



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

강 태 훈 교수 지도
석사학위 청구논문

페이지안 네트워크의 능력모수
복원에 관한 모의실험 연구

2015

성신여자대학교 대학원
교육학과
김 덕 정

베이지안 네트워크의 능력모수
복원에 관한 모의실험 연구

강 태 훈 교수지도

이 논문을 석사학위논문으로 제출함.

2015년 5월


성신여자대학교 대학원


교육학과


김 덕 정

인 준 서

김덕정의 석사학위 논문으로 인준함.

심사위원장 조영민 인 

심사위원 조운정 

심사위원 강태훈 

성신여자대학교 대학원

연구 요약

본 연구의 목적은 인지진단 모형 중 하나인 베이지안 네트워크를 적용한 모의실험 연구를 통하여 피험자 능력추정의 정확성을 확인하는 것이다. 이를 위해 3,000명이 응답한 실제 수학 검사자료를 사용하여 모의실험 연구를 실시하였다. 검사는 총 20문항으로 대수학과 삼각함수라는 두 개의 인지요소를 가지고 있다. 실제 검사 자료를 활용하여 생성문항모수(CPT)를 추출하고, 이원정규분포를 활용하여 진능력모수를 생성하였다. 모의실험 요인으로는 ‘검사의 길이’와 ‘인지요소 간 상관 정도’를 고려하였다.

베이지안 네트워크에서 추정되는 능력모수란 각 인지요소 간 숙달 확률 및 숙달 여부를 의미한다. 베이지안 네트워크를 통해 추정한 능력모수의 정확성을 확인하기 위해 진능력모수 복원 관점에서 진숙달확률과 추정된 숙달확률 간의 순위 상관계수와 RMSE, 진숙달여부와 추정된 숙달여부 간의 분류정확도를 비교하였다.

연구 결과는 다음과 같다. 첫째, 검사 길이가 길어질수록 진숙달확률(p)과 추정된 숙달확률(\hat{p}) 간 순위 상관계수가 높아지고 RMSE가 낮게 나타났다. 마찬가지로 검사 길이가 길어질수록 진숙달여부(m)와 추정된 숙달여부(\hat{m})의 분류정확도가 높아졌다. 이를 통해 베이지안 네트워크를 적용한 능력모수 추정 시 검사 길이가 길어질수록 능력모수를 정확하게 복원한다는 것을 알 수 있다.

둘째, 인지요소 간 상관이 높아질수록 진숙달확률(p)과 추정된 숙달확률(\hat{p})의 RMSE는 약간씩 줄어드는 것으로 나타났으나 일정한 패턴은 보이지 않았다. 마찬가지로 인지요소 간 상관 정도에 따라 진숙달여부(m)와 추정된 숙달여부(\hat{m})의 분류정확도는 약간의 변화를 보였으나 일정한 패턴은 보이

지 않았다. 이를 통해 베이지안 네트워크를 적용한 능력모수 추정에서 인지 요소 간 상관 정도는 진능력 복원에 별다른 영향을 미치지 않는다고 볼 수 있다.

본 연구에서는 베이지안 네트워크를 적용한 모의실험 연구를 통하여 피험자 능력모수 추정의 정확성을 진능력모수 복원 관점에서 확인하고자 하였다. 진숙달여부와 추정된 숙달여부의 분류정확도가 상당히 높은 것으로 나타났다으며, 이러한 결과는 베이지안 네트워크를 활용한 능력 추정이 피험자의 다차원적인 능력을 비교적 정확하게 추정하는 것을 확인하는 계기가 되었다. 나아가 교육측정분야에서 피험자의 능력모수 추정 시 베이지안 네트워크의 적용 가능성을 확인할 수 있었다.

향후 베이지안 네트워크는 진단평가 또는 형성평가에서 피험자의 정확한 능력 또는 문항 모수 추정 시에 적용될 수 있으며, 피험자의 학업성취 프로파일 특성을 분석하는데 활용할 수 있을 것이다.

주제어 : 베이지안 네트워크, 능력모수 추정, 인지진단모형, 숙달 확률, 분류 정확도, 모의실험 연구

목 차

연구 요약

| | |
|--|----|
| I. 서론 | 1 |
| 1. 연구 목적 및 필요성 | 1 |
| 2. 연구 내용 | 4 |
| II. 이론적 배경 및 선행 연구 | 5 |
| 1. 이론적 배경 | 5 |
| 1) 확률 | 5 |
| 2) 베이지안 네트워크 | 9 |
| 3) 인지진단이론 | 16 |
| 2. 선행 연구 | 18 |
| III. 연구 방법 | 25 |
| 1. 모의실험 설계 | 25 |
| 2. 모의실험 자료의 생성 및 추정 | 28 |
| 3. 베이지안 네트워크의 능력모수 추정 결과에 대한 평가 | 29 |
| IV. 연구 결과 | 31 |
| 1. 진속달확률과 추정된 속달확률 간 순위 상관계수와 RMSE | 31 |
| 1) 대수학의 순위 상관계수와 RMSE | 31 |

| | |
|--|----|
| 2) 삼각함수의 순위 상관계수와 RMSE | 33 |
| 3) 두 인지요소의 진속달확률과 추정된 속달확률의 RMSE 분산분석 ... | 35 |
| 2. 속달여부 분류 정확도 | 37 |
| 1) 대수학의 속달여부 분류정확도 | 37 |
| 2) 삼각함수의 속달여부 분류정확도 | 38 |
| 3) 두 인지요소의 진속달여부와 추정된 속달여부의 분류정확도 분산분석 | 39 |
| V. 결론 및 논의 | 42 |
| 1. 결론 | 42 |
| 2. 논의 | 43 |

참고문헌

ABSTRACT

부 록

표 목 차

| | |
|--|----|
| <표 II-1> 인지요소의 숙달여부에 따른 문항 반응 | 13 |
| <표 II-2> 다차원 인지요소 숙달여부에 따른 문항 반응 | 15 |
| <표 III-1> 수학 진단평가의 CTT 문항통계치 | 26 |
| <표 III-2> 수학 진단평가의 Q행렬 | 27 |
| <표 IV-1> 대수학의 진숙달확률과 추정된 숙달확률의 순위 상관 평균 | 32 |
| <표 IV-2> 대수학의 진숙달확률과 추정된 숙달확률의 RMSE 평균 | 33 |
| <표 IV-3> 삼각함수의 진숙달확률과 추정된 숙달확률의 순위 상관 평균 | 34 |
| <표 IV-4> 삼각함수의 진숙달확률과 추정된 숙달확률의 RMSE 평균 | 34 |
| <표 IV-5> 검사 길이와 인지요소 간 상관에 따른 RMSE 분산분석 결과(대수학) | 35 |
| <표 IV-6> 검사 길이와 인지요소 간 상관에 따른 RMSE 분산분석 결과(삼각함수) | 36 |
| <표 IV-7> 대수학의 진숙달여부와 추정된 숙달여부의 분류정확도 평균 | 37 |
| <표 IV-8> 삼각함수의 진숙달여부와 추정된 숙달여부의 분류정확도 | 38 |
| <표 IV-9> 검사 길이와 인지요소 간 상관에 따른 분류정확도 분산분석 결과(대수학) | 40 |
| <표 IV-10> 검사 길이와 인지요소 간 상관에 따른 분류정확도 분산분석 결과(삼각함수) | 40 |
| <표 V-1> 부정적인 문항모수(추측도, 부주의오류) | 45 |

그림 목차

| | |
|--|----|
| [그림 II-1] 연속 확률변수를 확률 밀도 함수로 변환하는 과정 | 8 |
| [그림 II-2] 베이지안 네트워크의 기본 모형 예시 | 11 |
| [그림 II-3] 조건부확률 분포로 표현한 베이지안 네트워크 예시 | 20 |
| [그림 II-4] 베이지안 네트워크 모델의 예시 | 21 |
| [그림 IV-1] 대수학과 삼각함수의 검사 길이에 따른 속달확률 RMSE | 36 |

I. 서 론

1. 연구 목적 및 필요성

우리는 복잡한 세상을 확률과 함께 살고 있다. 일기 예보를 통해 내일 날씨에 맞는 옷을 선택하고, 정기적인 검진을 통해 본인의 건강에 이상이 없는지 확인한다. 그러나 비가 내릴 확률이 적다고 예보와 달리 갑자기 비바람이 몰아칠 수 있고, 건강에 이상 없다는 검진 결과를 받았지만 갑자기 암에 걸려 3개월 정도밖에 살 수 없을 수도 있다. 이처럼 우리가 살아가고 있는 이 세상은 복잡하고 불확실한 것으로 가득 차 있어 발생하는 수많은 사건에 대해 결과를 쉽게 예측할 수 없으며, 예측한다 하더라도 이를 100% 확신하기 어렵다.

Miller와 Miller(2003)는 통계학이란 불확실한 상황 속에서 의사결정에 관한 문제를 다루는 학문이라 하였다. 이를 다시 말하면 어떤 결과를 예측하기 위한 배경 지식과 원인 변수들을 충분히 확보하지 못하면 예측의 불확실성은 점점 커질 수밖에 없다는 것이고, 반대로 사전에 정보가 충분히 수집되었다면 예측의 정확성을 높일 수 있다는 것이다. 예를 들어 오늘 날씨와 비슷한 과거의 온도와 습도 데이터를 활용해서 비가 올 확률을 계산하거나, 암환자들이 병원을 찾기 전 경험했던 증상들을 광범위하게 조사하여 건강 검진 환자의 암 발병 확률을 계산하는 경우와 같이 가용한 정보의 활용 및 분석을 통하여 예측 결과의 정확성을 높이고자 할 때 통계학은 의사 결정 시 매우 유용한 수단으로서 기능할 수 있다.

교실의 상황을 떠올려보면, 미분 능력을 측정하는 문항에서 방정식과 함수를 잘 알고 있는 학생은 그렇지 못한 학생보다 이 문항에 정답할 확률이 더 높을 것이다. 이처럼 확률은 기존의 지식과 정보에 의해 추정되며 주어

진 조건, 여기서는 피험자 개개인의 인지 숙달 정도에 따라 정답할 가능성의 정도가 변화하는 특징이 있다.

베이저안 추론(Bayesian inference)은 관찰된 자료와 더불어 과거의 경험 및 사전 지식 등과 같은 사전 정보를 활용하여 미지의 모수를 추정하기 때문에 고전적 추론보다 더 정확하게 추론 할 수 있다는 장점이 있다. 실제로 베이저안 네트워크는 예측의 정확성이 높아 의사 결정 시 유의미한 정보를 제공해야 하는 의료, 기상예측 등의 분야에서 널리 활용되고 있다(정용규 외, 2001; 김달호, 2013; 남기훈, 2014; Acid 외, 2004; Elsner, Bossak, 2001).

최근 교육측정 분야에서는 피험자 개개인에 대한 다양한 정보를 제공하는 장점 때문에 인지진단이론에 대한 관심이 급증하고 있다. 인지진단이론이란 피험자가 문항에 정답하기 위해 요구되는 특정 인지요소(attribute)의 숙달 정보를 제공하는 측정 이론이다. 인지진단이론은 인지요소 숙달 여부에 대한 구체적인 정보를 제공하여 피험자의 이해 상태를 정확하게 파악할 수 있고, 이러한 정보를 기반으로 교수·학습활동에 직접적으로 활용할 수 있다(김수진, 송미영, 김선희, 2008; 김수진, 2010; 김명화, 2012; 김지효, 2013; DiBello, Stout, & Roussos, 1995; Tatsuoka, 1983, 1995; Hartz, 2002).

강력한 인지진단 모형 중 하나인 베이저안 네트워크는 다른 인지진단 모형에 비해 변수 간 관계 규정과 다양한 형태의 인지요소 간 관계 등을 유연하게 표현할 수 있는 장점에도 불구하고, 폭넓은 활용이 이루어지지 못하고 특정 분야에서만 제한적으로 활용되고 있다(강태훈, 2014). 특히 교육 분야에서는 베이저안 네트워크 관련 연구를 찾아보기 쉽지 않다. 근래 국내외 교육 분야에서 베이저안 네트워크가 활용된 사례를 살펴보면 다음과 같다.

김성호(1994), 김성호와 김은희(1995)는 과제 수행과 과제 수행 능력 구조(관계)를 표현할 수 있는 네트워크 모형화를 탐색하는데 중점을 두고 베이저안 네트워크를 활용하였다. 박상희 외(2009)는 결측치가 많은 영과잉자료,

소표본 자료를 자주 접하게 되는 연구 집단의 특수성을 고려하여 장애 판정에 활용할 수 있는 베이지안 분석 방법을 소개하고, 특수교육분야에서의 적용가능성을 탐색한 바 있다. 황선아(2005), 나선웅, 김경수, 최용석(2012)은 일차원성 가정으로 인해 학습의 위계를 무시한다는 문항반응이론의 제한점을 보완하기 위해 베이지안 네트워크를 적용한 컴퓨터기반 적응형 평가(CAT)에 대해 연구하였다. 그 결과 베이지안 네트워크를 적용한 CAT는 문항반응이론의 제한점을 해결하면서 학습자의 실제 능력치에 수렴하는 속도도 증가했다고 보고하였다.

Vomlel(2004)의 연구는 베이지안 네트워크를 적용한 컴퓨터 기반 적응 검사(CAT가)가 비교 방법보다 적은 수의 문항으로 피험자의 능력을 정확하게 진단하는 결과를 보여주었다. Almond 외(2007)는 베이지안 네트워크는 진단 평가 측면에서 특별한 기술 없이 사용 가능하고, 능력 노드 사이의 숙달 정보 제공하며 확률로 이루어진 네트워크 그래픽을 제공하는 등과 같은 여러 장점을 소개하면서 교육측정분야 실무자에게 베이지안 네트워크의 사용을 제안하였다. García, Amandi, & Campo(2007)는 대학생을 대상으로 웹 기반의 환경에서 학생과 시스템이 교수학습 과정에서 상호작용하는 패턴을 분석하여 학생들의 학습습관을 모델링하였다. 베이지안 네트워크를 적용하여 개인의 학습유형을 진단하고, 학습유형에 적합한 학습 방법을 안내 및 보조 학습자료를 제공하는 역할을 할 수 있다고 소개하였다.

강태훈(2014)의 연구에서는 기존의 고전검사이론과 문항반응이론에 의한 문항분석 결과와 베이지안 네트워크를 활용한 문항 분석 결과를 비교하였다. 그 결과 베이지안 네트워크를 통한 문항분석 결과는 고전검사이론과 문항반응이론의 결과와 상당한 수준으로 일치한 결과를 얻었다고 보고한 바 있다.

앞에서 살펴본 교육 분야에서의 베이지안 네트워크 관련 선행연구들은 다

양한 정보 제공, 사전 정보를 확률 변수로 이용으로 인한 예측의 정확성 등과 같은 장점들을 언급했지만, 정작 일반 교사나 검사 결과 사용자들의 관심이라고 할 수 있는 피험자의 능력, 즉 피험자의 인지요소 숙달정도가 정확하게 추정될 수 있는지에 대한 관심은 거의 없는 것을 알 수 있다.

이에 따라, 본 연구에서는 인지진단 모형 중 하나인 베이지안 네트워크가 실제로 피험자의 능력 모수를 정확하게 추정하는지 진능력모수 복원(recovery) 관점에서 확인해 보고자 한다. 이를 위해 다양한 모의실험 조건 속에서 베이지안 네트워크를 적용한 자료 생성 및 분석 과정을 적용하여 피험자의 능력 모수 추정의 정확성을 평가해 보고자 한다.

2. 연구 내용

본 연구의 주된 목적은 피험자의 성취도 검사 자료가 주어졌을 때, 베이지안 네트워크 방법론을 활용하여 능력 모수를 추정 시 추정의 정확성을 확인하는데 있다. 이를 위해 진행되는 모의실험 연구 내용은 다음과 같다.

첫째, 베이지안 네트워크를 활용하여 실제 검사 자료로부터 문항모수를 추정한다.

둘째, 앞에서 추정된 문항모수를 생성모수(generating parameters)로 하고, 다변인정규분포에 기반한 진능력모수(true ability parameters)를 생성한 다음, 가상의 피험자 모의실험 자료(simulated datasets)를 생성한다.

셋째, 모의실험 자료에 베이지안 네트워크를 적용하여 피험자의 능력 모수를 추정한 후, 이를 진능력 모수와 비교하여 복원의 정확성을 평가한다.

넷째, 모의실험 연구 결과를 정리하여 교육 측정 분야에서 베이지안 네트워크 적용 가능성을 논의한다.

II. 이론적 배경 및 선행 연구

1. 이론적 배경

1) 확률(Probability)

확률이란 어떤 사건이 일어나리라고 기대되는 불확실성의 정도를 수학적 언어로 표현한 것이다. 확률은 기존의 지식과 정보에 의해 정해지며 주어진 조건에 따라 사건이 일어날 가능성의 정도가 0과 1사이에서 변화하는 성질이 있다.

주사위나 동전을 던지는 통계적 실험과 같이 각 실험 결과가 일어날 가능성이 동일한 경우 사건 A가 일어난 확률은 다음과 같이 정의할 수 있다.

$$P(A) = \frac{\text{사건 } A \text{에 속하는 원소의 개수}}{\text{표본공간(sample space)의 전체 원소의 개수}}$$

확률의 공리(axiom, 이론에서 가장 기초가 되는 명제)적 정의는 세 가지로 구성되어 있으며, 이 세 가지 공리를 만족할 때 $P(A)$ 를 사건 A의 확률이라 한다.

표본공간 S에서의 임의의 사건 A에 대하여

(a) 모든 x 의 값에 대하여 $0 \leq P(A) \leq 1$ 이다.

(b) $P(S) = 1$ 이다.

(c) 상호 배반인 사건 A, B에 대하여 $P(A \cup B) = P(A) + P(B)$ 이다.

(a) 임의 사건 A에 대하여 확률은 음수가 될 수 없으며, 0보다 작을 수 없고 1보다 클 수 없다. 즉 A사건이 일어날 가능성은 0과 1사이에 존재한다. (b) 어떤 실험을 할 때 발생한 사건이 표본 공간의 원소 중 하나일 확률은 1이다. (c) 상호 배반인 두 사건의 합집합의 확률은 각 사건의 확률의 합과 같다. 상호 배반 사건이란 동시에 일어날 수 없는 사건을 의미하는데, 가설 검정 시 활용하는 귀무가설(H_0)과 연구가설(대립가설, H_a)이 상호 배반적 관계의 대표적인 예라 할 수 있다. (c)의 공리를 확장하면 상호 배반인 개별 사건들이 발생할 확률을 모두 더하면 1이 된다.

일반적으로 확률은 크게 객관적 확률(objective probability)과 주관적 확률(subjective probability)로 구분할 수 있다.

객관적 확률은 어떤 사건이 발생할 가능성을 사전적으로 판단하여 얻는 확률과 사후적으로 상당히 많은 실험을 반복함으로써 어떤 사건이 발생할 상대도수를 관찰하여 얻는 확률로 구분할 수 있다. 동전을 한 번 던졌을 때 일어날 가능성이 동일한 경우, 앞면이 나올 확률이 $\frac{1}{2}$ 이라는 것은 발생 가능성을 사전적으로 판단하여 얻은 확률이라 한다. 우리나라 90세 이상 노인의 평균 수명과 같이 모집단이 무엇인지 알지 못할 때 표본을 이용하여 모집단의 특성을 파악하는 것을 상대도수에 기반한 확률이라 한다.

다음으로 주관적 확률이란 의사결정자가 자신의 지식이나 경험에 근거하여 주관적으로 어떤 사건이 일어날 가능성에 부여한 일정한 값이다. 주관적 확률은 개인의 경험에 바탕을 둔 주관적 판단에 의존하기 때문에 부정확하다고 생각할 수 있으나 관련 자료를 지속적으로 수집·활용한다면 확률의 정확성을 높일 수 있다. 내일 비가 올 확률 80% 이상, 또는 암에 걸릴 확률 3% 미만, 미분 문제에 정답할 확률 79% 등이 주관적 확률의 예라 할 수 있다.

어떤 사건 x가 일어났을 때 사건 y가 일어날 확률을 조건부확률(conditional probability)이라 한다. 예를 들어 덧셈과 뺄셈의 인지요소를 알

고 있다는 전제 하에서 성취도 검사의 빨셈 관련 문항에 정답할 사건이 일어날 확률을 조건부확률이라 하며, 이는 $P(y | x)$ 로 나타낼 수 있다. 조건부 확률분포(conditional probability distribution)는 한 확률 변수의 값이 주어졌을 때 다른 확률변수의 분포를 의미하고, $X=x$ 일 때 Y 의 조건부확률분포는 식 (1)로 나타낼 수 있다.

$$f(y|x) = \frac{f(x,y)}{f(x)} \quad f(x) > 0 \quad \text{식 (1)}$$

베이저안 통계학에서 많이 사용되는 확률 분포 중 본 연구에서 사용한 분포를 중심으로 정리해 보았다. 대표적인 이산형 확률분포는 베르누이 분포, 이항 분포가 있고, 대표적인 연속형 확률분포에는 균일분포, 정규분포 등이 있다.

