Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.1 pp. 67 77

Bayesian Multiple Comparisons for *K*-Exponential Populations with Type-II Censored Data by Fractional Bayes Factors

문경애[1]), 김달호[2])

Abstract

We propose the Bayesian testing for the equality of K-exponential populations means with Type-II censored data. Specially we use the fractional Bayesian factors suggested by O'Hagan (1995) based on the noninformative priors for the parameters. And, we investigate the usefulness of the proposed Bayesian testing procedures via both real data analysis and simulations and compare the classical likelihood ratio(LR) test with the proposed Bayesian test.

주제어 : 2종 중단, 모형선택, 무정보적 사전분포, 부분 베이즈요인, 지수분포,

1. 서론

베이지안 다중검정방법 (multiple hypothesis test) 은 일반적으로 여러 통계모형에서 만족할만한 결과를 주는 것으로 알려져 있다. 베이지안 검정법은 고려중인 모형에 대한 사후확률을 계산하여 가장 높은 확률을 갖는 모형을 선택하기 때문에 귀무가설의 기각여부에만 관심을 가지는 고전적인 검정과는 달리 좀 더 구체적인 모형을 선택할 수 있다는 장점이 있다.

베이지안 검정에서는 모수에 대한 사전분포 (prior distribution) 를 가정해야 하는데 흔히 모형에 대한 사전정보의 부족이나 여러 가지 여건으로 인해 무정보적 사전분포를 가정하는 경우가 많다. 그러나 무정보적 사전분포 (noninformative prior) 는 부적절 분포 (improper distribution) 인 경우가 대부분이어서 이 때 얻은 베이즈요인 (Bayes factor) 은 임의의 상수에 영향을 받아 모형 선택에 어려움을 주었다. 많은 학자들이 이런 어려움을 해결하기 위한 방법을 연구하였다. (Geisser와 Eddy (1979), Spiegalhalter와 Smith (1982), San Martini와 Spezzaferri (1984)).

Berger와 Pericchi (1996,1998) 는 자료 y를 트레이닝 표본 (training sample) 이라

1. 강원도 동해시 지흥동 산 119번지 동해대학교 컴퓨터공학과 전임강사
E-mail : diana62@mail.donghae.ac.kr
2. 대구시 북구 산격동 370번지 경북대학교 통계학과 부교수

불리는 $y(l)$과 이를 제외한 나머지 표본 $y(-l)$으로 나누어 임의의 상수를 소거하는 방법인 내재적 베이즈요인 (intrinsic Bayes factor; IBF) 을, O'Hagan (1995) 은 부분 베이즈요인 (fractional Bayes factor;FBF) 을 제안하였다.

IBF는 최소 트레이닝 표본 (minimal training sample) 의 선택에 영향을 받게 되므로 안정성이 떨어진다. 이런 영향을 제거하기 위하여 제안된 산술 내재적 베이즈요인 (arithmetic IBF; AIBF) 은 소표본인 경우와 비내포 (non-nested) 모형인 경우에는 안정적이지 못하고 표본의 수가 큰 경우에는 계산 시간이 많이 걸린다는 단점이 있다. 이런 이유로 Berger와 Pericchi (1998) 는 비내포 모형이나 소표본에서도 잘 적용되는 중위수 내재적 베이즈요인 (median IBF; MIBF) 을 제안하였다. 그러나 내재적 베이즈요인은 근본적으로 표본의 재사용이라는 문제점을 안고 있다.

한편, O'Hagan (1995) 이 제안한 FBF는 IBF에 비해 계산 시간이 적게 걸릴 뿐만 아니라 비내포 모형인 경우에도 쉽게 이용할 수 있다는 점과 트레이닝 표본을 고려하지 않아도 된다는 장점이 있다. 또한 두 모형이 비내포 모형인 경우 두 모형 M_1과 M_2에 대한 산술 내재적 베이즈요인을 구할 때, M_1의 M_2에 대한 AIBF와 M_2의 M_1에 대한 AIBF의 값은 역수의 관계가 성립하지 않지만 부분 베이즈요인인 경우에는 역수의 관계가 성립한다는 장점도 있다.

본 논문에서는 2종 중단된 K개 지수분포의 모수에 대한 다중 검정 방법으로서 O'Hagan (1995) 이 제안한 부분 베이즈요인을 이용한 베이지안 검정법을 제안한다.

기존의 고전적인 검정법은 우도비 검정통계량 (likelihood ratio test statistic) 을 이용하여 K개 지수분포의 평균 상등에 관한 검정을 실시한다.

