응용통계연구 제 **14**권 2호, *2001*년, *pp. 449–462*

부분 베이즈요인을 이용한 K개 로그정규분포의 상등에 관한 베이지안 다중검정

문경애[1)] 김달호[2)]

요 약

베이지안 다중검정방법 (multiple hypothesis test)은 여러 통계모형에서 성공적인 결과를 주는 것으로 알려져 있다. 일반적으로, 베이지안 가설검정은 고려중인 모형에 대한 사후확률을 계산하여 가장 높은 확률을 갖는 모형을 선택하기 때문에 귀무가설의 기각 여부에만 관심을 가지는 고전적인 분산분석 검정과는 달리 좀 더 구체적인 모형을 선택할 수 있는 장점이 있다. 이 논문에서는 독립이면서 로그정규분포를 따르는 $K(\geq 3)$개 모집단의 모수에 대한 가설 검정방법으로 O'Hagan (1995)이 제안한 부분 베이즈 요인을 이용한 베이지안 방법을 제안한다. 이 때 모수에 대한 사전분포로는 무정보적 사전분포를 사용한다. 제안한 검정 방법의 유용성을 알아보기 위하여 실제 자료의 분석과 모의실험을 이용하여 고전적인 검정방법과 그 결과를 비교한다.

주요용어: 베이지안 다중검정, 부분베이즈요인, 형태, 무정보적 사전분포, 모형선택.

1. 서론

신뢰수명검정모형이나 가속수명시험모형 등에 널리 이용되는 로그정규분포의 확률밀도함수는 다음과 같다.

$$f(x|\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{1}{2\sigma^2}(\log x-\mu)^2\right\},\ 0<x<\infty. \tag{1.1}$$

여기서 $-\infty<\mu<\infty,\ 0<\sigma$이다. 식 (1.1)을 확률밀도함수로 가지는 확률변수를 $X\sim LN(\mu,\sigma^2)$으로 나타내자. 일반적으로 로그정규분포를 따르는 확률변수에 로그변환을 하면 평균 μ와 분산 σ^2을 가지는 정규분포를 따르게 된다. 이런 이유로 로그정규분포의 모수에 대한 추정과 검정은 주로 정규분포를 이용하여 다루어진다. 그러나 베이지안 추정과 검정 문제에서는 최우추정법의 불변성이 만족되지 않기 때문에 로그정규분포에 대한 베이지안 검정 방법을 다루어 보고자 한다. 또한 세 개이상의 로그정규모집단의 모수비교에 대한 베이지안 검정은 아직 연구된 바가 없다.

독립인 두 모집단의 모수의 동일성에 대한 베이지안 가설검정과 관련한 연구를 간략하게 소개하면 다음과 같다. Dal Ho Kim, Sang Gil Kang 과 Seong W. Kim (2000)은 제2종 중

1) (240-713) 강원도 동해시 지흥동 산 119번지, 동해대학교 컴퓨터공학과, 전임강사
E-mail: diana62@mail.donghae.ac.kr
2) (702-701) 대구광역시 북구 산격동 370번지, 경북대학교 통계학과, 부교수
E-mail: dalkim@bh.knu.ac.kr

단된 자료를 이용하여 지수분포를 따르는 두 모집단의 평균이 같은지에 대한 베이지안 검정법을 제안하였다. 또한, Seong W. Kim과 Hyunsoo Kim (2000)은 지수분포를 따르는 두 집단의 평균을 비교하기 위한 방법으로 부분 베이즈요인의 값과 근사적으로 일치하는 일반적인 베이즈요인의 값을 주는 적절사전분포 (proper prior)인 내재적 사전분포 (intrinsic prior)를 이용한 방법을 제안했고, Jongsig Bae, Hyunsoo Kim과 Seong W. Kim (2000)은 정규분포를 따르는 두 모집단의 평균을 비교하기 위한 내재적 사전분포를 제안했다.

세 개이상의 로그정규모집단에 대한 모수를 비교하기위해, 기존의 고전적인 검정방법을 적용한다면, 각 모집단에서 뽑혀진 표본에 로그변환을 한 다음, 세 개이상의 모집단 평균 차이를 한꺼번에 검정할 수 있게 해주는 분산분석법을 이용한다.

$X_i \sim LN(\mu_i, \sigma^2)$, $i = 1, 2, \cdots, K$인 각각의 로그정규분포에서 뽑혀진 확률표본을 X_{ij}, $j = 1, 2, \cdots, n_i$,라 두고 $Y_{ij} = \log X_{ij}$라고 두자. 각 모집단의 분산이 같으면서 미지인 경우에 사용되는 검정통계량은

$$F_0 = \frac{SSA/(K-1)}{SSE/(N-K)}$$

이며 여기에서 $SSA = \sum_{i=1}^{K} n_i(\overline{Y}_i - \overline{Y})^2$, $SSE = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2$이다. 이 때 유의수준 α에서 $F_0 \geq F(\alpha, K-1, N-K)$이면 가설 M_1을 기각하게 된다. 여기에서 $F(\alpha, K-1, N-K)$는 자유도를 $(K-1, N-K)$로 가지는 F-분포에서 분포의 꼬리 면적이 α가 되는 점이다.

위의 고전적인 분산분석법은 평균이 동일하다는 귀무가설의 채택, 기각 여부를 결정한 후, 필요하다면 Turkey 검정이나, 최소유의차검정 등을 이용하여 다중비교검정 (multiple comparison)을 실시한다.

지금까지 고전적인 검정 방법에 대해서 간략하게 알아보았는데 이 논문에서 우리는 같은 문제를 베이지안 방법을 이용하여 해결하려고 한다. 즉 모수에 대한 최소한의 사전분포만을 이용하여 K개 모집단의 평균이 같은지를 검정하고자 하는 것이다.

일반적으로 K의 수가 작을 경우에는 직접 사후확률을 계산하는 것이 가능하다. 예를 들어 $K = 4$인 경우는 15개, $K = 5$인 경우에는 52개, $K = 6$인 경우는 203개, $K = 7$인 경우는 877개에 이르는 가설에 대한 사후확률을 계산하면 된다. 그러나 집단의 수가 1개씩만 증가하더라도 가설의 수는 엄청나게 증가하며 이들 가설에 대한 사후확률은 직접 계산이 가능하나 힘든 작업이 될 것임에 틀림없다. 그러나 Gopalan 과 Berry (1998)에 소개된 형태 (configuration)의 정의를 응용하여 사후확률을 계산한다면, 사후확률을 계산하는 알고리즘 측면에서 간략하게 표현이 가능하리라 생각한다.

