대한수학교육학회지 수학교육학연구 제 19 권 제 2 호
Journal of Educational Research in Mathematics
Vol. 19, No. 2, 233-256. May 2009

조건부확률 개념의 교수학적 분석과 이해 분석[1)]

이 정 연* · 우 정 호**

이 연구는 조건부확률 개념에 대한 교수학적 분석을 시도하고, 학생들의 조건부확률 개념의 이해에 관하여 분석하였다. 수학적, 역사 발생적, 심리학적, 인식론적 관점의 분석을 통하여, 조건사건을 표본공간으로 하는 확률이라는 대상 개념, 사전확률이 사후확률로 변화하는 확률 수정의 과정 개념, 조건사건의 확인, 사건의 시간 순서와 조건관계의 구분, 인과관계와 조건관계의 구분, 가추적 사고의 이해가 조건부확률 이해의 핵심적인 요소임을 확인하였다. 또한, 고등학생과 대학생의 지필 검사와 면담을 통하여 학생들의 이해와 오개념에 대해 분석하였다. 이를 토대로 교육과정에의 시사점을 도출하였다.

Ⅰ. 서 론

확률 이론과 직관 사이의 불일치, 확률 개념의 복잡한 특성으로 인하여 학생들은 수학교과의 어느 영역보다도 확률 단원을 어려워하는 듯하다. 우리는 불확실한 사건에 대해서 종종 불확실성이 "…에 달려있다"는 대답을 하곤 한다. 이처럼 조건이 주어진 불확실성을 수량화한 조건부확률은 확률 추론의 바탕에 놓여있는 개념으로 확률을 이해하는데 핵심적인 역할을 하는 기본 개념이다. 또한, 조건부확률과 종속·독립개념은 확률을 다른 측도론과 구분지어 주는 개념으로서 확률론의 핵심 개념이므로, 조건부확률 개념에 대한 올바른 직관과 이해는 매우 중요하다.

그러나 우리나라의 수학교육 과정에서는 집합론과 조합론의 바탕에서 조건부확률 개념의 일부만을 다루며, 확률 계산의 하나로 간단히 다루어지고 있다. 이는 학생들이 조건부확률을 형식적인 확률 계산 도구로 파악하고, 그에 내재된 아이디어를 이해하지 못하는 결과를 가져올 수 있다. 따라서 이 연구에서는 다양한 측면을 갖는 조건부확률 개념이 무엇이며, 이를 학습한 학생들이 어떻게 이해하고 있는가를 분석한다.

Kapadia(1986)는 '확률의 교수현상학'을 연구하기 위해서는 확률에 대한 지식을 알고 있어야 하고, 확률의 역사를 알아야 하며, 교과서를 분석하고, 수학적인 확률 개념의 획득과정을 이해하기 위해 학생들의 학습 과정을 관찰해야 한다고 하였다. 우정호(2004)는 학교수학을 그대로 가르치면 본질을 감춘 '형식적인 닫힌' 지식의 교육이 되기 쉬우며, 철학적, 수학적, 인식론적, 역사-발생적, 심리학적, 언어학적, 실용

* 개포중학교, khskme@hanmail.net
** 서울대 명예교수, wjh@snu.ac.kr
1) 이 논문은 이정연의 석사학위논문의 일부를 요약 재구성한 것임

적, 교육학적 분석을 아우르는 교수학적 분석을 통해 학교수학을 '열어' 의미 있는 지식으로 가르쳐야 한다고 하였다. 따라서 이 연구에서는 먼저 조건부확률 개념에 대한 교수학적 분석을 시도하고, 이 개념에 대한 고등학생과 대학생의 이해 정도를 분석하여, 조건부확률 이해의 특징을 파악하고자 한다.

II. 교수학적 분석

1. 수학적 분석

조건부확률은 확률론에서 확률적 종속 관계를 설명하는데 필요하며, 일반적으로 다음과 같이 정의된다.

정의 [조건부확률] 주어진 확률공간 (Ω, $\mathcal{F}$, P)에서 E_1, $E_2 \in \mathcal{F}$ 이고 $P(E_2) > 0$ 이라 하자. 사건 E_2가 일어났다는 조건에서 E_1가 일어날 확률을 E_2에 대한 E_1의 조건부확률이라 하며, $P(E_1|E_2)$로 표시하고 $P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$로 정의한다(김해경, 2001: 51).

이와 같이 E_2를 조건사건으로 하는 조건부확률이란 실험결과 전체의 집합이 E_2로 국한되는, E_2가 새로운 표본공간이 되는 확률을 의미한다. 즉, 조건부확률은 조건사건이 새 표본공간이 되는 새로운 확률측도로서 정의된다.[2] 조건부확률은 실험의 결과에 관한 부분적인 정보가 있는 경우에 확률을 계산하는데 유용하며, 어떤 부분적인 정보가 없을 때에도 구하고자 하는 확률을 쉽게 알아낼 수 있는 수단이 된다는 점에서 양면적인 중요한 가치를 지닌다(박기호, 2001). 또한 조건부확률은 다음과 같이 확률의 세 공리를 만족하므로, 확률의 모든 성질을 갖게 된다.

사건 A와 사건 B가 표본공간 S의 사건이고 $P(A) > 0$일 때,

(1) $0 \le P(B|A) \le 1$

(2) $P(S|A) = 1$

(3) 사건 $B_i(i = 1, 2, 3, \cdots)$가 서로 배반인 사건이면, $P(\bigcup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$이다.

$P(B|A)$는 표본공간 S에서 축소된 공간 A에 관한 B의 상대 확률로서 정의된다.

그런데 사건 A의 확률을 구할 수는 없지만 적당한 사건 B에 대하여 $P(A|B)$와 $P(A|B^c)$를 구할 수 있는 경우가 있다. 이러한 경우 개념적으로 풍부하고 많은 응용성을 가지며, 복잡하게 얽힌 사건에 대한 확률을 계산하기 위하여 전확률 정리가 사용된다. 이것은 $P(A)$가 B가 일어난다는 조건과 일어나지 않는다는 조건하에 A의 확률의 평균치임을 시사한다[3]. 베이즈 정리는 조건부확률을 일반화한 것이다. 베이즈 정리에서 $P(E_j)$를 사건 E_j의 사전확률이라 하고, $P(E_j|F)$를 사건 E_j의 사후확률이라 한다. 베이즈 정리의 기본 초점은 사전확률이 나중에 얻어지는 경험적인 증거에 근거해서 수정될 수 있다는 것이다. 사전확률이란 대체로 의사 결정자에 의해서 특정 사상에 부여된 주관적 확률이다. 경험적인 증거나 과거 연구의 결과로 얻

---

2) 확률공간 (Ω, $\mathcal{F}$, P)에서 F를 $\mathcal{F}$에서 임의로 정해진 한 사건, $\mathcal{F}_F = F \cap \mathcal{F} = \{F \cap E \mid E \in \mathcal{F}\}$라 하자. 집합함수 $P^*: \mathcal{F}_F \to R^1$를 $P^*(F \cap E) = \frac{P(F \cap E)}{P(F)}$, $P(F) > 0$로 정의하면 $(F, \mathcal{F}_F, P^*)$는 확률공간이 된다.

3) $P(A) = P(A|B)P(B) + P(A|B^C)P(B^C)$

어지는 사후확률은 사전확률보다 의사 결정을 하는데 보다 더 나은 근거를 제공한다.

사건 A와 사건 B가 표본공간 Ω의 사건일 때, A의 발생에 관한 지식이 B의 발생 확률에 영향을 미치지 않으면 'B는 A에 독립'이다. 즉, $P(B|A)=P(B)$이면 B는 A에 독립이다. 한편, $P(B|A)=\frac{P(A\cap B)}{P(A)}$이므로 독립이라는 조건에 의하여 $P(A\cap B)=P(A)P(B)$를 얻게 된다. 똑같은 방법으로 $P(B)>0$일 때, $P(A|B)=P(A)$이면 'A는 B에 독립'이라고 말할 수 있다. 따라서 사건의 독립성은 표본공간의 모든 사건들의 집합 위에서 '대칭관계'를 이룬다. 독립과 종속은 확률론을 측도론의 일반적 이론과 구별짓게 하는 개념으로 확률론의 핵심 개념이다 (Dahl, 1994: 34)

2. 역사-발생적 분석

과거부터 일상생활에서 조건에 의해 불확실성이 영향을 받는 조건부확률 상황에 자주 직면할 수 있었지만, '조건부확률'이라는 용어와 표기법은 20세기에 들어와서야 등장했으며, 이에 내포된 역추론의 사고도 그리 오래된 아이디어가 아니다. 17세기 말, Jacob Bernoulli (1654-1705)는 20년 동안 사후확률의 계산문제를 연구하였는데, 이 결과는 Nicholas Bernoulli에 의하여 1713년에 『Ars Conjectandi』라는 저서로 출간되었다. 사후확률에 관련하여 우리는 Thomas Bayes의 1764년 논문에 주목할 필요가 있다. Bayes의 확률에 관한 이 논문에는 우리가 현재 조건부확률의 법칙이라고 부르는 정리와 증명이 포함되어 있었다. Bayes는 조건부확률을 다음과 같이 두 가지로 구분하였다.

> (ㄱ) 이어서 일어나는 두 사건 중에서 첫 번째 사건이 일어날 확률이 a/N이고, 두 사건이 함께 일어날 확률이 P/N이라면, 첫 번째 사건이 일어났다는 가정에서 두 번째 사건이 일어날 확률은 P/a이다.
>
> (ㄴ) 이어서 일어나는 두 사건 중에서 두 번째 사건이 일어날 확률이 b/N이고, 두 사건이 함께 일어날 확률은 P/N이며, 두 번째 사건이 일어난 것을 발견했다고 하자. 따라서 첫 번째 사건도 일어났다고 내가 추측할 때, 이 추측이 맞을 확률은 P/b이다(Bayes, 1764: 378-381).

(ㄱ)은 Bayes가 쓴 에세이의 세 번째 명제이고 (ㄴ)은 다섯 번째 명제에 해당한다. "이어서 일어나는" 두 사건 중 첫 번째 사건을 A라 하고 두 번째 사건을 B라고 한다면, 두 명제는 현대식 용어로 다음과 같이 서술될 수 있다.

(1) $P(B|A)=\frac{P(A\cap B)}{P(A)}$

(2) $P(A|B)=\frac{P(A\cap B)}{P(B)}$

(1)과 (2)는 하나의 일반적인 조건부확률 법칙의 두 가지 예시일 뿐이라고 생각할 수 있지만, Bayes는 (1)과 (2)를 사건의 시간 순서에 따라서 서로 다른 진술로 구분하고 각각을 증명하였으며, 이 개념을 베이즈 정리로 형식화하였다.

