Journal for History of Mathematics
Vol. 35 No. 6 (Dec. 2022), 149–170
http://dx.doi.org/10.14477/jhm.2022.35.6.149

Bayes' Excuse for the Introduction of Prior Uniform Distribution

베이즈의 사전균등분포의 도입에 대한 변명

Park Sun-Yong 박선용

This study discusses in terms of historical genesis whether it is reasonable for Bayes to introduce a prior uniform distribution. In this study, we try to analyze the way he dealt with postulates, lemmas, and propositions in Bayes' essay and to understand its characteristics. The results of the study show that Bayes used random variables for two parameters and that the two random variables were converted to each other through cumulative distribution. Furthermore, it is revealed that the introduction of prior uniform distribution can be justified by this way.

Keywords: Bayes, prior uniform distribution, cumulative distribution, random variable; 베이즈, 사전균등분포, 누적분포, 확률변수.

MSC: 01A50 ZDM: A35

1 서론

베이지안 추론이 오랜 기간 비판을 받아왔던 가장 큰 이유는 '선험적인 어림짐작'으로 치부되는 '사전 균등분포'에 대한 설정 때문이라 해도 과언이 아닐 것이다. 아무것도 모르는 상태를 균등한 상태로 정량화하는 것은 '처음에 확률(분포)를 대충 추측하면서 무지를 지식으로 가장하는 것'으로 간주될 수 있기에, 베이지안 추론에 대한 비판은 일견 타당해 보인다 [2, 4, 5, 6, 7].

이와 관련해, 통계사학자인 스티글러(S. M. Stigler)는 통계학 역사에서 처음 고안된 것으로 알려진 베이즈의 사전 균등분포는 오히려 타당한 것이었다고 주장한다.[1)] 사전 균등분포가 획일적이고 무분별하게 사용된 관행에는 문제의 소지가 있지만, 적어도 베이지안 추론이 태동되었던 당시에는 균등분포가 매우 신중하게 사용되었다고 주장하였다. 즉, 이후에

이 연구는 2021학년도 영남대학교 학술연구조성비에 의한 것임.
Park Sun-Yong: Dept. of Math. Edu., Yeungnam Univ. E-mail: polya@yu.ac.kr
Received on Nov. 5, 2022, revised on Dec. 16, 2022, accepted on Dec. 21, 2022.

1) 스티글러는 관찰 결과에 기초해서 원인의 확률을 찾아가는 추론에 대해 베이즈가 아니라 니콜라스 손더슨(Nicoholas Saunderson)이 최초로 생각해냈을 수 있다는 의견을 제기한 적이 있다 [10].

벌어졌던 '천편일률적인 사전 균등분포의 사용과 그에 따른 비판'에 대한 책임이 베이즈에게 없다는 것이다.

구체적으로, 스티글러 [11, pp. 282–283]는 1763년에 발표된 베이즈의 〈우연의 원리로 문제를 해결하는 에세이(An essay toward solving a problem in the doctrine of chances)〉에서의 핵심적인 균등분포는 모수 θ(확률변수 Θ)의 분포가 아니라 n회 독립시행에서 사건의 성공횟수 X의 분포라고 말하면서 다음과 같이 베이즈에게 면죄부를 부여하고자 하였다.

> 베이즈가 균등분포를 따른다고 한 것은 θ가 아니라 바로 X였던 것이다! 그의 논리는 다음과 같은 것이다. 당구대의 경우 우리가 말했듯이 θ가 균등분포를 따르는 것은 명확하다. 그 수학적 결과 가운데 하나는 다음과 같이 X의 주변분포도 균등분포가 된다는 것이다.[2)]

$$
\begin{aligned}
\mathrm{P}(X=p) =& \mathrm{P}(0<\theta<1, X=p)\\
=& \int_0^1 \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta\\
=& \frac{1}{n+1}, \text{ 결과는 } p\text{에 무관함.}
\end{aligned}
$$

스티글러의 통계사학계에서의 학문적 영향력에도 불구하고 이러한 주장은 큰 반향을 일으키지는 못한 듯하다. 베이즈가 모수 θ의 사전분포를 균등분포로 설정했던 이유를 설득력 있게 제시하지 못한 것처럼 보였기에, 스티글러의 주장은 베이즈가 짊어진 비판의 멍에를 벗겨내는 데에 실패한 것처럼 보인다. 과학 저널리스트인 맥그레인(S. B. McGrayne)은 이와 같은 상황을 다음과 같이 묘사한다.

> 오늘날 일부 역사학자들은 베이즈가 균등 확률 개념을 최초로, 이른바 '이전'의 공굴리기가 아닌 자기 자료(즉, 나중에 던진 공들의 위치)에 적용했을지도 모른다는 말을 함으로써 그의 죄를 면제해 주고자 한다. 하지만 이것 역시 어림짐작이다. 그리고 그 문제는 수많은 통계학자들에게 아무 소용이 없다. 왜냐하면 표면 어느 지점에서나 공이 설 수 있도록 수평을 잘 맞춘 탁자를 상정한다면, 이렇게 하나 저렇게 하나 수학적인 결과는 동일하기 때문이다 [8, p. 31].

사실, 스티글러의 '베이즈의 균등분포의 사용은 적절했다.'는 주장은 그의 학문적 커리어에서 유래를 찾아보기 힘들 정도로 굉장히 강하게 제기되었던 사항이다. 그의 논문이나 저술에서 거의 유일하게 느낌표(!)를 붙여가면서 "베이즈가 균등분포를 따른다고 한 것은 θ가

2) 아래의 수식 표현 중에서 $P(0<\theta<1, X=p)$의 원래 표현은 $P(0<\theta<1)\cap(X=p)$이다. 확률변수 Θ와 X가 구별되는 것을 고려하여 결합사건에 대한 확률의 형태로 나타낸 것인데, 표현의 일관성을 위해 이 논문 전체에서 결합사건의 형태로 표현했음을 밝힌다.

아니라 바로 X 였던 것이다!" 라고 하며 이런 조건에선 베이즈의 추론이 타당하다고 말한다. 그러면서, 베이즈에 대해 내려지는 '전혀 모르는 것을 지식인 것처럼 만든 장본인' 이란 비판에 대해 다음과 같이 강하게 방어하였다.

> 피셔를 비롯한 여러 사람들이 베이즈의 주장이라고 생각했던(내가 보기에는 죄다 틀린 생각이다) 것에 대해 가했던 반론들은 이 주장에 적용될 수 있는 것들이 아니다. ... 어쨌든 그의 논증 과정은 지나치게 이항분포라는 한 가지 경우에만 국한된 것이며, 결국 "아무것도 알지 못한다." 는 조건을 심각하게 제약하는 셈이 되는데 여기서 우리는 베이즈가 집어넣은 "전혀(absolutely)" 라는 표현이 그냥 경솔하게 쓰인 것이 아님을 알 수 있다 [11, p. 284].

앞서 말했지만, 이러한 스티글러의 항변은 통계학계에서 '베이즈가 모수 θ의 사전분포를 균등분포로 설정한 것' 에 대해 대체적으로 납득하지 못함으로써[3] 그 힘을 잃고 말았다고 하겠다. 맥그레인의 평가에서 드러나듯, 베이즈가 수평을 잘 맞춘 탁자를 제시한 것 자체가 '모수 θ의 사전분포는 균등분포이다.' 를 미리 선언한 것이기에, 베이즈의 치밀함에 대한 스티글러의 주장은 기껏해야 절반 정도 수긍할 수밖에 없는 듯하다.

물론, 스티글러 본인도 "(평평한) 당구대의 경우 우리가 말했듯이 θ가 균등분포를 따르는 것은 명확하다." 고 말함으로써 이러한 지적 사항에 대해 잘 주지하고 있음을 밝혔다. 즉, 그는 '베이즈가 모수 θ의 사전 균등분포를 처음부터 공준으로 제시하면서 논의를 전개했다.' 는 사실 자체에 대해선 부정할 수 없었던 것이다.

결국, 베이즈의 사전 균등분포와 관련된 모든 학문적 논쟁은 "베이즈가 왜 평평하지 않은 탁자를 제외하고서, 굳이 평평한 탁자를 상상하면서 사고실험을 진행하였는가?" 라는 물음으로 귀결된다고 할 수 있다. 즉, '베이즈가 모수 θ가 균등분포를 따른다고 한 것이 나름 타당했다.' 는 주장이 설득력을 지니기 위해서는 "베이즈가 왜 평평한 탁자를 도입했을까?" 에 대해 규명해야 하는 것이다.

본 연구에서 관심을 기울인 지점이 바로 이곳이다. 베이즈가 평평하지 않은 탁자가 아닌 평평한 탁자 위에서 사고실험을 진행했던 이유를 개연적으로 추측해내는 것이 이 연구의 내용이고 그것을 통해 베이즈가 사전 균등분포를 적합하게 도입했음을 밝히는 것이 이 연구의 목적이라 하겠다. 연구자는 베이즈가 어떤 의도나 생각을 가지고서 평평한 탁자를 선택했고 그 탁자 위에 공을 굴리는 사고실험을 진행했는지를 밝히고자 하는 것이다.

"현대의 확률 학도들이 베이즈의 관점을 이해하고자 노력한다면 많은 것을 얻을 수 있을

3) 스티글러는 오늘날의 관점에서 "모든 자연수 n에 대해 $f_X(x) = \frac{1}{n+1}$(단, $x = 0, 1, \cdots, n$)이다." 로부터 "확률변수 Θ가 균등분포를 따른다." 를 유도할 수 있다고 하였다. 그는 "베이즈가 알지 못했던 것은 미세한 수학적 논점에 따라 그 주장이 부분적으로 달라진다는 것이다 [11, p. 284]." 라고 말하며, 베이즈가 이러한 유도 절차를 제시하지 못하였다고 하였다.

것이라 믿는다 [9, p. 1075]."는 쉐퍼(G. Shafer)의 조언대로, 이 연구에서는 '베이즈가 평평한 탁자를 선택했던 이유'에 대해 이해하고자 하는 것이다.

이를 위해, 베이즈 자신이 사전 균등분포에 대해 어떤 생각을 했는지에 대해 우선 살펴보도록 하자.