본 논문에서는 K개의 로그정규분포의 모수들에 대한 다중검정 방법으로서 O'Hagan (1995)이 제안한 부분 베이즈요인 (fractional Bayes factor;FBF)을 이용한 베이지안 검정법을 제안한다. 2절에서는 FBF에 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 모의실험을 실시한다.

2. 부분 베이즈요인

$\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)$을 확률밀도함수 $f(\mathbf{y}|\theta)$를 갖는 모집단으로부터 추출된 확률표본이라고 하자. 여기에서 $\theta \in \Theta$이며, 유한한 차원을 가진다. 만약 모형 $M_i : \theta \in \Theta_i$, $\Theta_i \subset \Theta$, $i = 1, 2, \cdots, Q$에 대한 모형선택 (model selection)을 한다고 했을 때 베이지안 관점에서는 모형 M_i에서의 모수 θ에 대한 사전분포와 모형 M_i가 사실일 사전확률을 각각 $\pi_i(\theta)$와 p_i로 가정하고, 모형 M_i가 사실일 사후확률을 계산하여 가장 높은 사후확률 모형을 선택한다. 이 때의 사후확률은 다음과 같이 계산된다.

$$P(M_i|\mathbf{y}) = (\sum_{j=1}^{Q} \frac{p_j}{p_i} B_{ji})^{-1}. \tag{2.1}$$

여기에서 B_{ji}는 모형 M_j의 모형 M_i에 대한 베이즈요인 (Bayes factor)이라고 부르며 정의는 다음과 같다.

$$B_{ji} = \frac{m_j(\mathbf{y})}{m_i(\mathbf{y})} = \frac{\int_{\Theta_j} f(\mathbf{y}|\theta)\pi_j(\theta)d\theta}{\int_{\Theta_i} f(\mathbf{y}|\theta)\pi_i(\theta)d\theta}. \tag{2.2}$$

여기에서 $m_i(\mathbf{y})$는 모형 M_i에서의 $\mathbf{Y}$에 대한 주변확률밀도함수 (marginal or predictive density of $\mathbf{Y}$)이다.

베이지안 검정에서는 흔히 모형 M_i에 대한 사전정보의 부족이나 여러 가지 여건으로 인해 사전분포를 무정보적 사전분포로 가정하는 경우가 많다. 그러나 무정보적 사전분포는 부적절 분포 (improper distribution)인 경우가 대부분이다. 이 무정보적 사전분포를 $\pi_i^N(\theta)$라고 두면, 식 (2.2)은

$$B_{ji}^N = \frac{m_j^N(\mathbf{y})}{m_i^N(\mathbf{y})} = \frac{\int_{\Theta_j} f(\mathbf{y}|\theta)\pi_j^N(\theta)d\theta}{\int_{\Theta_i} f(\mathbf{y}|\theta)\pi_i^N(\theta)d\theta} \tag{2.3}$$

이 되며, 이 때 식 (2.3)에는 부적절 분포의 사용으로 인한 임의의 상수가 포함되어 있다. 부적절 분포를 사용하여 계산한 베이즈 요인에 포함된 임의의 상수로 인해 식 (2.1)를 사용한 베이즈 모형 선택 문제에는 어려움이 있었다. 이러한 어려움을 해결하기 위한 제안으로 Berger와 Pericchi (1996,1998)의 내재적 베이즈요인 (intrinsic Bayes factor; IBF), O'Hagan (1995)의 부분 베이즈요인 등이 있다. 이 중 IBF는 자료 y를 트레이닝 표본 (training sample)이라 불리는 $\mathbf{y}(l)$과 이를 제외한 나머지 표본 $\mathbf{y}(-l)$으로 나누어 임의의 상수를 를 소거하는 방법을 제안하였는데 이것은 베이즈요인을 계산할 때 $\mathbf{y}(l)$에 대한 사후분포를 θ의 사전분포처럼 이용하는 방법이다. 그러나 이 때의 베이즈요인은 최소 트레이닝 표본 (minimal training sample)의 선택에 영향을 받게 되어 이런 영향을 제거하고 안정성을 높이기 위해 산술 내재적 베이즈요인 (arithmetic IBF; AIBF)을 사용하였다. 그러나 AIBF는 소표본인 경우와 비내포 (non-nested) 모형인 경우에는 안정적이지 못하고 표본의 수가 큰 경우에는 계산 시간이 많이 걸린다. 그런 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 안고있다.

한편, O'Hagan (1995)이 제안한 FBF는 IBF에 비해 계산 시간이 적게 걸릴 뿐만 아니라 비내포 모형인 경우에도 쉽게 이용할 수 있으며, 트레이닝 표본을 고려하지 않아도 되는 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 FBF인 경우에는 역수의 관계가 성립한다는 장점이 있다. O'Hagan (1995)이 제안한 FBF는 다음과 같이 정의한다.

정리 2.1 모형 M_1의 모형 M_2에 대한 부분 베이즈요인은 다음과 같이 주어진다.

$$B_{12}^b(\mathbf{y}) = \frac{q_1(b,\mathbf{y})}{q_2(b,\mathbf{y})}. \tag{2.4}$$

여기에서 $i=1,2$에 대하여

$$q_i(b,\mathbf{y}) = \frac{\int_{-\infty}^{\infty} \pi_i(\theta_i) f_i(\mathbf{y}|\theta_i) d\theta_i}{\int_{-\infty}^{\infty} \pi_i(\theta_i) f_i(\mathbf{y}|\theta_i)^b d\theta_i} \tag{2.5}$$

이며 b는 미리 정해진 상수이다.

위의 부분 베이즈요인을 계산할 때, θ_i에 대한 사전분포가 미지의 상수를 가지는 부적절 분포가 되더라도, 미지의 상수는 서로 상쇄되어 나타나지 않는다.

O'Hagan (1995) 는 식 (2.4)과 식 (2.5)에 포함되어 있는 상수 b의 선택은 연구자의 사전분포에 대한 관심에 따라서 세가지 방법을 제시하였다. 이 연구에서는, m을 최소 트레이닝 표본의 크기라 한다면, $b = m/n$, 인 경우를 사용하기로 한다. 이 경우는 FBF를 계산할 때 널리쓰이는 경우이기도 하다.

3. 로그정규분포의 상등에 관한 베이지안 다중검정

로그정규분포를 따르는 K개의 모집단을 각각 $X_p \sim LN(\mu_p, \sigma_p^2),\ \ p=1,\cdots,K$라고 했을 때, $\sigma_1=\sigma_2=\cdots=\sigma_K=\sigma$이면서 σ의 값은 모른다고 가정하자. 이 때 검정(선택)하고자 하는 가설(모형)은 다음과 같다.

