Journal of the Korean
Data & Information Science Society
2000, Vol. 11, No. 2, pp. 269 ~ 277

로그정규분포의 상등에 관한 베이지안 검정

문경애 [1] · 신임희 [2]· 김달호 [3]

요 약

독립이면서 로그정규분포를 따르는 두 모집단의 평균 차이에 대한 검정으로 Berger와 Pericchi(1996, 1998)가 제안한 내재적 베이즈 요인(intrinsic Bayes factor)을 이용한 베이지안 방법을 제안한다. 이 때 모수에 대한 사전분포로는 무정보적 사전분포(noninformative prior)를 사용한다. 제안한 검정 방법의 유용성을 알아보기 위해 실제 자료의 분석과 모의실험을 이용하여 고전적인 검정 방법과 그 결과를 비교한다.

주제어: 로그정규분포, 무정보적 사전분포, 내재적 베이즈 요인, 트레이닝 표본

1. 서론

신뢰수명검정모형이나 가속수명시험모형 등에 널리 이용되는 로그정규분포의 확률밀도함수는 다음과 같다.

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

여기서 $-\infty < \mu < \infty, 0 < \sigma$이다. 식 (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.

모수 μ_1,σ_1와 μ_2,σ_2를 가지는 독립인 두 로그정규분포에서 다음과 같은 가설이 관심의 대상이다.

$$M_1 : \mu_1 = \mu_2 \ vs.\ M_2 : \mu_1 \neq \mu_2$$

[1](240-600) 강원도 동해시 지흥동 산119 동해대학 사무자동화과 전임강사
[2](705-718) 대구시 남구 대명4동 3056-6 대구카톨릭대학교 의과대학 의학통계학교실 조교수
[3](702-701) 대구시 북구 산격동 1370 경북대학교 통계학과 조교수

이 때 $\sigma_1 = \sigma_2 = \sigma$라고 가정하자.

기존의 고전적인 검정 방법은 각 모집단에서 뽑혀진 표본에 로그변환을 한다음 정규분포에서의 두 모평균 차이에 대한 검정 방법인 t-검정을 이용한다.

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

$$|t| = \frac{|\overline{Y}_1 - \overline{Y}_2|}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},$$

이며 여기에서 $\overline{Y}_1 = \sum_{i=1}^{n_1} Y_{1i}/n_1$, $\overline{Y}_2 = \sum_{i=1}^{n_2} Y_{2i}/n_2$, 그리고

$$S_p^2 = \frac{\sum_{i=1}^{n_1}(Y_{1i}-\overline{Y}_1)^2 + \sum_{i=1}^{n_2}(Y_{2i}-\overline{Y}_2)^2}{n_1+n_2-2}$$

이다. 이 때 유의수준 α에서 $|t| \geq t_{\frac{\alpha}{2}}$이면 가설 M_1을 기각하게 된다. 여기에서 t_α는 자유도를 (n_1+n_2-2)으로 가지는 t-분포에서 분포의 꼬리면적이 α가 되는 점이다.

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

본 논문에서는 두 로그정규분포의 모수들에 대한 검정 방법으로서 Berger와 Pericchi (1996, 1998)가 제안한 내재적 베이즈 요인(intrinsic Bayes factor; IBF)을 이용한 베이지안 검정법을 제안한다. 2절에서는 IBF에 대해 간략하게 소개를 하고 3절에서는 관심의 대상이 되는 가설에 대해 무정보적(noninformative) 사전분포를 이용한 베이지안 검정 방법을 제시한다. 마지막으로 4절에서는 실제 자료를 이용하여 제안한 검정 방법과 고전적 방법을 비교해 보고 제안한 검정 방법이 올바른 결정을 내리는지를 알아보기 위하여 모의실험을 실시하였다.

2. 내재적 베이즈 요인

$\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)$을 확률밀도함수 $f(\mathbf{y}|\theta)$를 갖는 모집단으로부터 추출된 확률표본이라고 하자. 여기에서 $\theta \in \Theta$이며, 유한(finite)한 차원을 가진다.