2 베이즈의 사전 균등분포의 도입

베이즈가 사전분포로 균등분포를 도입하는 것에 대해 어떤 생각을 하고 있었을까? 우리는 〈우연의 원리로 문제를 해결하는 에세이〉의 서문에서 그 일부를 찾을 수 있다. 비록 프라이스가 원래의 서문을 수정하고 그의 의견까지 덧붙여서 작성한 형태의 서문이긴 하지만, 서문의 초반부에는 베이즈 본인의 사전 균등분포 도입에 대한 생각이 비교적 상세하게 담겨 있다.[4)]

> 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. ... 그는 그것(모수가 되는 확률)이 두 등차의 간격 사이에 놓이게 될 확률이 동일하다고 가정해야만 하는 규칙이 있는 듯하다[5)]는 말을 덧붙였다; 만약 이것(균등 사전분포)이 허락된다면 나머지 사항들은 확률의 원리에 따른 절차에 의해 일반적 방법으로 쉽게 계산할 수 있을 것이라 하였다. ... 그러나 이후에 그(베이즈)는 자신이 먼저 제시한 공준(균등 사전분포와 관련된 사항)이 아마 어떤 사람들에게는 합리적으로 보이지 않을 수도 있겠다고 생각하였다; 그래서 그는 논쟁의 여지가 있는 어떠한 사항도 수학적 추론으로 끌어들이지 않고서, 그가 생각하기에 문제에 대한 해법이 포함되는 방식으로 명제들을 다른 형태로 제시하고 그가 (사전 균등분포와 관련해) 왜 그렇게 생각했는지를 주석에 추가해 적기로 결정했다 [1, pp. 370–371].

베이즈는 사전분포로 균등분포를 도입하는 것에 대해 자신의 견해를 분명하게 밝혔다고 할 수 있다. 그는 모수인 확률[6)]이 어떤 범위에서 실현되는 확률을 알아내기 위해 사전 균등분포를 도입하는 것이 반드시 필요하지만, '어떤 사람들은 사전 균등분포의 도입 또는 그것을 제기하는 공준에 대해 비합리적으로 여길 수도 있겠다.'고 생각하였던 것이다.

그래서 장차 이루어질 비판을 예견하고 마치 그것에 대해 미리 방어하듯이, 베이즈는 사전

4) 이 인용문에서 괄호 안의 표현은 연구자가 보충한 것이다.

5) it appeared to him that the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees.

6) 베이지는 모수인 확률을 어떤 확정된 값이 아니라 확률변수로 다루었다.

균등분포에 대한 공준을 제시한 이후엔, 명제 8과 9[7]에서 '사전 균등분포'와 관련해 기존의 서술을 변형하였고 명제 9 이후에는 "자신이 왜 사전에 균등분포를 설정했는지"에 대해 독립된 주석을 통해 그 이유를 따로 제시하고자 했다. 기본적으로, 베이즈는 어떤 특수한 조건을 만족하는 상황에서는 사전 균등분포를 설정하는 것에 문제가 전혀 없다고 생각했던 것이다.

이제, 베이즈의 '사전 균등분포' 공준에 대한 고찰을 시작으로 하여, 그가 그러한 사전 균등분포를 제시했던 이유가 무엇이었는지에 대해 순차적으로 탐색해 보자 [11, pp. 385–387].

공준1. 평평하게 만들어진 정사각 테이블 또는 평면 ABCD가 있을 때, 공 O와 W 중 하나를 그 위에 던졌을 때[8] 그 공이 테이블 위의 어떤 곳에 반드시 위치하고, 테이블 위의 어떤 두 영역의 면적이 같으면 그 공이 두 영역에 놓일 확률은 서로 같을 것이라고 가정한다.

공준2. 나는 공 W를 첫 번째로 던질 것인데, 그 공이 놓인 지점을 통과하는 직선 중에서 AD에 평행한 직선을 그었을 때 그 직선은 CD, AB와 각각 s, o에서 만난다고 가정한다; 그리고 이후에 공 O를 $p+q$회 또는 n회 던질 것을 가정하고, 공을 한 번 던졌을 때 그 공이 AD와 os 사이에 놓이게 되는 것을 한 번의 시행에서 사건 M이 일어난 것이라 부르는 것을 가정한다. 이러한 공준의 가정 하에서 다음의 렘마가 제시된다.

렘마1. 점 o가 선분 AB위의 어떤 두 점 사이에 놓이게 될 확률은 전체 선분 AB에 대한 그 두 점 사이의 길이의 비이다. ... (중략) ...

렘마2. 공 W를 던져서 선분 os가 그어졌을 때, 한 번의 시행에서 사건 M의 확률은 선분 AB에 대한 Ao의 비이다.

7) 이 연구의 3장에 그 내용이 제시되어 있다.

8) 이후에 '공 O' 표현은 등장하지 않고 '공 W'만이 언급된다. 여기서, 공 W를 던지는 실험은 실제로는 이루어지지 않는다. 이는 가상의 선 os을 통해 모수(확률변수)를 도입하는 절차일 뿐이며, 공 W를 탁자 위에 던지는 결과(탁자 위에 놓은 위치)는 알 수 없는 것이다. 즉, 모수는 알 수 없는 대상인 것이다. 그에 따라, 공준2에서 공 O를 탁자 위에 n회 던지는 것도 가상의 사고실험이라 하겠다. 베이즈는 이러한 특징을 명제 9와 직후의 주석에서 밝힌다. 하지만 실제적 실험의 관점에서 베이즈에 대한 비판이 제기되곤 한다 [7, p. 549]. 이런 측면들을 고려하여, 이 연구에서는 베이즈의 에세이에서의 실험을 '사고실험'이라 부르기로 한다.

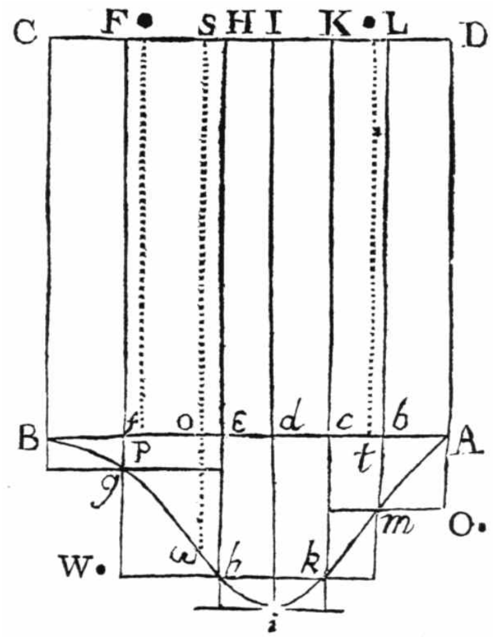

Figure 1. Bayes's Table1; 베이즈의 테이블1

공준1, 공준2, 렘마1, 렘마2에서 미지의 모수는 무엇일까? 이 질문을 제기한 이유는 모수와 그에 따른 확률변수가 마치 두 개인 것처럼 보이기 때문이다. 한 확률변수는, 미지의 모수가 'o의 위치' 또는 'AB 위에서 o의 상대적 위치'를 나타내는 'AB에 대한 Ao의 비인 $\frac{\mathrm{A}o}{\mathrm{AB}}$'인 것처럼 보인다. 다른 확률변수는, 미지의 모수가 '한 번의 시행에서 사건 M이 일어날 확률'인 것처럼 보인다. 사실, 이후에 살펴보겠지만 베이즈는 명제 8과 9에서 'o의 (상대적) 위치'와 '사건 M의 확률'을 모두 모수로 사용하였다.

이 두 모수를 조금 더 분명하게 구분해보도록 하자. 첫 번째는, 공 W를 던졌을 때의 공이 위치한 곳에서 AB와 CD에 내린 수선의 발을 각각 o, s라 할 때 o 또는 $\frac{\mathrm{A}o}{\mathrm{AB}}$을 모수로 삼는 상황이다. 두 번째는, 앞에서의 미지의 o의 위치 또는 $\frac{\mathrm{A}o}{\mathrm{AB}}$가 주어졌다는 가정하에서 어떤 사건이 일어날 확률을 새로운 모수로 정의하는 상황이다. 구체적으로, 공을 탁자 위에 한 번 던졌을 때 그 공이 AD와 os사이에 놓이게 되는 사건이 일어날 확률을 모수로 삼는 것인데, 그 공이 탁자 위에 놓인 곳에서 AB와 CD에 내린 수선의 발을 각각 t, t'라 할 때 "t가 o와 A사이에 위치하게 될 확률 또는 '$0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \frac{\mathrm{A}o}{\mathrm{AB}}$ 일 확률"이 새로운 모수가 된다.

최초의 공이 탁자 위에 놓이는 곳에 대한 가로 방향의 (상대적) 위치인 'o의 위치' 또는 '$\frac{\mathrm{A}o}{\mathrm{AB}}$'는 명백히 미지의 모수라 할 만하다. 그런데 베이즈가 정작 관심을 기울인 것은 이 모수에 대한 것이라 할 수 없다. 베이즈는 '어떤 사건이 일어날 확률', 정확히 말해 '어떤 사건이 일어날 확률이 어떤 범위에서 실현될 확률'을 알아내는 방법에 관심을 기울였다는 점에서 오히려 두

번째 모수가 그에게는 진정한 모수라 할 수 있다.

베이즈 에세이의 서문에서 살펴보았듯, 베이즈의 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 것〉이다. 탁자 위에서 공을 던지는 사고실험의 상황 속에서, 베이즈에게 있어 진정한 모수는 바로 '한 번의 시행에서 사건 M이 일어날 확률'이다.

이러한 논의를 바탕으로 하여, 베이즈가 왜 사전분포로 균등분포를 제시하였는지에 대해 탐색해 보도록 하자. 이를 위해, 잠시 동안 '탁자가 평평하다.'는 전제를 담고 있는 공준1과 렘마1이 가정되지 않은 것처럼 간주하자. 공준1과 렘마1을 가정하지 않은 상태에서, 진정한 모수인 '한 번의 시행에서 사건 M이 일어날 확률'을 다루는 렘마2는 무엇을 약속하는 것이라 할 수 있을까? 즉, "공 W를 던져서 선분 os가 그어졌을 때, 한 번의 시행에서 사건 M의 확률은 선분 AB에 대한 Ao의 비이다."는 어떤 의미를 지니는 것일까?

렘마2를 더 구체적으로 진술하면, "공 W를 던져서 (가상의) 선분 os가 그어졌을 때, 어떤 공을 한번 던졌을 때 그 공이 AD와 os사이에 놓이게 되는 사건 M이 일어날 확률은 선분 AB에 대한 Ao의 비이다."라 할 수 있다. 여기서 '공 W를 던져서 선분 os가 그어졌을 때'의 의미는 첫 번째 모수 o 또는 $\frac{\mathrm{A}o}{\mathrm{AB}}$가 이미 주어졌다는 것이다. 이런 점에서 '사건 M의 확률'은 일종의 조건부확률이다.