K개 지수분포의 표본의 크기는 각각 $n_i,\ i=1,2,\cdots,K$라 하고 각 지수분포는 모수 $\lambda_i,\ i=1,2,\cdots,K$를 가진다고 가정하자. 각 모집단에서의 2종 중단된 자료는 $y_{i1}\le y_{i2}\le\cdots\le y_{ir_i}$이고 나머지 (n_i-r_i)개의 자료는 y_{ir_i}와 같은 값을 가진다. 이때 우도비 검정통계량은 다음과 같이 주어진다.

$$\Lambda=\left(2\sum_{i=1}^{K}r_i\right)\log\hat{\theta}-2\sum_{i=1}^{K}\left(r_i\log\hat{\theta}_i\right).$$

여기에서 $\hat{\theta}=\sum_{i=1}^{K}T_i/\sum_{i=1}^{K}r_i,\quad \hat{\theta}_i=T_i/r_i$이고, $T_i=\sum_{j=1}^{r_i}y_{ij}+(n_i-r_i)y_{ir_i}$이다.

이 때 우도비 검정통계량 Λ는 근사적으로 자유도 $d.f.=(K-1)$을 가지는 카이제곱 (*Chi-square*) 분포를 따른다.

Dal Ho Kim, Sang Kil Kang 과 Seong W. Kim (2000), Seong W. Kim (2000), Seong W. Kim 과 D. Sun (2000) 은 2종 중단된 자료를 이용하여 지수분포를 따르는 두 모집단의 평균 비교에 대한 베이지안 검정법을 제안하였다. Seong W. Kim과 Hyunsoo Kim (2000) 은 지수분포를 따르는 두 집단의 평균을 비교하기 위한 방법으로 부분 베이즈요인의 값과 일반적인 베이즈요인의 값을 근사적으로 일치시키는 내재적 사전분포를 제안했고 Jongsig Bae, Hyunsoo Kim과 Seong W. Kim (2000) 은 정규분포를 따르는 두 모집단의 평균을 비교하기 위한 내재적 사전분포를 제안하였다.

일반적으로 K의 수가 작을 경우 $(K\ge 3)$ 에는 직접 사후확률을 계산하는 것이 가능하나 K의 수가 4 이상이 되면 사후확률을 계산하는 것이 매우 어려워진다. 예를

들어 $K=4$이면 가설의 수가 15개, $K=5$인 경우에는 52개, $K=6$인 경우는 203개에 이르는 가설의 사후확률을 계산하여야 한다. 이와 같이 모집단의 수가 1개씩만 증가하더라도 가설의 수는 엄청나게 증가하며 이들 가설에 대한 사후확률은 계산은 가능하나 힘든 작업이 될 것이다. 그러나 Gopalan 과 Berry (1998) 에 소개된 형태(configuration) 의 정의를 응용하여 사후확률을 계산한다면, 알고리즘 측면에서 간략하게 표현이 가능하리라 생각한다.

본 논문의 2절에서는 부분 베이즈요인에 대해 간략하게 소개하고 3절에서는 관심의 대상이 되는 가설에 대해 무정보적 사전분포 (noninformative prior) 를 이용한 베이지안 검정 방법을 제시한다. 마지막으로 4절에서는 제안한 검정 방법이 올바른 결정을 내리는지를 알아보기 위하여 실제자료와 모의실험을 통하여 그 결과를 알아보고 고전적인 검정법과 비교해 본다.

2. 부분 베이즈요인

$\boldsymbol{Y}=(Y_1, Y_2, \cdots, Y_n)$을 확률밀도함수 $f(\boldsymbol{y}\mid\theta)$를 갖는 모집단으로부터 추출된 확률표본이라고 하고 $\theta\in\Theta$이며, θ는 유한한 차원을 가진다고 가정하자.

베이지안 관점에서는 모형 $M_l: \theta\in\Theta_l,\ \Theta_l\subset\Theta,\ l=1,2,\cdots,Q$에 대한 모형선택(model selection)을 한다고 했을 때 모형 M_l에서의 모수 θ에 대한 사전분포와 모형 M_l가 참일 사전확률을 이용하여 모형 M_l가 사실일 사후확률을 계산하여 가장 높은 사후확률을 가지는 모형을 선택한다. 이 때의 사후확률은 다음과 같이 계산된다.

$$P(M_l\mid\boldsymbol{y})=\left(\sum_{u=1}^{Q}\frac{p_u}{p_l}B_{ul}\right)^{-1}. \tag{1}$$