모수 $\mu_i, \sigma,\ i=1,2,\cdots,K(\geq 3)$를 가지는 독립인 K개의 로그정규분포에서

$$\begin{aligned}
M_1 :&\quad \mu_1 = \mu_2 = \cdots = \mu_K,\\
M_2 :&\quad \mu_1 \neq \mu_2, \mu_2 = \mu_3 = \cdots = \mu_K,\\
M_3 :&\quad \mu_1 \neq \mu_3, \mu_2 = \mu_4 = \cdots = \mu_K,\\
&\quad \vdots\\
M_Q :&\quad \mu_1 \neq \mu_2 \neq \cdots \neq \mu_K.
\end{aligned}$$

또한 $\sigma_1=\sigma_2=\cdots=\sigma_K=\sigma$라고 가정하자. Berge (1971)에 의하면, 이 때 검정 가능한 가설의 수는

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

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

위의 각 가설에 Gopalan과 Berry (1998)의 형태 (configuration)의 개념을 도입하여 사전분포와 우도함수를 나타내기로 한다.

K개의 모수 μ_i를 임의의 r개의 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 $\mu_1=\mu_3\neq\mu_2=\mu_4\neq\mu_5$인 가설을 의미하며 모수가 서로 다른 세 그룹($r=3$)이 존재한다는 뜻이다.

각 그룹에 속하는 첨자들의 집합을 I_j라고 두면

$$I_j=\{i\mid s_i=j\}\ ,\ \ i=1,2,\cdots,K,\ j=1,2,\cdots,r,\ r=1,2,\cdots,K$$

가 된다. 즉 모형이 $\mu_1=\mu_3\neq\mu_2=\mu_4\neq\mu_5$인 경우 각 그룹에 속 하는 첨자들의 집합은 각각 $I_1=\{1,3\}$, $I_2=\{2,4\}$, $I_3=\{5\}$로 주어진다.

서로 다른 r개의 그룹에 속하는 모수를 $\mu_1^*,\mu_2^*,\cdots,\mu_r^*$라고 두 고 각 그룹에 속하는 표본의 크기를 다음과 같이 나타내자.

$$N(I_j)=\{n_i\mid i\in I_j\}\ ,\ \ i=1,2,\cdots,K,\ j=1,2,\cdots,r,\ r=1,2,\cdots,K.$$

이때 모형이 $\mu_1=\mu_3\neq\mu_2=\mu_4\neq\mu_5$으로 주어졌다면 $\mu_1=\mu_3=\mu_1^*,\ \ \mu_2=\mu_4=\mu_2^*,\ \mu_5=\mu_3^*$이 되며 $N(I_1)=\{n_1,n_3\},N(I_2)=\{n_2,n_4\},N(I_3)=\{n_5\}$가 됨을 알 수 있다.

앞에서 제시한 가설들을 검정하기 위해 사용하는 각 모형에 대한 사전분포로 다음과 같은 무정보적 사전분포를 생각한다.

$$\pi_1^N(\mu_1^*,\sigma)=\pi_2^N(\mu_1^*,\mu_2^*,\sigma)=\cdots=\pi_Q^N(\mu_1^*,\mu_2^*,\cdots,\mu_K^*,\sigma)=\frac{1}{\sigma}. \tag{3.1}$$

각각의 로그정규분포에서 나온 확률표본을 $X_{ij},\ j=1,\cdots,n_i,\ i=1,\cdots,K$라 두 고 $N=\sum_{i=1}^{K}n_i,\ \ Y_{ij}=\log X_{ij},\ \ T_i=\prod_{j=1}^{n_i}x_{ij},\ \ i=1,2,\cdots,K$라고 나타내자.

먼저 $x_{ij},\ j=1,2,\cdots,n_i,\ i=1,2,\cdots,K$을 이용하여 $r=1$인 경우 즉, $M_1:\mu_1=\mu_2=\cdots=\mu_K=\mu_1^*$에 대한 우도함수 (likelihood function)를 구해보면

$$L(\mu_1^*,\sigma)=(\sqrt{2\pi}\sigma)^{-N}(T_1T_2\cdots T_K)^{-1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i\in I_1}\sum_{j\in N(I_1)}(y_{ij}-\mu_1^*)^2\right\}$$

과 같다. 이 때 $I_1=\{1,2,\cdots,K\}$, $N(I_1)=\{n_1,n_2,\cdots,n_K\}$이고, $\sum_{j\in N(I_1)}$는 같은 모수값을 갖는 표본들의 합을 의미한다.

다음으로 $r=2$에 대한 우도함수는 아래와 같이 구해진다.

$$\begin{aligned}L(\mu_1^*,\mu_2^*,\sigma) &= (\sqrt{2\pi}\sigma)^{-N}(T_1T_2\cdots T_K)^{-1}\\ &\times\ \exp\left\{-\frac{1}{2\sigma^2}[\sum_{i\in I_1}\sum_{j\in N(I_1)}(y_{ij}-\mu_1^*)^2+\sum_{i\in I_2}\sum_{j\in N(I_2)}(y_{ij}-\mu_2^*)^2]\right\}.\end{aligned}$$

K개의 모수가 서로 다른 임의의 r, $r=1,2,\cdots,K$개의 그룹으로 나누어진다면, 우도함수는,

$$\begin{aligned} L(\mu_1^*,\mu_2^*,\cdots,\mu_r^*,\sigma) &= (\sqrt{2\pi}\sigma)^{-N}(T_1T_2\cdots T_K)^{-1} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{r}\sum_{i\in I_t}\sum_{j\in N(I_t)}(y_{ij}-\mu_t^*)^2\right\} \end{aligned}$$

이 된다.