그런데 '어떤 공을 한번 던졌을 때 그 공이 AD와 os사이에 놓이게 되는 사건'은 '그 공이 탁자 위에 놓인 곳에서부터 AB와 CD에 내린 수선의 발을 각각 t, t'라 할 때 t가 o와 A사이에 위치하게 되는 사건'이라고 할 수 있다. 두 번째 모수인 '사건 M의 확률'은 't가 o와 A 사이에 위치하게 될 확률' 또는 '$0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \frac{\mathrm{A}o}{\mathrm{AB}}$일 확률'이다.

논의의 명확성을 위해, 첫 번째 모수인 'o의 (상대적) 위치'를 $\frac{\mathrm{A}o}{\mathrm{AB}} = \theta$로 나타내고, '$t$가 o와 A사이에 위치하게 되는 것'을 '$o < t < A$'와 같이 기호화하여, 두 번째 모수인 '사건 M의 확률'은 $\mathrm{P}(o < t < A)$, $\mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \frac{\mathrm{A}o}{\mathrm{AB}})$, $\mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \theta)$ 등으로 표시하자.

그러면, '렘마2'는 $\mathrm{P}(o < t < A) = \mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \theta) = \theta$를 규정한다고 할 수 있다. 이러한 약속은 사건 M의 확률을 $\frac{\mathrm{A}o}{\mathrm{AB}}(= \theta)$으로 정하는 것인데, 이는 '$o$의 (상대적) 위치'를 '사건 M의 확률'로 간주하는 것이라 할 수 있다. 그리고 $\mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \theta)$와 θ를 동일시함으로써, 기존의 모수 θ에 새로운 의미를 부여한다고 하겠다 [3].

물론, 이처럼 첫 번째 모수와 두 번째 모수를 일치시키는 작업 이후에도 여전히 모수는 θ이다. 그렇다면, θ의 의미를 'o의 (상대적) 위치'로부터 '한 번의 시행에서 사건 M이 일어날 확률'을 나타내는 것으로까지 그 의미를 확장 시켰을 때, 우리의 궁극적 관심사인 '모수 θ의 사전분포'에는 어떤 일이 발생하게 되는 것일까?

우리는 '탁자가 평평하다.'고 전제하지 않고서 논의를 전개하였기에, 첫 번째 모수만 제시된

상태에서는 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$의 확률분포(또는 확률밀도함수)인 $f_\Theta(\theta)$ (단, $0 \le \theta \le 1$)가 어떤 것인지에 대해 전혀 알 수 없다. 즉, 모수 θ에 대한 정보가 전혀 없으므로 그것에 대한 분포도 정할 수 없다. 하지만 모수 θ를 '사건 M의 확률'과 동일시하면 $f_\Theta(\theta)$를 특정 지을 수 있게 된다.

구체적으로, 사건 M과 그 확률의 의미에 의해 $\mathrm{P}(0 < \frac{At}{AB} < \theta) = \int_0^\theta f_\Theta(z)dz = F_\Theta(\theta)$가 성립하는데, 렘마2는 $\mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \theta) = \theta$를 규정하므로, 이 렘마2에 의해 $F_\Theta(\theta) = \theta$가 성립하게 된다. 그러면, $\frac{d(F_\Theta(\theta))}{d\theta} = f_\Theta(\theta) = 1$이 성립하게 된다. 즉, 모수 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$를 '사건 M의 확률'과 동일시하게 되면 '$f_\Theta(\theta) = 1$ (단, $0 \le \theta \le 1$)'와 같이 사전 균등분포가 유도된다.

그런데 공준1과 렘마1이 주어지지 않았을 때, 렘마2는 상대적 위치로서의 모수 θ가 주어진 상황에서 그 모수 θ의 의미를 '한 번의 시행에서 어떤 사건이 일어날 확률'로 정하는 것과 같다고 할 수 있다. 즉, '한 번의 시행에서 어떤 사건이 일어날 확률' 자체를 이제 모수 θ로 정하는 것이다.

이와 관련해, 앞서 제기했듯, 우리가 다루는 '사건 M의 확률'은 일종의 조건부확률임을 기억할 필요가 있다. X_1을 성공할 때와 실패할 때에 각각 1과 0을 할당하는 확률변수(베르누이 시행에 따른 확률변수)라고 하면, '사건 M의 확률'은 $f_{X_1|\Theta}(1|\theta)$으로 나타낼 수 있는 것이다. 그러면, 렘마2는 $f_{X_1|\Theta}(1|\theta) = \int_0^\theta f_\Theta(t)dt = F_\Theta(\theta)$일 때 $f_{X_1|\Theta}(1|\theta)$라고 약속하는 것과 같다. 물론, 이러한 관점에서도 '사건 M이 일어날 확률'인 $f_{X_1|\theta}(1|\theta)$을 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$로 정하는 것이 바로 렘마2라 하겠다.

렘마2는 '상대적 위치'를 나타내는 모수 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$와 '사건 M의 확률'을 나타내는 모수 $\mathrm{P}(o < t < A)$, $\mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \frac{\mathrm{A}o}{\mathrm{AB}})$, $\mathrm{P}(0 < \frac{\mathrm{A}t}{\mathrm{AB}} < \theta)$, $\int_0^\theta f_\Theta(z)dz$, $F_\Theta(\theta)$, $f_{X_1|\Theta}(1|\theta)$를 일치시키는 약속이다.

이러한 논의를 종합해, 이 단계에서 잠정적으로 내릴 수 있는 추측은 다음과 같다: 베이즈가 모수 'o의 (상대적) 위치'인 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$을 '사건 M이 일어날 확률'로 취급하고자 했기에 그의 사고실험에서는 '평평한 탁자 조건'이 도입되었던 것이다. 즉, o의 (상대적) 위치'인 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$에 대해 '$f_\Theta(\theta) = 1$'이 되기 위한 필요충분조건은 '탁자가 평평한 것'이기 때문에, 공리적 전개양식을 따르는 전통에 따라 문제의 상황을 우선적으로 설정하는 작업을 위해 '평평한 탁자' 조건을 밝히는 공준1을 맨 처음에 내세웠을 것으로 추측된다.

비록 베이즈는 '공준1⇒ 렘마1⇒ 렘마2'와 같은 연역적 전개 방식으로 그 결과를 제시했지만, 이러한 저술 이전에 '렘마2⇒ 공준1, 렘마1'와 같이 분석적으로 전개하는 것도 생각했을 것이고, '어떤 사건이 일어날 확률(사건 M이 일어날 확률)이 어떤 범위에서 실현될 확률'을 알아내는 방법을 탐구하는 과정에서 이미 '사전 균등분포를 도입해야 할 불가피성'을 인식했을 것으로 보인다.

그런데 우리는 이러한 추측에 대한 한 가지 단서를 렘마2 이후의 명제에서 찾을 수 있다.

명제 8과 9에서 'o의 (상대적) 위치'와 '사건 M이 일어날 확률'의 두 모수를 함께 사용할 뿐만 아니라 전자에서 후자로 전환하는 모습을 볼 수 있는 것이다. 즉, 베이즈의 에세이 서문에서 '문제에 대한 해법이 포함되는 방식으로 명제들을 다른 형태로 제시하는 모습' 중 한 가지가 'o의 (상대적) 위치'와 '사건 M이 일어날 확률'을 동시에 모수로 사용하다가 그 둘을 하나로 통합시키는 모습일 것이라 예측된다.[9)]

한편, 사건 M이 일어날 확률인 $f_{X_1|\Theta}(1|\theta)$, $\int_0^\theta f_\Theta(t)dt$, $F_\Theta(\theta)$를 θ와는 완전히 다르고 구별되는 독립된 모수로 간주한다면, 베이즈의 논의에서 '평평한 탁자' 조건은 요구되지 않는다고 볼 수 있다. 그러나 이후에 살펴보겠지만, 사건 M이 일어날 확률을 나타내는 '확률변수 $F(\Theta)$'에 대한 사전분포는 $f_\Theta(\theta)$의 형태에 관계없이 항상 균등분포를 이루게 된다. 즉, 평평한 탁자를 도입하지 않더라도 베이즈가 제시한 문제 상황에서는 본질적으로 사전 균등분포가 도입될 수밖에 없는 것이다.

탁자의 평평함에 관계없이, 만약 베이즈가 제시한 문제 상황에서 사건 M이 일어날 확률을 나타내는 '확률변수 $F(\Theta)$'에 대한 사전분포가 균등분포일 수밖에 없다면 그리고 베이즈가 이 사실을 인지하고 있었다면, 이것은 "베이즈가 모수 'o의 (상대적) 위치'인 $\frac{\mathrm{A}o}{\mathrm{AB}}(=\theta)$을 '사건 M이 일어날 확률'로 취급하고자 했기에 그의 사고실험에서 '평평한 탁자 조건'이 도입되었을 것이다."는 추측을 낳는다.

이와 관련해, 결국 이 연구에서의 핵심적 관건은 베이즈가 밝힌 사전 균등분포의 도입의 불가피성에 대한 것이라 하겠는데, 이에 대한 고찰을 시작하기 위해, 베이즈가 명제 8과 9 그리고 따름정리 등에서 사전 균등분포를 어떤 방식으로 다루었는지에 대해 다루기로 하자.

3 사전 균등분포의 도입에 따른 의문점을 해결하기 위한 방식

우리는 '(사전 균등분포의) 문제에 대한 해법이 포함되는 방식으로 명제들을 다른 형태로 제시하는 모습'이 무엇인지를 확인하기 위해, 그 전개되는 형태에 초점을 맞추어, 명제 8과 9에 대해 살펴보도록 하자(Figure 2 참조[10)]).

9) 물론, 우리는 베이즈가 명제 8과 9를 수정하여 다른 형태로 제시하기 이전에 그 명제들을 다룬 최초의 방식에 대해선 알 수가 없다.

10) Figure 2는 원문에는 없는데, 독자의 이해를 위해, 연구자가 원문의 렘마1의 Figure 1에 있는 그림에 색을 칠하여 제시한 것이다.

