교육학박사학위논문

인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단평가 모형 연구

2019년 2월

서울대학교 대학원

수학교육과

한 수 연

인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단평가 모형 연구

지도교수 유 연 주

이 논문을 교육학박사학위논문으로 제출함
2018년 10월

서울대학교 대학원
수학교육과
한 수 연

한수연의 박사학위논문을 인준함
2019년 1월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

국문초록

인지진단평가는 특정 영역의 학습에 필요한 지식이나 기능, 능력 등을 나타내는 인지요소들에 대한 학생들의 숙달 여부를 파악하는 것을 목적으로 하는 평가이다. 인지진단평가는 학생들의 인지적 강점과 약점에 대한 구체적인 진단적 피드백을 통해 교수-학습에 유용한 정보를 제공할 수 있다는 점에서 많은 관심을 받아왔다. 인지진단평가의 실행을 위한 통계적 모형을 인지진단모형이라고 한다. 지금까지 인지요소들 사이의 관계 또는 인지요소와 문항 응답 사이의 관계에 대한 서로 다른 가정에 따라 다양한 인지진단모형들이 제안되어 왔으며, 그 중 일부 모형들은 인지요소 사이의 위계적 관계를 의미하는 '인지요소 위계'를 고려한다. 인지요소 위계를 고려한 인지진단모형은 수학의 계통성을 반영한 인지진단평가에 적용될 가능성을 지닌다.

인지요소 위계는 흔히 그래프 구조의 형태로 표현되므로, 그래프 기반의 확률 모형인 베이지안 네트워크는 인지요소 위계를 고려한 인지진단모형으로 활용될 수 있다. 그러나 베이지안 네트워크를 인지진단모형으로 고려했던 선행 연구들은 엄밀한 의미에서 인지요소 위계의 모델링에는 초점을 두지 않았다. 따라서 본 연구에서는 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단모형인 BN-H 모형을 정의하고, BN-H 모형을 중심으로 인지요소 위계를 고려한 인지진단평가의 방법론을 고안하고자 하였다. 특히, 인지요소 위계를 고려한 인지진단평가에서 다루어야 하는 중요한 문제인 위계 추정의 방법과 인지요소 위계를 고려한 인지진단 컴퓨터 적응검사의 방법을 제안하고자 하였다.

본 연구에서는 BN-H 모형을 '인지요소 위계를 베이지안 네트워크의 그래프 구조로 나타내고, 인지요소 위계에 따라 일반적인 베이지안 네트워크 모형의 특정 모수를 0으로 제약한 모형'으로 정의하였다. 그리고 BN-H 모형은 인지요소 위계를 고려한 기존의 모형들에 비해 구조 모형은 더 간명한 반면 측정 모형은 더 일반적이라는 특징이 있음을 확인하였다.

BN-H 모형을 기반으로 한 위계 추정의 방법으로는 베이지안 네트워크의 구조 학습 알고리즘에 기반한 기존의 위계 추정 방법을 개선한 SEM-H 방법을 제안하였다. 첫 번째 모의실험에서는 SEM-H 방법이 기존 방법보다 향상된 위계 추정 결과를 도출함을 확인하였으며, 두 번째 모의실험에서는 다양한 조건 하에서 SEM-H 방법의 위계 추정 성능을 평가하였다. 그리고 두 가지 실제 수학 검사 자료에 SEM-H 방법을 적용하여 위계 추정을 실시한 결과로부터 SEM-H 방법의 실제적인 활용성을 확인하였다.

다음으로 BN-H 모형을 기반으로 한 인지진단 컴퓨터 적응검사의 방법인 BN-H-CAT을 제안하고 그 절차를 제시하였다. BN-H-CAT은 인지요소 위계를 고려한다는 점과 베이지안 네트워크 기반의 모형을 인지진단모형으로 사용한다는 점에서 기존의 인지진단 컴퓨터 적응검사와 구별되는 특징을 지닌다. 모의실험에서 BN-H-CAT과 인지요소 위계를 고려하지 않은 기존의 인지진단 컴퓨터 적응검사를 비교한 결과, 실제로 인지요소 위계가 존재하는 상황에서 BN-H-CAT은 BN-H 모형을 이용하여 인지요소 위계를 모델링함으로써 인지진단 컴퓨터 적응검사의 효율성을 향상시킬 수 있음을 확인하였다.

마지막으로 본 연구의 시사점을 논의하고, 본 연구에서 제안한 인지진단평가 방법론을 바탕으로 하는 후속 연구 방향을 제시하였다.

주요어: 인지진단평가, 인지진단모형, 인지요소 위계,
베이지안 네트워크, 위계 추정, 인지진단 컴퓨터 적응검사

학 번: 2011-21567

목 차

국문초록 ····· i
목차 ····· iii
표 목차 ····· vi
그림 목차 ····· viii

Ⅰ. 서론 ····· 1
1. 연구의 필요성 및 목적 ····· 1
2. 연구문제 ····· 5
3. 영어 약어 색인 ····· 6

Ⅱ. 이론적 배경 ····· 7
1. 인지진단이론 ····· 7
1.1. Q-행렬 ····· 7
1.2. 인지진단모형 ····· 9
1.3. 인지요소 위계를 고려한 인지진단모형 ····· 20
1.4. 인지요소패턴의 추정 ····· 26
2. 베이지안 네트워크 ····· 28
2.1. 베이지안 네트워크의 정의 ····· 28
2.2. 베이지안 네트워크의 추론 ····· 34
2.3. 베이지안 네트워크의 학습 ····· 38
3. 컴퓨터 적응검사 ····· 46

Ⅲ. BN-H: 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단모형 ····· 50
1. 베이지안 네트워크 기반 인지진단모형 ····· 50
1.1. BN 모형 ····· 51
1.2. BN 모형에서 인지요소 사이의 관계의 의미 ····· 56
2. BN-H 모형 ····· 58
2.1. BN-H 모형 ····· 58
2.2. BN-H 모형과 기존 모형들의 관계 ····· 63
2.3. BN-H 모형을 적용한 인지진단평가 ····· 65
3. 요약 및 논의 ····· 67

Ⅳ. SEM-H: BN-H 모형을 적용한 위계 추정 방법 ····· 69
1. SEM 알고리즘과 위계 추정 ····· 70
1.1. SEM 알고리즘 ····· 70
1.2. COMMAND 알고리즘 ····· 73
2. SEM-H ····· 79
3. 모의실험을 통한 SEM-H의 검증 ····· 84
3.1. 모의실험 1 ····· 84
3.2. 모의실험 2 ····· 101
4. 수학 검사 자료에 대한 SEM-H의 적용 ····· 118
4.1. 분수 뺄셈 검사 ····· 118
4.2. 일차방정식과 일차부등식 검사 ····· 135
5. 요약 및 논의 ····· 142

Ⅴ. BN-H-CAT: BN-H 모형을 적용한 인지진단 컴퓨터 적응검사 ····· 145
1. 인지진단 컴퓨터 적응검사 ····· 146
1.1. 문항 선택 방법 ····· 146

1.2. 검사 종료 규칙 ····· 153
2. BN-H-CAT ····· 154
2.1. BN-H-CAT의 설계 ····· 154
2.2. BN-H-CAT의 실행 ····· 159
3. 모의실험 ····· 163
3.1. 모의실험 설계 ····· 163
3.2. 모의실험 결과 ····· 170
4. 요약 및 논의 ····· 184

Ⅵ. 요약 및 결론 ····· 186
1. 요약 ····· 186
2. 결론 ····· 188

참고문헌 ····· 193
부록 ····· 205
Abstract ····· 209

표 목 차

<표 Ⅰ-1> 영어 약어 색인 ····· 6
<표 Ⅱ-1> Q-행렬의 예 ····· 8
<표 Ⅱ-2> 예시 검사에 대한 일반적인 잠재집단모형의 모수 ····· 12
<표 Ⅱ-3> 예시 검사의 Q-행렬 ····· 16
<표 Ⅱ-4> 예시 검사에 대한 DINA 모형의 모수 ····· 17
<표 Ⅱ-5> 예시 검사에 대한 DINO 모형의 모수 ····· 17
<표 Ⅱ-6> 예시 검사에 대한 LCDM의 모수 ····· 19
<표 Ⅱ-7> 예시 검사에 대한 DINA-H 모형의 모수 ····· 23
<표 Ⅱ-8> 예시 검사에 대한 DINO-H 모형의 모수 ····· 23
<표 Ⅱ-9> 예시 검사에 대한 HCDM의 모수 ····· 25
<표 Ⅲ-1> 예시 검사의 Q-행렬 ····· 52
<표 Ⅲ-2> 예시 검사에 대한 BN-H 모형 하에서의 $P(A_1, A_2, A_3)$ ····· 62
<표 Ⅲ-3> 예시 검사에 대한 BN-H 모형의 모수 ····· 63
<표 Ⅳ-1> 불완전 자료에서의 BN 구조 학습을 위한 SEM 알고리즘(Friedman, 1998; Koller & Friedman, 2009) ····· 71
<표 Ⅳ-2> 위계 추정을 위한 COMMAND 알고리즘(Chen et al., 2016) ····· 74
<표 Ⅳ-3> 본 연구에서 제안하는 위계 추정을 위한 SEM-H 알고리즘 ····· 80
<표 Ⅳ-4> 모의실험 1의 요인 ····· 90
<표 Ⅳ-5> 모의실험 1의 위계 추정 결과 ····· 94
<표 Ⅳ-6> 인지요소 위계 구조별 각 표본 크기에 대한 위계 추정 결과 평균 95
<표 Ⅳ-7> 인지요소 위계 구조별 각 Q-행렬에 대한 위계 추정 결과 평균 ·· 97
<표 Ⅳ-8> 인지요소 위계 구조별 각 문항 질에 대한 위계 추정 결과 평균 ·· 99
<표 Ⅳ-9> 인지요소 3개인 위계 구조에 대한 위계 추정 결과 ····· 108
<표 Ⅳ-10> 인지요소 4개인 위계 구조에 대한 위계 추정 결과 ····· 112
<표 Ⅳ-11> 인지요소 5개인 위계 구조에 대한 위계 추정 결과 ····· 116
<표 Ⅳ-12> 대분수 뺄셈 검사의 문항과 Q-행렬(K. K. Tatsuoka, 1990) ···· 121
<표 Ⅳ-13> 본 연구에서 활용한 분수 뺄셈 검사의 인지요소, 문항 및 Q-행렬 ····· 126
<표 Ⅳ-14> 일차방정식과 일차부등식 검사의 Q-행렬(윤지영, 2015) ····· 136

<표 Ⅳ-15> 본 연구에서 활용한 일차방정식과 일차부등식 검사의 인지요소와 Q-행렬 ····· 138
<표 Ⅳ-16> 일차방정식과 일차부등식 검사의 모수에 대한 위계 관계 판정 결과 ····· 140
<표 Ⅴ-1> 고정-길이 검사의 ACCR 결과: 구조 1, 문제은행의 질이 높은 경우 ····· 172
<표 Ⅴ-2> 고정-길이 검사의 ACCR 결과: 구조 1, 문제은행의 질이 낮은 경우 ····· 173
<표 Ⅴ-3> 고정-길이 검사의 ACCR 결과: 구조 2, 문제은행의 질이 높은 경우 ····· 174
<표 Ⅴ-4> 고정-길이 검사의 ACCR 결과: 구조 2, 문제은행의 질이 낮은 경우 ····· 175
<표 Ⅴ-5> 고정-길이 검사의 PCCR 결과 ····· 176
<표 Ⅴ-6> 변동-길이 검사의 검사 길이에 대한 기술통계량 및 PCCR 결과: 구조 1의 경우 ····· 180
<표 Ⅴ-7> 변동-길이 검사의 검사 길이에 대한 기술통계량 및 PCCR 결과: 구조 2의 경우 ····· 181

그 림 목 차

[그림 Ⅱ-1] 세 인지요소가 이루는 위계 구조의 예(a) 및 가능한 인지요소패턴과 불가능한 인지요소패턴(b) ····· 21
[그림 Ⅱ-2] 유향 비순환 그래프(DAG)의 예 ····· 29
[그림 Ⅱ-3] 베이지안 네트워크의 DAG와 CPT ····· 32
[그림 Ⅱ-4] 4개의 변수로 이루어진 베이지안 네트워크의 예 ····· 36
[그림 Ⅱ-5] 4개의 변수로 이루어진 베이지안 네트워크의 모수 표현 ····· 40
[그림 Ⅱ-6] v-structure와 skeleton의 예 ····· 44
[그림 Ⅱ-7] 동형인 베이지안 네트워크 구조의 예 ····· 45
[그림 Ⅲ-1] 예시 검사에 대한 BN 모형(인지요소들이 서로 독립인 경우) ····· 54
[그림 Ⅲ-2] 예시 검사에 대한 BN 모형(인지요소 사이의 관계가 가정된 경우) ····· 55
[그림 Ⅲ-3] 인지요소 사이의 관계들의 포함 관계 ····· 57
[그림 Ⅲ-4] 예시 검사에 대한 BN-H 모형 ····· 61
[그림 Ⅲ-5] BN-H 모형을 적용한 인지진단평가의 결과표의 예 ····· 66
[그림 Ⅳ-1] 동형인 BN 구조들로 이루어진 동형 클래스의 예 ····· 76
[그림 Ⅳ-2] 모의실험 1에서 가정한 인지요소 위계 구조 ····· 85
[그림 Ⅳ-3] 모의실험 1에서 가정한 단일 Q-행렬 ····· 88
[그림 Ⅳ-4] 모의실험 1에서 가정한 복합 Q-행렬 ····· 89
[그림 Ⅳ-5] 표본 크기에 따른 위계 추정 결과의 그래프 ····· 96
[그림 Ⅳ-6] Q-행렬에 따른 위계 추정 결과의 그래프 ····· 98
[그림 Ⅳ-7] 문항의 질에 따른 위계 추정 결과의 그래프 ····· 100
[그림 Ⅳ-8] 모의실험 2에서 고려한 인지요소 위계 구조의 유형 ····· 102
[그림 Ⅳ-9] 모의실험 2에서 가정한 인지요소 위계 구조 ····· 103
[그림 Ⅳ-10] 모의실험 2에서 가정한 복합 Q-행렬 ····· 104
[그림 Ⅳ-11] 인지요소 3개인 위계 구조에 대한 위계 추정 결과의 그래프 ···· 109
[그림 Ⅳ-12] 인지요소 4개인 위계 구조에 대한 위계 추정 결과의 그래프 ···· 113
[그림 Ⅳ-13] 인지요소 5개인 위계 구조에 대한 위계 추정 결과의 그래프 ···· 117
[그림 Ⅳ-14] 대분수 뺄셈 검사의 방법 B에 대한 인지요소(K. K. Tatsuoka, 1990) ····· 120

[그림 Ⅳ-15] 분수 뺄셈 검사 자료에 대한 SEM-H의 구조 학습 결과 ····· 128
[그림 Ⅳ-16] $A_1 \rightarrow A_4$와 $A_3 \rightarrow A_4$의 위계 관계 판정을 위한 모수 추정 결과 ·· 132
[그림 Ⅳ-17] 분수 뺄셈 검사에 대해 SEM-H 방법을 적용하여 최종적으로 추정된 인지요소 구조 ····· 134
[그림 Ⅳ-18] 일차방정식과 일차부등식 검사의 인지요소와 위계 구조(윤지영, 2015) ····· 135
[그림 Ⅳ-19] 일차방정식과 일차부등식 검사에 대한 SEM-H 적용 결과 ····· 139
[그림 Ⅴ-1] BN-H 모형의 구조모형 ····· 157
[그림 Ⅴ-2] BN-H-CAT에서 BN-H 모형의 갱신 ····· 161
[그림 Ⅴ-3] 모의실험에서 가정한 인지요소 위계 구조 ····· 164
[그림 Ⅴ-4] 모의피험자들의 진 인지요소패턴 분포 ····· 165
[그림 Ⅴ-5] 고정-길이 검사 실험에서 MPWKL 방법의 PCCR 결과 ····· 177
[그림 Ⅴ-6] 변동-길이 검사 실험에서 MPWKL 방법의 평균 검사 길이 ····· 182

Ⅰ. 서론

1. 연구의 필요성 및 목적

‘학습을 위한 평가’(assessment for learning; Black & Wiliam, 1998)의 관점에서 평가는 교수 학습을 돕기 위해 자료를 수집하고 해석하여 활용하는 일련의 과정으로, 교사와 학생 모두에게 유용한 정보를 제공하는 것을 목적으로 하는 활동이다(NCTM, 2000; McMillan, 2008). 이 목적을 달성하기 위해 학습에 필요한 특정 지식(knowledge)이나 기능(skill), 능력(ability) 등에 대한 학생의 숙달 여부를 파악하여 학습 상태에 대한 구체적인 진단 정보를 제공할 수 있는 인지진단이론(cognitive diagnosis theory)을 적용한 평가가 최근 들어 주목받고 있다(DiBello, Stout, & Roussos, 1995).

