Journal of Korean
Data & Information Science Society
2005, Vol. 16, No. 3, pp. 539~547

Bayesian Testing for the Equality of Two Inverse Gaussian Populations with the Fractional Bayes Factor[1)]

Ko, Jeong Hwan[2)]

Abstract

We propose the Bayesian testing for the equality of two independent Inverse Gaussian population means using the fractional Bayesian factors suggested by O'Hagan(1995). As prior distribution for the parameters, we assumed the noninformative priors. In order to investigate the usefulness of the proposed Bayesian testing procedures, the behaviors of the proposed results are examined via real data analysis.

Keywords : fractional Bayes factor, Inverse Gaussian population, Reference prior

제 1 절 서론

인구통계학, 노동 및 재정관련 그리고 심장병학 자료, 신뢰수명검정모형에 널리 이용되는 역가우시안분포(inverse Gaussian distribution)의 확률밀도함수는 다음과 같다.

$$f(x : \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right] \quad ,0 < x < \infty. \tag{1.1}$$

여기서 $0 < \lambda < \infty$, $0 < \mu < \infty$이다. 모수 μ는 분포의 평균을 나타내고 λ은 척도모수이다. 식 (1.1)을 확률밀도함수로 가지는 확률변수를 $X \sim IG(\mu, \lambda)$으로 나타내자. Chhikara와 Folks(1989)은 역가우시안분포의 통계적 성질을 자세히 소개하였다.

모수 μ_1, λ_1와 μ_2, λ_2를 가지는 독립인 두 역 가우시안분포에서 관심 대상인 가

---

1) This paper is supported in part by an Andong National University Research Grant for 2003.
2) Professor, Department of Information Statistics, Andong National University, Andong, 760-749, Korea. E-mail : jhko@andong.ac.kr

설은 다음과 같다.

$$H_1 : \mu_1 = \mu_2 \ vs. \ H_2 : \ \mu_1 \neq \mu_2$$

이 때 $\lambda_1 = \lambda_2 = \lambda$라고 가정하자.

기존의 논문에서 두 집단의 평균차이에 대한 고전적인 검정방법인 t-검정을 이용한다. 여기에서 역 가우시안 분포에 대한 고전적인 검정방법은 Chhikara(1975)에 의해 다음과 같이 정의되어 있다. 두 모집단의 모수 $\lambda_1 = \lambda_2 = \lambda$이 같으면서 미지인 경우에 기각역은

$$\left| \frac{[n_1 n_2 (n_1 + n_2 - 2)]^{1/2} (\overline{X} - \overline{Y})}{[\overline{X}\,\overline{Y} (n_1 \overline{X} + n_2 \overline{Y})(V_1 + V_2)]^{1/2}} \right| > t_{1-\alpha/2}$$

이다. 여기서 $t_{1-\alpha/2}$ 는 자유도를 $(n_1 + n_2 - 2)$으로 가지는 t-분포에서 $100(1-\alpha/2)$가 되는 점이다. 이 논문에서 같은 문제를 베이지안 방법을 이용하여 해결하려고 한다

역가우시안분포와 관련한 베이지안 추론은 Banerjee와 Bhattacharyya(1979), Padgett (1981) 등이 모수에 대한 점추정과 신뢰도추정, 최고사후밀도구간 등의 문제를 고려하였다.

베이지안 가설검정(Bayesian hypothesis testing)이나 모형선택문제(model selection problem)에서 가설을 비교하기 위해서 베이즈요인(Bayes factor)의 계산이 필요하다. 베이즈 접근법에서는 모수에 대한 사전분포(prior distribution)가 필수적인데, 사전분포로는 흔히 무정보 분포(noninformative distribution)를 가정하는 경우가 많다. 그런데 무정보 분포는 부적절한 분포(improper distribution)가 되는 경우가 많으며, 이러한 경우, 베이즈 요인은 미지의 상수(undetermined constants)를 포함한다. 이와같은 문제를 해결하기 위하여 몇 가지의 방법이 제안되었다.

Berger와 Pericchi(1996)는 내재적 베이즈요인(intrinsic Bayes factor, IBF)를 정의하고 IBF가 가설검정이나 모형선택문제에 유용함을 밝혔다. 그들은 최소트레이닝표본(minimal training sample)을 이용하여 미지의 상수를 상쇄시키는 접근법을 이용하였다. 그러나 그들이 제안한 IBF는 소표본인 경우와 비내포(non-nested)인 경우에는 안정적이지 못하고 적용하기가 어렵다. 또한 계산량도 많은 결점이 있다. 그래서 Berger와 Pericchi(1998)는 비내포 모형이나 소표본에서도 잘 적용되는 중위수 내재적 베이즈요인(Median IBF, MIBF)를 제안하였다.

