2002 Proceedings of the Autumn Conference, Korean Statistical Society

왜 베이지안 인가?

이 군희
(서강대학교, 경영학과)

요 약

본 발표에서는 베이지안이 생각하는 확률의 개념을 상호교환성(exchangeability)의 가정아래 어떻게 확장되어 해석되는지를 소개하고, 빈도학자들의 접근방법과 비교함으로서 베이지안에서 생각하는 확률이 어떠한 특징을 가지고 있는지를 설명하고자 하였다. 또한 Efron에 의하여 지적된 베이지안의 네 가지 문제점에 대하여 논의하고 특별히 과학적 객관성(scientific objectivism)의 한계점과 이러한 한계점을 베이지안에서 어떻게 해결하고 있는지에 대하여 논의하였다. 일반적으로 과학적 객관성에 대한 한계점은 빈도학자들의 방법론에서도 존재하게 된다. 즉, 연구자가 가설을 설정하고 이에 맞는 실험설계를 하고 유의수준을 설정하고 p값을 이용하여 의사결정을 내리는 모든 단계에서 연구자의 주관성이 들어갈 수밖에 없게 된다는 것이다. 베이지안 방법론에서는 이러한 비객관적인 체계를 인정하고 파악하여 사전확률(prior)에 포함시킴으로서 이를 객관적인 자료인 가능도함수(likelihood function)와 혼합하여 추론이나 의사결정을 진행하게 된다. 마지막으로 베이지안 학자들의 최근 객관적인 사전확률에 대한 다양한 형태의 연구를 소개하는 것으로 발표를 마무리하고자 한다.

1. 서론

우리는 불확실한 현상이나 확신을 가지고 예측할 수 없는 결과가 나타나는 경우를 랜덤(random)이라고 한다. 이러한 랜덤을 구체화하고 정량화하기 위하여 확률이라는 개념이 소개되었다. 빈도학자(frequentist)들은, 동전을 여러 번 던지는 경우, 앞면이 나타나는 수와 뒷면이 나타나는 수가 정확하지는 않지만 비슷할 것이라는 많은 사람들이 공감하는 사실을 기초로 확률을 해석하여왔다. 하지만 현상을 이해하는 측면에서 이러한 논리는 매우 제한적으로 응용되어 왔다.

예를 들어, 동전을 20번 던져서 모두 앞면이 나타났다고 가정하여보자. 이 경우 던져진 동전이 어떻게 만들어 졌는지, 또는 과거에 이 동전을 던져서 어떠한 결과가 나타났는지에 대한 사전정보가 없다면, 이 동전은 (1) 공정한 동전이 아니거나 (2) 공정한 동전이지만 우연히 20번의 앞면이 나타났을 것이라고 판단할 수 있을 것이다. 이 경우, 일부 통계학자들은 (2)번의 가설에 확신을 가지고 현상을 설명하려고 하겠지만 일반인들은 동전이 잘못된 것이 아닌지에 대한 의심을 하는 (1)번 가설에 확신을 가질 것이다. 하지만 확실한 것은 21번째 던져진 동전의 결과가 어떻게 나타날 지에는 불확실성이 존재한다는 것이다. 만일 20번 동전을 던진 경험적인 증거가

있고, 동전이 공정한지에 대하여 정확하게 알지 못한다면 de Finetti 정리를 이용하여 다음 번 동전이 나타날 불확실성을 표현할 수 있다. de Finetti 정리를 알기 쉽게 설명하면 다음과 같다.

상호교환적[1]인 이항사건 (exchangeable binary event)을 나타내는 n 회의 베르누이 시행에서, 만일 확률변수 x_j를 j번째 시행에 대한 이항 확률변수라 한다면, 확률변수열 $\{x_1, x_2, \cdots, x_n\}$ 의 예측확률(predictive probability)은 다음과 같이 계산된다.

$$\begin{aligned}\Pr(x_1, x_2, ,\cdots, x_n) &= \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\, dF(\theta) \\ &= \int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\, dF(\theta)\end{aligned}$$

여기서 $F(\theta)$ 는 0에서 1구간의 θ에 대한 누적분포함수가 된다. 이러한 형태의 예측확률은 이항분포와 $F(\theta)$ 의 혼합형태이며, θ를 고정시키고 이항분포를 이용하여 확률을 계산한 뒤, 이 값을 모든 가능한 θ에 대하여 가중평균을 계산하여 얻은 수치로서의 의미를 가지게 된다.