여기에서 p_l는 모형 M_l가 참일 사전확률이며 B_{ul}는 모형 M_u의 모형 M_l에 대한 베이즈요인 (Bayes factor) 으로서 정의는 다음과 같다.

$$B_{ul}=\frac{m_u(\boldsymbol{y})}{m_l(\boldsymbol{y})}=\frac{\int_{\Theta_u}f(\boldsymbol{y}\mid\theta)\pi_u(\theta)d\theta}{\int_{\Theta_l}f(\boldsymbol{y}\mid\theta)\pi_l(\theta)d\theta}. \tag{2}$$

여기에서 $\pi_l(\theta)$는 모수 θ의 사전분포이고 $m_l(\boldsymbol{y})$는 모형 M_l에서의 $\boldsymbol{Y}$에 대한 주변확률밀도함수 (marginal or predictive density) 이다.

무정보적 사전분포를 $\pi_l^N(\theta)$라고 두면, 식(2)는

$$B_{ul}^N=\frac{m_u^N(\boldsymbol{y})}{m_l^N(\boldsymbol{y})}=\frac{\int_{\Theta_u}f(\boldsymbol{y}\mid\theta)\pi_u^N(\theta)d\theta}{\int_{\Theta_l}f(\boldsymbol{y}\mid\theta)\pi_l^N(\theta)d\theta} \tag{3}$$

이 되며, 이 때 식 (3)에는 부적절 분포의 사용으로 인한 임의의 상수가 포함되어 있다. 이런 어려움을 해결하기 위해 O'Hagan (1995) 이 제안한 부분 베이즈요인은 다음과 같이 정의한다.

[정의 1.] 모형 M_1의 모형 M_2에 대한 부분 베이즈요인은 다음과 같이 주어진다.

$$B^{b}_{12}(\mathbf{y}) = \frac{q_1(b,\mathbf{y})}{q_2(b,\mathbf{y})}. \quad (4)$$

여기에서 $l = 1,2$에 대하여

$$q_l(b,\mathbf{y}) = \frac{\int_{-\infty}^{\infty}\pi_l(\theta_i)f_l(\mathbf{y}\mid\theta_l)d\theta_l}{\int_{-\infty}^{\infty}\pi_l(\theta_l)[f_l(\mathbf{y}\mid\theta_l)]^b d\theta_l} \quad (5)$$

이며 b는 미리 정해진 상수이다. 위의 식 (4)를 이용하여 부분 베이즈요인을 계산하면 θ_l에 대한 사전분포가 미지의 상수를 가지는 부적절 분포가 되더라도, 미지의 상수는 서로 상쇄되어 나타나지 않음을 알 수 있다.

O'Hagan (1995) 는 식 (4)와 식 (5)에 포함되어 있는 상수 b를 선택하는 방법으로 연구자의 사전분포에 대한 관심에 따라 세 가지 방법을 제시하였다. 이 논문에서는, n을 표본의 크기, m을 최소 트레이닝 표본의 크기라 한다면, $b = m/n$인 방법 사용하기로 한다. 이 경우는 부분 베이즈요인을 계산할 때 널리 쓰이는 방법이기도 하다.

3. 2종 중단된 지수분포의 베이지안 다중검정

모수 λ를 가지는 지수분포의 확률밀도함수는 다음과 같이 주어진다.

$$f(y) = \lambda\exp(-\lambda y), \qquad y\ge 0,\ \lambda>0. \quad (6)$$

2종 중단된 자료는 n개의 자료 중에서 처음으로 관찰되는 r개의 관찰값으로써 $y_1 \le y_2 \le \cdots \le y_r$로 나타내고 나머지 $(n-r)$개의 자료는 y_r과 같은 값을 가진다. 그러면 관찰값 $(Y_1, Y_2, \cdots, Y_r)$의 결합확률밀도함수와 Y_p, $1\le p\le r$의 주변확률밀도함수는 다음과 같이 주어진다.

$$f(y_1,\cdots,y_r\mid\lambda) = \frac{n!}{(n-r)!}\lambda^r\exp(-\lambda(\sum_{i=1}^{r}y_i + (n-r)y_r)), \quad (7)$$

$$f(y_p\mid\lambda) = \frac{n!}{(p-1)!(n-p)!}\lambda(\exp(-\lambda y_p))^{n-p+1}(1-\exp(-\lambda y_p))^{p-1}. \quad (8)$$

K개 지수분포의 표본의 크기는 각각 n_i, $i=1,2,\cdots,K$라 하고 각 지수분포는 모수 λ_i, $i=1,2,\cdots,K$를 가진다고 가정했을 때, 각 모집단에서 2종 중단된 자료는 $y_{i1}\le y_{i2}\le\cdots\le y_{ir_i}$이고 나머지 $(n_i - r_i)$개의 자료는 y_{ir_i}이다.