이 때, 주변밀도함수는,

$$\begin{aligned} &\int_0^\infty\int_{-\infty}^\infty\cdots\int_{-\infty}^\infty \pi^N(\mu_1^*,\cdots,\mu_r^*,\sigma)L(\mu_1^*,\cdots,\mu_r^*,\sigma)d\mu_1^*\cdots d\mu_r^* d\sigma \\ =& \frac{\Gamma\left(\frac{N-r}{2}\right)}{2\pi^{(N-r)/2}\ T_1T_2\cdots T_K\ \sqrt{N_{I_1}N_{I_2}\cdots N_{I_r}}\ s_r^{N-r}} \end{aligned} \tag{3.2}$$

이다. 여기에서 $s_r^2=\sum_{t=1}^r\sum_{i\in I_t}\sum_{j\in N(I_t)}(y_{ij}-\overline{y}_{I_t})^2$, $r=1,2,\cdots,K$이고, $t=1,2,\cdots,r$에 대하여, $N_{I_t}=\sum_{j\in I_t}n_j$, $\overline{y}_{I_t}=\frac{\sum_{i\in I_t}\sum_{j\in N(I_t)}y_{ij}}{N_{I_t}}$ 이다. 또한, 미리 정하여진 b에 대해

$$\begin{aligned} &\int_0^\infty\int_{-\infty}^\infty\cdots\int_{-\infty}^\infty \pi^N(\mu_1^*,\cdots,\mu_r^*,\sigma)[L(\mu_1^*,\cdots,\mu_r^*,\sigma)]^b d\mu_1^*\cdots d\mu_r^* d\sigma \\ =& \frac{\Gamma\left(\frac{Nb-r}{2}\right)}{2\pi^{(Nb-r)/2}\ b^{Nb/2}\ (T_1T_2\cdots T_K)^b\ \sqrt{N_{I_1}N_{I_2}\cdots N_{I_r}}\ s_r^{Nb-r}} \end{aligned} \tag{3.3}$$

이다. 그러므로 K개의 모수가 임의의 r개의 그룹으로 나누어진다면, 식 (2.5)은

$$q(b,\mathbf{x})=\frac{\Gamma\left(\frac{N-r}{2}\right)\pi^{N(b-1)/2}\ b^{Nb/2}\ (T_1T_2\cdots T_K)^{(b-1)}\ s_r^{N(b-1)}}{\Gamma\left(\frac{Nb-r}{2}\right)} \tag{3.4}$$

로 표현된다.

임의의 두 가설

(1) M_i : K개의 모수가 임의의 r_1개의 서로 다른 그룹으로 나누어진다.

(2) M_j : K개의 모수가 임의의 r_2개의 서로 다른 그룹으로 나누어진다.

라고 가정했을 때, 두 모형에 대한 FBF는 식 (3.4)에 의하여 다 음과 같이 구해진다.

$$B_{12}^b(\mathbf{x})=\frac{q_1(b,\mathbf{x})}{q_2(b,\mathbf{x})}=\frac{\Gamma\left(\frac{N-r_1}{2}\right)\Gamma\left(\frac{Nb-r_2}{2}\right)}{\Gamma\left(\frac{N-r_2}{2}\right)\Gamma\left(\frac{Nb-r_1}{2}\right)}\left(\frac{s_{r_1}}{s_{r_2}}\right)^{N(b-1)}. \tag{3.5}$$

여기에서

$$s_{r_1}^2=\sum_{t=1}^{r_1}\sum_{i\in I_t}\sum_{j\in N(I_t)}(y_{ij}-\overline{y}_{I_t})^2,\quad s_{r_2}^2=\sum_{t=1}^{r_2}\sum_{i\in I_t}\sum_{j\in N(I_t)}(y_{ij}-\overline{y}_{I_t})^2.$$

위의 식 (3.5)를 이용하여 임의의 모형 $M_i,\ i = 1, 2, \cdots, Q$가 사실일 사후확률을 계산하여 사후확률이 가장 큰 모형선택을 한다.

4. 예제와 모의실험을 통한 비교

제안된 방법의 유용성을 알아보기 위하여 실제의 자료와 인위적인 자료를 이용하기로 한다.

먼저 모집단을 세 개로 가정하고 모의실험을 행하여 제안한 방법의 유용성을 알아보고자 한다.

예제 4.1: 로그정규분포를 따르는 세 개의 모집단을 각각 $LN(\mu_1, \sigma_1^2)$, $LN(\mu_2, \sigma_2^2)$와 $LN(\mu_3, \sigma_3^2)$라 두고 $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$ 이면서 σ의 값은 모른다고 가정하자. 이 때 고려할 수 있는 가설은 다음과 같이 주어진다.

$$M_1 : \mu_1 = \mu_2 = \mu_3, \qquad M_2 : \mu_1 = \mu_2 \neq \mu_3, \qquad M_3 : \mu_1 = \mu_3 \neq \mu_2,$$
$$M_4 : \mu_1 \neq \mu_2 = \mu_3, \qquad M_5 : \mu_1 \neq \mu_2 \neq \mu_3.$$

즉, 모집단의 수는 $K = 3$이고 검정 가능한 모형의 수는 $Q = 5$가 된다.

다음에 주어진 23개의 자료는 ball bearing의 내구성을 실험한 자료이다. 이 자료는 와이블 분포나 로그정규분포에 적합하다고 알려져 있다. Lawless (1982)는 로그정규분포에 대한 확률그림을 이용하여 이 자료가 로그정규분포를 따른다고 가정하고 분석하였다. 이 자료를 임의의 세 그룹으로 나누어 제안한 검정 방법을 적용해 보고자 한다.

집단 1	33.00, 45.60, 48.4, 55.56, 68.64, 105.84, 128.04
집단 2	17.88, 41.52, 51.84, 54.12, 84.12, 93.12, 105.12, 173.40
집단 3	28.92, 42.12, 51.96, 67.8, 68.88, 98.64, 127.92

위와 같이 로그정규분포를 따르는 집단을 세 개의 집단으로 분리하더라도 세 집단간에는 $\mu_1 = \mu_2 = \mu_3$, $\sigma_1 = \sigma_2 = \sigma_3$가 성립될 것이므로 이를 이용하여 앞에서 주어진 가설을 고려해 보았다.

이 때 $b = m/n$이라고 가정하고 각 모형이 사실일 사전확률을 1/5로 가정할 경우, 식 (2.1)을 이용하여 모형 $M_j, j = 1, 2, 3, 4, 5$에 대한 사후확률을 각각 계산해 보면

$$P\{M_1|\mathbf{x}\} = 0.52435, \quad P\{M_2|\mathbf{x}\} = 0.14909, \quad P\{M_3|\mathbf{x}\} = 0.14977,$$
$$P\{M_4|\mathbf{x}\} = 0.14902, \quad P\{M_5|\mathbf{x}\} = 0.02778$$

이 되어 모형1, 즉 $M_1 : \mu_1 = \mu_2 = \mu_3$이 선택됨을 알 수 있다.