만일 $F(\theta)$ 의 형태를 Bayes 또는 Laplace가 선호하는 균등분포로 가정한다면[2], $dF(\theta)$는 $d\theta$가 되며, 모든 r에 대하여 다음과 같은 예측확률이 계산된다.

$$\Pr(\sum x_i = r) = \int_0^1 \binom{n}{r} \theta^r (1-\theta)^{n-r}\, d\theta = \frac{1}{n+1}\ .$$

다시 말하면, 9번 동전을 던지는 경우 앞면이 한번도 나타나지 않을 확률은 9번 앞면이 나타날 확률과 동일하게 10%라는 것이다.

de Finetti 정리의 다른 특징이라면 불확실한 모수와 관련된 주관적인 주장에 대한 설명이 가능하다는 것이다. 예를 들어, 주사위 두 개를 던져서 눈의 합이 7이 되는 경우가 자주 나타나기 때문에, 연구자가 주사위 눈의 합이 7이 나오는 확률을 1/4 에서 3/4 라고 확신한다면, de Finetti 정리를 이용하여 5번 연속 7이 나타날 예측확률을 다음과 같이 계산할 수 있다.

$$\int_{1/4}^{3/4} 2\,\theta^5\, d\theta = \frac{91}{1536} = 0.0592448$$

2. 베이지안에 대한 비판

Efron(1984)은 베이지안이 많은 과학자로부터 호응을 받지 못하고 있는 이유를 4가지로 요약하여 설명하였다. 우선 Fisher에 의한 접근 방법은 쉬우며, 명확한 분석 결과를 제공하여 주

---

1) 상호교환성(exchangeability)은 확률변수의 무한열(infinite sequence) $\theta_1, \theta_2, \cdots, \theta_q, \cdots$ 에 대한 결합분포가 확률변수의 무한열에 대한 subscript의 어떠한 permutation에 대하여서도 변하지 않는 것을 의미한다.

2) 이것을 Laplace의 불충분한 이유의 원리라고 한다.

는 반면에, 베이지안은 복잡하고 쉽게 접근할 수 없다는 것이 첫 번째 이유이었으며, 두 번째로 통계에서 적용되는 실험 계획에 관련된 측정 -> 요약 -> 비교 -> 추론에서 베이지안은 너무 추론에만 집중되어 연구가 진행되고 있다는 것이다. 세 번째로 연구 상황에 따른 개별적인 방법론을 Neyman-Pearson-Wald 연구에서는 제공하지만[3] (예를 들면, UMVUE) 베이지안은 이에 대한 융통성이 없다는 것이다. 마지막으로 과학적 객관성에 대한 측면에서 빈도학자들이 많은 기여를 하고 있으며, 상대적으로 베이지안 학자들이 매우 불리한 위치에 있다는 것이다.

이러한 Efron의 비판에 대하여 Lindley(1986)는 베이지안 방법론 자체가 Fisherian 방법론보다 훨씬 체계화되고 명확한 방법론을 제시하고 있다고 반론하고 있다. 만일 우리가 H라는 사실을 알고 X라는 사실을 모르는 경우, $\Pr(X|H)$을 베이즈 정리에 의하여 자동적으로 계산할 수 있다는 간단한 원리로부터 추론이 시작된다고 설명하고 있다. 하지만 이러한 간단한 원리를 현실에 어떻게 적용할 것인가의 부분이 어려운 부분이며, 샘플링 이론가들은 이러한 문제를 너무 단순하고 기계적으로 생각하고 있다고 비판하고 있다(Press, 1989). 또한 베이지안이 너무 추론 중심의 이론으로 발전되고 있다는 측면에 대해서는 이러한 포인트는 베이지안 의사결정론(decision theory)에서 판단해야하는 문제라고 Lindley는 일축하고 있다.