모수 λ_i, $i=1,2,\cdots,K$를 가지는 독립인 K개의 2종 중단된 지수분포에서 각 모집단의 평균을 서로 비교한다고 가정했을 때 관심의 대상이 되는 모형은 다음과 같다.

M_1: $\lambda_1 = \lambda_2 = \cdots = \lambda_K$,

M_2: $\lambda_1 \neq \lambda_2, \lambda_1 = \lambda_3 = \cdots = \lambda_K$,

$\vdots$

M_Q: $\lambda_1 \neq \lambda_2 \neq \cdots \neq \lambda_K$.

이 때 가능한 모형의 수는 다음과 같이 주어진다.

$Q = B_K$, $K \geq 2$,

여기에서 $B_{t+1} = \sum_{i=0}^{t} \binom{t}{i} B_i$, $t = 0, 1, 2, \cdots$, $B_0 = 1$이다.

위의 각 모형에 Gopalan 과 Berry (1998) 의 형태 (configuration) 의 개념을 도입하여 사전분포와 우도함수를 나타내기로 한다.

K개의 모수 λ_i를 임의의 p개의 서로 다른 그룹으로 분류하기 위하여 $S = \{s_1, s_2, \cdots, s_K\}$라 두자. 이때 s_i, $i = 1, 2, \cdots, K$는 같은 그룹에 속하는 모수를 나타내는 첨자를 의미한다. 예를 들어 $K = 5$인 경우 $S = \{1, 2, 1, 2, 3\}$으로 주어졌다면 이는 $\lambda_1 = \lambda_3 \neq \lambda_2 = \lambda_4 \neq \lambda_5$인 모형을 의미하며 모수가 서로 다른 세 그룹($p = 3$)이 존재한다는 뜻이다.

서로 다른 p개의 그룹에 속하는 모수를 $\lambda_1^*, \lambda_2^*, \cdots, \lambda_p^*$라고 두고 각 그룹에 속하는 각 그룹에 속하는 첨자들의 집합을 I_j, 중단된 표본의 크기를 $N(I_j)$, 첨자의 크기를 $n(I_j)$라 한다면 $i = 1, 2, \cdots, K$, $j = 1, 2, \cdots, p$, $p = 1, 2, \cdots, K$에 대해서 다음과 같이 주어진다.

$$I_j = \{ i \mid s_i = j \},$$

$$N(I_j) = \sum_{i \in I_j} r_i,$$

$$n(I_j) = \text{total number of } i, \; i \in I_j.$$

이때 모형이 $\lambda_1 = \lambda_3 \neq \lambda_2 = \lambda_4 \neq \lambda_5$으로 주어졌다면 $\lambda_1 = \lambda_3 = \lambda_1^*$, $\lambda_2 = \lambda_4 = \lambda_2^*$, $\lambda_5 = \lambda_3^*$이 되고 각 그룹에 속하는 첨자들은 각각 $I_1 = \{1, 3\}$, $I_2 = \{2, 4\}$, $I_3 = \{5\}$로 주어진다. 또한 각 그룹의 표본의 크기는 각각 $N(I_1) = r_1 + r_3$, $N(I_2) = r_2 + r_4$, $N(I_3) = r_5$ 이며 $n(I_1) = 2$, $n(I_2) = 2$, $n(I_3) = 1$이 됨을 알 수 있다.

K개의 2종 중단된 지수분포의 모수를 비교하기 위해 앞에서 제시한 모형들을 검정하기 위한 사전분포로 다음과 같은 무정보적 사전분포를 가정한다.

$$\pi^N(\lambda_1^*, \lambda_2^*, \cdots, \lambda_p^*) = \frac{1}{\lambda_1^* \lambda_2^* \cdots \lambda_p^*}, \quad p = 1, 2, \cdots, K. \tag{9}$$

K개의 모수가 임의의 p개의 그룹으로 나누어진다고 가정했을 때 식 (7)을 이용하여 우도함수를 구해보면 다음과 같다.

$$L(\lambda_1^*, \lambda_2^*, \cdots, \lambda_p^*) = \prod_{i=1}^{K} \frac{n_i!}{r_i!(n_i - r_i)!} \times \prod_{k=1}^{p} \lambda_k^{*N(I_k)} \times \exp\Big[-\sum_{k=1}^{p}\Big[\lambda_k^* \sum_{i \in I_k}\Big(\sum_{j=1}^{r_i} y_{ij} + (n_i - r_i) y_{ir_i}\Big)\Big]\Big]. \tag{10}$$

먼저 식 (9)**에 주어진 사전분포와 식** (10)**의 우도함수를 이용하여 부분 베이즈요인을 구하기 위한 식**(5)**의 분자식을 구한다**.

