일관성의 원리를 충족하는 새로운 형태의 베이즈 P-값의 제안

황형태[1)]

요 약

이 연구에서는 일관성의 원리를 충족하는 새로운 형태의 베이즈 P-값으로 LR형 베이즈 P-값을 제안하고, 그 성질에 대하여 검토해보고자 한다. 제안된 베이즈 P-값은 가능도 비의 단순한 함수의 형태로 표현되어 쉽게 계산될 수 있다는 장점을 갖고 있으며, 검정방법으로서 일관성의 원리를 만족한다.

주요용어 : 일관성의 원리, 베이즈 P-값, 베이즈요인

1. 서론

Schervish(1996)는 일반적인 가설검정방법들에 대하여 일관성의 원리(coherence property)를 다음과 같이 정의하였다.

일관성의 원리 : 가설 H가 가설 H'을 내포하고 있다고 하자. 즉, H가 참이면 H'도 참이라고 하자. 어떤 검정방법에 의하여 가설 H'이 기각될 때 가설 H도 항상 기각된다면 그 검정방법은 일관성이 있다고 한다. 또한, 가설들에 대한 어떤 증거의 측도(measure of support, or measure of evidence)에 대하여 H'에 대한 값이 항상 H에 대한 값 이상인 성질을 가질 때 그 증거의 측도는 일관성이 있다고 한다.

Schervish(1996)는 고전통계학에서 일양최강력불편검정법(UMPU tests)에 의한 P-값이 일관성의 원리를 충족하지 못한다는 사실을 보인 바 있다.

한편 Lavine and Schervish(1999)는 베이즈요인 역시 일관성의 원리를 충족하지 못함을 보임으로써 베이즈요인이 갖고 있는 논리적 약점을 지적하였으며, 비일관성의 문제로 인하여 베이즈요인의 사용이 위험한 결과(dangerous consequences)를 초래할 수도 있음을 경고하였다.

최근에 Hwang(2001)과 Hwang and Oh(2002)는 각각 유의수준의 개념을 갖는 베이즈 가설검정 방법과 그에 따른 베이즈 P-값의 개념을 소개한 바 있다. 그러나 그들의 검정방법 역시 일관성의 원리는 충족하지 못하게 된다.

이 연구에서는 Hwang and Oh(2002)가 제안한 베이즈 P-값을 새로운 관점에서 수정하여 일관성의 원리를 충족하는 새로운 유형의 베이즈 P-값으로 LR형 베이즈 P-값을 제안하고, 그 성질에 대하여 검토한다.

2. 연구 배경

Schervish(1996)가 지적한 바와 같이 고전통계학에서 일양최강력불편검정법에 의한 P-값이

1) (140-714) 서울 용산구 한남동 단국대학교 정보컴퓨터학부 정보통계학전공 교수

비일관적이라는 사실은 다음의 [예 2-1]에서 확인할 수 있다.

[예 2-1] 정규모집단 $N(\mu, \sigma^2)$에서 다음과 같은 두 가지 가설검정을 고려해보자.

$$H_0: \mu=0 \quad v.s. \quad H_1: \mu>0 \tag{2.1}$$

$$H_0: \mu=0 \quad v.s. \quad H_1': \mu\neq 0 \tag{2.2}$$

여기에서 가설 H_1은 가설 H_1'을 내포하지만 단측검정과 양측검정의 특성상 동일한 관측치에 대한 일양최강력불편검정법에 의한 P-값은 가설 H_1에 대한 값보다 가설 H_1'에 대한 값이 더 작은 값을 갖게된다. 이렇게 되면 가설 H_1을 채택하면서도 가설 H_1'은 채택하지 못하는 경우가 발생할 수 있게된다.

한편 Lavine and Schervish(1999)는 베이즈요인 역시 일관성의 원리를 충족하지 못함을 지적하였다. 다음의 예를 들어보기로 하자.