Efron(1986)이 지적하였듯이 가설 검정 과정에서 베이지안과 빈도학자 사이에는 서로 다른 큰 특징은 귀무가설(또는 대립가설)을 지지하거나 혹은 반대하는 증거를 서로 다른 관점에서 수량화하고 있다는 것이다. 예를 들어, 평균이 0이라는 귀무가설과 평균이 2라는 귀무가설이 있고, 자료의 평균이 매우 큰 값이 나타났다면, 빈도학자들은 귀무가설을 기각하는 정도를 높게 계산하여 주지만 이러한 수치 자체는 절대적인 의미를 갖는 것이 아니라 상대적인 의미의 척도가 되는 것이다. 하지만, 베이지안에서는 이러한 문제를 사후오즈 (posteriori odds)로서 설명한다. 즉, '대립가설에 대한 사후오즈가 7이다' 라는 식으로 표현하는 절대적 의미의 척도를 사용하며, 이러한 오즈의 개념은 다른 가설검정에서 나타나는 7이라는 의미와 동일하게 해석된다. 베이지안 가설검정에서 많이 사용되는 오즈비에 대한 로그값은 결국 사전확률에 대한 오즈비의 로그값에 로그 가능도함수(likelihood function)를 더한 값이 되며, 가능도 함수는 자료가 어떻게 구성되어 있는지를 나타내는 값이 된다.

이 논문에서는 베이지안에서 가장 많은 공격대상이 되고 있는 과학적 객관성에 관련된 베이지안 학자들의 연구에 대하여 설명하고 마무리를 짓고자 한다.

3. 객관성에 대한 연구

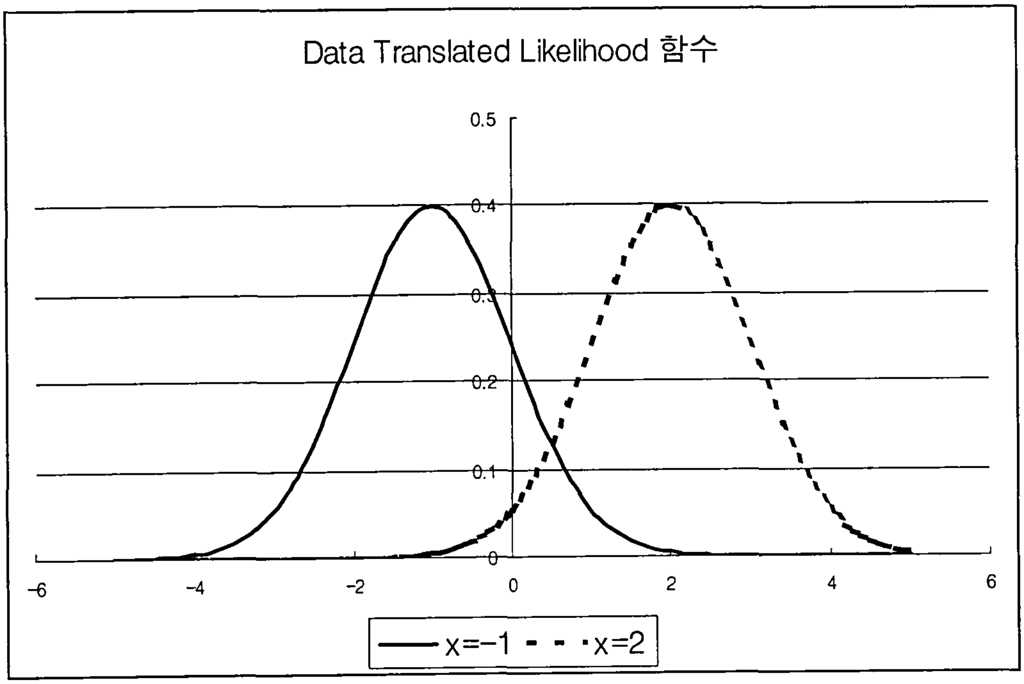
베이지안 방법의 추론이나 의사결정문제는 문제의 특성에 따라 주관적 분석을 요구하는 경우도 있고 객관적 분석을 요구하는 경우도 있다. 주관적 분석이 베이지안 추론을 통하여 가능하다는 이유만을 가지고 베이지안 추론 자체가 비과학적인고 주관적인 분석 방법이라고 취급되고 비판받는 것은 합당하지 못하다. 주관적 추론이나 의사결정 문제를 해결하고자 할 경우에는 주관적인 견해가 포함되어 있는 사전확률(prior)을 사용하여 추론을 진행하면 되고, 객관적인 추론이나 의사결정문제를 해결하고자 하는 경우에는 객관적인 사전확률을 사용하여 추론을 진

