The Korean Communications
in Statistics Vol. 8, No. 3, 2001
pp. 787-795

A Bayesian Hypothesis Testing Procedure Possessing the Concept of Significance Level[1)]

Hyungtae Hwang[2)]

Abstract

In this paper, Bayesian hypothesis testing procedures are proposed under the non-informative prior distributions, which can be thought as the Bayesian counterparts of the classical ones in the sense of using the concept of significance level. The performances of proposed procedures are compared with those of classical procedures through several examples.

Keywords : Bayesian hypothesis testing procedure, classical hypothesis testing procedure, Bayes factor, non-informative prior distribution, significance level

1. 서 론

먼저, 빈도론적 관점(Frequentist's viewpoint)에 의한 고전적 통계학(Classical Statistics)과 베이즈 관점(Bayesian viewpoint)에 의한 베이즈 통계학(Bayesian Statistics)을 간단하게 비교・검토하기 위하여 다음의 [예1-1]을 살펴보자.

[예1-1] 항아리 I과 항아리 II가 아래의 [표1-1]과 같이 각각 1,000개씩의 공들을 포함하고 있다고 하자.

[표1-1] 항아리 I과 II가 포함하고 있는 공들 (단위 : 개)

공 / 항아리	빨간 공	검은 공	흰 공	계
항아리 I	45	55	900	1,000
항아리 II	5	990	5	1,000

이제, 두 개의 항아리 중에서 임의로 하나의 항아리가 선택되어 어떤 통계학자에게 보내졌다고 하자. 어떤

1) This work was supported by Dankook University Research Fund in 2000
2) Professor, Department of Computer Science and Statistics, Dankook University, Seoul, 140-714, Korea, E-mail : hthwang@dankook.ac.kr

항아리가 자신에게 주어졌는지 알 수 없는 통계학자는 다음과 같이 θ를 정의하였다.

$$\theta = \begin{cases} 1 \text{ , 만일 주어진 항아리가 항아리 } I \text{ 이면,} \\ 2 \text{ , 만일 주어진 항아리가 항아리 } II \text{ 이면.} \end{cases}$$

통계학자는 주어진 항아리로부터 하나의 공을 임의로 추출할 예정이며 추출되는 공의 색을 관측치로 하려고 한다. 통계학자의 입장에서는 θ의 값이 미지의 값이지만 이미 고정되어 있고, 그 값이 관측치의 분포를 결정하므로 θ는 미지의 모수로 간주된다. 이제 θ의 값이 1인가를 유의수준 $\alpha = 0.1$에서 검정하기 위하여, $H_0: \theta = 1$ 을 귀무가설로 하고, 대립가설을 $H_1: \theta = 2$ 로 설정하였다.

우선, 고전적 통계학의 최강력검정(most powerful test)에서 유의수준 0.1의 기각역이 {빨간 공, 검은 공}임은 쉽게 구해진다. 이제 이 항아리에서 하나의 공을 임의추출한 결과 빨간 공이 추출되었다고 하자. 빨간 공은 기각역에 속하므로 관측치 빨간 공에 대하여, 귀무가설은 기각되고 상당한 증거에 의하여 대립가설이 채택됨으로써, 주어진 항아리는 항아리 II였다는 추론결과를 얻게되며, 이때의 p-값은 유의수준과 같은 0.1이 된다.

그러나 [표1-1]을 자세히 살펴보면, 이 결론은 정상적인 것으로 받아들이기 어렵다. 왜냐하면, 항아리 II에 있는 빨간 공의 비율은 항아리 I에 비하여 1/9밖에 되지 않으므로, 상식적으로 관측치 빨간 공이 항아리 II에 대한 상당한 증거라고는 여겨지지 않기 때문이다.

전통적 가설검정 방법의 문제는 여기에서 그치지 않는다. 이제, 똑같은 상황에서 검은 공과 흰 공을 한데 묶어서 빨갛지 않은 공으로 분류하였다고 하자. 이 때는 고전적 통계학의 최강력검정에서 동일한 관측치 빨간 공에 대한 p-값은 1.0으로 되어 유의수준 0.1에서의 최강력검정은 당연히 귀무가설을 기각하지 않는다. 즉, 두 항아리에서 빨간 공의 비율에 변함이 없고, 관측치도 여전히 빨간 공인데도 불구하고, 검은 공과 흰 공을 합친다든가 또는 검은 공과 흰 공을 재질에 따라 더 세분한다든가 등의 나머지 공들을 분류하는 방법에 따라서 p-값이 극단적으로 달라지게 되며, 따라서 검정의 결과가 얼마든지 뒤집어질 수 있는 바람직하지 않은 성질을 보여주는 것이다.