Hacking(1975)이 지적하였듯이, 이러한 관계를 Bayes가 최초로 발견했다고 볼 수는 없으며, 이전부터 알려진 사실을 Bayes가 최초로 형식화한 것으로 볼 수 있다. 역사적으로 (1)과 (2)의 서로 구분되는 두 아이디어가 조건부확률이라는 단일 개념으로 통합되는 과정을 거쳤음을 확인할 수 있다. 이보다 앞선 확률게임에 관한 Abraham De Moivre(1667-1754)의 저작에는 현재의 조건부확률 정리에 대한 최초의 계산이 포함되어 있다. De Moivre는 (1)의 아이디어, 즉 Bayes가 제시한 세 번째 명제의 질문에만 관심이 있었으며, '사건이 일어난 시간'을 매우 중요하게 다루었다. 어떤 사건의 확률이 그보

다 앞서 일어난 사건에 의해 어떻게 변할 수 있는가에 관한 질문만을 고려했던 것이다.

18세기 Simpson(1710-1761)은 관측되는 물체의 위치와 관측된 값과의 차이 즉, 오차에 주목함으로써 역확률의 아이디어를 출현시켰다. Simpson, Lambert, Lagrange 등이 독립적으로 연구한 오차의 확률분포 연구는 결과에 대한 원인을 추정할 수 있다는 '역추론'의 아이디어를 낳았고, 이러한 배경에서 1774년 Laplace는 '역확률'의 원리를 제시하게 되었다(Stigler, 1986). Laplace는 역확률의 원리에 대하여 다음과 같이 언급하였다.

> 하나의 사건이 n개의 서로 다른 원인에 의하여 발생되어지면, 사건이 주어졌을 때의 원인들의 확률은 각 원인이 주어졌을 때 사건의 확률과 같으며, 이 확률값은 그 원인이 주어졌을 때의 사건의 확률값을 각 원인 하에서의 사건의 확률값의 합으로 나눈 값과 같다(이용구, 1991: 4).

이 이론은 곧 원인의 확률이 사전에 균등하게 주어졌을 때의 '베이즈 정리'라고 볼 수 있다.[4] 즉, Laplace는 베이즈 정리의 특수한 경우를 제공한 것이다. Laplace의 역확률의 아이디어는 Bayes에 비해 한꺼번에 증명 없이 주어졌고, 동확률의 가정을 언급하지 않고 포함시켰다는 문제점이 있었으나, 조건부확률의 현대적 개념에 더 가까이 접근할 수 있는 계기를 만들어 주었다.

18세기에는 조건부확률을 계산하기 위해 여러 가지 공식들이 개발되었으며, 먼저 일어난 사건이 주어졌을 때 어떤 사건의 확률이 변한다는 아이디어뿐 아니라, 어떤 가설이 주어졌을 때의 사건의 확률 계산에 대해서도 수학적 확률이 적용되었다. 그러나 19세기까지도 Bayes의 (1), (2)의 두 아이디어는 단일 개념의 특별한 경우로 여겨지지 못하였고, 조건부확률의 표기법도 개발되지 못하였다(Shafer, 1982: 1087). Hausdorff는 19세기까지 서로 구분되었던 조건부확률의 두 아이디어를 '상대 확률(relative probability)'이라는 단일 개념으로 통합할 것을 주장하였고, 이는 상황을 바꾸어 주는 분수령이 되었다. 사건의 시간 순서에 따라 두 가지로 구분되었던 조건부확률 개념은 하나의 단일 개념으로 통합되어왔음을 알 수 있다.

3. 심리학적 분석

Tarr and Jones(1997)는 조건부확률과 독립에 관한 중학교 학생들의 사고를 평가하고 예측하기 위한 틀을 개발하였다. 이 틀은 Jones et al.(1997)의 연구를 바탕으로 개발한 것으로서 개념의 수준을 4수준(주관적 사고 수준, 전환적 사고 수준, 비형식적 양적 사고 수준, 수적인 사고 수준)으로 구분하였다.

이들에 따르면, 1수준의 학생들은 주관적 판단에 의존하며, 자신이 사건의 결과를 통제할 수 있다고 믿고, 확률 판단과 관련된 양적인 정보를 무시하는 경향이 있다. 주관적인 판단이나 개인적인 경험에 의존하는 경향과 잘못된 방식으로 규칙성을 부여하는 경향 때문에 독립성과 조건부확률 개념을 의미 있는 방식으로 이해하지 못한다. 2수준의 학생들은 조건부확

---

4) 즉, $A_1, A_2, \cdots, A_n$을 n개의 원인이라고 하고 E를 관심을 갖는 사건이라고 할 때, Laplace의 언급은 $\frac{P(A_i|E)}{P(A_j|E)} = \frac{P(E|A_i)}{P(E|A_j)}$ 또는 $P(A_i|E) = \frac{P(E|A_i)}{\sum P(E|A_j)}$를 의미한다. 이는 오늘날의 베이즈 정리 $P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum P(E|A_i)P(A_i)}$에서 모든 i에 대하여 $P(A_i) = \frac{1}{n}$을 적용한 것과 같다.

률의 판단에 양적인 정보를 적합하게 사용하지만 때때로 관련이 없는 특징에 의한 혼동을 겪는다. 3수준의 학생들은 조건부확률 판단에 영향을 주는 양적 정보의 역할을 인식하고, 정확한 수치적 확률값을 사용하지는 않지만 학생들은 상대 빈도와 비율을 사용해서 복원 또는 비복원 상황의 조건부확률을 판단할 수 있다. 표본공간의 구성을 이해하기 때문에 복원상황에서는 독립사건을 이해할 수 있지만, 때때로 연속된 독립 시행에서 한 사건이 계속 일어남을 관찰한 후에는 다시 대표성 전략으로 회귀하기도 한다. 4수준의 학생들은 확률 상황을 이해할 때 자연스럽게 사건에 수치적인 확률값을 할당하게 된다. 확률 판단에서 수의 역할을 정확히 이해하기 때문에 두 사건이 독립이냐 종속이냐를 판단할 때 표본공간을 고려해야 한다는 것을 인식한다. 비복원 상황에서 사건이 반드시 일어나기까지 최소한 몇 번을 추출해야 하는가를 결정할 수 있으며, 복원상황에서는 아무리 추출을 많이 하더라도 반드시 일어나는 사건은 존재할 수 없다는 것을 이해하게 된다.

Tarr and Jones(1997)의 조건부확률 개념 수준에 비추어 본다면, 조건부확률 이해의 핵심은 복원 상황과 비복원 상황을 구분하고, 비복원 상황에서 사건의 확률이 변한다는 것을 인식하는 데에 있다. 또한 표본공간의 구성에서 부분보다는 부분과 전체의 관련성을 파악해야 한다. 예를 들면, 비복원 추출 이후에도 확률을 구하려는 사건의 경우의 수가 변화하지 않은 경우에 전체 표본공간의 경우의 수는 조건사건에 의해서 수정된다는 것을 이해해야 하며, 비복원 추출 이후 모든 사건의 확률이 변화함을 인식해야 한다. 또한, 조건사건에 대한 새로운 표본공간을 구성할 수 있어야 하며 가능한 결과들을 빠짐없이 체계적으로 생성할 수 있는 능력이 필요하다. 대표성 전략이나 주관적 판단으로의 회귀 경향을 극복해야 하며, 조건부확률을 양화하고 수적인 확률로 표현할 수 있는 능력이 필요하다.

4. 인식론적 분석

대칭성과 이유 불충분의 원리의 가정에 의한 확률의 고전적 관점과 관찰에 의존하는 빈도적 관점, 확률을 순전한 논리적 진술로 보는 논리적 관점은 일반적으로 객관 확률에 해당한다. 반면, 개인적 확률 또는 주관적 확률은 인간이 어떤 대상에 대한 확신의 정도에 확률 해석의 기초를 둔다. 개인적 확률은 베이즈 정리와 베이즈적 추론에 스며들어 있는 철학적 기초가 되는 사고이다(Salsburg, 2003: 189). 베이즈적 논리는 개인의 마음속에 존재하는 '개인적인 사전확률'에서 출발한다. 개인이 현실을 관찰해 나가면서 혹은 실험을 해나가면서 데이터를 접하게 되면, 이 데이터를 이용하여 확률을 수정하고, '사후확률'을 만들어내는 것이다.

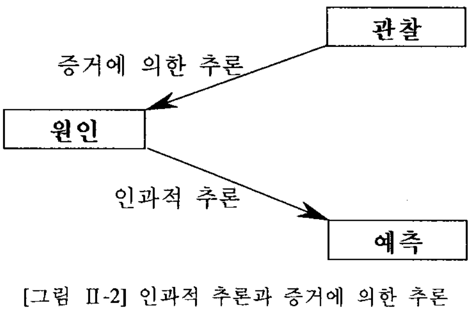
$\boxed{\text{사전확률}} \Rightarrow \boxed{\text{데이터}} \Rightarrow \boxed{\text{사후확률}}$

[그림 II-1] 베이즈적 논리

예를 들면, 모든 까마귀가 검은색인지를 확인하기 위해서, "모든 까마귀는 검은색이다"라는 진술이 사실일 가능성에 대해 사전확률 $P(A)$를 가지고 출발한다고 하자. 만약 까마귀에 대해서 아는 바가 전혀 없다면, 출발점에서는 50:50의 확률을 부여하게 된다. 그러나 까마귀에 관한 데이터를 축적해 가면서 새로이 발견한 까마귀가 검은색(B)이었다면, 사후확률 $P(A|B)$는 처음보다 높아질 것이다. 이때의 사후확률은 새로운 관찰이 이루어진다면 새로운 사후확률에 대한 사전확률이 된다. 고전 통계학자들은 베이즈 정리를 조건부확률 계산의 일반식으로만 생각하

는 반면, 베이즈적 통계학자들은 베이즈 정리를 하나의 방법론적 체계로 구축하고 사전적 정보를 주관·객관의 문제에 관계없이 적극 수용하여 추론이나 의사 결정에 사용한다. 베이즈적 관점에서 베이즈 정리는 사전 지식이 경험의 증가에 따라 개선되어 더욱 정확한 지식을 만들어 내도록 하는 것이다. 조건부확률은 객관적 관점에서 본다면 확률 계산 공식 그 이상도 이하도 아니지만, 주관적 관점을 통합한다면 지식이나 경험의 증가에 따라 새로운 지식을 만들어 낼 수 있으며 처음의 믿음을 계속해서 수정해 나가는 원리로 작용하게 된다.

