응용통계연구 제 16권 2호, 2003년, pp. 335-349

시뮬레이션을 통한 베이즈요인에 의한 모형선택의 비교연구 : 포아송, 음이항모형의 선택과 정규, 이중지수, 코쉬모형의 선택

오미라 [1] 윤소영 [2] 심정욱 [3] 손영숙 [4]

요 약

본 논문에서는 포아송분포 대 음이항분포, 그리고 정규분포, 이중지수분포 대 코쉬분포에 대한 모형선택을 위하여 베이지안 방법을 사용한다. 각 모수에 대한 사전분포로는 무정보 부적절 사전분포의 가정 하에, 베이지안 모형선택을 위하여 O'Hagan (1995)의 부분적 베이즈요인을 이용하였다. 실제자료와 모의실험자료의 분석을 통하여 부분적 베이즈요인의 유용성을 Berger와 Pericchi (1996, 1998)의 내재적 베이즈요인들과 함께 비교 검토해 본다.

주요용어: 무정보 부적절 사전분포, 베이지안 모형선택, 부분적 베이즈요인, 내재적 베이즈요인.

1. 서론

음이 아닌 정수의 값을 가지는 확률변수의 분포로서 포아송분포와 음이항분포를 생각해 보자. 음이항분포 $NB(\gamma, p)$의 분산 $\frac{\gamma(1-p)}{p^2}$ 는 항상 평균 $\frac{\gamma(1-p)}{p}$ 보다 크다. 이와 비교해 보면 포아송분포 $Poi(\lambda)$는 평균과 분산이 λ로서 동일하다. 그러나 실제로 시행횟수 혹은 발생건수들의 분포는 흔히 분산이 평균보다 큰 경우가 많다. 그림 1.1은 두 모형의 평균을 같게 하여 비교해 보았다. 그림 1.1에서 평균이 1인 경우와 평균이 2인 경우를 보면 음이항분포가 포아송분포보다 꼬리부분이 더 두텁고 길기 때문에 음이항분포가 포아송분포보다 분산이 더 크다는 것을 알 수 있다. 음이항분포, $NB(\gamma, p)$에서 $\gamma = 1$일 때의 음이항분포는 기하분포가 되므로 기하분포는 음이항분포에 속한다고 할 수 있다.

1) (500-757) 광주광역시 북구 용봉동 300, 전남대학교 자연과학대학 통계학과, 박사과정
E-mail : omr@chonnam.ac.kr
2) (442-835) 경기도 수원시 팔달구 인계동 1133-9, 경기통계사무소 경제조사과
E-mail : syyun@nso.go.kr
3) (500-757) 광주광역시 북구 용봉동 300, 전남대학교 자연과학대학 통계학과, 교수
E-mail : jwsim@chonnam.ac.kr
4) (500-757) 광주광역시 북구 용봉동 300, 전남대학교 자연과학대학 통계학과, 교수
E-mail : ysson@chonnam.ac.kr

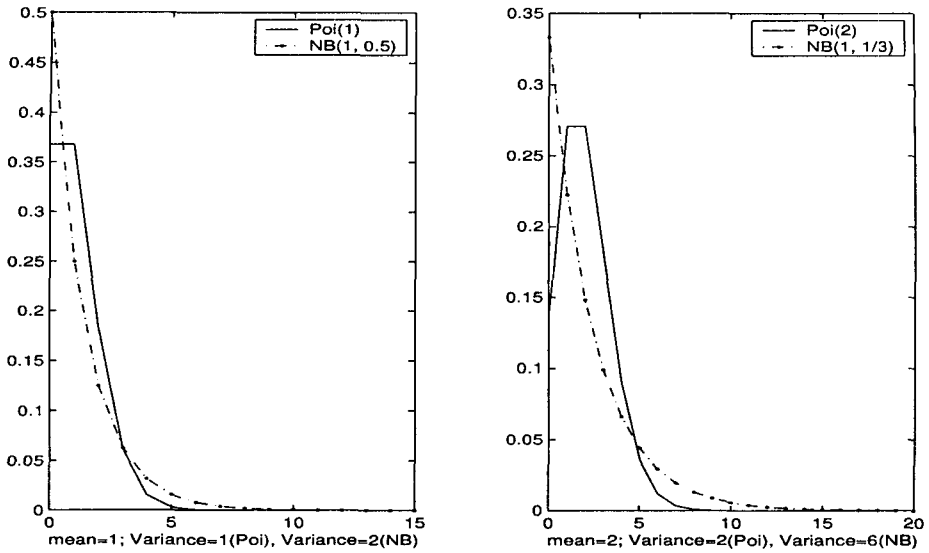

그림 1.1: 평균이 같은 포아송분포와 음이항분포의 비교

본 논문에서는 첫 번째 문제로서 양의 이산형의 자료에 대하여 아래와 같은 포아송 확률모형(M_1)과 음이항 확률모형(M_2)

$$\begin{cases} M_1 : X \sim Poi(\lambda), & f_x(\boldsymbol{x} \mid \lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,1,2,\ldots, \quad \lambda > 0\ , \\ M_2 : X \sim NB(1,p), & f_x(\boldsymbol{x} \mid p) = p(1-p)^x, \quad x = 0,1,2,\ldots, \quad 0 < p < 1 \end{cases} \tag{1.1}$$

중 하나의 모형을 선택하는 문제를 다루고자 한다.

연속형분포에서 위치-척도의 형태를 가지면서 좌우대칭인 분포로는 정규분포, 이중지수분포, 코쉬분포가 있다. 정규분포, 이중지수분포, 코쉬분포를 살펴보면 공통적으로 실수구간을 확률분포의 공간으로 가지며 또한 분포의 중앙값을 중심으로 대칭이다. 그러나 일반적으로 정규분포, 이중지수분포, 코쉬분포 차례로 분포의 꼬리가 두텁다고 알려져 있다. 그림 1.2는 위치모수값 0, 척도모수값 1인 경우의 세 분포의 확률밀도함수를 나타낸 것이다.

본 논문에서는 두 번째 문제로서 아래의 정규 확률모형(M_1), 이중지수 확률모형(M_2), 그리고 코쉬 확률모형(M_3)

$$\begin{cases} M_1 : X \sim N(0,\sigma^2), & f_x(\boldsymbol{x} \mid \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}, & -\infty < x < \infty, \quad \sigma > 0\ , \\ M_2 : X \sim DE(0,\rho), & f_x(\boldsymbol{x} \mid \rho) = \dfrac{1}{2\rho}\, e^{-\frac{|x|}{\rho}}, & -\infty < x < \infty, \quad \rho > 0\ , \\ M_3 : X \sim Cau(1,\tau), & f_x(\boldsymbol{x} \mid \tau) = \dfrac{1}{\pi}\dfrac{\tau}{\tau^2 + \pi^2}, & -\infty < x < \infty, \quad \tau > 0 \end{cases} \tag{1.2}$$

중 하나의 모형을 선택하는 문제를 다루고자 한다.