이제, 베이즈 관점에서 생각해 보자. 주어진 항아리에 대한 아무런 사전정보가 없으므로, 귀무가설과 대립가설 양쪽에 각각 1/2씩의 사전확률을 부여하자. 그러면 빨갛지 않은 공들을 분류하는 방법에 관계없이, 관측치 빨간 공이 주어졌을 때 귀무가설이 참일 사후확률은 0.9이며, 이는 귀무가설이 참임을 강력하게 시사한다고 볼 수 있다. 이 결과는 [표1-1]을 살펴본 후의 소감과 잘 일치하며, 현재의 관측치 빨간 공 이외의 빨갛지 않은 공의 재분류방법에 전혀 영향을 받지 않는다는 점에서도 전통적 가설검정 방법에 비하여 고무적이다.

위의 [예1-1]에서 드러난 문제점이야말로 통계학에 대한 빈도론적 접근이 지니고 있는 하나의 제한점으로 지적될 수 있다. 기본적으로, 빈도론적 관점에서의 확률적 진술은 현재의 관측치에 대한 것이 아니라 미래에 가상적으로 반복될 실험에 대한 것이기 때문에, 현재의 관측치가 빨간 공이고 각 항아리의 빨간 공의 비율이 일정하다고 하더라도, 나머지 빨갛지 않은 공의 분류 방법은 검정의 결과에 영향을 미치게 된다. 즉, 현재의 관측치가 빨간 공이라고 하더라도 미래에 가상적으로 반복될 실험에서는 얼마든지 빨갛지 않은 공이 관측될 수 있으며, 확률적 진술은 미래의 가상적 실험에 대한 것이기 때문에 빨갛지 않은 공의 분류 방법은 검정의 결과에 영향을 미칠 수밖에 없는 것이다.

여기에 반하여, 베이즈 관점은 이와 같은 문제점을 발생시키지 않는다. 왜냐하면, 사후분포는 사전분포와 현재의 관측치에 대한 우도함수에 의해서만 결정되며, 베이즈 통계학에서의 추론은 오직 사후분포를 통해서만 이루어지기 때문이다. 빈도론적 관점에 대한 또 다른 비판들은 Berger(1985) 등에서 찾아볼 수 있다.

이 논문에서는 무정보적 사전분포(non-informative prior distribution)의 가정 하에 전통적 가설검정 방법

의 속성을 계승하는 베이즈 가설검정 방법을 제안하고자 한다. 이 논문에서 제안하는 베이즈 가설검정 방법은 기본적으로 Lindley(1965)가 단일점 귀무가설(point null)에 대한 양측검정에 적용한 아이디어에 기초하였으며, 가설의 형식에 따라 적절한 형태의 베이즈 신뢰영역을 이용함으로써 가설검정 방법을 제시하였다.

몇 가지 예를 통하여, 전통적 가설검정 방법이 안정적으로 쓰이고 있는 통상적인 경우에 있어서는, 그 해석 방법은 비록 다르지만 추론결과의 면에서만 볼 때 제안된 베이즈 가설검정 방법이 전통적 가설검정 방법의 경우와 일치하는 경향이 있음을 보여주었다. 하지만, [예1-1]에서와 같이 빈도론적 관점이 정상적인 기능을 보여주지 못하는 경우에 있어서는, 전통적 가설검정과 차별화 되는 바람직한 성능을 보여주고 있음을 또 다른 예를 통하여 보여주었다.

