Journal for History of Mathematics http://dx.doi.org/10.14477/jhm.2024.37.1.001
Vol. 37 No. 1 (Feb. 2024), 1–17

A consideration of the real meanings of introducing Bayesian inference into school mathematics curriculum

베이즈 추론을 수학과 교육과정에 도입하는 것의 실제 의미에 대한 일고찰

PARK Sun-Yong 박선용

In this study, we identified the intellectual triggers for Bayesian inference and what key ideas contributed to its occurrence and discussed the practical implications of introducing Bayesian inference into the school mathematics curriculum by reflecting them. The results of the study show that the need for statistical inference about the parameter itself served as a trigger for the occurrence of Bayesian inference, and the most important idea for the occurrence of that inference was to regard the parameter itself as a probability variable rather than any fixed value. On the other hand, these research results suggest that the meaning of introducing Bayesian inference into the secondary mathematics curriculum is 'statistics education that expands the scope of uncertainty'.

Keywords: Bayesian inference, school mathematics curriculum; 베이즈 추론, 학교수학, 교육과정.

MSC: 01A50 ZDM: A34

1 서론

베이즈 추론은 정보의 추가에 따라 확률을 업데이트하는 도구라는 측면에서 인공지능 시대의 도래와 함께 학계 및 산업계 전반에서 각광받고 있다. 물론, 사회적 필요와 그 유용성에 근거하여 베이즈 추론은 우리나라 수학교육계에서도 주목하고 있다고 하겠다. 이와 관련해, 정치봉은 베이즈 추론의 학교수학에의 도입에 따른 우려와 함께 그 교육방향에 대한 제안을 다음과 같이 한 바 있다.

> (베이즈 추론은) 인공지능 개발 분야에서 광범위하게 열광하며 사용되고 있기에 필자로서 조심스럽게 느끼는 점이 있다. 학교수학에서 확률통계를 가르치

PARK Sun-Yong: Dept. of Math. Edu., Yeungnam Univ. E-mail: polya@yu.ac.kr
Received on Oct. 11, 2023, revised on Jan. 21, 2024, accepted on Jan. 22, 2024.

> 는 교육과정에서는 빈도, 콜모고로프 공리계, 확률변수와 분포, Law of large numbers, 중심극한정리 등을 중심으로 구성되어 있다. (중략) 학교수학 교육과정에서는 베이즈의 역확률 문제를 다루지 않는다. 조건부확률을 베이즈공식을 통해서 사후확률을 계산적으로 다룬다. 베이즈적인 추론 형식체계의 예는 배운다. 하지만 이런 특별한 예제를 통해 이론체계를 갖추고 일반화되고 다양한 모형으로 발전해가는 수학교육으로 나아가지는 못하고 있다. 베이즈 확률 추론을 통해서 수리모형과 함께 탐구형 수학학습이 학교 현장에서 이루어질 수 있기를 기대해본다 [9, pp. 21–27].

자세히 밝히지 않았지만 이 인용문에서 ‘조심스럽게 느끼는 점’은 아마도 ‘모든 학생이 인공지능 개발자가 될 필요가 없는데 학교수학에서 굳이 베이즈 추론을 다룰 필요가 있느냐?’라는 문제 제기와 먼저 관련될 것이다. 그런데 이 문제제기는 ‘베이즈 추론이 학교수학에서 다룰만한 가치가 있는 소양 또는 지식이다.’와 같은 수학교육계 전반의 합의가 이루어져야 해소될 수 있는 사항이라고 할 수 있다.

하지만 이러한 사회적 합의가 도출되기 위해서는 ‘베이즈 추론을 통해 가르칠 수 있는 소양이 무엇이냐?’에 대한 규명이 먼저 이루어질 필요가 있다. 그런데 이 물음은 베이즈 추론이 무엇이냐에 대한 것과 함께 그 추론에 대한 교육을 통해 가르칠 수 있는 것이 무엇인지를 묻는 것이라 할 수 있다. 베이즈 추론이 학교수학에 도입되는 것에 대한 정당화나 학계의 합의가 이루어지기 위해서라도 먼저 이루어져야 하는 연구는 ‘베이즈 추론의 본질’ 자체에 대한 교수학적 분석이라 할 것이다.

한편, 이 인용문에서 ‘조심스럽게 느끼는 점’ 중 다른 한 가지는 산업계의 요구에 부합해 ‘최신의 지식을 가장 효율적으로 신속하게 가르치는 교육’으로 변질될 수 있다는 우려 또는 그에 대한 경계일 것이라 생각한다. 역사적으로 수학교육 현대화 운동, 일명 ‘새수학 운동’의 부작용을 경험했던 수학교육계에서 이처럼 조심스러운 입장을 나타내는 것은 자연스러운 현상이다. 수학교육의 궁극적 목적을 과학기술의 발달이 아니라 학생의 수학적 역량 및 경험의 성장이라 여기는 발달적 교육관의 입장에서 볼 때, ‘산업계의 요구만 전적으로 반영하여 이루어지는 수학교육’은 분명 경계의 대상이라 하겠다.

이러한 맥락에서 정치봉은 ‘최신의 베이즈 이론을 조기에 도입하는 수학교육’이 아닌 ‘특별한 예제를 통해 이론체계를 갖추고 일반화되고 다양한 모형으로 발전해 가는 수학교육’을 제안했을 것이다. 이 연구에서도 이 부분에 주목하고자 한다. 최신의 완성된 베이즈 추론의 이론을 하향 초등화해서 학생에게 부과하여 가르치는 것이 아니라 학생이 베이즈 추론이 발생했던 과정을 경험하게 하려면 어떻게 해야 하느냐가 이 연구의 초점인 것이다.

사실, ‘베이즈 추론을 통해 가르칠 수 있는 소양’이 바로 ‘학생이 베이즈 추론이 발생했던 과정을 경험함으로써 얻을 수 있는 것’이라는 점에서 베이즈 추론을 학교수학에 도입

하기에 앞서 우선적으로 이루어져야 할 작업은 발생적 그리고 교육적 관점에서 '베이즈 추론의 본질'을 밝히는 것이라 하겠다. 이를 위해 이 연구에서는 다음과 같은 연구문제를 설정하였다.

1. "베이즈 추론이 발생하게 된 지적 계기와 그 추론의 발생에 기여한 핵심적 아이디어는 무엇인가?"
2. "베이즈 추론의 발생 계기와 그 추론에 담긴 아이디어를 반영해 베이즈 추론을 학교수학에 도입한다고 하면, 학교수학에 어떤 변화를 요구하는 것일까?"

이제, 베이즈 추론이 발생하게 된 지적인 계기부터 탐색하면서 베이즈 추론을 가르친다는 것의 실제 의미를 고찰하기 시작할 것인데, 이러한 일련의 작업은 베이즈 추론을 수학과 교육과정에 도입하려고 할 때 예상되는 난관의 실체가 무엇인지를 드러내는 것이기도 할 것이다.

2 베이즈 추론이 발생하게 된 지적 계기

베이즈가 새롭게 시도하고자 했던 추론은 어떤 점에서 새로운 것일까? 이 질문은 베이즈 추론이 출현하기 이전의 통계적 추론이 지닌 한계가 무엇이었는지를 묻는 것과 근본적으로 같다. 그리고 이 한계가 바로 베이즈 추론이 나타내게 된 일종의 지적인 계기와 깊은 관련이 있다고 할 것이다 [5, 6].

한편, 베이즈 이전에 불확정성을 계량화하는 데에 있어 가장 앞서 있었던 연구는 드 무아브르(Abraham De Moire, 1667∼1754)의 작업이다. 그러기에, 여기서 드 무아브르의 연구가 지닌 장점과 함께 그 한계가 무엇인지를 명확히 파악하는 것이 필요하다고 하겠다.