만약 모형 $M_i : \theta \in \Theta_i$, $\Theta_i \in \Theta, i = 1,2,\cdots,q$에 대한 모형선택(model selection)을 한다고 했을 때 베이지안 관점에서는 모형 M_i에서의 모수 θ에 대한 사전분포와 모형 M_i가 사실(true)일 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.

$$P(M_i|\mathbf{y}) = \left(\sum_{j=1}^{q}\frac{p_j}{p_i}B_{ji}\right)^{-1} \tag{2}$$

위의 식에서 B_{ji}는 모형 M_j의 모형 M_i에 대한 베이즈 요인(Bayes factor)이라고 부르며, 정의는 다음과 같다.

$$B_{ji} = \frac{m_j(\mathbf{y})}{m_i(\mathbf{y})} = \frac{\int_{\Theta_j} f(\mathbf{y}|\theta)\pi_j(\theta)d\theta}{\int_{\Theta_i} f(\mathbf{y}|\theta)\pi_i(\theta)d\theta}, \tag{3}$$

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 (3)은

$$B_{ji}^N = \frac{m_j^N(\mathbf{y})}{m_i^N(\mathbf{y})} = \frac{\int_{\Theta_j} f(\mathbf{y}|\theta)\pi_j^N(\theta)d\theta}{\int_{\Theta_i} f(\mathbf{y}|\theta)\pi_i^N(\theta)d\theta}. \tag{4}$$

이 되며, 이 때 식 (4)에는 부적절 분포의 사용으로 인한 임의의 상수가 포함되어 있다.

부적절 분포를 사용하여 계산한 베이즈 요인에 포함된 임의의 상수로 인해 식 (2)를 사용한 베이즈 모형 선택 문제에는 어려움이 있었다. 이러한 어려움을 해결하기 위한 제안으로 Berger와 Pericchi(1996, 1998)의 IBF, O'Hagan(1995)의 FBF(fractional Bayes factor) 등이 있다. 이 중 IBF는 자료 $\mathbf{y}$를 트레이닝 표본(training sample)이라 불리는 $\mathbf{y}(l)$과 이를 제외한 나머지 표본 $\mathbf{y}(-l)$으로 나누어 임의의 상수를 소거하는 방법을 제안하였는데 이것은 베이즈 요인을 계산할 때 $\mathbf{y}(l)$에 대한 사후분포를 θ의 사전분포처럼 이용하는 방법이다. 이 때 트레이닝 표본인 $\mathbf{y}(l)$는 다음의 조건을 만족해야 한다.

$$0 < m_i^N(\mathbf{y}(l)) < \infty, \quad i = 1, 2, \cdots, q. \tag{5}$$

트레이닝 표본 $\mathbf{y}(l)$을 이용한 베이즈 요인은 다음과 같이 계산한다. 먼저 베이즈 요인을 계산하기 위해 $\mathbf{y}(l)$에 대한 사후분포 $\pi_i^N(\theta|\mathbf{y}(l))$를 θ의 사전분포처럼 이용한다면 베이즈 요인은 다음과 같이 구해진다.

$$B_{ji}(\mathbf{y}(l)) = \frac{\int_{\Theta_j} f(\mathbf{y}(-l)|\theta, \mathbf{y}(l))\pi_j^N(\theta|\mathbf{y}(l))d\theta}{\int_{\Theta_i} f(\mathbf{y}(-l)|\theta, \mathbf{y}(l))\pi_i^N(\theta|\mathbf{y}(l))d\theta} = B_{ji}^N \times B_{ij}^N(\mathbf{y}(l)) \tag{6}$$

여기에서 B_{ji}^N는 식 (4)에 주어져 있고

$$B_{ij}^N(\mathbf{y}(l)) = \frac{m_i^N(\mathbf{y}(l))}{m_j^N(\mathbf{y}(l))}$$

이다.