한편, O'Hagan(1995)이 제안한 부분 베이즈요인(fractional Bayes factor, FBF)는 IBF에 비해 계산량도 적고, 비내포 모형인 경우에도 쉽게 이용할 수 있는 장점이 있다. 또, IBF의 경우에 비내포인 두 모형 H_1과 H_2에 대한 산술 내재적 베이즈요인

(arithmetic IBF, AIBF)를 구할 때, H_1의 H_2에 대한 AIBF와 H_2의 H_1에 대한 AIBF의 값이 역수의 관계가 성립하지 않는다. 이는 AIBF의 정의에 의해 명확하다. 그러나 FBF인 경우에는 역수의 관계가 성립하여 이러한 문제점은 없다.

본 논문에서는 두 역가우시안분포의 모수에 대한 검정방법으로 O′Hagan(1995)이 제안한 부분 베이즈요인을 이용한 베이지안 검정법을 제안한다. 2절에서는 부분 베이즈요인에 대해 간략하게 소개하고, 3절에서는 관심의 대상이 되는 가설에 대하여 무정보적 사전분포를 이용한 베이지안 검정방법을 제시한다. 마지막으로 4절에서는 실제자료를 이용하여 제안한 검정방법을 알아본다.

제 2 절 Fractional Bayes Factor에 대한 소개

$\underset{\sim}{X}=(X_1, X_2, \cdots, X_n)$을 확률밀도함수 $f(x|\theta)$를 갖는 모집단으로부터 추출된 확률표본이라고 하자. 여기에서 $\theta \in \Theta$이며 유한한 차원을 가진다. 만약 모형 H_i : $\theta \in \Theta_i$, $\Theta_i \in \Theta$, $i=1, 2, \cdots, q$에 대한 모형선택을 한다고 했을 때 베이지안 관점에서는 모형 H_i에서의 모수 θ에 대한 사전분포와 모형 H_i가 사실일 사후확률을 계산하여 가장 높은 사후확률모형을 선택한다. 이 때의 사후확률은 다음과 같이 계산한다.

$$P(H_i|\underset{\sim}{x}) = \left(\sum_{j=1}^{q} \frac{p_j}{p_i} B_{ji} \right)^{-1}. \tag{2.1}$$

여기에서 B_{ji}는 모형 H_j의 H_i에 대한 베이즈요인이라고 부르며 정의는 다음과 같다.

$$B_{ji} = \frac{m_j(\underset{\sim}{x})}{m_i(\underset{\sim}{x})} = \frac{\int_{\Theta_j} f(\underset{\sim}{x}|\theta)\pi_j(\theta)\,d\theta}{\int_{\Theta_i} f(\underset{\sim}{x}|\theta)\pi_i(\theta)\,d\theta} \quad . \tag{2.2}$$

여기에서 $m_j(\underset{\sim}{x})$는 모형 H_i에서의 $\underset{\sim}{x}$에 대한 주변확률밀도함수(marginal or predictive density of $\underset{\sim}{x}$)이다.

베이지안 검정에서는 흔히 모형 H_i에 대한 사전정보의 부족이나 여러가지 여건으로 인해 사전분포를 무정보 사전분포로 가정하는 경우가 많다. 그러나 무정보 사전분포는 부적절한 분포(improper distribution)인 경우가 대부분이다. 이 무정보사전분포를 $\pi_j^N(\theta)$라고 두면, 식 (2.2)은

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

이 되며, 이 때 식 (2.3)에는 부적절한 분포의 사용으로 인한 임의의 상수가 포함되어 있다. 부적절한 분포를 사용하여 계산된 베이즈요인에 포함된 임의의 상수로 인해 식 (2.1)를 사용한 베이즈 모형선택 문제에는 어려움이 있었다. 이러한 어려움을 해결하기 위한 제안으로 Berger와 Pericchi(1996, 1998)의 내재적 베이즈요인(IBF), O′Hagan (1995)의 부분 베이즈요인(FBF)등이 있다. 이 중 IBF는 자료 y를 트레이닝 표본이라 불리는 $y(l)$과 이를 제외한 나머지 표본 $y(-l)$으로 나누어 임의의 상수를 소거하는 방법을 제안하였는데 이것은 베이즈요인을 계산할 때 $y(l)$에 대한 사후분포를 θ의 사전분포처럼 이용하는 방법이다. 이 때 베이즈요인은 최소트레이닝표본의 선택에 영향을 받게 되어 이런 영향을 제거하고 안정성을 높이기 위해 산술 내재적 베이즈요인 (AIBF)을 사용하였다. 그러나 AIBF는 소표본인 경우와 비내포모형인 경우에는 안정적이지 못하고 계산 시간도 많이 걸린다는 결점이 있다. 그런 이유로 Berger와 Pericchi(1998)는 비내포 모형이나 소표본에서도 잘 적용되는 중위수 내재적 베이즈요인(MIBF)을 제안하였다. 그러나 IBF는 근본적으로 표본의 재사용이라는 문제점을 안고 있다.