잘 알려져 있듯이, 통계적 추론과 관련해 드 무아브르 업적의 집결체는 '시행수의 증가에 따른 이항분포의 정규근사라는 결과'라고 할 수 있다. 그는 이 결과를 1733년에 〈해석학 논문집(Miscellanea Analytica)〉의 주석으로 발표하였고 1738년과 1756년에 〈우연론(Doctrine of Chances)〉의 2판과 3판에도 포함시켰다. 그렇다면, 그는 이 결과를 통계적 추론과 관련해 어떻게 활용하였을까?

오늘날의 관점에서, 드 무아브르가 제시했던 통계적 추론을 '이항분포의 정규근사라는 결과'와 관련해 기술하면 다음과 같다 [1, 4, 8, 12].

① 모수(모비율, 모평균, 참값)가 알려져 있다고 하자.
② 충분히 큰 횟수 n의 시행이 이루어진 것처럼 가정하자. 즉, 실제의 실험이 이루어지지 않았더라도 아주 큰 시행이 n회에 걸쳐 이루어진 것처럼 간주한다.
③ 성공횟수 X의 분포, 성공률을 나타내는 확률변수 $\dfrac{X}{n}$의 분포는 정규분포에 근사한다.

④ 실험의 관측값인 성공횟수 또는 성공률이 어떤 범위 안에 있을 확률을 알아낼 수 있다.

드 무아브르는 자신의 방법을 적용해 모비율이 $\frac{1}{2}$일 때 $3,600$회의 실험에서 성공횟수가 $1,750$회 이상이고 $1,850$회 이하일 확률을 구하는 추론방식을 다음과 같이 기술하였다.

> 특별한 문제에 이 방법을 적용하려면, 실험의 실시 횟수, 또는 계획한 횟수를 나타내는 수의 제곱근을 가지고 사건이 일어나거나 일어나지 않은 빈도를 추정할 필요가 있다. 이 제곱근은 이미 네 번째 따름 정리에서 시사한 것과 같이 우리가 추정을 통제하는 단위(모듈러스, modulus)와 같다. 예를 들어, 실험 횟수가 $3,600$회일 때 사건이 $1,850$회 넘게 일어나지도 않고 $1,750$회 미만으로 일어나지도 않을 확률을 알아내려면 (물론, 중간의 합 $1,850$으로부터 같은 거리면 어떤 두 수를 택해도 된다) 우선 $1,850$과 $1,750$의 거리를 구해서 그 거리의 반을 계산한다. 즉 지금의 경우 $50 = s\sqrt{n}$에서 $3,600 = n$이라 가정했으므로 $\sqrt{n} = 60$이고, $50 = 60s$에서 $s = 50/60 = 5/6$가 된다. 따라서 우리가 이런 시행을 무한히 거듭하여 $\frac{5}{6}\sqrt{n}$까지의 구간 안에 있는 항들을 합한 것과 모든 항을 합한 것의 극한 비율을 구하면 원하는 확률에 상당히 가까운 값을 얻게 된다 [13, p. 208 재인용].

드 무아브르의 통계적 추론의 방법이 지닌 뚜렷한 특징은 그 방법이 가설-연역적이라는 데에 있다. 즉, 미지의 모수가 마치 고정된 것처럼 가정한 상태에서 그 모수를 추정하는 확률변수(실험에서의 성공 횟수 또는 성공률)의 분포를 만들고서, 이 분포를 이용하여 실제의 성공 횟수(또는 관찰값)가 어떤 범위 안에 있게 될 확률을 연역적으로 구한다고 하겠다. 정확히 말해, 실험을 실제로 하지 않은 상태에서 선험적으로 그 확률을 구할 수 있는 것이다.

위 예에서는, 만약 어떤 사건이 일어날 모비율 또는 참값이 $\frac{1}{2}$이라는 가정 하에서 $3,600$회 실험 중에서 그 사건이 일어날 횟수가 $1,750$회 이상이고 $1,850$회 이하일 확률을 '이항분포의 정규분포 근사'를 통해 실제로 실험을 하지 않고도 알아낼 수 있다고 하겠다. 다시 말해, X가 $3,600$회 실험에서 어떤 사건의 성공횟수를 나타내는 이항확률변수이고 Θ를 모비율이라 할 때, 드 무아브르의 방법은 $P(1,750 \le X \le 1,850 \,|\, \Theta = \frac{1}{2})$와 같은 값을 알게 해 준다.

하지만 이러한 드 무아브르의 통계적 추론의 방법이 지닌 특징은 뚜렷한 단점 혹은 한계로 작용한다. 그의 방법에서 실험의 관측값이 어떤 범위 안에 있을지를 추정하는 것에 머문다는 것에 주목하게 되면, 드 무아브르의 방법이 우리가 궁금해 하는 것에 전혀 답을 주지 못하는 것을 곧장 파악할 수 있다. 즉, 우리의 관심사는 실험의 어떤 관측값이 실제

로 나왔을 때 그 값을 통해 모비율 또는 그것이 어떤 범위 안에 있을지를 추정하는 것이라 하겠는데, 드 무아브르의 통계적 추론의 방법은 이에 대해 전혀 답을 주지 못하는 것이다. 이러한 한계에 대해, 스티글러는 다음과 같이 언급하였다.

> 드 무아브르의 결과는 두 가지 측면에서 당시 사람들이 알고 싶어 했던 추론 문제에 적절한 답을 제시하지 못했다. (중략) 하지만 거꾸로 참값 $a/(a+b)$는 모르고 관측한 비의 값만 안다고 할 때 이 방법으로 미지의 참값이 관측한 비의 값으로부터 정해진 같은 거리 안에 있게 될 확률은 계산할 수 없다. (중략) 그의 한계는 얻은 데이터의 배후에 숨은 미지의 확률에 대한 확률적인 추론보다는, (참값은 이미 알고 있다고 가정하고) 데이터에서 계산하는 비나 승수 등에 대한 확률적인 추론만을 생각했다는 것이다 [13, p. 213].

드 무아브르의 통계적 추론은, 모수가 알려져 있지 않을 때 그것이 마치 알려진 것처럼 가정하고서 관측값이 그 미지의 고정된 모수와 관련해 어떤 범위 안에 있을 확률을 알려주지만, 실제 실험 이후에 미지의 모수 자체가 어떤 범위 안에 있을 확률에 대해서는 전혀 다루지 못한다고 하겠다.

구체적으로, 그의 방법은 어떤 고정된 모비율 $\Theta = \theta$가 주어져 있을 때, n회의 이항 실험에서 어떤 사건의 성공횟수 X와 관련해 확률값 $P(x_1 \le X \le x_2 \,|\, \Theta = \theta)$또는 확률값 $P(\frac{x_1}{n} \le \frac{X}{n} \le \frac{x_2}{n} \,|\, \Theta = \theta)$을 알려줄 수 있지만, 미지의 모비율이 관측한 비의 값으로부터 정해진 거리 안에 있게 될 확률 $P(\frac{x_3}{n} - a \le \Theta \le \frac{x_3}{n} + b \,|\, \frac{X}{n} = \frac{x_3}{n})$에 대해서는 알려주는 바가 전혀 없는 것이다. (단, $0 < a < \frac{x_3}{n}$, $0 < b < 1 - \frac{x_3}{n}$)

이런 점에서, 드 무아브르의 연구를 넘어서기 위해서는 그 추론의 대상 자체가 '모수', '참값', '모비율', '모평균'이 되는 통계적 추론의 방법의 등장이 요청되었다고 하겠다.