식 (6)에서 모형 M_j의 모형 M_i에 대한 베이즈 요인 B_{ji}^N 계산시 부적절 사전분포를 사용하여 생성된 임의의 상수를 c_j/c_i라 하면, 트레이닝 표본 $\mathbf{y}(l)$을 이용한 베이즈 요인

$B_{ij}^N(\mathbf{y}(l))$에서도 임의의 상수 c_i/c_j가 생성되어 B_{ji}^N에 $B_{ij}^N(\mathbf{y}(l))$을 곱하면 임의의 상수는 서로 상쇄된다. 이러한 트레이닝 표본을 사용하기 위해서는 몇 개의 트레이닝 표본을 사용할 것인가를 결정해야 하는데 베이지안 가설검정을 위한 최소 트레이닝 표본의 정의는 다음과 같다.

Definition 1 트레이닝 표본 $\mathbf{y}(l)$가 식 *(5)*를 만족할 경우 **사후분포가 진이다**(proper)라 하고, 임의의 트레이닝 표본이 proper하면서 그의 부분 표본이 식 (5)를 만족하지 못할 경우 이 트레이닝 표본을 **최소**(minimal)라고 한다.

Beger와 Pericchi(1996, 1998)는 최소 트레이닝 표본을 이용한 베이지안 검정 방법을 제시하였는데, 이 때의 베이즈 요인 $B_{ji}(\mathbf{y}(l))$은 최소 트레이닝 표본의 선택에 영향을 받게 된다. 이런 영향을 제거하고 안정성을 높이기 위해 다음과 같은 베이즈 요인을 사용하였다.

Definition 2 모형 M_j의 M_i에 대한 산술 내재적 베이즈 요인(arithmetic intrinsic Bayes factor; AIBF)과 중위수 내재적 베이즈 요인(median intrinsic Bayes factor; MIBF)은 각각 다음과 같이 주어진다.

$$B_{ji}^{AI} = B_{ji}^N \cdot \frac{1}{L}\sum_{l=1}^{L} B_{ij}^N(\mathbf{y}(l)), \tag{7}$$

$$B_{ji}^{MI} = B_{ji}^N \cdot \underset{1 \le l \le L}{Median}\, B_{ij}^N(\mathbf{y}(l)), \tag{8}$$

이다. 여기에서 L은 모든 가능한 최소 트레이닝 표본의 수이다.

식 (7)이나 식 (8)에 주어진 B_{ji}^{AI} 혹은 B_{ji}^{MI}를 식 (2)의 B_{ji}대신 대입하여 각 모형에 대한 사후확률을 구한 후 가장 높은 사후확률을 가지는 모형을 선택하게 된다.

3. 로그정규분포의 상등에 관한 베이지안 검정

로그정규분포를 따르는 두 모집단을 각각 $X_1 \sim LN(\mu_1, \sigma_1^2)$과 $X_2 \sim LN(\mu_2, \sigma_2^2)$라고 했을 때, $\sigma_1 = \sigma_2 = \sigma$이면서 σ의 값은 모른다고 가정하자. 이 때 우리가 검정(선택)하고자 하는 가설(모형)은 다음과 같다.

$$M_1 : \mu_1 = \mu_2 \ vs.\ M_2 : \mu_1 \neq \mu_2$$

가설 M_1에서 $\mu_1 = \mu_2 = \mu$라고 두면, 위의 가설들을 검정하기 위해 사용하는 사전분포로 다음과 같은 무정보적 사전분포를 생각한다.

$$\pi_1^N(\mu, \sigma) = \pi_2^N(\mu_1, \mu_2, \sigma) = \frac{1}{\sigma}, \quad 0 < \sigma < \infty, \ -\infty < \mu < \infty. \tag{9}$$

각각의 로그정규분포에서 나온 확률표본을 $X_{ij},\ j=1,2,\cdots,n_i,\ i=1,2$라 두고 $N=n_1+n_2,\ Y_{ij}=\log X_{ij},\ T_i=\prod_{j=1}^{n_i} x_{ij},\ i=1,2$라고 나타내자.