인지진단이론에서 진단적 정보의 제공은 인지진단모형(cognitive diagnosis model) 또는 진단분류모형(diagnostic classification model)이라고 하는 통계적 모형의 발전에 의해 가능하게 되었다(박세진, 이현숙, 2017; 김성은, 박윤수, 이영선, 2012; Rupp, Templin, & Henson, 2010). 인지진단모형은 검사가 측정하고자 하는 지식, 기능 등을 인지요소(attribute)로 설정하고, 이들을 측정하는 문항에 대한 피험자의 응답으로부터 인지요소에 대한 숙달 여부를 추정하여 피험자의 인지 상태를 파악함으로써 학생의 학습 능력이나 수준에 대한 보다 구체적인 정보를 제공할 수 있게 해준다. 인지진단모형은 피험자가 문항에 응답하는 방식의 기저에 대한 가정에 따라 다르게 정의될 수 있으며, 이에 따라 그 동안 다양한 인지진단모형들이 발전되어 왔다(예를 들어, K. K. Tatsuoka, 1990; Junker & Sijtsma, 2001; Hartz, 2002; Templin & Henson, 2006).

인지진단모형에 관한 일련의 연구에서 인지요소들 사이의 구조적 관계들을 모형에 포함시키려는 시도가 이루어져 왔는데, 그중에서도 특히 인지요소들이 위계적 관계(hierarchical relationship)를 이룰 수 있음을

가정하고 이를 모델링하는 모형들이 발전되었다(Leighton, Gierl, & Hunka, 2004; Su, 2013; Templin & Bradshaw, 2014). 여기서 위계적 관계란 특정 인지요소의 숙달을 위해 그에 앞서는 선행 인지요소의 숙달이 필수적인 경우를 뜻하는 것으로, Levy와 Mislevy(2016)는 이와 같은 위계 관계를 강한 선행 관계(hard prerequisite relationship)라고 표현하였으며, Leighton 외(2004)는 한 검사의 인지요소들이 위계적 관계를 이루는 경우 인지요소 위계(attribute hierarchy)가 존재한다고 표현하였다. 인지요소 위계를 가정한 인지진단모형 연구들에서는 교수-학습의 과정은 어떤 순서로 학습이 이루어져야 하는지에 관한 이론들이 반영된 위계적 구조를 따르는 교육과정을 기반으로 이루어진다는 점을 들어 인지요소들 간에 위계적 관계가 존재할 수 있음을 가정한다. 이 경우 인지요소 위계는 학습의 위계와 밀접한 관련이 되는 개념으로도 이해될 수 있다.

학습 위계(learning hierarchy) 또는 학습 순서(learning sequence)를 밝히는 일은 과거부터 연구자들의 오랜 관심사였다. 예를 들어, “Thorndike는 현재 배우는 것이 이후 학습과 어떤 연관이 되어있고, 또 이후 학습을 위해 무엇을 배워야 하는지 알고 있어야 학습의 촉진을 돕는 교수 실행이 이루어질 수 있음을 주장하였다”(Baroody, Cibulskis, Lai, & Li, 2004, p. 231). Gágne 외(2005)는 학습 과제 분석(learning-task analysis)의 맥락에서 과제의 해결을 위한 지적 기능들(intellectual skills)이 위계적일 수 있으며 어떤 과제의 해결을 위해 필수적인 일련의 기능들을 연속적으로 수행하여 과제가 해결될 수 있을 때 그 과제의 해결을 위한 학습 위계가 있다고 표현하였다. Gágne 외(2005)는 특히 학습 위계를 설명하기 위해 수학 문제의 상황인 자연수의 뺄셈에 관한 일련의 과제를 예로 들면서 “받아내림이 없는 뺄셈”, “여러 번의 받아내림이 필요한 경우의 뺄셈”, “연속된 받아내림이 필요한 뺄셈”, “빼지는 수(피감수)에 0이 존재하는 경우의 뺄셈”과 같은 일련의 기능들의 숙달은 선행 기능에 대한 학습 이전에는 이루어질 수 없음을 밝혔다.

실제로 수학은 어떤 내용의 학습을 위해 특정 아이디어나 기능들이 먼저 학습되어야 하는 경우가 다수 존재하는 매우 연속적이고 위계적인

분야에 속한다(Clements & Sarama, 2009, p. 5). 따라서 수학의 특정 내용과 관련한 지식, 기능 등의 숙달 여부를 진단하고자 하는 수학 인지진단평가에서 인지요소 위계를 가정한 인지진단모형의 적용을 고려해볼 수 있다. 예를 들어, Gágne 외(2005)의 자연수의 뺄셈 과제의 경우에 대해서도 각각의 기능들을 인지요소로 설정하면 그들 사이의 위계를 모형에 반영한 인지진단모형을 적용함으로써 과제를 수행한 학습자들에 대한 진단이 이루어질 수 있다. 그러나 지금까지 이루어진 수학 인지진단평가 관련 선행 연구들은 대체로 인지요소들의 관계를 고려하지 않고 개별 인지요소에 대한 전체 피험자 집단의 숙달 양상을 확인하는 것에 주로 초점을 두고 있으며(예를 들어, 김선희, 김수진, 송미영, 2008; 김희경, 김부미, 2011; 이현숙, 고호경, 2014; 박지현, 김수진, 2015; Lee, Park, & Taylan, 2011; Choi, Lee, & Park, 2015), 수학 인지진단평가의 맥락에서 인지요소 위계를 반영한 인지진단모형을 고려한 연구는 찾아보기 어렵다.

인지요소 위계는 흔히 그래프 구조로 표현되므로, 인지요소 위계를 모델링하는 하나의 모형으로 베이지안 네트워크(Bayesian Network; Pearl, 1988)라고 하는 그래프 기반의 통계적 모형을 적용하는 것을 고려할 수 있다. 베이지안 네트워크는 복잡한 확률분포를 그래프 구조를 통해 간결하게 표현할 수 있다는 장점을 지닌 모형으로, 실제로 몇몇 연구들은 베이지안 네트워크를 인지진단모형으로 활용하는 방법을 제안하였다(Mislevy, 1995; Culbertson, 2016; Levy & Mislevy, 2016). 그러나 베이지안 네트워크를 인지요소 위계의 모델링에 활용하는 방안에 관한 연구는 많지 않다. 이에 베이지안 네트워크를 이용하여 위계를 고려한 인지진단모형을 어떻게 정의할 수 있는지를 밝히는 것이 본 연구의 첫 번째 목적이다.

인지요소 위계를 고려한 인지진단모형을 활용한 평가에서 다루어야 하는 중요한 문제는 인지요소 위계의 타당성 여부이다. 인지요소 위계를 고려한 인지진단모형에 대한 선행 연구들은 인지요소 위계가 참이라는 가정 하에 모형을 전개한다. 그러나 실제로 우리는 참인 인지요소 위계

를 알 수 없으므로 위계를 잘못 설정하는 경우가 생길 수 있으며, 인지요소 위계가 오지정(misspecification)되는 경우 진단의 정확도가 떨어질 수 있다는 문제가 발생한다(Templin, Henson, Templin, & Roussos, 2008; Liu, 2018). 따라서, 인지요소 위계를 고려한 인지진단모형을 활용할 때는 설정된 인지요소 위계가 타당한지를 검증할 필요가 있다. 이러한 필요성에 따라, 본 연구에서는 인지요소 위계의 타당화를 위해 실제 응답 자료를 이용하여 위계를 확인하는 과정을 '위계 추정'으로 정의하고 베이지안 네트워크를 이용한 위계 추정 방법을 발전시키는 것을 두 번째 목적으로 설정하였다.

한편, 최근 인지진단모형에 대한 관심이 높아짐에 따라 컴퓨터 적응검사(computerized adaptive testing)의 실행을 위한 측정 모형으로 인지진단모형을 적용한 인지진단 컴퓨터 적응검사(cognitive diagnostic computerized adaptive testing)에 대한 연구 또한 활발히 이루어지기 시작하였다(C. Tatsuoka, 2002; C. Tatsuoka & Ferguson, 2003; Cheng, 2009; Huebner, 2010; Wang, 2013). 컴퓨터 적응검사는 컴퓨터를 기반으로 각 피험자에게 자신의 능력 수준에 맞는 문항을 선별적으로 선택하여 제공하는 개별 맞춤형 검사 기법으로, 짧은 시간 안에 적은 수의 문항으로 측정하고자 하는 여러 특성들을 정확하고 효율적으로 측정하는 것을 목적으로 한다(이혜옥, 2005). 컴퓨터 적응검사에 베이지안 네트워크를 적용하려는 시도는 몇 차례 있었으나(Almond & Mislevy, 1999; Millán, Pérez-de-la-Cruz, & Suárez, 2000), 이들 연구는 인지진단모형이 아닌 IRT 모형을 가정하거나 인지진단모형의 인지요소 위계를 가정하지 않았다. 실제로 인지요소들 사이에 위계적 관계가 존재하는 경우 그 관계들이 제공하는 추가적인 정보로 인해 추정의 효율성이 높아질 수 있을 것이라 예상되나 이와 관련한 연구는 아직 이루어지지 않았다. 이에 본 연구에서는 위계를 고려한 베이지안 네트워크 기반 인지진단모형을 적용한 인지진단 컴퓨터 적응검사의 방법을 정의하고, 그 방법이 위계 관계를 고려함으로 인해 인지진단 컴퓨터 적응검사의 효율성에 어느 정도의 영향을 미치는지 확인해보고자 하였다.

2. 연구문제

본 연구는 베이지안 네트워크를 기반으로 인지요소 위계를 고려한 인지진단평가에 대한 방법론을 구체화하는 것을 목적으로 하며, 이를 위해 다음과 같은 연구문제를 설정하고 이에 답하고자 한다.

연구문제 1. 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단 모형(이하 BN-H 모형)은 무엇이며 BN-H 모형을 적용한 인지진단평가는 어떻게 이루어지는가?

연구문제 2. BN-H 모형을 적용한 위계 추정 방법은 무엇이며 어떻게 이루어지는가?

연구문제 3. BN-H 모형을 적용한 인지진단 컴퓨터 적응검사는 어떻게 이루어지며 인지진단 컴퓨터 적응검사의 효율성에 어떠한 영향을 미치는가?

Ⅱ장에서는 연구문제의 해결을 위한 초석을 다지는 것을 목표로, 본 연구의 핵심 주제인 인지진단이론, 베이지안 네트워크, 컴퓨터 적응검사의 기본 개념 및 원리를 검토한다. Ⅲ장에서는 BN-H 모형의 정의와 특징을 밝히고 BN-H 모형을 적용한 인지진단평가의 과정을 제시함으로써 연구문제 1에 답하고자 한다. Ⅳ장에서는 인지요소 위계를 고려한 인지진단평가에서 반드시 다루어야 하는 위계 추정의 문제를 BN-H 모형을 적용하여 해결하고자 하였으며, 이를 위해 구체적인 위계 추정 방법인 SEM-H를 제안하고 모의실험을 통해 SEM-H 방법의 성능을 검증한 결과와 실제 수학 검사 자료에 SEM-H 방법을 적용한 결과를 분석함으로써 연구문제 2에 답하고자 한다. Ⅴ장에서는 인지진단평가의 한 적용분야인 인지진단 컴퓨터 적응검사에 BN-H 모형을 적용한 BN-H-CAT을 제안하고, 모의실험을 통한 기존의 인지진단 컴퓨터 적응검사와의 비교·분석을 통해 연구문제 3에 답하고자 한다. 마지막으로 Ⅵ장에서는 이상의 논의를 요약하며 본 연구를 통해 얻은 시사점과 본 연구의 제한점을 밝히고, 후속 연구를 위한 제언을 하고자 한다.

3. 영어 약어 색인

본 연구의 전반에 걸쳐 사용되는 영어 약어는 <표 Ⅰ-1>과 같다.

<표 Ⅰ-1> 영어 약어 색인

약어(Abbreviation)	원어(Original name)
ACCR	Attribute correct classification rate
AHM	Attribute hierarchy method
BIC	Bayesian information criterion (score)
BME	Bayes modal estimation
BN	Bayesian Network; 베이지안 네트워크
BN-H	BN with hierarchy (model)
BN-H-CAT	Computerized adaptive testing with BN-H model
CAT	Computerized adaptive testing; 컴퓨터 적응검사
CDM	Cognitive diagnosis model; 인지진단모형
CD-CAT	Cognitive diagnostic computerized adaptive testing; 인지진단 컴퓨터 적응검사
CPT	Conditional probability table (of BN)
DAG	Directed acyclic graph (of BN)
DCM	Diagnostic classification model; 진단분류모형
DINA	Deterministic inputs, noisy “and” gate (model)
DINA-H	DINA with hierarchy (model)
DINO	Deterministic inputs, noisy “or“ gate (model)
DINO-H	DINO with hierarchy (model)
EAP	Expected a posteriori (estimation)
EM	Expectation-maximization (algorithm)
HDCM	Hierarchical diagnostic classification model
IRT	Item response theory; 문항반응이론
LCDM	Log-linear cognitive diagnosis model
MAP	Maximum a posteriori (estimation)
MCMC	Markov chain Monte Carlo (method)
MLE	Maximum likelihood estimation
MMLE	Marginal maximum likelihood estimation
PCCR	Pattern correct classification rate
r-TDER	Revised true discovery edge rate
SEM	Structural expectation-maximization (algorithm)
SEM-H	SEM for hierarchy (algorithm)
TDER	True discovery edge rate
TPER	True positive edge rate

Ⅱ. 이론적 배경

1. 인지진단이론

인지진단이론은 피험자가 학습에 필요한 어떤 지식이나 기능, 능력, 인지 과정(cognitive process) 등과 같은 잠재적 특성들을 얼마나 숙달하고 있는지를 진단하는 검사를 위한 측정이론으로, 각 잠재 특성을 측정하는 문항들로 이루어진 검사에 대한 응답 자료를 토대로 각 특성에 대한 피험자들의 숙달 여부를 추정하는 것을 목적으로 한다(성태제, 2018). 인지진단이론에서 검사 문항의 해결을 위해 요구되는 잠재적 특성들을 '인지요소(attribute)', 문항과 인지요소의 관계를 수학적으로 표현한 것을 Q-행렬(K. K. Tatsuoka, 1990)이라 하며, Q-행렬을 기반으로 여러 인지요소에 대한 피험자의 숙달 수준을 추정하기 위해 사용되는 통계적 모형을 인지진단모형 또는 진단분류모형이라고 한다. 인지진단모형의 적용을 통해 피험자의 인지 상태를 진단하고 학습에 대한 강점과 약점을 파악함으로써 구체적인 진단적 정보를 제공할 수 있도록 설계된 검사를 인지진단평가(cognitive diagnostic assessment)라고 한다(Leighton & Gierl, 2007; Wang & Chang, 2018). 이 절에서는 먼저 Q-행렬이 무엇인지 살펴보고, 대표적인 인지진단모형들과 인지요소 위계를 가정한 인지진단모형들을 검토한다. 마지막으로 인지진단모형을 토대로 피험자의 인지요소 숙달 수준을 추정하는 방법들을 살펴본다.