[예 2-2] 모형 $x \mid \theta \sim N(\theta, 1)$ 으로부터 $x=0.5$의 관측값을 얻었으며, 이를 근거로 다음과 같은 가설들에 대한 검정을 실시한다.

$$H_0: \theta\in[-1,0] \quad v.s. \quad H_1: \theta\in(0,1] \tag{2.3}$$

$$H_0: \theta\in[-1,0] \quad v.s. \quad H_1': \theta\in(0,2] \tag{2.4}$$

사전분포에 대하여는 각 가설로 주어지는 θ의 각 영역내부에서 θ에 대한 균등분포를 가정하기로 한다. 이때 식(2.3)과 식(2.4)에서 주어진 가설들에 대한 베이즈요인의 값을 계산해보면 각각 $B_1=0.6310$과 $B_2=0.7738$을 얻게되어 $B_1<B_2$가 된다. 그런데 작은 베이즈요인의 값은 대립가설에 대한 강한 믿음으로 해석되므로, 가설 $H_1: \theta\in(0,1]$이 가설 $H_1': \theta\in(0,2]$를 내포함에도 불구하고 동일한 관측값 $x=0.5$에 대한 베이즈요인의 값들은 H_1'보다 H_1에 더 강한 지지를 보여주는 비일관성의 논리적 모순을 드러내고 있다.

Lavine and Schervish(1999)이 지적한 바와 같이, 베이즈요인의 이러한 비일관성은 가설에 대한 지지의 측도로서 베이즈요인이 자체적으로 갖고 있는 구조적 결함에 기인한다고 볼 수 있다. 이 점을 살펴보기 위하여 일반적으로 다음과 같은 가설들을 고려해 보자.

$$H_0: \theta\in\Theta_0 \quad v.s. \quad H_1: \theta\in\Theta_1 \tag{2.5}$$

이때 베이즈요인은 보통 다음의 식으로 표현될 수 있다.

$$B=\frac{\int_{\theta\in\Theta_0} l(\theta \mid x)\rho_0(\theta)\,d\theta}{\int_{\theta\in\Theta_1} l(\theta \mid x)\rho_1(\theta)\,d\theta} \tag{2.6}$$

여기에서, $l(\theta \mid x)$는 가능도함수(likelihood function)을 뜻하며, ρ_i는 Θ_i의 범위에서 제한된 사전분포를 뜻한다. 즉, ρ_i는 $p(\theta)$가 모수 θ의 사전밀도함수이고, $\pi_i=P[\theta\in\Theta_i]$일 때 다음과 같이 정의된다.

$$\begin{aligned} \rho_0(\theta)&=p(\theta)/\pi_0 \quad \text{for } \theta\in\Theta_0 \\ \rho_1(\theta)&=p(\theta)/\pi_1 \quad \text{for } \theta\in\Theta_1 \end{aligned} \tag{2.7}$$

식(2.6)을 살펴보면 베이즈요인은 각 Θ_i내에서의 평균적인 가능도(likelihood)의 비로 해석될 수 있는데, 이 점이 바로 베이즈요인이 가설검정에서 증거의 측도로써 기능하는데 구조적인 문제를 발생시킬 수 있다. 예를 들어, 어떤 사람의 나이가 40대라는 주장에 대한 증거의 측도가 그 사람의 나이가 40세 이상이라는 주장에 대한 증거의 측도보다 더 큰 값을 가져서는 곤란할

것이다. 하지만 만일 각 주장의 영역에 속하는 각 나이에 대한 어떤 가능도들의 평균을 증거의 측도로 선택한다면, 그 사람이 50세 이상에서의 가능도가 상당히 낮은 경우에는 40대라는 주장에 대한 증거의 측도가 40세 이상이라는 주장에 대한 증거의 측도보다 더 큰 값을 갖게될 위험이 있다는 것이다. 사실 베이즈요인의 이러한 문제점은 앞서 단측가설과 양측가설의 문제에서 비롯되는 고전적 P-값의 비일관성의 문제보다 더욱 심각한 구조적 결함으로 보여진다.