3 베이즈 추론에 담긴 핵심적인 아이디어

베이즈가 새롭게 도입한 통계적 추론은 드 무아브르 연구의 한계에 대한 일종의 도전이라 할 수 있다. 우리는 베이지의 사후 저작인 ⟨우연의 원리로 문제를 해결하는 에세이⟩의 서문에서 이를 확인할 수 있다. 친구인 리처드 프라이스가 원래의 서문을 수정한 것이긴 하지만, 이 서문에는 베이즈가 어떤 의도를 가지고 새로운 통계적 추론의 방법을 만들었는지가 선명하게 드러나 있다.

> 그(베이즈)가 이 에세이에 작성했던 서문에는, 그가 말하길, 에세이의 주제에 대해 생각할 때 그의 의도는, 제시된 상황에서 어떤 사건이 일어날 확률에 대해서는 아무것도 모르지만 같은 상황에서 그 사건이 몇 번을 성공했고 몇 번을

> 실패했는지가 제시되었다고 할 때, 그 사건이 일어날 확률에 대해 판단을 내리는 방법을 찾는 것이라 한다 [1, pp. 370–371; 10, p. 152 재인용].

앞 절에서 제기했듯이, 드 무아브르의 방법은 어떤 고정된 모비율 $\Theta = \theta$가 주어져 있을 때, n회 이항 실험에서 어떤 사건의 성공횟수 X와 대해 $P(\frac{x_1}{n} \le \frac{X}{n} \le \frac{x_2}{n} \,|\, \Theta = \theta)$을 알게 해 주지만, 그의 방법은 $P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n})$ (단, $0 < c < d < 1$이고 $x = 0, 1, 2, \cdots, n$)에 대해 알려주는 바가 없다. 위의 인용문에 드러나듯이, 베이즈는 바로 $P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n})$을 알아내는 방법을 고안하고자 했던 것이다. 그렇다면, 그는 이 확률을 알아내는 추론 방법을 어떻게 고안해 낼 수 있던 것일까?

우리는 베이즈가 그 추론방법을 고안했던 경로를 발견술적 측면에서 거슬러 올라가면서 분석적으로 탐색해보고자 한다. 그가 알아내고자 했고 실제로도 최종적으로 제시했던 결과물인 $P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n})$ 또는 $P(c \le \Theta \le d \,|\, X = x)$로부터 출발해서 거꾸로 진행하여 초기 조건에 이르는 과정을 거치면서, 그에게 어떤 사고와 아이디어가 요청되었고 그가 이때 어떤 처리를 했을지에 대해 추측 및 고찰해보도록 하자.

우선, 모비율이 어떤 범위 안에 있다는 것, 즉 $c \le \Theta \le d$를 확률과 관련시켜 다루기 위해서는 어떤 사고가 필요한 것일까? 이를 위해서는 모비율을 어떤 '고정 확정된 수'라는 관점으로부터 벗어나는 유연한 사고가 요구된다고 할 수 있다. 어떤 고정된 모비율 θ를 추정하기 위해 도입된 추정량 $\frac{X}{n}$와 그것의 분포만을 가지고는 근본적으로 $P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n})$를 알아낼 수 없는 상황에서, 우선적으로 해야 할 것은 "모수를 정확히 알 수 없지만 어떤 하나의 값으로 고정되어 있다."는 관점을 고수하는 것으로부터 벗어나야 하는 것이다.

이 상황에서 베이즈는 어떤 탈출구를 마련했을까? 베이즈는 $P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n})$의 값을 알아내는 것은 모비율 자체에 대한 직접적인 추론을 하는 것이기에, 이러한 직접적인 추론 방식을 도입하기 위해 모비율을 확률변수로 간주하기 시작하였다. 즉, 모수는 관찰할 수도 없고 알 수도 없는 대상, 즉 일종의 불확정성을 지닌 값이기에 모수를 오히려 확률변수로 다루면서 모수에 대한 직접 추론을 하고자 했던 것이다. 그렇다면, 모비율 Θ를 확률변수로 도입한 상황에서 $P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n})$을 구하기 위해서는 무엇을 알아내야 하는 것일까?

$P(c \le \Theta \le d \,|\, \frac{X}{n} = \frac{x}{n}) = P(c \le \Theta \le d \,|\, X = x)$이고, 이 확률은 일종의 조건부 확률값이고 모비율 Θ는 0과 1사이의 값을 취하는 연속적인 확률변수이다. 이러한 사항을 고려하면, $P(c \le \Theta \le d \,|\, X = x) = \int_c^d f_{\Theta \,|\, X}(\theta \,|\, x) d\theta$임을 알 수 있다. 그러기에, 베이즈가 제시하고자 했던 것, 즉 어떤 미지의 사건이 몇 번을 성공했고 몇 번을 실패했는지가 제시되었다고 할 때 그 사건이 일어날 확률에 대해 확률적 판단을 내리기 위해서는 $X = x$가 주어졌을 때 모비율 Θ에 대한 조건부 확률밀도함수 $f_{\Theta \,|\, X}(\theta \,|\, x)$를 먼저 구할 필요가

있었다. 그러면, 조건부 확률밀도함수 $f_{\Theta\,|\,X}(\theta\,|\,x)$를 구하기 위해서는 이제 무엇을 알아내야 하는 것일까?

$f_{\Theta\,|\,X}(\theta\,|\,x) = \dfrac{f_{X,\Theta}(x,\theta)}{f_X(x)}$ 이다. 그렇다면, 확률변수 Θ와 X의 결합확률함수 $f_{X,\Theta}(x,\theta)$ 그리고 확률변수 X의 확률질량함수 $f_X(x)$을 각각 알아야 할 것이다. 그런데 $f_{X,\Theta}(x,\theta)$와 관련하여 이항분포 $f_{\Theta\,|\,X}(\theta\,|\,x) = \dfrac{f_{X,\Theta}(x,\theta)}{f_X(x)} = _n C_x\theta^x(1-\theta)^{n-x}$의 공식은 이미 알려져 있으므로, 즉 $f_{X,\theta}(x,\theta) = f_X(\theta)_n C_x\theta^x(1-\theta)^{n-x}$이기에, $f_{X,\Theta}(x,\theta)$을 알기 위해서는 확률변수 X의 확률밀도함수 $f_X(\theta)$을 알아야 한다.

정리하자면, $P(c \le \Theta \le d\,|\,\frac{X}{n} = \frac{x}{n})$을 구하기 위해서는 어떤 미지의 사건이 일어날 확률에 대해 아무것도 모르는 상황에서 그 미지의 사건이 일어날 확률 Θ의 확률밀도함수 $f_\Theta(\theta)$ 그리고 n회의 이항 실험에서 그 사건의 성공횟수 X의 확률질량함수 $f_X(x)$을 알아야 하는 것이다. 그리고 반대로 이 두 가지만 알아낼 수 있다면 $P(c \le \Theta \le d\,|\,\frac{X}{n} = \frac{x}{n})$의 값을 제시할 수 있는 것이다. 그렇다면, 베이즈는 어떤 사건이 일어날 확률에 대해 아무것도 모르는 '무정보'의 상황에서 $f_\Theta(\theta)$와 $f_X(x)$의 정보를 어떻게 제시할 수 있었을까?