1.1. Q-행렬

Q-행렬은 각 문항이 측정하는 인지요소에 대한 정보를 수학적으로 표현한 것으로, J개의 문항 $I_1, I_2, \cdots, I_J$으로 이루어진 검사가 K개의 인지요소 $A_1, A_2, \cdots, A_K$를 측정할 때 그 검사의 Q-행렬은 $J \times K$ 행렬로 표현된다. Q-행렬의 각 원소 q_{jk}는 j번째 문항 I_j의 해결을 위해 k번째 인지

요소 A_k가 필요한지의 여부를 나타내며, 문항 I_j가 인지요소 A_k를 필요로 하는 경우 1, 그렇지 않은 경우 0의 값을 가진다.

$$\boldsymbol{Q}=(q_{jk}),\quad q_{jk}=\begin{cases}1, & \text{문항 } I_j\text{가 인지요소 } A_k\text{를 필요로 하는 경우}\\ 0, & \text{그렇지 않은 경우}\end{cases}$$

예를 들어, <표 Ⅱ-1>의 Q-행렬은 문항 I_1은 인지요소 A_1과 A_4를, 문항 I_2는 인지요소 A_1과 A_2를, 문항 I_3는 인지요소 A_3를 필요로 하는 문항임을 나타낸다.

<표 Ⅱ-1> Q-행렬의 예

	인지요소 A_1	인지요소 A_2	인지요소 A_3	인지요소 A_4
문항 I_1	1	0	0	1
문항 I_2	1	1	0	0
문항 I_3	0	0	1	0
…	…	…	…	…

인지진단이론은 검사의 각 문항에 대한 피험자의 응답 자료를 토대로 각 인지요소에 대한 숙달 여부를 진단하는 것을 목적으로 하므로, 인지진단모형의 적용에서는 인지요소와 Q-행렬을 올바르게 규명하는 것이 필수적이다. 실제로 각 문항의 해결에 필요한 인지요소들에 대한 정의가 제대로 이루어지지 않거나, 각 문항이 측정하는 인지요소가 Q-행렬에 올바르게 표현되지 않은 경우, 그와 같은 Q-행렬을 바탕으로 얻어진 통계적 결과들은 타당하지 않을 수 있다(de la Torre, 2008; Rupp & Templin, 2008a). 따라서 인지진단모형을 검사에 적용할 때에는 검사 개발 단계에서부터 인지진단모형을 고려하여 인지요소와 Q-행렬을 신중하게 명세화하여야 하며(Rupp et al., 2010), 검사 실시 이후에는 수집된 응답 자료를 토대로 통계적 방법을 적용하여 Q-행렬을 타당화하는 과정이 이루어져야 한다(김성은, 2015; de la Torre, 2008; Chiu, 2013).

1.2. 인지진단모형

인지진단모형은 일련의 인지요소에 대한 피험자의 숙달 여부를 추정하기 위해 인지진단평가에서 활용되는 통계적 모형으로, 인지진단모형에서 인지요소는 보통 범주형 잠재 변수(categorical latent variable)로, 문항 응답은 관찰 변수(observable variable)로 표현된다(Rupp et al., 2010). 인지진단모형에서 이 두 변수는 이분형(dichotomous) 또는 다분형(polytomous)의 형태가 모두 가능하나(Rupp & Templin, 2008b), 이 절에서는 이분 인지요소 변수와 이분 문항 변수를 가정한 모형들을 중심으로 살펴볼 것이다.

이분형 인지요소가 가정된 경우 피험자의 각 인지요소의 숙달 정도는 숙달/비숙달의 두 수준으로 구분되며, K개의 독립적인 이분 인지요소들이 가정된 검사에서 각 인지요소에 대한 숙달 여부를 차례로 나열한 인지요소패턴의 종류는 2^K개가 된다. 또한, 이와 같은 검사에 인지진단모형을 적용하여 피험자들의 각 인지요소에 대한 숙달 여부를 추정한다는 것은 결국 피험자들의 모든 인지요소에 대한 숙달 여부를 나열한 인지요소패턴을 추정하는 것과 동일하므로, 이 경우 인지진단모형은 피험자들을 인지요소패턴에 따라 2^K개의 잠재집단(latent class)으로 분류하는 모형이라고 할 수 있다. 이러한 의미에서 이분형 변수들을 가정한 인지진단모형은 제약 잠재집단모형(restricted latent class model; Haertel, 1989)의 일종으로 볼 수 있는데, 이는 일반적인(unrestricted) 잠재집단모형의 특정 모수들을 어떻게 재매개화(reparameterization) 하는지에 따라 서로 다른 인지진단모형이 도출될 수 있다는 의미이다(Rupp et al., 2010). 다시 말해 지금까지 제안되어온 여러 인지진단모형들은 일반적인 잠재집단모형의 특수한 모형들인 것으로 볼 수 있다. 따라서 잠재집단모형의 일반적 형태를 먼저 살펴봄으로써 인지진단모형에 대한 이해를 도울 수 있을 것으로 보인다.

1.2.1. 제약 잠재집단모형으로서의 인지진단모형

일반적인 잠재집단모형의 수학적 표현을 살펴보기 위해, K개의 인지요소에 대한 숙달 여부를 나열한 인지요소패턴을 다음과 같이 나타내자:

$$\boldsymbol{\alpha}_c = (\alpha_{c1}, \alpha_{c2}, \cdots, \alpha_{cK}), \quad c = 1,2,\cdots, C \tag{2.1}$$

식 (2.1)에서 c는 잠재집단을 나타내고 C는 가능한 잠재집단의 수(2^K)를 의미하며, $\alpha_{ck}\,(k = 1,2,\cdots, K)$는 잠재집단 c에 속한 피험자의 k번째 인지요소에 대한 숙달 여부(0/1)를 나타낸다. 그리고 어떤 피험자의 인지요소패턴이 $\boldsymbol{\alpha}_c$일 확률, 다시 말해 어떤 피험자가 잠재집단 c에 속할 확률인 집단소속 확률(class membership probability) $P(\boldsymbol{\alpha}_c)$을 나타내는 모수를 v_c라 하자. 이때 각 피험자들의 인지요소패턴은 $C(= 2^K)$개의 가능한 인지요소패턴 중 하나로 결정되므로

$$\sum_{c=1}^{C} v_c = 1 \tag{2.2}$$

이 성립한다. 따라서 일반적인 잠재집단모형에서 추정해야 하는 집단소속 확률 $P(\boldsymbol{\alpha}_c)$에 대한 모수는 $C-1$개이다(Rupp et al., 2010).

다음으로, J개의 검사 문항에 대해, 인지요소패턴 $\boldsymbol{\alpha}_c$를 갖는 피험자 i가 각 문항 j의 정답을 맞힐 확률을 아래와 같이 나타내자:

$$\pi_{jc} = P(X_{ij} = 1 | \boldsymbol{\alpha}_c) \tag{2.3}$$

식 (2.3)의 확률은 피험자 i뿐 아니라 인지요소패턴 $\boldsymbol{\alpha}_c$를 갖는 모든 각 피험자들이 문항 j의 정답을 맞힐 확률로 해석될 수 있으며, 다른 의미로는 인지요소패턴 $\boldsymbol{\alpha}_c$를 갖는 피험자들에 대한 문항 j의 난도(difficulty)라고 할 수 있다. 그러면 문항 간 지역독립성 가정이 지켜졌을 때, 인지요소패턴 $\boldsymbol{\alpha}_c$를 갖는 피험자 i의 이분 문항반응패턴 $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iJ})$에 대한 조건부 확률(conditional probability)은 다음과 같이 표현된다:

$$P(\boldsymbol{X}_i = \boldsymbol{x}_i | \boldsymbol{\alpha}_c) = \prod_{j=1}^{J} \pi_{jc}^{x_{ij}} (1 - \pi_{jc})^{1 - x_{ij}} \tag{2.4}$$

이는 인지요소패턴이 $\boldsymbol{\alpha}_c$인 피험자 i에게서 문항반응패턴 $\boldsymbol{x}_i$가 관찰된 것이었을 가능성의 정도를 나타내는 우도(likelihood)이다(Rupp et al., 2010).

이러한 기호적 표현들을 토대로 잠재집단모형에서 피험자 i의 이분 문항반응패턴 $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iJ})$에 대한 주변확률(marginal probability)은 다음과 같이 표현된다:

$$\begin{aligned} P(\boldsymbol{X}_i = \boldsymbol{x}_i) &= \sum_{c=1}^{C} P(\boldsymbol{\alpha}_c)\, P(\boldsymbol{X}_i = \boldsymbol{x}_i | \boldsymbol{\alpha}_c) \\ &= \underbrace{\sum_{c=1}^{C} v_c}_{\text{구조모형}} \underbrace{\prod_{j=1}^{J} \pi_{jc}^{x_{ij}} (1-\pi_{jc})^{1-x_{ij}}}_{\text{측정모형}} \end{aligned} \tag{2.5}$$

식 (2.5)에서 v_c는 피험자 i가 특정 인지요소패턴을 가질 확률을 나타내는 모수로, 이는 인지진단모형에서 인지요소 변수들 사이의 관계를 가정하는 경우에 그 관계를 표현하는 부분이 될 수 있으므로 v_c에 대한 모형을 **구조모형**이라고 할 수 있다. 또한, $\pi_{jc}^{x_{ij}}(1-\pi_{jc})^{1-x_{ij}}$는 인지요소패턴 $\boldsymbol{\alpha}_c$를 갖는 피험자로부터 문항 j에 대한 응답 x_{ij}가 관찰될 확률을 표현하며, 이는 관찰변수인 문항 변수와 잠재변수인 인지요소변수의 관계에 대한 가정을 직접적으로 나타내는 부분이므로 $\pi_{jc} = P(X_{ij} = 1 | \boldsymbol{\alpha}_c)$에 대한 모형을 **측정모형**이라고 할 수 있다(Rupp et al., 2010).[1)]

1) Rupp 외(2010)는 '구조모형'과 '측정모형'이라는 용어를 구조방정식모형(structural equation model) 연구들로부터 차용하였음을 밝히고 있다.

잠재집단모형의 모수에 대한 이해를 돕기 위해, 3개의 독립적인 인지요소를 측정하는 3개의 문항으로 이루어진 예시 검사에 대한 잠재집단모형의 모수를 <표 Ⅱ-2>에 나타내었다.

<표 Ⅱ-2> 예시 검사에 대한 일반적인 잠재집단모형의 모수

c	$\boldsymbol{\alpha}_c$	구조모수 (v_c)	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	π_{11}	π_{21}	π_{31}
2	(0,0,1)	v_2	π_{12}	π_{22}	π_{32}
3	(0,1,0)	v_3	π_{13}	π_{23}	π_{33}
4	(0,1,1)	v_4	π_{14}	π_{24}	π_{34}
5	(1,0,0)	v_5	π_{15}	π_{25}	π_{35}
6	(1,0,1)	v_6	π_{16}	π_{26}	π_{36}
7	(1,1,0)	v_7	π_{17}	π_{27}	π_{37}
8	(1,1,1)	v_8	π_{18}	π_{28}	π_{38}

잠재집단모형에서 3개의 독립적인 인지요소를 측정하는 검사에 대해 가능한 인지요소패턴의 수는 $2^3=8$이며, 집단소속 확률 $P(\boldsymbol{\alpha}_c)$를 나타내는 구조모수 v_c의 수는 $8-1=7$이다($\sum_{c=1}^{8} v_c = 1$이므로 7개의 모수가 추정되면 나머지 하나는 그에 따라 결정된다). 또한, 각 문항의 문항모수는 잠재집단의 수만큼 존재하므로 총 문항모수의 수는 $8\times 3=24$이다.

그동안 제안되어온 여러 인지진단모형들은 식 (2.5)의 구조모형 또는 측정모형에 어떠한 가정을 부여하여 재매개화하는지에 따라 서로 구분되는 모형들로 볼 수 있다. 예를 들어, 인지진단평가에서 자주 활용되어 온 DINA 모형이나 Fusion 모형(Hartz, 2002) 등은 잠재집단모형의 측정모형에 대한 제약이 부여된 모형으로, 이들은 인지요소패턴 $\boldsymbol{\alpha_c}$를 가진 피험자 i가 문항 j의 정답을 맞힐 확률을 나타내는 문항반응함수(item response function) $\pi_{jc} = P(X_{ij} = 1|\boldsymbol{\alpha}_c)$가 서로 다른 형태로 재매개화된 모형들이다. 반면 이후에 살펴보게 될 인지요소 위계를 고려한 모형들은 잠재집단모형의 구조모형에 대한 제약이 부여된 모형으로, 인지요소 간의 위계 관계라는 가정에 따라 v_c를 재매개화한 모형이라고 볼 수 있다.

측정모형에 대한 제약이 부여된 인지진단모형들은 인지요소 조합과 문항 응답의 관계에 대한 가정에 따라 크게 비보상적(non-compensatory) 모형과 보상적(compensatory) 모형으로 구분된다(Henson, Templin, & Willse, 2009; 김명연, 강태훈, 2016). 비보상적 모형은 피험자가 어떤 문항의 해결에 요구되는 모든 인지요소를 숙달하였을 경우에만 해당 문항에 대해 일정 수준 이상의 정답 확률을 가지는 것을 가정한 모형이다. 반면 보상적 모형은 피험자가 어떤 문항의 해결에 필요한 여러 인지요소 중 일부를 숙달하지 않았더라도 나머지 인지요소를 숙달하였다면 해당 문항에 대한 정답 확률이 높아질 수 있음 가정한다.

이 절의 나머지 부분에서는 측정모형에 대한 제약이 부여된 인지진단모형들에 초점을 두고, 인지진단모형 중 비보상적 모형과 보상적 모형을 각각 대표하는 DINA 모형과 DINO 모형을 차례로 살펴본 뒤, 비보상적 모형과 보상적 모형을 모두 표현할 수 있는 포괄적 모형에 해당하는 LCDM을 검토한다.

1.2.2. DINA 모형과 DINO 모형

DINA 모형(Haertel, 1989; Junker & Sijtsma, 2001)은 어떤 피험자가 문항이 요구하는 모든 인지요소를 숙달하였을 때만 이들의 결합에 의해 그 문항의 정답을 맞힐 수 있다는 결합적 규칙을 따르는 비보상적 모형이다.

