대한수학교육학회지<학교수학>제3권 제1호
Journal of the Korea Society of Educational Studies in Mathematics
School Mathematics, Vol.3, No.1, 101~108 March 2001

확률화 응답 기법을 활용한 확률적 독립의 이해

최 경 호 (전주대학교)
김 래 선 (전주대학교)

I. 서 론

1992년 제 6차 고교 교육과정 개편을 통하여, 현 고등학교 수학과목의 교육과정은 공통 수학, 수학 I, 그리고 수학 II로 구성되어 있다. 수학 I의 내용에는 확률부분이 있는데, 이는 확률의 정의를 비롯하여 확률에 관계된 다양한 내용을 담고 있다. 이들 내용 중 학습자 입장에서 비교적 이해하기 어렵다고 여겨지는 확률적 독립(probabilistic independence) 등의 몇 가지 확률이론을 참여실습을 통하여 즐겁게 학습할 수 있는 방법을 소개하고자 한다. 학습자 입장에서 확률적 독립 등에 대한 이해에 어려움을 겪는 이유중의 하나는 이를 가르치는 선생님들 중, 통계학을 전공으로 깊이 있게 공부한 분이 적은 때문이라 생각된다. 따라서 본 논문은 일선에서 수학을 담당하는 이러한 선생님들에게 도움을 주고자 한다. 부언하면, 수업시간에 확률화응답기법(randomized response technique : RRT)을 활용하는 연습을 통하여, 교과과정의 일부인 확률이론을 습득시킬 수 있다. 예를 들어 RRT를 이용한 추정을 함에 있어 필요한 공식을 유도하는 과정에서 사건(event)의 상호배반(mutually disjoint), 상호독립(mutually independence) 등의 기초적 확률이론을 자연스럽게 이해시킬 수 있다.

확률화응답기법은 응답자가 정직하게 응답하기 어려운, 주로 개인의 사생활과 관련된 민감한 사안(sensitive issue)에 대한 조사시 거짓응답 등으로 인한 비표본오차를 줄이기 위하여 1965년에 Warner에 의하여 제안된 조사기법이다. 본질적으로 RRT란, 응답자에게 신분보호를 위한 확률장치(randomizing device)를 통해서 어떤 확률 하에서 질문을 선택할 수 있는 기회를 부여하여 응답의 편의를 없애거나 줄일 수 있도록 고안된 간접질문방식으로 이후 Greenberg et al.(1969), Moors(1971) 등에 의하여 개선, 발전되어 오고 있다.

동기개발과 흥미를 위하여 RRT를 활용하는 연습과정에서 학생들로 하여금, '민감질문의

구성'과 '확률장치의 구성' 그리고 '응답획득의 과정'을 직접 수행토록 하면 학습의 효과가 증대되리라 생각한다. 본 논문에서는 위의 과정을 Loynes(1976)의 강요형 모형(forced answer model)을 이용하여 설명하고자 한다.

II. 교과서에 표현된 독립의 정의

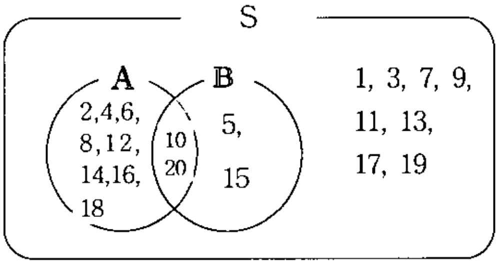
현재 일선고교에서 사용하고 있는 수학 I 교과서의 종류는 수 가지에 이른다. 그런데 이들 교과서에 나타난 확률적 독립에 대한 표현은 거의 대동소이한 바, 이 중 몇 가지를 선택하여 정리하면 다음과 같다.