한편 최근에 Hwang(2001)은 주어진 신뢰수준을 갖는 신뢰영역(credible region)을 이용하여 유의수준의 개념을 갖는 베이즈 가설검정방법을 제안하였으며, Hwang and Oh(2002)는 이에 기반한 베이즈 P-값을 제안한 바 있다. 먼저 이들의 연구결과를 간략하게 요약하여 소개한다.

식(2.5)에서 주어진 가설들을 검정하기 위하여 모수 θ에 대한 무정보적 사전분포를 가정하고, 주어진 관측치 x에 대하여 $C_{1-\alpha}(x)$ 를 θ에 대한 하나의 $100(1-\alpha)\%$ 신뢰영역이라고 하자. 이때 Hwang(2001)이 제안한 유의수준 α의 베이즈 가설검정 방법은 다음과 같다.

$$C_{1-\alpha}(x) \subset \Theta_1 \text{ 일 때만 } H_0\text{를 기각한다.} \tag{2.8}$$

신뢰영역 $C_{1-\alpha}(x)$ 의 형태를 선택하는데 Hwang(2001)이 제안한 원칙들은 다음과 같다.

첫째, 단측검정에서는 대립가설에서 지정하는 모수의 영역과 같은 방향의 단측 신뢰영역을 사용한다. 즉, 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.

이와 같이 Hwang(2001)에 의하여 제안된 베이즈 가설검정방법을 바탕으로, Hwang and Oh (2002)는 주어진 관측치 x에 대하여 베이즈 가설검정 방법이 귀무가설을 기각하는 최소의 유의수준 α를 관측치 x에 대한 베이즈 P-값으로 정의하였다. 이들은 여러 가지 예를 통하여 이와 같이 정의된 베이즈 P-값이 기존의 고전적 P-값이나 또는 베이즈요인에 비하여 검정방법으로써 많은 강점들을 갖고 있음을 보여주었다. 그러나 일관성의 원리라는 측면에서 살펴보면, [예2-1]과 같은 경우에서는 이 방법들 역시 고전적 가설검정방법과 마찬가지로 일관성의 원리를 충족하지 못하고 있음을 알 수 있다.

3. LR형 베이즈 P-값의 제안 및 일관성의 검토

편의상 앞 절의 식(2.5) 이후의 모든 표기들은 이 절에서도 그대로 사용하기로 한다.

먼저, 신뢰영역 $C_{1-\alpha}(x)$ 의 형태를 선택함에 있어서 Hwang(2001)이 제안했던 두 가지 원칙들 가운데 두 번째 원칙을 제외하고, 항상 Θ_1의 원소들을 Θ_0의 원소들에 우선하여 신뢰영역 $C_{1-\alpha}(x)$ 에 포함시키도록 한다. 이렇게 되면 식(2.8)에서 주어진 H_0의 기각조건은 다음과 같이 표현될 수 있음을 쉽게 알 수 있다.

$$P[\theta \in \Theta_1 \mid x] \geq 1-\alpha \text{ 일 때만 } H_0\text{를 기각한다.} \tag{3.1}$$

따라서, $\alpha \geq 1 - P[\theta \in \Theta_1 \mid x] = P[\theta \in \Theta_0 \mid x]$일 때만 H_0를 기각하게 되어 H_0를 기각하는 최소의 유의수준 α가 $P[\theta \in \Theta_0 \mid x]$이므로, Hwang and Oh(2002)의 정의에 따라 이 경우의 베이즈 P-값은 $P[\theta \in \Theta_0 \mid x]$, 즉, 귀무가설에 대한 사후확률임을 알 수 있다.