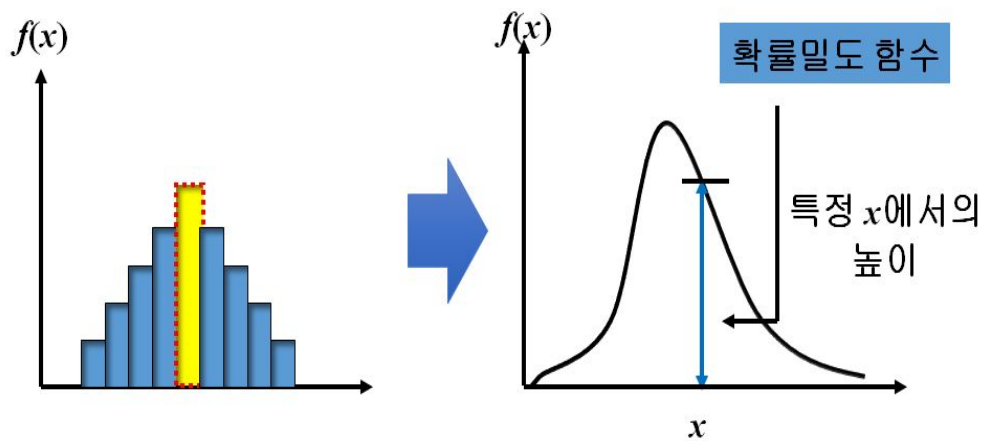
① 베르누이 분포(bernoulli distribution)란 확률 실험의 결과가 흡연 여부, 합격 여부 등과 같이 0과 1의 값을 갖는 사건에서 한 번의 실험을 실시했을 때($X=1$) 성공 또는 실패할 확률을 나타내는 분포이다. 베르누이 분포에서 성공의 확률은 p 로 표현한다면, 실패의 확률은 확률의 공리적 정의 (C)를 활용하여 $1-p$ 로 표현할 수 있다. 확률변수 X 는 식 (2)와 같이 베르누이 분포를 따른다.

$$P(X=x) = f(x) = p^x(1-p)^{1-x}, \quad x = 0,1 \quad \text{식 (2)}$$

② 베르누이 분포가 단 한 번의 실험에 의한 확률 분포라면, 베르누이 시행을 n 번 반복한 결과에 의한 확률 분포를 이항분포(binomial distribution)라 한다. 이항분포는 베르누이 시행을 독립적으로 n 번 반복했을 때 성공의 횟수를 X 라고 한다면, X 가 식 (3)과 같이 이항분포를 따를 때 $X \sim B(n, p)$ 로 나타낼 수 있다.

$$P(X=x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1 \quad \text{식 (3)}$$

이산 확률변수와 달리 연속 확률변수는 정수로 셀 수 없기 때문에 성공 횟수 대신 성공 구간의 면적을 확률로 구한다. 연속된 확률변수는 [그림 II-1]과 같이 확률 밀도 함수로 변환하여 구한다.



[그림 II-1] 연속 확률변수를 확률 밀도 함수로 변환하는 과정

③ 확률변수 X 의 확률밀도 함수가 식 (4)와 같을 때, X 는 균일분포 (uniform distribution)를 따른다고 하며 $X \sim U(a, b)$ 로 나타낸다.

$$f(x) = \frac{1}{b-a}, \quad a < x < b \quad \text{식 (4)}$$

④ 확률변수 X 의 확률밀도함수가 식 (5)로 주어지면, X 는 평균이 μ 이고,

분산이 σ^2 인 정규분포(normal distribution)를 따른다고 하며 이를, $X \sim N(\mu, \sigma^2)$ 로 나타낼 수 있다.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad \text{식 (5)}$$

정규분포의 형태는 평균 μ 과 분산 σ^2 에 의해 완전히 결정되며, 평균이 0이고, 분산이 1인 경우를 표준정규분포(standard normal distribution)라고 하고, $X \sim N(0, 1)$ 로 표현한다.

2) 베이저안 네트워크(Bayesian Network)

베이즈정리에 기반한 베이저안 추론은 조건부확률의 정의를 확장하여 유도한 것으로 주관적 확률에 근거해 만들어진 이론이다. 베이저안 확률의 목표는 사전확률과 조건부확률로부터 사후확률(Posterior Probability)을 구하는 것이다. 모수치 θ 를 추정하기 위하여 y 라는 응답 자료를 얻었다고 가정할 때, $p(y)$ 는 y 라는 자료를 얻을 확률이고, $p(\theta)$ 는 y 라는 자료를 얻기 전에 이미 알고 있는 사전 정보, 과거 데이터와 같이 미리 주어진 사전 정보를 의미한다. 조건부 확률은 모수치 θ 가 주어졌을 때 y 가 발생할 확률로 $p(y | \theta)$ 로 나타낸다. 사후확률이란 새로운 정보를 통하여 사전확률에 조정을 가한 확률이라 할 수 있는데, 여기에서는 y 라는 자료가 주어졌을 때 모수치 θ 를 추정하는 것으로 $p(\theta | y)$ 의 확률분포를 나타낸다.

이처럼 사전확률과 조건부확률을 이용하여 사후확률을 구하는 것을 베이즈정리(bayes theorem)라고 하며, 식 (6)과 같이 나타낼 수 있다.

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)} \quad \text{식 (6)}$$

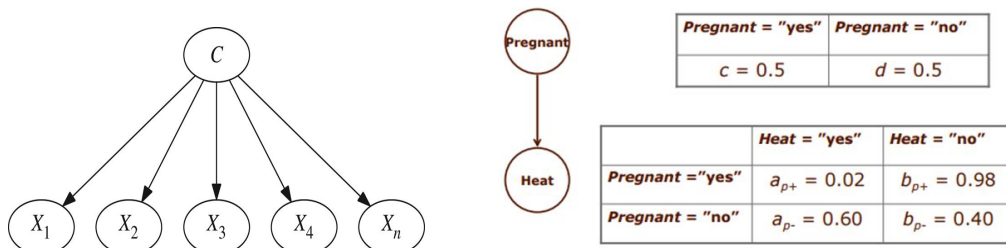
베이저안 통계의 구조는 크게 세 가지이다. 먼저 (a) 모수의 사전분포 결정하고, (b) 데이터의 확률모형과 사전분포를 이용하여 사후분포의 계산한 다음, (c) 사후분포를 이용하여 모수에 대한 추론을 하는 과정이다. 베이저안 확률에서 중요한 개념은 ‘주관적 사전 확률에 의한 사전분포’이다. 사전 확률은 모수에 대한 주관적 견해로서 믿음의 정도(degree of belief)를 수량화한 것이기 때문에 객관적으로 결정하기 쉽지 않다는 제한점이 존재한다. 그렇지만 사전 확률은 모수를 정확하게 추정하는 과정에서 매우 중요한 역할을 하고 있기 때문에, 정확한 사후 확률을 추정하기 위해서는 적절한 사전 분포를 선택하는 것이 중요할 것이다.

베이저안 추론은 전통적 통계학보다 결과의 해석이 쉽고, 많은 실제 문제들에서 모수의 사전 정보(예를 들어 모집단 분산 추정)를 쉽게 구할 수 있으며, 복잡한 상황에서 모수를 쉽게 추론할 수 있다는 장점이 있다.

전통적 통계학에서 많이 쓰이는 최대가능도추정(maximum likelihood method, MLE)은 가능도함수가 복잡하거나 많은 최소극대값을 가지는 경우가 많아 추정치의 계산이 어려울 뿐 아니라 일반적으로 추정량의 성질이 나쁠 때가 많다. 이와 대조적으로 베이저안 통계학에서는 사전분포와 가능도함수를 이용하여 사후분포를 추정하는 마르코프 연쇄 몬테카를로(markov chain monte carlo, MCMC) 기법을 사용한다. MCMC 방법은 모수의 사후 분포를 쉽게 구할 수 있을 뿐 아니라 사후 분포를 통한 모수의 추론이 전통적 방법보다 더 정확하다는 장점이 있다(Gilks, Roberts, 1996; Gamerman, Lopes, 2006).

베이저안 네트워크(Bayesian Network)는 여러 가지 변수들의 결합된 확률

을 그래픽 기반의 모형으로 표현한다. [그림 II-2]의 왼쪽 그래프와 같이 하나의 인지요소(C)와 각각의 문항들(X_1, X_2, \dots, X_n)은 잠재적인 테이블 $P(C | X_1, \dots, X_n)$ 와 연결되어 있으며 변수와 함께 지시선은 방향성 비순환 그래프 (directed acyclic graphs, DAGs)를 형성하고 있다. 베이저안 네트워크는 [그림 II-2]의 오른쪽 그래프와 같이 주어진 문제를 구성하는 변수들 간 존재하는 확률적 의존관계의 방향을 호(arc)로 나타내고, 각 변수들이 갖는 조건부 확률(conditional probability)을 계산함으로써 변수들간의 인과관계를 고려할 수 있다.



[그림 II-2] 베이저안 네트워크의 기본 모형 예시(Jensen, 1996)

베이저안 네트워크가 갖는 큰 장점은 다수의 독립적인 변수들 간의 관계를 명확하게 표현할 수 있다는 점이다. 이는 변수 간의 관계를 비교적 간략한 시각화를 유지하면서 무선 변수들간의 복잡한 관계를 모델링하여 도식화된 수학적 구조를 제공하는 것이다. 그렇기 때문에 베이저안 네트워크는 특별한 기술 없이도 사용이 가능하고, 시각화된 결과는 해석에 용이하다(Nachman, 2004; Jensen, 1996; Gogate et al, 2012).

베이저안 네트워크의 유형에는 가장 일반적인 형태인 naive Bayesian network(NBN), 트리 형태로 확장한 tree augmented naive Bayesian

network(TAN), NBN을 확장한 BN augmented naive Bayesian network (BAN), 가장 기본적인 형태인 general Bayesian network(GBN) 등이 있으며, 본 연구에서는 GBN을 사용하여 피험자의 능력모수를 추정하였다. GBN은 다른 베이지안 네트워크 유형과는 달리 종속변수와 독립변수들의 차이를 두지 않고 모든 변수(노드)들 간의 상호의존성을 하나의 베이지안 네트워크로 표현하는 것이 특징이다.

베이지안 네트워크는 양질의 정보를 제공한다. 네트워크 구조라는 질적 성분과 각 노드에서 정의되는 조건부확률표(conditional probability table, CPT)라는 양적 성분을 가지고 있다.

베이지안 네트워크의 질적 정보인 네트워크 구조는 검사자료에 대한 분석에서 인지요소와 문항 간의 관계 또는 인지요소들 간 상호관계를 의미한다. 이러한 인지요소와 문항 간 구조는 자료를 통하여 추정될 수도 있지만, 일반적으로 학교 현장의 교사 또는 교수진과 같은 내용전문가에 의해서 정의된다. 학습자들의 인지 능력을 정확히 파악하기 위해서는 인지 속성의 바탕이 되는 Q행렬을 정의하는 것이 중요하다. Q행렬의 인지 속성 즉, 인지요소의 숙달여부 분석 결과를 토대로 피험자 개인의 강·약점을 파악할 수 있고, 이에 따라 필요한 영역을 중심으로 보충학습을 진행할 수 있다. 따라서 인지 속성과 문항 간의 관계를 설명하는 Q행렬의 구성은 피험자의 능력을 정확히 진단하는 데 있어서 매우 중요한 절차라 할 수 있다.

베이지안 네트워크의 양적 성분인 CPT는 노드 상호 간의 관계를 구체적인 수치로 나타낸 것이다. 예를 들어 a1, a2라는 두 개의 인지요소의 숙달여부를 진단하는 검사에서 'a1을 숙달하지 못한 경우 특정 문항을 맞힐 확률' 혹은 'a1과 a2의 인지요소를 모두 숙달한 경우 특정 문항을 맞힐 확률' 등이 조건부 확률값이라 할 수 있다.

네트워크 구조를 결정하고 CPT 추정이 완료되면, 검사 실시 결과인 피험

자의 응답 반응에 따라 각 인지요소의 숙달확률을 추정할 수 있다. 피험자 개개인의 숙달 확률은 일종의 문항모수 역할을 하는 CPT 내 조건부확률값을 이용하여 확인할 수 있으며 각 검사 문항의 곤란도 및 변별도, 추측도와 같은 문항 모수를 추정할 수 있다.

하나의 인지요소를 측정하는 문항 i 에 대하여 인지요소 숙달여부에 따른 문항 반응은 <표 II-1>과 같이 나타낼 수 있다. 인지 요소의 숙달 여부에 따라 일반적인 문항 반응을 보이는 피험자라는 가정 아래, 문항 i 와 관련된 인지요소를 숙달하지 못한 학생은 오답(p_{00})이라는 반응을 보일 것이고, 반대로 인지요소를 숙달한 학생은 정답(p_{11})이라는 반응을 보일 것이다.

<표 II-1> 인지요소의 숙달여부에 따른 문항 반응(강태훈, 2014 p.903)

| | 문항 i | |
|------|----------|----------|
| | 오답(=0) | 정답(=1) |
| 미숙달자 | p_{00} | p_{01} |
| 숙달자 | p_{10} | p_{11} |

문항 분석 시 가장 기본이 되는 문항 모수는 문항곤란도와 문항변별도이다. 고전검사이론(Classical Test Theory, CTT)에서 일반적으로 사용하는 문항 곤란도(Item Difficulty)란 문항의 쉽고 어려운 정도를 나타내는 것으로 전체 피험자 중 해당 문항에 정답한 피험자의 비율(정답자 수/전체 피험자 수)로 정의할 수 있다. 문항곤란도(P)값이 클수록 쉬운 문항, 작을수록 어려운 문항으로 해석할 수 있으며 문항을 평가하는 절대적인 기준은 없으나, 대개 P 값이 0.2보다 작거나 0.8보다 클 경우 너무 어렵거나 너무 쉬운 문항으로 해석한다(황정규, 1998).

<표 II-1>에서 문항곤란도란 인지요소를 숙달한 학생이 정답할 확률로

정의할 수 있다. 즉, 전체 피험자 집단에서 해당 인지요소를 숙달한 학생 중 정답한 비율 즉, p_{11} 의 값이 문항곤란도를 의미한다.

CTT에서는 피험자를 능력에 따라 상위능력집단과 하위능력집단을 구분하여 각 집단의 정답 비율 차이를 비교하여 문항 변별도(Item Discrimination)를 계산한다(Johnson, 1951). 상위 집단능력에서의 정답 비율과 하위 능력집단에서의 정답 비율을 비교하여 변별도를 추정하는 방법은 식 (7)로 나타낼 수 있다.

$$DI = \frac{R_U - R_L}{N} \quad \text{식 (7)}$$

N : 각 집단의 피험자 수, R_U : 상위 능력 집단, R_L : 하위 능력 집단

피험자의 능력을 나누는 기준은 상위 27%, 하위 27%로 집단을 구분하는 방식이 널리 쓰이고 있으며, 상위 50% 하위 50%로 구분하는 방식, 혹은 준거점수가 아닌 총 피험자 수가 같도록 집단을 구분하는 방식 등이 있다(황정규, 1998; Ebel, 1965; Kelly, 1939).

능력이 높은 집단과 낮은 집단의 정답률 차이가 크다면 그 문항은 집단의 능력을 잘 구분하는 변별도가 높은 문항으로 해석할 수 있으며, 반대로 능력이 높은 집단과 낮은 집단의 정답률이 비슷하다면 그 문항의 변별도는 0에 가까워 집단을 변별하는 기능이 없는 문항으로 해석할 수 있다. 즉, 문항 변별도 DI 값이 클수록 변별력이 높은 문항, 작을수록 변별력이 낮은 문항이다. 문항곤란도와 마찬가지로 문항변별도 역시 절대적인 기준은 없으나, 대개 DI 값이 0.2이상일 경우 변별도가 있는 문항이라 한다(황정규, 1998).

CTT에서 문항변별도를 상위능력집단과 하위능력집단의 정답 비율 차이로 정의한다면, 베이지안 네트워크에서의 문항변별도란 <표 II-1>에서 p_{11}

과 p_{01} 의 차이라고 할 수 있다. 즉, 인지요소를 숙달한 피험자가 정답할 확률과 미숙달한 피험자가 정답할 확률의 차이가 문항변별도인 것이다.

$$BN_DI = p_{11} - p_{01} \quad \text{식 (8)}$$

그러나 두 개 이상의 인지요소를 고려해야 하는 다차원성의 검사라면 문항 분석을 위해 고려해야 되는 모수가 많아진다는 단점이 존재한다. 예시로 $31 \div 4$ 라는 문항에서 빨셈, 빌려오기, 나눗셈이라는 3개의 인지요소를 고려해야 한다면 CPT 행의 수는 $2^3=8$ 이 되어야 한다. 고려해야 할 인지요소 수가 더 많아진다면 문항변별도나 문항곤란도를 추정하기가 다소 복잡해 질 수 있다. 그러나 <표 II-2>와 같이 2 잠재집단(2 latent class, 2LC) 모형을 활용한다면 다차원의 인지요소를 고려하는 문항에서도 단일차원과 마찬가지로 간단하게 변별도와 곤란도를 산출할 수 있다(강태훈, 2014).

<표 II-2> 다차원 인지요소 숙달여부에 따른 문항 반응(강태훈, 2014, p.904)

| | 문항 i | |
|--------------------|----------|----------|
| | 오답(=0) | 정답(=1) |
| 미숙달자 ¹⁾ | p_{00} | p_{01} |
| 숙달자 ²⁾ | p_{10} | p_{11} |

¹⁾ 해당 문항을 해결하기 위해 요구되는 인지요소 중 하나라도 숙달하지 못한 자

²⁾ 해당 문항을 해결하기 위해 요구되는 모든 인지요소를 숙달한 자
(여기서는, 빨셈, 빌려오기, 나눗셈의 인지요소를 모두 숙달한 자)

<표 II-2>와 같이 3개의 인지요소를 필요로 하는 다차원성 문항 i 에서 문항곤란도란 3개의 인지요소를 모두 숙달한 피험자 중 문항에 정답한 비율

(p_{11})을 의미한다. 3개의 인지 요소 중 어느 하나라도 숙달하지 못한 자 즉, 인지요소를 하나도 숙달하지 못하거나 1개 또는 2개의 인지요소를 숙달한 피험자를 미숙달자로 분류한다. 문항변별도는 미숙달자가 정답할 비율(p_{01})과 숙달자가 정답할 비율(p_{11})의 차이를 계산하여 구할 수 있다.

3) 인지진단이론(Cognitive Diagnosis Theory)

계산이 쉽고 분석 결과를 쉽게 이해할 수 있다는 장점 때문에 CTT는 현장에서 많이 쓰이고 있지만, 문항의 곤란도와 변별도가 집단의 특성에 따라 달라지며 수험생의 능력도 시험의 특성에 따라 다르게 추정된다는 단점이 있다(성태제, 2009). CTT와 차별화되는 문항반응이론(Item Response Theory, IRT)은 단일차원성의 가정에 따라 단일한 연속체 위에 피험자의 능력을 표시할 수 있으며, 문항의 곤란도도 동일한 척도상에 함께 존재하여 문항 모수와 피험자 모수가 동일한 척도상의 값으로 표현될 수 있다(황정규 외, 2011, p.241). 그러나 IRT는 일차원성과 지역독립성이라는 강한 가정으로 인하여 문항 분석 시 적용이 어렵다는 단점이 있다. 또한 교육심리검사에서 하나의 문항이 하나의 능력을 측정하는 것보다 다양한 차원의 능력을 재고 있는 경우가 일반적이라 현실적으로 일차원성의 가정을 충족시키기 어렵다는 제한점이 있다.

검사 결과는 피검사자인 학생의 학습에 도움이 되어야 하며 이를 지도하는 교사에게도 유용한 정보를 제공하여야 하는데 인지진단이론은 이러한 요구를 충분히 수용할 수 있다. Embretson(1999)과 Stout(2002)는 인지진단모형을 21세기 새로운 평가 패러다임이라고 선언하기도 하였으며, 실제로 인지진단모형은 교육측정분야에서 활발하게 논의되고 있다(김희경 외, 2012, p.8 재인용).

인지진단이론은 피험자의 지식 및 기능에 대한 이해 상태를 상세하게 파악하여 피험자의 학습을 돕고자 개발된 측정이론으로(김수진, 송미영, 김선희, 2008; 김명화, 2012) 학생의 강·약점을 파악하고, 맞춤형 보정 학습을 제공하거나 향후 학력향상을 위한 학습계획을 수립하기 위해서 기존 평가 결과보다 더 상세한 정보를 제공한다. 인지진단이론은 기존에 보고되는 평가 결과인 총점 또는 성취수준 정보를 포함하여 피험자가 특정 인지요소를 숙달하였는지 여부, 숙달하였다면 어느 정도 숙달하였는지를 알 수 있는 숙달확률 정보를 구체적으로 제공한다.

인지요소에 대한 이해가 이루어지고 나면 피험자의 인지능력을 정확하게 파악하기 위하여 Q행렬을 작성한다. Q행렬이란 검사를 구성하는 문항들과 그 문항에 정답하기 위해 필요한 인지요소간의 관계를 말하는데, 일반적으로 Q행렬은 학교 현장의 교사 또는 교수진과 같은 내용전문가에 의해서 정의된다. 피험자의 능력 파악이나 문항 분석을 위해 Q행렬을 정확하게 설정하는 것은 능력 모수 및 문항 모수 추정에 기본이 되는 과정이기 때문에 매우 중요한 과정이다(김수진, 송미영, 김선희, 2008; 김수진, 2010; Tatsuoka, 1983, 1990 재인용). 정확한 Q행렬이 만들어지지 않으면 아무리 좋은 인지진단모형을 적용하여 분석한다고 해도 통계적 결과가 타당하지 않을 수 있다. Q행렬의 작성함에 있어서 특정 인지요소와 문항 간 관계를 제대로 명시하지 못한다면 특정한 문항들을 해결하는데 요구되는 인지요소를 충분히 숙달했다 하더라도 숙달하지 못한 것으로 판단되거나, 인지요소에 숙달하지 못했음에도 숙달한 것으로 판단되어 피험자의 강·약점을 잘못 파악하게 되는 치명적인 오류를 범할 수 있기 때문이다.

인지진단이론의 관심과 함께 Rule-Space Model, Fusion Model, DINA(Deterministic-Input Noisy And gate) Model, DINO(Deterministic-Input Noisy Or gate) Model 등과 같은 다양한 인지진단모형의 연구가 활

발하게 이루어지고 있다. 이와 같은 모형을 적용한 연구들의 궁극적인 목적은 피험자의 인지적 능력에 대한 정확한 추정을 통하여 타당하고 유용한 정보를 제공하여 교사의 수업이나 학생의 학습에 직접적인 도움을 주는 것이라 할 수 있다.

2. 선행 연구

불확실한 상황 속에서 유의미한 정보를 해석하여 의사결정을 해야 하는 의학, 기상학 등의 분야에서 베이지안 네트워크는 예측의 정확성이 높다는 장점 때문에 활발하게 활용되고 있다.