다음은 고전적인 가설 검정법을 이용하여 위에 주어진 가설을 검정해 보고자 한다. 위의 세 집단의 자료에 로그변환을 하면 세 집단은 정규분포를 따르게 되고 위에서 제시한 가설은 세 집단의 평균 차이에 대한 검정이 되므로 분산분석법으로도 검정이 가능하게 된다. 먼저 세 집단의 분산이 동일 ($\sigma_1 = \sigma_2 = \sigma_3$)한지를 알아보기 위하여 Bartlett 검정을 한 결과, $p-$값이 0.614가 되어 세 집단의 분산이 동일하다는 결론을 얻을 수 있었다. 이 때 세 집

단에 대한 평균 비교를 위해 분산분석을 실시한 결과, 검정통계량 F값은 0.108이 되고 p-값은 0.8982가 되어 모형 M_1이 선택됨을 알 수 있었다.

예제 4.2: 제안한 검정 방법이 올바른 결정을 내리는지를 알아보기 위하여 각 모형에 대한 사후확률을 구하기 위해 100번을 반복하여 모의실험을 실시하였다.

먼저 모집단이 3개인 경우의 모의실험에서 $\sigma_1 = \sigma_2 = \sigma_3 = 1$로 가정하였고, (μ_1, μ_2, μ_3)의 값은 (1,1,1), (1,1,2), (1,2,1), (2,1,1) 그리고 (1,2,3)로 가정하였다. 이는 $K = 3$인 경우, 표현 가능한 가설의 종류와 동일하다. 이 때 b는 m/n인 경우를 가정하였다. 이것은 Beger와 Pericchi (1996)가 제안한 최소 트레이닝 표본의 선택과 일치하는 경우이며 일반적으로 FBF를 계산할 때 많이 이용되는 경우이기 때문이다.

표 4.1은 가설 $M_1 : \mu_1 = \mu_2 = \mu_3$을 참이라고 가정했을 때 고전적인 검정 방법을 이용하여 나머지 대립 가설에 대한 p-value를 계산해 놓은 결과이고 표 4.2는 앞에서 제시한 FBF를 이용하여 각 모형 $M_j,\ j = 1, 2, 3, 4, 5$에 대한 사후확률을 계산해 놓은 결과이다. 표 4.2의 사후확률로부터 우리가 제안한 방법이 표본의 수에 관계없이 올바른 선택을 한다는 사실을 알 수 있다.

예제 4.3: 모집단을 5개로 가정하고 로그정규분포를 따르는 5개의 모집단을 각각 $LN(\mu_1, \sigma_1^2)$, $LN(\mu_2, \sigma_2^2)$, $LN(\mu_3, \sigma_3^2)$, $LN(\mu_4, \sigma_4^2)$와 $LN(\mu_5, \sigma_5^2)$라 두고 $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma$이면서 σ의 값은 역시 모른다고 가정하자. 이 때 고려할 수 있는 가설의 수는 $Q = 52$이며 다음과 같이 주어진다.

$$
\begin{array}{lll}
M_1 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5, & M_2 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \neq \mu_5, & M_3 : \mu_1 = \mu_2 = \mu_3 = \mu_5 \neq \mu_4, \\
M_4 : \mu_1 = \mu_2 = \mu_4 = \mu_5 \neq \mu_3, & M_5 : \mu_1 = \mu_3 = \mu_4 = \mu_5 \neq \mu_2, & M_6 : \mu_1 = \mu_3 = \mu_4 = \mu_5 \neq \mu_1, \\
M_7 : \mu_1 = \mu_2 = \mu_3 \neq \mu_4 = \mu_5, & M_8 : \mu_1 = \mu_2 = \mu_3 \neq \mu_4 \neq \mu_5, & M_9 : \mu_1 = \mu_2 = \mu_4 \neq \mu_3 = \mu_5, \\
M_{10} : \mu_1 = \mu_2 = \mu_4 \neq \mu_3 \neq \mu_5, & M_{11} : \mu_1 = \mu_2 = \mu_5 \neq \mu_3 = \mu_4, & M_{12} : \mu_1 = \mu_2 = \mu_5 \neq \mu_3 \neq \mu_4, \\
M_{13} : \mu_1 = \mu_3 = \mu_4 \neq \mu_2 = \mu_5, & M_{14} : \mu_1 = \mu_3 = \mu_4 \neq \mu_2 \neq \mu_5, & M_{15} : \mu_1 = \mu_3 = \mu_5 \neq \mu_2 = \mu_4, \\
M_{16} : \mu_1 = \mu_3 = \mu_5 \neq \mu_2 \neq \mu_4, & M_{17} : \mu_1 = \mu_4 = \mu_5 \neq \mu_2 = \mu_3, & M_{18} : \mu_1 = \mu_4 = \mu_5 \neq \mu_2 \neq \mu_3, \\
M_{19} : \mu_2 = \mu_3 = \mu_4 \neq \mu_1 = \mu_5, & M_{20} : \mu_2 = \mu_3 = \mu_4 \neq \mu_1 \neq \mu_5, & M_{21} : \mu_2 = \mu_3 = \mu_5 \neq \mu_1 = \mu_4, \\
M_{22} : \mu_2 = \mu_3 = \mu_5 \neq \mu_1 \neq \mu_4, & M_{23} : \mu_2 = \mu_4 = \mu_5 \neq \mu_1 = \mu_3, & M_{24} : \mu_2 = \mu_4 = \mu_5 \neq \mu_1 \neq \mu_3, \\
M_{25} : \mu_3 = \mu_4 = \mu_5 \neq \mu_1 = \mu_2, & M_{26} : \mu_3 = \mu_4 = \mu_5 \neq \mu_1 \neq \mu_2, & M_{27} : \mu_1 = \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5, \\
M_{28} : \mu_1 = \mu_3 \neq \mu_2 \neq \mu_4 \neq \mu_5, & M_{29} : \mu_1 = \mu_4 \neq \mu_2 \neq \mu_3 \neq \mu_5, & M_{30} : \mu_1 = \mu_5 \neq \mu_2 \neq \mu_3 \neq \mu_4, \\
M_{31} : \mu_2 = \mu_3 \neq \mu_1 = \mu_4 \neq \mu_5, & M_{32} : \mu_2 = \mu_3 \neq \mu_1 = \mu_5 \neq \mu_4, & M_{33} : \mu_2 = \mu_3 \neq \mu_1 \neq \mu_4 \neq \mu_5, \\
M_{34} : \mu_2 = \mu_4 \neq \mu_1 = \mu_3 \neq \mu_5, & M_{35} : \mu_2 = \mu_4 \neq \mu_1 = \mu_5 \neq \mu_3, & M_{36} : \mu_2 = \mu_4 \neq \mu_1 \neq \mu_3 \neq \mu_5, \\
M_{37} : \mu_2 = \mu_5 \neq \mu_1 = \mu_3 \neq \mu_4, & M_{38} : \mu_2 = \mu_5 \neq \mu_1 = \mu_4 \neq \mu_3, & M_{39} : \mu_2 = \mu_5 \neq \mu_1 \neq \mu_3 \neq \mu_4, \\
M_{40} : \mu_3 = \mu_4 \neq \mu_1 = \mu_2 \neq \mu_5, & M_{41} : \mu_3 = \mu_4 \neq \mu_1 = \mu_5 \neq \mu_2, & M_{42} : \mu_3 = \mu_4 \neq \mu_1 \neq \mu_2 = \mu_5, \\
M_{43} : \mu_3 = \mu_4 \neq \mu_1 \neq \mu_2 \neq \mu_5, & M_{44} : \mu_3 = \mu_5 \neq \mu_1 = \mu_2 \neq \mu_4, & M_{45} : \mu_3 = \mu_5 \neq \mu_1 = \mu_4 \neq \mu_2, \\
M_{46} : \mu_3 = \mu_5 \neq \mu_1 \neq \mu_2 = \mu_4, & M_{47} : \mu_3 = \mu_5 \neq \mu_1 \neq \mu_2 \neq \mu_4, & M_{48} : \mu_4 = \mu_5 \neq \mu_1 = \mu_2 \neq \mu_3, \\
M_{49} : \mu_4 = \mu_5 \neq \mu_1 = \mu_3 \neq \mu_2, & M_{50} : \mu_4 = \mu_5 \neq \mu_1 \neq \mu_2 = \mu_3, & M_{51} : \mu_4 = \mu_5 \neq \mu_1 \neq \mu_2 \neq \mu_3, \\
M_{52} : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5. & &
\end{array}
$$