DINA 모형은 크게 세 가지 핵심 요소로 이루어져 있다(Rupp et al., 2010, pp. 117-118). 첫 번째 요소는 결정론적 투입(deterministic inputs)에 해당하는 것으로, 피험자 i가 문항 j의 정답을 맞힐 수 있을 것이라 기대되는지의 여부를 나타내는 잠재변수 η_{ij}이다. 두 번째 요소는 η_{ij}를 결정하는 결합적 방식을 나타내는 "and" gate로, 이에 따라 DINA 모형에서는 잠재변수 η_{ij}가 다음과 같이 표현된다.

$$\eta_{ij} = \prod_{k=1}^{K} (\alpha_{ik})^{q_{jk}} \tag{2.6}$$

식 (2.6)에서 k는 인지요소를 의미하며, α_{ik}는 피험자 i의 인지요소패턴 $\boldsymbol{\alpha}_i$의 k번째 원소로, 피험자 i가 인지요소 k를 숙달하였는지의 여부(1/0)를 나타낸다. q_{jk}는 Q-행렬의 j행 k열의 원소로 문항 j의 해결에 인지요소 k가 필요한지의 여부(1/0)를 나타낸다. η_{ij}는 $q_{jk}=1$인 모든 인지요소 k에 대해서 $\alpha_{ik}=1$이어야만 1의 값을 가질 수 있으므로, η_{ij}는 피험자 i가 문항 j가 필요로 하는 모든 인지요소를 숙달하였는지의 여부를 나타낸다.

DINA 모형의 결정론적 투입의 관점에서 문항에 대한 응답은 η_{ij}에 의해 결정되어야 한다. 즉 $\eta_{ij}=1$인 피험자들은 문항 j의 정답을 맞혀야 하며 $\eta_{ij}=0$인 피험자들은 문항 j의 정답을 맞히지 못해야 한다. 그러나 실제 문항 응답은 잡음(noise) 요인에 의해 기대되는 반응과는 다른 결과가 도출될 수 있다. DINA 모형은 이 가능성을 허용하여 실제 문항 응답이 확률적으로 결정됨을 가정하는데, 이 점이 DINA 모형의 세 번째 요소에 해당한다. 구체적으로, DINA 모형은 각 문항 j에 대해서 그 문항이 요구하는 모든 인지요소를 숙달하였으나($\eta_{ij}=1$) 정답을 맞히지 못하

는 경우의 확률을 나타내는 부주의(slip) 모수 s_j와 그 문항이 요구하는 모든 인지요소들을 숙달하지 못하였음에도($\eta_{ij}=0$) 정답을 맞히는 경우의 확률을 나타내는 추측(guess) 모수 g_j를 각각 설정한다. DINA 모형의 부주의 모수 s_j와 추측 모수 g_j는 각각 다음과 같다.

$$s_j = P(X_{ij}=0|\eta_{ij}=1) \tag{2.7}$$

$$g_j = P(X_{ij}=1|\eta_{ij}=0) \tag{2.8}$$

피험자 i의 문항 j에 대한 잠재변수 η_{ij}와 문항 j에 대한 부주의 모수, 추측 모수를 모두 고려하면, 피험자 i가 문항 j의 정답을 맞힐 확률은 식 (2.9)와 같이 표현되며, 이를 DINA 모형의 문항반응함수라 한다.

$$P(X_{ij}=1|\eta_{ij}) = (1-s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}} \tag{2.9}$$

DINA 모형의 문항반응함수에 따르면, 문항 j가 요구하는 인지요소를 모두 숙달한($\eta_{ij}=1$) 피험자가 문항의 정답을 맞힐 확률은 $1-s_j$이며, 그렇지 않은($\eta_{ij}=0$) 피험자가 문항의 정답을 맞힐 확률은 g_j이다.

반면, DINO 모형(Templin & Henson, 2006)은 어떤 피험자가 문항이 요구하는 인지요소 중 적어도 하나 이상 숙달하면 그 문항의 정답을 맞힐 수 있다는 비결합적 규칙을 따르는 보상적 모형이다. DINO 모형은 DINA 모형의 보상형 버전으로 볼 수 있으며, DINA 모형과 유사하게 각 문항 수준에서 부주의 모수와 추측 모수를 허용한다.

DINO 모형도 DINA 모형과 마찬가지로 세 가지 핵심 요소로 이루어진다(Rupp et al., 2010, pp. 132-133). 첫 번째 요소는 DINA 모형과 동일한 결정론적 투입으로, 피험자 i가 문항 j의 정답을 맞힐 수 있을 것이라 기대되는지의 여부를 나타내는 잠재변수를 DINA 모형과는 다른 기호를 사용하여 ξ_{ij}로 나타낸다. 두 번째 요소는 ξ_{ij}를 결정하는 비결합적 방식을 나타내는 “or” gate로, 이에 따라 DINO 모형의 잠재변수 ξ_{ij}는 다음과 같이 표현된다.

$$\xi_{ij} = 1 - \prod_{k=1}^{K} (1-\alpha_{ik})^{q_{jk}} \tag{2.10}$$

ξ_{ij}는 피험자 i가 문항 j에 필요한 인지요소를 적어도 하나 이상 숙달하였으면 1의 값을 가지며, 그렇지 않은 경우 0의 값을 가진다.

세 번째 요소는 실제 문항 응답의 확률적 결정에 대한 가정으로, 문항 j의 부주의 모수와 추측 모수는 다음과 같이 표현된다.

$$s_j = P(X_{ij} = 0 | \xi_{ij} = 1) \tag{2.11}$$

$$g_j = P(X_{ij} = 1 | \xi_{ij} = 0) \tag{2.12}$$

부주의 모수 s_j는 문항 j가 필요로 하는 인지요소를 적어도 하나 이상 숙달한($\xi_{ij} = 1$) 피험자가 문항 j의 정답을 맞히지 못할 확률을 나타내며, 추측 모수 g_j는 문항 j가 필요로 하는 인지요소를 하나도 숙달하지 못한($\xi_{ij} = 0$) 피험자가 문항 j의 정답을 맞힐 확률을 나타낸다. 최종적으로, 피험자 i의 문항 j에 대한 잠재변수 ξ_{ij}와 문항 j에 대한 부주의 모수, 추측 모수를 모두 고려하면 DINO 모형의 문항반응함수는 다음과 같이 주어진다.

$$P(X_{ij} = 1 | \xi_{ij}) = (1-s_j)^{\xi_{ij}} g_j^{1-\xi_{ij}} \tag{2.13}$$

DINA 모형과 DINO 모형의 모수에 대한 이해를 돕기 위해, <표 Ⅱ-2>에서 가정했던 문항 3개, 인지요소 3개로 이루어진 예시 검사에 대해 두 모형의 모수가 각각 어떻게 표현되는지 살펴보자. 예시 검사의 Q-행렬이 다음의 <표 Ⅱ-3>과 같다고 하자.

<표 Ⅱ-3> 예시 검사의 Q-행렬

	인지요소 A_1	인지요소 A_2	인지요소 A_3
문항 1	0	1	0
문항 2	1	1	0
문항 3	0	1	1

이와 같은 Q-행렬을 갖는 검사에 대해 DINA 모형과 DINO 모형의 모수를 나타내면 각각 <표 Ⅱ-4>, <표 Ⅱ-5>와 같다. DINA 모형과 DINO 모형은 일반적으로 구조모수를 따로 추정하지는 않으나, 모든 피험자들의 인지요소패턴을 추정하고 나면 그에 따라 인지요소패턴의 분포 $P(\boldsymbol{\alpha}_c)$가 결정되므로 이를 구조모수 v_c로 나타내었다. <표 Ⅱ-4>와 <표 Ⅱ-5>로부터 DINA 모형과 DINO 모형은 각 문항의 문항모수가 2개씩이며 인지요소패턴 $\boldsymbol{\alpha}_c$의 인지요소 조합에 따라 각 문항을 맞힐 확률이 두 모수 중 하나로 표현됨을 확인할 수 있다.

<표 Ⅱ-4> 예시 검사에 대한 DINA 모형의 모수

c	$\boldsymbol{\alpha}_c$	구조모수 (v_c)	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	g_1	g_2	g_3
2	(0,0,1)	v_2	g_1	g_2	g_3
3	(0,1,0)	v_3	$1-s_1$	g_2	g_3
4	(0,1,1)	v_4	$1-s_1$	g_2	$1-s_3$
5	(1,0,0)	v_5	g_1	g_2	g_3
6	(1,0,1)	v_6	g_1	g_2	g_3
7	(1,1,0)	v_7	$1-s_1$	$1-s_2$	g_3
8	(1,1,1)	v_8	$1-s_1$	$1-s_2$	$1-s_3$

<표 Ⅱ-5> 예시 검사에 대한 DINO 모형의 모수

c	$\boldsymbol{\alpha}_c$	구조모수 (v_c)	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	g_1	g_2	g_3
2	(0,0,1)	v_2	g_1	g_2	$1-s_3$
3	(0,1,0)	v_3	$1-s_1$	$1-s_2$	$1-s_3$
4	(0,1,1)	v_4	$1-s_1$	$1-s_2$	$1-s_3$
5	(1,0,0)	v_5	g_1	$1-s_2$	g_3
6	(1,0,1)	v_6	g_1	$1-s_2$	$1-s_3$
7	(1,1,0)	v_7	$1-s_1$	$1-s_2$	$1-s_3$
8	(1,1,1)	v_8	$1-s_1$	$1-s_2$	$1-s_3$

1.2.3. LCDM

LCDM(Henson et al., 2009)은 앞서 살펴본 DINA 모형과 DINO 모형을 비롯한 다수의 모형을 아우를 수 있는 포괄적인 모형으로, 분산분석(ANOVA)과 유사한 모델링 방식을 취하며 주효과와 상호작용효과 모수를 포함한다(이연숙, 2013).

LCDM의 설명을 위해, 앞서 살펴본 예시 검사의 문항 2와 같이 두 개의 인지요소 1과 2를 측정하는 어떤 문항 j를 가정하고, 그 문항의 Q-행렬의 원소들을 나열한 벡터를 $\boldsymbol{q}_j = (1,1,0)$이라 하자. 어떤 피험자 i의 인지요소패턴은 $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})$와 같이 나타내자. 이때, LCDM에서 피험자 i가 문항 j의 정답을 맞힐 확률을 나타내는 문항반응함수는 다음과 같이 정의된다.

$$P(X_{ij}=1|\boldsymbol{\alpha}_i) = \frac{\exp(\lambda_{j,0} + \lambda_{j,1,(1)}\alpha_{i1} + \lambda_{j,1,(2)}\alpha_{i2} + \lambda_{j,2,(1,2)}\alpha_{i1}\alpha_{i2})}{1+\exp(\lambda_{j,0} + \lambda_{j,1,(1)}\alpha_{i1} + \lambda_{j,1,(2)}\alpha_{i2} + \lambda_{j,2,(1,2)}\alpha_{i1}\alpha_{i2})} \quad (2.14)$$

식 (2.14)에 포함된 문항 j의 문항모수는 총 4개로, 1개의 절편(intercept) 모수와 2개의 주효과(main effect) 모수, 1개의 상호작용효과(interaction effect) 모수로 구성된다. 절편 모수($\lambda_{j,0}$)는 문항 j가 필요로 하는 인지요소를 하나도 숙달하지 못한 피험자가 문항의 정답을 맞힐 로그-오즈(log-odds)를 나타낸다. 주효과 모수($\lambda_{j,1,(1)}, \lambda_{j,1,(2)}$)는 문항 j가 측정하는 인지요소 1과 2를 각각 숙달했을 때 문항의 정답을 맞힐 로그-오즈의 증가량을 의미하며, 상호작용효과 모수($\lambda_{j,2,(1,2)}$)는 인지요소 1과 2를 동시에 숙달함으로써 문항의 정답을 맞힐 로그-오즈의 증가량을 표현한다. 또한, λ의 첫 번째 아래첨자는 문항을, 두 번째 아래첨자는 모수의 유형을(0: 절편; 1: 주효과; 2: 상호작용효과), 세 번째 아래첨자는 주효과와 상호작용효과가 적용되는 인지요소가 무엇인지를 나타낸다.

LCDM의 일반적인 문항반응함수는 식 (2.15)와 같이 표현된다.

$$P(X_{ij}=1|\boldsymbol{\alpha}_i)=\frac{\exp(\lambda_{j,0}+\boldsymbol{\lambda}_j^T\boldsymbol{h}(\boldsymbol{\alpha}_i,\boldsymbol{q}_j))}{1+\exp(\lambda_{j,0}+\boldsymbol{\lambda}_j^T\boldsymbol{h}(\boldsymbol{\alpha}_i,\boldsymbol{q}_j))} \tag{2.15}$$

식 (2.15)에서 $\boldsymbol{q}_j$는 문항 j의 Q-행렬 원소들로 이뤄진 q-벡터이며, 절편모수 $\lambda_{j,0}$는 앞에서와 동일한 의미를 가진다. 벡터 $\boldsymbol{\lambda}_j$는 문항 j의 절편모수를 제외한 문항모수들을 나타내는 $(2^K-1)\times 1$ 벡터이며, $\boldsymbol{h}(\boldsymbol{\alpha}_i,\boldsymbol{q}_j)$는 인지요소패턴 $\boldsymbol{\alpha}_i$에 대해 문항 j의 각 문항모수들이 존재하는지를 크기 $(2^K-1)\times 1$인 벡터로 도출하는 지시함수(indicator function)로, 결과적으로 $\boldsymbol{\lambda}_j^T\boldsymbol{h}(\boldsymbol{\alpha}_i,\boldsymbol{q}_j)$는 다음의 식 (2.16)과 같이 표현된다:

$$\boldsymbol{\lambda}_j^T\boldsymbol{h}(\boldsymbol{\alpha}_i,\boldsymbol{q}_j)=\sum_{k=1}^{K}\lambda_{j,1,(k)}\alpha_{ik}q_{jk}+\sum_{k=1}^{K-1}\sum_{l>k}\lambda_{j,2,(k,l)}\alpha_{ik}q_{jk}\alpha_{il}q_{jl}+\cdots \tag{2.16}$$

한편, LCDM은 잠재집단모형의 일종으로서 구조모수 v_c를 포함하며, 그 의미는 일반적인 잠재집단모형에서와 동일하다.

앞에서 다뤘던 예시 검사에 대해 LCDM의 모수를 나타내면 <표 Ⅱ-6>과 같다(Q-행렬은 <표 Ⅱ-3> 참고). <표 Ⅱ-6>을 통해 알 수 있듯이 일반적인 LCDM에서 K'개의 인지요소를 필요로 하는 문항은 $2^{K'}$개의 문항모수를 가지게 된다.