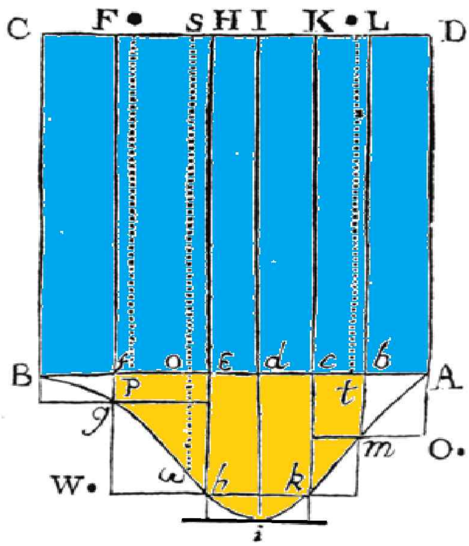

Figure 2. Bayes's Table2; 베이즈의 테이블2

명제 8

BA 위에 도형 B$ghikm$A를 세운다고 하면 그 도형의 성질은 이와 같은데, 밑변 BA를 Ao와 Bo처럼 두 부분으로 분할하고 분할점인 o에서 수선을 그어서 그 도형의 끝과 만나는 점을 ω이라 하고; y, θ, r은 각각 AB에 대한 $o\omega$, Ao, Bo의 비를 나타내는 것[11)]이고 E는 $(\theta+r)^{p+q}$의 이항 전개식에서 $\theta^p r^q$항의 계수이고 $y = \mathrm{E}\theta^p r^q$라고 한다.[12)] 가령 공 W를 던지기 이전에, 점 o가 선분 AB의 두 점인 f와 b의 사이에 있는 동시에 사건 M이 $p+q$번의 시행 중에서 p번 성공하고 q번 실패하게 되는 확률은 CA에 대한 $fghikmb$의 비와 같게 되는데, CA는 AB를 한변으로 하는 정사각형이며 도형 $fghikmb$은 선분 AB에서부터 그은 수선 fg, bm의 사이와 도형 B$ghikm$A이 겹치는 부분이다 [1, p. 388].

명제 8은 $\mathrm{P}(f < o < b, X = p) = \frac{fghikmb}{\square\mathrm{CA}}$을 나타낸다고 하겠는데, 베이즈는 소진법 (method of exhaustion)을 두 차례 사용하는 방식으로 이에 대해 증명하였다. $\mathrm{P}(f < o < b, X = p) = \frac{fghikmb}{\square\mathrm{CA}}$인 경우와 $\mathrm{P}(f < o < b, X = p) > \frac{fghikmb}{\square\mathrm{CA}}$인 경우로부터 모두

11) 베이즈의 에세이 원문 [1]에서는 "밑변 BA을 Ab와 Bb처럼 두 부분으로 분할하고 분할점인 b에서 수선을 그어서 그 도형의 끝과 만나는 점을 m이라 하고; y, θ, r은 각각 AB에 대한 bm, Ab, Bb의 비를 나타내는 것이고"와 같은 예시를 통해 도형의 성질이 기술되어 있다. 하지만 여기에서는 명제 8에 대한 일관적 진술을 도모하는 동시에 확률변수를 명확히 드러내기 위해 b 표현을 o 표현으로 대체하였다.

12) 에세이 원문에는 θ 표현이 아니라 x 표현으로 쓰여 있다. 서론에서부터 사용된 기호와의 일관성 문제를 고려해, 이 논문에서는 x를 θ로 바꾸었음을 밝힌다. 또한, 원문에서는 $(\theta+r)^{p+q}$이 아니라 $(a+b)^{p+q}$와 같은 표현이 사용되는데 '$(a+b)^{p+q}$에서의 b'와 '도해에서의 b'가 서로 같은 것이라는 오해를 줄 소지가 많아, 연구자가 표현을 바꾸었다.

모순을 유도하는데, 이러한 기호를 사용해 그 증명의 흐름을 간략하게 나타내면 다음과 같다. (단, 확률변수 X는 사건 M이 $p+q$번의 시행 중에서 성공하는 횟수를 나타내며, □와 ▯은 각각 정사각형과 직사각형을 나타내기로 한다.)

$\mathrm{P}(f<o<b, X=p) > \frac{fghikmb}{\square\mathrm{CA}}$라고 가정하자. 그러면, $fghikmb$보다 큰 도형 D가 존재하여, $\frac{\mathrm{D}}{\square\mathrm{CA}} = \mathrm{P}(f<o<b, X=p) > \frac{fghikmb}{\square\mathrm{CA}}$가 성립하게 된다. 즉, $\mathrm{D} > fghikmb$라 하겠다.

$fghikmb$를 다 덮으면서 그것에 근사하는 직사각형들이 있을 때, 충분히 세밀한 분할을 하게 되면 '직사각형들 (면적의) 합' 중에서 $fghikmb$보다 크지만 D보다 작은 것을 찾을 수 있다.

예를 들어, $\mathrm{P}(f<o<b, X=p) = \frac{\mathrm{D}}{\square\mathrm{CA}} > \frac{\text{▯}fh+\text{▯}ei+\text{▯}ci+\text{▯}bk}{\square\mathrm{CA}} > \frac{fghikmb}{\square\mathrm{CA}}$라 하자.

Ao가 Ae와 같을 때에, 렘마2에 의해, 어떤 한 번의 시행에서 사건 M이 일어날 확률은 AB에 대한 Ae의 비이고, 명제1의 따름정리[13)]에 따라, 그 사건이 일어나지 않을 확률은 AB에 대한 Be의 비이다.

밑변 fe의 경우에, Ao가 Ae와 같을 때에 최대 세로값 eh를 갖는다고 하자. 이 때, $\mathrm{P}(X=p\,|\,\mathrm{A}o=\mathrm{A}e) = \binom{p+q}{p}(\frac{\mathrm{A}e}{\mathrm{AB}})^p(\frac{\mathrm{B}e}{\mathrm{AB}})^q = y = \frac{eh}{\mathrm{AB}}$를 만족하는데, 이 eh가 밑변 fe위에 세워진 세로의 값 중에서 제일 크다고 하였으므로, $\mathrm{P}(X=p\,|\,f<o<e) < \frac{eh}{\mathrm{AB}}$이다.

그런데 렘마1에 의해 $\mathrm{P}(f<o<e) = \frac{eh}{\mathrm{AB}}$이다.

$\mathrm{P}(f<o<e, X=p) = \mathrm{P}(f<o<e)\cdot\mathrm{P}(X=p|f<o<e)$인 것을 이용하면, $\mathrm{P}(f<o<e, X=p) < \frac{ef}{\mathrm{AB}}\cdot\frac{eh}{\mathrm{AB}} = \frac{\text{▯}fh}{\square\mathrm{CA}}$이다. 즉, $\mathrm{P}(f<o<e, X=p) < \frac{\text{▯}fh}{\square\mathrm{CA}}$이다.

위와 유사한 방식[14)]에 의해, $\mathrm{P}(e<o<d, X=p) < \frac{\text{▯}ei}{\square\mathrm{CA}}$, $\mathrm{P}(d<o<c, X=p) < \frac{\text{▯}ci}{\square\mathrm{CA}}$, $\mathrm{P}(d<o<b, X=p) < \frac{\text{▯}bk}{\square\mathrm{CA}}$이 성립하게 된다. 일련의 논의를 종합해보면, $\mathrm{P}(f<o<e, X=p) + \mathrm{P}(e<o<d, X=p) + \mathrm{P}(d<o<c, X=p) + \mathrm{P}(d<o<b, X=p) = \mathrm{P}(f<o<b, X=p) < \frac{\text{▯}fh+\text{▯}ei+\text{▯}ci+\text{▯}bk}{\square\mathrm{CA}}$이 유도된다. 하지만 이러한 결과는 $\mathrm{P}(f<o<b, X=p) = \frac{\mathrm{D}}{\square\mathrm{CA}} > \frac{\text{▯}fh+\text{▯}ei+\text{▯}ci+\text{▯}bk}{\square\mathrm{CA}}$와 모순을 일으킨다.

$\mathrm{P}(f<o<b, X=p) < \frac{fghikmb}{\square\mathrm{CA}}$인 가정으로부터도 모순을 유사한 방식으로

13) 상호배타적인 사건의 확률 또는 여사건의 확률에 대한 성질을 지칭한다.

14) $\text{▯}fh, \text{▯}ei, \text{▯}ci, \text{▯}bk$의 각각의 밑변 fe, ed, ec, eb의 위에 세워진 세로값들은 $\mathrm{AB}\cdot y$인데, 각각의 밑변에서 그 세로값들 중에 최대값이 존재한다는 것을 이용하는 방식을 지칭한다.

유도할 수 있다. 따라서 $\mathrm{P}(f < o < b, X = p) = \frac{fghikmb}{\square\mathrm{CA}}$이 성립한다.

이러한 명제 8에 대한 증명이 전개되는 과정에서 두드러지게 나타나는 특징은 무엇일까? 그것은 모수의 변경 및 일치 작업이라고 할 수 있다. 최초의 모수는, 명제 8의 진술 자체의 "점 o가 선분 AB의 두 점인 f와 b의 사이에 있는" 표현에서 알 수 있듯이, '점 o의 (상대적) 위치' 라고 할 수 있다.

하지만 베이즈는 이 정리에 대한 증명 속에서, 렘마2를 활용하여 '한 번의 시행에서 사건 M이 일어날 확률' 을 새로운 모수로 삼는다고 하겠다. 즉 "Ao가 Ae와 같을 때에, 렘마2에 의해, 어떤 한 번의 시행에서 사건 M이 일어날 확률은 AB에 대한 Ae의 비이다 [1, p. 389]." 는 진술을 통해, 실제로 관심을 가지는 모수가 '사건 M이 일어날 확률' 인데 이것이 '점 o의 상대적 위치' 와 동일한 것으로 간주하겠다는 사항을 밝혔다고 하겠다.

한편, 베이즈는 명제 8의 따름정리에서 $\mathrm{P}(f < o < b, X = p) < \frac{fghikmb}{\square\mathrm{CA}}$에서 b, f에 각각 A, B를 대입하여, $\mathrm{P}(B < o < A, X = p) = \frac{\mathrm{B}ghikm\mathrm{A}}{\square\mathrm{CA}} = \frac{\mathrm{A}i\mathrm{B}}{\square\mathrm{CA}}$가 성립함과 함께, 이 성질이 주변확률분포에 대한 것임을 말하였다. 즉, $\mathrm{P}(\mathrm{B} < o < \mathrm{A}, X = p) = \mathrm{P}(X = p)$임을 다음과 같이 제시하였다.