우리는 현실 속에서 어떤 지식을 추론해 냄에 있어서 현실을 모델링하는 방법을 사용하곤 한다. 모델링을 인과적 모델링과 증거에 의한 모델링으로 나눌 수 있는데, 인과적 모델링이란 원인에서 결과가 유발되는 규칙에 대한 모델링이고 증거에 의한 모델링이란 결과에서 원인을 찾아가는 규칙에 대한 모델링이다(Poole, 2000: 154). 예를 들어, 감기가 걸렸을 때 나타나는 증상들을 규명하는 모델링은 인과적 모델링이고, 여러 증상들로부터 감기에 걸렸다는 결론을 얻어낼 수 있는 규칙에 대한 모델링은 증거에 의한 모델링이다. 이 두 가지 모델링에 바탕을 두고 이루어지는 추론에는 인과적 추론과 증거에 의한 추론이 있다. 두 가지 추론은 다음과 같이 나타낼 수 있다.



[그림 II-2] 인과적 추론과 증거에 의한 추론 (Poole, 2000: 155)

가추법은 인과적 모델링과 증거에 의한 추론을 연계시킨 의사결정 방법이라고 볼 수 있다. 만일 모델이 x라는 것으로부터 y를 이끌어낼 수 있는 규칙을 설명하고 있다면, 가추법은 그 모델을 이용하여 y로부터 x를 추론해 내는 방법이다. 이를테면, '비가 오면 땅이 젖는다'라는 명제는 '비가 온다'가 x이고 '땅이 젖는다'가 y라고 할 때 인과적 모델은 'x이면 y이다'라는 규칙이다. 이에 대해 가추법은 이 모델로부터 우리가 가지고 있는 지식 '땅이 젖어있다'는 y로부터 x를 추론해내는 것이다. 우리 교과서에서는 조건부확률의 공식으로 $P(A|B)=\frac{P(A\cap B)}{P(B)}$ 을 지도하는 있는데, 이 식을 변형하면 $P(A|B)=\frac{P(B|A)P(A)}{P(B)}$ 를 얻을 수 있다. 이 식은 베이즈 정리의 단순한 형태이며, 조건부확률의 식을 변형한 것뿐이므로 유도과정이나 식의 결과가 의미 없어 보이지만, 논리적 추론의 관점에서 본다면 중요한 의미가 있다. 베이즈 정리로부터 조건부확률에 내재되어 있는 가추적 사고의 관점을 읽어낼 수 있다.

$P(B|A)$에서 A를 가정, B를 결과라는 관점에서 본다면, 가정 A가 참인 상황에서 B라는 결과가 나타날 가능성을 의미하므로 $P(B|A)$는 인과적 모델을 의미한다고 볼 수 있으며, 가정 A가 결과 B를 얼마나 잘 예언하는지를 나타내는 지수라고 볼 수 있다. 그리고 $P(A)$는 아무런 자료도 주어지지 않은 상황에서 가정 A가 나타날 가능성을 뜻한다. 이는 가정 A가 우리에 의해 믿어지는 정도이다. 분모 $P(B)$는 조건부확률이 확률의 정의에 맞도록 확률값을 0과 1사이로 조정해주는 역할을 한다. 좌변 $P(A|B)$는 결과 B가 일어난다고 가정하였을 때 가정 A가 참이 될 가능성으로서, 결과로부터 가정을 추론해내는 증거에 의한 추론의 과정을 의미한다. 즉, $P(B|A)$는 인과적 모델에 관한 것

이고 $P(A|B)$는 증거에 의한 추론에 관련된 것으로서 이들을 연계하여 추론하는 것이 곧 가추법이라고 볼 수 있다. 따라서 베이즈 정리는 가추법적 사고의 전형적인 모습을 보여준다.

가추법은 어떤 결과가 주어졌을 때 그 원인을 추론해내는 방법으로 일상생활에서도 흔히 사용되는 자연스러운 추론방법일 뿐 아니라, 주어진 결과의 원인을 추리해내는 가장 효과적인 방법이기도 하다. 가추법은 학습자들에게 전혀 생소한 것이 아니며 학습자들의 일상생활 속에서 의식적으로나 무의식적으로 학습자의 현실과 깊이 관련되어 의미를 갖는 추론 형태이다. 베이즈 정리라는 길지 않은 등식 속에 인과적 모델과 증거에 의한 추론이라는 가추법의 가장 핵심적인 두 요소가 포함되고, 더불어 그들이 어떤 관계를 맺고 있는 지까지 규명되어 있는 것이다.

5. 조건부확률 개념의 교수학적 분석

교수학적 분석을 통해서 조건부확률 개념의 이해의 요소들을 다음과 같이 정리해 보았다.

첫째, 조건부확률 개념을 조건사건을 표본공간으로 하는 상대 확률로서 이해해야 한다. 조건부확률 $P(E|F)$는 사건 F가 일어난다는 가정 하에 사건 E가 일어나는 경우의 확률로서, 표본공간 전체의 집합이 F로 국한되는, F가 표본공간이 되는 시행에서의 확률을 의미한다. 수학적 분석에서 살펴보았듯이, $P(E|F)$는 확률의 세 가지 공리를 모두 만족하므로 조건부확률에서도 일반 확률의 모든 성질들이 성립하게 된다. 이러한 관점에서 조건부확률 $P(E|F)$는 $\frac{n(E\cap F)}{n(F)}$로 정의될 수 있다.

둘째, 조건부확률 $P(E|F)$를 $\frac{n(E\cap F)}{n(F)}$라는 '대상'으로서 뿐만 아니라, 사전확률 $P(E)$가 F라는 정보가 주어짐으로써 사후확률 $P(E|F)$로 변화한다는 '과정'으로서도 이해할 필요가 있다. Tarr and Jones의 조건부확률 개념 수준에 비추어 볼 때, 비복원 상황에서 하나의 사건이 조건으로 주어짐으로써 모든 사건의 확률이 변함을 인식하는 것이 수준 구분에서 중요하다. 확률상황에 주어진 정보는 선험적·이론적 정보와 후험적·경험적 정보로 구분된다. 조건부확률은 객관적 관점에서 본다면 확률 계산 공식일 뿐이지만, 주관적 관점에서 본다면 지식이나 경험의 증가에 따라 처음의 믿음을 계속 수정해 감으로써 새로운 지식을 만들어낼 수 있는 원리가 된다. 따라서 조건부확률을 객관적 '대상'이 아니라 주관 확률의 관점에서 변화하는 '과정'으로서 이해할 필요가 있다.

셋째, 조건부확률의 이해에는 조건이 되는 조건사건과 구하고자 하는 목적사건을 구분할 수 있는 능력이 필수적이다. Falk(1986)는 조건부확률 상황의 이해는 획득한 정보에 대하여 어떻게 생각하는가에 달려있으므로, 주어진 정보에 대해 주의를 기울여야 한다는 점을 지적하였다. 특히, 실생활의 조건부확률 상황은 무작위성이 명확히 언급되어 있지 않고, 조건사건이 관련이 없다고 생각되거나 조건사건을 확인하기 어려운 경우들이 많아서 조건부확률의 판단이 쉽지 않은 경우가 많다. 따라서 확률 판단에 영향을 주는 조건사건이 무엇인지를 확인하고 그 영향의 정도를 파악하는 능력이 필요하다.

넷째, 확률 추론은 일반적인 추론과 다르므로 '사건이 일어나는 시간'은 확률에 직접적인 영향이 없다는 것을 이해해야 한다. 역사적으로도 시간 순서에 따라 두 가지로 구분되었던 조건부확률 개념이 단일 개념으로 통합되어 왔음을 알 수 있었다. 조건부확률에서는 조건사건과 목적사건이 동시에 일어날 수도 있고, 조

건사건이 목적사건 보다 먼저 또는 나중에 일어날 수도 있다. 따라서 조건사건이 나중에 일어나지만 사전확률을 변화시키는 사례들을 다룸으로써 '사건이 일어나는 시간'과 조건사건은 관련이 없다는 것을 이해할 수 있어야 한다.

다섯째, 조건부확률의 조건관계를 인과관계와 구분할 수 있어야 한다. 조건관계는 확률에 영향을 미치는 정보를 제공하는 사건들 간의 관계인 반면, 인과관계는 두 사건 사이에 원인과 결과의 관계가 있다는 것이다. 인식론적 분석에서 알 수 있듯이 A를 가정, B를 결과라고 보는 관점에서 $P(B|A)$는 인과적 모델링을 의미하지만, $P(A|B)$는 증거에 의한 모델링을 의미한다. 실지로 A는 B를 초래하고 그 반대는 성립하지 않는 한 방향의 인과관계가 주어진 경우에도, 두 방향의 조건부확률이 존재하는 것이다. 또한, 두 사건이 서로의 확률에 영향을 주는 조건관계라고 할지라도 인과적 관련은 없을 수 있다. 인과관계와 구분해서 조건관계라는 것이 존재한다는 것을 인식하고, 결과가 조건이 되는 진단적 상황을 이해할 필요가 있다.

여섯째, 인식론적 분석에서 살펴보았듯이 조건부확률과 베이즈 정리에 내포된 가추적 사고를 이해해야 한다. 조건부확률과 베이즈 정리는 원인에서 결과로의 모델을 역으로 사용해서 결과에서 원인을 추론하는 대표적인 예이므로, 거꾸로 사고하는 가역적 사고가 요구된다.