<표 Ⅱ-6> 예시 검사에 대한 LCDM의 모수

c	$\boldsymbol{\alpha}_c$	v_c	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	$\lambda_{1,0}$	$\lambda_{2,0}$	$\lambda_{3,0}$
2	(0,0,1)	v_2	$\lambda_{1,0}$	$\lambda_{2,0}$	$\lambda_{3,0}+\lambda_{3,1,(3)}$
3	(0,1,0)	v_3	$\lambda_{1,0}$	$\lambda_{2,0}+\lambda_{2,1,(2)}$	$\lambda_{3,0}+\lambda_{3,1,(2)}$
4	(0,1,1)	v_4	$\lambda_{1,0}$	$\lambda_{2,0}+\lambda_{2,1,(2)}$	$\lambda_{3,0}+\lambda_{3,1,(2)}+\lambda_{3,1,(3)}+\lambda_{3,2,(2,3)}$
5	(1,0,0)	v_5	$\lambda_{1,0}+\lambda_{1,1,(2)}$	$\lambda_{2,0}+\lambda_{2,1,(1)}$	$\lambda_{3,0}$
6	(1,0,1)	v_6	$\lambda_{1,0}+\lambda_{1,1,(2)}$	$\lambda_{2,0}+\lambda_{2,1,(1)}$	$\lambda_{3,0}+\lambda_{3,1,(3)}$
7	(1,1,0)	v_7	$\lambda_{1,0}+\lambda_{1,1,(2)}$	$\lambda_{2,0}+\lambda_{2,1,(1)}+\lambda_{2,1,(2)}+\lambda_{2,2,(1,2)}$	$\lambda_{3,0}+\lambda_{3,1,(2)}$
8	(1,1,1)	v_8	$\lambda_{1,0}+\lambda_{1,1,(2)}$	$\lambda_{2,0}+\lambda_{2,1,(1)}+\lambda_{2,1,(2)}+\lambda_{2,2,(1,2)}$	$\lambda_{3,0}+\lambda_{3,1,(2)}+\lambda_{3,1,(3)}+\lambda_{3,2,(2,3)}$

1.3. 인지요소 위계를 고려한 인지진단모형

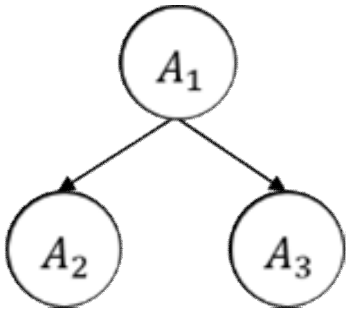
앞에서 다룬 인지진단모형들은 잠재집단모형의 측정모형에 제약을 부여한 모형들로, 인지요소 조합에 따라 문항 응답이 어떻게 결정되는지에 대한 서로 다른 가정을 토대로 문항반응함수를 표현한 것이었다. 이 절에서는 인지요소들 사이의 관계를 가정하고 그에 따라 잠재집단모형의 구조모형에 제약을 부여한 모형들에 초점을 두며, 그 중에서도 특히 인지요소들 사이에 '위계 관계'를 가정한 모형을 중심으로 살펴본다. 이를 위해 먼저 인지요소 위계의 의미를 확인하고 인지요소 위계를 고려한 선행연구들을 간략히 검토한다. 다음으로, 앞에서 검토한 인지진단모형들에 인지요소 위계에 대한 가정이 추가된 인지진단모형들을 살펴본다.

1.3.1. 인지요소 위계

인지요소 A_i가 인지요소 A_j의 숙달을 위해 필수적인 경우 두 인지요소 A_i와 A_j는 위계 관계를 가진다고 하며, A_i를 A_j의 선행 인지요소라 한다(Levy & Mislevy, 2016). 이와 같은 위계 관계를 기호 $A_i \rightarrow A_j$로 나타낸다. 또한, 어떤 검사의 인지요소들이 가지는 관계들이 모두 위계 관계인 경우 '인지요소 위계(attribute hierarchy)'가 존재한다고 표현한다(Leighton et al., 2004).

어떤 검사에 인지요소 위계가 가정되면, 피험자들이 가질 수 있는 가능한 인지요소패턴의 수가 줄어들게 된다. 예를 들어, 앞에서 다룬 예시 검사의 세 인지요소들 사이에 [그림 Ⅱ-1]의 (a)와 같은 인지요소 위계가 존재하는 경우, A_1이 A_2와 A_3의 선행 인지요소이므로 인지요소 A_1을 숙달하지 못한 피험자는 A_2나 A_3를 숙달할 수 없다. 따라서 이 상황에서 세 가지 인지요소패턴 $(0,0,1), (0,1,0), (0,1,1)$은 불가능하므로 가능한 인지요소패턴의 수가 $2^3 = 8$에서 5로 줄어든다([그림 Ⅱ-1]의 (b) 참고).

(a)



(b)

(A_1, A_2, A_3)
(0,0,0)
(0,0,1)
(0,1,0)
(0,1,1)
(1,0,0)
(1,0,1)
(1,1,0)
(1,1,1)

[그림 Ⅱ-1] 세 인지요소가 이루는 위계 구조의 예시(a) 및 가능한 인지요소패턴과 불가능한 인지요소패턴(b)

K. K. Tatsuoka(1993; 1995)는 위계 관계를 가지는 인지요소들을 인지진단이론에 처음 도입하였다. K. K. Tatsuoka는 인지요소들이 이루는 위계 관계의 구조를 '지식 구조(knowledge structure)'라 표현하였으며, 행렬 대수를 통해 지식 구조를 수리적으로 나타내고 인지 진단에 활용하는 규칙장 모형(rule space model)을 제시하였다. Leighton 외(2004)는 규칙장 모형을 변형한 AHM을 제안하였으며, 인지요소들이 이루는 위계 관계를 '인지요소 위계'라 칭하고 인지요소 위계가 있을 때 인지요소패턴을 분류하는 여러 가지 방법론들을 제안하였다(Leighton et al., 2004; Gierl, Cui, & Hunka, 2008). Su(2013)은 DINA 모형과 DINO 모형에 인지요소 위계에 대한 제약을 부여한 모형인 DINA-H 모형과 DINO-H 모형을 제시하였으며, Templin과 Bradshaw(2014)는 LCDM에 인지요소 위계에 대한 제약을 부여하여 LCDM의 내재모형(nested model)으로 HDCM을 정의하였다.

1.3.2. DINA-H 모형과 DINO-H 모형

DINA-H 모형과 DINO-H 모형(Su, 2013)은 앞의 1.2.2절에서 살펴본 DINA 모형과 DINO 모형에 인지요소 위계에 대한 가정이 각각 추가된 모형으로, 측정모형에 해당하는 문항반응함수의 수학적 표현은 기존의 모형과 동일하지만 인지요소 위계에 따라 피험자들이 나타낼 수 있는 인지요소패턴이 제한된다는 점에서 차이가 있다(Su, 2013).

구체적으로 K개의 인지요소를 측정하는 검사를 가정했을 때, 기존의 DINA 모형과 DINO 모형에서 가능한 인지요소패턴의 수는 2^K개인 반면, DINA-H 모형과 DINO-H 모형에서 인지요소 위계에 따라 줄어든 R개의 인지요소패턴만 가능하다. 따라서 기존의 DINA, DINO 모형에서와 달리 DINA-H, DINO-H 모형은 인지요소패턴 추정의 과정에서 피험자들의 인지요소패턴을 R개의 인지요소패턴 중 하나로 추정하게 된다. 기존의 DINA, DINO 모형의 모수는 총 $2J+2^K-1$개(J: 문항 수)인 반면 DINA-H, DINO-H 모형의 모수는 총 $2J+R-1$개다. DINA-H, DINO-H 모형 하에서 가능한 인지요소패턴의 수 R은 인지요소 위계의 구조에 따라 결정된다(Su, 2013, pp. 49-50).

앞에서 다뤘던 <표 Ⅱ-3>과 같은 Q-행렬을 갖는 예시 검사에 [그림 Ⅱ-1]과 같은 인지요소 위계가 가정된 경우에 대해 DINA-H 모형과 DINO-H 모형의 모수를 나타내면 각각 <표 Ⅱ-7>, <표 Ⅱ-8>과 같다. DINA-H, DINO-H 모형에서 측정모형을 이루는 문항모수는 기존의 모형과 동일하게 문항 당 2개씩 총 6개인 것을 확인할 수 있으며, 인지요소 위계에 따라 가능한 인지요소패턴의 수가 5개로 줄어듦에 따라 구조모형을 이루는 구조모수 v_c가 4개로 줄어든 것을 확인할 수 있다(표에 제시된 구조모수는 $v_1 \sim v_5$로 5개이지만, 이들의 합은 항상 1이기 때문에 실질적으로 추정해야 하는 구조모수는 4개이다). DINA-H 모형과 DINO-H 모형은 인지요소 위계에 대한 가정에 따라 불가능한 인지요소패턴에 대응하는 구조모수 v_c를 0으로 제약한 모형이라고 볼 수 있다.

<표 Ⅱ-7> 예시 검사에 대한 DINA-H 모형의 모수

c	$\boldsymbol{\alpha}_c$	구조모수 (v_c)	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	g_1	g_2	g_3
2	(1,0,0)	v_2	g_1	g_2	g_3
3	(1,0,1)	v_3	g_1	g_2	g_3
4	(1,1,0)	v_4	$1-s_1$	$1-s_2$	g_3
5	(1,1,1)	v_5	$1-s_1$	$1-s_2$	$1-s_3$

<표 Ⅱ-8> 예시 검사에 대한 DINO-H 모형의 모수

c	$\boldsymbol{\alpha}_c$	구조모수 (v_c)	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	g_1	g_2	g_3
2	(1,0,0)	v_2	g_1	$1-s_2$	g_3
3	(1,0,1)	v_3	g_1	$1-s_2$	$1-s_3$
4	(1,1,0)	v_4	$1-s_1$	$1-s_2$	$1-s_3$
5	(1,1,1)	v_5	$1-s_1$	$1-s_2$	$1-s_3$

1.3.3. HDCM

HDCM(Templin & Bradshaw, 2014)은 앞의 1.2.3절에서 살펴본 LCDM에 인지요소 위계에 대한 가정이 추가된 모형이다. HDCM은 주어진 인지요소 위계에 따라 LCDM의 문항모수와 구조모수 일부를 0으로 제약한 모형으로, LCDM의 내재모형이라 할 수 있다.

LCDM의 설명에서 다루었던 예를 다시 가져와서, 두 개의 인지요소 1과 2를 측정하는 어떤 문항 j를 가정하고, 그 문항의 Q-행렬의 원소들을 나열한 벡터를 $\boldsymbol{q}_j = (1,1,0)$이라 하자. 그리고 이번에는 인지요소 1이 인지요소 2의 선행 인지요소라는 위계 관계를 가정하자. HDCM에서는 이와 같은 인지요소의 위계 관계를 인지요소 2가 인지요소 1에 내재되어 있다고 표현하며 이를 $\alpha_{2(1)}$으로 나타낸다. 즉, 아래 첨자의 괄호 안의 인지요소가 선행 인지요소이다. HDCM에서 피험자 i의 인지요소패턴은 $\boldsymbol{\alpha}_i^* = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})$로 나타내며, 여기서 위 첨자는 이 인지요소패턴이 인지요소 위계에 따라 불가능한 인지요소패턴들은 제외하고 남은 패턴들 중 하나임을 나타낸다. HDCM에서 피험자 i가 문항 j의 정답을 맞힐 확률을 나타내는 문항반응함수는 다음과 같이 표현된다.

$$P(X_{ij}=1|\boldsymbol{\alpha}_i^*) = \frac{\exp(\lambda_{j,0} + \lambda_{j,1,(1)}\alpha_{i1} + \lambda_{j,2,(2(1))}\alpha_{i1}\alpha_{i2})}{1+\exp(\lambda_{j,0} + \lambda_{j,1,(1)}\alpha_{i1} + \lambda_{j,2,(2(1))}\alpha_{i1}\alpha_{i2})} \tag{2.17}$$

식 (2.17)을 LCDM에서의 문항반응함수 식 (2.14)와 비교해보면, 인지요소 위계로 인해 인지요소 1의 숙달 없이 인지요소 2만 숙달하는 경우가 없으므로 그에 대한 주효과모수인 $\lambda_{j,1,(2)}$가 제외되었으며, 인지요소 1과 2를 동시에 숙달하는 경우에 대응하는 상호작용효과 모수($\lambda_{j,2,2(1)}$)의 세 번째 아래 첨자가 인지요소 위계를 반영하여 2(1)로 표현되어 있음을 알 수 있다. HDCM에서 상호작용효과 모수($\lambda_{j,2,2(1)}$)는 인지요소 1이 숙달되어 있는 상태에서 인지요소 2를 숙달함으로써 문항의 정답을 맞힐 로그-오즈가 얼마나 증가하는지를 의미하는 것으로 볼 수 있다.

앞에서 다뤘던 <표 Ⅱ-3>과 같은 Q-행렬을 갖는 예시 검사에 [그림 Ⅱ-1]과 같은 인지요소 위계가 가정된 경우(문항 1은 인지요소 A_1을, 문항 2는 인지요소 A_1과 A_2를, 문항 3은 인지요소 A_2와 A_3을 필요로 하며, 인지요소 위계는 $A_1 \rightarrow A_2, A_1 \rightarrow A_3$로 가정된 경우)에 대한 HDCM의 모수를 나타내면 <표 Ⅱ-9>와 같다. LCDM의 모수(<표 Ⅱ-6>)와 비교해보면, 구조모수의 경우 DINA-H, DINO-H 모형에서와 마찬가지로 인지요소 위계를 반영하여 그 수가 줄어들었으며, 문항모수의 경우 위계관계를 가진 두 인지요소를 동시에 필요로 하는 문항 2의 경우에서 인지요소 A_2의 주효과모수가 제외되고, 상호작용효과의 모수는 $\lambda_{2,2,2(1)}$로 표기가 달라진 것을 확인할 수 있다.

<표 Ⅱ-9> 예시 검사에 대한 HCDM의 모수

c	$\boldsymbol{\alpha}_c^*$	v_c	문항 1	문항 2	문항 3
1	(0,0,0)	v_1	$\lambda_{1,0}$	$\lambda_{2,0}$	$\lambda_{3,0}$
2	(1,0,0)	v_2	$\lambda_{1,0}$	$\lambda_{2,0}+\lambda_{2,1,(1)}$	$\lambda_{3,0}$
3	(1,0,1)	v_3	$\lambda_{1,0}$	$\lambda_{2,0}+\lambda_{2,1,(1)}$	$\lambda_{3,0}+\lambda_{3,1,(3)}$
4	(1,1,0)	v_4	$\lambda_{1,0}+\lambda_{1,1,(2)}$	$\lambda_{2,0}+\lambda_{2,1,(1)}+\lambda_{2,2,2(1)}$	$\lambda_{3,0}+\lambda_{3,1,(2)}$
5	(1,1,1)	v_5	$\lambda_{1,0}+\lambda_{1,1,(2)}$	$\lambda_{2,0}+\lambda_{2,1,(1)}+\lambda_{2,2,2(1)}$	$\lambda_{3,0}+\lambda_{3,1,(2)}+\lambda_{3,1,(3)}+\lambda_{3,2,(2,3)}$

1.4. 인지요소패턴의 추정

인지진단평가에서 각 피험자의 인지요소패턴을 추정하기 위해 MLE, MAP, EAP 추정 방법을 이용할 수 있다(Huebner & Wang, 2011, pp. 409-411).