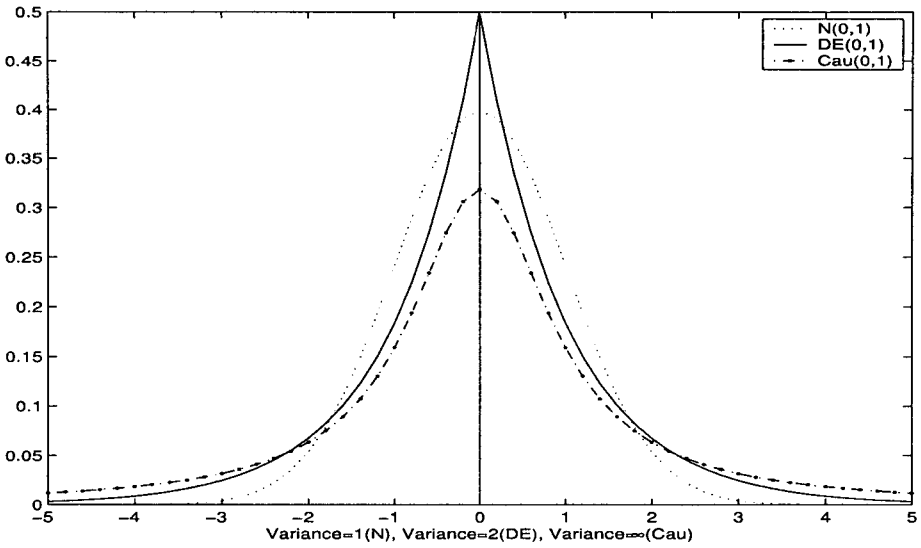

그림 1.2: 위치와 척도가 같은 정규분포, 이중지수분포와 코쉬분포의 비교

식 (1.1), 식 (1.2)와 같이 서로 다른 두 개 이상의 모형을 비교하기에는 고전적인 검정방법으로는 쉽지 않다. 고전적인 검정방법은 가설이 설정한 모형과 같은지, 그렇지 않는지를 판단하며, 가설에 설정된 모형은 하나라는 제약이 따른다. 그러나 베이지안 방법은 두 개 이상의 서로 다른 모형들을 동시에 설정하여 검정할 수 있고, 또한 각 모형에 대한 사후확률을 계산하여 모형의 상대적인 중요도에 의해 모형선택이 이루어진다. 이러한 이유로 가설검정이라는 용어보다는 베이지안 모형선택 또는 베이지안 모형비교라는 표현이 선호된다.

본 논문에서는 각 확률모형의 모수들에 대하여 무정보 사전분포(noninformative prior distribution)를 가정한다. 이때의 무정보 사전분포들은 부적절(improper)하기 때문에 O'Hagan (1995)의 부분적 베이즈요인(fractional Bayes factor ; FBF)을 이용하여 베이지안 모형선택을 수행하고자 한다. 이와 관련된 선행연구로는 식 (1.1)에 대해서는 표본크기가 $n = 10$이고, 평균이 서로 같은 두 이산형의 자료를 가지고, 식 (1.2)에 대해서는 표본크기가 $n = 8$인 하나의 자료를 가지고서 Bertolino와 Racugno (1996)는 Berger와 Pericchi (1996)의 산술 내재적 베이즈요인(arithmetic intrinsic Bayes factor ; AIBF), Berger와 Pericchi (1998)는 AIBF와 중위수 IBF(median IBF ; MIBF)를 계산하여 모형선택을 하였다.

본 논문에서는 또 다른 디폴트 베이즈요인인 FBF를 계산하여 실제자료분석과 모의실험을 통하여 디폴트 베이즈요인들의 유용성을 살펴보았다.

본 논문의 2절에서는 부분적 베이즈요인을 소개하였고, 3절에서는 Berger와 Pericchi (1998)가 사용한 무정보 부적절 사전분포의 가정 하에 포아송 확률모형 대 음이항 확률모형의 모형선택, 정규, 이중지수 확률모형 대 코쉬 확률모형의 모형선택을 위한 부분적 베이즈요인의 계산법을 설명한다. 4절에서는 실제자료와 모의실험자료를 가지고 베이지안 모

형선택의 과정을 수행해 본다.

2. 부분적 베이즈 요인

확률표본 $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$가 추출될 가능성이 있는 다음과 같은 q개의 모형을 고려해 보자.

$$M_j : \boldsymbol{X} \sim f_j(X \mid \theta_j), \quad \theta_j \in \Theta_j, \quad j = 1, 2, \ldots, q,$$

여기서 θ_j와 Θ_j는 모형 M_j하에서의 모수벡터와 모수공간이다. 확률표본 $\boldsymbol{X}$와 상수 b의 함수를 다음과 같이 정의하자.

$$B_{ji}(\boldsymbol{X} \mid b) = \frac{m_j(\boldsymbol{X} \mid b)}{m_i(\boldsymbol{X} \mid b)} \quad i, j = 1, 2, \ldots, q, \quad j \neq i, \tag{2.1}$$

여기서

$$m_j(\boldsymbol{X} \mid b) = \int_{\Theta_j} \pi_j(\theta_j) \cdot L_j^b(\theta_j \mid \boldsymbol{X}) d\theta_j, \tag{2.2}$$

$\pi_j(\theta_j)$는 모형 M_j하에서의 모수 θ_j의 사전분포, $b(0 < b \leq 1)$는 상수, $L_j^b(\theta_j|\boldsymbol{X}) = \{\prod_{k=1}^n f_j(X_k \mid \theta_j)\}^b$, 그리고 특히 $L_j^1(\theta_j|\boldsymbol{X})$는 확률모형 M_j의 우도함수(likelihood function)를 나타낸다.

베이지안 가설검정이나 베이지안 모형선택을 위해서는 베이즈요인이 사용된다. 모형 M_i에 대한 모형 M_j의 베이즈요인 B_{ji}는 다음과 같이 정의된다.

$$B_{ji} = B_{ji}(\boldsymbol{X} \mid b = 1) = \frac{m_j(\boldsymbol{X} \mid b = 1)}{m_i(\boldsymbol{X} \mid b = 1)}, \tag{2.3}$$

여기서 $m_j(\boldsymbol{X} \mid b = 1)$는 모형 M_j에서의 주변 또는 예측 확률밀도함수(marginal or predictive density)라고 한다.