<표 1> 교과서에 표현된 독립의 정의

정의 / 출판사	확률적 독립에 대한 정의 및 곱셈법칙
교학사 (1999)	두 사건 A, B에 대하여 $P(B\|A)=P(B)$ 또는 $P(A\|B)=P(A)$일 때, 사건 A와 B는 상호독립이라 하고, 이들 사건 A, B를 독립사건이라 한다. 두 사건 A, B가 독립일 때, 다음이 성립한다. $P(A\cap B)=P(A)\cdot P(B)$
지학사 (1998)	두 사건 A, B가 있을 때 조건부확률 $P(B\|A)$는 일반적으로 $P(B)$와 같지 않다. 그러나 $P(B\|A)=P(B)$이면, 즉 A가 일어난 일이 B가 일어나는 확률에 영향을 주지 않으면 B는 A와 독립이라고 한다. 두 사건 A와 B가 독립이기 위한 필요충분조건은 $P(A\cap B)=P(A)\cdot P(B)$이다.
천재교육 (1997)	두 사건 A, B에 대하여 한 사건이 일어나는지 일어나지 않는지에 관계없이 다른 한 사건이 일어날 확률이 일정할 때, 즉 $P(B\|A)=P(B\|A^c)$일 때, A와 B는 상호 독립이라 하며, 두 사건, A, B가 서로 독립일 때, $P(A\cap B)=P(A)\cdot P(B)$가 성립한다
중앙교육 연구소 (1999)	두 사건 A, B에 대하여 한 사건이 일어나는지 안 일어나는지에 따라 다른 한 사건이 일어날 확률에 영향을 주지 않을 때, 이 두 사건은 서로 독립이라고 하며, 서로 독립인 사건을 독립사건이라고 한다. 두 사건 A, B가 서로 독립일 때, 사건 B가 일어날 확률은 $P(B)=P(B\|A)=P(B\|A^c)$이며 다음이 성립한다. $P(A\cap B)=P(A)\cdot P(B)$

한편, 이들 교과서에서 정의된 확률적 독립에 대한 이해를 돕고자 이용되고 있는 예제를 보면 거의 벤다이어그램이나 주사위와 동전을 이용하고 있다. 예제의 한 보기를 보면 다음과 같다.

[예제-교학사(1999)] 20이하의 자연수에서 임의로 하나의 수를 뽑을 때 그 수가 짝수일 사건을 A, 5의 배수일 사건을 B라 하면



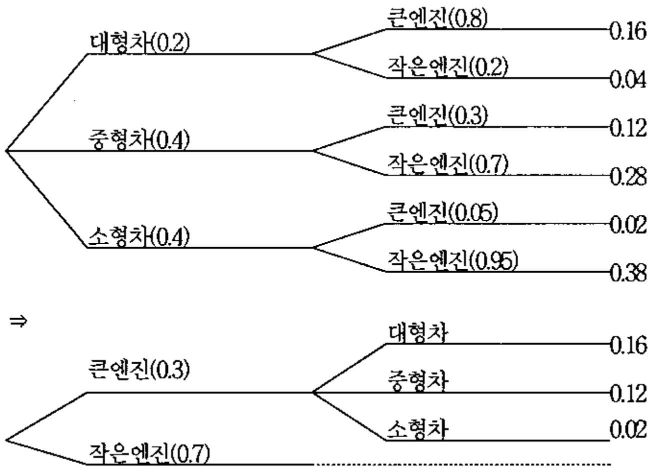
$P(A)=\frac{10}{20}$, $P(B)=\frac{4}{20}$, $P(A\cap B)=\frac{2}{20}$ 이다. 따라서 사건 A와 B는 독립이다.

이들 예제들의 풀이과정을 보면 다분히 독립의 정의로부터 유도된 확률의 곱셈정리를 이용하여 독립여부를 계산결과로부터 확인하는 과정일 뿐, 확률적 독립의 보다 정확한 의미를 이해시키기에는 부족하다. 참고로 Wonnacott와 Wonnacott(1990) 나타난 확률적 독립의 정확한 정의를 보면 다음과 같다.

「두 사건 A와 B에 대하여 $P(A|B)=P(A)$이면 사건 A는 B에 독립이다. 한편 의미적으로 보았을 때, 사건 A가 B에 독립이면 조건부확률로부터 사건 B도 A에 독립임이 유도된다. 따라서 사건 A가 B에 독립일 때, 사건 B는 A에 항시 독립이다.」

이상에서 볼 때, 고교 수학과정에서 확률적 독립에 대한 이해를 시키기 위해서는 계산만을 수행하여 결과를 확인하는 과정을 수행해서는 부족하며, 논리적인 방법이나 흥미로운 실습을 통하여 의미를 이해하도록 하는 것이 중요하다. 또한 독립의 정의를 이용하여 독립성 여부를 확인하는 경우에도, 단순한 계산보다는 다음과 같은 확률나무 그림을 활용하면 독립의 의미를 논리적으로 이해하는데 많은 도움이 되리라 생각한다.