2. 전통적인 형식의 가설검정과 베이즈요인

Lehmann(1986)은 그의 저서 서문에서 그의 저서가 통계적 가설검정을 주제로 함에도 불구하고, '베이즈 신뢰구간은 다루지만 베이즈 검정방법은 그 역할이 제대로 잘 정의되어 있지 않음을 들어 다루지 않는다'고 진술한 바 있다. 물론, 베이즈 통계학에서도 Kim and Son (2001)과 같이 베이즈요인의 개념을 이용하여 가설검정을 수행하고 있기는 하지만, Lehmann(1986)은 베이즈요인을 이용한 가설검정 방법이 전통적 가설검정 방법에 대응하여 비교될 수 있는 방법으로는 생각하지 않았던 것이다. Cox and Hinkley(1974, p392) 또한 전통적 가설검정 방법에 대응되는 베이즈 가설검정 방법이 없는 점이 가설검정에 있어서 베이즈 접근의 약점(weakness)이라고 기술한 바 있다.

먼저 전통적 가설검정의 특징을 간단하게 고찰해 보자. 전통적 가설검정은 기본적으로, '주어진 자료가 대립가설에 비하여 귀무가설이 참이 아니라는 상당한 근거를 제시하는가'에 대한 판단으로 이해되어야 한다. 따라서 당연히 귀무가설과 대립가설은 대칭적인 구조로 설정되지 않으며, 검정통계량의 기각역은 귀무가설이 참이라는 가정하에서 결정하게 된다. 전통적 가설검정의 결과, 귀무가설이 기각되었을 경우에는 '주어진 자료에 대립가설에 비하여 귀무가설이 참이 아님을 시사하는 상당한 증거가 포함되어 있다'는 의미로 이해되어야 하고, 반대로 귀무가설이 기각되지 않았을 경우에는 '주어진 자료에 대립가설에 비하여 귀무가설이 참이 아님을 시사하는 증거가 충분하지 않다'는 의미로 이해되어야 한다. 즉, 전통적 가설검정에서는 '두 가설 중 어떤 가설이 참인가'를 대칭적인 입장에서 고려하는 것이 아니라, '주어진 자료가 대립가설에 비하여 귀무가설이 참이 아님을 시사하고 있는가'가 주요 관심이라고 보아야 할 것이다.

한편, 베이즈 통계학에서도 베이즈요인의 개념을 이용하여 가설검정을 수행하고 있으나, 전통적 가설검정 방법에 비하여 유의수준 등 오류의 확률에 대한 관리가 결여되어 있고, 기각역의 설정에 대한 기준이 명확하게 제시되지 못하는 등의 문제점이 우선 지적될 수 있다.

그러나, 베이즈요인에 의한 가설검정과 전통적인 형식의 가설검정과의 차이점은 보다 더 근본적인데서 찾아볼 수 있다. 기본적으로 베이즈요인은 1을 중심으로 곱에 대하여 대칭적인 성질을 갖는다는 점에 주목할 필요가 있다. 즉, 귀무가설과 대립가설을 뒤바꾸는 경우 베이즈요인의 값은 원래의 값의 역수가 된다. 이와 같은 베이즈요인의 성질은 자연스럽게 귀무가설과 대립가설의 관계를 대칭적인 구조로 설정한다. 베이즈요인의 작은 값이 '귀무가설이 참이 아니라는 상당한 근거'를 제공한다면, 대칭적으로 그의 역수인 베이즈요인의 큰 값은 '대립가설이 참이 아니라는 상당한 근거'를 제공하게 된다. 따라서 베이즈요인에 의한 방법은 '대립가설에 대하여 귀무가설을 검정'하는 전통적인 형식의 가설검정과는 달리, '귀무가설과 대립가설 중에서 하나를 택하는(또는 아무것도 택하지 않는)' 선택(choice)의 문제에 더 가깝다고 보아야 한다. 두 방법은 구조적으로 비교될 수 있는 성질이 아닌 것이다.

베이즈요인에 의한 가설검정의 또 다른 문제점을 예시하기 위하여, Lindley(1957)에 의하여 제시된 Lindley's paradox의 한 형태로서 다음의 예를 인용하도록 한다.