> **따름정리.** 공 W를 던지기 이전에, 점 o가 A와 B 사이에 또는 선분 AB상의 어떤 곳에 있는 동시에 사건 M이 $p+q$번의 시행 중에서 p번 성공하고 q번 실패하게 되는 확률은 CA에 대한 AiB의 비와 같다. 그런데 점 o가 A와 B 사이에 어떤 곳에 있을 것이 확실하다. 따라서 공 W를 던지기 이전이라도, 사건 M이 $p+q$번의 시행 중에서 p번 성공하고 q번 성공하게 되는 확률은 CA에 대한 AiB의 비와 같다 [11, p. 391].

오늘날 관점에서 보면, 렘마2 및 명제 8과 그것의 따름정리는 다음을 나타낸다고 하겠다; X_1, X_2를 각각 베르누이 시행 및 n회의 독립시행에 따른 확률변수이고, $\frac{\mathrm{A}o}{\mathrm{AB}} = \theta$이고 확률변수 Θ의 확률밀도함수가 $f_\Theta(\theta)$라고 하자.

〈렘마 2〉: 일종의 조건부확률인 '사건 M의 확률' 을 $f_{X_1|\Theta}(1|\theta) = \int_0^\theta f_\Theta(t)dt = F_\Theta(\theta)$와 같이 정의한다고 할 때, 렘마 2는 $f_{X_1|\Theta}(1|\theta)$라고 추가적으로 새로이 약속하는 것과 같다.

〈명제 8〉: $\mathrm{P}(f < o < b, X = p) = \mathrm{P}(\frac{\mathrm{A}b}{\mathrm{AB}} < \frac{\mathrm{A}o}{\mathrm{AB}} < \frac{\mathrm{A}f}{\mathrm{AB}}, X = p)$이므로, '사건 M의 확률' 을 모수 θ와 동일한 것으로 놓았을 때,

$$\begin{aligned}\mathrm{P}(\theta_1 < \theta < \theta_2, X = p) &= \frac{fghikmb}{\square\mathrm{CA}} \\ &= \frac{(\overline{\mathrm{AB}})^2 \int_{\theta_1}^{\theta_2} \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta}{\square CA} \\ &= \int_{\theta_1}^{\theta_2} \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta\end{aligned}$$

이다. 즉, $\mathrm{P}(\theta_1 < \theta < \theta_2, X = p) = \int_{\theta_1}^{\theta_2} \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta$ 이다.

〈명제 8의 따름정리〉: $\mathrm{P}(0 < \theta < 1, X = p) = \mathrm{P}(X = p) = \int_0^1 \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta = \frac{\mathrm{A}i\mathrm{B}}{\square\mathrm{CA}}$ 이다.

그런데 명제 8의 따름정리 직후에, 베이즈는 명제 9와 그것의 따름정리를 제시함으로써, "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 것"이 어떤 모습인지를 드디어 제시한다. 몰론, 이것이 베이즈가 자신의 연구에서 밝히고자 했던 바로 그 방법이다.

명제 9

점 o의 위치가 발견되기 이전에 마치 사건 M이 $p+q$번의 시행 중에서 p번 성공하고 q번 실패하게 되는 것처럼 가정했을 때, 이것으로부터 나는 점 o가 가령 f와 b의 예처럼 선분 AB상의 어떤 두 점 사이에 놓여있게 되는 것, 결과적으로, 어떤 한 시행에서 사건 M의 확률이 AB에 대한 Ab의 비와 AB에 대한 Af의 비 사이에 있게 되는 것에 대해 추측한다: 내가 옳을 확률은 전체도형 AiB에 대한 이전에 기술한 도형[15)]의 비인데, 이 도형은 선분 AB 위의 점 f와 b에서 수선을 그었을 때 그 두 수선 사이와 도형 AiB가 겹치는 부분이다.

이처럼 두 연결되는 사건들이 있는데, 첫 번째는 점 o가 f와 b의 사이에 놓여있게 될 사건이다; 두 번째는 사건 M이 $p+q$번의 시행 중에서 p번 성공하고 q번 실패하게 되는 사건이다; 그리고 (명제 8의 따름정리에 의해) 두 번째 사건의 원래 확률은 CA에 대한 AiB의 비이고, (명제 8에 의해) 두 사건이 모두 일어날 확률은 CA에 대한 $fghikmb$의 비이고; 따라서 (명제 5에 의해)[16)] 두 번째 사건이 우선 발견되었다고 할 때, 이로부터 나는 첫 번째 사건도 역시 일어났다고 추측하는데, 내가 옳을 확률은 AiB에 대한 $fghikmb$의 비이다.

따름정리. 동일한 상황이 전제되었다고 할 때, 나는 사건 M의 확률이 0과 AB에 대한 Ab의 비의 사이에 놓이게 되는 것을 추측하는데, 내가 옳을 확률은 AiB에 대한 Abm의 비이다 [11, pp. 391–392].

여기서, 이 명제 9와 그것의 따름정리를 기호적으로 다음과 같이 나타낼 수 있을 것이다.

15) $fghikmb$

16) E_1, E_2를 각각 원인사건(논리적으로 먼저 일어나는 사건)과 결과사건(논리적으로 나중에 일어나는 사건)이라 할 때, $\mathrm{P}(\mathrm{E}_1|\mathrm{E}_2) = \frac{\mathrm{P}(\mathrm{E}_1,\mathrm{E}_2)}{\mathrm{P}(\mathrm{E}_2)}$ 임을 뜻한다.

〈명제 9〉[17]:

$$\begin{aligned} \mathrm{P}(f < o < b | X = p) =& \frac{\mathrm{P}(f < o < b, X = p)}{\mathrm{P}(X = p)} \quad \text{(명제 5에 의해)} \\ =& \frac{fghikmb}{\square\mathrm{CA}} / \frac{\mathrm{A}i\mathrm{B}}{\square\mathrm{CA}} \quad \text{(명제 8과 그 따름정리에 의해)} \\ =& \frac{fghikmb}{\mathrm{A}i\mathrm{B}} \end{aligned}$$

〈명제 9의 따름정리〉 $\mathrm{P}(0 < \frac{\mathrm{A}o}{\mathrm{AB}} < \frac{\mathrm{A}b}{\mathrm{AB}} | X = p) = \mathrm{P}(0 < \theta < \frac{\mathrm{A}b}{\mathrm{AB}} | X = p) = \frac{\mathrm{A}bm}{\mathrm{A}i\mathrm{B}}$이다. (렘마 2와 명제 9)

여기서 우리는 명제 9와 그것의 따름정리 사이에서 드러나는 한 가지 변화에 주목할 필요가 있다. 명제 9에서는 "첫 번째는 점 o가 f와 b의 사이에 놓여있게 될 사건이다."는 표현에서 알 수 있듯이 '점 o의 (상대적) 위치'를 모수로 간주하였지만, 명제 9의 따름정리에서는 '사건 M의 확률이 0과 AB에 대한 Ab의 비의 사이에 놓이게 되는 것'의 표현에서 드러나듯 '사건 M의 확률'을 모수로 삼는 것이다. 이러한 변화는, 명제 9에서의 진술에서 "첫 번째는 점 o가 f와 b의 사이에 놓여있게 될 사건이다."는 진술을 '첫번째는 사건 M의 확률이 AB에 대한 Af의 비와 AB에 대한 Ab의 비 사이에 놓이게 되는 사건이다."는 진술로 바꾸는 것에 해당한다고 하겠다.

그렇다면, 이와 같은 방식이 왜 사전균등분포의 도입을 정당화하는 것일까? 사실, 베이즈가 제시한 방식은 'o의 (상대적) 위치'인 $\frac{\mathrm{A}o}{\mathrm{AB}} (= \theta)$를 '어떤 특정한 사건이 일어날 확률'로 다루는 것에 불과한 듯이 보이기도 한다. 즉, 모수는 '어떤 특정한 사건이 일어날 확률'인데 그것을 'o의 (상대적) 위치'인 $\frac{\mathrm{A}o}{\mathrm{AB}} (= \theta)$로 대체해 놓은 것에 불과한데, 이것이 왜 사전 균등분포의 도입을 정당화하느냐는 의문이 제기된다고 하겠다.

그렇다면, 베이즈는 자신이 제시한 문제 상황에서 사전 균등분포의 도입이 왜 당연시되는지에 대해 무엇이라 주장하였을까? 이제, 베이즈가 제시한 주석에 대한 분석을 통해 이에 대해 고찰해보도록 하자.

4 베이즈가 직접 밝힌 '사전균등분포의 도입의 정당성'

이 시점에서의 관심사를 정리하면 다음과 같다: 어떤 사건의 확률을 나타내는 확률변수 T와 그것의 확률밀도함수 $f_T(t)$가 주어졌는데, 이 사건의 확률에 대해서는 전혀 알 수가 없기에 n회의 독립적인 이항시행에서 이 사건이 일어날 횟수 X의 분포는 균등분포라 할 수 있다. 즉,

17) 베이즈가 에세이에서 $\int_0^1 \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta = \frac{1}{n+1}$ (단, $p = 0, 1, 2, \cdots, n$) 임을 명확히 밝혔다는 사항을 고려하면, 명제 9는 현대적인 기호로 다음과 같이 나타낼 수 있다.
$$\mathrm{P}(f < o < b | X = p) = \frac{fghikmb}{\mathrm{A}i\mathrm{B}} = \frac{\square\mathrm{CA} \int_{\theta_2}^{\theta_2} \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta}{\square\mathrm{CA} \int_0^1 \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta} = (n+1) \int_{\theta_1}^{\theta_2} \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta$$

$P(X=x)=\frac{1}{1+n}$ (단, $x=0,1,2,\cdots,n$)이다. 그러면, 이때에 확률변수 T에 대한 확률밀도 함수 $f_T(t)$는 균등분포 $f_T(t)=1$ (단, $0<t<1$)이 되는가? $f_{X|T}(x|t)=\binom{n}{p}t^x(1-t)^{n-x}$ 이기에, 이 의문을 수학적으로 표현하면 다음과 같을 것이다.

$$f_X(x)=\int_0^1 f_{X,T}(x,t)dt=\int_0^1 f_T(t)\binom{n}{x}t^x(1-t)^{n-x}dt=\frac{1}{n+1}\text{일 때, } f_T(t)=1\text{인가?}$$

그런데 $\int_0^1\binom{n}{x}t^x(1-t)^{n-x}dt=\frac{1}{n+1}$은 미적분학의 규칙으로 이미 잘 알려진 사실이기에, 제기된 의문은 $x=0,1,2,\cdots,n$에 대해 $\int_0^1 f_T(t)\binom{n}{x}t^x(1-t)^{n-x}dt=\frac{1}{n+1}=\int_0^1\binom{n}{x}t^x(1-t)^{n-x}dt$일 때, $f_T(t)=1$이 성립하는지에 대한 것이다. 즉, 이산확률변수 X의 분포가 균등분포라 할 때 연속확률변수 T의 분포도 균등분포가 되느냐에 대한 것이다. 베이즈는 다음과 같은 주석을 통해 이에 대한 의견을 제시하였다.