먼저 $Y_{ij},\ j=1,2,\cdots,n_i,\ i=1,2$를 이용하여 모형 $M_1:\mu_1=\mu_2=\mu$에 대한 우도함수(likelihood function)를 구해보면

$$L_1(\mu,\sigma)=\sqrt{2\pi}\sigma^{-N}(T_1T_2)^{-1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(y_{ij}-\mu)^2\right\} \tag{10}$$

과 같으며, 모형 M_2에 대한 우도함수는

$$L_2(\mu_1,\mu_2,\sigma)=\sqrt{2\pi}\sigma^{-N}(T_1T_2)^{-1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2\right\} \tag{11}$$

이 된다.

식 (9)에 주어진 사전분포와 식 (10)와 식 (11)의 우도함수를 이용한 주변확률밀도함수는 다음과 같이 된다. 먼저 모형 M_1에 대한 주변확률밀도함수는

$$\begin{aligned} m_1^N(\mathbf{x}) &= \int_0^\infty\int_{-\infty}^\infty L_1(\mu,\sigma)\pi_1^N(\mu,\sigma)d\mu d\sigma \\ &= \frac{\Gamma(\frac{N-1}{2})}{2\pi^{\frac{N-1}{2}}T_1T_2\sqrt{N}S_1^{\frac{N-1}{2}}} \end{aligned}$$

이 되고, 여기에서 $S_1=\sum_{i=1}^2\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2,\ \bar{y}=\sum_{i=1}^2\sum_{j=1}^{n_i}y_{ij}/N$이다. 또한 모형 M_2에 대한 주변확률밀도함수는

$$\begin{aligned} m_2^N(\mathbf{x}) &= \int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty L_2(\mu_1,\mu_2,\sigma)\pi_2^N(\mu_1,\mu_2,\sigma)d\mu_1 d\mu_2 d\sigma \\ &= \frac{\Gamma(\frac{N-2}{2})}{2\pi^{\frac{N-2}{2}}T_1T_2\sqrt{n_1n_2}S_2^{\frac{N-2}{2}}} \end{aligned}$$

이고, 여기에서 $S_2=\sum_{i=1}^2\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$이고, $\bar{y}_i=\sum_{j=1}^{n_i}y_{ij}/n_i$이다.

다음은 최소 트레이닝 표본을 구하고 최소 트레이닝 표본에 대한 주변확률밀도함수를 계산한다. 각각의 로그정규분포에서 뽑혀진 두 개의 서로 다른 관찰값 $X(l)=(X_{1k},X_{2l})$를 트레이닝 표본으로 정의할 경우, $m_1^N(x_{1k},x_{2l})=1/2x_{1k}x_{2l}(y_{1k}-y_{2l})$이 되지만 $m_2^N(x_{1k},x_{2l})=\infty$가 되어 조건 (5)를 만족하지 않게 된다. 그러나 세 개의 서로 다른 관찰값 $X_1(l)=(X_{1k_1},X_{1k_2},X_{2l})$ 또는 $X_2(l)=(X_{1k},X_{2l_1},X_{2l_2})$를 트레이닝 표본으로 사용하면 조건 (5)를 만족하게 되어 최소 트레이닝 표본이 됨을 알 수 있다.

먼저 최소 트레이닝 표본이 $X_1(l)=(X_{1k_1},X_{1k_2},X_{2l})$인 경우, 모형 M_1과 모형 M_2에 대한 주변확률밀도함수를 구해보면 다음과 같다.

$$m_1^N(x_{1k_1},x_{1k_2},x_{2l})=\frac{1}{2\pi\sqrt{3}x_{1k_1}x_{1k_2}x_{2l}S_{11}},$$

$$m_2^N(x_{1k_1}, x_{1k_2}, x_{2l}) = \frac{1}{x_{1k_1}x_{1k_2}x_{2l}2\sqrt{2S_{12}}}.$$

여기에서

$$\begin{aligned} S_{11} &= y_{1k_1}^2 + y_{1k_2}^2 + y_{2l}^2 - \frac{(y_{1k_1} + y_{1k_2} + y_{2l})^2}{3}, \\ S_{12} &= y_{1k_1}^2 + y_{1k_2}^2 - \frac{(y_{1k_1} + y_{1k_2})^2}{2} \end{aligned}$$

이다.