표 4.3은 가설 $M_1 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$를 참이라고 가정했을 때 나머지 대립 가설에 대한 p-value를 구해 놓은 결과이며, 표 4.4는 모형 $M_1 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$, M_{29} :

표 4.1: $M_1 : \mu_1 = \mu_2 = \mu_3$이 참일 때의 p-value

	(μ_1, μ_2, μ_3)				
(n_1, n_2, n_3)	(1,1,1)	(1,1,2)	(1,2,1)	(2,1,1)	(1,2,3)
(10,10,10)	0.46143	0.00259	0.00164	0.00179	0.00676
(20,10,10)	0.47138	0.00185	0.00167	0.00007	0.00088
(20,20,20)	0.50173	0.00000	0.00000	0.00000	0.00002
(30,20,20)	0.49359	0.00000	0.00000	0.00000	0.00000
(30,30,30)	0.51594	0.00000	0.00000	0.00000	0.00000

표 4.2: FBF를 이용한 M_i, $i = 1, 2, 3, 4, 5$에 대한 사후확률

	(s_1, s_2, s_3)					
(μ_1, μ_2, μ_3)	(n_1, n_2, n_3)	(1,1,1)	(1,1,2)	(1,2,1)	(2,1,1)	(1,2,3)
(1,1,1)	(10,10,10)	0.43010	0.18015	0.17578	0.16823	0.04574
	(20,10,10)	0.45951	0.16054	0.18090	0.15795	0.04110
	(20,20,20)	0.52645	0.16009	0.14358	0.14277	0.02710
	(30,20,20)	0.54717	0.14312	0.15106	0.13480	0.02385
	(30,30,30)	0.57603	0.12260	0.15834	0.12360	0.01943
(1,1,2)	(10,10,10)	0.00930	0.77454	0.00608	0.01935	0.19073
	(20,10,10)	0.00085	0.79782	0.00069	0.00414	0.19650
	(20,20,20)	0.00008	0.82689	0.00015	0.00025	0.17263
	(30,20,20)	0.00001	0.83974	0.00004	0.00062	0.15959
	(30,30,30)	0.00000	0.86402	0.00000	0.00000	0.13598
(1,2,1)	(10,10,10)	0.00659	0.02622	0.76181	0.00570	0.19968
	(20,10,10)	0.00385	0.00201	0.80569	0.01064	0.17781
	(20,20,20)	0.00001	0.00006	0.83329	0.00054	0.16611
	(30,20,20)	0.00000	0.00001	0.84364	0.00013	0.15622
	(30,30,30)	0.00000	0.00000	0.86883	0.00000	0.13117
(2,1,1)	(10,10,10)	0.00682	0.00977	0.01222	0.76561	0.20558
	(20,10,10)	0.00074	0.00386	0.00374	0.78827	0.20338
	(20,20,20)	0.00002	0.00011	0.00011	0.83567	0.16409
	(30,20,20)	0.00000	0.00002	0.00001	0.84755	0.15242
	(30,30,30)	0.00000	0.00000	0.00000	0.86747	0.13253
(1,2,3)	(10,10,10)	0.01786	0.00527	0.30553	0.31171	0.35964
	(20,10,10)	0.00670	0.00230	0.26369	0.35475	0.37256
	(20,20,20)	0.00038	0.00008	0.24551	0.21640	0.53764
	(30,20,20)	0.00031	0.00006	0.19431	0.17541	0.62990
	(30,30,30)	0.00000	0.00000	0.09571	0.09304	0.81125

표 4.3: $M_1 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$이 참일 때의 p-value

	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$			
$(n_1, n_2, n_3, n_4, n_5)$	(1,1,1,1,1)	(1,2,3,1,4)	(2,1,3,1,2)	(1,2,3,4,5)
(10,10,10,10,10)	0.52693	0.00000	0.00361	0.00000
(20,20,20,10,10)	0.47243	0.00000	0.00001	0.00000
(30,30,30,30,30)	0.54173	0.00000	0.00000	0.00000