근거 중심의 의학 분야에서는 이전 환자로부터 얻은 질병 관련 데이터로부터 베이지안 네트워크를 적용하여 미래의 환자가 병에 걸릴 확률을 계산하고 질병 및 치료 계획을 진단한다. 정용규, 진훈, 김인철(2001)은 여러 측면을 복합적으로 고려해야 하는 의료 분야의 특성상 불임의 원인을 파악하기 힘들고, 임신가능성을 바로 찾아내기 어렵다고 지적하며, 이러한 문제점을 해결하기 위해 베이지안 네트워크를 활용하여 실제 불임환자들의 검사자료와 임신과의 관계를 규명하였다. 이를 통해 불임 원인 및 임신 가능성을 예측하였으며, 연구 결과 95.96% 정도의 정확한 예측 결과를 보였다고 보고하였다. 정용규, 이연주, 원재강(2012)은 과거 환자의 진찰 데이터 집합과 유방암 암세포 속성의 데이터베이스를 분석하여 유방암에 걸릴 확률을 계산하였다. 실제 베이지안 네트워크를 적용한 유방암의 분류는 86.5%의 정확도를 보였으며, 베이지안 네트워크를 적용하지 않은 방법보다 약 2.1%가 더 높은 진단 정확도를 확인하였다. 이를 통하여 의료 분야에서 베이지안 네트워크의 적용은 증거 기반의 의료서비스를 제공하며, 진찰결과에 대한 신빙성을 증가시킬 수 있다고 보고하였다. Acid 외(2004)는 스페인 응급병

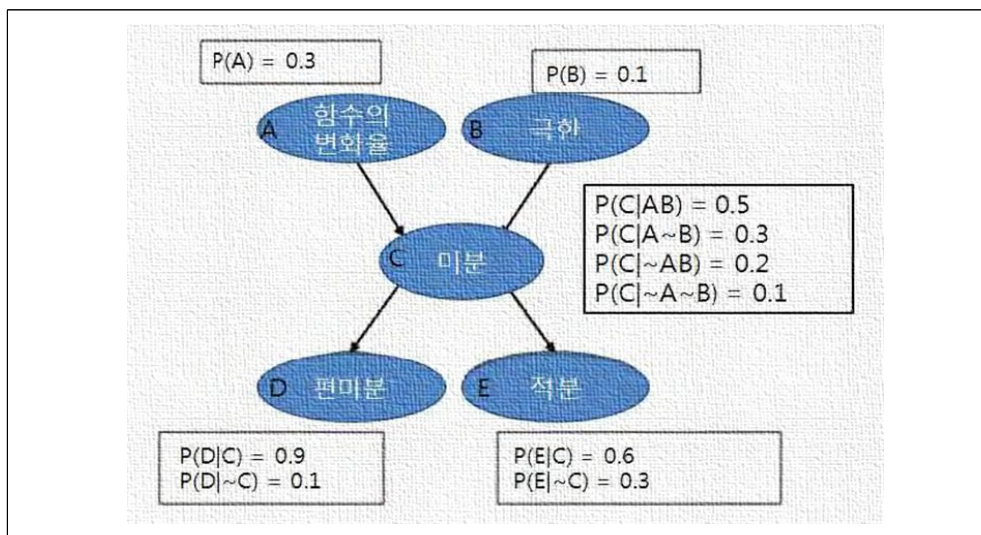
원을 대상으로 관리자 관점에서 응급 의료서비스에 영향을 미치는 요인을 파악하기 위해 베이지안 네트워크를 활용하였다. 약 3만 명 환자의 응급실 방문 원인, 입원시간과 병명, 주치의, 퇴원 등의 응급진료 기록을 사전확률로 활용하여 학습알고리즘을 실행하였으며, 변수 간의 관계를 직접 비교함으로써 더 나은 의료 서비스를 제공하기 위한 건강관리 시스템 구축을 위해 잠재적 용도의 베이지안 네트워크 활용을 제안하였다.

기상학 분야에서도 베이지안 네트워크의 활용을 쉽게 찾아볼 수 있다. Elsner와 Bossak(2001)은 1851년~2000년까지 관측된 기상 기록을 분석한 자료를 베이지안 네트워크에 적용하여 미국의 허리케인 활동을 예측하였다. 베이지안 네트워크를 통한 예측은 실제로 허리케인 발생과 유사한 결과를 보였으며, 이 연구는 향후 30년 동안 미국 내 강력한 허리케인이 발생할 숫자 및 시기를 예고하기도 하였다. 남기훈(2014)은 복합재난 위험성 평가를 위해 20년 간 국내에서 발생한 자연재난을 분석하였다. 이 자료를 바탕으로 베이지안 네트워크를 이용하여 태풍과 호우로 인해 발생하는 국내 복합재난에 대한 모형을 구축하였으며, 모형의 정확도는 약 81%로 나타났다. 또한, 위험관리 측면에서 베이지안 네트워크를 활용하여 복합재난과 관련된 특성을 분석하였다.

교육측정분야에서도 베이지안 네트워크를 활용한 연구가 조금씩 늘어나는 추세에 있다. 이러한 연구를 소개하자면 다음과 같다.

김성호(1994), 김성호와 김은희(1995)는 과제 수행과 과제수행 능력 구조(관계)를 표현할 수 있는 네트워크 모형을 탐색하는데 중점을 두고 베이지안 네트워크 모형을 개발하는 과정을 보여주었다. 황선아와 최용석(2005)은 기존의 CAT가 먼저 풀었던 문항이 다음 문항의 응답에 영향을 미치는 문맥효과(context effect)를 무시한 채 평가 문항을 구성하였기 때문에 피험자의 능력을 제대로 측정하지 못한다는 점을 지적하였다. 이에 황선아(2005)

는 베이저안의 조건부확률을 적용하여 기존의 50문항에서 평균 27개의 문항으로 피험자의 능력을 정확하게 추정하였다고 보고하였다. 나선웅과 김경수, 최용석(2012), 나선웅(2012)은 일차원성 가정으로 인해 학습의 위계를 무시한다는 IRT의 제한점을 보완하기 위한 목적으로 베이저안 네트워크를 적용한 컴퓨터기반 적응형평가(CAT)를 연구하였다. 베이저안 네트워크를 적용한 CAT는 IRT의 문제점을 해결하면서 피험자 개개인의 수준에 맞는 측정이 가능함은 물론, 피험자의 실제 능력치에 수렴하는 속도도 증가하였다고 보고하였다.

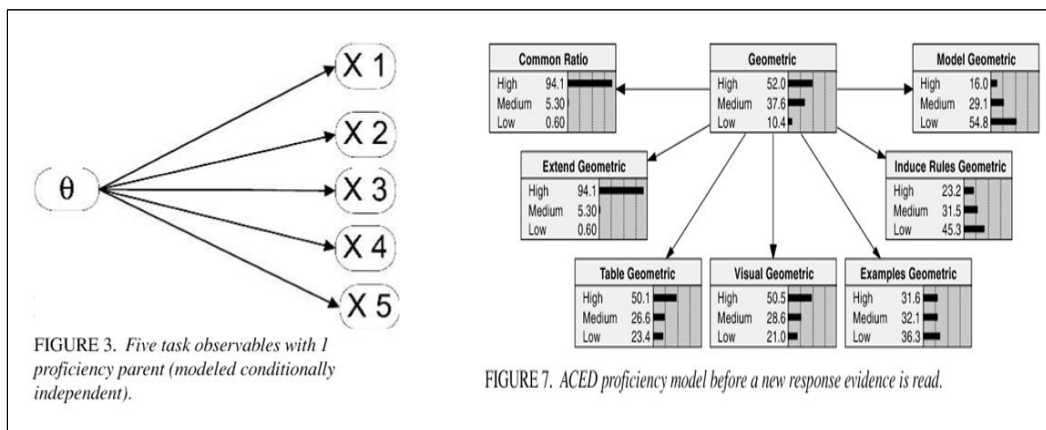


[그림 II-3] 조건부확률 분포로 표현한 베이저안 네트워크 예시(나선웅, 2012 p.17)

박상희 외(2009)는 결측치가 많은 영과잉자료, 중도절단자료, 소표본 자료를 자주 접하게 되는 특수교육 집단이라는 연구 대상의 특수성을 고려하여 장애 적격성 판정에 활용할 수 있는 베이저안 분석 방법을 소개하고 특수교육분야에서의 적용 가능성을 탐색하였다. 주관적인 사전분포를 사용하는 베이저안 네트워크는 교사 또는 해당 분야 전문가의 주관적인 의견을 적극 반영하여 추론의 정확성을 높일 수 있으며, MCMC 방법을 사용하기 때문에

소표본 자료에서도 분석이 가능하다는 장점을 언급하며 특수교육분야에 최적화된 방법론임을 보고하였다.

Almond 외(2007)는 진단평가 측면에서 근거 중심의 평가 방법인 베이지안 네트워크의 활용 예시를 제시하였다. 베이지안 네트워크는 특별한 기술 없이 사용 가능하고, 결과 해석이 쉬우며, 능력 노드 사이의 속달 정보를 제공할 뿐만 아니라, 확률로 이루어진 네트워크 그래픽을 제공하는 등과 같은 장점을 언급하며 교육측정분야의 실무자에게 베이지안 네트워크의 사용을 제안하였다.



[그림 II -4] 베이지안 네트워크 모델의 예시(Russell G. Almond 외, 2007)

Vomlel(2004)은 베이지안 네트워크를 응용한 컴퓨터기반적응검사(CAT)의 성능을 확인하였다. CAT에서 많이 사용되는 평균, 오름차순, 내림차순의 문항 순서보다 베이지안 네트워크를 적용한 시험이 더 적은 수의 문항으로 피험자의 능력을 정확하게 측정하는 결과를 보여주었다. García 외(2007)는 개개인이 학습 습관과 정보를 처리하는 방법이 각기 다르다는 가정 하에 컴퓨터공학 전공생을 대상으로 웹 기반의 교육환경에서 상호작용하는 과정을 분석한 자료를 베이지안 네트워크에 적용하여 학습유형을 모델링하였다. 개개인의 학습 유형에 따라 적절한 학습 방법을 선택하고 보조학습을 제공하는

데 활용될 수 있다고 제안하였다.

그리고 강태훈(2014)은 베이지안 네트워크를 활용한 문항 분석 방법의 타당성을 검증하기 위해 기존의 CTT와 IRT에 의한 결과와 베이지안 네트워크를 활용한 분석 결과를 비교하였다. 다차원성이 존재하는 수학 검사의 분석 결과, 베이지안 네트워크와 CTT, IRT의 곤란도 상관이 .62~.68, 변별도 상관이 .89~.91로 상당한 수준으로 일치한 결과를 얻었다고 보고하였다.

앞에서 살펴본 교육 분야에서의 베이지안 네트워크 관련 선행연구들을 정리해보면, 베이지안 네트워크는 질적·양적 자료 모두에 모두 손쉽게 적용될 수 있고, 확률적 결과 해석이 합리적으로 제시될 수 있으며, 결측치가 존재하는 경우에도 주어진 자료 안에서 최선의 추정이 가능하다는 장점이 있다고 보고하였다. 그러나 위 연구는 베이지안 네트워크의 소개와 장점을 제시하였을 뿐 베이지안 네트워크를 활용하여 정작 일반 교사나 검사 결과 사용자들의 관심이라고 할 수 있는 피험자의 능력, 즉 피험자가 인지요소를 숙달했는지(숙달여부), 숙달했다면 어느 정도 숙달했는지(숙달확률)를 정확하게 추정할 수 있는지에 대한 연구는 거의 이루어지지 않았음을 알 수 있었다.

IRT나 CTT에 의한 피험자의 능력 추정은 상대적 숙달정도에 대한 정보를 제공해 주긴 하지만, 학생 개개인의 인지상태에 대한 정보를 제공해주지는 못한다(김성훈, 2005). 반면 베이지안 네트워크는 능력모수 θ 이외에 θ 와 관련된 하위 능력과 그 능력간의 관계를 조건부확률을 이용하여 추정해 준다. 각 분야에서 응용되고 있는 강력한 인지진단모형인 베이지안 네트워크를 교육측정분야에 적용하고자 할 때, 교사나 측정전문가의 관심이라 할 수 있는 피험자의 능력 즉, 피험자의 인지요소 숙달 정도가 얼마나 정확하게 추정되는지 확인할 필요가 있다.

피험자의 능력을 추정하고자 할 때 능력 추정의 정확성 또는 능력 모수 복원에 영향을 주는 조건을 정리해 보면 다음과 같다.

강태훈, 박찬호, 김인숙(2011)의 연구에서 3모수 로지스틱 모형을 적용하여 실제 검사자료를 통해 피험자의 성취수준 분류일관도와 분류정확도에 영향을 미치는 요인을 파악하였는데, 검사의 길이가 길어질수록, 수행등급 개수가 증가할수록 양호한 분류일관도와 분류정확도를 얻었다고 보고하였다.

인지진단 모형의 연구를 보면, DINA모형에서는 부정적인 문항모수의 값이 작을수록 피험자의 분류정확도는 높아진다고 보고하였다(Huebner, Wang, 2011; Cui, Gierl, & Chang, 2012; de la Torre, Hong, & Deng, 2010; 김지효, 2013;). 부정적인 문항모수란 실제 문항에 반응 시 잡음(noisy)으로 인해 피험자가 기대와는 다른 반응을 보이는 것으로 인지요소를 숙달하지 못했음에도 불구하고 추측으로 인해 정답하거나, 인지요소를 숙달하였음에도 부주의에 의한 실수로 문항에 오답하는 반응을 말한다.

검사의 인지요소와 관련된 분류정확도는 연구에 따라 차이가 있다. Gierl, Leighton, & Tan(2006)은 인지요소 간 상관성이 높아질수록 구조가 일차원과 유사해지기 때문에 분류 정확도와 일관성이 낮아진다고 보고하였다. 그러나 Cui, Gierl, & Chang(2012)의 연구에서는 인지요소 간 관계가 종속적일수록 분류의 정확성과 일관성이 높아진다고 보고하였다. 검사의 인지요소의 수에 따른 분류정확도에서도 연구에 따라 차이가 있는데, Chiu, Douglas, & Li(2009), Cui, Gierl, & Chang(2012)은 검사의 인지요소 수가 많을수록 분류일관성과 분류정확도가 낮아진다고 보고하였다. 같은 맥락으로 김지효(2013)의 연구에서도 DINA, DINO 모형에서 인지요소의 수가 증가할수록 분류정확도는 감소한다고 보고하였다. 그러나 Huebner와 Wang(2011)은 인지요소 수가 많을수록 피험자의 능력을 정확하게 분류할 수 있다고 보고하였다. 아직까지 인지요소 개수에 대한 범위에 대해서는 논란이 있다. 인지요소의 개수가 적으면 평가 결과를 보고할 때 학업성취에 대한 세부적 정보가 줄어들어 그만큼 유용성이 떨어지게 되며, 인지요소의 개수가 많으면 인지진

단모형을 적용할 때 추정해야 하는 모수가 많아서 분석의 정확도가 떨어지게 된다(김희경 외, 2012, p.115). 이는 관련 연구가 많지 않기 때문에 향후 인지 요소 간 상관이나 인지요소 수에 따른 피험자의 분류정확성에 대한 연구를 수행할 필요가 있다.

본 연구에서는 베이지안 네트워크를 적용한 모의실험 연구를 통하여 피험자 능력 추정의 정확성을 확인하고자 하였다. 이를 위해 ‘인지요소 간 상관 정도’와 ‘검사의 길이’라는 요인을 고려하여 모의실험 연구를 수행하였다.

Ⅲ. 연구 방법

본 연구에서는 모의실험 연구를 통하여 검사 자료에 베이지안 네트워크를 적용할 때 피험자의 능력모수 추정이 정확하게 이루어질 수 있는지를 진능력모수 복원 관점에서 평가하였다. 이를 위해 인지요소 간 상관 정도와 검사의 다양한 길이를 모의실험 요인으로 채택하여 조건이 달라짐에 따라 피험자의 능력모수 복원에 어떠한 영향을 미치는지 확인해 보았다.

1. 모의실험 설계

베이지안 네트워크를 활용한 모의실험 자료 생성을 위해서는 생성문항모수 확보가 필요하다. 이를 위하여 본 연구에서는 이론적 분포로부터 문항모수를 무선 표집하는 방법 대신 실제 검사 자료로부터 추정된 문항모수를 활용하였다. 본 연구에서는 강태훈(2014)의 연구에서 이미 활용된 바가 있는 실제 수학 검사 자료($n=3,000$)를 재사용하였다. 2005년 미국 중부의 한 대학 신입생들에게 실시한 수학 검사는 대수학(algebra) 10문항과 삼각함수(trigonometry) 10문항으로 구성되어 있으며, 이 검사의 전체 평균은 $9.72(SD=5.18)$, 대수학 10문항의 평균은 $4.96(SD=2.81)$, 그리고 삼각함수 10문항의 평균은 $4.75(SD=2.81)$ 로 나타났다. CTT에 의한 문항 통계치는 <표 III-1>에 제시된 바와 같다. 여기서 문항 곤란도는 정답률을 의미하며, 문항 변별도는 상위 27% 집단과 하위 27% 집단 간 정답률 차이를 의미한다.

<표 III-1> 수학 진단평가의 CTT 문항통계치(강태훈, 2014, pp.912-913)

| item | 곤란도 | 변별도 | item | 곤란도 | 변별도 |
|------|------|------|------|------|------|
| 1 | 0.61 | 0.48 | 11 | 0.42 | 0.74 |
| 2 | 0.52 | 0.72 | 12 | 0.56 | 0.62 |
| 3 | 0.54 | 0.55 | 13 | 0.62 | 0.68 |
| 4 | 0.38 | 0.61 | 14 | 0.53 | 0.65 |
| 5 | 0.41 | 0.57 | 15 | 0.44 | 0.56 |
| 6 | 0.42 | 0.50 | 16 | 0.41 | 0.48 |
| 7 | 0.51 | 0.80 | 17 | 0.48 | 0.49 |
| 8 | 0.48 | 0.58 | 18 | 0.50 | 0.57 |
| 9 | 0.56 | 0.55 | 19 | 0.45 | 0.53 |
| 10 | 0.55 | 0.63 | 20 | 0.35 | 0.66 |

20개 문항 중 곤란도 측면에서 가장 쉬운 문항은 13번이고, 가장 어려운 문항은 20번이다. 문항변별도 측면에서 가장 높은 변별도를 나타내는 문항은 7번이고, 1번과 16번 문항은 상대적으로 상위집단과 하위집단을 잘 변별하지 못하는 변별도가 낮은 문항이다. 하지만 모든 곤란도와 변별도가 황정규(1998)가 말하는 대략적 판단 기준인 0.2 이상이기 때문에 특별히 문제가 있는 문항은 없다고 해석할 수 있다.

탐색적 요인분석 및 내용전문가로부터 도출된 수학 진단평가의 Q행렬은 <표 III-2>와 같이 두 개의 인지요소로 구성되어 있다. 대수학(A1)의 능력을 요구하는 문항이 8개, 삼각함수(A2)의 능력을 요구하는 문항이 7개, 대수학과 삼각함수의 능력을 모두 갖추어야 풀 수 있는 문항은 5개(◎로 표시)이다.

<표 III-2> 수학 진단평가의 Q행렬(강태훈, 2014)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| A1 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | | | | | ○ | | | ○ |
| A2 | | | | | | | ○ | | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| A1 & A2 | | | | | | | ◎ | | ◎ | | ◎ | | | | | | ◎ | | | ◎ |

대수학(A1) : 1,2,3,4,5,6,8,10번
 삼각함수(A2): 12,13,14,15,16,18,19번
 대수학과 삼각함수(A1&A2) : 7,9,11,17,20번

베이지안 네트워크를 적용하여 능력모수를 추정함에 있어 모수 복원의 정확성에 영향을 미칠 것으로 생각되는 다음의 두 가지 요인을 채택하여 모의 실험을 실시하였다.

첫째, 인지 요소 간 상관 정도이다. 인지 요소 간 상관에 따라 피험자 능력 추정의 정확성이 어떻게 달라지는지를 확인하기 위해 대수학(A1)과 삼각함수(A2) 간 상관을 0, 0.4, 그리고 0.8의 세 가지 조건을 고려하여 모의실험 자료를 생성하였다.

둘째, 검사의 길이를 달리하였다. 검사 문항의 수가 늘어남에 따라서 피험자 능력 모수 추정의 정확성이 얼마나 향상되는지를 확인해 보았다. 문항 수는 일반적으로 시행되고 있는 성취도 검사의 문항 수를 참고하여, 검사 길이는 20개, 40개, 그리고 60개 세 가지 조건을 고려하였다. 이때 검사의 길이가 늘어남에 따라 인지요소(Q행렬)와 문항모수(beta)는 변하지 않고 문항의 개수만 늘어나도록 변화를 주었다. 이는 마치 본래의 20개 문항검사에 대한 동형검사가 추가적으로 두 개 만들어진 후, 세 개 검사 중 하나만 사용한 경우(20개 문항), 두 개만 사용한 경우(40개 문항), 그리고 세 개 다 사용한 경우(60개 문항) 피험자 능력 추정의 정확성이 증가하는지를 확인하기 위한 고려라 할 수 있다. 이를 통해 피험자의 능력모수 추정의 정확성을 확

인하는데 집중하고자 하였다.

결과적으로 능력모수 복원의 정확성을 확인하기 위한 모의실험 연구를 위해 본 연구에서는 총 9가지(인지요소 간 상관 3가지 * 검사 길이 3가지) 모의실험 조건을 설계하였으며, 각 조건마다 3,000명이 응답한 10개의 모의실험 자료를 생성하였다.

2. 모의실험 자료의 생성 및 추정

본 연구의 주된 관심은 피험자의 능력모수를 추정하여 진능력모수 복원의 정확성을 확인하는데 있기 때문에, 데이터를 생성함에 있어서 문항모수는 별도로 추정하지 않고 앞에서 제시한 생성문항모수(generating item parameters)를 그대로 사용하면서 능력모수만을 추정하는 방법을 선택하였다. 능력모수 추정을 위한 모의 데이터 생성 시 WinBUGS(Spiegelhalter, Thomas, best, & Lunn, 2003) 통계소프트웨어패키지를 사용하였다.

모의실험 자료를 생성하기 위한 과정은 다음과 같다.

1) 강태훈(2014) 연구의 실제 자료를 활용하여 베이지안 네트워크 하에서의 각 문항의 모수라 할 수 있는 문항모수(beta)를 추정하였다. 그 결과는 <부록 3>에 제시된 바와 같다. MCMC로 추정하였기 때문에 문항모수와 약간의 차이는 있었으나 거의 유사하게 나타났다.