주석

내가 M이라고 불렀던 것과 같은 사건에 대해 우리가 알고 있는 것이 일정한 시행 결과 사건이 일어난 횟수와 일어나지 않는 횟수뿐이라고 할 때, 앞서 나온 정리를 쓰면 당연히 그 횟수를 가지고 대체 사건의 확률이 얼마쯤일지 짐작할 수 있다. 또 거기서 언급한 면적을 계산하는 통상적인 방법을 쓰면 그 짐작한 값이 옳을 확률도 알 수 있다. 그런데 아래에서 살펴보듯이 그러한 규칙은 시행해보기 전에는 그 확률에 대해 아무것도 알지 못하는 어떤 사건에서도 사용할 수 있다; 가령 일정한 횟수의 시행에서 사건이 어떤 특정한 횟수만큼 일어날 것이라고 생각할 이유가 전혀 없는 경우가 그러하다. 결국 그 사건에 대해서는 사건의 확률이 애초에 고정되어 있지 않았고, 그래서 확률을 정할 때는 일정한 횟수의 시행 결과 그 사건이 굳이 어떤 특별한 횟수만큼 일어날 것이라고 할 이유는 없다고 생각해야 옳다.

그런데 M이라는 사건이 바로 그러한 경우다. 각 시행에서 사건의 확률을 정하게 되는 공 W를 굴리기 전에서는 (정리8에 의해) $p+q$회 또는 n회의 시행에서 p번의 성공과 q번의 실패가 일어날 확률은 AiB와 CA의 비가 되는데 p가 어떤 값이든 일단 $p+q$ 또는 n이 정해지고 나면 이 비는 같은 값이다. 그 값은 뉴턴의 미적분법인 유율법을 써서 AiB의 면적을 구하면 알 수 있다. 그리고 결과적으로 o의 위치를 알거나 n회의 시행에서 사건 M이 일어난 횟수를 알기 전에는 나는 그 사건이 다른 모든 횟수가 아닌 어떤 하나의 특정한 횟수만큼 일어날 것이라고 생각할 이유는 전혀 없다.

그러므로 앞으로 나는 사건 M에 관한 규칙은 당연히 그에 대해 실험을 하거나 관측하기 전에는 그 확률에 대해 아무것도 알지 못하는 어떤 사건과 관련해서도

역시 쓰일 수 있는 규칙이라고 간주할 것이다. 이제부터 나는 그런 사건을 미지의 사건(unknown event)이라고 부를 것이다 [11, p. 282 재인용] 18).

이러한 베이즈의 답변 속에는 $\mathrm{P}(X = x) = \frac{1}{n+1}$ 로부터 $f_T(t) = 1$을 유도하는 모습이 전혀 보이지 않는다. 사실, 이 증명은 적률생성함수의 유일성에 의해 증명할 수 있는 사항이다. 구체적으로, $\int_0^1 f_T(t)\binom{n}{x}t^x(1-t)^{n-x}dt = \frac{1}{n+1} = \int_0^1 \binom{n}{x}t^x(1-t)^{n-x}dt$에서 $x = n$을 대입하면, 어떠한 자연수 n에 대해서도 $\int_0^1 f_T(t)t^n dt = \frac{1}{n+1} = \int_0^1 t^n dt$이 성립하게 된다. 그런데 이것은, $K{\sim}U(0,1)$인 확률변수 K에 대해 $E(T^n) = E(K^n)$을 뜻하므로, 확률변수 T의 적률생성함수가 $K{\sim}U(0,1)$인 확률변수 K에 대한 적률생성함수와 서로 같다는 것을 함의한다. 그러면 적률생성함수의 유일성에 의해 $T{\sim}U(0,1)$이 성립하므로 우리가 원하는 결론에 이르게 된다.

하지만 베이즈가 줄곧 제기한 것은, 이러한 연역적 설명 또는 정당화가 아니라, 사건의 확률에 대한 정보가 전혀 없어 $\mathrm{P}(X = x) = \frac{1}{n+1}$인 상황에서는 탁자에서의 사건 M을 다루는 방식을 그대로 적용하면 된다는 것뿐이다. 베이즈는 "그런데 M이라는 사건이 바로 그러한 경우다."라는 선언과 함께 "앞으로 나는 사건 M에 관한 규칙은 당연히 그에 대해 실험을 하거나 관측하기 전에는 그 확률에 대해 아무것도 알지 못하는 어떤 사건과 관련해서도 역시 쓰일 수 있는 규칙이라고 간주할 것이다 [1, p. 282]"라고 반복해서 말할 뿐이다.

이와 같은 국면에서, 베이즈가 $\mathrm{P}(X = x) = \frac{1}{n+1}$인 상황에서는 탁자에서의 사건 M을 다루는 방식을 그대로 적용하면 된다고 왜 줄기차게 주장하는지에 대해 탐색해볼 필요가 있다. 물론, 이 고찰을 진행함에 있어 우리에겐 여전히 의문스러운 사항이 한 가지 있다. 베이즈는 평평한 탁자 위에서의 사고실험만을 보여주었다는 점에서, 그는 $f_T(t) = 1$로부터 $\mathrm{P}(X = x) = \frac{1}{n+1}$을 유도했을 뿐이라는 것이 바로 그것이다. 즉, 표면적으로 사건 M을 다루는 방식 안에는 $\mathrm{P}(X = x) = \frac{1}{n+1}$로부터 $f_T(t) = 1$을 유도한 것을 전혀 볼 수 없고, 우리가 원하는 것과 비교해 그 반대방향으로의 모습만 볼 수 있을 뿐이다.

이러한 측면을 고려하여, 우리는 탁자가 평평하다는 조건을 제거한 상태에서 '사건 M을 다루는 방식'의 특징에 대해 살펴보려고 한다. 즉, $f_T(t) = 1$을 가정하지 않은 상태에서 베이지의 논의가 어떻게 전개되는지에 대해 고찰하려고 한다.

정사각형 모양의 탁자 위에 가상의 세로선을 긋는데, 탁자는 평평하지 않을 수도 있다고 하자. 가상의 세로선과 정사각형 ABCD와 만나는 점을 o, s라고 하자. 즉, 가상의 세로선이 AB와 CD와 각각 만나는 점을 o, s라 하자. 물론, 가상의 세로선이 어디에 놓일지 알 수 없다고 하자. 이때, 가상의 세로선의 가로방향으로의 상대적 위치 $\frac{\mathrm{A}o}{\mathrm{AB}}$를 θ로 놓고서 확률변수로 다루도록 하자(단, $0 < \theta < 1$). 확률변수 Θ는 세로선 os의 정사각형 ABCD안에서의 가로방

18) 조재근의 역서 [11]에서 이 주석이 충실하게 번역되어 있어, 이 번역문을 재인용하였다.

향으로의 상대적 위치를 나타내는 확률변수라고 할 수 있는데, 이에 대응하는 확률밀도함수는 $f_\Theta(\theta)$ 라고 하자.

하나의 점처럼 간주할 수 있는 공 1개를 탁자 위에 던진다고 할 때, 이 공은 반드시 탁자 안에 놓이게 된다고 하자. 이때, 이 공이 탁자 위에서 가상의 선분 os 의 오른쪽에 놓이게 되는 사건을 사건 M이라 하자. 여기서, 사건 M이 일어날 확률은 가상의 세로선이 놓이는 상대적 위치 θ를 알 수 없는 것과 마찬가지로 전혀 알 수가 없다(Figure 2 참조).

그런데 사건 M이 일어날 확률은 상대적 위치 θ에 대한 함수 $\int_0^\theta f_\Theta(s)ds$ 로 표현할 수 있다. 즉, 사건 M이 일어날 확률은 $\mathrm{P}(\Theta \le \theta)$ 또는 $F_\Theta(\theta)$ 라 할 수 있다. 여기서, 사건 M이 일어날 확률을 확률변수 T 로 다루도록 하자. 즉, $T = F(\Theta)$ 로 놓고서, 확률변수 T 와 그에 대한 확률밀도함수 $f_T(t)$ 를 다루도록 하자. 이와 같은 상황에서, 다음 세 가지 질문에 대해 각각 순차적으로 점검해 보도록 하자.

> 첫째, $\mathrm{P}(X = x) = \frac{1}{n+1}$ (단, $x = 0, 1, 2, \cdots, n$) 이 성립하는가?
>
> 둘째, $f_T(t) = 1$ (단, $0 < t < 1$)이 성립하는가?
>
> 셋째, 미지의 사건의 확률을 나타내는 확률변수 T 는 항상 $T = F(\Theta)$ 로 나타낼 수 있는가? 즉, 어떤 (다른) 확률변수에 대한 누적분포의 형태로 나타낼 수 있는가?

우선, 첫 번째 의문부터 다루어보자. 구체적으로, 상대적 위치를 나타내는 확률변수 Θ 의 매개를 통해, 즉 $T = F(\Theta)$ 또는 $\Theta = F^{-1}(T)$ 와 같은 변수변환을 했을 때 n 회 독립시행에서 사건 M의 성공횟수 X 의 분포가 균등분포가 되는지를 살펴보자.

① $f_{X|\Theta}(x|\theta) = \binom{n}{x}[F_\Theta(\theta)]^x[1 - F_\Theta(\theta)]^{n-x}$

② $f_{\Theta,X}(\theta, x) = \binom{n}{x}[F_\Theta(\theta)]^x[1 - F_\Theta(\theta)]^{n-x} f_\Theta(\theta)$

③ $f_X(x) = \int_0^1 \binom{n}{x}[F_\Theta(\theta)]^x[1 - F_\Theta(\theta)]^{n-x} f_\Theta(\theta)d\theta = \frac{1}{n+1}$

왜냐하면 $F_\Theta(\theta) = t$ 라고 놓고 치환적분을 이용하면, $f_\Theta(\theta)d\theta = dt$ 이고

$f_X(x) = \int_0^1 \binom{n}{x}[F_\Theta(\theta)]^x[1 - F_\Theta(\theta)]^{n-x} f_\Theta(\theta)d\theta = \int_0^1 \binom{n}{x} t^x(1-t)^{n-x}dt = \frac{1}{n+1}$ 이기 때문이다.