(예 1) 어떤 자동차 회사는 2000년도 상반기에 판매한 승용차에 대하여 다음과 같은 자료를 수집하였다. 대형차 중에서 80%는 큰 엔진을 장착했고, 중형차 중에서 70%는 작은 엔진을 장착했으며, 소형차 중에서 95%는 작은 엔진을 장착했다. 판매승용차의 20%는 대형차이며, 40%는 중형차이고, 40%는 소형차이다. 무작위로 선정한 승용차가 중형차라는 사건(A)은 무작위로 선정한 승용차가 큰 엔진을 장착한 차(B)라는 사건과 독립인가?



위의 확률나무 그림으로부터 $P(A|B)=\frac{0.12}{0.3}=0.4$, $P(A)=0.4$, 따라서 사건 A와 B는 독립사건이다.

III. 확률화응답기법의 활용

사건의 상호배반과 확률적 독립을 쉽고 흥미롭게 이해시키기 위한 일환으로, 확률화응답기법을 활용하는 과정에 대해서 알아보도록 하자. 본 논문에서는 Loynes(1976)이 제시한 강요형 모형을 이용하도록 하겠다.

사회적 관념상 인정되기 어려운 사안을 민감속성이라고 하는데, 학생들로 하여금 민감속성을 선택하여 민감질문을 만들도록 한다. 민감속성을 A(예컨대, 흡연이나 본드흡입 그리고 성 경험 등)라 할 때, 다음은 민감질문의 한 예가 된다.

"당신은 지난달(주)에 A 행동을 한 적이 있습니까?

다음 응답자의 신분보호를 위하여 확률장치를 구성하는데, 주위에서 흔히 구하기 쉬운 트럼프 카드 등을 이용할 수 있다. 한 예로 20장으로 구성된 카드를 준비하여, 8장에는 '예'라고 기입하고 나머지 12장에는 '민감질문에 응답'이라고 적은 다음 잘 섞는다. 실험과정에서 참여자는 구성된 확률장치 중에서 임의의 하나를 본인만 알도록 선택한 다음 확인 후, 다시 잘 섞어 놓는다. 이때 조사자는 참여학생으로부터 확률장치를 통한 "예" 또는 "아니오."의 응답만을 얻게된다. 참고로 확률장치를 구성하기 위한 도구로는 주사위, 지폐, 회전판 등이 이용될 수 있다.

이러한 실험(조사)에서 알고자하는 값은 모집단(학생집단)내의 민감속성(A)을 갖는 학생의 모비율이다. 이제 이를 추정하기 위하여 사건 B를, 정직하게 응답한다는 가정아래 확률장치를 사용하지 않고 민감질문을 직접질문 했을 때 민감속성 A를 갖는다고 응답할 사건이라고 하자. 그러면 추정의 대상은 확률장치를 통하여 얻어진 응답으로부터 사건 B의 확률 즉, $P(B)$를 추정하는 것이다.

이제 n명으로 구성된 실험참여 학생 중에서 임의의 한 학생이 확률장치를 통하여 선택된 카드에 대해서 정직하게 응답한다고 가정했을 때, “예”라고 응답할 경우를 생각해 보자. 이는 ‘예’라고 적힌 카드를 선택하거나, 민감속성을 가진 학생이 ‘민감질문에 응답’이라고 적힌 카드를 선택하는 경우이다. 전자의 경우를 나타낼 확률을 $P(Y)$라 하고, 후자의 경우를 나타낼 확률을 $P(S)$라 하자.