2) 추정된 문항모수를 진문항모수로 설정한 후, 피험자 당 두 개의 잠재적 능력 변수값을 이원 정규분포(binary normal distribution)를 활용하여 생성하였다. 이 때 각 변수는 평균 0, 표준편차 1로 설정하였으며, 변수 간 공분산을 이용하여 능력 간 상관계수를 0, 0.4, 그리고 0.8 로 상정하였다.

3) 인지요소 $A1$, $A2$ 에 대한 피험자 3,000명의 진속달확률을 구하기 위하여, 앞에서 생성된 각 표준점수의 백분위를 사용하였다. 즉 2)의 이원 정규

분포 상에서 한 잠재변수에 대한 어떤 피험자의 표준점수로서의 능력모수 값이 0이라면, 그에 해당하는 인지요소 숙달확률은 표준정규분포 상의 백분위인 50%라고 보았다.

4) 마지막으로, 생성문항 모수 및 피험자의 진숙달확률을 이용하여 베이지안 네트워크 하에서 각 피험자가 각 문항에 정답 반응할 수 있는 확률을 계산하였다. 다음 최소값 0과 최대값 1을 가지는 균일분포(uniform distribution)에서 무작위로 생성된 값과 비교하여 이 값보다 정답 확률이 크면 정답(1)한 것으로, 반대로 정답 확률보다 큰 값일 경우는 오답(0)한 것으로 처리하여 모의실험 자료를 생성하였다.

3. 베이지안 네트워크의 능력모수 추정 결과에 대한 평가

베이지안 네트워크를 적용한 능력모수 추정의 정확성을 평가하기 위한 기준으로 다음의 방법을 사용하였다.

첫째, 진능력모수와 추정된 능력모수의 결과가 얼마나 비슷한 추정치를 산출하는지 확인하기 위해 각 인지요소에 대한 모의 피험자들의 진숙달확률(p)과 추정된 숙달확률(\hat{p}) 간 순위 상관계수와 평균 제곱근 편차(root mean square deviation, RMSD)를 계산하였다. 진능력모수(true θ)값에 대한 추정된 능력모수의 복원이 정확하다면 양자 간의 순위 상관은 높고, RMSE는 작게 산출될 것으로 기대된다. RMSE를 계산하는 식은 다음과 같다.

$$\text{RMSE} = \sqrt{\sum_{j=1}^n \frac{(p - \hat{p})^2}{N}} \quad \text{식 (9)}$$

모의 조건에 따른 RMSE의 차이가 통계적으로 유의미한지 확인하기 위해 분산분석(2-way ANOVA)을 실시하였다.

둘째, 진숙달여부(m)와 추정된 숙달여부(\hat{m})의 분류정확도이다. 특정 인지 요소의 숙달했는지 여부에 대한 판단 기준으로는 숙달확률 0.5를 사용하였다(Li, 2011; Lee 외, 2011; 김희경, 2012). 즉, 숙달확률 0.5를 기준으로, 0.5 미만이면 인지요소를 숙달하지 못한 것으로 판단하여 미숙달로 분류하고, 0.5이상이면 숙달로 분류하였다. 진숙달여부(m)와 추정된 숙달여부(\hat{m})간 피험자의 숙달과 미숙달에 대한 분류정확도를 확인하고, 어느 정도 일치한 결과를 보이는지에 따라 복원이 정확하게 이루어졌는지 평가하였다. 모의 조건에 따른 분류정확도의 차이가 통계적으로 유의미한지 확인하기 위해 분류정확도 비율을 로짓(Logit)으로 변환하여 분산분석(2-way ANOVA)을 실시하였다.

IV. 연구 결과

본 연구는 진능력모수와 추정된 능력모수를 비교하여 베이지안 네트워크를 적용한 능력모수 복원의 정확성을 평가하였다. 앞에서 언급한 바와 같이 능력모수 복원의 정확성을 평가하기 위한 기준은 1) 피험자의 진속달확률(p)과 모의피험자의 추정된 속달확률(\hat{p})의 순위 상관계수와 RMSE 값을 비교하고, 2) 진속달여부(m)와 추정된 속달여부(\hat{m})의 분류정확도를 비교하는 것이다.

1. 진속달확률과 추정된 속달확률 간 순위 상관계수와 RMSE

인지요소 간 상관 정도와 검사 길이를 달리하여 구성된 각 모의실험 조건 속에서 3,000명의 피험자가 응답한 모의실험 자료를 10개씩 생성한 바 있다. 각 자료에 대하여 인지요소별로 피험자의 속달확률을 추정하였으며, 이를 진속달확률과 비교하여 복원의 정확성을 평가하였다.

1) 대수학의 순위 상관계수와 RMSE

첫 번째 인지요소인 대수학(A1)에 대해 10번 추정한 모의피험자의 진속달확률(p)과 추정된 속달확률(\hat{p})간 순위 상관계수 평균과 표준편차는 <표 IV-1>과 같이 .49~.51로 나타났다.

모의실험 요인의 하나인 검사의 길이에 따른 순위 상관계수의 변화를 살펴보면, 상관이 0일 때 검사의 길이가 길어질수록 순위 상관계수가 조금씩 높아지는 것을 확인할 수 있었다. 이는 문항의 수가 많아질수록 \hat{p} 가 p 를 정확하게 복원하고 있다고 해석할 수 있다. 상관이 .4와 .8을 가진 조건에서는 검사 길이

에 따라 순위 상관계수가 변화하였지만, 검사 길이가 길어질수록 꾸준히 증가하거나 감소하는 일정한 패턴은 나타나지 않았다.

또 다른 모의 실험요인인 인지 요소 간 상관정도는 검사 길이와 함께 능력 추정의 정확성에 대하여 약간의 차이가 나타났다. 20개 문항 조건에서는 인지 요소 간 상관이 높아질수록 p 와 \hat{p} 간 순위 상관이 높게 나타났으나, 40개와 60개 문항을 가진 모의실험 조건에서는 인지요소 간 상관에 따른 일정한 패턴은 나타나지 않았다.

<표 IV-1> 대수학(A1)의 진속달확률과 추정된 속달확률의 순위 상관 평균

| | | 20개 | 40개 | 60개 |
|----|--------|-----------------|-----------------|-----------------|
| A1 | $r=.0$ | .485 (0.014) | .500 (0.015) | .505 (0.012) |
| | $r=.4$ | .491 (0.017) | .498 (0.015) | .497 (0.012) |
| | $r=.8$ | .503 (0.016) | .497 (0.013) | .498 (0.015) |

()는 표준편차

A1을 10번 추정한 p 와 \hat{p} 의 RMSE 평균과 표준편차를 살펴보면 <표 IV-2>에 제시한 바와 같이 0.237~0.244의 범위를 나타냈다. RMSE는 진속달확률과 추정된 속달확률 간 추정의 차이, 즉 추정의 오차를 의미하며 RMSE 값이 작을수록 \hat{p} 가 p 를 정확하게 복원한 것으로 해석할 수 있다.

검사 길이에 따른 RMSE 값을 살펴보면, 문항 수가 늘어날수록 추정의 오차가 줄어드는 것을 확인할 수 있었다. 그러나 인지 요소 간 상관 정도가 높아질수록 RMSE는 약간의 변화가 나타났지만, 그 차이는 극히 미미하여 거의 같다고 표현할 수 있을 정도였다.

<표 IV-2> 대수학(A1)의 진속달확률과 추정된 속달확률의 RMSE 평균

| | | 20개 | 40개 | 60개 |
|----|--------|--------------------|--------------------|--------------------|
| A1 | $r=.0$ | 0.2444 (0.0033) | 0.2389 (0.0031) | 0.2373 (0.0031) |
| | $r=.4$ | 0.2439 (0.0031) | 0.2379 (0.0022) | 0.2373 (0.0022) |
| | $r=.8$ | 0.2419 (0.0027) | 0.2383 (0.0024) | 0.2373 (0.0024) |

()는 표준편차

2) 삼각함수의 순위 상관계수와 RMSE

두 번째 인지요소인 삼각함수(A2)에 대해 10번 추정 한 진속달확률(p)과 추정된 속달확률(\hat{p})간 순위 상관계수의 평균과 표준편차는 <표 IV-3>과 같이 .48~.51의 범위를 나타냈다.

모의실험 요인의 하나인 검사의 길이에 따른 상관계수의 변화를 살펴보면 동일한 조건일 때 검사의 길이가 길어질수록 순위 상관이 조금씩 높아지는 것을 확인할 수 있었다. 이러한 결과는 문항의 수가 많아질수록 두 추정치간의 정확도가 높아진다는 의미이며, 이는 검사 길이가 길어질수록 \hat{p} 가 p 를 유사하게 복원한다고 말할 수 있다.

또 다른 요인인 인지 요소 간 상관에서는 20개에서는 두 개의 인지 요소 간 상관이 높아질수록 p 와 \hat{p} 간 순위 상관이 조금씩 높게 나타났으나, 40개와 60개 문항의 조건에서는 인지요소 간 상관에 따른 일정한 패턴은 나타나지 않았다.

<표 IV-3> 삼각함수(A2)의 진속달확률과 추정된 속달확률의 순위 상관 평균

| | | 20개 | 40개 | 60개 |
|----|--------|-----------------|-----------------|-----------------|
| A2 | $r=.0$ | .480 (0.010) | .499 (0.020) | .511 (0.017) |
| | $r=.4$ | .485 (0.010) | .504 (0.012) | .509 (0.014) |
| | $r=.8$ | .490 (0.012) | .498 (0.020) | .508 (0.013) |

()는 표준편차

A2를 10번 추정된 p 와 \hat{p} 의 RMSE 평균과 표준편차를 살펴보면 <표 IV-4>에 제시한 바와 같이 0.235~0.246의 범위를 나타냈다. RMSE는 진속달확률과 추정된 속달확률 간 차이를 나타내는데 RMSE 값이 작을수록 \hat{p} 가 p 를 정확하게 복원한 것으로 해석할 수 있다.

검사 길이에 따른 RMSE 값을 살펴보면 문항 수가 늘어날수록 추정의 오차가 줄어드는 것을 확인할 수 있었다. 인지 요소 간 상관 정도가 높아질수록 RMSE는 약간의 변화가 나타났지만 일정한 패턴은 나타나지 않았으며 그 차이는 극히 미미하여 거의 같다고 표현할 수 있을 정도였다.

<표 IV-4> 삼각함수(A2)의 진속달확률과 추정된 속달확률의 RMSE 평균

| | | 20개 | 40개 | 60개 |
|----|--------|--------------------|--------------------|--------------------|
| A2 | $r=.0$ | 0.2456 (0.0019) | 0.2394 (0.0039) | 0.2348 (0.0037) |
| | $r=.4$ | 0.2453 (0.0023) | 0.2393 (0.0038) | 0.2349 (0.0038) |
| | $r=.8$ | 0.2450 (0.0023) | 0.2395 (0.0040) | 0.2349 (0.0037) |

()는 표준편차

3) 두 인지요소의 진속달확률과 추정된 속달확률의 RMSE 분산분석

모의실험 조건에 따라 추정된 p 와 \hat{p} 의 RMSE 값이 통계적으로 유의미한 차이인지를 확인하기 위하여 모의실험 조건에 따라 계산된 90개의 RMSE 값을 종속변수로 하고, 검사 길이(3)와 인지요소 간 상관정도(3)를 독립변수로 두고 이원 분산분석(2 way-ANOVA)을 실시하였다.

첫 번째 인지요소인 대수학(A1)의 결과를 살펴보면 <표 IV-5>와 같이 검사 길이에 따른 RMSE 차이는 통계적으로 유의한 것으로 나타났으며($F(2, 81)=41.977, p<.001$), scheffe 사후 검증을 통해 검사 길이가 길어질수록 RMSE 값이 작아지는 것을 확인할 수 있다($i= 20 > 40 = 60$). 즉, 검사 길이가 길어질수록 속달확률 간 추정오차가 줄어들어 p 가 \hat{p} 를 정확하게 복원하는 것으로 해석할 수 있다. 인지요소 간 상관에 따른 RMSE 차이는 통계적으로 유의하지 않는 것으로 나타났다($F(2, 81)=1.015, p=.367$)

<표 IV-5> 검사 길이와 인지요소 간 상관에 따른 RMSE 분산분석 결과(대수학)

| A1 | 제 III 유형 제곱합 | 자유도 | 평균 제곱 | F | 유의확률 |
|-------------|-----------------|-----|-------|--------|----------|
| 검사 길이 | .001 | 2 | .000 | 41.977 | <.001*** |
| 인지요소간 상관 | .000 | 2 | .000 | 1.015 | .367 |
| i * r(상호작용) | .000 | 4 | .000 | .805 | .526 |

* $p<.05$, ** $p<.01$, *** $p<.001$

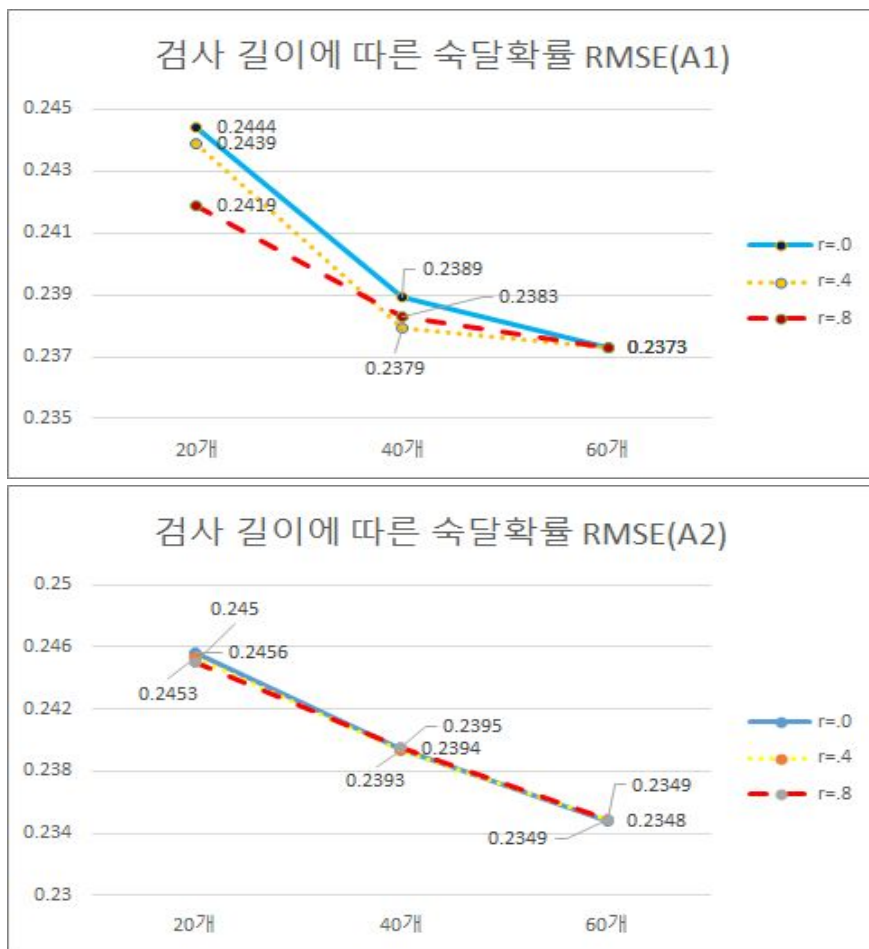
두 번째 인지요소인 삼각함수(A2)에서도 <표 IV-6>와 같이 검사 길이에 따른 RMSE 차이는 통계적으로 유의미한 것으로 나타났으며($F(2, 81)=71.381, p<.001$), scheffe 사후 검증을 통해 검사 길이가 길어질수록 RMSE 값이 작아지는 것을 확인할 수 있다($i= 20 > 40 > 60$). 이를 통해 검사 길이가 길어질수록 속달확률 간 복원의 정확성이 높아지는 것을 확인할 수 있었다. 인지요소 간 상관에 따른 RMSE의 차이는 통계적으로 유의하지 않았다($F(2, 81)=0.021, p=.979$).

<표 IV-6> 검사 길이와 인지요소 간 상관에 따른 RMSE 분산분석 결과(삼각함수)

| $A2$ | 제 III 유형 제공합 | 자유도 | 평균 제곱 | F | 유의확률 |
|----------------|-----------------|-----|-------|--------|----------|
| 검사 길이 | .002 | 2 | .001 | 71.381 | <.001*** |
| 인지요소 간 상관 | .000 | 2 | .000 | .021 | .979 |
| $i * r$ (상호작용) | .000 | 4 | .000 | .049 | .995 |

* $p < .05$, ** $p < .01$, *** $p < .001$

‘검사 길이’에 따른 대수학(A1)과 삼각함수(A2)의 숙달확률 RMSE의 변화는 [그림 IV-1]과 같이 나타낼 수 있다.



[그림 IV-1] 대수학(A1)과 삼각함수(A2)의 검사 길이에 따른 숙달확률 RMSE

2. 숙달여부 분류정확도

능력모수 복원의 정확도를 평가하기 위해 진숙달여부(m)와 추정된 숙달여부(\hat{m})의 분류정확도를 확인하였다. 숙달확률이 0.5 미만이면 피험자가 인지요소를 숙달하지 못한 미숙달로 분류하였고, 숙달확률이 0.5 이상이면 피험자가 인지요소를 숙달한 것으로 판단하였다. 모의피험자의 숙달여부가 m 과 \hat{m} 에서 일관되게 분류한 비율을 분류정확도로 보고 모의실험 조건에 따른 분류정확도 결과를 확인하였다.

1) 대수학의 숙달여부 분류정확도

첫 번째 인지요소인 대수학(AI)과 관련된 숙달여부 분류 정확도를 확인하기 위해 진숙달여부(m)와 추정된 숙달여부(\hat{m})를 10번 추정하였다. 모의피험자 3,000명의 m 과 \hat{m} 에서 일관되게 분류한 피험자 수와 비율은 <표 IV-7>과 같다.

<표 IV-7> 대수학(AI)의 진숙달여부와 추정된 숙달여부의 분류정확도 평균

| | | 20개 | 40개 | 60개 |
|----|--------|----------|----------|----------|
| AI | $r=.0$ | 2,161 | 2,219 | 2,259 |
| | | (72.02%) | (73.95%) | (75.29%) |
| | $r=.4$ | 2,169 | 2,232 | 2,246 |
| | | (72.30%) | (74.40%) | (74.88%) |
| | $r=.8$ | 2,183 | 2,233 | 2,239 |
| | | (72.76%) | (74.42%) | (74.62%) |

m 과 \hat{m} 의 분류정확도는 약 72.0~75.3% 정도로 비교적 정확하게 분류된 것을 확인할 수 있었다.

상관이 0일 때의 분류정확도를 살펴보면 문항 수가 증가할수록 분류정확도가 조금씩 높아지는 것을 확인할 수 있다(72.0% < 74.0% < 75.3%). 다른 조건에서도 비슷하게 문항 수가 증가할수록 숙달여부 분류정확도는 높아지는 것을 확인할 수 있었다. 그러나 인지요소 간 상관이 높아질수록 분류정확도가 일정하게 높아지거나 또는 낮아지는 패턴은 발견되지 않았다. 이는 숙달여부 분류정확도를 기준으로 인지 요소 간 상관 정도는 진능력 복원에 별다른 영향을 미치지 않는다고 해석할 수 있다.

2) 삼각함수의 숙달여부 분류정확도

두 번째 인지요소인 삼각함수(A2)와 관련된 숙달여부 분류 정확도를 확인하기 위해 진숙달여부(m)와 추정된 숙달여부(\hat{m})를 10번 추정하였다. 첫 번째 인지요소(A1)와 마찬가지로 숙달확률 0.5를 기준으로 피험자를 숙달과 미숙달 집단으로 분류하였으며, 3,000명의 m 과 \hat{m} 에서 일관되게 분류한 피험자 수와 비율은 <표 IV-8>과 같다.

<표 IV-8> 삼각함수(A2)의 진숙달여부(m)와 추정된 숙달여부(\hat{m})의 분류정확도

| | | 20개 | 40개 | 60개 |
|----|--------|----------|----------|----------|
| A2 | $r=.0$ | 2,157 | 2,231 | 2,260 |
| | | (71.89%) | (74.35%) | (75.33%) |
| | $r=.4$ | 2,158 | 2,232 | 2,259 |
| | | (71.95%) | (74.41%) | (75.29%) |
| | $r=.8$ | 2,160 | 2,230 | 2,262 |
| | | (72.00%) | (74.32%) | (75.41%) |

m 과 \hat{m} 의 분류정확도는 약 71.9~75.4% 정도로 피험자의 숙달여부를 비

교적 정확하게 분류된 것을 확인할 수 있었다.

상관이 0일 때 분류정확도를 살펴보면 문항 수가 증가할수록 분류정확도가 꾸준히 높아지는 것을 확인할 수 있으며($71.9\% < 74.4\% < 75.3\%$), 다른 조건에서도 문항 수가 증가할수록 분류정확도가 높아지는 것을 확인할 수 있었다. 그러나 인지요소 간 상관이 높아질수록 분류정확도가 일정하게 높아지거나 또는 낮아지는 패턴은 발견되지 않았다. 이는 숙달여부 분류정확도를 기준으로 인지 요소 간 상관 정도는 진능력 복원에 별 다른 영향을 미치지 않는다고 해석할 수 있다.

3) 두 인지요소의 진숙달여부와 추정된 숙달여부의 분류정확도 분산분석

모의실험 조건에 따라 추정된 m 과 \hat{m} 의 분류정확도가 통계적으로 유의미한 차이인지 확인하기 위하여 로짓으로 변환한 분류정확도 90개를 종속변수로 하고 검사 길이(3)와 인지요소 간 상관정도(3)를 독립변수로 두어 이원 분산분석(2way-ANOVA)을 실시하였다.

첫 번째 인지요소인 대수학(AI)의 결과를 살펴보면 <표 IV-9>와 같이 검사 길이에 따른 분류정확도 차이는 통계적으로 유의미한 것으로 나타났으며 ($F(2, 81)=72.484, p<.001$), scheffe 사후 검증을 통해 검사 길이가 길어질수록 숙달여부 분류정확도가 높아지는 것을 확인할 수 있었다($i= 20 > 40 > 60$). 즉, 검사 길이가 길어질수록 \hat{m} 이 m 을 유사하게 복원한다고 해석할 수 있다. 그러나 검사 길이에 따른 숙달여부 분류정확도 차이는 통계적으로 유의하지 않았다($F(2, 81)=0.295, p=.745$).