베이즈는 그의 독특한 사고실험 과정에서 미지의 사건이 일어날 확률을 다른 확률변수의 누적분포함수로 나타내는 일반적 방법이 있음을 알게 되고 이를 통해 $f_\Theta(\theta)$와 $f_X(x)$의 정보를 제시할 수 있었던 듯하다. 이렇게 말한 이유는, 그가 임의의 $\theta \in [0,1]$에 대해 $f_\Theta(\theta) = 1$임과 $x = 0,1,2,\cdots,n$에 대해 $f_X(x) = \dfrac{1}{n+1}$임을 제시했지만 이에 대한 수학적 정당화를 제공하지는 않았기 때문이다. 하지만 그가 미지의 사건을 누적분포함수에 의해 다룬 방식 안에 이미 그 정당화의 윤곽이 이미 선명하게 들어있다고 할 수 있다.

간략하게, 베이즈가 $f_\Theta(\theta)$와 $f_X(x)$을 알아내기 위해 수행했을 것으로 보이는 사고실험을 베이즈의 테이블과 현대적 기호를 이용하여 추측해서 나타내면 다음과 같다 [1, 7].

> 어떤 정사각형1) 테이블 위에 하나의 공을 내려놓았을 때 그 공이 멈추게 되는 '가로방향의 상대적 위치'가 있을 것이다. 테이블 위에 공이 놓이게 되는 상대적 위치는 알 수 없는 것이다. 예를 들어, Figure 12) 에서와 같이, 어떤 공을 테이블 위에 던질 때 그 공이 멈춘 지점을 통과하는 직선 중에서 AD에 평행한 직선을 그었을 때 그 직선은 CD, AB와 각각 s, o에서 만난다고 하자. 이때, AB에 대한 Ao의 비인 상대적 위치의 값 $\frac{\mathrm{A}o}{\mathrm{AB}}$를 확률변수 L이라고 하자.
>
> 이런 상황에서, 어떤 다른 공을 테이블 위에 던졌을 때 그 공이 $ADso$에 놓이게 되는 사건 M은 그 사건에 대한 정보를 전혀 알 수 없는 미지의 사건이다. 이때, 이 사건이 일어날 확률 θ을 생각해 보자.

1) 가로와 세로 방향의 기준만 정해질 수 있다면, 여기서의 논의는 어떤 하나로 연결된 형태의 단일 테이블에도 같은 방식의 논의가 적용될 수 있다.

2) 베이즈의 에세이 [1]에 제시되어 있는 그림

여기서, 확률변수 L의 임의의 확률밀도함수 $f_L(l)$에 대해 $\theta = \int_0^l f_L(\omega)d\omega = F_L(l)$이라 할 수 있다. 즉, 미지의 사건 M이 일어날 확률을 나타내는 확률변수를 Θ라고 하면 $\Theta = F(L)$인 것이다. 그런데 확률변수의 변환과 누적분포함수의 성질에 의해 $\theta = F_L(l) = F_\Theta(\theta) = \int_0^\theta f_\Theta(s)ds$가 성립한다. 미분을 하면, 임의의 $\theta \in [0,1]$에 대해 $f_\Theta(\theta) = 1$임을 알 수 있다. 그러면, $x = 0, 1, 2, \cdots, n$에 대해 $f_X(x) = \int_0^1 {}_nC_x\theta^x(1-\theta)^{n-x}d\theta = \dfrac{1}{n+1}$임을 알 수 있다.

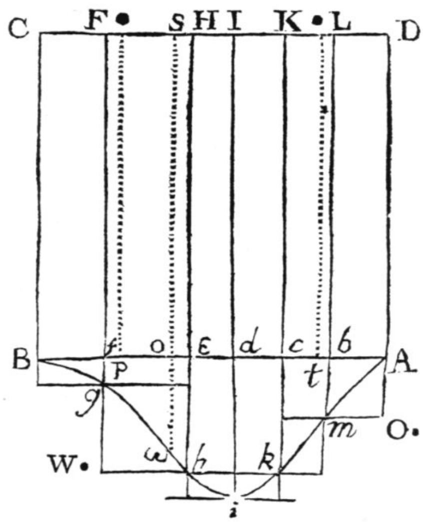

Figure 1. Table of Bayes; 베이즈의 테이블

그런데 베이즈가 종합에 의해 최종적으로 제시한 자신의 추론을 제시했을 때에는 위의 두 확률변수 L, Θ를 하나로 줄였다고 할 수 있다. 확률변수 L의 임의의 확률밀도함수 $f_L(l)$에 대해 , $f_\theta(\theta) = 1$, $f_X(x) = \dfrac{1}{n+1}$이 성립하기에 $f_L(l) = 1$(단, $0 < l < 1$)로 두어서 $L = \Theta$가 되도록 하는 초기상태를 설정하였다. 즉, 그는 평평한 정사각형 테이블을 가정하고서 종합적으로 자신의 추론을 제시하였다. 그 종합된 내용은 다음과 같다.[3)]

① 임의의 $\theta \in [0,1]$에 대해 $f_\Theta(\theta) = 1$이고 $x = 0, 1, 2, \cdots, n$에 대해 $f_X(x) = \frac{1}{n+1}$이다.

② $f_{X|\Theta}(x|\theta) = \frac{f_{X,\theta}(x,\theta)}{f_\Theta(\theta)} =_n C_x\theta^x(1-\theta)^{n-x}$이다.

③ $f_{X,\Theta}(x,\theta) = f_\Theta(\theta)_nC_x\theta^x(1-\theta)^{n-x} =_n C_x\theta^x(1-\theta)^{n-x}$이다.

④ $f_{\Theta|X}(\theta|x) = \frac{f_{X,\theta}(x,\theta)}{f_X(\theta)} = (n+1)_nC_x\theta^x(1-\theta)^{n-x}$이다.

3) Figure 2의 우측은 $n = 10$일 때의 결합확률함수 $f_{X,\Theta}(x,\theta)$의 그래프를 나타낸 것이다. 11개 각각의 곡선 아래의 면적은 $\frac{1}{11}$이다.

⑤ $P(c \leq \Theta \leq d|X = x) = \int_c^d f_{\Theta|X}(\theta|x)d\theta = \int_c^d (n+1)_n C_x \theta^x (1-\theta)^{n-x}$ 이다.

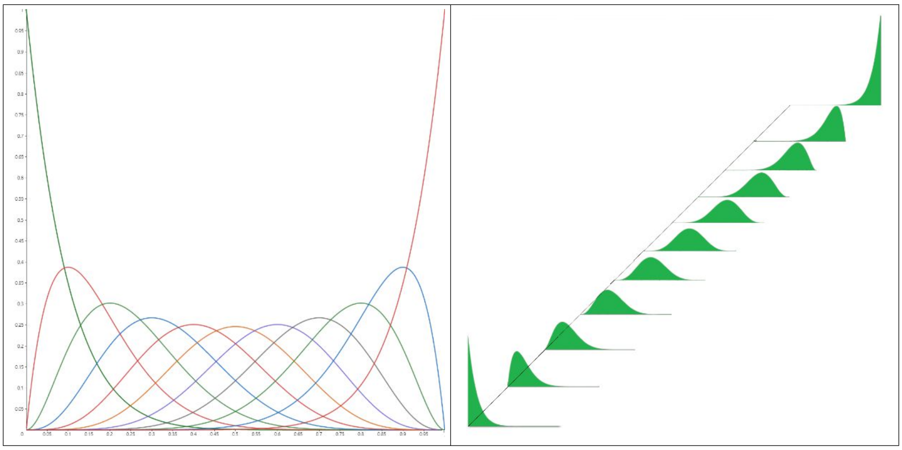

Figure 2. joint probability distribution($n = 10$); 결합확률분포($n = 10$)

마침내, 베이즈에 의해 드 무아브르의 통계적 추론의 한계가 극복된 것이라 할 수 있다. 그런데 우리는 드 무아브르의 통계적 추론이 지닌 한계를 오늘날의 신뢰구간을 이용한 추정에서도 여전히 포착할 수 있다. 이에 대한 논의를 전개하기 위해, 매우 동떨어진 것처럼 보이는 '슈뢰딩거의 고양이 사고실험' 의 결과를 확률 및 통계적 관점에서 다음 절에서 고찰하려고 한다. 이 분석은 베이즈 추론을 학교수학에 도입하고자 할 때, 우리가 무엇에 주목해야 하는지를 보여줄 것이다.