3) 예를 들어 nuisance parameter를 고려한 어떤 관심있는 parameter에 대한 UMVUE를 찾고자 하는 경우 관심있는 parameter가 어떠한 함수 형태인가에 따라 UMVUE를 새롭게 계산하여야 한다. 이와 같은 현상을 베이지안에서는 incoherence 문제라고 한다.

행하면 된다. 그렇다면 객관적인 사전확률이란 무엇을 의미하는 것일까?

Jeffreys(1961)는 관측된 자료가 변하는 경우, 가능도함수의 모양은 변하지 않으면서 위치만 바뀔 경우를 data translated 가능도 함수라고 정의하고 이에 따라 만들어진 모수에 균등분포를 할당하면 객관적인 사전확률이라고 설명하였다 (<그림 1> 참조).

이러한 Jeffreys의 아이디어는 모수가 한 개일 경우에는 완벽하게 설명이 되지만 모수가 두 개 이상으로 확장될 경우에는 상당한 문제점이 있는 것으로 나타났다. 이러한 문제점을 보완하기 위하여 Bernardo와 Berger (1984)는 모수가 두 개 이상일 경우, 사후확률함수와 가장 멀리 떨어져 있는 사전확률함수의 형태가 객관적 사전확률함수라고 정의하고, 두 함수간의 거리를 측정하는 Kullback-Liebler Divergence를 최대화시키는 사전확률함수를 찾는 방법을 제안하여 'reference prior'라고 부르게 되었다. 또한 이러한 과정을 통하여 얻은 사전확률함수는 모수가 한 개일 경우, Jeffreys가 완벽하게 정의된 객관적 사전확률과 일치한다는 사실도 확인할 수 있었다.



<그림 1> Jeffreys의 객관적 사전확률 기준

또한 빈도학자들이 생각하는 95% 신뢰구간에 대한 성질을 만족하는 사전확률을 찾으려는 노력으로 matching prior에 대한 연구가 활발하게 진행되고 있다.

참 고 문 헌

Bernardo J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, New York, John Wiley and Sons.

Box G. E. P. and Tiao G. C. (1973), *Bayesian Inference in Statistical Analysis*, Massachusetts, Addison-Wesley Publishing.

DeFinetti, B. (1974) *Theory of Probability* (Vol. 1) London, John Wiley

DeFinetti, B. (1975) *Theory of Probability* (Vol. 2) London, John Wiley

Efron, B. (1986), "Why Isn't Everyone a Bayesian?" (with discussion), *The American Statistician*, Vol. 40, No. 1, 1-11.

Jeffreys, H (1961), *Theory of Probability*, Oxford, Claredon Press.

Lindley, D. V. (1983), "Theory and Practice of Bayesian Statistics", *The Statistician,* 32, 1-11.

Press, S. J. (1989), *Bayesian Statistics*, New York, John Wiley and Sons.

Savage, L. J. (1972), *The Foundations of Statistics*, New York: Dover Publications

Tanner M. A. (1996), *Tools for Statistical Inference*, New York, Springer