<표 IV-9> 검사 길이와 인지요소 간 상관에 따른 분류정확도 분산분석 결과(대수학)

| A1 | 제 III 유형 제공합 | 자유도 | 평균 제공 | F | 유의확률 |
|-------------|-----------------|-----|-------|--------|----------|
| 검사 길이 | .282 | 2 | .141 | 72.484 | <.001*** |
| 인지요소간 상관 | .001 | 2 | .001 | .295 | .745 |
| i * r(상호작용) | .016 | 4 | .004 | 2.077 | .091 |

* $p < .05$, ** $p < .01$, *** $p < .001$

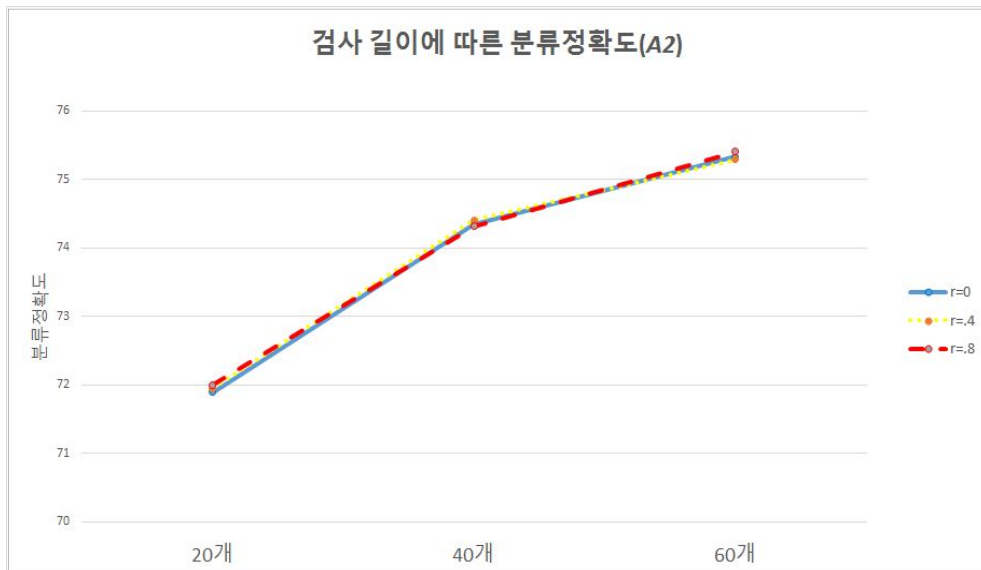
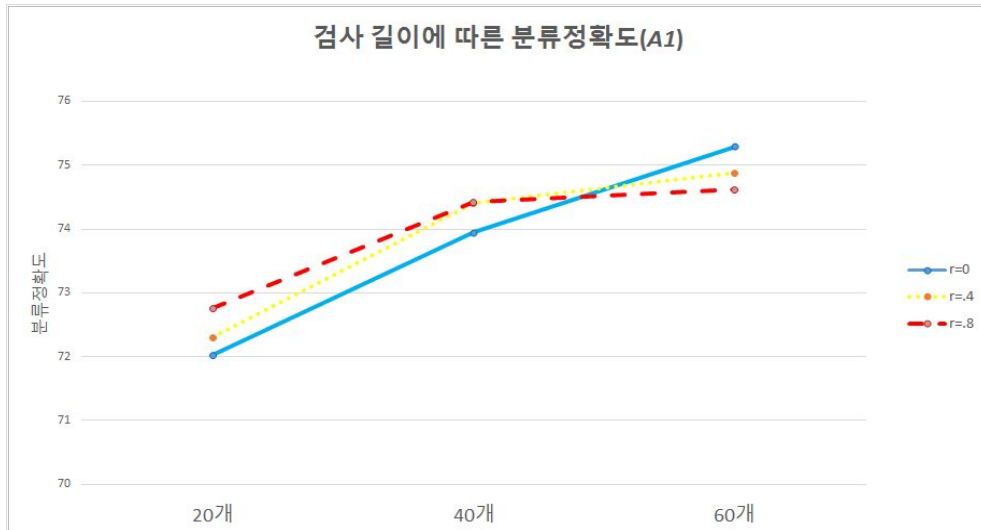
두 번째 인지요소인 삼각함수(A2)에서도 <표 IV-10>와 같이 검사 길이에 따른 분류정확도 차이는 통계적으로 유의미한 것으로 나타났으며($F(2, 81) = 102.729$, $p < .001$), scheffe 사후 검증을 통해 검사 길이가 길어질수록 숙달 여부 분류정확도가 높아지는 것을 확인할 수 있었다($i = 20 > 40 > 60$). 즉, 검사 길이가 길어질수록 \hat{m}_i 가 m 을 유사하게 복원한다고 해석할 수 있다. 그러나 상관정도에 따른 숙달 여부 분류정확도 차이는 통계적으로 유의하지 않았다($F(2, 81) = 0.029$, $p = .972$).

<표 IV-10> 검사 길이와 인지요소 간 상관에 따른 분류정확도 분산분석 결과(삼각함수)

| A2 | 제 III 유형 제공합 | 자유도 | 평균 제공 | F | 유의확률 |
|-------------|-----------------|-----|-------|---------|----------|
| 검사 길이 | .489 | 2 | .245 | 102.729 | <.001*** |
| 인지요소 간 상관 | .000 | 2 | .000 | .029 | .972 |
| i * r(상호작용) | .000 | 4 | .000 | .039 | .997 |

* $p < .05$, ** $p < .01$, *** $p < .001$

대수학(A1)과 삼각함수(A2)의 검사 길이에 따른 숙달 여부 분류정확도의 변화는 [그림 IV-2]와 같이 나타낼 수 있다.



[그림 IV-2] 대수학(A1)과 삼각함수(A2)의 검사 길이에 따른 분류정확도

V. 결론 및 논의

1. 결론

베이지안 네트워크에서 추정되는 능력모수란 피험자가 각 인지 속성을 가지고 있을 확률 즉, 숙달확률과 숙달여부를 의미한다. 본 연구에서는 강력한 인지이론모형 중 하나인 베이지안 네트워크를 적용한 모의실험 연구를 통하여 피험자 능력 추정의 정확성을 확인하고자 하였다. 이를 위해 실제 피험자 3,000명이 응답한 수학 검사 자료로부터 생성문항모수(CPT)를 추출하고, 이원정규분포를 활용하여 진능력모수를 생성하였다. 생성된 진능력모수를 활용하여 각 모의실험 조건에서 10개씩의 모의실험 자료를 생성하였고, 생성된 자료로부터 각 모의 피험자의 능력 모수를 추정하였다.

베이지안 네트워크를 통해 추정한 능력모수의 정확성을 확인하기 위해 진능력모수 복원의 관점에서 p 와 \hat{p} 의 순위 상관계수와 RMSE, m 과 \hat{m} 의 분류정확도를 비교하였다.

먼저 모의실험 조건에 따른 진숙달확률(p)과 추정된 숙달확률(\hat{p})의 변화를 살펴보면 첫째, 검사 길이가 길어질수록 p 와 \hat{p} 의 순위 상관계수가 높아지고 RMSE 값이 작아졌으며 이러한 차이는 통계적으로 유의한 것으로 나타났다($p < .001$). 이러한 결과는 검사 길이가 길어질수록 숙달확률 추정의 정확도가 높아진다고 해석할 수 있다. 둘째, 인지요소 간 상관성이 높아질수록 p 와 \hat{p} 의 RMSE는 약간씩 줄어드는 것으로 나타났으나 이러한 차이는 통계적으로 유의하지 않았다($p < .05$).

정리하면, 숙달확률의 복원 관점에서 검사 길이가 길어질수록 \hat{p} 가 p 를 정확하게 복원하는 것을 확인할 수 있었으며 인지요소 간 상관 정도는 진숙달

확률 복원에 별다른 영향을 미치지 않는 것으로 나타났다.

다음으로 모의실험 조건에 따라 진속달여부(m)와 추정된 속달여부(\hat{m})의 분류정확도가 어떻게 달라지는지 살펴보면 첫째, 검사 길이가 길어질수록 m 과 \hat{m} 의 속달여부 분류정확도가 높아지는 것을 확인할 수 있었다. 분류정확도 비율을 로짓으로 변환하여 분산분석을 실시한 결과, 이러한 차이는 통계적으로 유의미하여($p < .001$) 문항 수가 길어질수록 속달여부 분류정확도가 높아지는 것을 확인할 수 있었다. 둘째, 인지요소 간 상관 정도에 따라 m 과 \hat{m} 의 분류정확도는 일정한 패턴을 보이지 않았으며, 상관 정도에 따른 차이는 통계적으로 유의하지 않았다($p < .05$).

정리하면, 속달여부 분류정확도 관점에서 검사 길이가 길어질수록 \hat{m} 이 m 을 정확하게 추정하는 것을 확인할 수 있었으며 인지요소 간 상관 정도는 진속달여부 복원에 별다른 영향을 미치지 않는 것으로 나타났다.

본 연구에서는 베이지안 네트워크를 적용한 모의실험 연구를 통하여 피험자 능력 추정의 정확성을 진능력모수의 복원의 관점에서 확인하고자 하였으며, 연구 결과 진속달여부와 추정된 속달여부의 분류정확도가 상당히 높은 것으로 나타났다. 이러한 결과는 베이지안 네트워크를 활용한 능력 추정이 피험자의 다차원적인 능력을 비교적 정확하게 추정하는 것으로 해석할 수 있다. 다만 일부 조건에서 속달확률과 속달여부 분류정확성이 상대적으로 낮게 나타났으며 이러한 결과에 대해 논의되어야 할 부분을 정리하면 다음과 같다.

2. 논의

베이지안 네트워크는 질적·양적 자료 모두 손쉽게 적용될 수 있고, 확률적

결과 해석이 합리적으로 제시될 수 있으며, 결측치가 존재하는 경우에도 주어진 자료 안에서 최선의 추정이 가능하다는 장점을 가지고 있다. 그럼에도 불구하고, 교육 분야에서 베이지안 네트워크를 적용함에 있어서 피험자의 능력이 얼마나 정확하게 추정될 수 있는지 평가하는 연구는 매우 드물었다. 본 연구는 이와 같은 문제의식 하에서 베이지안 네트워크의 적용 시 피험자 능력모수 추정이 얼마나 정확하게 이루어질 수 있는지를 확인하고자 숙달확률 간 순위상관계수와 RMSE, 숙달여부 분류정확도라는 기준을 사용하여 능력모수 복원의 정확성을 평가하였다. 본 연구에서 도출된 결론을 바탕으로 능력 모수 추정 시 베이지안 네트워크를 활용하기 위해 논의되어야 할 부분을 정리하면 다음과 같다.

첫째, 일반적으로 **검사 길이**가 길어질수록 모수를 정확하게 추정하는 것으로 알려져 있다. 실제로 본 연구에서도 검사 길이가 길어질수록 숙달확률 간 순위 상관계수가 높아지고 RMSE 값이 작아지는 것을 알 수 있었다($p < .001$). 같은 맥락에서 검사 길이가 길어질수록 숙달여부 분류정확도가 높아지는 것을 확인할 수 있었다($p < .001$). 이를 통해 본 연구의 주목적이라 할 수 있는 베이지안 네트워크를 적용하여 능력모수 추정의 정확성을 확인할 수 있었다. 또한 베이지안 네트워크를 적용한 능력모수 추정은 IRT와 마찬가지로 검사 길이가 길어질수록 능력추정의 정확도가 높아진다는 것을 확인하는 계기가 되었다.

둘째, 본 연구에서는 **인지요소 간 상관**이 높을수록 숙달확률을 정확하게 복원하거나, 분류정확도가 높아지는 일정한 패턴을 보이지 않았다. 이는 본 연구에서 인지요소 간 상관은 진능력 복원에 별다른 영향을 미치지 않는다고 해석할 수 있다. 다만 Gierl, Leighton, & Tan(2006)은 인지요소 간 상관이 높아질수록 분류정확도와 일관성이 낮아진다고 보고하였고, Cui, Gierl, & Chang(2012)은 인지 요소 간 관계가 종속적일수록 분류정확도와 분류일치

도가 높다고 보고한 결과를 반영하여 인지요소 간 상관정도가 진능력모수 복원에 어떠한 영향을 미치는지를 확인하기 위한 추가적인 연구가 필요할 것으로 보인다.

셋째, 인지요소 숙달확률 추정은 베이저안 네트워크의 강점이라 할 수 있다. 그러나 본 연구에서 베이저안 네트워크를 통해 복원된 숙달확률은 .48 ~ .51의 범위를 보여 비교적 낮은 복원력을 보였다. 하나의 인지요소를 측정하는 문항 비율이 높을수록 분류정확도가 높아진다고 보고한 Chiu, Douglas, &Li(2009)의 연구 결과를 참고하더라도 본 연구에서 인지요소와 연결된 문항 수의 비율은 대수학을 측정하는 문항 8개, 삼각함수를 측정하는 문항 7개로 적절하다고 할 수 있다. 그럼에도 불구하고 숙달확률이 비교적 낮은 수치를 보인 원인으로서는 다음의 사항을 고려해 볼 수 있을 것이다.

1) 본 연구에서 사용한 생성문항모수는 <표 V-1>에 제시한 바와 같이 추측도와 부주의 오류와 같은 부정적인 문항모수치가 높은 문항이 상당수 포함되어 있었다.

<표 V-1> 부정적인 문항모수(추측도, 부주의오류)

| 문항번호 | 추측(p_{01}) | 부주의오류(p_{10}) | 문항번호 | 추측(p_{01}) | 부주의오류(p_{10}) |
|------|----------------|-------------------|------|----------------|-------------------|
| 1번 | 0.448 | 0.205 | 11번 | 0.197 | 0.168 |
| 2번 | 0.249 | 0.157 | 12번 | 0.301 | 0.198 |
| 3번 | 0.356 | 0.243 | 13번 | 0.338 | 0.113 |
| 4번 | 0.150 | 0.358 | 14번 | 0.256 | 0.200 |
| 5번 | 0.211 | 0.359 | 15번 | 0.225 | 0.342 |
| 6번 | 0.236 | 0.365 | 16번 | 0.236 | 0.427 |
| 7번 | 0.288 | 0.071 | 17번 | 0.346 | 0.282 |
| 8번 | 0.272 | 0.285 | 18번 | 0.279 | 0.279 |
| 9번 | 0.406 | 0.167 | 19번 | 0.246 | 0.361 |
| 10번 | 0.310 | 0.177 | 20번 | 0.154 | 0.276 |

*추측 또는 부주의 오류가 0.3이상인 문항은 음영 처리

인지요소를 숙달하지 못한 피험자가 정답할 확률인 추측도(p_{01})가 0.3이상인 문항이 7개, 인지요소를 숙달하였음에도 불구하고 오답할 비율인 부주의에 의한 실수(p_{10})가 0.3이상인 문항이 6개로 나타났다. 추측도와 부주의에 의한 실수가 0.5이상으로 심각하게 잘못된 문항은 없었지만, 연구 결과에서 비교적 높지 않은 분류정확도를 보인 원인으로는 추측도와 부주의 오류와 같은 부정적인 문항모수치가 높은 문항이 상당수 포함되어 있기 때문이라 할 수 있다. 이는 부정적인 문항모수(추측도, 부주의오류) 값이 작을수록 분류정확도가 높아진다고 보고한 김지효(2013), Chiu, Douglas, & Li(2009), Delatorre, Hong, & Deng(2010), Cui, Gierl, & Chang(2012)의 연구 결과와 일치한다. 2) 베이저안 네트워크를 포함한 인지진단모형은 기본적으로 숙달확률 자체보다 숙달여부에 중점을 두고 모형이 구성되었다. 선행연구들은 숙달확률보다는 주로 숙달여부의 분류정확도에 집중하여 연구를 수행하였다. 3) 심리측정이론에서 '진확률'에 대한 복원은 정확하게 이루어지기 어렵다. 예를 들어 3PLM에서 추측도 모수 복원의 정확도는 순위 상관계수 0.6을 넘기 어려운 현실이다(Kang과 Cohen, 2007). 비교적 낮은 숙달확률을 보인 원인을 종합해 봤을 때 향후 베이저안 네트워크로 능력모수를 추정할 경우, 추측도가 낮고 부주의 오류가 적은 양호한 문항 모수를 사용한다면 지금보다 더 높은 추정의 정확도를 보일 것이라 예상한다. 또한 향후 베이저안 네트워크를 통해 능력모수 추정 시 더 많은 문항 수와 더 다양한 문항정보 그리고 체계적으로 구성된 Q행렬과 같은 양질의 문항으로 이루어진 문항모수 사용이 전제되어야만 베이저안을 적용한 능력모수 추정 시 피험자 능력 추정의 정확도가 더욱 높아질 것으로 보인다.

베이저안을 통해 추정한 숙달여부 분류정확도는 비교적 정확하게 복원되었다. 본 연구에서 숙달여부 k개 일치도는 약 51.9~57.3%이고, k-1개 일치도는 91.4~94.0% 수준으로 나타나 높은 분류정확도를 보였다.

본 연구의 제한점과 후속 연구의 필요성을 정리하면 다음과 같다.

첫째, 현실적 상황을 보다 포괄적으로 고려하기 위하여 본 연구에서 다른 ‘검사의 길이’와 ‘인지요소 간 상관정도’ 모의실험 요인 이외에 추가적인 요인을 고려할 필요가 있다. 예를 들어, 피험자의 수나 피험자의 능력수준 분포에 따라 분류정확도가 어떻게 달라지는지, 인지요소별 문항 구성 비율을 고려하여 인지요소 수와 인지 요소를 구성하는 문항의 수에 따라 분류 정확도가 어떻게 달라지는지 확인하는 연구가 필요하다.

둘째, 본 연구는 능력모수가 정확하게 추정되는지를 확인하기 위해 검사 길이($i=20, 40, 60$)와 인지요소 간 상관($r=0, 0.4, 0.8$)을 고려한 9가지 모의실험 조건을 제시하였다. 그러나 한 요인 내에서 보다 많은 조건들을 추가적으로 고려하는 연구의 확장이 필요할 것으로 보인다. 예를 들면 검사 길이를 5개, 15개, 30개, 60개 등으로 조금 더 구체화하여 분류정확도 연구를 수행해 볼 수 있다. 또, 본 연구에서는 2개의 인지요소를 사용하였지만 현실에서는 보다 다양한 인지요소 수가 흔히 존재하기 때문에 인지 요소의 수를 달리하여 분류 정확도를 확인하는 연구가 필요하다.

셋째, 이 연구에서는 베이지안 네트워크를 적용하여 능력모수 추정의 정확성을 확인하는 연구이므로 문항모수는 별도로 추정하지 않고 실제 데이터로부터 추출한 생성문항모수를 사용하였다. 이 방법은 피험자의 능력 모수 추정의 정확성 확인에 집중하기 위하여 채택되었고, CAT나 CBT에서 능력 모수 추정 시 문항모수를 고정하는 방법을 일반적으로 사용한다. 하지만 현실에서는 문항모수를 미리 알 수 없는 경우도 많이 존재하고, 문항모수에 따라 능력모수도 달라질 수 있기 때문에, 향후 연구에서는 모의실험 자료에 대하여 베이지안 네트워크를 적용하여 문항모수와 능력모수를 함께 추정하면서 그 복원의 정확성을 평가하는 접근 방법도 고려해 볼 수 있을 것이다. 이러한 방식에서는 피험자의 수가 중요한 모의실험 요인으로 고려될 필요가

있다.

넷째, 이 연구에서는 능력모수 복원의 정확성을 평가함에 있어서 베이지안 네트워크만을 고려하였다. 그러나 동일 모의실험 자료에 대해 DINA, DINO, Fusion model, Rule-Space Model 등과 같은 다른 인지진단 모형이나 다차원 문항반응이론을 적용한 결과와 비교해보는 것도 유의미한 연구가 될 수 있다. 또한, 인지요소숙달 추정 방법 간의 차이를 비교해 볼 필요가 있다. 본 연구에서는 숙달 확률을 구한 후 0.5를 기준으로 인지요소 숙달 여부를 판단하는 MCMC 방법을 고려하였으나, 인지요소 숙달여부를 바로 추정하는 MLE, MAP 방법을 선택한 경우에도 본 연구와 유사한 결과가 도출되는지 추정 방법 간 차이를 비교하는 연구가 필요하다. 즉 다양한 모의실험 조건에 있어서 피험자의 능력을 추정하기 위하여 어떠한 모형이나 방법을 사용하는 것이 보다 적절한 것인지, 그리고 모형에 따라 추정 정확성의 차이가 존재하는지 비교 연구할 필요가 있다.

그 외에도 숙달여부 분류정확성과 효율적인 분석 시간을 고려한 최적의 인지요소 개수를 탐색하는 연구가 수행될 필요가 있으며, 덧붙여 본 연구에서는 수학 검사 자료를 사용하였지만 수학 외에도 과학이나 사회, 영어와 같은 다른 과목으로부터의 자료를 분석하여 과목의 특성이 능력 추정의 정확성에 어떠한 영향을 미치는지도 살펴볼 필요가 있다.

본 연구를 통해 교육측정분야에서 피험자의 능력모수 추정 시 베이지안 네트워크의 적용 가능성을 확인할 수 있었으며 이러한 결과를 바탕으로 향후 베이지안 네트워크가 활용될 수 있는 방안을 탐색해 보면 다음과 같다.

진단평가 또는 형성평가에서 피험자의 다차원적 능력을 효율적으로 추정하기 위하여 베이지안 네트워크를 활용할 수 있다. 다수의 독립적 변수 간의 관계를 그래픽으로 명확하게 표현할 수 있는 베이지안 네트워크의 장점을 적극 활용하여, 피험자 개인의 강·약점에 대한 구체적인 피드백을 도

식화하여 그래픽으로 제공해 줄 수 있다. 이를 제공받은 피험자는 본인이 중점적으로 보충해야 할 내용이 무엇인지, 완벽하게 숙달한 내용은 무엇인지 쉽게 파악할 수 있다. 피험자 개개인의 결과가 지속적으로 누적된다면 교사는 특정 인지요소 숙달정도의 추이를 확인하거나, 학생들의 이해 정도에 따른 효과적인 수업지도 방안을 수립 시 활용할 수 있을 것이다.