베이지안 검정에서는 실험의 초기단계에서 모수에 대한 특별한 사전정보가 없을 때 모수에 대한 사전분포를 무정보 분포를 가정하는 경우가 많다. 그러나 무정보 분포는 흔히 부적절한 경우가 많이 있다. 식 (2.3)에서 사전분포를 무정보 부적절 분포를 사용하면 임의의 상수를 포함하고 있기 때문에 보통의 베이즈요인을 사용할 수 없게 된다. 이러한 문제점을 해결하기 위해서 O'Hagan (1995)의 FBF와 Berger와 Pericchi (1996)의 IBF를 사용할 수 있다. FBF와 IBF같은 베이즈요인들은 Spiegelhalter와 Smith (1982)가 제안한 가상의 상수(imaginary constant)를 고려할 필요가 없고, 또한 공액(conjugate) 사전분포를 가정하는 경우 지정해 주어야 할 초사전 모수(hyperprior parameter)들을 고려할 필요가 없기 때문에 훨씬 더 간편하고 자동적이다. 따라서 FBF와 IBF를 '디폴트(default)'혹은 '자동적(automatic)'베이즈요인으로 부른다(Berger와 Mortera (1999)).

Berger와 Pericchi (1996)의 IBF는 최소시험표본(minimal training sample) $\boldsymbol{\mathcal{X}} = \{\boldsymbol{X}(l), l = 1, 2, \ldots, L\}$를 사용하여 부적절 분포를 적절한 사후분포로 전환시킨다. 최소시험표본은 모든 모형 M_j에 대해서 $0 < m_j(\boldsymbol{X}(l) \mid b = 1) < \infty$의 조건을 만족하는 모든 가능한 표본들 가운데 최소 크기의 표본들을 말한다.

O'Hagan (1995)의 FBF는 우도함수의 멱승 b을 사용하여 무정보 부적절 사전분포를 적절 사전분포로 변화시킬 수 있다. FBF는 다음과 같이 정의된다.

$$B_{ji}^{FBF} = B_{ji}(\boldsymbol{X} \mid b=1) \cdot B_{ij}(\boldsymbol{X} \mid b). \tag{2.4}$$

이때 상수 b의 값은 O'Hagan (1995)이 제시한 3가지 방법 중 사전분포의 불명확성에 대한 로버스트성(robustness)에 특별한 관심이 없을 때 사용하는 $b = \frac{m}{n}$(여기서 m은 최소시험표본의 크기)로 정의된 값을 흔히 사용한다.

모형에 M_j대한 각 모형의 사후확률(posterior probability)은 모형의 사전확률, p_j와 베이즈요인을 고려하여 다음과 같이 구할 수 있다.

$$P(M_i \mid \boldsymbol{X}) = \left\{ \sum_{j=1}^{q} \left(\frac{p_j}{p_i} \right) \cdot B_{ji}^{FBF} \right\}^{-1}, \quad i = 1, 2, \ldots, q.$$

최종적으로 각 모형에 대한 사후확률을 계산하여 가장 높은 사후확률을 갖는 모형을 선택한다.

3. 모형선택을 위한 부분적 베이즈요인의 계산

3.1. 포아송 대 음이항분포의 모형선택

식 (1.1)의 포아송 확률모형(M_1)과 음이항 확률모형(M_2)의 모수들에 대한 무정보 부적절 사전분포를 다음과 같이 정의하자(Berger와 Pericchi (1998)).

$$\begin{cases} \pi_1(\lambda) = \lambda^{-\alpha}, & \alpha \in [0,1], \quad \lambda > 0, \\ \pi_2(p) = p^{-\beta}(1-p)^{-\gamma}, & \beta, \gamma \in [0,1], \quad 0 < p < 1. \end{cases} \tag{3.1}$$

이제 $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$을 관측값이라 놓으면 각 모형의 우도함수는 다음과 같이 나타낼 수 있다.

$$\begin{cases} L_1(\lambda \mid \boldsymbol{x}) = \dfrac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}, \\ L_2(p \mid \boldsymbol{x}) = p^n (1-p)^{\sum_{i=1}^n x_i}. \end{cases} \tag{3.2}$$

각 모형에 대해서 식 (3.1)과 식 (3.2)을 이용하여 식 (2.2)을 계산한 결과는 다음과 같다.

$$\begin{cases} m_1(\boldsymbol{x} \mid b) = \dfrac{\Gamma(b\sum_{i=1}^n x_i - \alpha + 1)\; b n^{-(b\sum_{i=1}^n x_i - \alpha + 1)}}{(\prod_{i=1}^n x_i!)^b}, \\ m_2(\boldsymbol{x} \mid b) = \dfrac{\Gamma(bn - \beta + 1)\,\Gamma(b\sum_{i=1}^n x_i - \gamma + 1)}{\Gamma(b\sum_{i=1}^n x_i + bn - \beta - \gamma + 2)}. \end{cases} \tag{3.3}$$

이제 식 (3.3)을 식 (2.1), 식 (2.2)에 대입하여 식 (2.4)의 FBF를 계산할 수 있다.

3.2. 정규, 이중지수 대 코쉬분포의 모형선택

식 (1.2)의 정규 확률모형(M_1), 이중지수 확률모형(M_2), 코쉬 확률모형(M_3)의 모수들에 대한 무정보 부적절 사전분포를 다음과 같이 정의하자(Berger와 Pericchi (1998)).

$$\begin{cases} \pi_1(\sigma) = \sigma^{2\alpha-3}, & \alpha \in [0,1], \quad \sigma > 0, \\ \pi_2(\rho) = \rho^{\beta-2}, & \beta,\gamma \in [0,1], \quad \rho > 0, \\ \pi_3(\tau) = \tau^{1-2\gamma}, & \gamma \in [0,2], \quad \tau > 0. \end{cases} \tag{3.4}$$

이제 $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$을 관측값이라 놓으면 각 모형의 우도함수는 다음과 같이 나타낼 수 있다.

$$\begin{cases} L_1(\sigma \mid \boldsymbol{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n {x_i}^2}{2\sigma^2}\right), \\ L_2(\rho \mid \boldsymbol{x}) = (2\rho)^{-n} \exp\left(-\frac{\sum_{i=1}^n |x_i|}{\rho}\right), \\ L_3(\tau \mid \boldsymbol{x}) = \left(\frac{\tau}{n}\right)^{-n} \cdot \prod_{i=1}^n \left(\tau^2 + x_i^2\right)^{-1}. \end{cases} \tag{3.5}$$

각 모형에 대해서 식 (3.4)과 식 (3.5)을 이용하여 식 (2.2)을 계산한 결과는 다음과 같다.

$$\begin{cases} m_1(\boldsymbol{x} \mid b) = 2^{-\alpha}\, \pi^{-\frac{bn}{2}}\, \Gamma(\frac{bn}{2} - \alpha + 1)\, \left(b \sum_{i=1}^n {x_i}^2\right)^{-(\frac{bn}{2}-\alpha+1)}, \\ m_2(\boldsymbol{x} \mid b) = 2^{-bn}\, \Gamma(bn - \beta + 1)\, \left(b \sum_{i=1}^n |x_i|\right)^{-(bn-\beta+1)}, \\ m_3(\boldsymbol{x} \mid b) = \pi^{-bn} \int_0^\infty \tau^{1-2\gamma+bn} \cdot \prod_{i=1}^n \left(\tau^2 + x_i^2\right)^{-b} d\tau. \end{cases} \tag{3.6}$$

이제 식 (3.6)을 식 (2.1), 식 (2.2)에 대입하여 식 (2.4)의 FBF를 계산할 수 있다.