한편, O′Hagan(1995)는 IBF에 비해 계산시간이 적게 걸릴 뿐만 아니라 비내포 모형인 경우에도 쉽게 이용할 수 있으며, 트레이닝표본을 고려하지 않아도 되는 장점이 있는 FBF를 아래와 같이 제안하였다. 먼저, $b=m/n, m\le n$이라 두자. 그리고 H_1의 H_2에 대한 FBF를

$$B_b(x)=\frac{q_1(b,x)}{q_2(b,x)}, \tag{2.4}$$

여기에서 $i=1,2$에 대하여,

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

이고, 식 (2.4), (2.5)에 포함되어 있는 상수 b의 선택은 다음의 세가지 방법을 제시하였다. m을 Berger와 Pericchi(1996)가 정의한 최소트레이닝표본의 수라 한다면,

(1) $b=m/n$, (2) $b=n^{-1}\max(m,\sqrt{n})$ (3) $b=n^{-1}\max(m,\log n)$

으로 제안하였다. (1)은 사전분포의 불명확성에 대한 로버스트성(robustness)에 특별

히 관심이 없을 때 사용하고, (2)는 로버스트성이 관심이 있는 경우에 사용하며, (3)은 (1)과 (2)의 중간적 입장에서 사용할 것을 제시하였다.

위의 방법을 적용하면 θ_i에 대한 사전분포가 미지의 상수를 가지는 부적절 분포가 되더라도, 미지의 상수는 서로 상쇄되어 나타나지 않는다.

제 3 절 역가우시안분포의 상등에 관한 베이지안 검정

역가우시안분포를 따르는 두 모집단을 각각 $X_1 \sim IG(\mu_1, \lambda_1)$ 과 $X_2 \sim IG(\mu_2, \lambda_2)$라고 하고, $\lambda_1 = \lambda_2 = \lambda$이면서 λ의 값을 모른다고 가정하자. 이 때 우리가 검정하고자 하는 가설은 다음과 같다.

$$H_1 : \mu_1 = \mu_2 \quad vs. \quad H_2 : \mu_1 \neq \mu_2$$

위의 가설들을 검정하기 위해 사용하는 사전분포로 다음과 같은 무정보적 사전분포를 생각한다. 이 무정보적 사전분포는 각각의 가설에서 준거사전분포이다.

$$\Pi_1^N(\mu, \lambda) \propto \mu^{-\frac{3}{2}} \lambda^{-1}, \quad 0<\lambda<\infty, \quad 0<\mu<\infty,$$
$$\Pi_2^N(\mu_1, \mu_2, \lambda) \propto \mu_1^{-\frac{3}{2}} \mu_2^{-\frac{3}{2}} \lambda^{-1}, \quad 0<\lambda<\infty, 0<\mu_1<\infty, 0<\mu_2<\infty. \tag{3.1}$$

각각의 역가우시안분포에서 나온 확률변수를 $X_{ij}, j=1,2,\cdots, n_i, \ i=1,2$라 두고 $N = n_1 + n_2, \ W_i = \prod_{j=1}^{n_i} x_{ij}^3, \ i=1,2$라고 나타내자.

가설 $H_1 : \mu_1 = \mu_2 = \mu$에 대한 우도함수를 구해보면

$$L_1(\mu, \lambda) = \prod_{i=1}^{2} \prod_{j=1}^{n_i} f(x_{ij} : \lambda, \mu)$$
$$= \left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} w^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2\mu^2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \frac{(x_{ij}-\mu)^2}{x_{ij}}\right\} \tag{3.2}$$

과 같으며, 가설 H_2에 대한 우도함수는

$$L_2(\mu_1, \mu_2, \lambda) = \prod_{i=1}^{2} \prod_{j=1}^{n_i} f(x_{ij} : \mu_1, \mu_2, \lambda)$$

$$= \left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} w^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}\left[\sum_{j=1}^{n_1} \frac{(x_{1j}-\mu_1)^2}{\mu_1^2} + \sum_{j=1}^{n_2} \frac{(x_{2j}-\mu_2)^2}{\mu_2^2}\right]\right\} \tag{3.3}$$

이 된다.

