0823	0.8714 *	0.0549	0.0954	0.1503
-0.3	>	-0.5	M_3	0.1783	0.5169 *	0.3048	0.1128	0.1676	0.2505
-0.3	<	0.0	M_1	0.6294 *	0.3256	0.0451	0.2580	0.2151	0.0509
-0.3	<	0.5	M_1	0.9825 *	0.0151	0.0025	0.0469	0.0398	0.0071
-0.3	<	0.9	M_1	0.9996 *	0.0003	0.0001	0.0018	0.0012	0.0006
0.0	>	-0.9	M_3	0.0049	0.0089	0.9863 *	0.0126	0.0222	0.0348
0.0	>	-0.5	M_3	0.0507	0.2314	0.7179 *	0.0568	0.2000	0.2543
0.0	=	0.0	M_2	0.2358	0.5669 *	0.1973	0.1899	0.1404	0.1628
0.0	<	0.5	M_1	0.8107 *	0.1573	0.0320	0.2216	0.1792	0.0429
0.0	<	0.9	M_1	0.9926 *	0.0048	0.0026	0.0229	0.0146	0.0083
0.3	>	-0.9	M_3	0.0004	0.0008	0.9988 *	0.0020	0.0036	0.0055
0.3	>	-0.5	M_3	0.0058	0.0347	0.9595 *	0.0117	0.0643	0.0759
0.3	>	0.0	M_3	0.0743	0.4116	0.5141 *	0.0857	0.2104	0.2700
0.3	<	0.5	M_1	0.4324	0.4460 *	0.1215	0.2695	0.1958	0.0830
0.3	<	0.9	M_1	0.9210 *	0.0506	0.0285	0.1193	0.0754	0.0439
0.5	>	-0.9	M_3	0.0000	0.0001	0.9999 *	0.0003	0.0005	0.0008
0.5	>	-0.5	M_3	0.0008	0.0055	0.9937 *	0.0026	0.0163	0.0189
0.5	>	0.0	M_3	0.0294	0.2201	0.7505 *	0.0444	0.2234	0.2624
0.5	=	0.5	M_2	0.2045	0.5550 *	0.2405	0.1858	0.1240	0.1468
0.5	<	0.9	M_1	0.7348 *	0.1683	0.0970	0.2285	0.1428	0.0858
0.7	>	-0.9	M_3	0.0000	0.0000	1.0000 *	0.0000	0.0000	0.0000
0.7	>	-0.5	M_3	0.0002	0.0012	0.9986 *	0.0024	0.0132	0.0157
0.7	>	0.0	M_3	0.0072	0.0687	0.9241 *	0.0181	0.1333	0.1508
0.7	>	0.5	M_3	0.0749	0.4548	0.4703 *	0.1075	0.1925	0.2453
0.7	<	0.9	M_1	0.4863 *	0.3200	0.1937	0.2823	0.1706	0.1125
0.9	>	-0.9	M_3	0.0000	0.0000	1.0000 *	0.0000	0.0000	0.0000
0.9	>	-0.5	M_3	0.0001	0.0010	0.9989 *	0.0019	0.0134	0.0153
0.9	>	0.0	M_3	0.0014	0.0163	0.9824 *	0.0058	0.0630	0.0687
0.9	>	0.5	M_3	0.0233	0.2304	0.7464 *	0.0441	0.2378	0.2711
0.9	=	0.9	M_2	0.2372	0.4502 *	0.3127	0.2565	0.1438	0.1297

* : 자료로부터 지지되는 모형임.