Ⅲ. 학생들의 이해 분석

1. 연구 방법 및 절차

이 연구의 목적은 조건부확률에 대한 학생들의 이해를 분석하는데 있다. 이를 위하여 학교에서 조건부확률 단원을 배운 학생들을 잘 대표할 수 있으며 전형적이라고 생각되는 학생들

<표 Ⅲ-1> 조사문항 분류 및 출처

문항		주요 개념	출 처
조건부 확률 오개념 조사	1	조건사건을 확인하기 어려움	Falk(1989)의 연구
	2	사건의 시간 순서에 기인한 어려움	Shaughnessy(1992)의 연구
	3	독립성의 이해의 어려움	Shaughnessy(1981)의 Konold et al.(1993)의 연구
	4	사건의 시간 순서에 기인한 어려움	제 작
	5	조건사건을 확인하기 어려움	Falk(1989)의 연구
	6	사건의 시간 순서에 기인한 어려움	Carnell(1997)의 연구
	7, 8	조건관계와 인과관계의 혼동에 기인한 어려움	Pollatsek et al.(1987)의 연구
조건부 확률 이해 조사	9	조건사건의 이해	제 작
	10	복원상황과 비복원 상황의 구분	제 작
	11	조건부확률 공식의 적용	최용준 · 신현성(2002)
	12	조건에 의한 확률의 변화 (조건부확률의 비교)	MiC 교과서 Great Expectations
	13	곱의 법칙과 합의 법칙(전확률 정리)	제 작
	14	독립사건과 종속사건의 구분	임재훈 외(2002)
	15	베이즈 정리 (조건부확률 공식 이해 및 적용)	제 작
	16	독립사건과 종속사건의 구분	제 작

을 선정하는 유의 표집 또는 판단 표집 방법을 사용하였다. 조사 대상 학생들은 지역 간의 사회경제적 수준이 비슷한 두 지역에서 선정하였으며, 서울 지역 고등학교 1개와 포항 지역 고등학교 2개를 조사 대상 학교로 섭외하였다. 대학생은 서울 소재 대학교 수학·과학교육 전공 66명의 학생들이 조사에 참여하였다. 오개념을 가진 학생들과 공식을 사용하는 표준적인 방법으로 해결한 학생들을 포함하여 전형적이고 대표적인 반응을 보인 고등학생 4명과 대학생 5명을 선정하여 면담하였다.

이 연구에 사용한 조사 문항은 선행연구에서 사용한 문항을 분석 종합하여 8개의 조건부확률 오개념 조사문항을 선정하였고, 조건부확률 개념 분석과 고등학교 교과서에서 다루어지는 문제를 바탕으로 조건부확률 개념의 이해를 알아보기 위한 문항 8개를 선정하였다. 각 문항은 조건부확률과 종속·독립 개념에 관한 잘못된 인식의 경향과 이해의 정도를 드러낼 수 있도록 구성하였고, 지필검사의 취약성을 보완하기 위해 답을 선택한 이유와 풀이 과정을 반드시 기록하도록 하였다. 문제의 순서에 따른 결과의 차이를 배제하기 위해 홀수형과 짝수형으로 구분하였다. 지필검사 문항은 <부록 1>과 같다.

조사도구의 신뢰도 및 타당도를 확인하고, 적정 소요 시간과 문제 수준의 적절성, 문제 진술의 적절성 등을 알아보기 위하여 예비검사를 실시하였다. 학교에서 조건부확률 개념을 학습한 경험이 있는 서울 소재 고등학교 2학년 학생 13명에게 설문조사 및 면담을 실시하였으며, 이해에 어려움이 있는 것으로 드러난 문항을 수정 보완하였다.

검사의 신뢰성을 높이기 위하여 답안 작성 요령에 대해서 설명을 하고 해결 과정을 자세히 작성하도록 하였고, 조사에 소요되는 시간은 대략 50분~70분으로 하였다. 면담은 한 명 당 20~30분 동안 실시하였다. 분석을 위해 면담 내용을 모두 녹음하고 녹취하였다.

고등학생 대상 질문지 조사의 짝수형에는 40명, 홀수형에는 35명이 참여하였는데 이 중에서 60%이상 응답을 하지 않은 7명의 학생을 제외하고 68명(짝수형 37명, 홀수형 31명)에 대해서만 분석하였다. 확률 문제의 특성 상 해결 과정이 몇 종류로 구분되기 때문에 설문조사 자료에 나타난 학생들의 해결 과정을 몇 종류로 구분할 수 있었다. 반응을 문제 해결 패턴에 따라 분류하고 조직하였다.

<표 Ⅲ-3> 오개념 조사문항 분류

문제 분류	문항
조건사건에 대한 오개념	1, 5
사건의 시간 순서에 대한 오개념	2, 4, 6
조건관계와 인과관계에 대한 오개념	7, 8
독립성에 대한 오개념	3

2. 연구 결과

가. 조건부확률에 대한 오개념 분석

조사와 면담 결과를 통해 확인한 조건부확률

<표 Ⅲ-2> 오개념 조사의 문항 별 정답률

문항	1	2(1)	2(2)	3(1)	3(2)	3(3)	4(1)	4(2)	4(3)	5	6	7	8
정답률(%)	7.4	64.7	27.9	83.8	76.5	83.8	89.7	63.2	19.1	0	22.1	42.6	4.4
정답자 수	5	44	19	57	52	57	61	43	13	0	15	29	3
오답자 수	59	24	39	10	15	11	7	25	48	59	44	33	51
무응답	4	0	10	1	1	0	0	0	7	9	9	6	14
전체	68	68	68	68	68	68	68	68	68	68	68	68	68

에 대한 오개념은 크게 4가지로 분류할 수 있다. 분석에서 R은 연구자, S1-S4는 면담에 참여한 고등학생, S5-S9는 면담에 참여한 대학생이며, H는 고등학생의 지필답안이다.

1) 조건사건에 대한 오개념

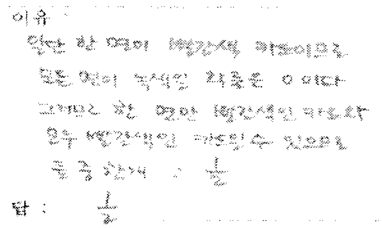
<표 Ⅲ-4> 1번과 5번 문항에 대한 반응 유형의 분포

문항	정답	오답 유형					
1	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	기타	무응답	전체
	5 (7.4)	26 (38.2)	16 (23.5)	6 (8.8)	11 (16.2)	4 (5.9)	68 100%
5	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	기타	무응답	전체
	0 (0.0)	44 (64.7)	5 (7.4)	8 (11.8)	2 (2.9)	9 (13.2)	68 100%

1번 문항의 경우, 세 장의 카드에는 여섯 개의 면이 있고, 여섯 개의 면은 각각 테이블 위에 놓일 확률이 동일하다. 표본공간이 $\{r_1/r_2,\ r_2/r_1,\ g_1/g_2,\ g_2/g_1,\ r/g,\ g/r\}$이므로 구하는 확률은 2/3이다.5)

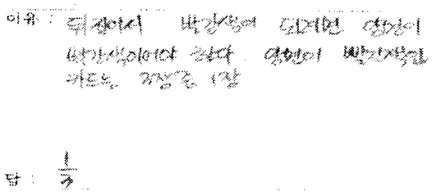
38.2%의 학생들은 이미 빨간색 면이 보였기 때문에 양면이 모두 녹색인 카드를 제외하고, 카드를 뒤집었을 때 빨간색 아니면 녹색이기 때문에 확률을 1/2이라고 추론하였다. '무작위로 선택한 카드의 무작위로 선택한 한 면이 빨간색'이라는 정보를 고려하지 않고, '이미 빨간색 면이 보였기 때문에 양면이 모두 녹색인 카드는 제외된다'는 것에 주목했다. 결국 '무작위성' 조건을 무시함으로써 뒤집었을 때 빨간색이 나올 확률이 더 크다는 것을 인식하지 못했으며, 이러한 오류는 대학생에서도 나타났다. 이들은 표본공간을 {red/red, red/green, green/green}으로 파악했음을 알 수 있다.

한편, 1/3이라고 한 23.5%의 학생들은 '테이블에 놓인 카드가 양면이 모두 빨간색이라는 것을 알기 때문에 양면이 모두 빨간색인 카드를 선택할 확률과 같다'고 생각했고, 1/4이라 답한 8.8%의 학생들은 결합사건의 확률을 구했다.

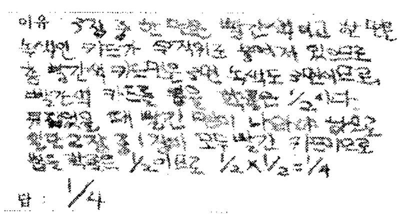


[그림 Ⅲ-1] 1번 문항에 대한 학생의 답안 (H6)

S9: 그래도 무작위라고 그랬어도, 세 개 중에 여하튼 하나를 뽑은 거잖아요? 뽑았더니 빨간색이라고 했으니까 이미 빨간색이니까, 카드 두 개로 좁혀지고. 둘 중에 뒤집었을 때 빨간색 아니면 녹색이니까 2분의 1이라고 했어요.



[그림 Ⅲ-2] 1번 문항에 대한 학생의 답안 (H1)



[그림 Ⅲ-3] 1번 문항에 대한 학생 답안 (H38)

$\{b_1/b_2,\ g_1/g_2,\ b/g,\ b_2/b_1,\ g_2/g_1,\ g/b\}$를 표본공간으로 볼 수 있는 5번 문항은 '무작위로 검사한 아이가 남자 아이로 판정된 사건'

5) 카드의 윗면을 첫 번째 자리의 문자로, 뒷면을 두 번째 자리의 문자로 짝을 지어 나타내어 표본공간 S를 나타낸 것이다. 예컨대 r/g는 테이블에 놓인 카드의 윗면이 빨간색이고 그 뒷면이 녹색인 카드이다.

이 조건사건이므로 검사한 아이가 남자 아이인 경우인 b_1/b_2, b_2/b_1, b/g를 고려한다면 확률은 2/3가 된다. 64.7%의 학생들이 '한명이 이미 남자 아이라고 했기 때문에 다른 한명의 성별만 결정하면 된다'고 생각해서 1/2이라고 판단하였다. 대부분 '태아의 세포에서 무작위로 선택한 결과가 남자 아이'라는 사건 대신에 '모두 여자 아이인 경우가 제외된다'는 것을 조건사건으로 생각하고 있었다.

한편, 11.8%의 학생들은 아이의 '순서'를 생각해서 표본공간을 {남/남, 남/여, 여/남, 여/여}라고 생각했고 조건을 무시하고 결합사건의 확률을 구하는 오류를 범했다. 표본공간을 {남/남, 여/여, 남/여}라고 생각한 학생들은 세 가지 가능한 경우 중에서 두 명이 모두 남자일 경우는 한 가지이므로 1/3이라고 답했다.

S3: 한 명이 남자라는 걸 알았잖아요. 그런데 남자라고 해서 나머지 조건에 영향을 미치지가 않는 것 같아요. 한 명이 남자라는 건 다른 한 명이 남잔지 여잔지 그거랑 상관이 없잖아요. 그러니깐 그건 조건이 안 되고. 그냥 여자랑 (남자랑) 태어날 가능성이 같으니까, 반반이니까 2분의 1이죠. 어차피 영향을 미치지 않으니까.

R : 그런데 여기서 테스트에서 세포를 무작위로 선택했다는 조건이 있지 않아요?

S5: (중략) 그건 상관이 없죠. 어차피 나머지 한 명이 문제이기 때문에. 1/2.

S7: 남자-남자, 남자-여자, 여자-남자, 여자-여자 이니까요.