또한, 최소 트레이닝 표본이 $X_2(l) = (X_{1k}, X_{2l_1}, X_{2l_2})$인 경우의 모형 M_1과 모형 M_2에 대한 주변확률밀도함수는

$$m_1^N(x_{1k}, x_{2l_1}, x_{2l_2}) = \frac{1}{2\pi\sqrt{3}x_{1k}x_{2l_1}x_{2l_2}S_{21}},$$

$$m_2^N(x_{1k}, x_{2l_1}, x_{2l_2}) = \frac{1}{x_{1k}x_{2l_1}x_{2l_2}2\sqrt{2S_{22}}}$$

이 된다. 여기에서

$$\begin{aligned} S_{21} &= y_{1k}^2 + y_{2l_1}^2 + y_{2l_2}^2 - \frac{(y_{1k} + y_{2l_1} + y_{2l_2})^2}{3}, \\ S_{22} &= y_{2l_1}^2 + y_{2l_2}^2 - \frac{(y_{2l_1} + y_{2l_2})^2}{2} \end{aligned}$$

이다.

Theorem 1 최소 트레이닝 표본 $X_1(l) = (X_{1k_1}, X_{1k_2}, X_{2l})$와 $X_2(l) = (X_{1k}, X_{2l_1}, X_{2l_2})$를 이용한 모형 $M_2 : \mu_1 \neq \mu_2$의 모형 $M_1 : \mu_1 = \mu_2$에 대한 AIBF와 MIBF는 각각 다음과 같다.

$$B_{21}^{AI} = B_{21}^N \frac{\sum_{l=1}^{L_1} B_{12}^N(x_1(l))/L_1 + \sum_{l=1}^{L_2} B_{12}^N(x_2(l))/L_2}{2}, \tag{12}$$

$$B_{21}^{MI} = B_{21}^N \frac{\underset{1\le l\le L_1}{Median} B_{12}^N(x_1(l)) + \underset{1\le l\le L_1}{Median} B_{12}^N(x_2(l))}{2}. \tag{13}$$

여기에서 $B_{21}^N = m_2^N(\mathbf{x})/m_1^N(\mathbf{x})$이며, $B_{12}^N(x_i(l)) = m_1^N(x_i(l))/m_2^N(x_i(l))$, $i = 1, 2$이다. 또한 L_i는 $x_i(l), i = 1, 2$에 대해서 모든 가능한 최소 트레이닝 표본의 수이다.

4. 예제

예제1: 다음에 주어진 23개의 자료는 볼 베어링(ball bearing)의 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 두 그룹으로 나누어 제안한 검정 방법을 적용해 보고자 한다.

집단 1	33.00, 45.60, 51.84, 51.96, 55.56, 67.80, 68.64, 68.64, 93.12, 105.84, 128.04
집단 2	17.88, 28.92, 41.52, 42.12, 48.40, 54.12, 68.88, 84.12, 98.64, 105.12, 127.92, 173.40

위와 같이 로그정규분포를 따르는 집단을 두 개의 집단으로 분리하더라도 두 집단간에는 $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$가 성립될 것이 예상되므로 이를 이용하여 아래의 가설을 고려해 보았다.