표 4.4: FBF를 이용한 $M_1, M_{29}, M_{35}, M_{52}$에 대한 사후확률

		$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$			
$(n_1, n_2, n_3, n_4, n_5)$	$(s_1, s_2, s_3, s_4, s_5)$	(1,1,1,1,1)	(2,1,3,1,2)	(1,2,3,1,4)	(1,2,3,4,5)
(10,10,10,10,10)	(1,1,1,1,1)	0.0402236	3.0629E-6	1.956E-11	1.536E-15
	(1,1,1,1,2)	0.0251026	1.1863E-6	0.0001107	4.802E-10
	(1,1,1,2,1)	0.0357463	0.0011572	3.6614E-9	2.249E-14
	(1,1,2,1,1)	0.0299485	0.0108984	3.648E-11	5.189E-16
	(1,2,1,1,1)	0.0360114	0.0001125	1.278E-11	4.646E-15
	(2,1,1,1,1)	0.0339105	4.1321E-6	9.874E-10	6.3998E-8
	(1,1,1,2,2)	0.0289621	0.0000122	1.618E-11	0.0008139
	(1,1,1,2,3)	0.0169428	0.0003212	0.0004744	0.0009164
	(1,1,2,1,2)	0.0234323	0.0120227	0.0167406	9.583E-12
	(1,1,2,1,3)	0.0151591	0.0333796	0.0586585	1.0665E-9
	(1,1,2,2,1)	0.0253116	2.6688E-6	1.602E-11	1.644E-14
	(1,1,2,3,1)	0.0175701	0.0438357	3.737E-9	1.374E-14
	(1,2,1,1,2)	0.0337505	9.0335E-6	6.6985E-9	4.978E-15
	(1,2,1,1,3)	0.0190991	0.0000338	0.0000416	3.5108E-9
	(1,2,1,2,1)	0.0240026	0.0362487	2.6214E-8	4.84E-16
	(1,2,1,3,1)	0.0235278	0.0179981	1.881E-8	3.55E-14
	(1,2,2,1,1)	0.0269621	0.0000265	1.187E-11	2.709E-15
	(1,2,3,1,1)	0.0182273	0.0273894	1.685E-11	2.898E-15
	(2,1,1,1,2)	0.0250292	5.7661E-6	1.57E-10	6.938E-16
	(2,1,1,1,3)	0.0168433	1.9057E-6	0.0003611	0.0071278
	(2,1,1,2,1)	0.0325126	0.0000247	0.0019155	1.672E-14
	(2,1,1,3,1)	0.0184731	0.0003701	0.0024267	1.0989E-7
	(2,1,2,1,1)	0.0321854	0.0058941	6.902E-12	1.831E-11
	(2,1,3,1,1)	0.0188442	0.0234178	1.0974E-9	4.3477E-8
	(2,2,1,1,1)	0.0276575	4.2504E-6	7.455E-10	0.0003985
	(2,3,1,1,1)	0.0189855	0.0000543	3.6735E-9	0.0042863
	(1,1,2,3,4)	0.0067267	0.0299244	0.0520224	0.1115084
	(1,2,1,3,4)	0.0089417	0.012814	0.0001476	0.0002884
	(1,2,3,1,4)	0.007334	0.0191598	0.3063903	1.9519E-9
	(1,2,3,4,1)	0.008913	0.1424035	4.0793E-9	1.206E-14
	(2,1,1,2,3)	0.0149859	0.0000175	0.1385273	2.917E-10
	(2,1,1,3,2)	0.0165551	0.0004472	2.4105E-9	1.32E-14
	(2,1,1,3,4)	0.0068309	0.0000989	0.0628674	0.1769417
	(2,1,2,1,3)	0.0155576	0.0356272	0.0001251	1.9309E-6
	(2,1,3,1,2)	0.0161787	0.2543813	6.2633E-9	1.714E-16
	(2,1,3,1,4)	0.0071903	0.1074462	0.0555161	0.0015499
	(2,1,2,3,1)	0.0192989	0.0075601	3.8919E-8	6.793E-12
	(2,1,3,2,1)	0.0191714	0.0070108	0.0009435	8.705E-15
	(2,1,3,4,1)	0.007541	0.0328064	0.0007183	3.0827E-8
	(2,2,1,1,3)	0.0144109	1.9894E-6	0.0000393	0.0469987
	(2,3,1,1,2)	0.020001	0.0000507	3.902E-11	7.033E-15
	(2,3,1,1,3)	0.0175145	4.6103E-6	1.4342E-8	2.3108E-8
	(2,3,1,1,4)	0.0081859	0.0000141	0.0000846	0.1899109
	(2,2,1,3,1)	0.0163992	0.0087155	0.020337	0.0001903
	(2,3,1,2,1)	0.0177291	0.0077816	0.1044103	3.89E-12
	(2,3,1,3,1)	0.0142397	0.0351636	0.0111403	5.7008E-7
	(2,3,1,4,1)	0.0080212	0.0129619	0.0567003	0.001205
	(2,2,3,1,1)	0.0160828	0.0058144	2.729E-10	0.0476228
	(2,3,2,1,1)	0.0209345	0.0077751	5.221E-12	0.0002637
	(2,3,3,1,1)	0.017016	0.0000124	3.553E-10	0.0358744
	(2,3,4,1,1)	0.0077457	0.0251778	1.0013E-9	0.0990583
	(1,2,3,4,5)	0.0020732	0.0336001	0.109301	0.2750419

표 4.4 (Continued)