R : 한명이 남자라는 것을 알았다는 것은 아무 상관이 없는 거예요?

S7: 네. 어차피 확률은... 둘 다 남자-남자이면 4가지 중에 한 가지이니까.

대부분의 학생들에게 조건사건은 문제에서 추론해낸 사건이 아니라, 정보를 제공하는 직접적인 사건이라는 것을 알 수 있었다. Falk (1986, 1989)는 실생활 문제에는 무작위성이 명확히 언급되어 있지 않고, 조건사건을 확인하기 어렵거나 확인하더라도 관련이 없다고 생각하기 쉬우므로 실생활 문제에 모호함이 있다는 것을 강조했다. 5번 문항이 실생활 문제에 더 가까웠기 때문에 검사를 '무작위로' 행했다는 것을 인식하기 어려웠을 것이다. 두 문항을 통해서 주어진 조건을 고려하지 않고 조건부확률 문제를 단순화시키는 경향을 확인할 수 있었다.

2) 사건의 시간 순서에 대한 오개념

조건사건이 목적사건보다 나중에 일어나는 경우에 조건사건의 확률적 영향을 간과하는 오개념이 나타났고, 조건사건이 먼저 일어나는 경우는 조건부확률과 곱사건의 확률을 혼동하는 경향이 크게 나타났다. 2(1)은 조건사건이 시간상 먼저 일어나는 문제이고 2(2)는 나중에 일어나는 문제였다. 2(1)의 정답률은 64.7%이었지만 2(2)의 정답률은 27.9%로 낮았고 두 문제 모두 맞힌 학생은 19.1%에 불과했다. 두 사건을 곱하는 오류는 2(2)보다 2(1)에서 흔하게 나타났다. 2(2)번 문항에서 많은 학생들이 '두 번째 뽑은 공이 흰색'이라는 조건은 확인하였으나 "첫 번째에 흰 공을 뽑는 것은 두 번째에 흰 공을 뽑았다는 것과는 상관이 없다"는 이유로 조건사건의 영향을 고려하지 않았다.

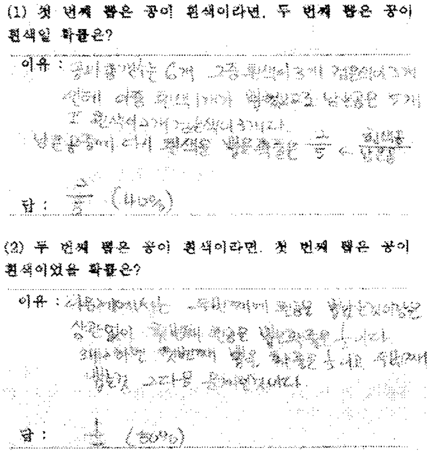
S8: 첫 번째는 두 번째 뽑은 거 일어나기 전이니까, 그냥 첫 번째 뽑은 공이 흰색이었을 확률은 그냥 6분의 3이죠.

R : 두 번째 뽑은 공은 더 나중에 일어나는 것이기 때문에?

S8: 예.

R : 그럼 두 번째 뽑은 공이 검은색이라면 어떻게 되요?

S8: 그래도 똑같죠. 그냥. 그래도 첫 번째 뽑은 공이 흰색일 확률은 2분의 1이죠.



[그림 Ⅲ-4] 2번 문항에 대한 학생의 답안 (H1)

2(2)번 문항에서 정답을 구한 학생들도 두 번째로 뽑은 공에 대한 조건으로 인한 표본공간의 변화를 인식하기보다는 공식을 사용하여 구했다. 정답자 중에서 공식을 사용한 학생은 2(1)번에서는 11.4%이었지만 2(2)번에서는 73.7%이었고, 공식을 사용하지 않고 남은 공의 개수를 생각해서 구한 학생은 2(1)번에서는 88.6%이었지만 2(2)에서는 26.3%에 불과했다. 조건사건이 먼저 일어나는 경우는 학생들의 사고과정에 자연스럽지만 조건사건이 나중에 일어나는 경우는 생각하기 어렵기 때문에 필연적으로 공식에 의존하게 된다는 것을 보여준다.

<표 Ⅲ-5> 2, 4, 6번 문항에 대한 반응 유형의 분포

문항	정답	오답 유형					
2(1)	$\frac{2}{5}$	$\frac{1}{5}\left(=\frac{3}{6}\times\frac{2}{5}\right)$	기타				전체
	44 (64.7)	23 (33.8)	1 (1.5)				68 (100%)
2(2)	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{1}{5}\left(=\frac{3}{6}\times\frac{2}{5}\right)$	기타	무응답		전체
	19 (27.9)	30 (44.1)	3 (4.4)	6 (8.8)	10 (14.7)		68 (100%)
2 (1),(2)	$\frac{2}{5}$, $\frac{2}{5}$	$\frac{2}{5}$, $\frac{1}{2}$	$\frac{2}{5}$,기타오답	$\frac{1}{5}$, $\frac{1}{2}$	$\frac{1}{5}$, $\frac{1}{5}$	기타	전체
	13 (19.1)	21 (30.9)	10 (14.7)	9 (13.2)	4 (5.9)	9 (13.2)	68 (100%)
4(1)	$\frac{2}{15}\left(=\frac{4}{10}\times\frac{3}{9}\right)$	기타					전체
	61 (89.7)	7 (10.3)					68 (100%)
4(2)	$\frac{1}{3}$	$\frac{2}{15}\left(=\frac{4}{10}\times\frac{3}{9}\right)$	기타				전체
	43 (63.2)	17 (25.0)	8 (11.8)				68 (100%)
4(3)	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{2}{15}\left(=\frac{4}{10}\times\frac{3}{9}\right)$	기타	무응답		전체
	13 (19.1)	31 (45.6)	7 (10.3)	10 (14.7)	7 (10.3)		68 (100%)
4 (2),(3)	$\frac{1}{3}$, $\frac{1}{3}$	$\frac{1}{3}$, $\frac{2}{5}$	$\frac{1}{3}$,기타오답	$\frac{2}{15}$, $\frac{2}{5}$	$\frac{2}{15}$, $\frac{2}{15}$	기타	전체
	11 (16.2)	25 (36.8)	7 (10.3)	5 (7.4)	7 (10.3)	13 (19.1)	68 (100%)
6	$\frac{4}{9}$	$\frac{1}{2}$	$\frac{2}{9}\left(=\frac{5}{10}\times\frac{4}{9}\right)$	기타	무응답		전체
	15 (22.1)	27 (39.7)	12 (17.6)	5 (7.4)	9 (13.2)		68 (100%)

4(3)번 문항에서도 시간 순서의 오류를 나타낸 학생이 45.6%였다. 4(2)번에서 조건사건을 파악하고 정답을 구한 학생의 절반 이상이 4(3)번 문제를 제대로 해결하지 못했으므로 두 문제를 모두 맞힌 학생은 16.2%에 불과했다. 6번 문항에서도 39.7%의 학생들이 시간 순서의 오개념을 나타냈고, 17.6%의 학생들이 곱사건의 확률과 혼동하였다. 학생 S9는 4(3)번 문제를 공식을 사용하지 않고는 해결할 수 없었으며, 공식을 사용하지 않고 추론할 것을 요구하자 시간 순서에 기인한 오류를 나타냈다. 그러나 6번 문제의 면담 과정에서 6(1)과 6(2)의 확률적 구조가 같음을 깨달았다.

S6: (3)번도 똑같이 (공식을 사용해서) 했어요.

R : 이걸 공식을 안 쓰고 하면 어떻게 될까요? 앞의 문제에서 첫 번째 공을 뽑으면 하나가 제외되고 표본공간이 변한다고 했잖아요. 그것처럼 이것은 을이 당첨제비를 하나 뽑으니까 표본공간이 변한다고 생각하지는 않았어요?

S6: 아무래도 사건 순서가 있으니까요. 을이 뽑았을 때 표본공간이 변한다고 생각하기는 좀... 시간 순서대로 생각하잖아요. 갑이 먼저 뽑고 을이 뽑는 거잖아요. 을이 먼저 뽑았을 때(을이 조건인 경우)를 생각하기가 힘들어요.

갑이 당첨 → $\frac{4}{10}$
제비 뽑을 확률

$\frac{4}{10}$에서 갑이 당첨제비를 하나 뽑았으므로 남은 제비는 9개 당첨제비는 3개가 된다.

3 ← 남은 당첨제비
9 ← 남은 제비

답 : $\frac{3}{9} = \frac{1}{3}$

이유 : 갑의 사건이 먼저 일어났으므로 이때에서는 을의 사건과는 무관하다 그러므로 $\frac{4}{10}$이다.
갑이 당첨제비 뽑을 확률.

답 : $\frac{4}{10} = \frac{2}{5}$

[그림 III-5] 4번 문항에 대한 학생의 답안 (H1)

S8: 이것도 나중에 일어난 거니까 처음 거랑 상관이 없잖아요. 그래서 2분의 1이라고 했는데요.

R : 처음 거랑 상관이 없다는 것은 어떤 뜻이에요?

S8: 그러니까, 얘가 나중에 뽑았을 때 그 확률이 그 다음에 무엇이 또 있다면 다음 거에는 영향을 미치겠지만, 그 전의 것에는 영향을 못 미치잖아요.

S9: 신기하네요. 둘이 답이 같은 게 신기해요. 두 번째에 초콜릿을 이미 하나 뽑았다는 것을 알면, 처음에 뽑은 물건으로 고려해야 할 것은 아홉 개 중에 네 개죠. 그렇게 되면... 그런데, 왠지 순서가 되게 중요하다고 생각을 했었는데...(중략) 지금 생각하니까 순서와는 상관이 없는 것 같아요. 왠지 두 번째 뽑는 건 첫 번째 뽑는 거에 별로 영향을 안줄 것 같다고 생각을 했는데...

3) 조건관계와 인과관계에 대한 오개념

두 사건 A, B가 정적상관이 있고 조건부확률 P(A|B)와 P(B|A)가 실제로는 같지만 A가 B의 원인인 경우에 조건관계와 인과관계를 혼동하는 어려움이 나타날 수 있다. 7번의 경우 두 확률이 같다는 반응은 42.6%이었고, 엄마의 눈

이 갈색일 때 딸의 눈이 갈색일 확률이 그 반대의 경우보다 크다고 생각한 학생들은 27.9%로 나타났다. 학생들은 두 조건부확률을 고려해서 각 모집단에서 갈색 눈을 가진 사람의 비율이 동일하기 때문에 두 확률이 같다고 생각했다. 두 확률이 같다는 대답이 많았지만 대체로 조건사건을 원인으로 파악할 수 있는 경우의 조건부확률이 더 크다고 생각하는 경향이 나타났다.