4. 모의실험 및 예제

앞 절에서 계산한 FBF의 유용성을 검토하기 위하여 모의실험과 실제자료분석을 수행하였다. 이와 함께 베이즈요인 중 Berger와 Pericchi (1996, 1998)의 IBF인 AIBF, MIBF, 기하 IBF(geometric IBF ; GIBF)와 비교 분석하였다. IBF의 계산식은 Bertolino와 Racugno (1996)의 Example 2와 Example 3의 계산식을 사용하였다. 사전분포의 정의식 (3.1)과 식 (3.4)에 나오는 α, β, γ는 Bertolino와 Racugno (1996) 그리고 Berger와 Pericchi (1998)의 값을 그대로 사용하였다. 모의실험 및 실제자료분석을 위하여 MATLAB (The MathWorks Inc., 1998)을 사용하였다.

4.1. 포아송 대 음이항분포

만약 사전분포 상수 α가 1일 때 자료 값이 0인 경우에는 최소시험표본을 구할 수 없으므로 자료에 0이 포함된 경우는 제외하여 구하였다. 또한 자료를 버리지 않고 모두 이용하기

위하여 α을 1에서 0.99로 아주 작게 변화시켰다. 식 (3.3)에서 최소시험표본크기는 $m = 1$이다. 두 개 모형의 비교이므로 사후확률의 값이 0.5보다 큰 모형이 적합한 모형으로 선택된다.

(모의실험) MATLAB의 확률변수 생성함수인 POISSRND와 NBINRND으로부터 각 30개의 포아송 확률변수와 음이항 확률변수를 생성하였다. 표 4.1은 평균이 1로서 같으며 분산은 각각 1과 2인 $Poi(1)$분포, $NB(1, 0.5)$분포에서 생성된 자료에 대한 결과를 나타내고, 표 4.2는 평균이 2로서 같으며 분산은 각각 2와 6인 $Poi(2)$분포, $NB(1, 1/3)$분포로부터 생성된 자료에 대하여 얻은 결과들을 나타낸 것이다. 실험의 정확도를 살펴보기 위하여 1000회 반복 실행하였다.
표 4.1과 표 4.2의 각 칸에 대한 첫 번째 줄은 참모형에 대한 사후확률의 평균을 나타내고, 두 번째 줄의 괄호 안에 제시된 것은 참모형에 대한 사후확률의 표준편차를 나타낸다. 먼저 표 4.1을 살펴보면 $Poi(1)$ 자료에서 $(\alpha, \beta, \gamma) = (0.99, 1, 0.5)$와 $(\alpha, \beta, \gamma) = (0.99, 0.5, 0.5)$일 때는 AIBF를 제외하고는 모두 참 모형을 선택하였다. 표 4.2는 사전분포의 변동에 상관없이 참 모형을 선택하였다. $\lambda = 1$에서 포아송 자료는 $\lambda = 2$인 자료보다 0을 더 많이 포함하고 있으므로 IBF중 특히 AIBF가 MIBF, GIBF보다 더 불안정하다는 것을 알 수 있었다. 표 4.1과 표 4.2로부터 분산비가 클수록 높은 확률로 참 모형을 선택함을 알 수 있다.

(예제 1) 다음의 자료는 Bertolino와 Racugno (1996) 그리고 Berger와 Pericchi (1998)에서 사용하였던 자료로서 $\boldsymbol{x}^{(1)}$과 $\boldsymbol{x}^{(2)}$ 는 평균이 1.2로 같다.

$$\boldsymbol{x}^{(1)} = (1,1,1,1,1,1,1,1,2,2)\ , \quad \boldsymbol{x}^{(2)} = (0,0,0,0,1,1,2,2,3,3).$$

이 자료에 대한 디폴트 베이즈요인들을 이용하여 계산한 각 모형의 사후확률의 결과는 표 4.3과 같다. 표 4.3을 살펴보면, 가정된 모든 사전분포들에 대하여 자료 $\boldsymbol{x}^{(1)}$는 M_1모형에 적합하지만 자료 $\boldsymbol{x}^{(2)}$는 대부분의 경우에서 모형 M_2의 사후확률이 대략 $0.4 \sim 0.6$으로 어느 모형에 우월하게 적합하다고 결론을 내리기는 힘들다. 그러나 FBF는 작은 확률의 차이이긴 하지만 M_1모형을 선택하고 있으며 AIBF와 GIBF는 $(\alpha, \beta, \gamma) = (0.99, 1, 0.5)$, $(\alpha, \beta, \gamma) = (0.99, 0.5, 0.5)$의 사전분포의 가정 하에서 높은 사후확률로 M_2모형을 선택하였다. $\boldsymbol{x}^{(2)}$자료에서 $\alpha = 1, \gamma = 0.5$일 경우에 $\alpha = 0.99$로 약간의 변화를 주었을 때 AIBF와 GIBF는 불안정한 상태가 됨을 알 수 있다. $\boldsymbol{x}^{(2)}$자료에는 0이 포함되고 있기 때문에 최소시험표본 중에서 포아송분포의 값이 작아져, AIBF와 GIBF에는 크게 영향을 미치나 MIBF는 크게 영향을 미치지 않기 때문에 더 안정적이다.