④ 따라서 $\mathrm{P}(X = x) = \frac{1}{n+1}$ (단, $x = 0, 1, 2, \cdots, n$)이 성립하게 된다.

다음으로, $T = F(\Theta)$, 즉 $\Theta = F^{-1}(T)$ 와 같은 변수변환을 했을 때 $f_T(t) = 1$ $(0 < t < 1)$ 이 성립하는지에 대해 검토하자. 사실, 첫 번째 의문을 다루는 과정에서 이것이 성립한다는 것은 매우 쉽게 드러난다. 왜냐하면 연속확률변수 T, Θ 사이의 변환에서 $F_\Theta(\theta) = F_T(t)$ 가 항상 성립하는데, $F_\Theta(\theta) = t$ 와 같은 치환이 주어졌으므로 $F_\Theta(\theta) = F_T(t) = t$, 즉 $f_T(1) = 1$ 이 성립하기 때문이다.

구체적으로, $F_\Theta(\theta) = t$와 같은 치환이 주어졌기에 $F_\Theta(\theta) = F_T(t) = t$가 성립하게 되는데, $F_T(t) = t$를 미분하여, 확률변수 T에 대한 확률밀도함수 $f_T(t)$에 대해 $f_T(1) = 1$이 유도되는 것이다.

일반적으로 말해, 임의의 연속확률변수 Θ의 누적분포확률변수 $F(\Theta)$에 대해, 다음에서 알 수 있듯, $F(\Theta) \sim U(0,1)$은 항상 성립한다. 즉, $F(\Theta)$는 항상 균등분포를 이루게 된다.

A. $T = F(\Theta)$ 라고 하자.

B. $F_T(t) = F_\Theta(\theta)$가 성립하는데, 양변을 t에 대해 미분하자.

C. 역함수의 미분법을 이용하게 되면

$f_T(t) = \frac{dF_T(t)}{dt} = \frac{dF_\Theta(\theta)}{dt} = \frac{dF_\Theta(\theta)}{d\theta}\frac{d\theta}{dt} = f_\Theta(\theta)\frac{1}{f_\Theta(\theta)} = 1$이 성립하게 된다.

즉, $f_T(t) = 1(0 < t < 1)$이 성립하게 된다.

D. 따라서 $F(\Theta) \sim U(0,1)$이 성립한다.

놀랍게도, 이러한 2번째 질문에 대한 고찰 결과는 베이즈가 사전균등분포를 도입한 것에 전혀 문제가 없다는 것, 즉 평평한 탁자를 가정한 것에 아무런 문제점이 들어있지 않다는 것을 직접적으로 나타낸다고 할 수 있다. 잠시, 그 이유를 생각해보자.

확률변수 Θ와 그것의 확률밀도함수 $f_\Theta(\theta)$에 관계없이, 즉 탁자의 형태나 그 표면의 상태에 관계없이 $F(\Theta) \sim U(0,1)$이 성립한다. 이러한 사실을 베이즈의 사고실험 과정에 적용하게 되면, 어떤 공이 탁자 위에 놓이게 되는 상대적 위치를 나타내는 확률변수 Θ의 확률밀도함수 $f_\Theta(\theta)$에 관계없이 $F(\Theta) \sim U(0,1)$이 성립한다고 할 수 있다. 즉, $f_\Theta(\theta) = 1$인지의 여부에 상관없이, $F(\Theta) \sim U(0,1)$이 성립하는 것이다.

정리하면, 평평한 탁자를 도입하지 않은 상황($f_\Theta(\theta) \neq 1$)에서도 $F(\Theta) \sim U(0,1)$이 성립하기에, 사전 균등분포의 도입은 이루어질 수밖에 없다고 하겠다. 다시 강조하지만, $f_\Theta(\theta) = 1$로 놓는 것의 여부에 관계없이 $F(\Theta) \sim U(0,1)$이 성립한다. 즉, $f_\Theta(\theta) = 1$이라 가정해야만 $F(\Theta) \sim U(0,1)$이 유도되는 것이 전혀 아니고, 임의의 확률밀도함수 $f_\Theta(\theta)$에 대해 $F(\Theta) \sim U(0,1)$이 성립하는 것이다.

그런데 $f_\Theta(\theta) = 1$로 놓게 되면, $\theta = F_\Theta(\theta) = F_T(t) = t$가 되어, 사건 M 일어날 확률을 나타내는 확률변수 T와 확률변수 Θ가 동일한 것으로 취급할 수 있게 된다. $F_\Theta(\theta) = t$를 θ로 대체할 수 있기에 확률변수를 1개로 통일할 수 있는 것이다. 이런 점에서, 베이즈는 평평한 탁자를 도입함으로써 단지 2개의 확률변수를 1개로 통합적으로 다루었을 뿐인 것이다.

그렇다면, 미지의 사건의 확률을 나타내는 확률변수 T를 $T = F(\Theta)$처럼 어떤 확률변수에 대한 누적분포로 나타내는 것이 항상 가능하다는 것만 보인다면, 우리는 베이즈에게 완벽한 면죄부를 줄 수 있을 것이다. 물론, 이것이 위에서 제기한 세 번째 질문과 관련된다.

그렇지만, 베이즈에게는 우리의 면죄부가 전혀 필요 없다. 그리고 우리는 이미 세 번째 질문에 대한 답을 이미 다루었다고 할 수 있다. 왜냐하면 그는 미지의 사건의 확률을 나타내는 확률변수 T를 어떤 (다른) 확률변수 Θ에 대한 누적분포의 형태, 즉 $T = F(\Theta)$로 나타내는 '만능의 방법'을 우리에게 이미 구체적으로 제시해주었기 때문이다. 이미 살펴보았듯, 이때 도입된 확률변수가 바로 '탁자 위에 놓인 상대적 위치를 나타내는 확률변수'이다.

베이즈가 보여주고자 했던 것은 '미지의 사건이 일어날 확률을 누적분포로 나타내는 구체적 방법'이었던 것이다! 그는 $\mathrm{P}(X = x) = \frac{1}{n+1}$ (단, $x = 0, 1, 2, \cdots, n$)이 성립하는 상황에서의 사건을 미지의 사건이라 하였다. $f_X(x) = \int_0^1 f_{X,S}(x,t)dt = \int_0^1 f_S(s)\binom{n}{x}s^x(1-s)^{n-x}ds = \frac{1}{n+1}$이 성립할 때의 확률변수 S가, 베이즈에게 있어, 미지의 사건이 일어날 확률을 의미한다. 그런데 탁자 위에 놓이게 되는 공의 상대적 위치를 나타내는 확률변수 Θ를 도입하면, $f_X(x) = \int_0^1 \binom{n}{x}[F_\Theta(\theta)]^x[1 - F_\Theta(\theta)]^{n-x}f_\Theta(\theta)d\theta = \frac{1}{n+1}$이 성립하기에, 베이즈는 이때의 확률변수 $F(\Theta)$가 미지의 사건이 일어날 확률을 나타내는 '임의의' 확률변수임을 통찰했고 그것을 우리에게 알려주고자 했던 것이다.

$f_X(x) = \int_0^1 \binom{n}{x}[F_\Theta(\theta)]^x[1 - F_\Theta(\theta)]^{n-x}f_\Theta(\theta)d\theta = \frac{1}{n+1}$이 성립하므로, 아래의 ①$'$ ~③$'$에서와 같이, $F(\Theta)$와 X의 결합분포를 이용해 X의 주변확률분포를 구할 때에도 $f_X(x) = \frac{1}{n+1}$은 마찬가지로 성립한다. 즉, $F(\Theta) = T$는 미지의 사건이 일어날 확률을 나타내는 '임의의' 확률변수인 것이다. 물론, 미지의 사건이 일어날 확률을 누적분포로 나타내는 방식 자체의 특성에 의해, $f_T(t) = 1$이 성립하는 것은 너무나 분명하다! 왜냐하면 $F_\Theta(\theta) = t$와 같은 치환이 주어진 상태에서 $t = F_\Theta(\theta) = F_T(t)$로부터 $f_T(t) = 1$이 곧바로 유도되기 때문이다.

①$'$ $f_{X|T}(x|t) = \binom{n}{x}t^x(1-t)^{n-x}$

②$'$ $f_{\Theta,T}(\theta,t) = \binom{n}{x}t^x(1-t)^{n-x}f_T(t)$

③$'$ $f_X(x) = \int_0^1 \binom{n}{x}t^x(1-t)^{n-x}f_T(t)dt = \frac{1}{n+1}$

5 결론 : 베이즈를 위한 변명

베이즈가 자신의 에세이를 통해 규명하려고 했던 것은, 어떤 사건이 일어날 확률 T에 대해 정보가 전혀 없어서 n회의 독립적인 이항시행에서 그 사건이 일어날 횟수 X의 확률분포가 균등분포인 상황에서부터 출발하여, 실제로 n회의 독립적인 이항시행을 시행하여 $X = x$라는 결과가 나왔을 때 그 사건이 일어날 확률 T가 어떤 범위 사이에 있을 확률을 구하는 방식이다. 즉, 그는 $\mathrm{P}(t_1 < T < t_2|X = x)$를 구하는 방법이라 하겠다.

이러한 베이즈의 작업과 관련해, 이 논문에서 근본적으로 의문을 제기했던 사항은 $f_X(x) = \frac{1}{n+1}\,(x = 0, 1, 2, \cdots, n)$ 와 같은 조건으로부터 어떤 사건이 일어날 확률 T 의 분포가 균등분포가 되는 것을 유도할 수 있느냐의 문제라 할 수 있다. 즉, 베이즈가 T 의 사전분포가 $f_T(t) = 1(0 < t < 1)$ 인 이유에 대해 명확히 밝혔느냐에 대한 것이다.

표면적으로 본다면, 베이즈는 $f_T(t) = 1$ 로부터 $f_X(x) = \frac{1}{n+1}$ 만을 유도했으며 그 역에 대한 것은 보이지 않은 것처럼 보인다. 즉 $f_X(x) = \frac{1}{n+1}$ 로부터 $f_T(t) = 1$ 을 유도한 것처럼 보이지 않는다. 물론, 베이즈 에세이의 명제 8에 대한 분석을 통해 이러한 특징은 분명히 드러난다.

그런데 프라이스가 작성한 베이지 에세이의 서문에는 "그(베이즈)가 문제에 대한 해법이 포함되는 방식으로 명제들을 다른 형태로 제시하고 그가 왜 그렇게 생각했는지를 주석에 추가해 적기로 결정했다 [1, p. 371]."고 명시되어 있다.