일반적으로 부모의 형질이 자식들에게 유전되기 때문에 엄마의 눈이 갈색 눈일 때는 딸의 눈도 갈색일 확률이 높다. (H24)

$$① = \frac{P(\text{엄마 딸 갈색})}{P(\text{딸 갈색})} \quad ② = \frac{P(\text{엄마 딸 갈색})}{P(\text{엄마 갈색})}$$

사람마다 갈색 눈을 가질 확률이 같다면, ① = ②이다.

[그림 Ⅲ-6] 7번 문항에 대한 학생의 답안 (H37)

마찬가지로 8번 문항에서도 두 확률이 같다는 반응이 많았으나, 학생들은 암이 걸렸을 때 양성반응이 나타날 확률을 그 반대의 경우보다 더 크다고 생각했다. 즉, 조건부확률의 비교에서 인과적 증거를 진단적 증거보다 더 중요하게 생각하는 경향을 확인할 수 있었다.

암에 걸렸다면 확실히 양성반응을 나타내는 반면 검사결과도 양성반응을 나타내도 암이 아닐수가 있다

[그림 Ⅲ-7] 8번 문항에 대한 학생의 답안 (H43)

4) 독립성에 대한 오개념

3번 문항은 동전을 잇달아 던져서 일어날 가능성이 가장 큰 것과 작은 것을 고르는 문항으로, 가능성이 가장 작은 것을 고르는 문제에서 오답률이 높게 나타났다. 학생들은 한쪽 면이 연속으로 나오는 경우보다 THHTH, HTHTH와

<표 Ⅲ-6> 7번과 8번 문항에 대한 반응 유형의 분포

문항	정답	오답 유형				
7	③같다	①이 더 크다.	②가 더 크다.	알 수 없다.	무응답	전체
	29 (42.6)	8 (11.8)	19 (27.9)	6 (8.8)	6 (8.8)	68 (100%)
8	알 수 없다.	①이 더 크다.	②가 더 크다.	③같다.	무응답	전체
	3 (4.4)	6 (8.8)	14 (20.6)	31 (45.6)	14 (20.6)	68 (100%)

<표 Ⅲ-7> 3번 문항에 대한 반응 유형의 분포

문항	정답	오답 유형					
3(1)	⑤ 같다.	①HHHHT	②THHTH	③THTTT	④HTHTH	무응답	전체
	57 (83.8)	0 (0.0)	7 (10.3)	0 (0.0)	3 (4.4)	1 (1.5)	68 (100%)
3(2)	⑤ 같다.	①HHHHT	②THHTH	③THTTT	④HTHTH	무응답	전체
	52 (76.5)	13 (19.1)	0 (0.0)	0 (0.0)	2 (2.9)	1 (1.5)	68 (100%)
3(3)	③ 같다.	①앞앞앞앞앞		②뒤앞앞앞뒤앞		무응답	전체
	57 (83.8)	0 (0.0)		11 (16.2)		0 (0.0)	68 (100%)

같이 불확실한 과정의 특성을 반영한 경우의 가능성이 더 크다고 생각했다. 표본이 모집단과 유사하고 표본의 추출과정이 무작위성을 대표해야 한다는 것을 아주 작은 표본에서도 기대하는 대표성 전략을 사용하고 있음을 알 수 있다. 사건의 독립성에 관한 오개념은 대학생도 가지고 있었다.

S9: 각각이 앞뒤가 나올 수 있는 확률은 2분의 1이잖아요. 앞이 나올 확률은 2분의 1, 뒤가 나올 확률도 2분의 1. 그러니까 앞이 한 번 나오고 나면은 아무래도 다음에는 뒤가 나올 확률이 좀 더 높지 않을까 생각해서.

R : 그래서 1번(HHHHT)이 나올 확률이 가장 드물다고 생각했어요?

S9: 예. 그러니까 번갈아서 나오는 것보다는 드물다고 생각했어요. 3번도 같은 맥락으로 생각했어요. 이것도 역시 (ㄱ)은 앞면 계속 나온 거고, (ㄴ)은 '뒤앞앞앞뒤앞'이런 식으로 나와서, 뒤가 한 번 나왔으면 앞이 나올 확률이 좀 더 높고. 그런 식으로.

나. 조건부확률에 대한 이해 분석

1) 조건사건의 이해 (9번 문항)

1/3(정답)이라고 답한 학생들은 대부분 공식을 이용해서 문제를 해결했고, 20.6%의 학생들만 조건사건의 표본공간을 고려해서 해결했다. 47%의 학생들이 문제에 주어진 사건의 확률을 곱해서 1/6이라고 하거나, 짝수이면서 3의 배수일 확률을 구했다. 조건사건과 목적사건이 동시에 일어나는 문제였기 때문에 조건부확률과 곱사건의 확률을 혼동하는 경향이 크게 나타난 것으로 추측할 수 있다.

2) 복원상황과 비복원상황의 구분 (10번 문항)

복원상황과 비복원 상황의 확률 변화를 이해하고 있는가를 알아보고자 한 문항으로, 고등학교 3학년 학생에게 비교적 쉬운 문제였지만 정답률은 80%이었다. 나머지 학생들도 확률의 변화는 인식하고 있었으나 문제에 주어진 사건의 확률을 곱하여 $P(A)\times P(B|A)$를 구한 학생들이 있었다. 주어진 사건들의 확률을 적절하지 않은 상황에서 곱하는 오류가 나타났다.

3) 조건부확률 공식의 적용 (11번 문항)

66.2%의 학생들이 조건부확률의 공식을 적용해서 정답을 구했고, 공식을 몰라서 답을 구하지 못한 7명의 학생을 제외하고 모두 공식을 이용해서 풀었다. 오답을 구한 학생들 중 6명은 조건부확률의 공식을 잘못 알고 있었다. $P(A|B)=\frac{P(A\cap B)}{P(A)}$로 공식을 잘못 적용한 학생들이 6명이나 있는 것으로 볼 때, 학생들이 조건부확률을 이해하지 않고 공식만 암기하고 있음을 알 수 있다.

<표 Ⅲ-8> 이해 조사의 문항 별 정답률

문항	9	10(1)	10(2)	11	12(1)	12(2)	12(3)	12(4)	13	14	15	16
정답률(%)	45.6	82.4	79.4	66.2	82.4	75.0	69.1	69.1	26.5	67.6	25.0	54.4
정답자 수	31	56	54	45	56	51	47	47	18	46	17	37
오답자 수	36	8	10	16	6	9	13	12	42	18	36	27
무응답	1	4	4	7	6	8	8	9	8	4	15	4
전체	68	68	68	68	68	68	68	68	68	68	68	68

4) 조건부확률의 비교 (12번 문항)

비조건부확률 문제였던 (1)의 정답률은 82.4%로 비교적 높았으나, 조건부확률을 비교하는 (3), (4)의 정답률은 70%에 미치지 못했다. (2)번은 P(X)와 P(X|A)를 비교하는 문제였으나, '모양과 맛은 연관이 없다'라고 생각해서 사탕의 모양에 대한 정보를 무시하는 학생도 있었으며, '막대모양의 딸기 맛 사탕은 8개이므로 확률은 1/5로 줄어든다'고 생각해서 결합사건의 확률을 구한 학생도 있었다. 약 20%의 학생들이 비조건부확률과 조건부확률을 비교하고, 조건부확률을 서로 비교하는 데에서 어려움을 겪었다. 조건부확률의 이해는 이원표로 제시된 범주형 자료를 정리하고 분석하는 데에 필수적이므로, 학생들이 자료를 해석하는 도구로서 조건부확률을 이용할 수 있도록 지도할 필요가 있다.

5) 곱의 법칙과 합의 법칙 (13번 문항)

정답을 구한 학생들은 모두 첫 번째 뽑힌 학생이 남자인 경우와 여자인 경우에 각각 확률을 구해서 더하는 방법을 사용했다.[6] 제비뽑기의 특성 상 언제 뽑든 확률이 같다는 것을 인식하지 못했으며, 두 번째로 뽑는 사건이기 때문에 첫 번째의 경우를 반드시 나누어서 따져야 한다고 생각하는 것으로 볼 때, 순차적이고 인과적으로 사고하고 있음을 알 수 있다.

6) 독립사건과 종속사건의 구분 (14번, 16번 문항)

14번 문항에서 '주사위를 던진다고 해서 동전의 앞면과 뒷면에 영향을 주지 않으므로 독립'이라고 답한 학생은 37명이었고, $P(A \cap B) = P(A)P(B)$이기 때문에 독립이라고 답한 학생은 9명이었다. 반면, 16번 문항에서 '여학생인 경우 알레르기가 있고 없는 확률과 남학생인 경우 알레르기가 있고 없는 확률이 다르므로 종속'이라고 답한 학생은 24명이었고, 공식을 이용해서 종속이라고 답한 학생은 13명이었다. 공식 사용의 차이는 14번 문항이 16번보다 사건의 관련성을 추론해내기 쉬웠기 때문이라고 볼 수 있다.

몇몇 학생들은 두 사건이 동시에 일어날 수 없으므로 홀수의 눈이 나오는 사건과 앞면이 나오는 사건은 독립이지만, 여학생일 사건과 알레르기가 있을 사건은 동시에 일어날 수 있으므로 종속이라고 대답했다. 이것은 학생들이 배반사건과 독립사건을 혼동한다는 것을 의미한다. 학생들은 독립사건과 배반사건의 개념을 어려워할 뿐만 아니라 두 개념을 혼동하기도 하며, 실지로 교과서에도 두 개념을 관련지어 가르치지 않고 있다. 또한 독립과 종속을 구분하는 데에 시간적으로 '동시에' 일어나는 사건인가 그렇지 않은가에 의존해서 판단하고 있으므로, 학생들은 무의식적으로 '시간'이라는 요인을 독립·종속의 개념과 관련시키고 있음을 알 수 있다. 반드시 공식만으로 종속관계와 독립관계를 구분할 수 있는 것은 아니므로, 학생들이 사건의 관련성을 파악할 수 있는 다양한 방법을 이해할 수 있게 할 필요가 있다.

7) 베이즈 정리 (15번 문항)

15번 문항은 베이즈 정리를 적용할 수 있는가를 알아보기 위한 문제로서, 25%의 학생들이 3/7(정답)을 구했고, 29%의 학생들은 결합사건의 확률을 구했다.