피험자의 인지요소패턴 추정 방법들은 기본적으로 각 인지요소패턴별 피험자 i의 문항응답패턴 $\boldsymbol{x}_i$에 대한 우도 함수(likelihood function)를 기반으로 하며, 이는 다음과 같다.

$$L(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{\alpha}_c) = \prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_c)^{x_{ij}}(1 - P_j(\boldsymbol{\alpha}_c))^{1-x_{ij}} \tag{2.18}$$

식 (2.18)에서 $P_j(\boldsymbol{\alpha}_c)$는 각 인지진단모형(측정모형)에서 정의한 문항반응함수이다. 예를 들어, DINA 모형에서 우도 함수는 다음과 같이 표현된다.

$$L(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{\alpha}_c) = \prod_{j=1}^{J} [(1-s_j)^{x_{ij}} s_j^{1-x_{ij}}]^{\eta_{ij}} [g_j^{x_{ij}}(1-g_j)^{1-x_{ij}}]^{1-\eta_{ij}} \tag{2.19}$$

우도 함수는 피험자 i의 인지요소패턴이 $\boldsymbol{\alpha}_c$일 때 문항응답패턴 $\boldsymbol{x}_i$가 관찰될 가능성의 정도를 나타내는 것으로, 피험자 i의 진(true) 인지요소패턴은 실제로는 알 수 없는 것이므로 이 우도함수를 활용하여 인지요소패턴의 추정이 이루어진다(Huebner & Wang, 2011, p. 410).

MLE을 이용한 인지요소패턴 추정에서는 각 인지요소패턴 $\boldsymbol{\alpha}_c$별로 피험자 i의 응답에 대한 우도를 계산한 뒤 가장 큰 우도 값을 갖는 인지요소패턴을 그 피험자의 인지요소패턴으로 추정하며, 이를 식으로 표현하면 다음과 같다.

$$\hat{\boldsymbol{\alpha}}_{i,MLE} = \arg max_c \{L(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{\alpha}_c)\} \tag{2.20}$$

MAP 추정은 피험자가 각 인지요소패턴을 가질 확률에 대한 사전 정보가 있는 경우, 즉 사전 확률(prior probability) $P(\boldsymbol{\alpha}_c), c = 1,2,...,C,$ 가 있는 경우에 이를 반영하여 인지요소패턴을 추정하는 방식이다. 구체적으로, 피험자 i의 문항응답패턴 $\boldsymbol{x}_i$가 주어졌을 때 피험자 i가 특정 인지

요소패턴 $\boldsymbol{\alpha}_c$를 가질 사후 확률(posterior probability) $P(\boldsymbol{\alpha}_c|\boldsymbol{X}_i=\boldsymbol{x}_i)$는 베이즈 정리에 의해 다음과 같이 도출된다.

$$P(\boldsymbol{\alpha}_c|\boldsymbol{X}_i=\boldsymbol{x}_i)=\frac{L(\boldsymbol{X}_i=\boldsymbol{x}_i|\boldsymbol{\alpha}_c)P(\boldsymbol{\alpha}_c)}{\sum_{c=1}^{C}L(\boldsymbol{X}_i=\boldsymbol{x}_i|\boldsymbol{\alpha}_c)P(\boldsymbol{\alpha}_c)} \tag{2.21}$$

MAP 추정은 이 사후 확률이 가장 큰 인지요소패턴을 피험자 i의 인지요소패턴으로 추정하는 방식으로, 다음과 같이 표현된다.

$$\hat{\boldsymbol{\alpha}}_{i,MAP}=\arg max_c\{P(\boldsymbol{\alpha}_c|\boldsymbol{X}_i=\boldsymbol{x}_i)\} \tag{2.22}$$

참고로, 만약 모든 c에 대해 인지요소패턴 $\boldsymbol{\alpha}_c$의 사전 확률 $P(\boldsymbol{\alpha}_c)$가 동일한 값으로 주어진 경우(즉, 이산 균등분포(discrete uniform distribution)가 가정된 경우), 식 (2.20)의 MLE 추정과 식 (2.22)의 MAP 추정은 동일한 추정 방식이 된다. 다시 말해, 사전 분포가 균등 분포인 경우의 MAP 추정은 MLE 추정과 동일하다(Huebner & Wang, 2011, p. 410).

EAP 추정은 식 (2.21)의 사후 확률을 이용하여 각 인지요소 k에 대한 주변확률을 구한 뒤 이를 추정에 활용하는 방식을 취한다. 구체적으로, 인지요소 k의 주변확률 $\tilde{\alpha}_{i,k}$는 가능한 모든 인지요소패턴들 중에서 인지요소 k의 값이 1인 인지요소패턴들의 사후 확률 $P(\boldsymbol{\alpha}_c|\boldsymbol{X}_i=\boldsymbol{x}_i)$의 합으로 계산되며, 이를 식으로 표현하면 다음과 같다.

$$\tilde{\alpha}_{i,k}=\sum_{c=1}^{C}P(\boldsymbol{\alpha}_c|\boldsymbol{X}_i=\boldsymbol{x}_i)\,I(\alpha_{ck}=1) \tag{2.23}$$

여기서 $I(\alpha_{ck}=1)$은 인지요소패턴 $\boldsymbol{\alpha}_c$의 k번째 인지요소가 1인지의 여부를 나타내는 지시함수로

$$I(\alpha_{ck}=1)=\begin{cases}1, & \boldsymbol{\alpha}_c\text{의 } k\text{번째 인지요소가 1인 경우}\\ 0, & \text{그렇지 않은 경우}\end{cases} \tag{2.24}$$

와 같다. 위 식에 따라 피험자 i의 각 인지요소 k에 대한 주변확률 $\tilde{\alpha}_{i,k}$이 도출되면, $\tilde{\alpha}_{i,k}$가 0.5보다 크면 피험자 i의 k번째 인지요소를 1로 추정하고 그렇지 않으면 0으로 추정하는 방식으로 피험자 i의 인지요소패턴을 추정할 수 있다. 이와 같은 추정 방식을 EAP 추정이라고 한다(Huebner & Wang, 2011, p. 411).

2. 베이지안 네트워크

2.1. 베이지안 네트워크의 정의

베이지안 네트워크(BN)는 여러 확률변수들이 이루는 복잡한 확률분포를 그래프 기반의 표현을 이용하여 나타내는 확률 그래프 모형(probabilistic graphical model)의 일종으로, 변수들 간의 관계와 그에 따른 조건부 독립성 가정을 유향 비순환 그래프(DAG)로 나타내고 변수들의 결합확률분포를 조건부 확률분포(conditional probability distribution)를 이용하여 간결하게 표현한 것이다. 베이지안 네트워크는 조건부 독립성, 조건부 분포 등과 같은 확률론의 개념을 그래프 이론에 따른 그래프 구조인 네트워크로 표현한 것이므로, 베이지안 네트워크의 정의를 제시하기에 앞서 확률론과 그래프 이론의 몇 가지 핵심 개념들을 살펴볼 필요가 있다.

2.1.1. 유향 비순환 그래프와 조건부 독립

이 절에서는 베이지안 네트워크의 정의를 이해하기 위해 필요한 개념들을 살펴보도록 한다.

• 유향 비순환 그래프(DAG)

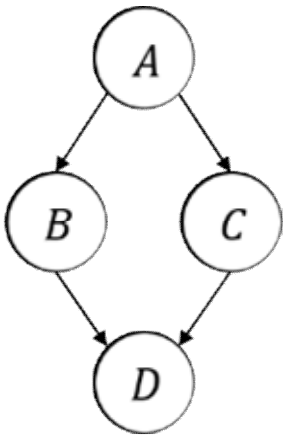
[그래프]

그래프 $K=(X,E)$란 일련의 노드(node) $X=\{X_1,X_2,\cdots,X_n\}$와 그들을 연결하는 변(edge)들로 이루어진 구조이다. 그래프의 서로 다른 두 노드 X_i, X_j에 대해, X_i에서 X_j로 향하는 유향 변(directed edge)이 존재하는 경우 $X_i \rightarrow X_j$가 E의 원소가 되며, X_i와 X_j가 방향이 없는 무향 변(undirected edge)으로 연결된 경우 $X_i - X_j$가 E의 원소가 된다.

[유향 비순환 그래프]

베이지안 네트워크가 가정하는 유향 비순환 그래프(DAG) $G=(X,E)$는 변의 집합 E가 유향 변들로만 이루어져 있으면서, 임의의 노드 X_i에 대해 유향 변들을 따라 X_i에서 출발하여 다시 X_i로 돌아오는 순환(cycle)이 없는 그래프를 의미한다.

$G=(X,E)$에서 두 노드 X_i, X_j가 유향 변 $X_i \rightarrow X_j$으로 연결되어 있을 때, 화살표를 보내는 노드 X_i를 X_j의 부모(parent)라고 하며, 화살표를 받는 노드 X_j를 X_i의 자식(child)이라고 한다. 또한, $G=(X,E)$의 두 노드 $X_{i'}$와 $X_{j'}$에 대해, 만약 $X_{i'}$에서 유향 변들을 따라 $X_{j'}$에 도달할 수 있으면 $X_{i'}$를 $X_{j'}$의 조상(ancestor)이라고 하며 $X_{j'}$는 $X_{i'}$의 자손(descendant)이라고 표현한다. 예를 들어, [그림 Ⅱ-2]의 DAG에서 노드 A는 노드 B의 부모이며, 노드 B는 노드 A의 자식이다. 또한 노드 A에서 유향 변들을 따라 노드 D에 도달할 수 있으므로, 노드 A는 노드 D의 조상이며 노드 D는 노드 A의 자손이 된다.



[그림 Ⅱ-2] 유향 비순환 그래프(DAG)의 예

• **조건부 독립(conditional independence)**

[사건의 독립과 조건부 독립]

확률 P가 정의된 사건 A, B에 대해

$$P(A|B) = P(A) \text{이거나 } P(B) = 0 \tag{2.25}$$

이면, 사건 A와 B는 서로 **독립**이라고 하며 이를 기호로 $P \models (A \perp B)$로 나타낸다. 또한, 사건 A와 B가 서로 독립이면

$$P(A, B) = P(A)P(B) \tag{2.26}$$

가 성립하며, 그 역도 성립한다.

확률 P가 정의된 사건 A, B, C에 대해

$$P(A|B, C) = P(A|C) \text{이거나 } P(B \cap C) = 0 \tag{2.27}$$

이면, 사건 C가 주어졌을 때 사건 A와 B는 **조건부 독립**이라고 하며 이를 기호로 $P \models (A \perp B|C)$로 나타낸다. 또한, 사건 C가 주어졌을 때 사건 A와 B가 조건부 독립이면

$$P(A, B|C) = P(A|C)P(B|C) \tag{2.28}$$

가 성립하며, 그 역 또한 성립한다.

[확률변수의 독립과 조건부 독립]

$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$를 각각 확률변수들의 집합이라 하고, 각 집합의 원소를 나열한 확률벡터의 성분들의 집합을 $Val(\boldsymbol{X})$, $Val(\boldsymbol{Y})$, $Val(\boldsymbol{Z})$라 하자.[2] $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$의 결합확률분포 $P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$와 모든 $\boldsymbol{x} \in Val(\boldsymbol{X}), \boldsymbol{y} \in Val(\boldsymbol{Y}), \boldsymbol{z} \in Val(\boldsymbol{Z})$에 대해

$$P \models ((\boldsymbol{X} = \boldsymbol{x}) \perp (\boldsymbol{Y} = \boldsymbol{y}) | (\boldsymbol{Z} = \boldsymbol{z})) \tag{2.29}$$

가 성립하면, $\boldsymbol{Z}$가 주어졌을 때 $\boldsymbol{X}$와 $\boldsymbol{Y}$는 **조건부 독립**이라고 하며, 이를 기호로 $\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}$ 로 나타낸다. 이때 $\boldsymbol{Z}$의 변수들에 대해 '관찰되었다'라고 표현하기도 한다. 또한, $\boldsymbol{Z}$가 주어졌을 때 $\boldsymbol{X}$와 $\boldsymbol{Y}$가 조건부 독립이면

$$P(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{Z}) = P(\boldsymbol{X}|\boldsymbol{Z})P(\boldsymbol{Y}|\boldsymbol{Z}) \tag{2.30}$$

가 성립하며, 그 역도 성립한다.

$\boldsymbol{Z} = \varnothing$ 인 경우 $\boldsymbol{X}$와 $\boldsymbol{Y}$는 **독립**이라고 하며, $\boldsymbol{X} \perp \boldsymbol{Y}$ 로 표현한다.

---

2) 예를 들어, $\boldsymbol{X} = \{X_1, X_2\}$이고 X_1과 X_2는 모두 0 또는 1의 값을 갖는 확률변수인 경우, $Val(\boldsymbol{X}) = \{(0,0), (0,1), (1,0), (1,1)\}$이다.

2.1.2. 베이지안 네트워크의 정의

베이지안 네트워크 $B=(G,P)$는 확률 변수 $X=\{X_1,X_2,\cdots,X_n\}$를 노드로 갖는 유향 비순환 그래프 $G=(X,E)$와 X에 대한 확률분포 P로 이루어지며, 다음과 같은 특징을 갖는 것으로 정의된다(Jensen & Nielsen, 2007; Koller & Friedman, 2009).

1) 그래프 G에서 각 변수 X_i의 부모인 변수들의 집합을 $Pa^G(X_i)$라 하고, X_i의 부모도 아니고 자손도 아닌 변수들의 집합을 $NDs^G(X_i)$라 하자. 이때 그래프 G는 다음과 같은 국지 독립성(local independencies)을 나타낸다.

$$\text{각 변수 } X_i\text{에 대해 } (X_i \perp NDs^G(X_i)|Pa^G(X_i))$$

다시 말해, 그래프 G에서 각 변수 X_i는 그 부모 변수들 $Pa^G(X_i)$이 주어졌을 때 자손이 아닌 변수들 $NDs^G(X_i)$과 조건부 독립이다.

2) $X=\{X_1,X_2,\cdots,X_n\}$의 결합확률분포 P는 다음과 같이 분해(factorize)된다.

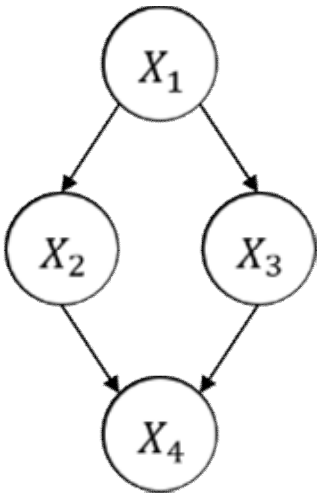
$$P(X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}P(X_i|Pa^G(X_i)) \tag{2.31}$$

식 (2.31)을 베이지안 네트워크에 대한 연쇄 법칙(chain rule)이라 한다. 식 (2.31)의 우변에 있는 각 인수 $P(X_i|Pa^G(X_i))$를 변수 X_i의 조건부 확률분포라 하며, X_i가 이산확률변수일 때 $P(X_i|Pa^G(X_i))$를 나타내는 표를 조건부 확률 표(CPT)라고 표현한다.

베이지안 네트워크의 핵심은 2)의 연쇄법칙으로, 확률 변수 $X=\{X_1,X_2,\cdots,X_n\}$의 결합확률분포 $P(X_1,X_2,\cdots,X_n)$가 각 변수들의 부모 변수가 주어졌을 때의 조건부 확률 $P(X_i|Pa^G(X_i))$들의 곱으로 표현되므로, 변수 X에 대한 그래프 구조와 함께 결합확률분포 P를 훨씬 간결하게 표현할 수 있다.

다음의 [그림 Ⅱ-3]은 베이지안 네트워크의 그래프 구조 G를 나타내는 DAG와 각 변수 $X_i(i=1,2,3,4)$들이 0 또는 1의 값을 갖는 이산 확률변수일 때, 각 변수들의 확률분포 P를 표현하는 CPT를 나타낸 것이다.