[Lemma 1.] K **개의 모수가 임의의** p **개의 그룹으로 나누어진다고 가정할 때 부분 베이즈요인을 구하기 위한 식** (5)**의 분자식은 다음과 같다**.

$$\prod_{i=1}^{K} \frac{n_i!}{r_i!(n_i - r_i)!} \times \prod_{k=1}^{p} \frac{\Gamma(N(I_k))}{\Big[\sum_{i \in I_k}\Big(\sum_{j=1}^{r_i} y_{ij} + (n_i - r_i) y_{ir_i}\Big)\Big]^{N(I_k)}}. \tag{11}$$

Proof.

$$\int_0^\infty \int_0^\infty \cdots \int_0^\infty L(\lambda_1^*, \cdots, \lambda_p^*) \cdot \pi^N(\lambda_1^*, \cdots, \lambda_p^*) d\lambda_1^* \cdots d\lambda_p^*$$

$$= \prod_{i=1}^{K} \frac{n_i!}{r_i!(n_i - r_i)!} \times \prod_{k=1}^{p} \int_0^\infty \lambda_k^{*N(I_k)-1} \times \exp\Big[-\sum_{k=1}^{p}\Big[\lambda_k^* \sum_{i \in I_k}\Big(\sum_{j=1}^{r_i} y_{ij} + (n_i - r_i) y_{ir_i}\Big)\Big]\Big] d\lambda_k^*$$

$$= \prod_{i=1}^{K} \frac{n_i!}{r_i!(n_i - r_i)!} \prod_{k=1}^{p} \frac{\Gamma(N(I_k))}{\Big[\sum_{i \in I_k}\Big(\sum_{j=1}^{r_i} y_{ij} + (n_i - r_i) y_{ir_i}\Big)\Big]^{N(I_k)}}.$$

다음으로 K **개의 모수가 임의의** p **개의 그룹으로 나누어질 때 부분 베이즈요인을 구하기 위한 식** (5)**의 분모식을 구한다**.

[Lemma 2.] K **개의 모수가 임의의** p **개의 그룹으로 나누어진다고 가정할 때 부분 베이즈요인을 구하기 위한 식** (5)**의 분모식은 다음과 같다**.

$$\Big[\prod_{i=1}^{K} \frac{n_i!}{r_i!(n_i - r_i)!}\Big]^b \times \prod_{k=1}^{p} \frac{\Gamma(bN(I_k))}{\Big[b \sum_{i \in I_k}\Big(\sum_{j=1}^{r_i} y_{ij} + (n_i - r_i) y_{ir_i}\Big)\Big]^{bN(I_k)}}. \tag{12}$$

Proof.

$$\int_0^{\infty}\int_0^{\infty}\cdots\int_0^{\infty}[L(\lambda_1^*,\cdots,\lambda_p^*)]^b\cdot\pi^N(\lambda_1^*,\cdots,\lambda_p^*)d\lambda_1^*\cdots d\lambda_p^*$$

$$=\left[\prod_{i=1}^{K}\frac{n_i!}{r_i!(n_i-r_i)!}\right]^b\prod_{k=1}^{p}\int_0^{\infty}\lambda_k^{*N(I_k)b-1}$$

$$\times\exp[-b\lambda_k^*\sum_{i\in I_k}(\sum_{j=1}^{r_i}y_{ij}+(n_i-r_i)y_{ir_i})]d\lambda_k^*$$

$$=\left[\prod_{i=1}^{K}\frac{n_i!}{r_i!(n_i-r_i)!}\right]^b\prod_{k=1}^{p}\frac{\Gamma(bN(I_k))}{[b\sum_{i\in I_k}(\sum_{j=1}^{r_i}y_{ij}+(n_i-r_i)y_{ir_i})]^{bN(I_k)}}.$$

임의의 두 모형

(1) M_l : K **개의 모수가 임의의** p_1**개의 서로 다른 그룹으로 나누어진다**.

(2) M_u : K **개의 모수가 임의의** p_2**개의 서로 다른 그룹으로 나누어진다**.

라고 가정하자.