(예제 2) $\boldsymbol{x}^{(1)}$는 Lawless (1987)의 <표 1>에 있는 처리그룹1에서 쥐에 대한 종양의 수를 조사한 것이다. $\boldsymbol{x}^{(2)}$는 Barnwal과 Paul (1988)의 <표 2>에 있는 제어그룹과 두 처리집단에서 죽은 태아의 수를 조사한 것이다.

$$\boldsymbol{x}^{(1)} = (1,0,2,1,4,3,6,1,1,5,2,1,5,2,5,2,3,4,5,5,1,2,6,0,1),$$
$$\boldsymbol{x}^{(2)} = (7,2,1,0,0,5,4,0,1,0,4,2,3,0,1).$$

여기서 $\boldsymbol{x}^{(1)}$과 $\boldsymbol{x}^{(2)}$는 각각 포아송분포(모수 최우추정치 $\widehat{\lambda} = 2.6522$)와 음이항분포(모수

표 4.1: 디폴트 베이즈요인들에 의한 참 모형의 사후확률

Prior			$Poi(1)$ data				$NB(1,0.5)$ data			
α	β	γ	FBF	AIBF	MIBF	GIBF	FBF	AIBF	MIBF	GIBF
1	1	0.5	0.8605	0.8297	0.8297	0.8297	0.8305	0.8489	0.8510	0.8485
			(0.1689)	(0.1928)	(0.1928)	(0.1928)	(0.2428)	(0.2347)	(0.2324)	(0.2350)
0.99	1	0.5	0.8628	0.4800	0.7791	0.6744	0.8233	0.9689	0.9271	0.9378
			(0.1802)	(0.3016)	(0.2935)	(0.2892)	(0.2508)	(0.0974)	(0.1942)	(0.1519)
1	1	1	0.8605	0.8297	0.8297	0.8297	0.7899	0.8274	0.8274	0.8274
			(0.1689)	(0.1928)	(0.1928)	(0.1928)	(0.2572)	(0.2288)	(0.2288)	(0.2288)
0.99	1	1	0.8667	0.8374	0.8376	0.8374	0.7690	0.8059	0.8057	0.8059
			(0.1781)	(0.2012)	(0.2010)	(0.2012)	(0.2830)	(0.2618)	(0.2619)	(0.2618)
1	0.5	0.5	0.8493	0.8146	0.8160	0.8146	0.7978	0.8272	0.8258	0.8271
			(0.1703)	(0.1959)	(0.1947)	(0.1958)	(0.2710)	(0.2519)	(0.2526)	(0.2519)
0.99	0.5	0.5	0.7811	0.4535	0.7133	0.5900	0.7868	0.9708	0.8967	0.9332
			(0.2636)	(0.3455)	(0.3318)	(0.3457)	(0.2449)	(0.0516)	(0.1888)	(0.1085)
0	0	0	0.8128	0.7577	0.7819	0.7975	0.8250	0.8745	0.8315	0.8410
			(0.2099)	(0.2496)	(0.2269)	(0.2218)	(0.2203)	(0.1849)	(0.2138)	(0.2097)

(모의실험자료 : 공통평균 1, 분산비 1 : 2)

표 4.2: 디폴트 베이즈요인들에 의한 참 모형의 사후확률

Prior			$Poi(1)$ data				$NB(1,0.5)$ data			
α	β	γ	FBF	AIBF	MIBF	GIBF	FBF	AIBF	MIBF	GIBF
1	1	0.5	0.9646	0.9477	0.9490	0.9480	0.9470	0.9557	0.9549	0.9555
			(0.0829)	(0.1155)	(0.1130)	(0.1150)	(0.1792)	(0.1650)	(0.1667)	(0.1652)
0.99	1	0.5	0.9820	0.8951	0.9724	0.9520	0.9489	0.9957	0.9657	0.9850
			(0.0499)	(0.2094)	(0.0730)	(0.1191)	(0.1302)	(0.0135)	(0.0987)	(0.0439)
1	1	1	0.9578	0.9422	0.9422	0.9422	0.9746	0.9839	0.9839	0.9839
			(0.1244)	(0.1465)	(0.1465)	(0.1465)	(0.0736)	(0.0519)	(0.0519)	(0.0519)
0.99	1	1	0.9370	0.9196	0.9196	0.9196	0.9258	0.9433	0.9433	0.9433
			(0.1692)	(0.1980)	(0.1980)	(0.1980)	(0.1874)	(0.1616)	(0.1616)	(0.1616)
1	0.5	0.5	0.9640	0.9485	0.9497	0.9486	0.9458	0.9611	0.9602	0.9611
			(0.1073)	(0.1327)	(0.1312)	(0.1326)	(0.1575)	(0.1327)	(0.1344)	(0.1328)
0.99	0.5	0.5	0.9542	0.8585	0.9343	0.9085	0.9579	0.9939	0.9716	0.9847
			(0.1131)	(0.2612)	(0.1555)	(0.2024)	(0.1317)	(0.0238)	(0.0960)	(0.0566)
0	0	0	0.9470	0.9228	0.9349	0.9383	0.9457	0.9613	0.9516	0.9516
			(0.1460)	(0.1836)	(0.1681)	(0.1605)	(0.1701)	(0.1388)	(0.1614)	(0.1589)

(모의실험자료 : 공통평균 2, 분산비 1 : 3)

표 4.3: 디폴트 베이즈요인들에 의한 각 모형의 사후확률(예제 1)

Prior			$\boldsymbol{x}^{(1)}$ data : $P\left(M_1\|\boldsymbol{x}^{(1)}\right)$				$\boldsymbol{x}^{(2)}$ data : $P\left(M_2\|\boldsymbol{x}^{(2)}\right)$			
α	β	γ	FBF	AIBF	MIBF	GIBF	FBF	AIBF	MIBF	GIBF
1	1	0.5	0.9724	0.9706	0.9696	0.9707	0.4165	0.4953	0.4845	0.4939
0.99	1	0.5	0.9725	0.9708	0.9698	0.9708	0.4155	0.9395	0.5282	0.8075
1	1	1	0.9735	0.9724	0.9724	0.9724	0.4067	0.5056	0.5056	0.5056
0.99	1	1	0.9736	0.9725	0.9726	0.9725	0.4057	0.5059	0.5063	0.5059
1	0.5	0.5	0.9678	0.9660	0.9660	0.9660	0.4556	0.5644	0.5590	0.5643
0.99	0.5	0.5	0.9679	0.9662	0.9662	0.9662	0.4546	0.9304	0.5767	0.8228
0	0	0	0.9635	0.9611	0.9674	0.9627	0.4882	0.6293	0.5481	0.5332

표 4.4: 디폴트 베이즈요인들에 의한 각 모형의 사후확률(예제 2)

Prior			$\boldsymbol{x}^{(1)}$ data : $P\left(M_1\|\boldsymbol{x}^{(1)}\right)$				$\boldsymbol{x}^{(2)}$ data : $P\left(M_2\|\boldsymbol{x}^{(2)}\right)$			
α	β	γ	FBF	AIBF	MIBF	GIBF	FBF	AIBF	MIBF	GIBF
1	1	0.5	0.9375	0.8868	0.8883	0.8877	0.9590	0.9818	0.9811	0.9817
0.99	1	0.5	0.9377	0.6466	0.8888	0.8522	0.9589	0.9987	0.9845	0.9945
1	1	1	0.9383	0.8944	0.8944	0.8944	0.9582	0.9813	0.9813	0.9813
0.99	1	1	0.9384	0.8948	0.8949	0.8948	0.9581	0.9813	0.9813	0.9813
1	0.5	0.5	0.9283	0.8733	0.8825	0.8739	0.9645	0.9855	0.9847	0.9855
0.99	0.5	0.5	0.9284	0.7057	0.8757	0.8413	0.9644	0.9982	0.9862	0.9948
0	0	0	0.9235	0.8472	0.9050	0.8952	0.9676	0.9894	0.9717	0.9792

