응용통계연구 제 7권 2호
1994년 9월 pp. 263-268

베이즈의 균일분포에 관한 소고

허명회[1)]

요 약

베이즈(Thomas Bayes)는 역사적인 1764년 논문에서 공 W의 위치 θ 에 대한 추론 문제를 생각한 바 있다. 이 때 그는 θ 에 대한 사전(prior)분포로서 균일(uniform)분포를 가정하였는데, 본 소고에서는 이 사전분포가 단순한 주관적 확률이 아닌 논리적 확률로 간주될 수 있음을 보일 것이다.

1. 베이즈의 1764년 논문

베이즈 통계학의 원조인 베이즈(Thomas Bayes)는 1764년 논문에서 다음의 문제를 풀이와 함께 제시하였다 (이에 대한 상세한 논의는 Stigler(1986) 또는 이용구(1991)를 참조할 것).

> 공 W가 사각평면에서 정지한 위치의 수평축 좌표 θ 에 대한 추론문제를 생각하여 보자 ($0 \le \theta \le 1$). 이 때 주어진 정보는 공 O를 독립적으로 n 번 굴렸을 때 공 O가 공 W의 왼쪽에 온 회수 r 과 오른 쪽에 온 회수 $n-r$ 이다. <그림 1> 참조.

후세에 당구대로 발전된 이 문제에서 베이즈는 θ 에 대한 사전(prior)분포로서 균일(uniform) 분포를 가정하고 관측된 r에서 얻어진 우도(likelihood)

$$\theta^r (1-\theta)^{n-r}$$

를 활용함으로써, θ 에 대한 사후(posterior)분포로 베타분포 Beta($r+1$, $n-r+1$)을 유도할 수 있었다.

본 소고의 주안점은 이 문제에서 가정된 θ 에 대한 사전분포의 확률적 성격이 단순한 주관적 확률에 그치는 것이 아니라 수리적으로 객관화될 수 있는 논리적 확률이라는 것을 보이는데 있다. 확률의 의미에 대한 참고문헌으로 Barnett(1981), Fine(1973)과 함께 허명회(1994)를 참조하라.

2. 균일분포의 유도

이제 <그림 1>에서 공 W가 정지한 위치 θ 가 균일분포를 따른다는 것을 수리적으로 보이도록 하겠다. 무한히 펼쳐진 가상적인 당구대에서 X 만큼 수평으로 이동시킬 수 있는 힘이 공에 가해진다고 하자. 이 때 X 는 개인에 따라 다를 수 있는 어떤 확률밀도함수 $f_X(x)$를 따른다고

---

1) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과 교수.

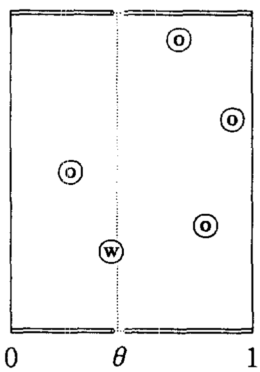
하자 (이 때 일반성을 잃지 않고 $x \ge 0$ 을 가정). <그림 1>의 당구대에는 수평축 1의 간격으로 양쪽에 충돌면이 있으므로, 경계면에서의 완전탄성충돌을 가정하는 경우, θ 와 X의 관계는 다음과 같게 된다.

X 를 2로 나누었을 때의 나머지를 "X mod 2"로 표기할 때

$$\theta = \min\{X \bmod 2,\ 2-(X \bmod 2)\},$$

예컨대 $X = 2.2$이면 $\theta = 0.2$이고 $X = 5.4$이면 $\theta = 0.6$이다.

<그림 1> 베이즈의 당구대



이 때 확률변수 X 가 상당히 큰 값을 갖는 경향이 있다고 하자. 다시 말하여 양쪽 경계면에 여러 번 충돌한 후 정지할 수 있게끔 공을 세게 치는 것을 전제로 하자. 이것을 수리적으로

$$f_X(x;N) = N^{-1}h(x/N), \quad 0 \le x < 2N$$

으로 표현하기로 한다. 여기서 N은 상당히 큰 자연수이고 $h(z)$는 구간 $[0, 2)$가 바탕(support)인 임의의 절대연속 확률밀도함수라고 하자. 다시 말하자면 확률변수 X가 공을 당구대의 오른쪽 경계면에 N 번까지 (왼쪽 경계면에 N-1 번까지) 충돌할 수 있을 만큼 확률적으로 크다고 하자는 것이다. 그러면

$$\begin{aligned} P_N(\theta \le \theta_0) &= \sum_{k=0}^{N} P_N(|X-2k|) \le \theta_0) \\ &= \sum_{k=0}^{N} \int_{2k-\theta_0}^{2k+\theta_0} N^{-1}h(x/N)\,dx \end{aligned}$$