[Theorem 1.] 모형 M_l**에 대한** M_u, $l,u=1,2,\cdots,Q$**에 대한 부분 베이즈요인은 다음과 같이 구해진다**.

$$B_{ij}^b(\boldsymbol{y})=\frac{\prod_{k_1=1}^{p_1}\Gamma(N(I_{k_1}))\prod_{k_2=1}^{p_2}\Gamma(bN(I_{k_2}))}{\prod_{k_2=1}^{p_2}\Gamma(N(I_{k_2}))\prod_{k_1=1}^{p_1}\Gamma(bN(I_{k_1}))}$$
$$\times\frac{\prod_{k_1=1}^{p_1}[\sum_{i_1\in I_{k_1}}(\sum_{j_1=1}^{r_{i_1}}y_{i_1j_1}+(n_{i_1}-r_{i_1})y_{i_1r_{i_1}})]^{N(I_{k_1})(b-1)}}{\prod_{k_2=1}^{p_2}[\sum_{i_2\in I_{k_2}}(\sum_{j_2=1}^{r_{i_2}}y_{i_2j_2}+(n_{i_2}-r_{i_2})y_{i_2r_{i_2}})]^{N(I_{k_2})(b-1)}}. \quad (13)$$

식 (13)**은** Lemma 1**과** Lemma 2**에 의해 주어진 식** (11)**과 식** (12)**을 이용하여 얻을 수 있다**.

4. 모의 실험

4**절에서는 모집단을 세 개로 가정하고 모의실험을 행하여 제안한 방법의 유용성을 알아보고 고전적인 검정법을 이용하여 얻은 결과와 비교해 보기로 한다**.

세 개의 모집단을 각각 $\exp(\lambda_1)$**과** $\exp(\lambda_2)$, $\exp(\lambda_3)$**이라 하자**. **이 때 고려할 수 있는 모형은 다음과 같이 주어진다**.

$M_1:\lambda_1=\lambda_2=\lambda_3,\quad M_2:\lambda_1=\lambda_2\neq\lambda_3,\quad M_3:\lambda_1=\lambda_3\neq\lambda_2,$

$M_4:\lambda_1\neq\lambda_2=\lambda_3,\quad M_5:\lambda_1\neq\lambda_2\neq\lambda_3.$

즉 모집단의 수는 $K+3$**이고 검정 가능한 모형의 수는** $Q=5$**이다**.

[예제 1] 다음에 주어진 자료는 보잉 720 **제트 비행기에서 냉방장치 고장의 시간간격을 나타내는 자료이다** (Proschan 1963) . **여기에서 각 비행기의 고장시간 간격은 독**

립이면서 지수분포를 따른다고 가정한다. 세 집단의 표본의 크기는 각각 $n_1 = 24$, $n_2 = 16$, $n_3 = 15$이고 $r_1 = 20$, $r_2 = 13$, $r_3 = 12$로 가정하였다.

Plane 1	3,5,5,13,14,15,22,22,23,30,36,39,44,46,50,72,79,88,97,102, 139,188,197,210
Plane 2	14,14,27,32,34,54,57,59,61,66,67,102,134,152,209,230
Plane 3	12,21,26,27,29,29,48,57,59,70,74,153,326,386,502

먼저 고전적인 가설 검정법인 우도비 검정을 이용하여 위에 주어진 가설을 검정해 본 결과 p-값이 0.48020이 되어 모형 M_1이 선택됨을 알 수 있었다.

다음으로 표 1에는 각 모형에 대한 사전확률을 1/5로 가정하고 앞에서 제안한 부분 베이즈요인을 이용하여 계산한 결과를 나타내었다. 고전적인 검정법에서와 마찬가지로 모형 1, 즉 $M_1 : \lambda_1 = \lambda_2 = \lambda_3$이 선택됨을 알 수 있다.

표 1. 모형 M_l, $l = 1,2,3,4,5$ 대한 사후확률

	$P\{M_1 \mid \mathbf{y}\}$	$P\{M_2 \mid \mathbf{y}\}$	$P\{M_3 \mid \mathbf{y}\}$	$P\{M_4 \mid \mathbf{y}\}$	$P\{M_5 \mid \mathbf{y}\}$
FBF	.46602	.13472	.12902	.22036	.04989

[예제 2] 각 모형에 제안한 검정 방법이 올바른 결정을 내리는지를 알아보기 위하여 모의실험을 실시하였는데 이 때의 계산은 100번 반복으로 이루어졌으며 각 모형에 대한 사전확률은 1/5로 가정하였다. 또한 모수 $(\lambda_1, \lambda_2, \lambda_3)$의 값은 (5,5,5), (5,5,10), (5,10,5), (5,10,10) 그리고 (5,10,20)으로 가정하고 각각 10%와 20%의 자료가 중단되었다고 가정하였다. 표 2와 3은 중단율이 각각 10%와 20%일 경우, 앞에서 제시한 FBF를 이용하여 각 모형 M_l, $l = 1,2,3,4,5$에 대한 사후확률을 계산해 놓은 결과이다. 표 2와 3으로부터, 제안한 FBF를 이용하면 표본의 크기가 비교적 작은 경우에도 올바른 모형을 선택한다는 것을 알 수 있었다.