가설 $H_1 : \mu_1 = \mu_2 = \mu$에 대해

$$m_1^N(x) = \int_0^\infty \int_0^\infty \Pi_1(\mu, \lambda)\, L_1(\mu, \lambda)\, d\lambda\, d\mu = \int_0^\infty \frac{\Gamma\left(\frac{n}{2}\right) \mu^{-\frac{3}{2}}}{(2\pi)^{\frac{n}{2}} w^{\frac{1}{2}} A^{\frac{n}{2}}}\, d\mu$$

여기에서 $A = \frac{1}{2\mu^2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \frac{(x_{ij}-\mu)^2}{x_{ij}}$ 이다. 또한, 미리 정하여진 b에 대해

$$\int_0^\infty \int_0^\infty \Pi_1(\mu, \lambda)\, [\, L_1(\mu, \lambda)]^b\, d\mu\, d\lambda = \frac{\Gamma\left(\frac{nb}{2}\right) S(b : x)}{(2\pi)^{\frac{nb}{2}} w^{\frac{b}{2}}}$$

이다. 여기에서 $S(b: x) = \int_0^\infty \mu^{-\frac{3}{2}} (Ab)^{-\frac{n}{2}}\, d\mu$이다.

그러므로

$$q_1(b : x) = \frac{\Gamma\left(\frac{n}{2}\right) S(1: x)}{(2\pi)^{\frac{n}{2}} w^{\frac{1}{2}}} \frac{(2\pi)^{\frac{nb}{2}} w^{\frac{b}{2}}}{\Gamma\left(\frac{nb}{2}\right) S(b : x)} \tag{3.4}$$

이다.

가설 $H_2 : \mu_1 \neq \mu_2$에 대해

$$m_2^N(x) = \int_0^\infty \int_0^\infty \int_0^\infty \Pi_2(\mu_1, \mu_2, \lambda)\, L_2(\mu_1, \mu_2, \lambda)\, d\lambda\, d\mu_1 d\mu_2$$

$$= \int_0^\infty \int_0^\infty \frac{\Gamma\left(\frac{n}{2}\right) \mu^{-\frac{3}{2}}}{(2\pi)^{\frac{n}{2}} w^{\frac{1}{2}} B^{\frac{n}{2}}}\, d\mu_1 d\mu_2$$

여기에서 $B = \sum_{j=1}^{n_1} \frac{(x_{1j} - \mu_1)^2}{x_{1j}} + \sum_{j=1}^{n_2} \frac{(x_{2j} - \mu_2)^2}{x_{2j}}$ 이다. 그리고,

$$\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} \Pi_2(\mu_1, \mu_2, \lambda)\,[L_2(\mu_1, \mu_2, \lambda)]^b \, d\lambda\, d\mu_1\, d\mu_2 = \frac{\Gamma\left(\frac{nb}{2}\right) T(b : x)}{(2\pi)^{\frac{nb}{2}}\, w^{\frac{b}{2}}}$$

이다. 여기에서 $T(b: x) = \int_0^{\infty} \mu_1^{-\frac{3}{2}} \mu_2^{-\frac{3}{2}} (Bb)^{-\frac{nb}{2}} d\mu_1\, d\mu_2$ 이다.
그러므로,

$$q_2(b : x) = \frac{\Gamma\left(\frac{n}{2}\right) T(1 : x)}{(2\pi)^{\frac{n}{2}}\, w^{\frac{1}{2}}} \; \frac{(2\pi)^{\frac{nb}{2}}\, w^{\frac{b}{2}}}{\Gamma\left(\frac{nb}{2}\right) T(b : x)} \tag{3.5}$$

이다.

가설 H_1의 H_2에 대한 FBF는 식(3.4)와 (3.5)에 의하여 다음과 같이 구해진다.

$$B_{12}^{\,b}(x) = \frac{q_1(b : x)}{q_2(b : x)} = \frac{S(1 : x)}{T(1 : x)} \cdot \frac{T(b : x)}{S(b : x)} \tag{3.6}$$

위의 식 (3.6)을 이용하여 가설 H_1과 가설 H_2가 참일 사후확률을 계산하여 모형 선택을 한다.

제 4 절 예제를 통한 비교

다음은 실제 자료를 통하여 식 (3.6)에 주어진 FBF을 이용하여 검정결과를 알아보고자 한다.