최우추정치 $\widehat{p} = 0.33$)를 따른다고 알려져 있다. $\boldsymbol{x}^{(1)}$의 평균은 2.6522, 분산은 3.7826이고, $\boldsymbol{x}^{(2)}$의 평균은 2, 분산은 4.7143이다. 이들 각 자료에 대해 디폴트 베이즈요인들을 이용하여 계산된 각 모형의 사후확률의 결과는 표 4.4와 같다. 표 4.4를 살펴보면, 자료 $\boldsymbol{x}^{(1)}$에 대해서는 모형 M_1이, 자료$\boldsymbol{x}^{(2)}$에 대해서는 모형 M_2가 적합함을 높은 사후확률로서 입증하고 있다.

4.2. 정규, 이중지수 대 코쉬분포

Bertolino와 Racugno (1996) 그리고 Berger와 Pericchi (1998)는 최소시험표본에 대해서 감마함수가 값을 갖도록 사전분포의 상수 α, β, γ값의 범위를 제한하였다. 식 (3.6)에서 최소시험표본크기는 $m = 1$이다. 식 (3.6)의 $m_3(\boldsymbol{x} \mid b)$에서 적분계산은 수치적분에 의해서 계산하였다.

(모의실험) 정규 확률변수는 MATLAB의 확률변수 생성함수인 NORMRND로부터 생성하였고, 이중지수 확률변수와 코쉬 확률변수는 RAND에서 난수를 생성하여 각 분포에 적합한 알고리즘에 의해 생성하였다. $N(0,1)$, $DE(0,1)$, $Cau(0,1)$분포로부터 생성한 각 크기 30인 자료를 500회 반복 시행하여 얻은 결과로서 표 4.5의 각 칸의 첫 번째 줄은 참 모형에 대한 사후확률의 평균을 나타내고, 두 번째 줄의 괄호 안에 제시된 것은 참 모형에 대한 사후확률의 표준편차를 나타낸다. 정규, 이중지수, 코쉬분포 모든 경우에서 참 모형을 잘 선택하고 있다. 특히 코쉬분포는 아주 높은 사후확률로서 M_3모형을 선택하였다.

(예제 3) 다음의 자료는 Bertolino와 Racugno (1996) 그리고 Berger와 Pericchi (1998)에서 사용하였던 자료이다.

$$\boldsymbol{x} = (-1, -0.4, -0.2, 0.001, 0.01, 0.1, 0.3, 1).$$

이 자료에 대하여 디폴트 베이즈요인들에 의한 각 모형의 사후확률의 결과는 표 4.6과 같다. 표 4.6을 살펴보면 $(\alpha, \beta, \gamma) = (1,1,1)$인 경우에 디폴트 베이즈요인들은 모형 M_2, 그리고 이외의 값들에서 AIBF는 모형 M_1을, FBF, MIBF와 GIBF는 모형 M_2을 선택하였다. 이 결과에서 AIBF는 사전분포의 변동에 민감하다는 것을 알 수 있다. 이와 같은 이유는 $\boldsymbol{x}$의 자료에 0.001나 0.01이 0에 가까운 수이기 때문에 최소시험표본을 구할 때는 MIBF, GIBF는 크게 변함이 없으나 AIBF는 크게 영향을 받기 때문에 불안정하다. 최소시험표본을 구할 때 0.001나 0.01를 제거하여 분석한 결과를 지면관계상 생략하였지만 모든 디폴트 베이즈요인들이 모형 M_2을 선택한다는 것을 알 수 있었다.

(예제 4) $\boldsymbol{x}^{(1)}$는 당근 한 상자의 무게를 잰 것으로 Hogg와 Tanis (2001, Example 4.4-9)에서 인용하였다. $\boldsymbol{x}^{(2)}$는 Bain과 Engelhardt (1973)의 두 지역간의 홍수량의 차이를 나타낸 자료 중에서 첫 번째 줄의 자료이다. $\boldsymbol{x}^{(3)}$는 Kouritzin (1998)의 <표 3>에 있는 소음추적 자료이다.

$$\begin{aligned}
\boldsymbol{x}^{(1)} &= (1.12, 1.13, 1.19, 1.26, 1.06, 1.31, 1.12, 1.23, 1.29, 1.17, 1.20, 1.11),\\
\boldsymbol{x}^{(2)} &= (1.97, 1.96, 3.60, 3.80, 4.79, 5.66, 5.76, 5.78, 6.27, 6.30, 6.76),\\
\boldsymbol{x}^{(3)} &= (27.07, 21.52, 14.45, 0.87, 57.27, 10.05, 14.332, 25.32, 28.176, 162.142).
\end{aligned}$$

여기서 $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$, $\boldsymbol{x}^{(3)}$는 각 문헌에서 각각 포아송분포, 이중지수분포, 코쉬분포을 따른다고 알려져 있다. 그러나 식 (1.2)의 설정된 모형들에서 위치 값을 0으로 고정하였으므 로 이 자료들의 위치 값을 0으로 표준화 시켰다. 즉, $\boldsymbol{x}^{(1)}$는 중위수 1.18을 뺀 값, $\boldsymbol{x}^{(2)}$는 중위수 5.66을 뺀 값 중 0을 포함하고 있으면 최소시험표본을 구할 수 없으므로 0인 값을 0.1로 대입한 것, $\boldsymbol{x}^{(3)}$는 중위수 23.42를 뺀 값이다. 변환된 자료를 이용하여 디폴트 베이즈요인들에 의한 각 모형의 사후확률의 결과는 표 4.7과 같다. 표 4.7을 살펴보면 $\boldsymbol{x}^{(1)}$는 모형 M_1을, $\boldsymbol{x}^{(2)}$는 모형 M_2을, 그리고 $\boldsymbol{x}^{(3)}$는 높은 사후확률을 가지고 모형 M_3을 선택하였다.