다음으로, 사전분포로서 Hwang(2001)과 Hwang and Oh(2002)에서 가정했던 무정보적사전분포의 가정 대신에 다음과 같은 방법으로 사전분포를 가정하도록 한다. 우선, 주어진 관측치 x에 대하여 Θ_0와 Θ_1에서의 MLE들이 존재한다고 가정하고, 이들을 각각 $\widehat{\theta}_0$과 $\widehat{\theta}_1$으로 표기

하자. 이때 우리가 가정할 사전분포는 $\pi_0 = \pi_1 = 1/2$과 $\rho_0(\hat{\theta}_0) = \rho_1(\hat{\theta}_1) = 1$으로써 주어진다. 즉, 귀무가설과 대립가설에 대하여 각각 $1/2$씩의 공평한 사전확률(impartial prior probability)을 부여하고, 각 가설내부의 사전분포로는 각 영역의 MLE 한 점에 집중된 포인트 매스(point mass)들을 가정한다. 사전분포가 주어진 자료에 의존한다는 점에서 경험적 사전분포(empirical prior distribution)의 일종으로 이해될 수 있다.

이와 같은 가정들 아래 얻게되는 베이즈 P-값을 LR형 베이즈 P-값(Type LR Bayesian p-value)이라고 하고 P_{LR}이라고 표기하자. 즉, 앞에서 언급한 바와 같이 P_{LR}은 주어진 가정 아래 귀무가설에 대한 사후확률이므로, 다음의 식과 같이 주어지게 된다.

$$\begin{aligned} P_{LR} &= P[\theta \in \Theta_0 \mid x] \\ &= \frac{\pi_0 \int_{\theta \in \Theta_0} l(\theta \mid x)\rho_0(\theta)\, d\theta}{\pi_0 \int_{\theta \in \Theta_0} l(\theta \mid x)\rho_0(\theta)\, d\theta + \pi_1 \int_{\theta \in \Theta_1} l(\theta \mid x)\rho_1(\theta)\, d\theta} \\ &= \frac{l(\hat{\theta}_0 \mid x)}{l(\hat{\theta}_0 \mid x) + l(\hat{\theta}_1 \mid x)} = \frac{1}{1+R} \end{aligned} \tag{3.2}$$

여기에서 $R = l(\hat{\theta}_1 \mid x) / l(\hat{\theta}_0 \mid x)$은 일반적으로 가능도 비(likelihood ratio)로 알려져 있는 통계적 양이다. 또한 R을 다음의 식(3.3)으로 표현하게 되면 이는 일반적으로 Θ_0에서나 Θ_1에서의 MLE들이 존재하지 않는 모든 경우를 포괄할 수 있게 될 것이다.

$$R = \frac{\sup_{\theta \in \Theta_1} l(\theta \mid x)}{\sup_{\theta \in \Theta_0} l(\theta \mid x)} \tag{3.3}$$

P_{LR}의 일관성에 관하여 검토해 보자. 식(3.2)에서 P_{LR}이 가능도 비 R의 감소함수이고, 식(3.3)에서 Θ_0가 넓어질 때 R의 값은 커지지 않으며, 반면에 Θ_1이 넓어질 때 R의 값은 작아지지 않는다는 사실로부터 P_{LR}은 일관성 있는 증거의 척도임을 알 수 있다.

4. 적용 예

편의상 고전적 P-값은 P_C로, Hwang and Oh(2002)의 베이즈 P-값은 P_{HO}로 표기한다.