4 '슈뢰딩거의 고양이' 와 신뢰구간 : 교육에서 주목해야 하는 것

슈뢰딩거의 고양이 사고실험[4)]에 대한 양자역학계에서의 논쟁이 베이즈 추론 및 그것을 교육과정에 도입하는 것과 무슨 관련이 있을지에 대한 의문을 제기할 수 있을 것이다. 하지만 앞으로 살펴보겠지만, 빈도주의 통계학의 관점에서 슈뢰딩거 고양이 실험의 결과를 해석한

4) 보어(N. Bohr), 하이젠베르크(W. K. Heisenberg), 보른(M. Born)을 주축으로 한 코펜하겐 학파는 "어떤 입자에 대해 위치의 오차와 운동량의 오차의 곱은 항상 어떤 수보다 크기에, 위치와 운동량을 동시에 정확하게 측정하는 것은 실제적으로 그리고 이론적으로 불가능하다." 는 불확정성의 원리를 제기하였다. 그런데 이를 반박하기 위한 목적으로, 슈뢰딩거(E. R. J. A. Schrödinger)는 어떤 특정한 상황 속에서 고양이의 생사여부를 판단하는 사고실험을 제안한 바 있다. 잘 알려져 있듯이, 슈뢰딩거와 더불어 아인슈타인은 '신은 주사위 놀이를 하지 않는다.' 고 말하면서 제 5차(1927년) 및 6차(1930년) 솔베이 회의에서부터 시작해 줄곧 코펜하겐 학파의 불확정성의 원리를 반박하려 했다. 특히, 그들은 이 고양이 사고실험을 통해 불확정성의 원리가 "이 실험 후의 관찰이 안 된 고양이는 산 것과 죽은 것의 중첩상태에 놓여 있다."는 역설을 낳을 수 있음을 제기하고 그 이론을 반박하고자 하였다. 당시 수준으로는 '빛을 통한 관찰에 의해 빛의 광자가 전자에 충돌하는 현상이 일어나므로 정확한 측정이 이루어지지 않아 전자의 위치와 운동량을 동시에 정확히 측정할 수 없다.'는 것을 인정하지만, 이 실험의 시사를 통해, 그들은 기술과 물리학의 발전(일명 '숨은 변수' 이론)에 의해 언젠가는 전자의 위치와 운동량 모두를 동시에 어떤 오차 범위 내에서 측정할 수 있음을 보여주는 이론이 나올 것이라 주장하였다[3, 7].

일명 '앙상블 해석' 에 대한 논의를 통해 우리는 베이즈 추론 도입이 요구되는 지점을 오히려 정확하게 찾아낼 수 있을 것이다.

물론, 우리는 코펜하겐 학파가 주장한 불확정성의 원리가 미시세계의 근본적인 물리적 특성인지는 명확히 알지 못한다. 또한, 이 원리는 이 연구의 범위를 벗어난 주제이기도 하다. 그러기에, 이 연구에서는 불확정성의 원리 자체와 슈뢰딩거의 고양이 실험에 대한 여러 해석의 타당성 등에 대해 분석 및 평가는 하지 않는다. 다만, 오직 고양이 사고실험의 결과인 고양이의 생사 여부의 판단에 대해서만 확률 및 통계의 관점에서 논의하려고 한다.

이를 위해 이 연구에서는, 슈뢰딩거 고양이 사고 실험에서의 미시세계의 측면을 제거함으로써 미시세계와 거시세계 사이의 연결성 측면을 고려하지 않은 상태에서, 오직 거시세계에서 '슈뢰딩거 고양이의 생사여부 판단' 에 대해서만 확률과 통계의 관점에서 다룬다. 그러면서, 슈뢰딩거 고양이의 생사여부 판단' 이면의 확률 및 통계적 기본가정에 대해 분석하고 이를 통해 교육적 시사점을 도출하고자 한다.

한편, 이 연구에서 다루는 내용과 목적에 부합하도록, 이 실험을 의도적으로 거시세계로 한정하여 단순화시키고 실험 속의 확률을 50%에서 95%로 조정시켜서 만든 실험을 슈뢰딩거의 입장에서 진술하면 다음과 같다 [2, 7].

> 어떤 상자 안에 타이머가 작동하는 장치에 의해 100% 살상률을 지닌 독가스가 어떤 일정한 시간 동안에 분출될 확률이 95%라고 하자. 여기서, 상자가 닫힌 상태에서 독가스가 분출되면 금세 임의의 생명체는 죽게 된다고 하자. 한 마리의 고양이를 그 상자 안에 놓고서 타이머를 작동시키고 그 일정한 시간이 지났을 때, 고양이는 어떤 상태일까? 상자를 열지 않아 고양이를 관찰할 수 없는 상황, 즉 고양이가 죽었는지의 여부를 판명할 수 없는 상황에서도 고양이는 죽었든지 아니면 안 죽었든지 둘 중의 하나이다. 즉, 그 고양이는 확률 1로 죽었든지 아니면 확률 0으로 살았든지 둘 중의 하나이다.

사실, 이러한 고양이 생사여부의 판단에 대한 슈뢰딩거의 주장은 빈도주의 통계학의 관점에서 제기된 일명 앙상블 해석에 의해 옹호되었다. 이 통계적 해석은 다음과 같다.

> 1억 번의 실험을 하게 되면 9천500만 번의 실험에서는 고양이가 죽을 것인데, 우리가 한 실험은 그러한 실험 중의 하나이다. 실험을 한 이후에는, 고양이의 생사 여부를 육안으로 명확히 확인할 수 없는 상황에서도, 고양이가 죽었든지 살았든지 둘 중의 하나일 뿐이며 '고양이가 죽어 있을 확률이 95%이다.' 와 같은 말은 전혀 할 수 없다. 이때의 확률은 1 또는 0일 뿐이다.

그런데 이 해석 방식은 우리에게 매우 익숙하다. 사실, 슈뢰딩거의 주장과 그것을 지지하는 앙상블 해석은 빈도주의자의 신뢰구간에 대한 해석에서 똑같은 모습으로 다시 등장한다.

> 1억 개의 신뢰구간을 구하는 실험을 하게 되면 9천500만 개의 신뢰구간 안에는 모수가 들어있을 것이다. 우리가 구한 95% 신뢰구간은 그러한 신뢰구간 중의 하나이다. 모수가 실제로 그 신뢰구간 안에 들어있는지를 확인할 수 없지만, 모수가 신뢰구간 안에 들어있든지 아니든지 둘 중의 하나이며 '이 신뢰구간 안에 모수가 들어있을 확률이 95%이다.' 와 같은 말은 전혀 할 수 없다. 이때의 확률은 1 또는 0일 뿐이다.