[예2-1] 표본 $x_1, \cdots, x_n$은 정규분포 $N(\theta, \varphi)$로부터의 크기 n인 확률표본이며, φ는 알고 있는 값이라고 하자. 검정하고자하는 가설은 다음과 같다.

$$H_0: \theta = \theta_0 \quad v.s. \quad H_1: \theta \neq \theta_0 \tag{2.1}$$

이 문제를 베이즈요인에 의하여 추론하기 위하여, 귀무가설과 대립가설에 각각 1/2 의 사전확률을 부여하고, 대립가설 하에서 θ에 대한 사전분포로서 $N(\mu, \psi)$를 가정하자. 그러면 관측치 $x_1, \cdots, x_n$에 대하여 베이즈요인 B는 다음과 같이 구해진다.

$$B = \frac{\{2\pi\varphi/n\}^{-1/2}\exp[-\frac{1}{2}(\bar{x}-\theta_0)^2/(\varphi/n)]}{\{2\pi(\psi+\varphi/n)\}^{-1/2}\exp[-\frac{1}{2}(\bar{x}-\mu)^2/(\psi+\varphi/n)]} \tag{2.2}$$

만일 대립가설 하에서의 사전정보가 거의 없다면 분석자는 대립가설 하에서의 사전분포에 대하여 무정보적 사전분포를 고려하게 될 것이며, 이는 식(2.2)에서 ψ의 값을 크게 함으로써 동등한 결과를 얻을 수 있다. 그러나 식(2.2)에서 ψ의 값을 크게 하면 크게 할수록 B의 값은 관측치에 관계없이 ∞로 커지게 됨을 쉽게 알 수 있다. 즉, 대립가설 하에서의 사전정보가 전혀 없다면, 식(2.1)로써 주어지는 형태의 가설에 대한 베이즈요인의 검정결과는 관측치에 관계없이 귀무가설을 채택하게 되는 문제점을 갖고 있는 것이다.

3. 베이즈 가설검정 방법의 제안

일반적으로, 모수 θ의 값이 표본에 의존하는 어떤 영역 $C_{1-\alpha}$에 속할 사후확률이 $1-\alpha$ 이상일 때 영역 $C_{1-\alpha}$를 모수 θ에 대한 $100(1-\alpha)\%$ 신뢰영역(credible region)이라고 하고, $100(1-\alpha)\%$ 신뢰영역 $C_{1-\alpha}$ 내에서의 사후밀도함수의 값들이 항상 $C_{1-\alpha}$ 밖에서의 사후밀도함수의 값들 이상일 때 $C_{1-\alpha}$를 모수 θ에 대한 $100(1-\alpha)\%$ HPD 영역(highest posterior density region)이라고 한다.

Lindley(1965)는 단일점 귀무가설에 대한 양측검정을 실시함에 있어서, 귀무가설에 의하여 지정된 모수의 값이 $100(1-\alpha)\%$ HPD 영역에 속하지 않을 때만 귀무가설을 기각하는 베이즈 검정방법을 제안하였다. 그는 자신이 제안한 검정방법이 '모호하고 확산된(vague and diffuse)' 사전분포를 가정하는 경우에만 유효하다는 점을 강조하였다.

Lindley(1965)의 방법은 전통적 가설검정에서의 유의수준의 개념을 계승하는 베이즈 가설검정 방법으로서, Zellner(1971), Lee(1989) 등에 의하여 다시 소개되고 제한적인 통계모형에 대하여 부분적으로 적용되기도 하였으나, 단일점 귀무가설에 대한 양측검정이라는 특별한 경우에만 적용된다는 제한점으로 해서 널리 사용되지는 못하였다. 이 절에서 제안하는 베이즈 검정방법은 Lindley(1965)의 방법을 일반적인 가설들에 적용되도록 일반화한 것이며, 사전분포는 주어진 모형에 적합하는 무정보적 사전분포만을 가정하도록 한다.

모수 θ에 대하여 표본이 주어져 있을 때, 검정하고자 하는 가설들은 다음과 같다.

$$H_0: \theta \in \Omega_0 \quad v.s. \quad H_1: \theta \in \Omega_1 \tag{3.1}$$

모수 θ에 대하여 무정보적 사전분포를 가정하고, $C_{1-\alpha}$를 θ에 대한 $100(1-\alpha)\%$ 신뢰영역이라고 하

자. 이때 제안되는 베이즈 가설검정 방법은 다음과 같다.

$$C_{1-\alpha} \subset \Omega_1 \text{ 일 때만 } H_0\text{를 기각한다.} \tag{3.2}$$

식(3.2)에 의한 가설검정 방법을 신뢰영역 $C_{1-\alpha}$에 대한 유의수준 α의 베이즈 가설검정 방법(Bayesian hypothesis testing procedure of significance level α with respect to $C_{1-\alpha}$)이라고 부르자. 여기에서 α를 유의수준이라 한 것은, 위의 검정방법이 적절한 형태의 $C_{1-\alpha}$에 대하여 유의수준 α의 전통적 가설검정 방법과 일치하는 경향을 흔히 보이기 때문이다.