또, 베이지안 네트워크를 적용하여 피험자의 학업성취 프로파일 특성 즉, 인지요소에 대한 숙달정도를 분석하는데 활용될 수 있다. 예를 들어 특정 인지요소에 대해 이해도가 다문화학생, 새터민학생 등과 같은 집단에서 현저히 낮게 나타났다면, 그 원인을 분석하고 집단의 특성을 고려한 차별화된 방안이 마련되어야 할 것이다.

참 고 문 헌

- 강태훈, 박찬호, 김인숙(2011). 검사 길이와 수행등급 개수에 따른 성취수준 분류 일관도 및 정확도 연구. *교육평가연구*, 24(4), 1017-1038.
- 강태훈(2014). 베이지안 네트워크를 활용한 교육 및 심리검사의 문항분석. *교육평가연구*, 27(4), 897-923.
- 고왕경(2003). *확률의 이해*. 서울:경문사
- 고길곤(2014). *통계학의 이해와 활용*. 서울:문우사
- 김달호(2013). *R과 winBUGS를 이용한 베이지안 통계학 2판*. 서울:자유아카데미.
- 김명화(2012). 기초학력 진단평가 결과보고에 인지진단 모델 적용 가능성 탐색. *한국교육학연구*, 18(3), 5-24.
- 김성훈(2005). 인지구조모형에 근거한 학생의 지식상태의 진단: 규칙장 이론의 적용. *교육학연구*, 43(1), 81-107.
- 김성호(1994). 과제수행능력과 과제수행과의 관계모형에 대한 통계적 접근 방법. *교육평가연구*, 7(2), 177-200.
- 김성호, 김은희(1995) 중1 수학 검사자료를 사용한 통계적 능력진단모형의 개발과 활용. *교육평가연구*, 8(1). 153-181.
- 김수진, 송미영, 김선희(2008). Fusion Model에 의한 수학 능력 진단을 위한 Q-행렬의 정교화. *교육평가연구*, 21(2). 115-139.
- 김수진(2010). Q-행렬 변화에 따른 혼합모형(Fusion Model) 모수 비교. *교육평가연구*, 23(2), 487-507.
- 김지효(2013). DINA와 DINO모형에서 응시생 분류 정확성에 영향을 미치는 요인 탐구 : 응시생 분류방법을 중심으로. 충남대학교 박사학위 논문.
- 김희경, 한정아, 최숙기, 김부미(2012). 인지진단모형을 적용한 학업성취

- 프로파일 분석 및 결과 보고 방안. 한국교육과정평가원 연구자료 RRE 2012-7.
- 나선웅(2012). 컴퓨터 적응형 검사를 위한 베이지안 네트워크 기법. 한양대학교 석사학위 논문.
- 나선웅, 김경수, 최용석(2012). 베이지안 네트워크를 적용한 컴퓨터 기반 적응형 평가. **정보과학회논문지:소프트웨어 및 응용**, 39(6), 497-506.
- 남기훈(2014). 베이지안 네트워크를 이용한 복합재난 위험성 평가에 관한 연구. 인제대학교 박사학위 논문.
- 박상희, 김남순, 이선재, 홍성두(2009). 특수교육연구에 있어 베이지안 통계 분석법의 적용 가능성에 대한 이론적 탐색. **특수아동교육연구**, 11(3), 341-356.
- 성태제(2009). **문항반응이론의 이해와 적용**. 서울:교육과학사.
- 심혜진(2015). 다분 문항반응모형에서 사후기대추정법의 능력모수 추정 정확성에 관한 연구 : EAPrp와 EAPss를 중심으로. 성신여자대학교 석사학위 논문.
- 정용규, 진훈, 김인철(2001). 베이지안 망을 이용한 불임요인 분석 및 가임 예측. **한국정보과학회 학술발표논문집**, 28(2), 7-9
- 정용규, 이연주, 원재강(2012). 베이지안 알고리즘을 이용한 유방암 진단 예측모델. **한국인터넷방송통신학회**, 12(2), 175-180.
- 황선아(2005). 베이지안 확률을 이용한 CAT의 효과분석. 한양대학교 교육대학원 석사학위논문.
- 황선아, 최용석(2005). 베이지안 확률을 적용한 CAT의 효과 분석. **한국인터넷정보학회**, 6(1), 463-466.
- 황정규(1998). **학교학습과 평가**. 서울:교육과학사.
- 황정규, 서민원, 최종근, 김민성, 양명희, 김재철, 강태훈, 이대식, 김준엽, 신종호, 박인우, 김동일(2011). **교육평가의 이해**. 서울:학지사.

- Acid, S., de Campos, L. M., Fernández-Luna, J. M., Rodriguez, S., Rodriguez, J. M., & Salcedo, J. L. (2004). A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial intelligence in medicine*, 30(3), 215-232.
- Almond, R. G., DiBello, L. V., Moulder, B., & Zapata Rivera, J. D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement*, 44(4), 341-359.
- Chiu, C. Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74(4), 633-665.
- Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49(1), 19-38.
- De La Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249.
- DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. *Cognitively diagnostic assessment*, 361-389.
- Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs, NJ: Prentice-hall.
- Elsner, J. B., & Bossak, B. H. (2001). Bayesian analysis of US hurricane climate. *Journal of Climate*, 14(23), 4341-4350.

- Embretson, S. E. (1999). *Cognitive psychology applied to testing*. In F. T. Durso(ed.), *Handbook of applied cognition*, pp.629-660. NewYork, NY:John Wiley & Sons Ltd.
- Hartz, S. M. (2002). *A Bayesian Framework for the Unified Model for Assessing Cognitive Abilites: Blending Theory with Practicality*. Doctoral dissertation, University of Illinois at Urbana-Champaign.
- Huebner, Wang (2011). A Note on Comparing Examinee Classification Methods for Cognitive Diagnosis Models. *Educational and Psychological Measurement*, 71(2), 407-419.
- Jensen, F. V. (1996). *Introduction to Bayesian Networks*. Advanced Herd Management.
[http://www.prodstyr.ihh.kvl.dk/vp/2009/slides/BayesNet -I-6.pdf](http://www.prodstyr.ihh.kvl.dk/vp/2009/slides/BayesNet-I-6.pdf).
- Johnson, A. P. (1951). Notes on a suggested index of item validation: the U-L index, *Journal of Educational Psychology*. 62, 499-504.
- Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. NEWYORK :CRC Press.
- García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*, 49(3), 794-808.
- Gierl, M. J., Leighton, J. P., & Tan, X. (2006). Evaluating DETECT classification accuracy and consistency when data display complex structure. *Journal of Educational Measurement*, 43(3), 265-289.
- Gilks, W. R., & Roberts, G. O. (1996). Strategies for improving MCMC. *Markov chain Monte Carlo in practice*, 6, 89-114.
- Gogate, V., Dechter, R., Bidyuk, B., Rindt, C., & Marca, J. (2012).

- Modeling transportation routines using hybrid dynamic mixed networks.* arXiv preprint arXiv:1207.1384.
- Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement*, 31(4), 331-358.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17-24.
- Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan fellow working papers in second or foreign language assessment*, 9, 17-46.
- Nachman, I. (2004). *Probabilistic modeling of gene regulatory networks from data*. Stanford University.
<http://noodle.cs.huji.ac.il/~nirf/Papers/NachmanPHD.pdf>
- Miller, I., Miller, M. (2003). *John E. Freund's Mathematical Statistics with Applications (8th Edition)*. Pearson Education India.
- Rina Dechter (2012). bayesian networks and belief propagation: from rumelhart to pearl to today.
<https://www.ics.uci.edu/~dechter/talks/ein-gedi-v6.pdf>
- Solow, A. R. (1988). A Bayesian approach to statistical inference about climate change. *Journal of climate*, 1(5), 512-521.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). WinBUGS User Manual(version1.4).
<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.
- Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika*, 67(4), 485-518.

- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of educational measurement*, 20(4), 345-354.
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively diagnostic assessment*, 327-359.
- Vomlel, J. (2004). *Bayesian networks in educational testing*. Int. J. Uncertain. Fuzziness and Knowledge-Based Systems, 12(supp01), 83-100.
- Lee, Y. S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing*, 11(2), 144-177.

ABSTRACT

A Simulation Study on the Recovery of the Ability Parameter of Bayesian Network

Deok-Jeong, Kim

Department of Education

The graduate school of Sungshin Women's University

This study attempted to confirm examinee's accuracy of the estimation for ability parameters utilizing the bayesian network through a simulation study. For this, actual test data acquired from 3,000 respondents was used. The test is composed of a total of 20 problems including 10 problems about cognitive attribute-algebra(A1) and 10 problems about trigonometry(A2). Actual data was utilized to extract generating item parameters(CPT) and utilize binary normal distribution and create true ability parameters. The created true ability parameters were utilized to estimate simulated examinee's ability parameters by differentiating the 'correlations between cognitive attributes' and 'test length.'

The ability parameter estimated in the bayesian network means an attribute proficiency ratio and mastery and non-mastery of cognitive attribute to each cognitive attribute. In an attempt to evaluate the accuracy of the estimated ability parameters applying the bayesian network, the true ability parameters and estimated ability parameters

were compared, and rank-order correlation coefficients and mastery and non-mastery of cognitive attributes classification accuracy between the attribute proficiency ratios were calculated.

The findings are as follows: First, when it comes to the estimation accuracy of the ability parameters according test length, the longer the test length became, the higher the correlation coefficient between the attribute proficiency ratios of the true ability parameter(p) and the ability parameter(\hat{p}) showed. Likewise, the longer the test length became, the higher the classification accuracy ratios of the mastery and non-mastery of cognitive attributes(m) of the true ability parameter and the mastery and non-mastery of cognitive attributes(\hat{m}) of the estimated ability parameter showed. Through this, the estimation of ability parameters applying the bayesian network was confirmed to be higher in estimation accuracy as the test length becomes longer.

Second, when it comes to the accuracy of parameter estimation according to the correlation between cognitive attributes, the attribute proficiency ratio of the true ability parameter(p) and ability parameter(\hat{p}) did not have a big difference according t the correlation between cognitive attributes. Likewise, the mastery and non-mastery of cognitive attributes(m) of the true ability parameter and the classification accuracy of mastery and non-mastery of cognitive attributes(\hat{m}) of the estimated ability parameter did not show a regular pattern according to the correlation. Through this, it was confirmed that the correlation between cognitive attributes did not have a big influence on the accuracy of the estimation for ability parameters. Through these results, the accuracy of

the estimation for ability parameters could be confirmed by utilizing the bayesian network.

Keywords : bayesian network, ability parameter, classification accuracy, cognitive attribute, simulation study

부 록

- 부록 1. 모의피험자 생성을 위한 WinBUGS 코드
- 부록 2. 능력모수 추정 정확성 확인을 위한 WinBUGS 코드
- 부록 3. 베이지안 네트워크를 통해 실제 자료로부터 추정된 생성문항모수
- 부록 4. 베이지안 네트워크에 의한 문항 분석 결과
- 부록 5. 진속달확률과 추정된 속달확률의 순위상관계수-대수학
- 부록 6. 진속달확률과 추정된 속달확률의 순위상관계수-삼각함수
- 부록 7. 진속달확률과 추정된 속달확률의 RMSE-대수학
- 부록 8. 진속달확률과 추정된 속달확률의 RMSE-삼각함수
- 부록 9. 진속달여부와 추정된 속달여부의 분류정확성-대수학
- 부록 10. 진속달여부와 추정된 속달여부의 분류정확성-삼각함수
- 부록 11. 모의실험 조건에 따른 대수학(A1) 속달여부 분류정확도
- 부록 12. 모의실험 조건에 따른 삼각함수(A2) 속달여부 분류정확도

[부록 1] 모의피험자 생성을 위한 WinBUGS 코드

```
# 2개의 인지요소를 가지고 있는 20개 문항의 능력모수 추정.
model
{
# student model
for (i in 1:N) {
  theta[i,1] ~ dbern(alpha1[i]);
  theta[i,2] ~ dbern(alpha2[i]);
  tau[i,1] <- theta[i,1] + 1;
  tau[i,2] <- theta[i,2] + 1;
# priors on conditionl probabilities among skills
  alpha1[i] ~ dunif(0,1);
  alpha2[i] ~ dunif(0,1);
}
mas1pro <- sum(theta[1:N,1]) / N;
mas2pro <- sum(theta[1:N,2]) / N;

# evidence model 1: items requiring skill1 only
# algebra(skill1): items 1,2,3,4,5,6,8,10
for (i in 1:N){
  r[i,1] ~ dbern(pi[1, tau[i,1]]);
  r[i,2] ~ dbern(pi[2, tau[i,1]]);
  r[i,3] ~ dbern(pi[3, tau[i,1]]);
  r[i,4] ~ dbern(pi[4, tau[i,1]]);
  r[i,5] ~ dbern(pi[5, tau[i,1]]);
  r[i,6] ~ dbern(pi[6, tau[i,1]]);
  r[i,8] ~ dbern(pi[8, tau[i,1]]);
  r[i,10]~ dbern(pi[10,tau[i,1]]);
}

# evidence model 2: items requiring skill2 only
# tri(skill2): items 12,13,14,15,16,18,19
for (i in 1:N){
  r[i,12]~ dbern(pi[12,tau[i,2]]);
  r[i,13]~ dbern(pi[13,tau[i,2]]);
  r[i,14]~ dbern(pi[14,tau[i,2]]);
  r[i,15]~ dbern(pi[15,tau[i,2]]);
  r[i,16]~ dbern(pi[16,tau[i,2]]);
  r[i,18]~ dbern(pi[18,tau[i,2]]);
  r[i,19]~ dbern(pi[19,tau[i,2]]);
}

# evidence model 3: items requiring both skill1 and skill2
# items 7,9,11,17,20
```

```

for (i in 1:N){
  theta[i,3] <- (theta[i,1]* theta[i,2]);
  tau[i,3] <- theta[i,3] + 1;
  r[i,7] ~ dbern(pi[7, tau[i,3]]);
  r[i,9] ~ dbern(pi[9, tau[i,3]]);
  r[i,11]~ dbern(pi[11,tau[i,3]]);
  r[i,17]~ dbern(pi[17,tau[i,3]]);
  r[i,20]~ dbern(pi[20,tau[i,3]]);
}
# CPT
beta [1,1] <-      0.4479   ;
beta [1,2] <-      0.7950   ;
beta [2,1] <-      0.2487   ;
beta [2,2] <-      0.8426   ;
beta [3,1] <-      0.3564   ;
beta [3,2] <-      0.7566   ;
beta [4,1] <-      0.1501   ;
beta [4,2] <-      0.6423   ;
beta [5,1] <-      0.2113   ;
beta [5,2] <-      0.6415   ;
beta [6,1] <-      0.2357   ;
beta [6,2] <-      0.6354   ;
beta [7,1] <-      0.2881   ;
beta [7,2] <-      0.9287   ;
beta [8,1] <-      0.2722   ;
beta [8,2] <-      0.7155   ;
beta [9,1] <-      0.4056   ;
beta [9,2] <-      0.8330   ;
beta [10,1]<-      0.3097   ;
beta [10,2]<-      0.8235   ;
beta [11,1]<-      0.1967   ;
beta [11,2]<-      0.8325   ;
beta [12,1]<-      0.3014   ;
beta [12,2]<-      0.8021   ;
beta [13,1]<-      0.3378   ;
beta [13,2]<-      0.8875   ;
beta [14,1]<-      0.2558   ;
beta [14,2]<-      0.8005   ;
beta [15,1]<-      0.2254   ;
beta [15,2]<-      0.6576   ;
beta [16,1]<-      0.2355   ;
beta [16,2]<-      0.5727   ;
beta [17,1]<-      0.3462   ;

```

```
beta [17,2]<- 0.7185 ;  
beta [18,1]<- 0.2790 ;  
beta [18,2]<- 0.7212 ;  
beta [19,1]<- 0.2459 ;  
beta [19,2]<- 0.6393 ;  
beta [20,1]<- 0.1538 ;  
beta [20,2]<- 0.7241 ;  
  
}  
}
```

[부록 2] 능력모수 추정 정확성 확인을 위한 WinBUGS 코드

```
# 진속달확률과 추정된 속달확률의 정확성 확인(조건: 20개 문항, 인지요소 간 상관 0.8)

model
{
  for (i in 1:N) {
    thetc[i,1:2] ~ dmnorm(mu[ ], TT[ , ]);

    alpha1[i] <- phi(thetc[i,1]);
    alpha2[i] <- phi(thetc[i,2]);
  }
  # student model
  for (i in 1:N) {
    theta[i,1] ~ dbern(alpha1[i]);
    theta[i,2] ~ dbern(alpha2[i]);
    tau[i,1] <- theta[i,1] + 1;
    tau[i,2] <- theta[i,2] + 1;
  }

  mas1pro <- sum(theta[1:N,1]) / N;
  mas2pro <- sum(theta[1:N,2]) / N;

  # 두 인지 요소 간 상관 계수 0.8
  corr[1,1] <- 1
  corr[1,2] <- 0.8
  corr[2,1] <- 0.8
  corr[2,2] <- 1
  TT[ 1:2, 1:2] <- inverse(corr[1:2, 1:2])

  # fixed item parameters
  beta [1,1] <- 0.4479 ;
  beta [1,2] <- 0.7950 ;
  beta [2,1] <- 0.2487 ;
  beta [2,2] <- 0.8426 ;
  beta [3,1] <- 0.3564 ;
  beta [3,2] <- 0.7566 ;
  beta [4,1] <- 0.1501 ;
  beta [4,2] <- 0.6423 ;
  beta [5,1] <- 0.2113 ;
  beta [5,2] <- 0.6415 ;
  beta [6,1] <- 0.2357 ;
}
```

```

beta [6,2] <-      0.6354      ;
beta [7,1] <-      0.2881      ;
beta [7,2] <-      0.9287      ;
beta [8,1] <-      0.2722      ;
beta [8,2] <-      0.7155      ;
beta [9,1] <-      0.4056      ;
beta [9,2] <-      0.8330      ;
beta [10,1]<-      0.3097      ;
beta [10,2]<-      0.8235      ;
beta [11,1]<-      0.1967      ;
beta [11,2]<-      0.8325      ;
beta [12,1]<-      0.3014      ;
beta [12,2]<-      0.8021      ;
beta [13,1]<-      0.3378      ;
beta [13,2]<-      0.8875      ;
beta [14,1]<-      0.2558      ;
beta [14,2]<-      0.8005      ;
beta [15,1]<-      0.2254      ;
beta [15,2]<-      0.6576      ;
beta [16,1]<-      0.2355      ;
beta [16,2]<-      0.5727      ;
beta [17,1]<-      0.3462      ;
beta [17,2]<-      0.7185      ;
beta [18,1]<-      0.2790      ;
beta [18,2]<-      0.7212      ;
beta [19,1]<-      0.2459      ;
beta [19,2]<-      0.6393      ;
beta [20,1]<-      0.1538      ;
beta [20,2]<-      0.7241      ;

  for (i in 1:N) {
    for (j in 1:K) {
      x.tmp[i,j] ~ dunif(0,1)
    }
  }

#evidence model 1 : items requiring skill 1 only algebra(skill 1): items 1, 2, 3, 4, 5, 6, 8, 10
for (i in 1:N){
  y.rep[i,1] <- step(beta[1,tau[i,1]] -x.tmp[i,1]) #beta > x.tmp = 1, else 0
  y.rep[i,2] <- step(beta[2,tau[i,1]] -x.tmp[i,2])
  y.rep[i,3] <- step(beta[3,tau[i,1]] -x.tmp[i,3])
  y.rep[i,4] <- step(beta[4,tau[i,1]] -x.tmp[i,4])

```

```

        y.rep[i,5] <- step(beta[5,tau[i,1]] -x.tmp[i,5])
        y.rep[i,6] <- step(beta[6,tau[i,1]] -x.tmp[i,6])
        y.rep[i,8] <- step(beta[8,tau[i,1]] -x.tmp[i,8])
        y.rep[i,10] <- step(beta[10,tau[i,1]] -x.tmp[i,10])

    }

#evidence model 2 : items requiring skill 2 only trigonometry(skill 2): items
12,13,14,15,16,18,19
for (i in 1:N){
    y.rep[i,12] <- step(beta[12,tau[i,2]] -x.tmp[i,12])
    y.rep[i,13] <- step(beta[13,tau[i,2]] -x.tmp[i,13])
    y.rep[i,14] <- step(beta[14,tau[i,2]] -x.tmp[i,14])
    y.rep[i,15] <- step(beta[15,tau[i,2]] -x.tmp[i,15])
    y.rep[i,16] <- step(beta[16,tau[i,2]] -x.tmp[i,16])
    y.rep[i,18] <- step(beta[18,tau[i,2]] -x.tmp[i,18])
    y.rep[i,19] <- step(beta[19,tau[i,2]] -x.tmp[i,19])

}

#evidence model 3 : items requiring skill 1 and skill 2 : items 7, 9, 11, 17, 20
for (i in 1:N){
    theta[i,3] <- (theta[i,1]*theta[i,2]):
    tau[i,3] <- theta[i,3]+1:
        y.rep[i,7] <- step(beta[7,tau[i,3]] -x.tmp[i,7])
        y.rep[i,9] <- step(beta[9,tau[i,3]] -x.tmp[i,9])
        y.rep[i,11] <- step(beta[11,tau[i,3]] -x.tmp[i,11])
        y.rep[i,17] <- step(beta[17,tau[i,3]] -x.tmp[i,17])
        y.rep[i,20] <- step(beta[20,tau[i,3]] -x.tmp[i,20])

}
}