[표 4.2] 추정된 자기회귀계수가 $\hat{\phi} = 0.87$인 AR(1)모형 자료에 대한 $M_1 : \phi < \phi_0$, $M_2 : \phi = \phi_0$, $M_3 : \phi > \phi_0$의 다중검정

$\hat{\phi}$		ϕ_0	추정모형	$P(M_1)$	$P(M_2)$	$P(M_3)$
0.87	>	0.70	M_3	0.0000	0.0000	1.0000 *
0.87	>	0.75	M_3	0.0000	0.0064	0.9935 *
0.87	>	0.78	M_3	0.0016	0.1791	0.8192 *
0.87	>	0.80	M_3	0.0063	0.5617 *	0.4320
0.87	>	0.85	M_3	0.0242	0.9279 *	0.0480
0.87	=	0.87	M_2	0.0419	0.9313 *	0.0268
0.87	<	0.90	M_1	0.1351	0.8514 *	0.0134
0.87	<	0.92	M_1	0.3572	0.6353 *	0.0075
0.87	<	0.95	M_1	0.8928 *	0.1062	0.0009
0.87	<	0.99	M_1	0.9993 *	0.0007	0.0000

* 자료로부터 지지되는 모형. 출처: Box와 Jenkins (1976). $n = 310$.

[표 4.3] 추정된 자기회귀계수가 $\hat{\phi} = 0$인 AR(1)모형 자료에 대한 $M_1 : \phi < \phi_0$, $M_2 : \phi = \phi_0$, $M_3 : \phi > \phi_0$의 다중검정

$\hat{\phi}$		ϕ_0	추정모형	$P(M_1)$	$P(M_2)$	$P(M_3)$
0	>	-0.8	M_3	0.0074	0.0762	0.9164 *
0	>	-0.6	M_3	0.0343	0.2640	0.7017 *
0	>	-0.4	M_3	0.0884	0.5066 *	0.4050
0	>	-0.2	M_3	0.1739	0.6163 *	0.2098
0	=	0	M_2	0.3355	0.5552 *	0.1093
0	<	0.2	M_1	0.6143 *	0.3375	0.0482
0	<	0.4	M_1	0.8718 *	0.1144	0.0138
0	<	0.6	M_1	0.9731 *	0.0242	0.0027
0	<	0.8	M_1	0.9954 *	0.0042	0.0004

* 자료로부터 지지되는 모형. 출처: Kmenta (1971). $n = 20$.

의의 상수 ϕ_0값이 모수 추정치 $\hat{\phi}$와 차이가 적은 경우는 다소 이론에 부합되지 않은 결과를 보였지만, 이는 모수의 추정값을 사용함으로서 발생되는 추정오차에 기인한 것으로 볼 수 있으며, 그 외에 ϕ_0가 추정오차 범위를 벗어난 경우는 이론에 잘 부합되는 결과를 보였다.

5. 결 론

본 논문은 정상 AR(1)모형의 자기회귀계수의 다중검정을 위한 베이지안방법을 논의하였다. 고전전인 접근방법과 비교할 때, 베이지안 방법은 다중 모형(가설)을 동시에 설정하여 검정한다는 장점과 검정 결과를 각 모형에 대한 확률로써 설명할 수 있다는 장점이 있다. 특히, 고전적인 방법으로는 접근하기 어려웠던 가설 $\phi = \phi_0(\phi_0$: 임의의 상수)에 대한 검정이 가능하다.

AR(1)모형에 대한 다중검정은 공통의 모수인 상수항 δ와 오차항의 표준편차 σ에 대해서만 부적절 사전분포를 가정했기 때문에, 가장 간단한 형태의 베이즈인자를 적용할 수 있었다. 또한 이때의 베이즈인자는 t-누적분포함수값을 계산할 수 있는 통계소프트웨어에 의해 쉽게 얻어진다. 이론적인 결과를 검토하기 위해 모의실험 자료와 실제 자료에 적용하였다. 지면상 본 논문에 제시하지는 않았지만 다양한 모수값과 표본의 크기에 대해 모의실험이 이루어졌다. 대부분의 결과는 본 논문에 제시된 내용과 유사하며, 특히 표본의 크기가 클수록 이론의 결과에 보다 더 부합되었다.