이 과정에서 실험참여자인 학생들은 ‘예’라고 적힌 카드나, ‘민감질문에 응답’이라고 적힌 카드 중에서 하나만 선택하도록 되어 있으므로, 사건의 상호배반(mutually disjoint)의 개념에 대한 설명이 가능하다. 한편 임의의 한 학생이 확률장치를 통하여 “예”라고 응답할 확률을 R이라 하면 이는 다음과 같다.

$$R = P(Y \cup S) = P(Y) + P(S) \qquad (3.1)$$

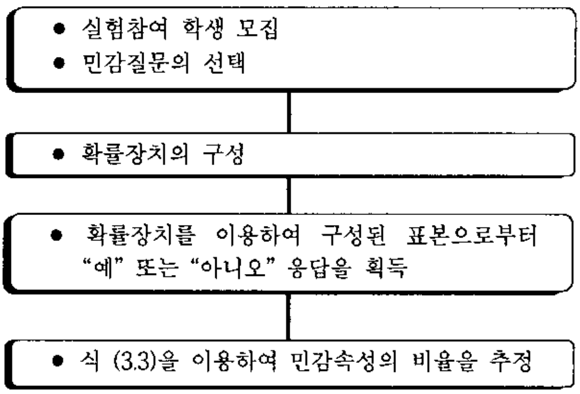
식 (3.1)을 자세하게 표현하기 위하여 확률장치내의 ‘예’라고 적힌 카드의 수를 y, 그리고 ‘민감질문에 응답’이라고 적힌 카드의 수를 s라 하자. 그러면 $P(Y) = \frac{y}{y+s}$ 이다. 한편 사건 T를 임의의 한 학생이 ‘민감한 질문에 응답’이라고 적힌 카드를 선택할 사건이라면 사건 S는 사건 T와 B의 교집합이다. 그런데 임의로 선택되는 카드의 내용이 실험참여 학생의 민감속성 보유여부에 영향을 미치지 않으므로, 이로부터 사건의 확률적 독립(probabilistic independence)의 개념에 대한 설명이 가능하다. 이로부터 $P(S) = P(T) \cdot P(B)$이며, $P(T) = \frac{s}{y+s}$ 이므로 $P(S) = (\frac{s}{y+s})P(B)$이다. 이를 이용하면 식 (3.1)의 R은 $R = \frac{y}{y+s} + (\frac{s}{y+s})P(B)$이 되며, 결국 구하고자 하는 $P(B)$는 다음과 같다.

$$P(B) = R\left(\frac{y+s}{s}\right) - \frac{y}{s} \qquad (3.2)$$

이제 N명의 모집단 중에서 단순임의추출된 n명의 표본(실험참여 학생)에 대해서 앞에서 구성된 확률장치를 사용하여, 이 중에 m명이 “예”라는 응답을 하였다면 $\hat{R} = \frac{m}{n}$ 이다. 따라서 구하고자 하는 $P(B)$의 추정치는 다음과 같다.

$$\hat{P}(B) = \frac{m}{n}\left(\frac{y+s}{s}\right) - \frac{y}{s} \qquad (3.3)$$

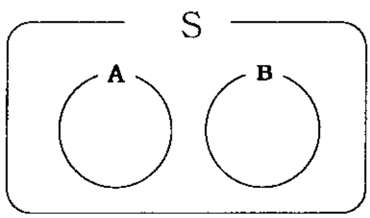
이제 식 (3.3)을 사용하여 민감속성의 비율을 추정할 수 있는데, 이러한 실험을 하는 과정에서 유의할 점이 있다. 첫째, 확률장치 구성 시, 민감질문이 적힌 카드의 수(s)를 몇 개로 해야 할 지에 대한 최적의 기준은 없다. 다만 실험 참여자의 신분보호를 고려하여 적절히 선택하도록 한다. 둘째, 모집단 내의 민감속성의 비율이 0에 가까울 때 식 (3.3)을 이용한 추정치는 음수가 나올 수도 있는데, 이 경우에는 n과 s의 크기를 조절하여 실험을 다시 실시하도록 해야한다. 참고로 확률화응답기법을 이용하여 실험을 실시하는 절차를 나타내보면 다음과 같다.



[그림 1] 확률화응답기법의 이용 절차

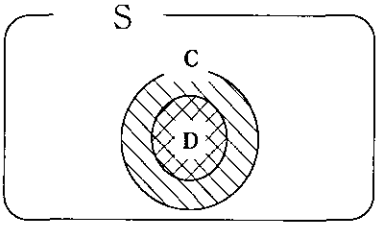
앞의 식 (3.3)을 유도하는 과정을 통하여 학생들에게 사건의 상호배반과, 독립의 개념에 대한 이해를 넓힐 수 있었다. 그러나 이에 대한 추가적인 설명이 요구된다면 다음의 상호배반과 독립의 관계를 이용함으로써, 효율적인 결과를 달성할 수 있으리라 생각한다.