$$M_1 : \mu_1 = \mu_2 \ vs. \ M_2 : \mu_1 \neq \mu_2$$

위의 가설에 대한 식 (12)와 식 (13)의 베이즈 요인을 계산해 보면

$$B_{21}^{AI} = 0.3189, \qquad B_{21}^{MI} = 0.2248$$

가 된다. 각 모형에 대한 사전확률을 1/2로 가정하고 위의 결과를 식 (2)를 대입하여 모형 M_1에 대한 사후확률을 각각 계산해 보면

$$P^{AI}\{M_1|\mathbf{x}\} = 0.7582, \qquad P^{MI}\{M_1|\mathbf{x}\} = 0.8164$$

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

다음은 고전적인 가설 검정법을 이용하여 위에 주어진 가설을 검정해 보고자 한다. 위의 두 집단의 자료에 로그변환을 하면 두 집단은 정규분포를 따르게 되어 위에서 제시한 가설은 모평균 차이에 대한 t검정방법으로도 검정이 가능하게 된다. 먼저 두 집단의 분산이 동일($\sigma_1 = \sigma_2$)한지를 알아보기 위하여 F-검정을 한 결과, F값이 $F_0 = 0.3618$이고 p-값은 0.1204로 되어 두 집단의 분산이 동일하다는 결론을 얻을 수 있었다. 이 때 모평균 차이에 대한 t검정을 실시한 결과, t값은 $t_0 = 0.2303$이 되고 p-값은 0.4101이 되어 역시 모형 M_1이 선택되었다.

예제2: 제안한 검정 방법이 올바른 결정을 내리는지를 알아보기 위하여 모의실험을 실시하였다. 이 때의 계산은 500번 반복으로 이루어졌으며 각 모형에 대한 사전확률은 1/2로 가정하였다. 이 모의실험에서 $\sigma_1 = \sigma_2 = 1$로 가정하였고, (μ_1, μ_2)의 값은 (-1,-1), (-1,0) 그리고 (-1,1)로 가정하였다.

표 1은 식 (12)와 식 (13)에 제시한 AIBF와 MIBF를 이용하여 가설의 사후확률을 계산할 경우 올바른 결정을 내리도록 하는지 알아보기 위하여 $M_1 : \mu_1 = \mu_2$와 $M_2 : \mu_1 \neq \mu_2$의 가설에서 모형 M_1에 대한 사후확률을 계산해 놓은 결과이다. 이 표로부터 제안된 AIBF와 MIBF를 이용하여 가설에 대한 사후확률을 계산할 경우 표본이 비교적 작은 경우에도 올바른 가설을 선택하게 한다는 것을 알 수 있었다.

표 1. $M_1 : \mu_1 = \mu_2$ v.s. $M_2 : \mu_1 \neq \mu_2$에 대한 사후확률

(μ_1, μ_2)	(n_1, n_2)	$P^{AI}\{M_1\|\mathbf{x}\}$	$P^{MI}\{M_1\|\mathbf{x}\}$	p-value
(-1,-1)	(10,10)	0.85250	0.88923	0.48422
	(20,10)	0.86068	0.89730	0.51501
	(20,20)	0.88433	0.91460	0.52244
	(30,20)	0.89023	0.92276	0.52704
	(30,30)	0.90036	0.92844	0.48925
(-1,0)	(10,10)	0.51799	0.59712	0.10907
	(20,10)	0.45662	0.48268	0.09733
	(20,20)	0.29303	0.33509	0.03976
	(30,20)	0.22226	0.25758	0.01449
	(30,30)	0.13516	0.15447	0.01016
(-1,1)	(10,10)	0.08292	0.13524	0.00305
	(20,10)	0.02987	0.04256	0.00123
	(20,20)	0.00305	0.00442	0.00003
	(30,20)	0.00027	0.00144	0.00000
	(30,30)	0.00005	0.00012	0.00000

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

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

Bayesian Testing for the Equality of Two Lognormal Populations

Kyoung Ae Moon [4]· Im Hee Shin [5]· Dal Ho Kim [6]

Abstract

We propose the Bayesian testing for the equality of two log-normal population means. Specifically we use the intrinsic Bayes factors suggested by Berger and Perichi (1996, 1998) based on the noninformative priors for the parameters. In order to investigate the usefulness of the proposed Bayesian testing procedures, we compare it with classical tests via both real data analysis and simulation.

Key Words and Phrases: lognormal distribution, noninformative prior, intrinsic Bates Factor, training sample

[4]Full time Instructor, Donghae Junior College, Donghae, 240-600, Korea.
[5]Assistant Professor, Division of Medical Statistics, Catholic University of Taegu, Taegu, 705-718, Korea.
[6]Assistant Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.