```

[부록 3] 베이지안 네트워크를 통해 실제 자료로부터 추정한 생성문항모수

| | | | | | | | |
|-----|-------|-------|-------|-----|-------|-------|-------|
| 1번 | 문항 반응 | | | 11번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.552 | | 0.448 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.205 | 0.795 | 요소 | 숙달 | 0.168 | 0.833 |
| 2번 | 문항 반응 | | | 12번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.751 | | 0.249 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.157 | 0.843 | 요소 | 숙달 | 0.198 | 0.802 |
| 3번 | 문항 반응 | | | 13번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.644 | | 0.356 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.243 | 0.757 | 요소 | 숙달 | 0.113 | 0.888 |
| 4번 | 문항 반응 | | | 14번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.85 | | 0.15 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.358 | 0.642 | 요소 | 숙달 | 0.2 | 0.801 |
| 5번 | 문항 반응 | | | 15번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.789 | | 0.211 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.359 | 0.642 | 요소 | 숙달 | 0.342 | 0.658 |
| 6번 | 문항 반응 | | | 16번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.764 | | 0.236 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.365 | 0.635 | 요소 | 숙달 | 0.427 | 0.573 |
| 7번 | 문항 반응 | | | 17번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.712 | | 0.288 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.071 | 0.929 | 요소 | 숙달 | 0.282 | 0.719 |
| 8번 | 문항 반응 | | | 18번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.728 | | 0.272 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.285 | 0.716 | 요소 | 숙달 | 0.279 | 0.721 |
| 9번 | 문항 반응 | | | 19번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.594 | | 0.406 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.167 | 0.833 | 요소 | 숙달 | 0.361 | 0.639 |
| 10번 | 문항 반응 | | | 20번 | 문항 반응 | | |
| | | 0 | 1 | | | 0 | 1 |
| | 인지 | 미숙달 | 0.69 | | 0.31 | 인지 | 미숙달 |
| 요소 | 숙달 | 0.177 | 0.824 | 요소 | 숙달 | 0.276 | 0.724 |

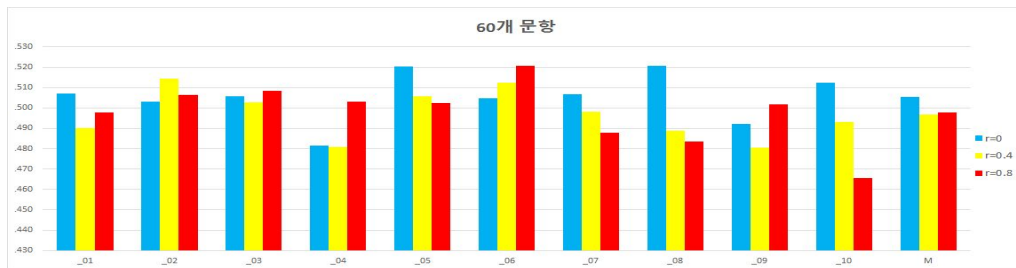
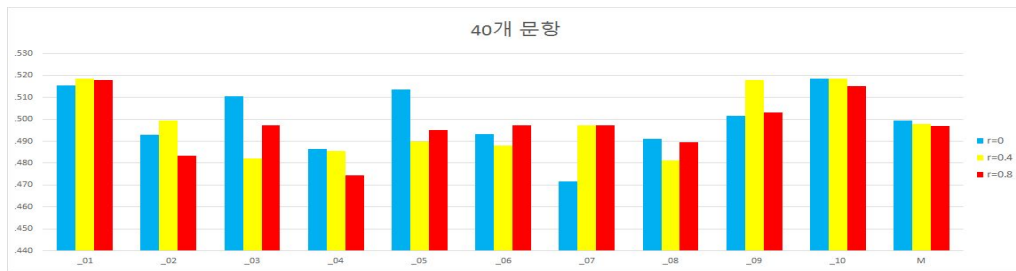
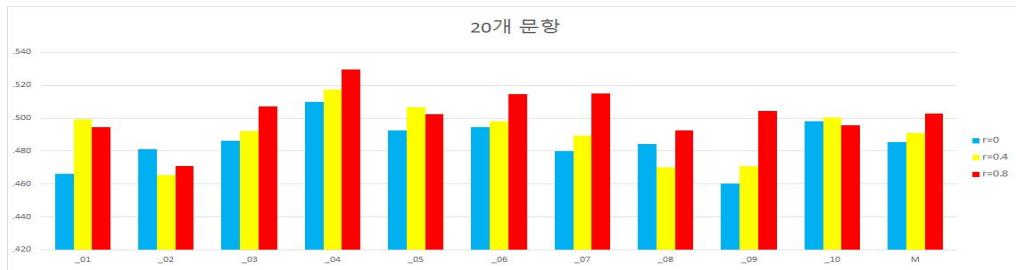
[부록 4] 베이저안 네트워크에 의한 문항 분석 결과

| 문항번호 | p_{00} | 추측(p_{01}) | 부주의오류(p_{10}) | 문항곤란도(p_{11}) | 문항변별도(D_i) |
|------|----------|----------------|-------------------|-------------------|----------------|
| 1번 | 0.552 | 0.448 | 0.205 | 0.795 | 0.347 |
| 2번 | 0.751 | 0.249 | 0.157 | 0.843 | 0.594 |
| 3번 | 0.644 | 0.356 | 0.243 | 0.757 | 0.400 |
| 4번 | 0.850 | 0.150 | 0.358 | 0.642 | 0.492 |
| 5번 | 0.789 | 0.211 | 0.359 | 0.642 | 0.430 |
| 6번 | 0.764 | 0.236 | 0.365 | 0.635 | 0.400 |
| 7번 | 0.712 | 0.288 | 0.071 | 0.929 | 0.641 |
| 8번 | 0.728 | 0.272 | 0.285 | 0.716 | 0.443 |
| 9번 | 0.594 | 0.406 | 0.167 | 0.833 | 0.427 |
| 10번 | 0.690 | 0.310 | 0.177 | 0.824 | 0.514 |
| 11번 | 0.803 | 0.197 | 0.168 | 0.833 | 0.636 |
| 12번 | 0.699 | 0.301 | 0.198 | 0.802 | 0.501 |
| 13번 | 0.662 | 0.338 | 0.113 | 0.888 | 0.550 |
| 14번 | 0.744 | 0.256 | 0.200 | 0.801 | 0.545 |
| 15번 | 0.775 | 0.225 | 0.342 | 0.658 | 0.432 |
| 16번 | 0.765 | 0.236 | 0.427 | 0.573 | 0.337 |
| 17번 | 0.654 | 0.346 | 0.282 | 0.719 | 0.372 |
| 18번 | 0.721 | 0.279 | 0.279 | 0.721 | 0.442 |
| 19번 | 0.754 | 0.246 | 0.361 | 0.639 | 0.393 |
| 20번 | 0.846 | 0.154 | 0.276 | 0.724 | 0.570 |

*추측 또는 부주의 오류가 0.3이상인 문항은 음영 처리

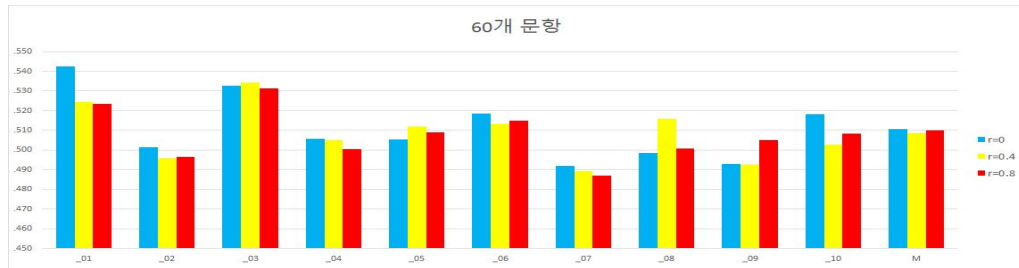
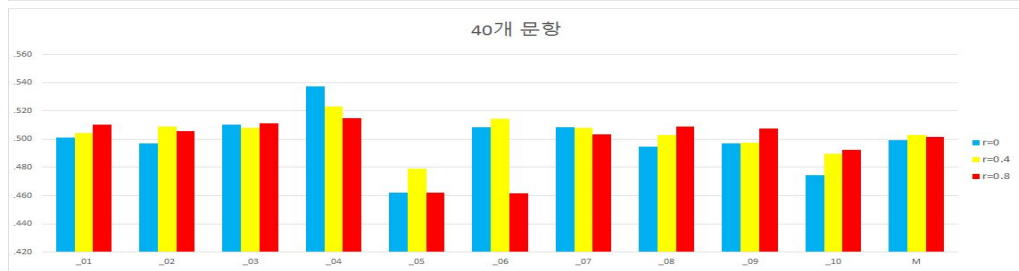
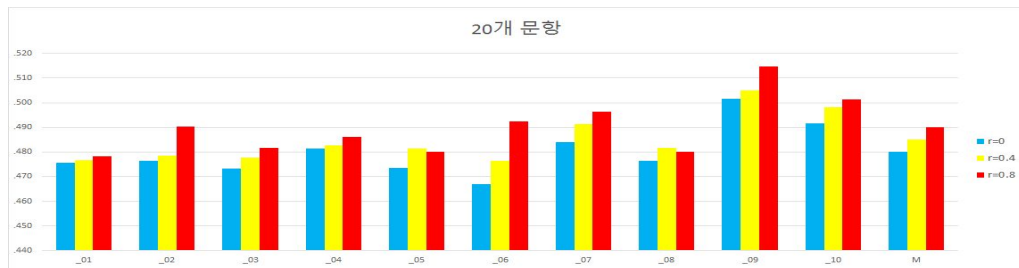
[부록 5] 진속달확률과 추정된 속달확률의 순위상관계수-대수학

| A1 | | \hat{p}_1 | \hat{p}_2 | \hat{p}_3 | \hat{p}_4 | \hat{p}_5 | \hat{p}_6 | \hat{p}_7 | \hat{p}_8 | \hat{p}_9 | \hat{p}_{10} | 평균 (SD) |
|------|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------------|-----------------|
| i=20 | r=0 | .466 | .481 | .486 | .510 | .493 | .494 | .480 | .484 | .460 | .498 | .485 (0.015) |
| | r=.4 | .499 | .466 | .492 | .517 | .507 | .498 | .490 | .470 | .471 | .500 | .491 (0.017) |
| | r=.8 | .494 | .471 | .507 | .530 | .503 | .515 | .515 | .493 | .504 | .496 | .503 (0.016) |
| i=40 | r=0 | .515 | .493 | .510 | .486 | .514 | .493 | .472 | .491 | .502 | .519 | .500 (0.015) |
| | r=.4 | .518 | .499 | .482 | .486 | .490 | .488 | .497 | .481 | .518 | .518 | .498 (0.015) |
| | r=.8 | .518 | .483 | .497 | .475 | .495 | .497 | .497 | .490 | .503 | .515 | .497 (0.013) |
| i=60 | r=0 | .507 | .503 | .506 | .481 | .520 | .505 | .507 | .521 | .492 | .512 | .505 (0.012) |
| | r=.4 | .490 | .514 | .503 | .481 | .506 | .512 | .498 | .489 | .481 | .493 | .497 (0.012) |
| | r=.8 | .498 | .507 | .508 | .503 | .502 | .521 | .488 | .483 | .502 | .466 | .498 (0.015) |



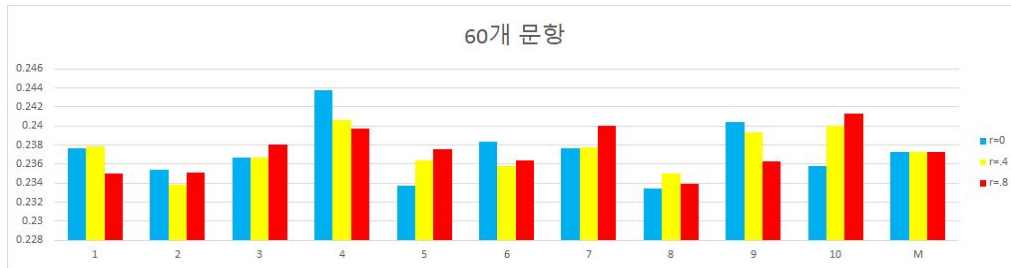
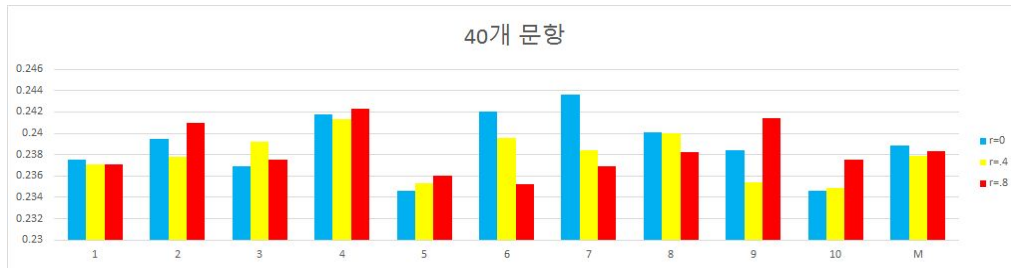
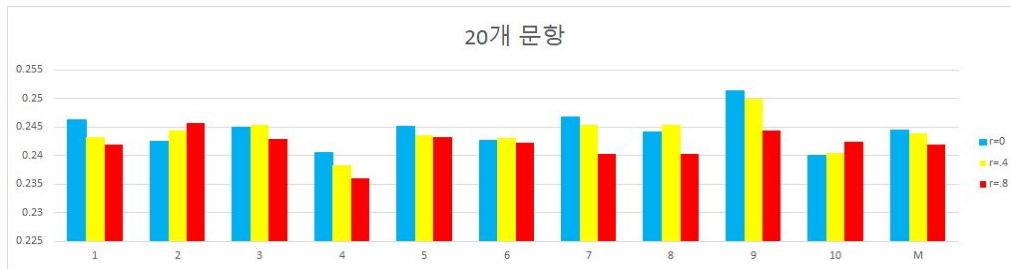
[부록 6] 진속달확률과 추정된 속달확률의 순위상관계수-삼각함수

| A2 | | \hat{p}_1 | \hat{p}_2 | \hat{p}_3 | \hat{p}_4 | \hat{p}_5 | \hat{p}_6 | \hat{p}_7 | \hat{p}_8 | \hat{p}_9 | \hat{p}_{10} | 평균 (SD) |
|------|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------------|-----------------|
| i=20 | r=.0 | .476 | .476 | .473 | .481 | .473 | .467 | .484 | .476 | .502 | .492 | .480 (0.010) |
| | r=.4 | .477 | .478 | .478 | .483 | .481 | .476 | .491 | .482 | .505 | .498 | .485 (0.010) |
| | r=.8 | .478 | .490 | .482 | .486 | .480 | .492 | .496 | .480 | .515 | .501 | .490 (0.012) |
| i=40 | r=.0 | .501 | .497 | .510 | .537 | .462 | .508 | .509 | .495 | .497 | .475 | .499 (0.020) |
| | r=.4 | .504 | .509 | .508 | .523 | .479 | .514 | .508 | .503 | .497 | .490 | .504 (0.012) |
| | r=.8 | .510 | .506 | .511 | .515 | .462 | .462 | .503 | .509 | .507 | .493 | .498 (0.020) |
| i=60 | r=.0 | .542 | .502 | .533 | .506 | .505 | .519 | .492 | .499 | .493 | .518 | .511 (0.017) |
| | r=.4 | .525 | .496 | .534 | .505 | .512 | .513 | .489 | .516 | .493 | .503 | .509 (0.014) |
| | r=.8 | .524 | .497 | .531 | .500 | .509 | .515 | .487 | .501 | .505 | .508 | .508 (0.013) |



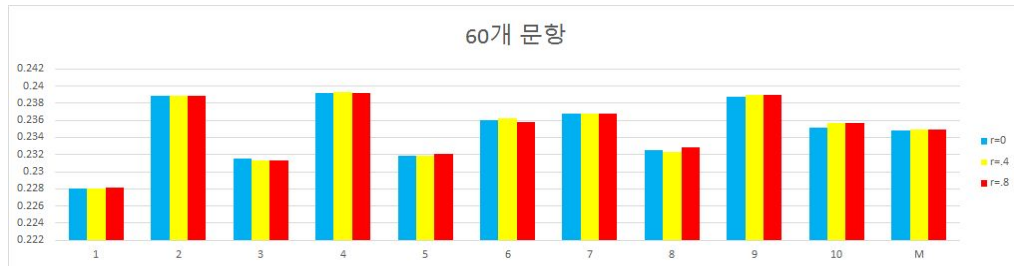
[부록 7] 진속달확률과 추정된 속달확률의 RMSE-대수학

| A1 | | \hat{p}_1 | \hat{p}_2 | \hat{p}_3 | \hat{p}_4 | \hat{p}_5 | \hat{p}_6 | \hat{p}_7 | \hat{p}_8 | \hat{p}_9 | \hat{p}_{10} | 평균 (SD) |
|------|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------------|--------------------------|
| i=20 | r=0 | 0.2463 | 0.2425 | 0.2450 | 0.2405 | 0.2451 | 0.2427 | 0.2467 | 0.2442 | 0.2513 | 0.2400 | .2444 (0.0033) |
| | r=.4 | 0.2432 | 0.2444 | 0.2454 | 0.2383 | 0.2436 | 0.2430 | 0.2454 | 0.2453 | 0.2499 | 0.2404 | .2439 (0.0031) |
| | r=.8 | 0.2420 | 0.2457 | 0.2429 | 0.2360 | 0.2432 | 0.2423 | 0.2403 | 0.2402 | 0.2444 | 0.2424 | .2419 (0.0027) |
| i=40 | r=0 | 0.2375 | 0.2395 | 0.2369 | 0.2418 | 0.2346 | 0.2420 | 0.2436 | 0.2401 | 0.2384 | 0.2346 | .2389 (0.0031) |
| | r=.4 | 0.2371 | 0.2378 | 0.2392 | 0.2413 | 0.2353 | 0.2396 | 0.2384 | 0.2400 | 0.2354 | 0.2349 | .2379 (0.0022) |
| | r=.8 | 0.2371 | 0.2410 | 0.2375 | 0.2423 | 0.2360 | 0.2352 | 0.2369 | 0.2382 | 0.2414 | 0.2375 | .2383 (0.0024) |
| i=60 | r=0 | 0.2377 | 0.2354 | 0.2367 | 0.2438 | 0.2337 | 0.2383 | 0.2377 | 0.2334 | 0.2404 | 0.2358 | .2373 (0.0031) |
| | r=.4 | 0.2379 | 0.2338 | 0.2367 | 0.2406 | 0.2364 | 0.2358 | 0.2378 | 0.2350 | 0.2393 | 0.2400 | .2373 (0.0022) |
| | r=.8 | 0.2350 | 0.2351 | 0.2381 | 0.2397 | 0.2376 | 0.2364 | 0.2400 | 0.2339 | 0.2363 | 0.2413 | .2373 (0.0024) |



[부록 8] 진속달확률과 추정된 속달확률의 RMSE-삼각함수

| A2 | | \hat{p}_1 | \hat{p}_2 | \hat{p}_3 | \hat{p}_4 | \hat{p}_5 | \hat{p}_6 | \hat{p}_7 | \hat{p}_8 | \hat{p}_9 | \hat{p}_{10} | 평균 (SD) |
|------|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------------|---------------------------|
| i=20 | r=.0 | 0.2472 | 0.2459 | 0.2470 | 0.2483 | 0.2452 | 0.2468 | 0.2443 | 0.2462 | 0.2436 | 0.2420 | 0.2456 (0.0019) |
| | r=.4 | 0.2473 | 0.2455 | 0.2468 | 0.2483 | 0.2454 | 0.2463 | 0.2430 | 0.2460 | 0.2429 | 0.2411 | 0.2453 (0.0023) |
| | r=.8 | 0.2471 | 0.2452 | 0.2460 | 0.2477 | 0.2458 | 0.2453 | 0.2430 | 0.2465 | 0.2423 | 0.2405 | 0.2450 (0.0023) |
| i=40 | r=.0 | 0.2422 | 0.2389 | 0.2377 | 0.2341 | 0.2463 | 0.2348 | 0.2385 | 0.2368 | 0.2411 | 0.2439 | 0.2394 (0.0039) |
| | r=.4 | 0.2420 | 0.2388 | 0.2377 | 0.2338 | 0.2459 | 0.2351 | 0.2387 | 0.2369 | 0.2408 | 0.2436 | 0.2393 (0.0038) |
| | r=.8 | 0.2414 | 0.2391 | 0.2378 | 0.2337 | 0.2467 | 0.2349 | 0.2393 | 0.2369 | 0.2414 | 0.2439 | 0.2395 (0.0040) |
| i=60 | r=.0 | 0.2280 | 0.2388 | 0.2315 | 0.2392 | 0.2319 | 0.2360 | 0.2368 | 0.2325 | 0.2387 | 0.2351 | 0.2348 (0.0037) |
| | r=.4 | 0.2280 | 0.2389 | 0.2313 | 0.2393 | 0.2319 | 0.2362 | 0.2368 | 0.2323 | 0.2390 | 0.2357 | 0.2349 (0.0038) |
| | r=.8 | 0.2282 | 0.2388 | 0.2313 | 0.2392 | 0.2321 | 0.2358 | 0.2368 | 0.2328 | 0.2390 | 0.2357 | 0.2349 (0.0037) |



[부록9] 진속달여부와 추정된 속달여부의 분류정확성-대수학

(n=30,000)

A1. 속달여부 분류 정확도

| i=20, r= 0 | | | i=20, r= 0.4 | | | i=20, r= 0.8 | | |
|------------|----------------|-------|--------------|----------------|-------|--------------|----------------|-------|
| | \hat{m} | | | \hat{m} | | | \hat{m} | |
| m | 10844 | 4266 | m | 10822 | 4266 | m | 10812 | 4227 |
| | 4129 | 10761 | | 4045 | 10867 | | 3945 | 11016 |
| | 21,605(72.02%) | | | 21,689(72.30%) | | | 21,828(72.76%) | |
| i=40, r= 0 | | | i=40, r= 0.4 | | | i=40, r= 0.8 | | |
| | \hat{m} | | | \hat{m} | | | \hat{m} | |
| m | 11052 | 3822 | m | 11139 | 3849 | m | 11131 | 3861 |
| | 3992 | 11134 | | 3830 | 11182 | | 3813 | 11195 |
| | 22,186(73.95%) | | | 22,321(74.40%) | | | 22,326(74.42%) | |
| i=60, r= 0 | | | i=60, r= 0.4 | | | i=60, r= 0.8 | | |
| | \hat{m} | | | \hat{m} | | | \hat{m} | |
| m | 11199 | 3719 | m | 11105 | 3770 | m | 10992 | 3778 |
| | 3695 | 11387 | | 3766 | 11359 | | 3836 | 11394 |
| | 22,586(75.29%) | | | 22,464(74.88%) | | | 22,386(74.62%) | |