예제 1. 다음에 주어진 20개의 자료는 고속 터빈 베어링(high-speed turbine bearing)의 성능을 비교하기 위하여 2개의 다른 합성물로 만든 실험결과이다. 실험에서 각 종류당 10개의 베어링을 실험하여 백만 사이클(millions of cycles)당 고장시간을 기록한 것이다.(Chhikara and Folks(1989))

집단 1	3.03 5.53 5.60 9.30 9.92 12.51 12.95 15.21 16.04 16.84
집단 2	3.19 4.26 4.47 4.53 4.67 4.69 5.78 6.79 9.37 12.75

위와 같이 역가우시안분포를 따르는 자료를 이용하여 아래의 가설을 검정하고자 한다.

$$H_1 : \mu_1 = \mu_2 \ vs. \ H_2 : \mu_1 \neq \mu_2 .$$

각 모형에 대한 사전확률을 1/2로 가정하고 식 (3.6)을 이용한 H_1의 H_2에 대한 FBF값과 사후확률은 다음과 같다

$$B^b_{21} = 3.1399, \qquad P(H_1 : x) = 0.2415$$

여기에서 b은 2절에서 제안한 (1)을 사용하였고 값은 3/20입니다. 위의 결과로부터 가설 H_2이 선택됨을 알 수 있다.

또한 고전적인 가설 검정법을 이용하여 위에 주어진 가설을 검정해 보고자 한다. 모평균 차이에 대한 t-검정을 실시한 결과 t-값은 2.4857이 되고, p-값은 0.0230이 되어 역시 H_2가 선택이 되었다.

예제 2. Gacula and Kubala(1975)는 냉장식품 M과 K에 대한 어떤 지각기관의 고장자료를 보고했다. 여기서 자료는 저장수명모형인 와이블분포와 로그정규분포에서 연구하였다. 위 분포들은 다음과 같이 주어진 고장자료에 대해 적합하다고 한다. 또한 각 생산품에 대한 자료는 역가우시안 분포로 잘 설명된다(Chhikara와 Folks(1989)).

집단 M	24 24 26 26 32 32 33 33 33 35 41 42 43 47 48 48 48 50 52 54 55 57 57 57 57 61
집단 K	21 23 25 38 43 43 52 56 61 63 67 69 70 75 86 107

위의 자료를 이용하여 다음과 같은 가설을 검정하고자 한다.

$$H_1 : \mu_1 = \mu_2 \ vs. \ H_2 : \mu_1 \neq \mu_2 .$$

각 모형에 대한 사전확률을 1/2로 가정하고 식 (3.6)을 이용한 H_1의 H_2에 대한 FBF값과 사후확률은 다음과 같다

$$B^b_{21} = 1.7110, \qquad P(H_1 : x) = 0.3689$$

여기에서 $b=m/n$은 3/42입니다. 위의 결과로부터 가설 H_2이 선택됨을 알 수 있다.
그리고 고전적인 가설 검정법을 이용하여 위에 주어진 가설을 검정해 보고자 한다. 모평균 차이에 대한 t-검정을 실시한 결과 t-값은 2.2460이 되고, p-값은 0.0303이 되어 역시 H_2가 선택이 되었다.

참고문헌

1. Banerjee, A. K. and Bhattacharyya, G. K.(1979). Bayesian Results for the Inverse Gaussian Distribution with Application, *Technometrics, Vol. 21, 247-251.*
2. Berger, J. O. and Pericchi, L. R.(1996). The Intrinsic Bayes Factor for Mddel Selection and Prediction, *Journal of American Statistical Association, 91, 109-122.*
3. Berger, J. O. and Pericchi, L. R.(1998). Accurate and Stable Bayesian Model Selection: The Median Intrinsic Bayes Factor, *Sankhya, B, 91, 1-18.*
4. Chhikara, R. S.(1975). Optimum Tests for the Comparison of Two Inverse Gaussian Distribution Means. *Austral. J. Statist., 17, 77-83.*
5. Chhikara, R. S. and Folks, J. L.(1989). *The Inverse Gaussian Distribution,* Marcel Dekker, Inc. New York and Basel.
6. Gacula, M, C., Jr., and Kubala, J. J.(1975). Statistical Models for the Shelf Life Falilures. *J. Food Sci., 40, 404-409.*
7. O'Hagan, A.(1995). Fractional Bayes Factor for Model Comparison, *Journal of the Royal Statistical Society, B, 57, 99-138.*
8. Padgett, W. J.(1981). Bayes Estimation of Reliability for the Inverse Gaussian Model, *IEEE Transactions on Reliability, Vol. R-30, 384-385.*
9. Seshadri, V.(1999). *The Inverse Gaussian Distribution: Statistical Theory and Applications.* Springser, New York.

[2005년 5월 접수, 2005년 7월 채택]