[예 4-1] 표본 $x_1, \cdots, x_n$은 정규분포 $N(\mu, 1)$로부터의 크기 n인 확률표본이고, 검정하고자 하는 가설들은 다음과 같다고 하자.

$$H_0: \mu \le 0 \quad v.s. \quad H_1: \mu > 0 \tag{4.1}$$

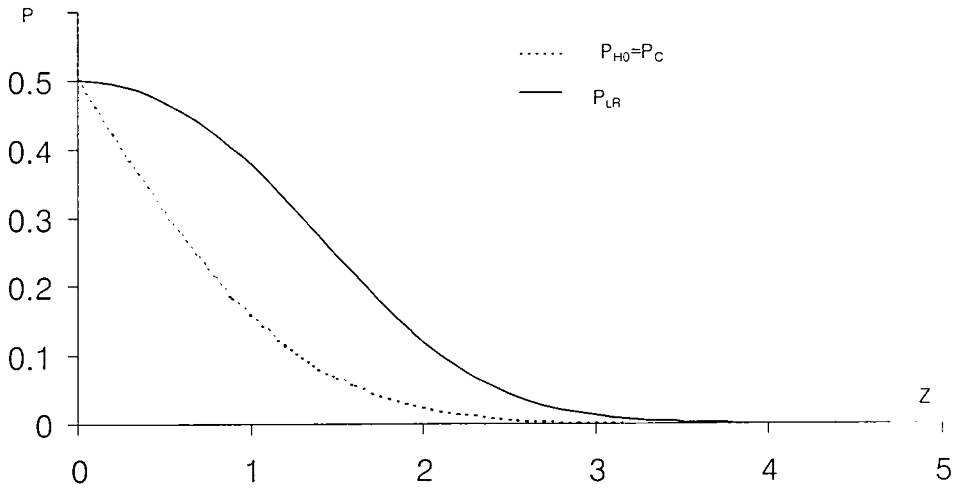
표본으로부터 계산된 $z = \frac{\bar{x} - 0}{1/\sqrt{n}}$의 관측값이 $z > 0$이라고 하면, P_{LR}은 다음과 같다.

$$P_{LR} = \frac{1}{1 + \exp(z^2/2)} \tag{4.2}$$

또한 이 경우에 P_{HO}는 가능도비 검정에 의한 P_C와 일치하는 값을 갖게 되며, Φ를 표준정규분포의 분포함수라고 할 때 다음 식으로 표현된다.

$$P_{HO} = P_C = 1 - \Phi(z) \tag{4.3}$$

[그림 1]에는 몇 가지 특별한 z값들에 대하여 P_{LR}과 $P_{HO} = P_C$의 값들을 그림으로 비교한 결과가 주어져 있다.



[그림 1] [예 4-1]에서 P_{LR}과 $P_{HO}=P_C$의 추이 비교

[그림 1]을 살펴보면 P_{LR}의 값이 $P_{HO}=P_C$의 값보다 전반적으로 큰 값으로 나타나고 있다. 즉, 식(4.1)과 같은 단측가설검정의 경우에 있어서는 P_{LR}은 P_{HO}나 P_C에 비하여 귀무가설의 기각에 상당히 신중한 경향이 있음을 보여주고 있으므로, P_{LR}을 사용할 때는 이 점에 유의해야 할 것으로 보인다.

다음의 [예4-2]는 Hwang and Oh(2002)에 의하여 제기되었던 예를 약간 변형해본 것이다.

[예4-2] $X \sim N(\mu, 1)$ 에 대하여 다음과 같은 단순가설들을 생각해 보자.

$$H_0 : \mu = 0 \quad v.s. \quad H_1 : \mu = \mu_1 \tag{4.4}$$

여기서, $\mu_1 > 0$은 정해진 양수이다. 이제 X의 관측값이 $x = 3$라고 하자.

먼저 고전적 P-값은 $\mu_1 > 0$의 값에 관계없이 다음과 같다.

$$P_C = 1 - \Phi(3) = 0.0013 \tag{4.5}$$

또한 Hwang and Oh(2002)에 의한 베이즈 P-값은 다음과 같다.

$$P_{HO} = \begin{cases} \dfrac{1}{1+\exp(3\mu_1 - \mu_1^2/2)}, & 0 < \mu_1 \le 6 \\ 1, & 6 < \mu_1 \end{cases} \tag{4.6}$$

다음으로, LR형 베이즈 P-값은 식(3.2)에 의하여 계산해 보면 다음과 같이 주어진다.

$$P_{LR} = \frac{1}{1+\exp(3\mu_1 - \mu_1^2/2)} \tag{4.7}$$

다음의 [표 2]는 식(4.5) ~ (4.7)을 이용하여 몇 가지 선택된 μ_1의 값들에 대한 각 P-값들을 비교하여 도표화 한 것이다.