(a) DAG



(b) CPT

$P(X_1=0)$	$P(X_1=1)$
0.2	0.8

X_1의 CPT

$P(\boldsymbol{X_2}\|X_1)$	$X_2=0$	$X_2=1$
$X_1=0$	0.7	0.3
$X_1=1$	0.2	0.8

X_2의 CPT

$P(\boldsymbol{X_3}\|X_1)$	$X_3=0$	$X_3=1$
$X_1=0$	0.8	0.2
$X_1=1$	0.5	0.5

X_3의 CPT

$P(\boldsymbol{X_4}\|X_2,X_3)$	$X_4=0$	$X_4=1$
$X_2=0,X_3=0$	0.1	0.9
$X_2=0,X_3=1$	0.2	0.8
$X_2=1,X_3=0$	0.3	0.7
$X_2=1,X_3=1$	0.4	0.6

X_4의 CPT

[그림 Ⅱ-3] 베이지안 네트워크의 DAG와 CPT

베이지안 네트워크의 정의 2)의 연쇄법칙은 1)의 국지 독립성 가정으로부터 도출되는 것이다. [그림 Ⅱ-3]의 예를 이용하여 설명해보자. 네 변수의 결합확률분포는 확률에 대한 연쇄법칙에 따라 다음과 같이 분해된다.

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3) \quad (2.32)$$

식 (2.32)은 네 변수에 대한 어떠한 가정이 없어도 조건부 확률의 정의에 의해 항상 성립하는 식이다. 이제 식 (2.32)의 우변의 세 번째 인수 $P(X_3|X_1, X_2)$에 주목해보자. [그림 Ⅱ-3]의 (a)에 주어진 그래프 구조 G에서 X_3의 부모는 X_1이므로 $Pa^G(X_3) = \{X_1\}$이며, X_3의 부모도 아니고 자손도 아닌 변수는 X_2이므로 $NDs^G(X_3) = \{X_2\}$이다. 변수 X_3에 대해 베이지안 네트워크의 정의 1)의 국지 독립성 가정을 적용하면 $(X_3 \perp X_2|X_1)$이 성립하며, 조건부 독립의 정의에 의해 다음이 성립한다.

$$P(X_3|X_2, X_1) = P(X_3|X_1) \quad (2.33)$$

다음으로 식 (2.32)의 우변의 마지막 인수 $P(X_4|X_1, X_2, X_3)$에 주목해보자. G에서 X_4의 부모는 X_2와 X_3이므로 $Pa^G(X_4) = \{X_2, X_3\}$이다. 그리고 X_4의 부모도 아니고 자손도 아닌 변수는 X_1이므로 $NDs^G(X_4) = \{X_1\}$이다. 이제 변수 X_4에 대해 국지 독립성 가정을 적용하면 $(X_4 \perp X_1|X_2, X_3)$가 성립하며, 조건부 독립의 정의에 의해 다음이 성립한다.

$$P(X_4|X_1, X_2, X_3) = P(X_4|X_2, X_3) \quad (2.34)$$

이제 식 (2.33)과 (2.34)의 결과를 식 (2.32)에 대입하면 다음과 같다.

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1)P(X_4|X_2, X_3) \quad (2.35)$$

그래프의 구조에 따르면 $Pa^G(X_2) = Pa^G(X_3) = \{X_1\}$, $Pa^G(X_4) = \{X_2, X_3\}$이므로, 이를 식 (2.35)에 대입하면 다음과 같이 베이지안 네트워크의 연쇄법칙이 성립함을 알 수 있다.

$$P(X_1, X_2, X_3, X_4) = \prod_{i=1}^{4} P(X_i|Pa^G(X_i)) \quad (2.36)$$

2.2. 베이지안 네트워크의 추론

베이지안 네트워크의 추론이란 베이지안 네트워크의 일부 변수에 대한 관찰값이 주어졌을 때 관찰되지 않은 변수들에 대한 사후 확률분포를 찾는 과정을 뜻한다(Nadkarni & Shenoy, 2001). 구체적인 과정은 다음과 같다(Kollter & Friedman, 2009, p. 287).

변수 집합 X에 대한 베이지안 네트워크에서, 관찰된 변수들을 나타내는 벡터를 $\boldsymbol{E}$라 하고 그에 대한 관찰값들의 벡터 $\boldsymbol{e}$가 주어졌다고 하자. 이때, 변수 집합 X에서 관찰되지 않은 변수들 중 일부 변수 $\boldsymbol{Y}$에 대한 사후 확률분포를 찾고자 하는 상황을 고려해보자. $\boldsymbol{Y}$의 사후 확률분포는 다음과 같이 정의된다.

$$P(\boldsymbol{Y}|\boldsymbol{E}=\boldsymbol{e}) = \frac{P(\boldsymbol{Y},\boldsymbol{e})}{P(\boldsymbol{e})} \tag{2.37}$$

식 (2.37)에서 분자가 가질 수 있는 각 값은 $P(\boldsymbol{y},\boldsymbol{e})$로 표현할 수 있으며, 이는 X에 대한 결합확률분포에서 $\boldsymbol{Y}=\boldsymbol{y}$이고 $\boldsymbol{E}=\boldsymbol{e}$에 해당하는 결합 확률들을 다 더함으로써 계산된다. 구체적으로, 변수 집합 X에서 관찰되지도 않았으며 사후 분포에 대한 관심도 없는 변수들을 $\boldsymbol{W}=X-\boldsymbol{E}-\boldsymbol{Y}$라고 하면, 식 (2.37)의 분자가 가지는 각 확률은 다음과 같이 계산된다.

$$P(\boldsymbol{y},\boldsymbol{e}) = \sum_{\boldsymbol{w}} P(\boldsymbol{y},\boldsymbol{e},\boldsymbol{w}) \tag{2.38}$$

여기서 $\boldsymbol{Y},\boldsymbol{E},\boldsymbol{W}$는 모두 베이지안 네트워크의 변수들이므로, 식 (2.38)의 각 확률 $P(\boldsymbol{y},\boldsymbol{e},\boldsymbol{w})$은 X에 대한 결합확률분포의 값들이다.

식 (2.37)의 분모의 확률 $P(\boldsymbol{e})$은 분자에 해당하는 $P(\boldsymbol{Y},\boldsymbol{e})$를 모든 $\boldsymbol{y}$에 대해 더함으로써 얻어질 수 있다.

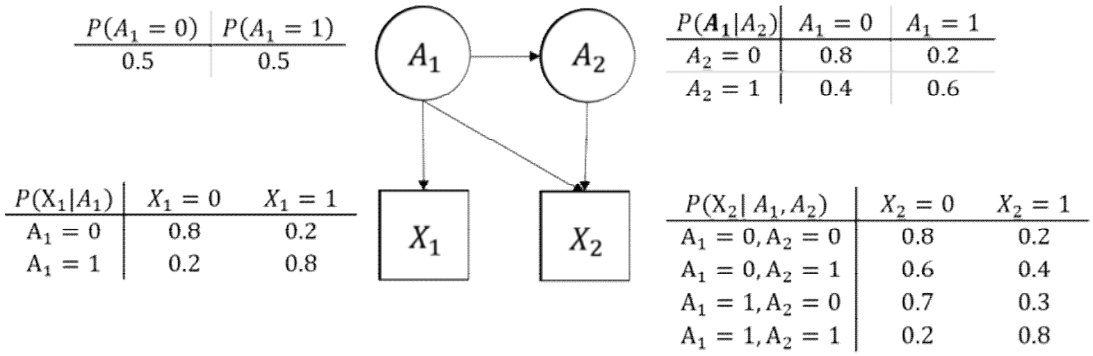
$$P(\boldsymbol{e}) = \sum_{\boldsymbol{y}} P(\boldsymbol{y},\boldsymbol{e}) \tag{2.39}$$

결과적으로 식 (2.38)과 식 (2.39)을 모두 계산한 뒤 각 $P(\boldsymbol{y}, \boldsymbol{e})$를 $P(\boldsymbol{e})$로 나누어주면, 식 (2.37)의 분포에 해당하는 각 확률 $P(\boldsymbol{y}|\boldsymbol{E}= e)$를 얻을 수 있다. 또한, 이 과정은 주변확률들의 벡터 $P(\boldsymbol{y}^1, \boldsymbol{e}), P(\boldsymbol{y}^2, \boldsymbol{e}), \cdots, P(\boldsymbol{y}^k, \boldsymbol{e})$ (단, $k = |Var(\boldsymbol{Y})|$)를 도출한 뒤 그 합이 1이 되도록 재정규화(renormalization)하는 과정과 동일하다. 이와 같은 방식으로 베이지안 네트워크의 추론이 이루어진다(Koller & Freidman, 2009, p. 287).

그러나 네트워크의 구조가 복잡해지고 변수의 수가 많아지면 위의 과정에서 주변확률 $P(\boldsymbol{y}, \boldsymbol{e})$의 계산은 매우 복잡하고 다루기 어려운 문제가 될 수 있다. 이와 같은 문제는 효율적인 계산을 목적으로 개발된 동적 프로그래밍 방식을 이용한 변수 제거(variable elimination) 알고리즘, 그래프 이론의 개념인 clique tree를 기반으로 한 sum-product message passing 알고리즘, sum-product divide 알고리즘 등의 방법을 이용하면 해결 가능함이 밝혀져 있다(Bertelé, & Brioschi, 1972, Shafer & Shenoy, 1990; Jensen, Olesen, & Andersen, 1990).

이후 3장에서 제시하는 베이지안 네트워크 기반의 인지진단모형에서는 베이지안 네트워크의 추론 과정을 통해 인지요소패턴의 추정이 이루어지며, 실제 검사에서는 R, MATLAB 등과 같은 소프트웨어를 이용하면 위의 알고리즘들을 활용한 인지요소패턴 추정이 이루어질 수도 있다. 그러나 이러한 알고리즘들의 계산 방법을 검토하는 것은 본 연구의 범위를 벗어나므로, 이 절에서는 비교적 간단한 베이지안 네트워크를 예시로 하여 앞에서 살펴본 베이지안 네트워크의 기본적인 추론 과정을 묘사해 보도록 하겠다.

4개의 변수로 이루어진 변수 집합 $X=\{A_{1,}A_{2,}X_{1,}X_{2}\}$에 대해 [그림 Ⅱ-4]와 같은 베이지안 네트워크가 주어져 있다고 가정하자. 이 중 두 변수 X_1과 X_2에 대해 $X_1=1,\ X_2=0$이 관찰되었다고 할 때, $\boldsymbol{A}=(A_1,A_2)$의 사후 확률분포 $P(A_1,A_2|X_1=1,X_2=0)$를 베이지안 네트워크의 추론 과정을 통해 계산해보자.



[그림 Ⅱ-4] 4개의 변수로 이루어진 베이지안 네트워크의 예

먼저, [그림 Ⅱ-4]와 같은 DAG가 주어져 있을 때, 네 변수들의 결합 확률분포는 다음과 같이 분해된다.

$$P(A_1,A_2,X_1,X_2)=P(A_1)P(A_2|A_1)P(X_1|A_1)P(X_2|A_1,A_2) \tag{2.40}$$

구하고자 하는 $\boldsymbol{A}=(A_1,A_2)$의 사후 확률분포 $P(A_1,A_2|X_1=1,X_2=0)$는 다음의 식 (2.41)과 같다.

$$P(A_1,A_2|X_1=1,X_2=0)=\frac{P(A_1,A_2,X_1=1,X_2=0)}{P(X_1=1,X_2=0)} \tag{2.41}$$

식 (2.41)의 분모는 식 (2.41)의 분자를 모든 순서쌍 (A_1,A_2)에 대해 더하면 얻어진다. 따라서 식 (2.41)의 각 확률을 구하기 위해 모든 순서쌍 (A_1,A_2)에 대해 식 (2.41)의 분자를 각각 계산하면 되고, 이는 [그림 Ⅱ-4]에 주어진 CPT를 이용하여 다음과 같이 계산된다.

$$
\begin{aligned}
&P(A_1=0, A_2=0, X_1=1, X_2=0) \\
&= P(A_1=0)P(A_2=0|A_1=0)P(X_1=1|A_1=0)P(X_2=0|A_1=0, A_2=0) \quad (2.42) \\
&= (0.5)(0.8)(0.2)(0.8) = 0.064
\end{aligned}
$$

$$
\begin{aligned}
&P(A_1=0, A_2=1, X_1=1, X_2=0) \\
&= P(A_1=0)P(A_2=1|A_1=0)P(X_1=1|A_1=0)P(X_2=0|A_1=0, A_2=1) \quad (2.43) \\
&= (0.5)(0.2)(0.2)(0.6) = 0.012
\end{aligned}
$$

$$
\begin{aligned}
&P(A_1=1, A_2=0, X_1=1, X_2=0) \\
&= P(A_1=1)P(A_2=0|A_1=1)P(X_1=1|A_1=1)P(X_2=0|A_1=1, A_2=0) \quad (2.44) \\
&= (0.5)(0.4)(0.8)(0.7) = 0.112
\end{aligned}
$$

$$
\begin{aligned}
&P(A_1=1, A_2=1, X_1=1, X_2=0) \\
&= P(A_1=1)P(A_2=1|A_1=1)P(X_1=1|A_1=1)P(X_2=0|A_1=1, A_2=1) \quad (2.45) \\
&= (0.5)(0.6)(0.8)(0.2) = 0.048
\end{aligned}
$$

식 (2.41)의 분모는 이 네 확률을 더하여 다음과 같이 계산된다.

$$
P(X_1=1, X_2=0) = \sum_{A_1}\sum_{A_2} P(A_1, A_2, X_1=1, X_2=0) = 0.236 \qquad (2.46)
$$

식 (2.42)~(2.45)에서 계산된 값을 식 (2.46)의 값으로 각각 나누어주면 사후 확률분포 $P(A_1, A_2|X_1=1, X_2=0)$가 다음과 같이 도출된다.

$$
P(A_1, A_2|X_1=1, X_2=0) = \begin{cases} 0.064/0.236 = 0.271, & A_1=0, A_2=0 \\ 0.012/0.236 = 0.051, & A_1=0, A_2=1 \\ 0.112/0.236 = 0.475, & A_1=1, A_2=0 \\ 0.048/0.236 = 0.203, & A_1=1, A_2=1 \end{cases} \qquad (2.47)
$$

2.3. 베이지안 네트워크의 학습

앞의 베이지안 네트워크에 대한 논의에서는 베이지안 네트워크의 그래프 구조인 DAG와 확률분포를 나타내는 CPT가 주어져 있는 것으로 가정하였다. 앞에서 살펴본 베이지안 네트워크의 추론에서도 주어진 DAG와 CPT를 이용하여 사후 확률을 찾는 방식을 취하였다.

베이지안 네트워크의 DAG와 CPT를 구성하는 방법에 대한 두 가지 접근법이 있다. 하나는 전문가의 도움으로 직접 베이지안 네트워크의 구조와 확률분포를 설정하는 방법이다. 그러나 이는 구조가 커질수록 복잡해지는 작업이 될 수 있으며, 여러 전문가들이 오랜 시간에 걸쳐 구성해야 할 수도 있는 다소 번거로운 일이다(Friedman, 1997). 또한, 전문가에 의해 구성된 구조와 분포가 실제로 타당한지에 대한 검증도 필요하다. 심지어는 베이지안 네트워크를 구성하고자 하는 영역에 대한 전문가가 존재하지 않는 경우도 있을 수 있다.