표 4.5: 디폴트 베이즈요인들에 의한 참 모형에서의 사후확률(모의실험자료)

Prior			$N(0,1)$ data			
α	β	γ	FBF	AIBF	MIBF	GIBF
1	1	1	0.7566	0.7215	0.7215	0.7215
			(0.2091)	(0.2253)	(0.2253)	(0.2253)
1	0.5	1.25	0.7296	0.7650	0.7164	0.7342
			(0.2196)	(0.2033)	(0.2205)	(0.2153)
1	0.75	0.75	0.7441	0.7377	0.7160	0.7244
			(0.2129)	(0.2068)	(0.2243)	(0.2222)
0.5	0.75	0.75	0.7745	0.7047	0.6831	0.6545
			(0.2021)	(0.2234)	(0.2456)	(0.2515)
Prior			$DE(0,1)$ data			
α	β	γ	FBF	AIBF	MIBF	GIBF
1	1	1	0.6262	0.6380	0.6380	0.6380
			(0.2127)	(0.2045)	(0.2045)	(0.2045)
1	0.5	1.25	0.5622	0.6048	0.6477	0.6350
			(0.1993)	(0.2176)	(0.2088)	(0.2121)
1	0.75	0.75	0.6620	0.5896	0.6629	0.6499
			(0.2159)	(0.2010)	(0.2086)	(0.2076)
0.5	0.75	0.75	0.6340	0.6222	0.7001	0.7089
			(0.2245)	(0.2012)	(0.2059)	(0.1984)
Prior			$Cau(0,1)$ data			
α	β	γ	FBF	AIBF	MIBF	GIBF
1	1	1	0.8480	0.8647	0.8647	0.8647
			(0.2860)	(0.2700)	(0.2700)	(0.2700)
1	0.5	1.25	0.8616	0.8707	0.8708	0.8708
			(0.2716)	(0.2664)	(0.2660)	(0.2660)
1	0.75	0.75	0.8361	0.8416	0.8593	0.8612
			(0.2983)	(0.2940)	(0.2782)	(0.2750)
0.5	0.75	0.75	0.8360	0.8415	0.8596	0.8616
			(0.2986)	(0.2943)	(0.2776)	(0.2742)

표 4.6: 디폴트 베이즈요인들에 의한 각 모형의 사후확률(예제 3)

Prior							
α	β	γ		FBF	AIBF	MIBF	GIBF
1	1	1	$P(M_1 \mid \boldsymbol{x})$	0.2664	0.1809	0.1809	0.1809
			$P(M_2 \mid \boldsymbol{x})$	0.4953	0.4792	0.4792	0.4792
			$P(M_3 \mid \boldsymbol{x})$	0.2383	0.3399	0.3399	0.3399
1	0.5	1.25	$P(M_1 \mid \boldsymbol{x})$	0.2849	0.4548	0.2054	0.2759
			$P(M_2 \mid \boldsymbol{x})$	0.5883	0.3328	0.4850	0.4420
			$P(M_3 \mid \boldsymbol{x})$	0.1268	0.2124	0.3096	0.2821
1	0.75	0.75	$P(M_1 \mid \boldsymbol{x})$	0.2716	0.6116	0.2096	0.2001
			$P(M_2 \mid \boldsymbol{x})$	0.5504	0.1582	0.5325	0.4262
			$P(M_3 \mid \boldsymbol{x})$	0.1781	0.2302	0.2578	0.3737
0.5	0.75	0.75	$P(M_1 \mid \boldsymbol{x})$	0.3117	0.6202	0.1322	0.0595
			$P(M_2 \mid \boldsymbol{x})$	0.5200	0.1661	0.5854	0.5012
			$P(M_3 \mid \boldsymbol{x})$	0.1682	0.2137	0.2823	0.4394

5. 맺음말

본 논문에서는 경쟁 모형의 모든 모수들에 대하여 무정보 사전분포를 가정하고, 포아송 대 음이항분포의 모형선택, 정규, 이중지수 대 코쉬분포의 모형선택을 위하여 O'Hagan (1995)의 FBF를 통한 각 모형의 사후확률을 계산하였다. 서로 다른 두 개 이상의 모형을 비교하기 위해 모의실험자료와 실제자료에 대해서 FBF뿐만 아니라 또 다른 디폴트 베이즈요인들인 AIBF, MIBF, GIBF를 이용하여 각 모형에 대해 반복 시행하여 구한 사후확률의 분포를 비교하였다. 포아송 대 음이항분포의 모형선택에서 표 4.1는 포아송분포가 사전분포가 변동될 때 AIBF가 불안정하였으나 표 4.2는 사전분포가 변동되어도 디폴트 베이즈요인에 상관없이 안정적이다는 것을 알 수 있었다. 정규, 이중지수 대 코쉬분포의 모형선택에서 표 4.6는 사전분포가 변동될 때 AIBF가 불안정하다. 여기에서 자료에 0이 포함된 경우는 최소시험표본에서 MIBF, GIBF는 크게 영향을 주지 않으나, AIBF는 영향을 주므로 불안정하다는 것을 알 수 있었다. 최소한 본 논문에서 분석한 모의실험이나 실제자료분석에서 FBF는 참 모형을 잘 선택하고 있다는 결론을 내릴 수 있었다.

표 4.7: 디폴트 베이즈요인들에 의한 추정 모형에서의 사후확률(예제 4)