가 된다. 함수 $h(z)$가 2번까지 미분가능하고 2계 미분 $h''(z)$의 절대값이 유계(bounded)되어 있다고 하면, N을 무한히 크게함에 따라

$$
\begin{aligned}
P(\theta \le \theta_0) &\equiv \lim_{N\to\infty} P_N(\theta \le \theta_0) \\
&= \lim_{N\to\infty} \sum_{k=0}^{N} \int_{2k-\theta_0}^{2k+\theta_0} N^{-1} h(x/N)\, dx \\
&= \lim_{N\to\infty} \sum_{k=0}^{N} \int_{(2k-\theta_0)/N}^{(2k+\theta_0)/N} h(z)\, dz \\
&= \lim_{N\to\infty} \sum_{k=0}^{N} \{2\theta_0/N \cdot h(2k/N) + O(N^{-3})\} \\
&= \theta_0 \int_0^2 h(z)\, dz \\
&= \theta_0, \qquad\qquad 0 \le \theta_0 \le 1
\end{aligned}
$$

가 된다. 이 결과는 개인에 따라 다를 수 있는 미지의 밀도함수 $h(\cdot)$에 관련없으므로, 그것이 어느 정도 매끄러운 것이기만 하면 (공을 여러번 양쪽 경계면에 충동시킬 수 있을 만큼 세게 치는 경우) 공의 최종 위치 θ 가 근사적으로 균일분포를 따르게 됨을 말하는 것이다.

그러므로 베이즈의 1764년 논문에서 가정한 θ 에 대한 사전(prior)분포로서의 균일분포가 개인적 또는 주관적 확률이라기 보다는 (어떤 견고한 기반을 갖는 전제로부터) 수리적으로 유도가능한 논리적 확률(logical probability)이라고 할 수 있다.

3. 실험 및 결과

실제의 당구대에서 공의 정지위치 θ 를 측정하는 실험을 하여 보았다. 실험의 목적은 관측자료가 과연 균일분포를 잘 따르는가를 확인하고자 하는 것이다. 당구공을 통상 4회-8회 충돌하도록 가능한한 세게 쳤으며 반복수는 400회였다. <표 1>은 관측자료의 기술통계치를 정리한 것이고 <그림 2>는 경험적 분포함수를 플롯한 것이다.