이해 분석의 결과, 조건사건을 고려하여 표본공간을 구성하기보다는 공식을 선호하며, 주어진 사건의 확률을 곱해서 결합사건의 확률을

---

6) $\frac{4}{8} \times \frac{4}{7} + \frac{4}{8} \times \frac{3}{7}$

구하는 오류가 많이 나타났다. 조건부확률을 서로 비교하는 문제는 내용이 기초적인 경우에도 정답률이 70%에 미치지 못했다. 시간에 따라 순차적이고 인과적으로 사고하는 경향이 나타났고, 많은 학생들이 확률의 종속과 독립을 직관적으로 이해하지 못했다. 또한, 조건부확률 공식은 알고 있어도 베이즈 정리를 적용하지 못했다.

분석 결과 모든 학생들이 하나 이상의 오개념을 갖고 있으며, 성취 수준이 우수한 대학생도 오개념을 가지고 있었다. 이는 조건부확률 개념에 숙달되었을 것이라 예상되더라도 개념을 정확하게 사용하고 있지는 못하다는 것을 보여준다. 분석을 통해 나타난 이해의 특징을 정리해보면 다음과 같다.

첫째, 학생들은 조건사건으로 문제에서 주어진 직접적인 자료를 사용하기 보다는 그들이 추론해낸 사건을 조건사건으로 사용하였다. 이것은 문제에서 조건사건을 확인하고 주어진 정보를 적절하게 사용하는 데에 어려움을 느낀다는 것을 보여준다.

둘째, 조건사건이 목적사건보다 시간상 나중에 일어나는 경우에 조건사건을 무시하는 오개념이 나타났다. 조건사건과 목적사건이 역순서로 나타날 때 나중에 일어나는 정보가 무엇이든 먼저 일어나는 사건에는 영향을 주지 않으므로 조건사건의 정보를 이용하지 않아도 된다고 생각했다. 면담 과정에서 학생들이 조건부확률 문제에서 인과적 추론을 사용하고 있음이 파악되었다.

셋째, 적절하지 않은 확률 계산 방법을 사용하였고, 조건부확률 문제해결 방법으로 공식을 선호하였다. 특히 조건사건이 먼저 일어나는 경우는 공식을 사용하지 않고도 표본공간을 고려해서 해결할 수 있었지만, 조건사건이 나중에 일어나는 경우에는 공식 사용의 비율이 높았다. 시간 순서에 따른 추론의 경향 때문에 시간적으로 나중에 일어나는 것이 조건인 상황을 부자연스럽고 어렵게 여겼다.

넷째, 조건부확률과 곱사건의 확률을 혼동하는 경향은 모든 문제에서 나타났지만 특히 조건사건이 먼저 일어나는 경우에 더 많이 나타났다. Pollatsek et al.(1987)의 지적처럼 학생들은 조건부확률을 곱사건의 확률과 분리해서 생각하지 않고, 문제에 따라 다르게 대답하고 있었으며 두 개념을 혼합해서 사용하였다. 또한 두 사건의 확률을 의미를 생각하지 않고 곱하는 경향이 나타났다. 어떠한 경우에 두 사건의 확률을 곱하고, 곱하지 않는가를 구분하기 어려워했다.

다섯째, 조건부확률에 주어진 두 사건이 인과관계를 나타낸다고 생각될 때, 진단적 추론보다 인과적 추론을 선호하는 것으로 나타났다. 몇몇 학생들은 7번 문항 같이 동일한 정보가 포함되어 있더라도 진단적 자료보다 인과적 자료의 확률을 더 크게 추정했다.

여섯째, 다른 영역의 선행지식을 부적절하게 사용하였다. 출생 성별의 생물학적 독립성에 관한 선행 지식 때문에 카드 문제보다 쌍둥이 문제에서 오개념이 더 많이 나타났으며, 유전법칙에 관한 지식 때문에 인과관계로 주어진 조건부확률을 더 크게 추정하는 오개념이 나타났다. 생물학적 원리들이 과학적으로는 옳다고 하더라도 확률적 상황에서는 의미가 없을 수 있기 때문에 많은 오류들을 야기할 수 있다.

일곱째, 조건부확률 문제를 단순화하는 경향이 있었다. 카드 문제와 쌍둥이 문제에서 표본공간을 실제보다 더 단순하게 생각하는 학생들이 많았다. 카드 문제에서 표본공간을 카드의 여섯 개의 면이라고 생각하지 않고, 단순히 주어진 카드 세 장으로 생각했다.

여덟째, 독립 개념을 부적절하게 사용하였다.

조건사건의 정보를 무시함으로써 조건사건과 목적사건이 서로 독립이 아닐 때에도 독립인 것처럼 다루는 경향이 있었으며, 동전 던지기와 같이 서로 독립인 상황을 마치 독립이 아닌 것처럼 다루는 경향이 나타났다.

Ⅳ. 결 론

이 연구에서는 조건부확률 개념을 다양한 관점에서 분석하고, 학습 경험이 있는 고등학생과 대학생을 대상으로 조건부확률의 이해와 오개념을 조사하였다.

수학적 분석에서는 조건부확률이 주어진 사건을 새로운 표본공간으로 하는 확률이라는 대상 개념임을 확인하였고, 아울러 베이즈 정리에 따라 사전확률이 경험적 증거에 근거해서 사후확률로 수정될 수 있다는 확률 수정의 과정 개념임을 확인하였다. 역사 발생적 분석에서는 '이어서 일어나는 두 사건' 중 먼저 일어나는 사건을 A, 나중 일어나는 사건을 B라고 할 때 P(B|A)와 P(A|B)가 별개로 다루어지다가 최근에 이르러서야 수학화가 이루어졌음을 확인하였다. 초기에는 먼저 일어나는 사건에 의해 나중 일어나는 사건의 변화만을 다루었으나, 점차 나중에 일어난 사건의 정보에 의해 먼저 일어난 사건의 확률의 변화를 인식하게 되었음을 알 수 있었다. 심리학적 분석에서는 조건부확률 개념 수준에 관하여 논의하였고, 인식론적 분석에서는 가추적 추론의 관점에서 조건부확률의 의미를 분석하였다. 조건부확률과 베이즈 정리는 경험의 증가에 따라 사전의 지식이 개선되어 더욱 정확한 지식을 만들어내기 위한 것으로서, 확률의 주관적 관점을 잘 반영하고 있음을 알 수 있었다. 이상의 교수학적 분석을 통하여 조건부확률 개념을 분석하여 6가지로 정리하였다.

지필 검사와 면담 분석을 통해 조건부확률을 배운 고등학생의 경우 모든 학생들이 하나 이상의 오개념을 갖고 있었으며, 성취 수준이 매우 우수한 대학생도 오개념을 가지고 있음을 확인할 수 있었다. 또한, 학생들이 가진 조건부확률 개념 이해의 8가지 특징을 추출하였다. 학생들은 조건사건으로 문제에 주어진 자료를 사용하기보다는 자신이 추론해낸 사건을 조건사건으로 사용하였으며, 조건사건과 목적사건의 시간 순서의 오개념을 가지고 있었다. 조건사건이 목적사건보다 시간상 먼저인 문제에서 조건부확률과 곱사건의 확률을 혼동하는 경향이 있었고, 조건부확률 문제해결 방법으로 표본공간의 구성보다는 공식을 선호하였다. 조건부확률의 비교 문제에서는 진단적 자료보다 인과적 자료에 대한 확률을 더 크게 추정하였으며, 다른 영역의 선행 지식을 부적절하게 사용하였다.

이와 같이 조건부확률의 개념 분석과 이해의 특징에 대한 연구 결과로부터 이후 조건부확률과 관련된 교육과정을 설계함에 있어 다음과 같은 몇 가지 시사점을 얻을 수 있었다.

첫째, 조건부확률의 공식과 같은 '대상' 관점뿐만 아니라 사전확률이 사후확률로 수정·개선되는 '과정'의 관점에서 이해할 수 있도록 지도되어야 한다. 둘째, 학생들이 조건부확률 오개념을 극복할 수 있는 교육과정 구성이 필요하다. 시간 순서와 조건관계는 관련이 없음을 인식시키고, 조건관계를 인과관계와 구분할 수 있는 문제들이 포함되어야 한다. 셋째, 문제해결 방법으로 조건부확률 공식에만 의존하기보다는 조건사건에 의해 변화되는 표본공간을 구성할 수 있게 지도해야 한다. 이를 위해서는 이원표, 수형도, 면적모델 등 다양한 해결 방법의 지도가 절실하다.

이 연구에서는 조사에 참여한 학생들의 표본이 작고, 한 대학에 속한 대학생들을 대상으로 하였으므로 결과를 일반화하는데 한계가 있다. 조사에 사용된 문항은 선행연구에서 사용되었거나 직접 제작한 것으로 표준화된 도구가 아니었다. 따라서 보완된 발전적인 후속 연구가 요구된다. 후속 연구로서 조건부확률 교육과정 설계에 관한 연구가 필요하며, 지도 시기와 관련하여 확률개념 수준에 관한 연구와 이를 위해 전 학년에 걸친 조건부확률 이해에 관한 연구가 필요하다.

참고문헌

김해경(2001). **통계적 추론(Statistical Inference).** 대우학술총서 510. 서울: 아카넷.

박기호(2001). **조건부확률의 효과적인 지도 방안 연구.** 건국대학교 석사학위 논문.

우정호(2004). '인간교육'을 위한 주요 교과로서의 수학교육. **대한수학교육학회 수학교육논총, 26,** 1-20.

이용구(1991). 역확률. **통계학사 콜로키움** (허명회 편). 서울: 자유아카데미.

임재훈 · 이경화 · 김진호 · 윤오영 · 반용호 · 조동석 · 이희종 · 박수연 · 한명주 · 남승진 (2002). **고등학교 수학 I.** 서울: (주)두산.

최용준 · 신현성(2002). **고등학교 수학 I.** 서울: (주)천재교육.

Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. Phil. Trans. Roy. Soc. London for 1763, 53, 370-418. Reprinted by W. E. Deming(1940), *Facsimiles of Two Papers By Bayes.* New York: Hafner Publishing Company.

Carnell, L. J. (1997). *Characteristics of Reasoning about Conditional Probability.* Doctorial dissertation, Greensboro University.

Dahl, H. (1994). Teaching Independence. *Teaching Statistics, 16*(2), 34-37.

Falk, R. (1986). Conditional probabilities: Insights and Difficulties. *The Second International Conference on the Teaching of Statistics,* 992-997.

________(1989). Inference under Uncertainty via Conditional probabilities. In R. Morris (Ed.), The Teaching of Statistics, *Studies in Mathematics Education. 7,* 175-184. UNESCO.