신뢰영역 $C_{1-\alpha}$는 다음과 같이 선택된다.

첫째, 단측검정에서는 대립가설에서 지정되는 모수의 영역과 같은 방향의 단측 신뢰영역을 사용한다. 예를 들어, 대립가설이 $H_1 : \theta > \theta_0$와 같은 형태일 때는 $C_{1-\alpha} = [t(x), +\infty)$의 형태로, 대립가설이 $H_1 : \theta < \theta_0$와 같은 형태일 때는 $C_{1-\alpha} = (-\infty, t(x)]$의 형태로 신뢰영역을 구한다.

둘째, 단측검정 이외에는 HPD 영역을 이용한다.

위에서 제안된 베이즈 가설검정 방법은 단일점 귀무가설에 대한 양측검정에 있어서는 Lindley(1965)의 검정방법과 일치하지만, 일반적인 형태의 가설에 대한 검정이 가능하도록 일반화한 것이며, 이를 위해 HPD 영역 대신에 검정의 형태에 따라 적절한 형태의 신뢰영역을 이용할 수 있도록 한 것이다.

베이즈 가설검정의 추론결과에 대한 해석은 다음과 같다.

만일 귀무가설이 기각되고 대립가설이 채택되었을 경우에는, 신뢰영역이 대립가설에 의한 모수의 영역에 포함되므로, 표본에 의하여 대립가설이 참이라는 상당한 증거가 나타난 것으로 해석될 수 있다. 반면에, 만일 귀무가설이 채택되는 경우에는 귀무가설이 참이라는 증거가 나타난 것이 아니라 귀무가설이 거짓이라는 증거가 충분하지 않은 것으로 해석되어야 한다. 이와 같은 해석방식은 기존의 전통적 가설검정의 해석방식과 속성을 같이 하는 것이며, 유의수준의 개념을 통하여 가설검정을 수행한다는 점에서도, 이 방법은 전통적 가설검정 방법의 속성을 계승하는 베이즈 가설검정 방법으로 간주될 수 있는 것이다.

그러나 그럼에도 불구하고 전통적 가설검정 방법은 빈도론적 관점에서, 베이즈 가설검정 방법은 베이즈 관점에서 출발한다는 기본적인 차이점이 간과되어서는 안된다. 이런 차이점으로 인하여, [예1-1]에서와 같은 전통적 가설검정 방법의 바람직하지 못한 성질은 다음 절의 [예4-2]와 [예4-3]에서 보는 바와 같이 베이즈 가설검정 방법에서는 나타나지 않는다. 또한 베이즈 가설검정의 결과 대립가설이 채택된 경우에, '대립가설이 참일 확률이 $1-\alpha$ 이상' 또는, '귀무가설이 참일 확률이 α 이하'라는 식의 쉬우면서도 통계수요자의 요구에 부응하는 자연스러운 확률적 진술이 가능한 반면에, 전통적 가설검정 결과에 대한 확률적 진술은 상대적으로 매우 어려우며 자연스럽지도 못하다는 차이점도 지적될 수 있다.

4. 몇 가지 예를 통한 베이즈 가설검정 방법의 성능 검토

이 절에서는 앞의 3절에서 제시된 베이즈 가설검정 방법을 몇 가지 예를 통하여 여러 가지 모형에 대하여 적용해 보고, 그 성능을 검토하도록 한다.

다음의 [예4-1]은 정규모형의 모수들에 대하여 베이즈 가설검정 방법을 적용한 것이다.