Prior			$\boldsymbol{x}^{(1)}$ data				
α	β	γ		FBF	MIBF	GIBF	AIBF
1	1	1	$P(M_1 \mid \boldsymbol{x}^{(1)})$	0.7465	0.7123	0.7123	0.7123
			$P(M_2 \mid \boldsymbol{x}^{(1)})$	0.2171	0.2422	0.2422	0.2422
			$P(M_3 \mid \boldsymbol{x}^{(1)})$	0.0364	0.0455	0.0455	0.0455
1	0.5	1.25	$P(M_1 \mid \boldsymbol{x}^{(1)})$	0.7368	0.7471	0.7063	0.7277
			$P(M_2 \mid \boldsymbol{x}^{(1)})$	0.2393	0.2233	0.2593	0.2404
			$P(M_3 \mid \boldsymbol{x}^{(1)})$	0.0239	0.0293	0.0344	0.0319
1	0.75	0.75	$P(M_1 \mid \boldsymbol{x}^{(1)})$	0.7384	0.7264	0.7096	0.7167
			$P(M_2 \mid \boldsymbol{x}^{(1)})$	0.2351	0.2424	0.2597	0.2488
			$P(M_3 \mid \boldsymbol{x}^{(1)})$	0.0264	0.0312	0.0307	0.0345
0.5	0.75	0.75	$P(M_1 \mid \boldsymbol{x}^{(1)})$	0.7761	0.7163	0.7058	0.6680
			$P(M_2 \mid \boldsymbol{x}^{(1)})$	0.2013	0.2575	0.2631	0.2916
			$P(M_3 \mid \boldsymbol{x}^{(1)})$	0.0226	0.0262	0.0311	0.0404
Prior			$\boldsymbol{x}^{(2)}$ data				
α	β	γ		FBF	MIBF	GIBF	AIBF
1	1	1	$P(M_1 \mid \boldsymbol{x}^{(2)})$	0.3125	0.2343	0.2343	0.2343
			$P(M_2 \mid \boldsymbol{x}^{(2)})$	0.5144	0.5265	0.5265	0.5265
			$P(M_3 \mid \boldsymbol{x}^{(2)})$	0.1761	0.2393	0.2393	0.2393
1	0.5	1.25	$P(M_1 \mid \boldsymbol{x}^{(2)})$	0.2627	0.3316	0.2649	0.2850
			$P(M_2 \mid \boldsymbol{x}^{(2)})$	0.4798	0.4847	0.5331	0.5185
			$P(M_3 \mid \boldsymbol{x}^{(2)})$	0.2575	0.1837	0.2020	0.1965
1	0.75	0.75	$P(M_1 \mid \boldsymbol{x}^{(2)})$	0.3101	0.3720	0.2250	0.2571
			$P(M_2 \mid \boldsymbol{x}^{(2)})$	0.5551	0.4720	0.5617	0.5384
			$P(M_3 \mid \boldsymbol{x}^{(2)})$	0.1348	0.1560	0.1833	0.2044
0.5	0.75	0.75	$P(M_1 \mid \boldsymbol{x}^{(2)})$	0.3551	0.3419	0.1645	0.1399
			$P(M_2 \mid \boldsymbol{x}^{(2)})$	0.5189	0.5138	0.6300	0.6234
			$P(M_3 \mid \boldsymbol{x}^{(2)})$	0.1260	0.1443	0.2055	0.2367
Prior			$\boldsymbol{x}^{(3)}$ data				
α	β	γ		FBF	MIBF	GIBF	AIBF
1	1	1	$P(M_1 \mid \boldsymbol{x}^{(3)})$	0.0042	0.0013	0.0013	0.0013
			$P(M_2 \mid \boldsymbol{x}^{(3)})$	0.1420	0.0861	0.0861	0.0861
			$P(M_3 \mid \boldsymbol{x}^{(3)})$	0.8538	0.9126	0.9126	0.9126
1	0.5	1.25	$P(M_1 \mid \boldsymbol{x}^{(3)})$	0.0056	0.0021	0.0018	0.0017
			$P(M_2 \mid \boldsymbol{x}^{(3)})$	0.2107	0.0807	0.0807	0.0807
			$P(M_3 \mid \boldsymbol{x}^{(3)})$	0.7838	0.9172	0.9175	0.9175
1	0.75	0.75	$P(M_1 \mid \boldsymbol{x}^{(3)})$	0.0059	0.0914	0.0018	0.0019
			$P(M_2 \mid \boldsymbol{x}^{(3)})$	0.2183	0.0757	0.1013	0.1062
			$P(M_3 \mid \boldsymbol{x}^{(3)})$	0.7758	0.8329	0.8969	0.8920
0.5	0.75	0.75	$P(M_1 \mid \boldsymbol{x}^{(3)})$	0.0072	0.0921	0.0004	0.0005
			$P(M_2 \mid \boldsymbol{x}^{(3)})$	0.2180	0.0757	0.1014	0.1063
			$P(M_3 \mid \boldsymbol{x}^{(3)})$	0.7748	0.8322	0.8981	0.8932

참고문헌

[1] Hogg, R. V. and Tanis, E. A. (2001). *Probability and Statistical Inference.* Sixth Edition. Prentice-Hall, Inc.

[2] Bain, L. J. and Engelhardt, M. (1973). Interval Estimation for the Two-parameter Double Exponential Distribution. *Technometrics.* Vol. 15. 875-887.

[3] Barnwal, R. K. and Paul, S. R. (1988). Analysis of one-way layout of count data with negative binomial variation. *Biometrika.* Vol. 75. 215-222.

[4] Berger, J. O. and Mortera, J. (1999). Default Bayes Factor Non-Nested Hypothesis Testing. *Journal of the American Statistical Association.* Vol. 94. 542-554.

[5] Berger, J.O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection Prediction. *Journal of the American Statistical Association.* Vol. 91. 109-122.

[6] Berger, J.O. and Pericchi, L. R. (1998). Accurate and Stable Bayesian Model Selection: The Median Intrinsic Bayes Factor. *Sankaya B.* Vol. 60. 1-18.

[7] Bertolino, J. and Racugno, W. (1996). Is the Intrinsic Bayes Factor Intrinsic?. *Metron.* Vol. 54. 5-15.

[8] Kouritzin, M. A. (1998). On Exact Filters for Continuous Signals with Discrete Observations. *IEEE transactions on automatic control.* Vol. 43. 709-715.

[9] Lawless, J. F. (1987). Regression Methods for Poisson Process Data. *Journal of the American Statistical Association.* Vol. 82. 808-815.

[10] O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparisons. *Journal of the Royal Statistical Society B.* Vol. 57. 99-138.

[11] Spiegelhalter, D. J. and Smith, A. F. M. (1982). Bayes Factors for Linear and Log-Linear Models with Vague Prior Information. *Journal of the Royal Statistical Society, B.* Vol. 44. 377-387.

[12] The MathWorks Inc. (1998). *MATLAB/Statistics Toolbox, Version 5.3.* Natick, MA.

[2002년 10월 접수, 2003년 6월 채택]

Comparative Study of Model Selection Using Bayes Factor through Simulation : Poisson *vs.* Negative Binomial Model Selection and Normal, Double Exponential *vs.* Cauchy Model Selection

Mi Ra Oh [1] So Young Yun [2] Jung Wook Sim [3] Young Sook Son [4]

ABSTRACT

In this paper, we use Bayesian method for model selection of poisson *vs.* negative binomial distribution, and normal, double exponential *vs.* cauchy distribution. The fractional Bayes factor of O'Hagan (1995) was applied to Bayesian model selection under the assumption of noninformative improper priors for all parameters in the models. Through the analyses of real data and simulation data, we examine the usefulness of the fractional Bayes factor in comparison with intrinsic Bayes factors of Berger and Pericchi (1996, 1998).

Keywords: Noninformative improper prior distribution; Bayesian model selection; fractional Bayes factor; intrinsic Bayes factor.

1) Doctoral course, Dept of Statistics, Chonnam University, Kwangju, Korea
E-mail : omr@chonnam.ac.kr
2) Economic Survey Division, Gyunggi Branch, Korea National Statistical Office
E-mail : syyun@nso.go.kr
3) Professor, Dept of Statistics, Chonnam University, Kwangju, Korea
E-mail : jwsim@chonnam.ac.kr
4) Professor, Dept of Statistics, Chonnam University, Kwangju, Korea
E-mail : ysson@chonnam.ac.kr