'고양이', '고양이의 죽음' 을 각각 '신뢰구간', '신뢰구간 안에 모수가 들어감' 으로 대응시키면 두 해석 방식이 완벽하게 동일하다는 것을 알 수 있다. 그런데 이 두 해석에서 공통적으로 발견되는 전형적 특징이 있다. '확률적 판단을 하지 않는다.' 는 것이 바로 그것이다. 실험 이후 어떤 특정한 고양이의 생사여부를 확인할 수 없는 상황에서 '고양이가 죽었을 확률' 을 다루지 않으며, 마찬가지로, 어떤 신뢰구간을 구한 후에 그 신뢰구간 안에 모수가 들어있는지 판단할 수 없는 상황에서 '신뢰구간 안에 모수가 들어있을 확률' 을 다루지 않는다.

그렇다면 왜 특정한 1마리의 고양이 생사여부와 특정한 1개의 신뢰구간의 모수 소유여부에 대해 확률을 부여하지 않는 것일까? 이 상황에서, 확률적 판단을 하지 않는 이유는 관찰에 따른 확인 여부에 상관없이 그 결과는 이미 결정되어 있다고 보기 때문이다.

이와 비교해, 베이즈 추론에서는 고양이 생사여부와 신뢰구간의 모수 소유여부가 불확정한 상황 속에서 확률적 판단을 할 수 있는 수단을 제공한다. 그 수단이란 불확정한 대상을 확률변수로 취급하는 것이다.

물론, 베이즈 추론에서도 '1번의 실험 후에 고양이가 죽었든지 아니면 살았든지 둘 중의 하나이다.' 는 진술은 인정한다. 하지만 고양이의 생사여부에 대한 확인이 불가능한 상황에서 고양이의 죽음을 확률적으로 다루는 방법을 제공해 줄 수 있다. 즉, 한 차례의 실험 후에 고양이가 죽을 확률은 어떤 고정된 값이 아닌 새로운 확률변수로 취급하고 그 확률분포를 통해 다루는 것이다. 이것은 흡사 어떤 선출직 선거의 투표가 끝난 후 개표가 이루어지지 않은 상황[5] 또는 부분적으로 개표가 이루어진 상황에서 어떤 특정 사람이 당선될 확률을 확률변수로 다루는 것과 같다.

구체적으로, 1번의 실험 후에 고양이가 죽을 확률을 어떤 고정된 모수가 아니라 확률변수 Θ로 놓고 그에 대한 사전분포를 생각해 보자. 그러면, 이미 제시된 어떤 일정한 시간 동안에 독가스가 분출될 확률인 95%, 즉 고양이가 죽을 확률로 제시된 95%는 확률변수 Θ가 이루는 분포의 기댓값으로 간주할 수 있다. 사전 데이터가 있다면 그것에 기초해 사전분포를 설정할 수도 있고, 아니면 기댓값이 0.95인 베타분포로서의 사전분포 $f_\Theta(\theta) = 10\theta^{18}$(단, $0 \leq \theta \leq 1$)를 도입할 수도 있을 것이다.

5) 투표가 마무리되었기에, 슈뢰딩거 고양이의 죽음 여부와 마찬가지로, 어떤 사람의 당선 여부는 이미 결정되어 있음에도 불구하고 그 사람의 당선 확률의 분포를 다룬다.

이처럼 모수를 나타내는 확률변수에 대한 분포를 도입할 경우에는, 관찰이 불가능한 상황에서도 $P(0.91 < \Theta < 0.99) = 0.66$와 같은 확률적 판단을 내릴 수 있다. 즉, "고양이가 죽었을 확률이 0.91과 0.99사이에 있을 확률이 0.66이다."와 같은 확률적 판단을 내릴 수 있는 것이다.

만약 7번의 추가실험에서 7마리의 고양이 중에서 5마리는 생존하고 2마리가 죽었다는 것을 확인할 수 있었다면, 이 결과를 반영한 사후분포는 $g_\Theta(\theta) = 7800\theta^{23}(1-\theta)^2$이 되는데, 우리는 이 분포를 이후의 실험에 대한 새로운 사전분포로 활용할 수도 있다. 그래서 추가 1회의 실험에서 고양이의 생존을 확인할 수 없는 경우에는, 이 분포를 활용해 고양이가 죽었을 확률로 기대되는 값은 0.89이고[6] $P(0.85 < \Theta < 0.93) = 0.5$와 같은 판단을 내릴 수 있다. 즉, "1번의 실험에서 고양이가 죽을 확률이 0.85와 0.93사이에 있을 확률이 0.5이다."와 같은 확률적 판단을 내릴 수 있는 것이다.

사실, 베이즈 통계학자들 역시도 어떤 하나의 기존 95% 신뢰구간 안에 모수가 들어있든지 없든지 둘 중의 하나라는 사실은 명확하게 인정한다. 신뢰구간에 대한 빈도주의적 해석 자체에 대해선 이의를 제기하지 않는 것이다. 즉, 고양이가 죽어있든지 살아있든지 둘 중의 하나라는 것을 부인하지 않는다. 단지, 그들은 이러한 상식적 주장에 대해 반대하려는 것이 아니라 '불확정한 값 또는 대상'을 확률변수로 또는 확률적으로 다루어야 한다고 주장한다. 즉, 모수를 알 길이 없다면, 그 모수를 고정 확정된 것으로 간주하지 않고 그것을 불확정한 것 자체로 취급하면서 확률변수로 다루어야 한다고 말한다.

한편, 앙상블 해석에서 취하는 빈도주의적 관점[7]은 추정량의 분포를 통해 도저히 알아낼 수 없는 것이 있다는 사실을 명확하게 확인시켜 준다. 즉, 어떤 관측이나 실험을 통해 얻은 관측값을 이용하더라도 모수와 관련해서 확률적 판단을 전혀 할 수 없다는 것이 바로 그것인데, 이는 추정량의 분포를 이용한 통계적 추론의 근본적 한계를 보여주는 것이기도 하다.[8]

예를 들어, 앞서 이미 살펴보았듯 하나의 관측값에 의해 만들어진 신뢰구간이 주어졌을 때 추정량의 분포를 이용해 모수가 그 신뢰구간에 들어있는 것에 대한 확률적 판단을 할 수 없다. 추정량의 분포만을 가지고서는 우리가 알고 싶은 것 자체를 알아낼 수 없는 것이다. 기호적으로 표현하자면, 모수와 그에 대응하는 추정량을 나타내는 확률변수를 각각 Y, X라고 할 때, 어떤 고정된 미지의 모수 y_0에 대한 추정량의 분포 $f_{X|Y}(x|y_0)$를 이용할 경우에는 관측값 x_0와 그 값을 이용한 어떤 신뢰구간을 구하더라도 그 모수 y_0와 관련하여 확률적 판단을 할 수가

6) 이때의 기댓값이 0.95가 아닌 것은 자료에 기초해서 분포를 수정했기 때문이다.

7) 앙상블 해석에서는 추정량의 분포를 수많은 실험을 통해 이루어진 관측값들이 이룰 분포로 간주하면서 실제로 구한 하나의 관측값(관측치, estimate)에 대해서는 확률적 판단을 하지 않는다는 입장을 취한다. 그래서 앙상블 해석에서는 전자의 관측 위치와 운동량의 추정량 분포를 통해서 특정한 전자 1개의 위치와 운동량을 이용한 확률적 판단을 할 수 없다는 사항을 통해, 코펜하겐 학파의 여러 주장에 대해 반박하려고 했다고도 볼 수 있다.

8) 이 진술은 앙상블 해석이나 빈도주의 통계학에서 취하는 관점의 타당성 자체를 언급하는 것이 아니며, 추정량 분포의 특성에 대해서만 언급하는 것이다. 또한, 이 연구가 베이즈 추론을 다루기에 그렇게 오해할 여지가 있지만, 여기서 '양자 베이즈주의'를 주장하려는 것은 더욱 아니다.

없다고 하겠다.