		$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$			
$(n_1, n_2, n_3, n_4, n_5)$	$(s_1, s_2, s_3, s_4, s_5)$	(1,1,1,1,1)	(2,1,3,1,2)	(1,2,3,1,4)	(1,2,3,4,5)
(30,30,30,30,30)	(1,1,1,1,1)	0.0898427	4.618E-19	1.52E-40	5.453E-54
	(1,1,1,1,2)	0.0378197	1.518E-19	7.419E-18	1.306E-32
	(1,1,1,2,1)	0.0340989	4.616E-14	1.4E-30	7.428E-50
	(1,1,2,1,1)	0.0402309	0.0000233	2.779E-36	1.204E-54
	(1,2,1,1,1)	0.0366146	4.492E-13	9.077E-41	1.545E-44
	(2,1,1,1,1)	0.0387499	1.917E-18	6.82E-32	3.32E-30
	(1,1,1,2,2)	0.0297882	6.195E-18	5.171E-40	2.206E-13
	(1,1,1,2,3)	0.0096668	1.574E-14	1.293E-10	4.012E-10
	(1,1,2,1,2)	0.0418308	1.6495E-6	0.0004475	7.572E-38
	(1,1,2,1,3)	0.0123871	0.0011917	0.0023079	3.354E-30
	(1,1,2,2,1)	0.0283548	2.471E-18	8.416E-41	9.901E-51
	(1,1,2,3,1)	0.0123926	0.0014221	3.861E-29	3.573E-50
	(1,2,1,1,2)	0.0442318	7.436E-17	8.868E-29	6.258E-52
	(1,2,1,1,3)	0.0130836	6.837E-14	1.3E-16	9.197E-28
	(1,2,1,2,1)	0.0500812	0.0002743	9.017E-29	3.665E-54
	(1,2,1,3,1)	0.0126902	0.0000516	5.793E-27	1.779E-43
	(1,2,2,1,1)	0.0382626	2.345E-19	6.545E-40	3.28E-50
	(1,2,3,1,1)	0.0125107	0.0001658	4.086E-37	2.665E-45
	(2,1,1,1,2)	0.0376693	1.232E-18	2.714E-39	1.283E-54
	(2,1,1,1,3)	0.0124564	1.15E-18	1.175E-12	9.891E-12
	(2,1,1,2,1)	0.0368688	2.486E-17	2.168E-12	1.168E-51
	(2,1,1,3,1)	0.0113453	1.592E-13	4.566E-13	7.948E-29
	(2,1,2,1,1)	0.0406226	3.2705E-6	4.542E-41	2.165E-40
	(2,1,3,1,1)	0.0135629	0.0002767	1.153E-29	6.056E-30
	(2,2,1,1,1)	0.0353411	8.153E-17	1.322E-31	3.312E-13
	(2,3,1,1,1)	0.0121529	6.763E-14	4.109E-30	9.892E-12
	(1,1,2,3,4)	0.0027795	0.0045851	0.0201378	0.0342759
	(1,2,1,3,4)	0.0029118	0.0016868	3.054E-11	9.474E-11
	(1,2,3,1,4)	0.0030953	0.0012443	0.7966185	2.443E-27
	(1,2,3,4,1)	0.0030151	0.1641947	2.143E-26	1.92E-44
	(2,1,1,2,3)	0.0129853	7.26E-18	0.008605	2.16E-33
	(2,1,1,3,2)	0.0108976	1.517E-14	2.376E-31	4.421E-49
	(2,1,1,3,4)	0.0026166	1.819E-14	0.0031157	0.0148424
	(2,1,2,1,3)	0.0133423	0.0104339	3.463E-17	9.623E-26
	(2,1,3,1,2)	0.0138402	0.593124	3.71E-28	4.983E-55
	(2,1,3,1,4)	0.0032324	0.1765013	0.0027473	1.371E-12
	(2,1,2,3,1)	0.0142064	9.8398E-6	1.5E-24	2.305E-40
	(2,1,3,2,1)	0.013968	5.6901E-6	3.531E-13	4.953E-52
	(2,1,3,4,1)	0.003209	0.00181	5.809E-14	1.493E-29
	(2,2,1,1,3)	0.0096347	1.164E-17	2.889E-18	0.000549
	(2,3,1,1,2)	0.0121036	6.393E-14	4.016E-40	1.339E-44
	(2,3,1,1,3)	0.0119046	1.113E-17	1.906E-27	6.52E-31
	(2,3,1,1,4)	0.0029036	7.846E-15	1.283E-13	0.0309692
	(2,2,1,3,1)	0.0122873	0.0006755	0.0011797	6.975E-14
	(2,3,1,2,1)	0.0118776	7.7229E-7	0.0073956	5.605E-35
	(2,3,1,3,1)	0.0155614	0.0027758	0.0001867	5.502E-24
	(2,3,1,4,1)	0.0030245	0.0015005	0.0013811	1.391E-12
	(2,2,3,1,1)	0.012993	5.1738E-6	4.538E-31	0.0009242
	(2,3,2,1,1)	0.0110888	0.0000161	8.42E-41	6.292E-14
	(2,3,3,1,1)	0.0103749	1.45E-17	1.09E-32	0.0000528
	(2,3,4,1,1)	0.00305	0.0011224	1.432E-29	0.0118956
	(1,2,3,4,5)	0.0004402	0.0368975	0.1558773	0.9064909

$\mu_1 = \mu_4 \neq \mu_2 \neq \mu_3 \neq \mu_5$, $M_{35} : \mu_2 = \mu_4 \neq \mu_1 = \mu_5 \neq \mu_3$, $M_{52} : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$ 을 참으로 가정하였을 때, 52개의 각 모형에 대한 사후확률을 FBF를 이용하여 계산해 놓은 결과이다.

앞에 주어진 예제 4.2에서와 마찬가지로 예제 4.3에서도 여러 경우의 모의실험을 행한 결과, 제안한 FBF를 이용하여 계산한 사후확률을 표 4.4로부터 살펴보면 비교적 표본의 크기가 작은 경우에도 항상 올바른 가설을 선택한다는 것을 알 수 있다.

5. 결론

이 논문에서는 세 개 이상의 로그 정규모집단의 모수 비교를 위한 베이지안 다중검정 방법을 제안하였다. 다중검정에서 관심의 대상이 되는 모집단의 수가 증가한다면, 비교해야할 가설의 수는 지수적으로 증가한다. 그러나 형태 (configuration)의 정의를 통하여 비교하고자 하는 모든 가설을 구체적으로 표현하고, 다중검정을 위한 우도함수와 베이즈 요인들을 계산할 수 있었다. 제안된 방법의 유용성을 실제자료와 모의실험을 통하여 알아 보았다. 모의실험 결과를 통하여 볼 때, 제안된 방법은 표본의 크기가 비교적 작은 경우에도 정확한 결과를 준다는 사실을 알 수 있었다.

참고문헌

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

[2] 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.

[3] Berge, C. (1971). *Principle of Combinatorics*, New York. Academic Press.

[4] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.

[5] 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.

[6] Berger, J.O. and Pericchi, L.R. (1998). Accurate and Stable Bayesian Model Selection : The Median Intrinsic Bayes Factor, *Sankhya*, B, **91**,1-18.

[7] Seong W. Kim (2000). Intrinsic priors for testing exponential means, *Statistics and Probability Letters*, **46**, 195-201.

[8] 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.

[9] 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.

[10] Gopalan, R. and Berry, D.A. (1998). Bayesian Multiple Comparisons Using Dirichlet Process Priors, *Journal of American Statistical Association*, **93**, 1130-1139.

[2001년 5월 접수, 2001년 9월 채택]

Bayesian Testing for the Equality of K-Lognormal Populations with the Fractional Bayes Factor

Kyoung Ae Moon[1)] Dal Ho Kim[2)]

ABSTRACT

We propose the Bayesian multiple hypotheses testing for the equality of K-Lognormal population means. Specially we use the fractional Bayesian factors suggested by O'Hagan (1995) based on the noninformative priors for the parameters. In order to investigate the usefulness of the proposed Bayesian multiple hypotheses testing procedures, we compare it with classical tests via both real data and simulated data.

Keywords: Bayesian multiple hypothesis test; Fractional Bayes factor; Configuration; Noninformative prior; Model selection.

1) Senior Lecturer, Department of Computer Engineering, Donghae University.
E-mail: diana62@mail.donghae.ac.kr
2) Associate Professor, Department of Statistics, Kyungpook National University.
E-mail: dalkim@bh.knu.ac.kr