1. 두 사건 A, B가 상호배반이면 이들은 독립적이지 않다.

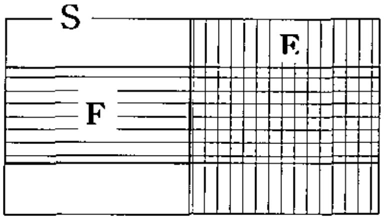


$P(A)>0,\ P(B)>0$ 그러나 $P(A|B) = \frac{P(A\cap B)}{P(B)} = 0$. 따라서 사건 A와 B는 독립이 아니다.

2. 두 사건이 상호배반이 아니면 독립적일 수도 있고, 아닐 수도 있다.



$P(C)<1$, $P(D)<1$ 그러나 $P(C|D)=1$. 따라서 사건 C와 D는 독립이 아니다.



$P(E)=\frac{1}{2}$, $P(F)=\frac{1}{2}$ 그런데 $P(F|E)=\frac{1}{2}$. 따라서 사건 E와 F는 독립이다.

IV. 결 론

고등학교 교육과정 수학 I에서 다루어지는 내용 중, 확률적 독립에 대한 개념은 처음으로 확률을 접하는 학생들에게 있어서는 그리 쉬운 개념은 아니다. 더욱이 각종 교과서에 나타난 이에 대한 설명은 공식을 이용하여 결과를 확인하는 방식으로 구성되어 있어, 개념에 대한 정확한 이해전달에 어려움을 주고 있다. 이에 본 논문에서는 확률화응답기법(randomized response technique : RRT)을 활용하는 연습을 통하여, 학습자 입장에서 비교적 이해하기 어렵다고 여겨지는 확률적 독립(probabilistic independence) 등의 몇 가지 확률이론을 참여실습을 통해 즐겁게 학습할 수 있는 방법을 소개함으로써, 일선에서 수학을 담당하는 선생님들에게 도움을 주고자 하였다. 앞에서 언급했듯이 동기개발과 흥미를 위하여 RRT를 활용하는 연습과정에서 학생들로 하여금, '민감질문의 구성'과 '확률장치의 구성' 그리고 '응답획득의 과정'을 직접 수행토록 하면 확률적 독립에 대한 이해의 증진과 더불어 학습의 효과가 증대되리라 기대된다.

참 고 문 헌

박두일, 신동선, 김기현, 박복현(1999). 수학 I, 서울: 교학사.

우정호(1998). 수학 I, 서울: 지학사.

윤옥경, 윤재한, 허원, 손문구, 송병희(1999). 수학 I, 서울: 중앙교육연구소.

이현구, 지동표, 김우철, 고성은, 박병욱, 장훈, 최용준(1997). 수학 I, 서울: 천재교육.

Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R. & Horvitz, D. G.(1969). The Unrelated Question Randomized Response Model Theoretical Framework, *Journal of American Statistical Association*, 64, 520-539.

Loynes, R. M.(1976). Asymptotically Optimal Randomizes Response Procedures, *Journal of American Statistical Association*, 71, 924-928.

Moors, J. J. A.(1971). Optimization of the Unrelated Question Randomized Response Model, *Journal of American Statistical Association*, 66, 627-629.

Warner, S. L.(1965). A Survey Technique for Eliminating Evasive Answer Bias, *Journal of American Statistical Association*, 60, 63-69.

Wonnacott, T. H. & Wonnacott, R. J.(1990). *Introductory Statistics*, John Wiley & Sons, Inc.

Understanding Probabilistic Independence using Randomized Response Technique

Choi, Kyung Ho (Jeonju University)
Kim, Rae Sun (Jeonju University)

Classroom exercise using the randomized response technique may be used to summarize a high school unit of instruction on probability. In this paper, we show that the derivation of the formula for this technique illustrates basic concepts in probability, such events being mutually disjoint and independent.