[예4-1] 표본 $x_1, \cdots, x_n$은 정규분포 $N(\mu, \sigma^2)$로부터의 크기 n인 확률표본이며, 모수 (μ, σ^2)에 대하여 다음과 같은 무정보적 사전분포를 가정하자.

$$h(\mu, \sigma^2) \propto 1/\sigma^2 \tag{4.1}$$

이 때, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 와 $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 은 모수 μ, σ^2 의 충분통계량이므로 주어진 자료는 $\bar{x}, s^2$ 로 축소될 수 있으며, 다음과 같은 사후분포들이 쉽게 구해질 수 있다.

$$\frac{\mu-\bar{x}}{s/\sqrt{n}} \mid \bar{x}, s^2 \sim t(n-1) \tag{4.2}$$

$$\frac{(n-1)s^2}{\sigma^2} \mid \bar{x}, s^2 \sim \chi^2(n-1) \tag{4.3}$$

이제 유의수준 α 에서 다음과 같은 가설들에 대한 베이즈 가설검정 방법들을 구해보자.

① $H_0 : \mu=\mu_0 \quad v.s. \quad H_1 : \mu\neq\mu_0$

식(4.2)로부터, 구간 $[\bar{x}-t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{x}+t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}]$ 는 μ 에 대한 $100(1-\alpha)\%$ HPD 영역이므로, 유의수준 α 의 베이즈 가설검정 방법은 μ_0 가 이 영역 밖에 있을 때, 즉,

$$\frac{|\bar{x}-\mu_0|}{s/\sqrt{n}} > t_{n-1,\alpha/2} \tag{4.4}$$

일 때 H_0 를 기각한다. 이 결과는 기존의 전통적 가설검정에서 유의수준 α 의 우도비검정 (likelihood ratio test)과 일치한다.

② $H_0 : \mu\leq\mu_0 \quad v.s. \quad H_1 : \mu>\mu_0$

식(4.2)로부터, 구간 $[\bar{x}-t_{n-1,\alpha}\frac{s}{\sqrt{n}}, +\infty)$ 는 μ 에 대한 $100(1-\alpha)\%$ 신뢰영역이므로, 유의수준 α 의 베이즈 가설검정 방법은 μ_0 가 이 영역 밖에 있을 때, 즉,

$$\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1,\alpha} \tag{4.5}$$

일 때 H_0 를 기각한다. 이 결과는 기존의 전통적 가설검정에서 유의수준 α 의 우도비검정과 일치한다.

③ $H_0 : \mu\geq\mu_0 \quad v.s. \quad H_1 : \mu<\mu_0$

식(4.2)로부터, 구간 $(-\infty, \bar{x}+t_{n-1,\alpha}\frac{s}{\sqrt{n}}]$ 는 μ 에 대한 $100(1-\alpha)\%$ 신뢰영역이므로, 유의수준 α 의 베이즈 가설검정 방법은 μ_0 가 이 영역 밖에 있을 때, 즉,

$$\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha} \tag{4.6}$$

일 때 H_0 를 기각한다. 이 결과는 기존의 전통적 가설검정에서 유의수준 α 의 우도비검정과 일치한다.

④ $H_0 : \sigma^2\leq \sigma_0^2 \quad v.s. \quad H_1 : \sigma^2> \sigma_0^2$

식(4.3)으로부터, 구간 $[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha}}, +\infty)$ 는 σ^2 에 대한 $100(1-\alpha)\%$ 신뢰영역이므로, 유의수준 α 의 베이즈 가설검정 방법은 σ_0^2 가 이 영역 밖에 있을 때, 즉,

$$\frac{(n-1)s^2}{\sigma_0^{\ 2}} \geq \chi^2_{\ n-1,\alpha} \tag{4.7}$$

일 때 H_0 를 기각한다. 이 결과는 기존의 전통적 가설검정에서 유의수준 α 의 우도비검정과 일치한다.

위의 [예4-1]에서는, 정규모형과 같이 전통적 가설검정 방법이 안정적으로 사용되고 있는 경우에 있어서는, 3절에서 제안된 베이즈 가설검정 방법이 검정의 결과 면에서 전통적 가설검정 방법과 일치하는 경향이 있음을 예시하였다. 그러나 [예1-1]에서와 같이 전통적 가설검정 방법이 이상기능을 보이는 경우에 베이즈 가설검정 방법의 성능을 검토하기 위하여, 다음의 [예4-2]와 [예4-3]을 계속하여 살펴보자.