사실, 베이즈 역시 사람들이 $f_T(t) = 1$ 와 같은 사전 균등분포를 제시하는 것에 대해 이상하게 여길 수 있다고 생각하였고, 그래서 그는 이 문제에 대한 해법이 포함되는 방식으로 명제들을 다루었고 주석에는 추가적인 설명을 제시하기로 했다고 할 수 있다.[19] 이와 관련해, 이 논문에서는 베이즈가 기술하고자 했던 '해법이 포함되는 방식'과 '왜 그렇게 생각했는지'를 분명히 드러낸 것이라 볼 수 있다.

베이즈가 공준, 렘마, 명제 등을 진술하면서 가장 두드러지게 드러나는 특징은 무엇일까? 그가 문제를 해결하기 위해 취한 접근방식의 특징 중 가장 현저한 것은 무엇이라고 할 수 있을까? 이 논문에서의 분석을 통해 일관되게 나타나는 사항은, 모수에 해당하는 확률변수가 2개이고 그 두 확률변수 사이의 변환 관계가 있다는 것이다.

구체적으로 말해, 첫 번째 확률변수는 어떤 공이 탁자 위에 놓이게 되는 가상의 상대적 위치를 나타내는 확률변수 Θ 이고 두 번째 확률변수는 그 공이 놓인 위치에 세로 수직선을 그었다고 할 때 어떤 공을 추가로 탁자 위에 던졌을 때 그 선의 오른쪽에 놓이게 되는 사건 M이 일어날 확률을 나타내는 확률변수 $F(\Theta)$ 라 하겠다. 그리고 두 확률변수는 $F_\Theta(\theta) = \int_0^\theta f_\Theta(s)ds$ 인 관계에 의해 변환된다고 하겠다.

그렇다면, 이러한 특징적 방식은 '사전균등분포의 도입이 정당한지'에 대한 의문에 대해 그 '해법이 포함되는 방식'이라 할 수 있을까? 이 연구에서의 결과는 '이 방식을 통해, 베이즈가 $f_T(t) = 1$ 일 수밖에 없는 이유를 이미 제시했다.'는 것을 지지한다고 할 수 있다.

베이즈가 제시한 방식은, 어떤 사건이 일어날 확률을 어떤 (다른) 확률변수 Θ 의 누적분포인 $F(\Theta)$ 로 나타내는 방식이다. 그렇다면, 어떤 점에서 이러한 방식이 우리의 의문에 대한 해법이 된다고 할 수 있을까?

19) 물론, 베이즈의 에세이가 사후 유작이기에 그것에 대한 진술이 온전히 기록되었는지에 대해선 의문의 여지가 있을 수 있다.

이 연구는, 이처럼 확률변수를 변환하는 방식이 $f_X(x) = \frac{1}{n+1}\,(x = 0, 1, 2, \cdots, n)$뿐만 아니라 $f_T(t) = 1\,(0 < t < 1)$이 성립한다는 것을 보장한다는 사실을 보여준다.

첫째, 상대적 위치를 나타내는 확률변수 Θ와 그것에 대한 확률밀도함수 $f_\Theta(\theta)$에 대해 $F_\Theta(\theta) = \int_0^\theta f_\Theta(s)ds$을 통해 새로운 확률변수 $F(\Theta) = T$를 도입한다고 할 때, $F(\Theta) = T$는 미지의 사건이 일어날 확률을 나타낸다고 할 수 있다. 즉, $f_X(x) = \frac{1}{n+1}\,(x = 0, 1, 2, \cdots, n)$이 아래에서와 같이 성립한다.

$$\begin{aligned} f_X(x) &= \int_0^1 \binom{n}{x} [F_\Theta(\theta)]^x [1 - F_\Theta(\theta)]^{n-x} f_\Theta(\theta) d\theta \\ &= \int_0^1 \binom{n}{x} t^x (1-t)^{n-x} dt \\ &= \frac{1}{n+1} \end{aligned}$$

둘째, $F(\Theta) = T$와 같은 도입에 의해 $f_X(x) = \int_0^1 \binom{n}{x} t^x (1-t)^{n-x} f_T(t)dt = \frac{1}{n+1}$이 성립할 뿐만 아니라, 누적분포를 이용하여 확률변수를 변환하는 방식 자체에 의해 $f_T(t) = 1\,(0 < t < 1)$일 수 밖에 없는 것이 너무나 분명하게 드러난다. 왜냐하면 두 연속확률변수 Θ, T 사이의 일대일 변환에 의해 $F_T(t) = F_\Theta(\theta)$이 성립하는데, 베이즈는 $F(\Theta) = T$와 같은 변환을 제시했기에 $F_T(t) = F_\Theta(\theta) = t$, 즉 $f_T(t) = 1\,(0 < t < 1)$을 제시한 것이나 마찬가지라 하겠다.

베이즈의 방식은, $f_T(t) = 1$로부터 $f_X(x) = \frac{1}{n+1}$을 먼저 유도하고 $f_X(x) = \frac{1}{n+1}$로부터 $f_T(t) = 1$을 따로 유도하는 방식이 아니었던 것이다. 베이즈는 미지의 사건을 $F(\Theta) = T$로 나타내는 구체적 방법을 제시함으로써, 그는 한꺼번에 $f_X(x) = \frac{1}{n+1}$와 $f_T(t) = 1$이 자명함을 보여주고자 했다고 하겠다. 즉, $F(\Theta) = T$와 같은 변환방식의 특징 자체에 의해, $f_X(x) = \frac{1}{n+1}$과 $f_T(t) = 1$이 유도된다.

연구자는, 베이즈가 주석을 통해 '왜 그렇게 생각하는지'를 설명할 때 말하고자 했던 것이 바로 이 '변환방식에 의한 자명함'이라 예측한다. 베이즈가 "그런데 사건 M이 바로 그러한 경우다. ... 그러므로 앞으로 나는 사건 M에 관한 규칙은 당연히 그에 대해 실험을 하거나 관측하기 전에는 그 확률에 대해 아무것도 알지 못하는 어떤 사건과 관련해서도 역시 쓰일 수 있는 규칙이라고 간주할 것이다. 이제부터 나는 그런 사건을 미지의 사건이라고 부를 것이다 [11, p. 282]"라고 말하면서, 사건 M의 확률을 다루는 방식 자체만을 강조하며 이 방식이 보편적으로 적용될 수 있는 규칙이라는 것을 언급하였을 뿐이다. 즉, 그는 사건 M의 확률을 다루는 방식을 언급하는 것만으로 모든 문제가 당연히 해소된다는 듯이 말했던 것이다.

이처럼, 베이즈가 사건 M의 확률을 다루는 방식 자체만을 언급하며 강조한 이유는 무엇일까? 연구자는, 사건 M의 확률을 다루는 방식이 어떤 (다른) 확률변수 Θ에 대한 누적분포인 $F(\Theta) = T$로 표현되는 방식이기에, $f_X(x) = \frac{1}{n+1}$을 쉽게 유도할 뿐만 아니라 $f_T(t) = 1$은 $F(\Theta) = T$와 같은 변환 자체에 의해 즉각적으로 설명될 수 있기 때문이라 예측한다. 숨겨진

비밀과 해법은 사건 M의 확률을 다루는 방식 속에 이미 들어있던 것이다.

이제, 서론에서 제기했던 최초의 의문 사항인 "왜 베이즈는 평평한 탁자라는 조건을 추가해서 사고실험을 하였을까?"에 대해 답해보도록 하자. 우리의 연구결과는, 사건 M이 일어날 확률을 누적분포인 $F(\Theta)$로 나타내는 방식에서는 Θ의 확률밀도함수 $f_\Theta(\theta)$에 관계없이 두 가지의 사전균등분포가 모두 유도된다. 즉, 평평한 탁자라는 조건이 없어도 그 방식 자체로부터 $f_X(x) = \frac{1}{n+1}$과 $f_T(t) = 1$이 성립하게 되는 것이다.

그렇다면, 베이즈는 구태여 평평한 탁자라는 조건 $f_\Theta(\theta) = 1\,(0 < \theta < 1)$을 왜 넣었을까? 연구자는 '확률변수를 줄여서 전체과정을 단순하려는 시도'였을 것이라 예측한다. 이미, $f_T(t) = 1$과 같이 '사건 M이 일어날 확률'에 대한 사전분포가 균등분포라는 것이 분명한 상황에서, $f_\Theta(\theta) = 1$을 제시함으로써 $t = F_T(t) = F_\Theta(\theta) = \theta$와 같이, 어떤 공이 탁자 위에 놓인 상대적 위치를 나타내는 확률변수 Θ와 사건 M이 일어날 확률 $F(\Theta)$를 동일한 것으로 취급함으로써 전체 과정에서 확률변수를 Θ 한 가지로 통일하려 했다고 하겠다 [3].

결론적으로 말해, 미지의 사건의 확률을 어떤 (다른) 확률변수에 대한 누적분포로 나타내는 방식이라는 '보석'에 있어, 탁자가 평평하다는 조건은 전혀 '흠'이 되지 않는다고 하겠다.

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. A. I. Dale, *A History of Inverse Probability from Thomas Bayes to Karl Pearson*, Springer, 1999.
3. A. W. F. Edwards, Commentary on the arguments of Thomas Bayes, *Scandinavian Journal of Statistics* 5 (1978), 116–118.
4. D. A. Gillies, Was Bayes a Bayesian?, *Historia Mathematica* 14 (1987), 325–346.
5. I. Hacking, *The Emergence of Probability*, Cambridge University Press, 1975.
6. A. Hald, *A History of Mathematical Statistics from 1750 to 1930*, John Wiley & 5. Sons, 1998.
7. V. J. Katz, *A History of Mathematics*, Harper Collins College Publishers, 1993.
8. S. B. McGrayne, *The Theory that Would not Die*, Yale University Press, 2011. (translated by Lee, G. S., 2013). 이경식(역), 불멸의 이론, 서울: 휴먼사이언스, 2013.
9. G. Shafer, Bayes's Two Arguments for the Rule of Conditioning, *The Annals of Statistics* 10 (1982), 1078–1089.
10. S. M. Stigler, Who discovered Bayes's Theorem?, *The American Statistician* 37 (1983), 290–296.
11. ________, *The History of Statistics*, Harvard University Press, 1986. (translated by Cho, J. G, 2002). 조재근(역), 통계학의 역사, 서울: 한길사, 2002.