[부록 10] 진속달여부와 추정된 속달여부의 분류정확성-삼각함수

(n=30,000)

| A2. 속달여부 분류 정확도 | | | n=3,000 | | |
|-----------------|-----------------------|-------|---------|-----------------------|-------|
| i=20, r= 0 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 10781 | 4163 | m | 10726 | 4218 |
| | 4270 | 10786 | | 4198 | 10858 |
| | 21,567(71.89%) | | | 21,584(71.95%) | |
| i=20, r= 0.4 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 10651 | 4293 | m | 11089 | 3872 |
| | 4106 | 10950 | | 3805 | 11234 |
| | 21,601(72.00%) | | | 22,323(74.41%) | |
| i=20, r= 0.8 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11042 | 3919 | m | 11178 | 3745 |
| | 3784 | 11255 | | 3670 | 11445 |
| | 22,297(74.32%) | | | 22,623(75.41%) | |
| i=40, r= 0 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11206 | 3707 | m | 11151 | 3734 |
| | 3695 | 11392 | | 3678 | 11437 |
| | 22,598(75.33%) | | | 22,588(75.29%) | |
| i=40, r= 0.4 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11178 | 3745 | m | 11178 | 3745 |
| | 3670 | 11445 | | 3670 | 11445 |
| | 22,623(75.41%) | | | 22,623(75.41%) | |
| i=40, r= 0.8 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11178 | 3745 | m | 11178 | 3745 |
| | 3670 | 11445 | | 3670 | 11445 |
| | 22,623(75.41%) | | | 22,623(75.41%) | |
| i=60, r= 0 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11206 | 3707 | m | 11151 | 3734 |
| | 3695 | 11392 | | 3678 | 11437 |
| | 22,598(75.33%) | | | 22,588(75.29%) | |
| i=60, r= 0.4 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11151 | 3734 | m | 11151 | 3734 |
| | 3678 | 11437 | | 3678 | 11437 |
| | 22,588(75.29%) | | | 22,588(75.29%) | |
| i=60, r= 0.8 | | | | | |
| | \hat{m} | | | \hat{m} | |
| m | 11178 | 3745 | m | 11178 | 3745 |
| | 3670 | 11445 | | 3670 | 11445 |
| | 22,623(75.41%) | | | 22,623(75.41%) | |

[부록 11] 모의실험 조건에 따른 대수학(A1) 숙달여부 분류정확도(1)

A1. 숙달여부 분류 정확도(1)

문항 수 20개, 상관이 0일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r1i1 | 0 | 1021 | 428 | 0 | 1094 | 435 | 0 | 1087 | 441 | 0 | 1088 | 408 | 0 | 1060 | 408 | 0 | 1106 | 433 | 0 | 1110 | 437 | 0 | 1090 | 412 | 0 | 1089 | 452 | 0 | 1099 | 412 |
| | 1 | 431 | 1120 | 1 | 416 | 1055 | 1 | 424 | 1048 | 1 | 405 | 1099 | 1 | 415 | 1117 | 1 | 403 | 1058 | 1 | 399 | 1054 | 1 | 412 | 1086 | 1 | 417 | 1042 | 1 | 407 | 1082 |
| 정확도 (비율) | 2141(71.4%) | | | 2149(71.6%) | | | 2135(71.2%) | | | 2187(72.9%) | | | 2177(72.6%) | | | 2164(72.1%) | | | 2164(72.1%) | | | 2176(72.5%) | | | 2131(71.0%) | | | 2181(72.7%) | | |

문항 수 20개, 상관이 0.4일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r2i1 | 0 | 1053 | 396 | 0 | 1047 | 482 | 0 | 1094 | 427 | 0 | 1106 | 406 | 0 | 1063 | 411 | 0 | 1095 | 423 | 0 | 1081 | 451 | 0 | 1071 | 445 | 0 | 1101 | 413 | 0 | 1111 | 412 |
| | 1 | 402 | 1149 | 1 | 436 | 1035 | 1 | 407 | 1072 | 1 | 379 | 1109 | 1 | 409 | 1117 | 1 | 396 | 1086 | 1 | 404 | 1064 | 1 | 418 | 1066 | 1 | 411 | 1075 | 1 | 383 | 1094 |
| 정확도 (비율) | 2202(73.4%) | | | 2082(69.4%) | | | 2166(72.2%) | | | 2215(73.8%) | | | 2180(72.7%) | | | 2181(72.7%) | | | 2145(71.5%) | | | 2137(71.2%) | | | 2176(72.5%) | | | 2205(73.5%) | | |

문항 수 20개, 상관이 0.4일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r3i1 | 0 | 1067 | 422 | 0 | 1036 | 454 | 0 | 1094 | 435 | 0 | 1097 | 408 | 0 | 1097 | 430 | 0 | 1084 | 405 | 0 | 1098 | 409 | 0 | 1068 | 400 | 0 | 1084 | 416 | 0 | 1087 | 448 |
| | 1 | 377 | 1134 | 1 | 419 | 1091 | 1 | 388 | 1083 | 1 | 384 | 1111 | 1 | 398 | 1075 | 1 | 402 | 1109 | 1 | 378 | 1115 | 1 | 421 | 1111 | 1 | 386 | 1114 | 1 | 392 | 1073 |
| 정확도 (비율) | 2201(73.4%) | | | 2127(70.9%) | | | 2177(72.6%) | | | 2208(73.6%) | | | 2172(72.4%) | | | 2193(73.1%) | | | 2213(73.8%) | | | 2179(72.6%) | | | 2198(73.3%) | | | 2160(72.0%) | | |

[부록 11] 모의실험 조건에 따른 대수학(A1) 숙달여부 분류정확도(2)

A1. 숙달여부 분류 정확도(2)

문항 수 40개, 상관이 0일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r1i2 | 0 | 1133 | 387 | 0 | 1125 | 369 | 0 | 1068 | 372 | 0 | 1121 | 375 | 0 | 1102 | 369 | 0 | 1108 | 382 | 0 | 1075 | 400 | 0 | 1124 | 396 | 0 | 1058 | 404 | 0 | 1138 | 368 |
| | 1 | 370 | 1110 | 1 | 393 | 1113 | 1 | 413 | 1147 | 1 | 412 | 1092 | 1 | 406 | 1123 | 1 | 388 | 1122 | 1 | 429 | 1096 | 1 | 405 | 1075 | 1 | 396 | 1142 | 1 | 380 | 1114 |
| 정확도 (비율) | 2243(74.8%) | | | 2238(74.6%) | | | 2215(73.8%) | | | 2213(73.8%) | | | 2225(74.2%) | | | 2230(74.3%) | | | 2171(72.4%) | | | 2199(73.3%) | | | 2200(73.3%) | | | 2252(75.1%) | | |

문항 수 40개, 상관이 0.4일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r2i2 | 0 | 1132 | 368 | 0 | 1122 | 385 | 0 | 1074 | 375 | 0 | 1126 | 398 | 0 | 1110 | 388 | 0 | 1099 | 392 | 0 | 1114 | 392 | 0 | 1120 | 412 | 0 | 1113 | 377 | 0 | 1129 | 362 |
| | 1 | 364 | 1136 | 1 | 376 | 1117 | 1 | 403 | 1148 | 1 | 384 | 1092 | 1 | 388 | 1114 | 1 | 398 | 1111 | 1 | 390 | 1104 | 1 | 393 | 1075 | 1 | 358 | 1152 | 1 | 376 | 1133 |
| 정확도 (비율) | 2268(75.6%) | | | 2239(74.6%) | | | 2222(74.1%) | | | 2218(73.9%) | | | 2224(74.1%) | | | 2210(73.7%) | | | 2218(73.9%) | | | 2195(73.2%) | | | 2265(75.5%) | | | 2262(75.4%) | | |

문항 수 40개, 상관이 0.8일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r3i2 | 0 | 1121 | 392 | 0 | 1102 | 391 | 0 | 1095 | 380 | 0 | 1084 | 395 | 0 | 1133 | 381 | 0 | 1087 | 389 | 0 | 1136 | 390 | 0 | 1122 | 387 | 0 | 1109 | 378 | 0 | 1142 | 378 |
| | 1 | 350 | 1137 | 1 | 382 | 1125 | 1 | 394 | 1131 | 1 | 402 | 1119 | 1 | 387 | 1099 | 1 | 393 | 1131 | 1 | 386 | 1088 | 1 | 379 | 1112 | 1 | 375 | 1138 | 1 | 365 | 1115 |
| 정확도 (비율) | 2258(75.3%) | | | 2227(74.2%) | | | 2226(74.2%) | | | 2203(73.4%) | | | 2232(74.4%) | | | 2218(73.9%) | | | 2224(74.1%) | | | 2234(74.5%) | | | 2247(74.9%) | | | 2257(75.2%) | | |

[부록 11] 모의실험 조건에 따른 대수학(A1) 숙달여부 분류정확도(3)

A1. 숙달여부 분류 정확도(3)

문항 수 60개, 상관이 0일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r1i3 | 0 | 1124 | 372 | 0 | 1140 | 342 | 0 | 1107 | 366 | 0 | 1102 | 384 | 0 | 1119 | 365 | 0 | 1151 | 375 | 0 | 1160 | 359 | 0 | 1120 | 367 | 0 | 1072 | 399 | 0 | 1104 | 390 |
| | 1 | 355 | 1149 | 1 | 376 | 1142 | 1 | 376 | 1151 | 1 | 380 | 1134 | 1 | 343 | 1173 | 1 | 374 | 1100 | 1 | 377 | 1104 | 1 | 368 | 1145 | 1 | 390 | 1139 | 1 | 356 | 1150 |
| 정확도 (비율) | 2273(75.8%) | | | 2282(76.1%) | | | 2258(75.3%) | | | 2236(74.5%) | | | 2292(76.4%) | | | 2251(75.0%) | | | 2264(75.5%) | | | 2265(75.5%) | | | 2211(73.7%) | | | 2254(75.1%) | | |

문항 수 60개, 상관이 0.4일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r2i3 | 0 | 1113 | 382 | 0 | 1146 | 368 | 0 | 1128 | 375 | 0 | 1057 | 386 | 0 | 1100 | 361 | 0 | 1161 | 345 | 0 | 1146 | 375 | 0 | 1113 | 381 | 0 | 1082 | 392 | 0 | 1059 | 405 |
| | 1 | 366 | 1139 | 1 | 373 | 1113 | 1 | 348 | 1149 | 1 | 387 | 1170 | 1 | 359 | 1180 | 1 | 372 | 1122 | 1 | 373 | 1106 | 1 | 393 | 1113 | 1 | 401 | 1125 | 1 | 394 | 1142 |
| 정확도 (비율) | 2252(75.1%) | | | 2259(75.3%) | | | 2277(75.9%) | | | 2227(74.2%) | | | 2280(76.0%) | | | 2283(76.1%) | | | 2252(75.1%) | | | 2226(74.2%) | | | 2207(73.6%) | | | 2201(73.4%) | | |

문항 수 60개, 상관이 0.8일 때 숙달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r3i3 | 0 | 1104 | 376 | 0 | 1135 | 360 | 0 | 1125 | 388 | 0 | 1048 | 382 | 0 | 1063 | 372 | 0 | 1127 | 351 | 0 | 1110 | 394 | 0 | 1089 | 373 | 0 | 1111 | 387 | 0 | 1080 | 395 |
| | 1 | 376 | 1144 | 1 | 381 | 1124 | 1 | 350 | 1137 | 1 | 386 | 1184 | 1 | 384 | 1181 | 1 | 387 | 1135 | 1 | 402 | 1094 | 1 | 410 | 1128 | 1 | 379 | 1123 | 1 | 381 | 1144 |
| 정확도 (비율) | 2248(74.9%) | | | 2259(75.3%) | | | 2262(75.4%) | | | 2232(74.4%) | | | 2244(74.8%) | | | 2262(75.4%) | | | 2204(73.5%) | | | 2217(73.9%) | | | 2234(74.5%) | | | 2224(74.1%) | | |

[부록 12] 모의실험 조건에 따른 삼각함수(A2) 속달여부 분류정확도(1)

A2. 속달여부 분류 정확도(1)

문항 수 20개, 상관이 0일 때 속달여부 분류정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r1i1 | 0 | 1080 | 438 | 0 | 1050 | 420 | 0 | 1072 | 418 | 0 | 1081 | 413 | 0 | 1115 | 437 | 0 | 1044 | 429 | 0 | 1107 | 413 | 0 | 1050 | 406 | 0 | 1061 | 388 | 0 | 1121 | 401 |
| | 1 | 414 | 1068 | 1 | 424 | 1106 | 1 | 443 | 1067 | 1 | 412 | 1094 | 1 | 395 | 1053 | 1 | 424 | 1103 | 1 | 435 | 1045 | 1 | 451 | 1093 | 1 | 435 | 1116 | 1 | 437 | 1041 |
| 정확도 (비율) | 2148(71.6%) | | | 2156(71.9%) | | | 2139(71.3%) | | | 2175(72.5%) | | | 2168(72.3%) | | | 2147(71.6%) | | | 2152(71.7%) | | | 2143(71.4%) | | | 2177(72.6%) | | | 2162(72.1%) | | |

문항 수 20개, 상관이 0.4일 때 속달여부 분류정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r2i1 | 0 | 1071 | 447 | 0 | 1047 | 423 | 0 | 1069 | 421 | 0 | 1075 | 419 | 0 | 1106 | 446 | 0 | 1040 | 433 | 0 | 1103 | 417 | 0 | 1038 | 418 | 0 | 1055 | 394 | 0 | 1122 | 400 |
| | 1 | 410 | 1072 | 1 | 420 | 1110 | 1 | 433 | 1077 | 1 | 405 | 1101 | 1 | 391 | 1057 | 1 | 413 | 1114 | 1 | 420 | 1060 | 1 | 447 | 1097 | 1 | 429 | 1122 | 1 | 430 | 1048 |
| 정확도 (비율) | 2143(71.4%) | | | 2157(71.9%) | | | 2146(71.5%) | | | 2176(72.5%) | | | 2163(72.1%) | | | 2154(71.8%) | | | 2163(72.1%) | | | 2135(71.2%) | | | 2177(72.6%) | | | 2170(72.3%) | | |

문항 수 20개, 상관이 0.4일 때 속달여부 분류정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r3i1 | 0 | 1069 | 449 | 0 | 1038 | 432 | 0 | 1064 | 426 | 0 | 1064 | 430 | 0 | 1090 | 462 | 0 | 1037 | 436 | 0 | 1093 | 427 | 0 | 1033 | 423 | 0 | 1045 | 404 | 0 | 1118 | 404 |
| | 1 | 402 | 1080 | 1 | 408 | 1122 | 1 | 424 | 1086 | 1 | 398 | 1108 | 1 | 381 | 1067 | 1 | 408 | 1119 | 1 | 410 | 1070 | 1 | 440 | 1104 | 1 | 413 | 1138 | 1 | 422 | 1056 |
| 정확도 (비율) | 2149(71.6%) | | | 2160(72.0%) | | | 2150(71.7%) | | | 2172(72.4%) | | | 2157(71.9%) | | | 2156(71.9%) | | | 2163(72.1%) | | | 2137(71.2%) | | | 2183(72.8%) | | | 2174(72.5%) | | |

[부록 12] 모의실험 조건에 따른 삼각함수(A2) 속달여부 분류정확도(2)

A2. 속달여부 분류 정확도(2)

문항 수 40개, 상관이 0일 때 속달여부 분류 정확도

| r1i2 | _01 | | | _02 | | | _03 | | | _04 | | | _05 | | | _06 | | | _07 | | | _08 | | | _09 | | | _10 | | |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|---------------|------|------|
| | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| | 0 | 1123 | 400 | 0 | 1097 | 402 | 0 | 1117 | 369 | 0 | 1106 | 348 | 0 | 1103 | 426 | 0 | 1099 | 369 | 0 | 1140 | 356 | 0 | 1093 | 397 | 0 | 1107 | 406 | 0 | 1130 | 373 |
| | 1 | 348 | 1129 | 1 | 396 | 1105 | 1 | 388 | 1126 | 1 | 382 | 1164 | 1 | 415 | 1056 | 1 | 356 | 1176 | 1 | 426 | 1078 | 1 | 363 | 1147 | 1 | 358 | 1129 | 1 | 416 | 1081 |
| 정확도 (비율) | 2252(75.1%) | | | 2202(73.4%) | | | 2243(74.8%) | | | 2270(75.7%) | | | 2159(72.0%) | | | 2275(75.8%) | | | 2218(73.9%) | | | 2240(74.7%) | | | 2236(74.5%) | | | 2211 73.7% | | |

문항 수 40개, 상관이 0.4일 때 속달여부 분류 정확도

| r2i2 | _01 | | | _02 | | | _03 | | | _04 | | | _05 | | | _06 | | | _07 | | | _08 | | | _09 | | | _10 | | |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| | 0 | 1120 | 403 | 0 | 1095 | 404 | 0 | 1115 | 371 | 0 | 1104 | 350 | 0 | 1103 | 426 | 0 | 1092 | 376 | 0 | 1139 | 357 | 0 | 1089 | 401 | 0 | 1107 | 406 | 0 | 1125 | 378 |
| | 1 | 346 | 1131 | 1 | 394 | 1107 | 1 | 385 | 1129 | 1 | 376 | 1170 | 1 | 406 | 1065 | 1 | 349 | 1183 | 1 | 420 | 1084 | 1 | 361 | 1149 | 1 | 358 | 1129 | 1 | 410 | 1087 |
| 정확도 (비율) | 2251(75.0%) | | | 2202(73.4%) | | | 2244(74.8%) | | | 2274(75.8%) | | | 2168(72.3%) | | | 2275(75.8%) | | | 2223(74.1%) | | | 2238(74.6%) | | | 2236(74.5%) | | | 2212(73.7%) | | |

문항 수 40개, 상관이 0.8일 때 속달여부 분류 정확도

| r3i2 | _01 | | | _02 | | | _03 | | | _04 | | | _05 | | | _06 | | | _07 | | | _08 | | | _09 | | | _10 | | |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| | 0 | 1114 | 409 | 0 | 1091 | 408 | 0 | 1110 | 376 | 0 | 1099 | 355 | 0 | 1094 | 435 | 0 | 1090 | 378 | 0 | 1134 | 362 | 0 | 1088 | 402 | 0 | 1103 | 410 | 0 | 1119 | 384 |
| | 1 | 340 | 1137 | 1 | 391 | 1110 | 1 | 384 | 1130 | 1 | 373 | 1173 | 1 | 405 | 1066 | 1 | 348 | 1184 | 1 | 418 | 1086 | 1 | 359 | 1151 | 1 | 359 | 1128 | 1 | 407 | 1090 |
| 정확도 (비율) | 2251(75.0%) | | | 2201(73.4%) | | | 2240(74.7%) | | | 2272(75.7%) | | | 2160(72.0%) | | | 2274(75.8%) | | | 2220(74.0%) | | | 2239(74.6%) | | | 2231(74.4%) | | | 2209(73.6%) | | |

[부록 12] 모의실험 조건에 따른 삼각함수(A2) 속달여부 분류정확도(3)

A2. 속달여부 분류 정확도(3)

문항 수 60개, 상관이 0일 때 속달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r1i3 | 0 | 1094 | 391 | 0 | 1143 | 367 | 0 | 1148 | 376 | 0 | 1069 | 370 | 0 | 1092 | 335 | 0 | 1144 | 351 | 0 | 1111 | 376 | 0 | 1073 | 367 | 0 | 1166 | 387 | 0 | 1166 | 387 |
| | 1 | 321 | 1194 | 1 | 389 | 1101 | 1 | 317 | 1159 | 1 | 400 | 1161 | 1 | 398 | 1175 | 1 | 363 | 1142 | 1 | 401 | 1112 | 1 | 394 | 1166 | 1 | 356 | 1091 | 1 | 356 | 1091 |
| 정확도 (비율) | 2288(76.3%) | | | 2244(74.8%) | | | 2307(76.9%) | | | 2230(74.3%) | | | 2267(75.6%) | | | 2286(76.2%) | | | 2223(74.1%) | | | 2239(74.6%) | | | 2257(75.2%) | | | 2257(75.2%) | | |

문항 수 60개, 상관이 0.4일 때 속달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r2i3 | 0 | 1094 | 391 | 0 | 1141 | 369 | 0 | 1148 | 376 | 0 | 1067 | 372 | 0 | 1092 | 335 | 0 | 1143 | 352 | 0 | 1110 | 377 | 0 | 1071 | 369 | 0 | 1163 | 390 | 0 | 1122 | 403 |
| | 1 | 320 | 1195 | 1 | 388 | 1102 | 1 | 316 | 1160 | 1 | 399 | 1162 | 1 | 399 | 1174 | 1 | 364 | 1141 | 1 | 398 | 1115 | 1 | 390 | 1170 | 1 | 356 | 1091 | 1 | 348 | 1127 |
| 정확도 (비율) | 2289(76.3%) | | | 2243(74.8%) | | | 2308(76.9%) | | | 2229(74.3%) | | | 2266(75.5%) | | | 2284(76.1%) | | | 2225(74.2%) | | | 2241(74.7%) | | | 2254(75.1%) | | | 2249(75.0%) | | |

문항 수 60개, 상관이 0.8일 때 속달여부 분류 정확도

| | _01 | 0 | 1 | _02 | 0 | 1 | _03 | 0 | 1 | _04 | 0 | 1 | _05 | 0 | 1 | _06 | 0 | 1 | _07 | 0 | 1 | _08 | 0 | 1 | _09 | 0 | 1 | _10 | 0 | 1 |
|-------------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|-------------|------|------|
| r3i3 | 0 | 1092 | 393 | 0 | 1141 | 369 | 0 | 1186 | 376 | 0 | 1067 | 372 | 0 | 1090 | 337 | 0 | 1143 | 352 | 0 | 1110 | 377 | 0 | 1068 | 372 | 0 | 1163 | 390 | 0 | 1118 | 407 |
| | 1 | 319 | 1196 | 1 | 389 | 1101 | 1 | 314 | 1162 | 1 | 398 | 1163 | 1 | 398 | 1175 | 1 | 361 | 1144 | 1 | 398 | 1115 | 1 | 391 | 1169 | 1 | 355 | 1092 | 1 | 347 | 1128 |
| 정확도 (비율) | 2288(76.3%) | | | 2242(74.7%) | | | 2348(78.3%) | | | 2230(74.3%) | | | 2265(75.5%) | | | 2287(76.2%) | | | 2225(74.2%) | | | 2237(74.6%) | | | 2255(75.2%) | | | 2246(74.9%) | | |