이와 대비해, 관측값 x_0와 대한 모수 Y의 사후분포 $f_{Y|X}(y|x_0)$을 이용할 경우에는 어떤 임의의 구체적인 값 y_1, y_2에 대해 $P(y_1 < Y < y_2 \mid X = x_0)$을 제시함으로써 모수와 관련한 확률적 판단을 내릴 수 있다. 이런 점에서, 어떤 관측값이 주어졌을 때 모수와 관련해 확률적 판단을 내리기 위해 요구되는 '숨은 변수'는 등잔 밑에 늘 있던 모수 자체라 하겠다.[9)]

이와 같이 확률과 통계의 관점에서 슈뢰딩거의 고양이 실험을 논의하는 과정을 통해, 우리는 자연스럽게 이 논의가 신뢰구간 추정과 밀접한 관련이 있음을 알 수 있었다. 이제, 이러한 분석결과를 학교수학에서의 신뢰구간 교육과 연관시켜보도록 하자.

이를 위해, 다음의 안경 소유의 예를 들겠다. 어떤 나라의 인구는 1억 명인데 그 중에 9천 5백만 명이 안경을 소유하고 있다고 하자. 그중에 1명을 표본으로 실제로 추출했을 때, 그 사람의 안경 소유를 확인할 수 없는 상황에서 그 사람이 안경을 가지고 있을 확률은 얼마라고 할 수 있는가? 이에 대해, 우리는 전통적인 빈도주의 통계학 및 상식에 따라 그 사람은 안경을 가지고 있든지 또는 그렇지 않던지 둘 중의 하나라고 가르친다. 즉, 그 확률은 1 또는 0 이라고 가르친다. 분명한 사실이다.

하지만 우리 학생들은 어떻게 느낄까? 만약 안경 소유여부를 확인할 수 없는 상황이라면 여전히 불확정한 상태이기에, 학생들은 그 확률이 0.95 일 것이라 잘못 생각하는 경향이 있기 마련일 것이다.

위의 예에서 '사람'을 '신뢰구간'으로 그리고 '안경'을 '모수'로 대체해 보면, 우리 학생들이 신뢰구간을 배우는 장면에서 직면한 인지적 어려움의 실체가 명확히 드러난다. 우리는 신뢰구간을 어떻게 가르치는가? 1억 개의 신뢰구간 중에서 9천 5백만 개의 신뢰구간 안에 모수를 가지고 있는데, 어떤 특정한 신뢰구간을 구했을 때 그 구간은 수많은 신뢰구간 중에 하나에 불과하고 그 구간 안에 모수가 들어있을 확률은 1 또는 0 이라 가르친다. 아이러니하게도, 그 신뢰구간 안에 모수가 들어있을 확률이라는 말 자체를 꺼내지 못하면서, 우리는 여전히 그 구간을 어떤 모수에 대한 95% 신뢰구간 또는 98% 신뢰구간 중의 하나라 부르도록 하고 있다. 이처럼 배운 학생들은 신뢰구간과 확률 사이의 관계에 대해 어떻게 여기게 될까?

학생들은 왠지 그 실현된 신뢰구간 안에 모수가 들어있을 확률이 0.95 라 말해야 할 것 같은데, 그것이 아니라고 하니, 교과서 또는 교재에 나온 표현을 외울 뿐이다. 사실, 학생들이 겪는 어려움의 본질은 근본적으로 모수의 불확정성에 기인하다고 할 수 있다. 설령 모수를 고정 확정된 값으로 간주하더라도 그 값이 무엇인지 모르는 상태이기 때문에, 구체적으로 구한 1 개의 신뢰구간 안에 불확정 상태의 그 모수가 들어있는지 여전히 알 수가 없는 것이다. 그러므로 '확률이 얼마이다.'는 말을 전혀 하지 못하게 하면서 '모수에 대한 몇 %의 신뢰구간'을

9) 아인슈타인이 제기한 '숨은 변수'에 대한 답을 의미하는 것은 아니다. 그의 '숨은 변수'는, 추정량 분포의 표준오차를 줄이기 위한 것이기에, 관측값을 이용하여 확률적 추론을 하는 것과는 거리가 있다.

논의하는 자체가 근본적으로 학생들을 인지적 혼란에 빠뜨릴 수밖에 없게 한다 하겠다.

이상의 슈뢰딩거의 고양이와 관련해 진척시켜온 논의는, 현재의 통계적 추론 교육을 개선하고자 하는 입장에서, 베이즈 추론의 도입과 관련해 어떤 교육적 변화가 이루어져야 하는지를 강하게 시사해 준다: 만약 학생들의 통계적 추론 역량을 실질적으로 향상시키기 위해 베이즈 추론을 학교수학에 도입하고자 한다면, 추정량뿐만 아니라 모수의 분포를 다루는 통계적 구간추정 교육이 이루어져야 한다. 왜냐하면 그렇게 하는 것이 베이즈 추론의 구별되는 실제 모습이기 때문이다.

5 결론

서론에서 우리는 "베이즈 추론이 발생하게 된 지적 계기와 그 추론의 발생에 기여한 핵심적 아이디어는 무엇인가?", "베이즈 추론의 발생 계기와 그 추론에 담긴 아이디어를 반영해 베이즈 추론을 학교수학에 도입한다고 하면, 학교수학에 어떤 변화를 요구하는 것일까?"와 같은 연구문제를 제기하였다.

이 연구에서는, 베이즈 추론이 출현하게 된 지적 계기가 '모수가 어떤 범위 안에 있을 확률을 구하는 방식이 존재하지 않은 상황'과 밀접한 관련이 있음을 드러내었다. 그리고 그 추론의 발생에 있어 모수를 확률변수로 취급한다는 아이디어가 결정적인 역할을 했다는 것을 보여주었다. 수학의 역사에서 상수를 문자로 나타내면서부터 추상적인 대수학이 등장할 수 있었듯이, 통계학의 역사에서도 모수를 확률변수로 나타내면서부터 새로운 통계적 추론이 등장하게 되었다고 하겠다.

한편, 우리나라의 학교수학 및 대학수학에서는 '모수가 어떤 범위 안에 있을 확률을 구하는 추론'을 거의 다루지 않는다. 그 대신에, 보편적으로 통계량 중에서 구간추정량으로서의 신뢰구간을 활용하여 통계적 추정을 하도록 한다.

매우 신기하게, 신뢰구간 이론이 나온 지 거의 1세기가 지났음에도 불구하고 학교수학에서 통계적으로 추론하는 방식은 베이즈 추론의 등장하기 이전에 사용했던 방식과 놀랍도록 비슷하다고 할 수 있다. 여러 통계 개념이 정교화된 것을 제외하고는, 오늘날 학교수학에서 다루는 구간추정 이론은 드 므아브르의 연구결과의 수준을 크게 벗어나지 못하고 있다. 사실, 이 신뢰구간 이론이 가진 원초적 한계는 그 이론이 등장했을 때부터 지금까지 줄곧 제기되고 있다.

네이만(J. Neyman)은 1934년에 영국의 왕립 통계 학회의 발표에서 〈대표하는 방법의 두 가지 다른 측면(On the two different aspects of the representative method)〉 논문을 통해, 오늘날의 신뢰구간(confidence interval)에 의한 구간추정 방식을 설명하고 제안하였다. 그런데 이때 사회를 본 바울리(G. M. Bowley)는 이 신뢰구간의 근원적 한계성을 다음과 같이 이미 제시한 바 있다.