[예4-2] 앞의 1절에서 주어진 [예1-1]의 경우에 대하여 제안된 베이즈 가설검정 방법을 이용하여 접근해보자. θ 는 1 또는 2의 값을 취할 수 있으므로 다음과 같은 무정보적 사전분포를 가정한다.

$$h(1) = h(2) = 1/2 \tag{4.8}$$

이 때, 주어진 관측치 빨간 공에 대한 θ 의 사후분포는 다음과 같다.

$$p(1 \mid \text{빨간 공}) = 0.9, \quad p(2 \mid \text{빨간 공}) = 0.1 \tag{4.9}$$

따라서 주어진 관측치 빨간 공에 대하여 1 보다 2 를 우선적으로 포함하는 θ 의 90% 단측 신뢰영역은 $\{1,2\}$가 되므로, 귀무가설 $H_0: \theta = 1$ 은 유의수준 $\alpha = 0.1$ 에서 기각되지 않는다. 이 결과는 전통적 가설검정의 결과와 상반되며, [표1-1]을 살펴본 후의 직관과 잘 일치한다. 또한 이 결과는 빨갛지 않은 공의 재분류방법에 전혀 영향을 받지 않는다는 점에서도 전통적 가설검정 방법에 비하여 우월하다고 볼 수 있다.

[예4-3] $z \sim N(\mu, 1)$ 에 대하여 다음과 같은 단순가설들을 생각해 보자.

$$H_0 : \mu = 0 \quad v.s. \quad H_1 : \mu = 0.01 \tag{4.10}$$

z 의 관측값이 $z = 3$ 이라고 하면, 전통적 가설검정의 결과는 매우 강력한 증거에 의하여 H_0 를 기각하고 H_1 를 채택한다. 왜냐하면 H_0 가 참일 때 z 의 값이 3 이상일 확률이 대략 0.13%로서 매우 작기 때문이다. 그러나 이 경우에 H_1 이 참이라고 하더라도 z 의 값이 3 이상일 확률은 대략 0.14% 로서 H_0 가 참인 경우와 별 차이가 없어서, 직관적으로 $z = 3$ 이라는 관측치가 H_1 에 비하여 H_0 가 거짓이라는 특별한 증거로는 보여지지 않기 때문에, 매우 강력한 증거에 의하여 H_1 을 채택한 검정의 결과는 정상적인 것으로 보기 어렵다.

이 문제를 베이즈 가설검정 방법에 의하여 접근해보자. 우선 모수공간이 $\{0,\ 0.01\}$이므로, 다음과 같은 무정보적 사전분포를 가정한다.

$$h(0) = h(0.01) = 1/2 \tag{4.11}$$

이에 따라 $z = 3$ 에 대하여 다음과 같은 μ 의 사후분포가 구해진다.

$$\begin{aligned} p(0 \mid z=3) &= \frac{\exp(-3^2/2)}{\exp(-3^2/2) + \exp(-2.99^2/2)} = 0.4925 \\ p(0.01 \mid z=3) &= \frac{\exp(-2.99^2/2)}{\exp(-3^2/2) + \exp(-2.99^2/2)} = 0.5075 \end{aligned} \tag{4.12}$$

따라서 $\alpha \leq 0.4925$ 인 모든 α 에 대하여, μ 에 대한 $100(1-\alpha)\%$ 신뢰영역은 $\{0, 0.01\}$ 로 주어지므로,

관측치 $z=3$ 에 대하여 유의수준이 0.4925 이하인 베이즈 검정방법은 귀무가설을 기각하지 않는다. 즉, 대립가설에 비하여 귀무가설이 거짓이라는 특별한 증거가 없다는 의미이다. 이 결과는 앞에서의 전통적 가설검정의 결과와 달리 직관에 잘 부합하는 결과로 볼 수 있다.