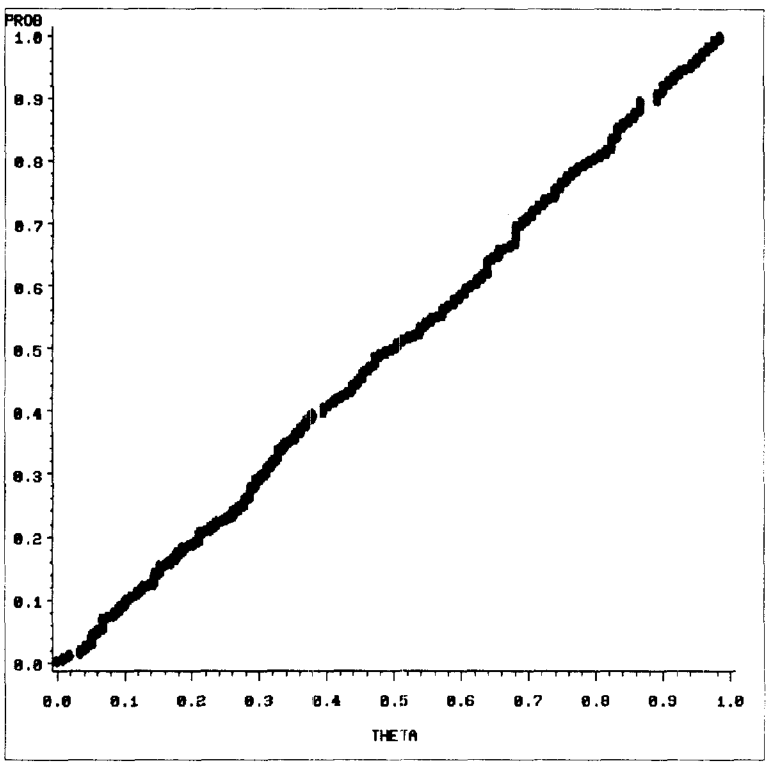
균일분포 가설을 검토하기 위하여 콜모고로프(Kolmogorov) 통계량

$$D = \max_{0\le\theta\le 1} |F(\theta) - \widehat{F}_n(\theta)|$$

를 계산해 보았다. 여기서 $F(\theta)$는 균일분포의 이론적 분포함수이고 $\widehat{F}_n(\theta)$는 표본에서 얻은 경험적 분포함수이다 (n = 400). D = 0.0305를 얻을 수 있었는데 이 값은 α = 20% 대표본 임계값인 $1.07/\sqrt{n}$ = 0.0535 와 비교하여 볼 때 통계적으로 유의하지 않다.

<표 1> 당구공의 정지위치를 관측한 결과와 균일분포의 대비

	자료 (크기 400)	균일분포 (이론)
평균	0.5014	0.5
분산	0.079	0.083
왜도	-0.0211	0
첨도	-1.2071	-1.2
최소값	0.0000	0
25% 분위수	0.2773	0.25
50% 분위수	0.5042	0.5
75% 분위수	0.7395	0.75
최대값	0.9832	1



<그림 2> 당구공의 정지위치를 관측한 표본자료의 경험적 분포함수

4. 맺음말

2절에서의 전제와 같지는 않지만 비슷한 전제하에서 균일분포를 수리적으로 유도한 선행연구들을 간략히 언급하고자 한다. Diaconis and Engel(1986)에 따르면 J. Kemperman이 1975년 미발표논문에서 임의의 어떤 분포함수 F를 갖는 확률변수 X 에 대하여 σX (*mod* 1)의 분포가 σ 가 무한히 커짐에 따라 균일분포로 수렴한다는 것을 보였다고 한다. 그리고 이와 유사하게 1966년 S.V. Nagaev와 A.B. Muhkin은 독립적인 확률변수 X_1, X_2,... 가 있을 때 부분합의 제수 1 나머지인

$$\sum_{j=1}^{n} X_j \pmod 1$$

이 균일분포로 수렴할 필요충분조건을 특성함수를 사용하여 표현하였다(Johnson and Kotz, 1970). 따라서 2절에 제시된 균일분포의 유도가 초유의 것이라고는 결코 할 수 없다. 그러나 2절의 그것이 미적분(calculus) 수준의 수학을 써서 얻어진 것이기 때문에 대한 학부의 저학년 수준 강의에서도 다루어 질 수 있을 것이다.

참 고 문 헌

[1] 이용구(1991). 역확률, 『통계학사 콜로키움』(허명회編) 4장, 자유아카데미, 서울.

[2] 허명회(1994). 14면 주사위의 확률, 『응용통계연구』, 제7권 1호. 113-119.

[3] Barnett, V.(1981). *Comparative Statistical Inference*, Second Edition, Wiley, New York.

[4] Bayes, T.(1764). An essay towards solving a problem in the doctrin of chances, *Philosophical Transactions of the Royal Society of London* **53**(1763), 370-418. Reprinted in *Studies in the History of Statistics and Probability, Volume 1* (edited by E.S. Pearson and Sir M. Kendall, 1970), 134-153.

[5] Diaconis, P. and Engel, E.(1986). Comment (to I.J. Good's "Some Statistical Applications of Poisson's work"), *Statistical Science* , Vol. 1, 171-174.

[6] Fine, T.L.(1973). *Theories of Probability: An Examination of Foundations.* Academic Press, New York.

[7] Johnson, N.L. and Kotz, S.(1970). *Continuous Univariate Distributions - 2.* Houghton Mifflin, Boston.

[8] Stigler, S.M.(1986). *The History of Statistics: The Measurement of Uncertainty before 1900,* Harvard University Press, Cambridge.

On Bayes' Uniform Prior

Myung-Hoe Huh[2)]

Abstract

Thomas Bayes assumed uniform prior for the location θ of a billiard ball W in his historic 1764 paper. In this study, following mathematical derivation of the uniform distribution from several assumptions that are plausible on the billiard table, it is argued that the probabilistic meaning of Bayes' uniform prior (especially in *Billiard Problem*) is not just subjective but logical.

2) Dept. of Statistics, Korea University. Anam-dong 5-1, Sungbuk-gu, Seoul 136-701, KOREA.