> 제가 말하고 있는 건 네이만 박사의 신뢰한계(confidence limit)에 대한 것인데, 저는 여기서 '신뢰'(confidence)라는 단어가 '신뢰라는 이름의 속임수'(confidence trick)는 아닌지 확실히 모르겠습니다. (중략) 이 연구가 우리에게 어떤 추가적인 설명을 해주고 있는 겁니까? 우리는 토드헌터(Todhunter: 19세기 후반의 확률 학자)가 알던 것에 비해 더 많은 것을 알고 있습니까? (중략) 이 연구가 우리를 우리가 가야 할 곳으로 인도하고 있습니까? 이렇게 될 가능성이 신뢰구간 안에 있는 겁니까? 제기 보기엔 그렇지 않은 것 같습니다. (중략) 이건 이 방법론이 처음 제기되었을 때부터 줄곧 느껴온 어려움이었습니다. 이론은 설득력이 별로 없습니다. 적어도 제가 설득될 때까지는 그 진술의 타당성을 회의할 것 같습니다 [11, pp. 178–9 재인용].

바울리는 신뢰구간 이론이 나올 때부터 이 이론에 의해 모수 또는 참값에 대해 확률적 판단을 내릴 수 없다는 것을 정확하게 지적했던 것이다. 즉, 어떤 구체적 신뢰구간을 구했을 때 확률적 판단을 전혀 내릴 수 없는 것이다. 게다가, 모수의 참값이 어떤 구간 내에 존재할 확률에 대해 전혀 이야기하지 못하면서, 그러한 거짓 인상마저 슬그머니 주고 있다고 하겠다. 그러기에, 오늘날 학교수학을 통해 신뢰구간을 다루더라도 학생들은 '베이즈가 새로운 추론을 고안하고자 할 때의 상황' 에 여전히 머물러 있다고 평가할 수 있는 것이다.

이상의 논의는 "베이즈 추론의 발생 계기와 그 추론에 담긴 아이디어를 반영해 베이즈 추론을 학교수학에 도입한다고 하면, 학교수학에 어떤 변화를 요구하는 것일까?"와 같은 연구문제에 대해 비교적 분명한 답을 준다.

우리의 연구결과는 베이즈 추론 이면의 아이디어와 그 발생의 계기를 반영하여 베이즈 추론을 학교수학 및 수학과 교육과정에 도입하기 위해서는 모수를 고정 확정된 것이 아니라 불확정한 것, 즉 확률변수로도 다룰 수 있어야 함을 보여준다. 또한, 슈뢰딩거의 고양이 실험에 대한 논의에서 살펴보았듯이, 기존의 신뢰구간 이론이 지닌 한계를 인식하는 것이 모수로서의 확률변수를 도입하는 주요한 계기 또는 디딤돌이 될 수 있다는 것도 보여준다. 물론, 이러한 것들은 베이즈 추론을 도입하는 과정에서 신뢰구간에 대한 교수-학습과의 조정 작업이 이루어져야 한다는 것을 강하게 시사한다고 하겠다.

이제, 베이즈 추론을 학교수학 및 수학과 교육과정에 도입하는 것의 실제 의미에 대해 정리하도록 하자.

현재 우리는 $P(A\,|\,B)$와 $P(B\,|\,A)$에 대한 계산을 통해 어떤 사건의 확률을 업데이트하는 교수-학습에 의해 베이즈 추론을 어렴풋이 가르친다고 할 수 있다. 하지만 이처럼 조건부 확률인 $P(A\,|\,B)$와 $P(B\,|\,A)$를 동시에 계산하면서 스팸메일 필터, 몬티홀 문제 등을 해결하는 방식에 의해서는, '계산' 이라는 나무에 가려 '추론' 이라는 숲을 보게 할 수 없다. 즉, 베이즈 추론의 본질을 전혀 가르칠 수 없다는 근본적 한계가 있다. 왜냐하면 베이즈 추론은 어떤 사

건에 대한 '조건부 확률'이 아니라 어떤 모수 확률변수의 '조건부 확률분포'를 가지고 하는 추론이기 때문이다.

결론적으로 말해, 베이즈 추론을 통해서 학생들이 새롭게 경험해야 하는 것은 모수가 고정된 상태에서 추정량의 분포를 통해 간접적으로 추론하는 것이 아니라 모수를 확률변수로 취급하면서 그것의 조건부 확률분포를 통해 직접적으로 추론하는 것이다. 물론, 이러한 접근은 기본적으로 불확정성의 범위를 측정값뿐만 아니라 모수로까지 확장하는 메타적인 사고를 요구한다. 당연히, 이것이 베이즈 추론을 학교수학에 도입하려고 할 때 마주치게 되는 어려움의 근본적 실체이기도 하다.

이 연구에서의 논의를 종합해 보면, 발생적 측면에서 볼 때 기존의 신뢰구간이 지닌 한계를 인식해 가면서 모수를 유연하게 확률변수로 취급하고 그 확률변수의 분포를 다루는 '베이즈 추론'의 교육이 건전하다고 하겠다. 즉, 그 아이디어의 본질적 측면을 고려한다면, 베이즈 추론에 대한 교육은 현재의 신뢰구간의 지도에 대한 개선 및 조정에 초점을 두어 이루어지는 것이 바람직하다고 하겠다. 물론, 이러한 교육의 도입과 성공은 베이즈가 느꼈을 '추정량의 분포를 통해서는 모수가 어떤 범위 안에 있을 확률을 알아낼 수 없다.'와 같은 문제의식과 그 해법에 대한 아이디어를 교육적으로 접목시킬 수 있느냐의 여부에 달려있을 것이다.

References

1. T. Bayes, An Essay toward Solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London* 53 (1763), 370–418.
2. M. L. Bhaumik, Is Schrödinger's Cat Alive? *Quanta* 6 (2017), 70–80.
3. Choi JunGon, *Quantum Mechanics*, Seoul: BumHan Press, 2010. 최준곤, 양자역학, 서울: 범한서적, 2010.
4. A. I. Dale, *A History of Inverse Probability from Thomas Bayes to Karl Pearson*, Springer, 1999.
5. A. W. F. Edwards, Commentary on the arguments of Thomas Bayes, *Scandinavian Journal of Statistics* 5 (1978), 116–118.
6. D. A. Gillies, Was Bayes a Bayesian?, *Historia Mathematica* 14 (1987), 325–346.
7. J. Gribbin, *In Search of Schrödinger's Cat: Quantum Physics and Reality*, Bantam Books, 1984.
8. I. Hacking, *The Emergence of Probability*, Cambridge University Press, 1975.
9. Jung ChiBong, The Story of Bayes, an Amateur Mathematician who was a Minister, *Newsletter of the Korean Society of Mathematical Education* 35(1) (2019), 20–28. 정치봉, 목사였던 아마투어 수학자 베이즈 이야기, 한국수학교육학회 뉴스레터 35(1) (2019), 20–28.
10. Park SunYong, Bayes' Excuse for the Introduction of Prior Uniform Distribution, *Journal for History of Mathematics* 35(6) (2022), 149–170. 박선용, 베이즈의 사전균등분포의 도입에 대한 변명, 한국수학사학회지 35(6) (2022), 149–170.
11. D. Salsburg, *The Lady Tasting Tea*, W. H. Freeman and Company , 2001. (translated by Choi JungGuo, 2003). 최정규(역), 천재들의 주사위, 서울: 뿌리와이파리, 2003.

12. S. M. STIGLER, Who discovered Bayes's Theorem?, *The American Statistician* 37 (1983), 290–296.
13. S. M. STIGLER, *The History of Statistics*, The Belknap Press of Harvard University Press, 1986. (translated by CHO, J. G, 2002). 조재근(역), 통계학의 역사, 서울: 한길사, 2002.