Hacking, I. (1975). *The Emergence of probability.* Cambridge University Press. University of Sheffield.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*(2), 101-125.

Kapadia, R. (1986). Didactical Phenomenology of Probability. *Proceedings of the Second International Conference on Teaching Statistics (ICOTS)* (2nd, University of Victoria, Canada, August, 1986), 260-264.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education,* 24, 392-414.

Pollatsek, A., Well, A. D., Konold, C., Hardiman, P., & Cobb, P. (1987). Understanding Conditional Probabilities.

Organizational Behavior & Human Decision Processes, 40(2), 255-269.

Poole, D. (2000). Learning, Bayesian probability, Graphical models and Abduction. In P. A. Flash and A. C. Kakas (Eds.), *Abduction and Induction,* 153-168. Dordrecht: Kluwer Academic Publishers.

Salsburg, D. (2003). 천재들의 주사위 (최정규 역.). 서울: 뿌리와 이파리. (원저 *The Lady Tasting Tea.* 2001 출판)

Shafer, G. (1982). Bayes' Two Arguments for the Rule of Conditioning. *The Annals of Statistics, 10*(4), 1075-1089.

Shaughnessy, J. M. (1981). Misconceptions of probability : From systematic errors to systematic experiment and decisions. In A. P. Shulte (Ed.). *Teaching Statistics and Probability* (1981 Yearbook of NCTM). Reston, VA: NCTM.

________(1992). Research in Probability and Statistics : Reflections and Directions. In D. A. Grouws (Ed.). *Handbook of Research on Mathematics Teaching and Learning,* 465-494. New York: Macmillan Publishing Company.

Stigler, S. M. (1986). The History of Statistics : *The Measurement of Uncertainty before 1990.* Harvard University Press, MA: Cambridge.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal, 9,* 39-59.

A Didactic Analysis of Conditional Probability

Lee, Jung Yeon (Gaepo Middle School)
Woo, Jeong Ho (Seoul National University)

The notions of conditional probability and independence are fundamental to all aspects of probabilistic reasoning. Several previous studies identified some misconceptions in students' thinking in conditional probability. However, they have not analyzed enough the nature of conditional probability.

The purpose of this study was to analyze conditional probability and students' knowledge on conditional probability. First, we analyzed the conditional probability from mathematical, historico-genetic, psychological, epistemological points of view, and identified the essential aspects of the conditional probability. Second, we investigated the high school students' and undergraduate students' thinking in conditional probability and independence. The results showed that the students have some misconceptions and difficulties to solve some tasks with regard to conditional probability. Based on these analysis, the characteristics of reasoning about conditional probability are investigated and some suggestions are elicited.

* **Key words** : conditional probability(조건부확률), didactic analysis(교수학적 분석), misconception(오개념), independence(독립), dependence(종속)

논문접수: 2009. 4. 2
논문수정: 2009. 5. 18
심사완료: 2009. 5. 22

<부록 1> 지필 검사 문항

1. 모자 안에 다음 세 장의 카드가 들어있다. 양면이 모두 빨간색인 카드, 양면이 모두 녹색인 카드, 한 면은 빨간색이고 다른 한 면은 녹색인 카드가 들어있다. 눈을 감고 모자에서 무작위로 카드 한 장을 꺼내서 테이블 위에 놓았더니 빨간색이었다. 그렇다면 이 카드를 뒤집었을 때에도 빨간색일 확률을 구하여라.

2. 주머니에 흰색 공 3개와 검은색 공 3개가 들어있다. 공을 하나 꺼낸 후 다시 집어넣지 않고 다시 공을 꺼낸다.
   (1) 첫 번째 뽑은 공이 흰색이라면, 두 번째 뽑은 공이 흰색일 확률은?
   (2) 두 번째 뽑은 공이 흰색이라면, 첫 번째 뽑은 공이 흰색이었을 확률은?

3. 정상적인 동전을 연속해서 다섯 번 던진다고 하자. 다섯 번을 순서대로 던져서 다음과 같은 결과를 얻었다. (H는 앞면이고 T는 뒷면이다.)
   (1) 다음 중 어떤 경우가 가장 많이 일어날 것 같은가?
   ① HHHHT ② THHTH ③ THTTT
   ④ HTHTH ⑤ 가능성이 모두 같다.
   (2) 위의 보기 중에서 어떤 경우가 가장 적게 일어날 것 같은가?
   (3) 정상적인 동전을 연속해서 여섯 번 던진다고 하자. 다음 ㉠과 ㉡ 중에서 어떤 경우가 더 잘 일어날 것 같은가?
   ㉠ 앞앞앞앞앞앞 ㉡ 뒤앞앞앞뒤앞
   ① ㉠ ② ㉡ ③ 가능성은 같다.

4. 10개의 제비 중 4개가 당첨제비라고 한다. 갑이 먼저 제비를 뽑고, 을이 두 번째로 제비를 뽑는다고 할 때, 다음을 구하여라. (단, 뽑은 제비는 다시 넣지 않는다.)
   (1) 갑과 을이 모두 당첨제비를 뽑을 확률은?
   (2) 갑이 당첨제비를 뽑았을 때, 을이 당첨제비를 뽑을 확률은?
   (3) 을이 당첨제비를 뽑았을 때, 갑이 당첨제비를 뽑았을 확률은?

5. 이란성 쌍둥이를 임신한 한 아주머니가 있다. 아이에게 이상이 있는가를 테스트하기 위해서 태아의 세포를 무작위로 선택하여 염색체 검사를 했더니 아이 중 한 명이 남자아이라는 것을 알았다. 남자아이와 여자아이가 태어날 가능성이 같다고 한다면, 두 아이 모두 남자아이일 확률은 얼마인가?

6. 주머니에 사탕 5개와 초콜릿 5개가 들어있다. 꺼낸 것을 다시 집어넣지 않고 하나씩 두 번을 뽑았다. 두 번째 뽑은 것이 초콜릿일 때, 첫 번째 뽑은 것도 초콜릿일 확률을 구하여라.

7. 딸이 한 명 있는 가족들의 집합이 있다고 하자. 보기의 확률 중에서 어느 것이 더 클 것 같습니까?

① 딸의 눈이 갈색일 때, 엄마의 눈이 갈색일 확률

② 엄마의 눈이 갈색일 때, 딸의 눈이 갈색일 확률

③ 두 확률은 같다.

8. 어떤 도시에 살고 있는 모든 거주자들을 대상으로 암 검사를 실시하였다. 양성반응은 암에 걸렸다는 것을 암시해주며, 음성반응은 암에 걸리지 않았다는 것을 암시해준다. 실제로 의학검사가 완전한 것은 아니므로 양성반응이 나타나지만 암이 아닌 경우도 있고, 암이지만 양성반응이 안 나타나는 경우도 있다. 보기의 확률 중에서 어느 것이 더 클 것 같습니까?

① 검사결과가 양성반응으로 나타난다면 그 사람은 암에 걸렸을 것이다.

② 암에 걸렸다면 그 사람의 검사결과는 양성반응으로 나타날 것이다.

③ 두 확률은 같다.

9. 한 개의 주사위를 던져 나온 눈이 짝수일 때, 그것이 3의 배수일 확률을 구하여라.

10. 봉지에 보라색 젤리 7개와 녹색 젤리 4개가 있다. 젤리를 하나 꺼냈더니 녹색이었다.

(1) 녹색 젤리를 먹지 않고 봉지에 넣고 다시 젤리를 꺼낸다고 할 때, 보라색 젤리를 선택할 확률은 얼마인가?

(2) 녹색 젤리를 먹고 나서 다시 젤리를 꺼낸다고 할 때, 보라색 젤리를 선택할 확률은 얼마인가?

(1)과 비교해서 확률은 어떻게 변하였는가?

11. 어느 시행에서 일어날 수 있는 사건 A, B에 대하여 $P(A)=0.6$, $P(B)=0.4$, $P(A|B)=0.3$일 때, $P(B|A)$를 구하여라.

12. 주머니 안에 막대사탕, 둥근사탕, 네모사탕의 세 종류의 사탕 40개가 담겨 있다. 딸기 맛과 오렌지 맛이 각각 20개씩 있다. 사탕의 맛과 모양을 구분해서 표와 같이 개수를 헤아려 보았다. 영미가 상원이 몰래 사탕을 하나 집고, 상원이가 그 사탕의 맛을 맞추는 게임을 하기로 했다.

	막대사탕	둥근사탕	네모사탕	합계
딸기 맛	8	2	10	20
오렌지 맛	5	10	5	20
합계	13	12	15	40

(1) 상원이가 "딸기 맛 사탕일거야"라고 추측하였을 때, 상원이가 영미의 사탕을 맞출 확률은 얼마인가?

(2) 영미는 "내가 집은 것은 막대모양의 사탕이야"라고 가르쳐 주었다. 상원이가 "딸기 맛 사탕일거야"라고 추측할 때, 영미가 가르쳐 준 모양은 상원이가 맞추는 것에 도움이 되는가?

(3) 영미가 어떤 모양을 집어서 가르쳐 주어야, 딸기 맛일 거라는 상원이의 추측에 가장 도움이 되겠는가?

(4) 영미가 어떤 모양을 집어서 가르쳐 줄 때, 딸기 맛일 거라는 상원이의 추측에 가장 도움이 되지 않겠는가?

13. 4명의 남학생과 4명의 여학생으로 이루어진 모임에서 동아리 대표 두 명을 뽑으려고 한다. 처음 한 명을 임의로 뽑고 나머지 7명 가운데 한 명을 임의로 뽑을 때, 두 번째 뽑힌 사람이 여학생일 확률을 구하여라.

14. 주사위 한 개와 동전 한 개를 동시에 던질 때, 주사위의 홀수의 눈이 나오는 사건을 A, 동전의 앞면이 나오는 사건을 B라고 하자. A와 B는 서로 독립인지 종속인지 조사하여라.

15. 통계에 의하면 어떤 지역에서 심장마비로 사망할 확률은 50세 미만이 0.02이고 50세 이상이 0.04라고 한다. 이 지역 인구의 60%가 50세 미만이라고 할 때, 심장마비로 사망한 사람이 50세 미만일 확률을 구하여라.

16. 상우네 반 학생은 총 36명인데, 남자가 19명이고 여자가 17명이다. 봄이라서 알레르기 때문에 많은 학생들이 기침과 재채기를 한다. 상우네 반 남학생 중에 12명이 알레르기가 있고, 여학생 중에 9명이 알레르기가 있다. 어떤 학생이 여학생일 사건을 A, 알레르기가 있을 사건을 B라고 하자. 사건 A와 B는 서로 독립인지 종속인지 조사하여라.