참고문헌

1. Berger, J. O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of American Statistical Association*, 91,109-122.
2. Berger, J. O. and Pericchi, L. R. (1998). Accurate and Stable Bayesian Model Selection : The Median Intrinsic Bayes Factor, *Sankhya*, B, 91,1-18.
3. Dal Ho Kim, Sang Kil Kang, and Seong. W. Kim (2000). Intrinsic Bayes Factor for Exponential Model Comparison with Censored Data, *Journal of the Korean Statistical Society*, 29:1, 123-135.
4. Geisser, B. and Eddy, W.F. (1979). A Predictive Approach to Model Selection, *Journal of the American Statistical Association*, 74, 153-160.
5. Gopalan, R. and Berry, D.A. (1998). Bayesian Multiple Comparisons Using Dirichlet Process Priors, *Journal of American Statistical Association*, 93, 1130-1139.
6. Jongsig Bae, Hyunsoo Kim and Seong. W. Kim (2000). Intrinsic Priors for

Testing Two Normal Means with the Default Bayes Factors, *The Journal of Korean Statistical Society*, 29:4, 443-454.

7. O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparison, *Journal of the Royal Statistical Society*, B, 57, 99-138.
8. Proschan, F. (1063). Theoretical Explanation of Observed Decreasing Failure Rate, *Technometrics* 5, 375-383.
9. San Martini, A. and Spezzaferri, F. (1984). A Predictive Model Selection Criterion, *Journal of Royal Statistical Society, B*, 46, 296-303.
10. Seong W. Kim (2000). Intrinsic priors for testing exponential means, *Statistics and Probability Letters*, 46, 195-201.
11. Seong W. Kim and D. Sun (2000). Intrinsic priors for model selection using an encompassing model with application to censored failure time data, *Lifetime Data Analysis*, 6, No. 3, 251-269.
12. Seong W. Kim and Hyunsoo Kim (2000). Intrinsic Priors for Testing Exponential Means with the Fractional Bayes Factor, *The Journal of Korean Statistical Society*, 29:4, 395-405.
13. Spiegelhalter, D.J. and Smith, A.F.M. (1982). Bayes Factor for Linear and Log-linear Models with Vague Prior Information, *Journal of Royal Statistical Society, Ser. B*, 44, 377-387.

표 2. FBF를 이용한 $M_l,\ l = 1, 2, 3, 4, 5$에 대한 사후확률(중단율10%)

$(\lambda_1, \lambda_2, \lambda_3)$	(n_1, n_2, n_3)	$P\{M_1 \mid \mathbf{y}\}$	$P\{M_2 \mid \mathbf{y}\}$	$P\{M_3 \mid \mathbf{y}\}$	$P\{M_4 \mid \mathbf{y}\}$	$P\{M_5 \mid \mathbf{y}\}$
(5,5,5)	(10,10,10)	.39466	.18623	.16951	.17343	.07617
	(20,10,10)	.40881	.17181	.17389	.17474	.07075
	(20,20,10)	.42628	.17205	.16963	.16858	.06346
	(20,20,20)	.45818	.15698	.17645	.15371	.05468
	(30,20,20)	.48693	.15575	.16545	.14332	.04854
	(30,30,20)	.48674	.14745	.16199	.15739	.04643
	(30,30,30)	.49123	.13955	.15479	.16787	.04656
(5,5,10)	(10,10,10)	.25033	.27576	.18176	.15179	.14036
	(20,10,10)	.29497	.30339	.13111	.15753	.11300
	(20,20,10)	.27412	.33908	.13487	.12600	.12592
	(20,20,20)	.19185	.43644	.13160	.09974	.14038
	(30,20,20)	.18159	.46791	.08464	.11043	.15543
	(30,30,20)	.18621	.47524	.10970	.08595	.14291
	(30,30,30)	.16227	.50034	.07654	.11964	.14121
(5,10,5)	(10,10,10)	.26550	.15027	.29655	.15749	.13018
	(20,10,10)	.27101	.12191	.34044	.13941	.12723
	(20,20,10)	.17249	.10755	.41441	.15609	.14946
	(20,20,20)	.20452	.12704	.42731	.10305	.13807
	(30,20,20)	.19482	.10851	.42006	.12744	.14918
	(30,30,20)	.13897	.07837	.53368	.10086	.14813
	(30,30,30)	.14704	.07902	.53565	.09892	.13937
(5,10,10)	(10,10,10)	.22290	.15428	.15775	.30895	.15611
	(20,10,10)	.20158	.15551	.14191	.36017	.14083
	(20,20,10)	.15672	.09715	.13860	.45410	.15344
	(20,20,20)	.17166	.09725	.10610	.46824	.15674
	(30,20,20)	.14806	.10626	.11941	.46742	.15885
	(30,30,20)	.15499	.08093	.14850	.46367	.15192
	(30,30,30)	.11804	.08069	.08260	.54619	.17248
(5,10,20)	(10,10,10)	.13605	.33637	.05996	.24938	.21824
	(20,10,10)	.07124	.29881	.03387	.35815	.23793
	(20,20,10)	.07860	.22172	.03038	.38365	.28565
	(20,20,20)	.02222	.30648	.00720	.32869	.33540
	(30,20,20)	.01954	.27085	.00541	.35496	.34924
	(30,30,20)	.01313	.26737	.00410	.29064	.42476
	(30,30,30)	.00561	.28341	.00117	.22015	.48967