마지막으로, 베이즈요인에 의한 가설검정의 문제점이 예시된 [예2-1]의 모형에 θ에 대한 무정보적 사전분포를 가정하여 베이즈 가설검정 방법을 적용하면, 기존의 전통적 가설검정에서의 우도비검정과 일치하는 통상적인 검정방법을 얻게 된다. 따라서, 대립가설 하에서의 사전정보가 전혀 없을 때 관측치에 관계없이 귀무가설을 채택하게 되는 베이즈요인의 문제점은 우리의 베이즈 가설검정 방법에서는 나타나지 않는다.

5. 결 론

이 논문에서는 무정보적 사전분포의 가정 하에서 유의수준의 개념을 갖는 베이즈 가설검정 방법을 제안하였다. 기본적으로, 충분한 증거가 있을 때만 대립가설을 채택할 것인가, 아니면 가능성이 더 큰 가설을 채택할 것인가의 검정방법의 선택은 분석자가 주어진 문제에 따라 판단해야 할 일이다. 보다 구체적으로, 두 가설이 대칭적인 경우, 즉, 제1종의 오류와 제2종의 오류의 심각성이 대등하다고 생각되는 문제에 있어서는 베이즈요인에 의한 가설검정 방법과 같이 가능성이 더 큰 가설을 채택하는 가설검정방법을 적용하는 것이 타당하다고 판단되며, 제1종의 오류가 상대적으로 치명적인 결과를 초래하는 문제에 있어서는 충분한 증거에 의해서만 대립가설을 채택하는 가설검정형식의 관점이 유효한 것으로 판단된다. 이 논문에서는 후자의 경우에 대하여 전통적 가설검정에 대응되는 베이즈 가설검정 방법을 제안한 것이다. 몇 가지 예를 통하여 제안된 베이즈 가설 검정 방법이 다음과 같은 특징들을 갖고 있음을 보여 주었다.

1) 유의수준의 개념을 통하여 전통적인 형식의 가설검정의 속성을 계승하였다.
2) 전통적 가설검정 방법이 안정적으로 기능하는 경우에는 전통적 가설검정법과 일치하는 경향을 볼 수 있었지만, 전통적 가설검정 방법이 이상기능을 보이는 예에서는 차별화 되어 바람직한 검정결과를 갖는 성능을 보여 주었다.
3) 전통적인 가설검정 방법에 비하여, 여타의 베이즈 추론들과 마찬가지로 검정 결과에 대하여 보다 자연스럽고 직관적이며 통계수요자의 요구에 부응하는 해석을 제공하게 되었다.
4) 기존의 베이즈요인에 의한 가설검정 방법에 비하여는, 유의수준의 개념을 통함으로써 기각역의 결정이 쉽고 명확하며, 베이즈요인의 기능이 문제가 있는 경우에도 바람직한 성능을 유지하는 등의 장점들을 갖고 있었다.

이 논문에서 제안된 베이즈 가설 검정 방법을 적용하기 위해서는 적절한 무정보적 사전분포가 전제되어야 하는데, 그렇지 못한 경우에는 이 방법의 적용이 곤란하다는 문제점을 갖고 있다. 그러나 그런 경우에도 만일 모수에 대한 사전정보가 존재하여 적당한 사전분포가 가정될 수 있다면, 가정된 사전분포를 무정보적 사전분포 대신 사용할 수 있을 것이다.

마지막으로 언급하고자 하는 것은, 이 논문에서 제안된 베이즈 가설검정 방법은 많은 경우에 전통적 가설검정 방법과 일치하는 검정결과를 보여주고 있으므로, 베이즈 가설검정을 수행하기 위한 별도의 통계 패키지 프로그램이 대부분 요구되지 않는다는 점이다.

References

[1] Berger, J. (1985). *Statistical decision theory and Bayesian analysis, Second edition*, Springer-Verlag, New York.

[2] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

[3] Kim, H.J. and Son Y.S. (2001). A Bayesian criterion for a multiple test of two multivariate normal population, *The Korean Communications in Statistics*, Vol 8, No. 1, 97-107.

[4] Lee, P.M. (1989). *Bayesian Statistics : an Introduction*, Edward Arnold, London.

[5] Lehmann, E.L. (1986). *Testing Statistical Hypotheses, Second edition*, John Wiley & Sons, New York.

[6] Lindley, D.V. (1957). A statistical paradox, *Biometrika* 44, 187-192

[7] Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge: University Press.

[8] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.