고전적인 방법의 가설검정은 단지 두 개의 가설을 검정하므로 검정력함수(power function of test)를 이용하여 검정력을 계산할 수 있다. 베이지안 방법의 가설검정에서도 단지 두 개의 가설을 검정하는 경우에는 모의실험을 통하여 검정력을 계산할 수 있으며, 이러한 경우에는 고전적 방법에 의한 검정력과 비교도 가능하다. 그러나 가설이 셋이상인 경우에는 가설검정에 대한 검정력을 구하기가 쉽지 않다. 따라서 본 논문에서는 그 대안으로써 각 모형에 대한 사후확률의 분포를 이용하여 비교하였다. 향후 연구과제로는 본 논문에서 다루지 않은 여러 다른 확률모형에 대해서도 이와 같은 다중검정 방법을 적용해 볼 수 있을 것이다.

참고문헌

[1] 김혜중, 한성실(1998). 회귀모형의 오차항의 1차 자기상관에 대한 베이즈 검정법, <응용통계연구>, 제11권 1호, 97-111.

[2] 한성실, 김혜중(1999). 다중베이즈요인에 의한 회귀모형 오차항의 1차 자기상관 검정, <응용통계연구>, 제12권 2호, 605-619.

[3] Berger, J. O. and Mortera, J. (1999). Default Bayes Factors for Non-Nested Hypothesis Testing, *Journal of the American Statistical Association*, Vol. 94, No. 446, 542-554.

[4] Berger, J. O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection

and Prediction, *Journal of the American Statistical Association*, Vol. 91, No. 433, 109-122.

[5] Bertolino, F., Piccinato, L., and Racugno, W. (1995). Multiple Bayes Factors for Testing Hypotheses, *Journal of the American Statistical Association*, Vol. 90, No. 429, 213-219.

[6] Box, G. P. and Jenkins, G. M. (1976). *Time Series Analysis: forecasting and control*, revised edition, Holden-Day, 196-528.

[7] Durbin, J. and Watson, G. S. (1951). Testing for Serial Correlation in Least Square Regression, *Biometrika*, Vol. 38, 159-177.

[8] Kmenta, J. (1971). *Elements of Econometrics*, New York, Wiley.

[9] Jeffreys, H. (1961). *Theory of Probability*, London: Oxford University Press.

[10] Marr, R. L. and Quesenberry, C. P. (1991). A NU Test for Serial Correlation of Residuals from One or More Regression Regimes, *Technometrics*, Vol. 33, No. 4, 441-458.

[11] O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparison, *Journal of the Royal Statistical Society*, B, Vol. 57, No. 1, 99-138.

[12] Spiegelhalter, D. J. and Smith, A.F.M. (1982). Bayes Factorsfor Linear and Log-linear Models with Vague Prior Information, *Journal of the Royal Statistical Society*, B, Vol. 44, No. 3, 377-387.

[13] The MATH WORKS Inc. (1998). *MATHLAB/Statistics Toolbox*, Version 5.2, Natick, MA.

[2002년 9월 접수, 2002년 12월 채택]

Bayesian Method for the Multiple Test of an Autoregressive Parameter in Stationary AR(1) Model

Kyungsook Kim[1)] Young Sook Son[2)]

ABSTRACT

This paper presents the multiple testing method of an autoregressive parameter in stationary AR(1) model using the usual Bayes factor. As prior distributions of parameters in each model, uniform prior and noninformative improper priors are assumed. Posterior probabilities through the usual Bayes factors are used for the model selection. Finally, to check whether these theoretical results are correct, simulated data and real data are analyzed.

Keywords: AR(1) model; autoregressive parameter; Bayes factor; noninformative prior; posterior probability.

1) Graduate Student, Department of Statistics, Chonnam National University.
E-mail: ksook620@stat.chonnam.ac.kr
2) Professor, Department of Statistics, Chonnam National University.
E-mail: ysson@chonnam.ac.kr