표 3. FBF를 이용한 $M_l,\ l=1,2,3,4,5$에 대한 사후확률(중단율 20%)

$(\lambda_1,\lambda_2,\lambda_3)$	(n_1,n_2,n_3)	$P\{M_1 \mid \boldsymbol{y}\}$	$P\{M_2 \mid \boldsymbol{y}\}$	$P\{M_3 \mid \boldsymbol{y}\}$	$P\{M_4 \mid \boldsymbol{y}\}$	$P\{M_5 \mid \boldsymbol{y}\}$
(5,5,5)	(10,10,10)	.35815	.19191	.18656	.17278	.09060
	(20,10,10)	.36676	.18461	.18986	.16849	.09028
	(20,20,10)	.44290	.17315	.16778	.15507	.06109
	(20,20,20)	.44560	.16379	.16279	.16693	.06088
	(30,20,20)	.45238	.16188	.15805	.17068	.05701
	(30,30,20)	.47047	.15636	.15504	.16672	.05140
	(30,30,30)	.46638	.17474	.16713	.13921	.05254
(5,5,10)	(10,10,10)	.22978	.31320	.16568	.14688	.14445
	(20,10,10)	.25023	.30572	.12782	.17663	.13961
	(20,20,10)	.26640	.33142	.14613	.12872	.12733
	(20,20,20)	.22461	.42484	.10589	.11487	.12979
	(30,20,20)	.19088	.41446	.09970	.15030	.14467
	(30,30,20)	.19794	.47182	.10159	.09413	.13452
	(30,30,30)	.16408	.50941	.08778	.09828	.14045
(5,10,5)	(10,10,10)	.24843	.15225	.29966	.15417	.14549
	(20,10,10)	.26700	.12535	.31502	.16299	.12964
	(20,20,10)	.18808	.09636	.38923	.18524	.14108
	(20,20,20)	.21216	.11537	.42268	.11351	.13629
	(30,20,20)	.18618	.07502	.46393	.12451	.15036
	(30,30,20)	.16361	.09665	.45148	.14655	.14171
	(30,30,30)	.12548	.08362	.53965	.08615	.16509
(5,10,10)	(10,10,10)	.24462	.16541	.14899	.30282	.13816
	(20,10,10)	.22313	.15466	.15785	.32299	.14138
	(20,20,10)	.18947	.11383	.16459	.38962	.14250
	(20,20,20)	.21191	.11778	.13108	.40747	.13176
	(30,20,20)	.14436	.09583	.13249	.47556	.15176
	(30,30,20)	.18197	.09275	.12948	.46484	.13096
	(30,30,30)	.16201	.08837	.09283	.50068	.15611
(5,10,20)	(10,10,10)	.09889	.30615	.04727	.29946	.24824
	(20,10,10)	.05437	.27170	.02867	.38387	.26140
	(20,20,10)	.06085	.26740	.03368	.37410	.26397
	(20,20,20)	.02762	.33015	.00914	.28209	.35099
	(30,20,20)	.01054	.27369	.00430	.35035	.36113
	(30,30,20)	.02292	.20143	.00632	.32916	.44017
	(30,30,30)	.00380	.24481	.00104	.30353	.44681