고전적 P-값인 P_C의 값이 P_{HO}나 P_{LR}과는 달리 $\mu_1 > 0$의 값에 관계없이 일정한 값을 갖는다는 사실은 다음과 같은 직관적인 관점에서 비판받을 만 하다.

[표 2] [예 4-2]에서 P_C, P_{HO} 및 P_{LR}의 비교

μ_1	P_C	P_{HO}	P_{LR}
0.01	0.0013	0.4925	0.4925
0.854	0.0013	0.1000	0.1000
1.236	0.0013	0.0500	0.0500
3.00	0.0013	0.0110	0.0110
4.764	0.0013	0.0500	0.0500
5.146	0.0013	0.1000	0.1000
6.00	0.0013	0.5000	0.5000
8.00	0.0013	1.0000	0.9997

먼저, $\mu_1=0.01$과 같이 0에 매우 가까운 경우를 생각해 보자. 고전적 P-값은 $H_0: \mu=0$이 참일 때 x의 값이 현재 관측값인 3 이상일 확률로 계산되며, 그 값이 0.0013으로 매우 작다는 사실을 근거로 하여 귀무가설을 기각하고 매우 강력한 증거에 의해서 대립가설 $H_1: \mu=0.01$을 채택하게 한다. 그러나 H_1이 참인 경우에도 x의 값이 3 이상일 확률은 0.0014로서 H_0가 참인 경우와 별 차이가 없으므로, 관측값 $x=3$이 H_0에 비하여 H_1이 참이라는 매우 강력한 증거라는 고전적 P-값의 결론은 직관적으로 받아들이기 어렵다. 이 경우에는 0.4925의 값으로써 귀무가설이 거짓인 증거가 별로 없다는 결론을 얻고 있는 P_{HO}나 P_{LR}가 직관에 더 부합하고 있음을 알 수 있는 것이다.

이번에는 $\mu_1=8.0$과 같이 $H_0: \mu=0$에서보다 관측값 $x=3$에서 더 멀리 있는 경우를 생각해 보자. 이 경우에도 고전적 P-값인 P_C는 역시 0.0013으로 매우 강력한 증거에 의하여 H_0를 기각하고 H_1을 채택하게 된다. 그러나 상식적으로 $x=3$의 관측값이 $H_0: \mu=0$에 비하여 $H_1: \mu=8.0$이 참이라는 강력한 증거로는 보여지지 않으므로, 고전적 가설검정의 결과는 받아들이기 어렵다. 반면에 이 경우에도 P_{LR}은 0.9997의 매우 큰 값으로 귀무가설을 채택함으로써 직관에 잘 부합하고 있음을 알 수 있는 것이다.

References

[4] Hwang, H.T. (2001). A Bayesian hypothesis testing procedure possessing the concept of significance level, *The Korean Communications in Statistics*, Vol. 8, No. 3, 787-795

[5] Hwang, H.T. and Oh, H.J. (2002). A study on Bayesian p-values, *The Korean Communications in Statistics*, Vol. 9, No. 3, 725-732

[8] Lavine, M. and Schervish, M.J. (1999). Bayes factors: What they are and what they are not, *The American Statistician*, 53, No. 2, 119-122

[11] Schervish, M.J. (1995). *Theory of Statistics*, New York: Springer-Verlag.

[12] Schervish, M.J. (1996). P-values: What they are and what they are not, *The American Statistician*, 50, No. 3, 203-206