이와 같은 문제점을 해결하는 접근법은 기존에 수집된 자료를 이용하여 베이지안 네트워크의 구조와 확률분포를 찾는 방법으로, 이를 베이지안 네트워크의 학습(learning)이라고 한다. 베이지안 네트워크의 학습은 기본적으로 수집된 자료가 실제로 어떤 구조와 어떤 확률분포를 갖는 베이지안 네트워크에서 도출된 것이었을 가능성이 가장 높은지를 찾아내는 과정으로 볼 수 있다. 이때 참인 구조는 이미 알려진 것으로 가정하고 그 구조에 대한 확률분포를 나타내는 모수들의 참값을 추정하는 과정을 '모수 학습(parameter learning)'이라고 하며, 참인 구조를 찾는 과정을 '구조 학습(structure learning)'이라고 한다(Almond, Mislevy, Steinberg, Yan, & Williamson, 2015). 일반적으로 구조 학습 과정에서는 모수도 동시에 추정된다.

베이지안 네트워크의 학습은 수집된 자료가 베이지안 네트워크의 모든 변수에 대해 수집된 완전(complete) 자료인지, 일부 변수에 대해서만 수집되거나 결측이 있는 불완전(incomplete) 자료인지에 따라 그 방식이 달라지며, 일반적으로 불완전 자료를 이용한 학습 방법이 더 복잡하다.

그러나 앞으로 다루게 될 인지진단평가의 맥락에서는 피험자들의 인지상태를 나타내는 인지요소들이 관찰될 수 없는 잠재변수이므로 필연적으로 불완전 자료를 이용한 학습의 과정이 요구된다. 따라서 이 절에서는 우선 불완전 자료가 주어진 경우의 모수 학습 방법의 원리를 살펴볼 것이다. 구조 학습 방법의 경우, 이 절에서는 기본적인 원리와 개념을 중심으로 검토하며, 불완전 자료가 주어진 경우의 구조 학습 방법은 Ⅳ장의 연구와 직접적으로 연관되므로 Ⅳ장에서 자세히 살펴볼 것이다.

2.3.1. 모수 학습

모수 학습이란 베이지안 네트워크 $B=(G,P)$의 그래프 $G=(\boldsymbol{V},E)$는 알려져 있으나 확률분포 P는 모르는 상황에서, 분포 P로부터 관찰된 자료를 이용하여 확률분포 P의 CPT들의 확률 값을 추정하는 과정이다. 불완전 자료가 주어진 상황이란 베이지안 네트워크의 변수 집합 $\boldsymbol{V}$의 일부 변수에 대해서만 관찰 값이 주어진 경우이다(전체 변수에 대한 관찰자료가 주어졌으나 결측이 존재하는 경우는 다루지 않는다).

관찰되지 않은 변수들의 집합을 $\boldsymbol{A}$, 관찰된 변수들의 집합을 $\boldsymbol{X}$라 하고($\boldsymbol{V}=\boldsymbol{A}\cup\boldsymbol{X}$), 그래프 G에서 $\boldsymbol{X}$의 부모는 $\boldsymbol{A}$에 속하는 것으로 가정하자.[3] 이는 전체 변수 $\boldsymbol{V}$의 결합확률분포가 다음과 같이 분해될 수 있다는 의미이다.

$$P(\boldsymbol{V})=P(\boldsymbol{A})P(\boldsymbol{X}|\boldsymbol{A}) \tag{2.48}$$

다음으로, $\boldsymbol{A}$의 CPT에 대한 모수들의 집합을 γ라 하고, 관찰된 변수들($\boldsymbol{X}$)의 CPT에 대한 모수들의 집합을 π라 하자. 그러면 $\boldsymbol{A}$의 확률분포는 모수 γ에 의해 결정되고, $\boldsymbol{X}$의 확률분포는 π에 의해 결정되므로 식 (2.48)을 다음과 같이 표현할 수 있다.

$$P(\boldsymbol{V}|\gamma,\pi)=P(\boldsymbol{A}|\gamma)P(\boldsymbol{X}|\boldsymbol{A},\pi) \tag{2.49}$$

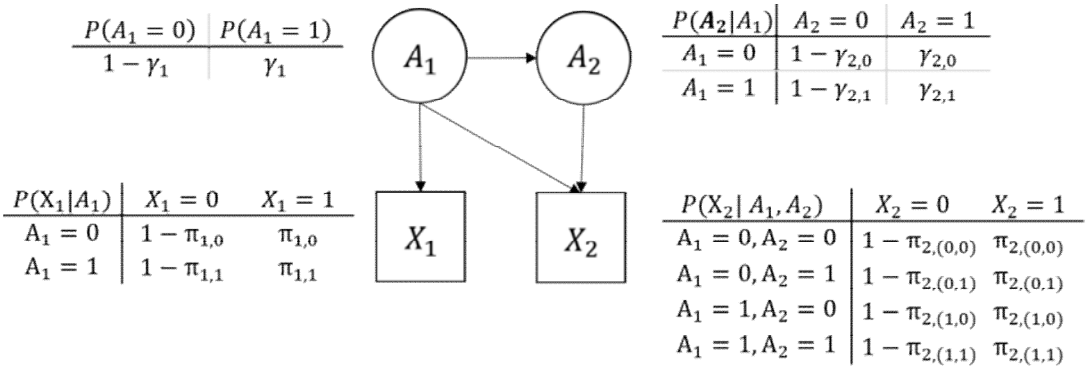
---

3) Ⅲ장에서 다루게 될 베이지안 네트워크 기반 인지진단모형의 상황을 묘사하기 위해 가정한 것으로, $\boldsymbol{A}$는 인지요소, $\boldsymbol{X}$는 문항에 해당한다.

다음 [그림 Ⅱ-5]는 위에서 가정한 상황에 대한 예시로, $\boldsymbol{A}=\{A_1, A_2\}$, $\boldsymbol{X}=\{X_1, X_2\}$, $\boldsymbol{\gamma}=\{\gamma_1, \gamma_{2,0}, \gamma_{2,1}\}$, $\boldsymbol{\pi}=\{\pi_{1,0}, \pi_{1,1}, \pi_{2,(0,0)}, \pi_{2,(0,1)}, \pi_{2,(1,0)}, \pi_{2,(1,1)}\}$이며

$$P(A_1, A_2, X_1, X_2|\boldsymbol{\gamma}, \boldsymbol{\pi}) = P(A_1, A_2|\boldsymbol{\gamma})P(X_1, X_2|A_1, A_2, \boldsymbol{\pi}) \tag{2.50}$$

가 성립한다.



[그림 Ⅱ-5] 4개의 변수로 이루어진 베이지안 네트워크의 모수 표현

불완전 자료가 주어진 경우의 모수 학습 방법은 EM 알고리즘을 이용한 방법과 MCMC 추정을 이용한 방법의 두 가지가 있다. 이 절에서는 EM 알고리즘을 이용한 방법을 살펴본다.

EM 알고리즘을 이용한 방법은 모수 $\boldsymbol{\gamma}$와 $\boldsymbol{\pi}$에 대한 MLE 추정을 원칙으로 하며, 이는 $\boldsymbol{X}$에 대해 관찰된 자료 $\boldsymbol{x}^*$의 우도

$$P(\boldsymbol{X}=\boldsymbol{x}^*|\boldsymbol{\gamma}, \boldsymbol{\pi}) \tag{2.51}$$

를 최대로 하는 $\hat{\boldsymbol{\gamma}}$와 $\hat{\boldsymbol{\pi}}$를 찾는 것이다. 그러나 변수 $\boldsymbol{A}$에 대해서는 관찰값이 없으므로, 식 (2.51)의 확률은 $\boldsymbol{A}$를 나열한 확률벡터의 모든 성분들 $Val(\boldsymbol{A})$에 대해서 $P(\boldsymbol{A}, \boldsymbol{X}=\boldsymbol{x}^*|\boldsymbol{\gamma}, \boldsymbol{\pi})$를 더하여(marginalize) 구할 수 있고, 식 (2.49)의 결과까지 적용하면 식 (2.51)의 우도를 다음과 같이 나타낼 수 있다.

$$P(\boldsymbol{X}=\boldsymbol{x}^{*}|\boldsymbol{\gamma},\boldsymbol{\pi}) = \sum_{\boldsymbol{a}\in Val(\boldsymbol{A})} P(\boldsymbol{A},\boldsymbol{X}=\boldsymbol{x}^{*}|\boldsymbol{\gamma},\boldsymbol{\pi})$$
$$= \sum_{\boldsymbol{a}\in Val(\boldsymbol{A})} P(\boldsymbol{A}|\boldsymbol{\gamma})P(\boldsymbol{X}=\boldsymbol{x}^{*}|\boldsymbol{A},\boldsymbol{\pi}) \qquad (2.52)$$

식 (2.52)을 주변 우도(marginal likelihood)라 하며, 단위 변환을 위해 주변 우도에 로그를 취한 것을 로그 주변 우도(log marginal likelihood)라 하며 다음과 같이 표현한다.

$$l(\boldsymbol{\gamma},\boldsymbol{\pi}|\boldsymbol{X}=\boldsymbol{x}^{*}) = \log\left[\sum_{\boldsymbol{a}\in Val(\boldsymbol{A})} P(\boldsymbol{A}|\boldsymbol{\gamma})P(\boldsymbol{X}=\boldsymbol{x}^{*}|\boldsymbol{A},\boldsymbol{\pi})\right] \qquad (2.53)$$

식 (2.53)의 로그 주변 우도를 최대로 하는 모수 $\hat{\boldsymbol{\gamma}}$와 $\hat{\boldsymbol{\pi}}$를 추정하는 것을 MMLE이라고 하며, MMLE를 가능하게 해주는 것이 EM 알고리즘이므로 MMLE/EM 알고리즘을 이용한 추정이라고 부르기도 한다.

EM 알고리즘을 통한 추정은 다음과 같이 이루어진다.

1) 추정하고자 하는 모수 $\boldsymbol{\gamma}$와 $\boldsymbol{\pi}$의 초기값 $\boldsymbol{\gamma}^{0}$와 $\boldsymbol{\lambda}^{0}$를 설정한다.

2) $t=0,1,2,3,\cdots$에 대해 다음의 과정을 반복한다.

E단계: 주어진 $\boldsymbol{\gamma}^{t},\boldsymbol{\pi}^{t}$를 이용하여 로그 주변 우도를 구한다.

$$l^{t}(\boldsymbol{\gamma},\boldsymbol{\pi}|\boldsymbol{X}=\boldsymbol{x}^{*}) = \log\left[\sum_{\boldsymbol{a}\in Val(\boldsymbol{A})} P(\boldsymbol{A}|\boldsymbol{\gamma}^{t})P(\boldsymbol{X}=\boldsymbol{x}^{*}|\boldsymbol{A},\boldsymbol{\pi}^{t})\right]$$
$$= E_{\boldsymbol{A}}\left[\log P(\boldsymbol{X}=\boldsymbol{x}^{*}|\boldsymbol{A},\boldsymbol{\pi}^{t})|\boldsymbol{\gamma}^{t},\boldsymbol{\pi}^{t},\boldsymbol{X}\right] \qquad (2.54)$$

식 (2.54)의 과정은 관찰된 자료 $\boldsymbol{x}^{*}$의 우도(식 (2.51))의 $\boldsymbol{A}$에 대한 기댓값(expectation)을 구하는 것과 동일하므로 E단계라고 한다.

M단계: $l^{t}(\boldsymbol{\gamma},\boldsymbol{\pi}|\boldsymbol{X}=\boldsymbol{x}^{*})$를 최대로 하는 모수 $\hat{\boldsymbol{\gamma}}$와 $\hat{\boldsymbol{\pi}}$을 찾아서 $\boldsymbol{\gamma}^{t+1}$와 $\boldsymbol{\pi}^{t+1}$로 설정한다.

3) $\boldsymbol{\gamma}^{t},\boldsymbol{\pi}^{t}$의 변화량이 일정한 기준 이하가 되면 2)의 반복을 종료하고 모수 추정치 $\hat{\boldsymbol{\gamma}}$와 $\hat{\boldsymbol{\pi}}$를 얻는다.

이와 같은 과정을 모수적 EM(parametric EM)이라고도 한다.

2.3.2. 구조 학습

구조 학습이란 어떤 베이지안 네트워크 $B^* = (G^*, P^*)$의 구조 $G^* = (\boldsymbol{X}, E^*)$와 분포 $P^*(\boldsymbol{X})$가 모두 알려져 있지 않은 상황에서, 어떤 분포 $P^*(\boldsymbol{X})$를 따르는 변수 $\boldsymbol{X}$에 대해 관찰된 자료 $\boldsymbol{x}^*$를 이용하여 베이지안 네트워크의 구조 G^*와 확률분포 P^*를 찾는 과정을 의미한다.

베이지안 네트워크의 구조 학습 방법은 크게 두 가지로 구분된다(Jensen & Nielsen, 2007, p. 230). 첫 번째 방법은 제한-기반(constraint-based) 구조 학습으로, 이 방법은 기본적으로 베이지안 네트워크가 일련의 독립성을 나타낸 것이라는 관점이 반영된 방법이다. 제한-기반 구조 학습 방법은 자료에서 조건부 독립성과 종속성을 검증한 뒤 그들을 가장 잘 나타내는 네트워크를 찾는 방식으로 이루어진다(Koller & Friedman, 2009, p. 785).

두 번째 방법은 점수-기반(score-based) 구조 학습으로, 이 방법은 베이지안 네트워크가 어떤 통계적 모형을 구체화한 것이라는 관점이 반영된 방법이며 구조 학습의 문제를 모형 선택의 문제로 다룬다. 이를 위해, 잠재적으로 가능한 베이지안 네트워크 구조들의 집합인 탐색 공간(search space)을 정의하고, 각 구조가 관찰된 자료와 얼마나 적합한지에 대한 측도인 점수 함수(scoring function)라는 것을 설정한 뒤, 탐색 공간의 구조들 중에서 점수가 가장 높은 구조를 찾는 것을 목표로 한다. 그러나 탐색 공간은 변수의 수가 늘어남에 따라 크기가 기하급수적으로 늘어나는 공간이므로, 탐색 공간에서 점수가 최대인 구조를 효율적으로 찾기 위해 적절한 탐색 절차(search procedure)를 활용한다(Koller & Friedman, 2009, p. 785). 이 절에서는 이후 Ⅳ장에서 다루는 불완전 자료에 대한 구조 학습 방법에서 활용되는 점수-기반 구조 학습 방법의 원리를 간단히 살펴보도록 한다.

점수-기반 구조 학습에서는 탐색 공간의 가능한 구조들에 점수를 부여한다. 점수는 각 구조로부터 자료가 도출되었을 가능성이 얼마나 되는

지를 반영한 수치이다. 일반적으로 많이 사용되는 점수 함수는 BIC(Schwarz, 1978)로, 이는 구조가 자료와 얼마나 적합한지를 측정함과 더불어 구조의 복잡성에 대해서도 고려한다. 구조 학습을 위한 BIC 점수는 다음과 같이 정의된다.

$$BIC(G|\boldsymbol{x}^*) = \log_2 P(\boldsymbol{x}^*|\hat{\theta}_G, G) - \frac{size(G)}{2} log_2(N) \quad (2.55)$$

여기서 G는 탐색 공간에 있는 후보 구조를 나타내며, $\boldsymbol{x}^*$는 관찰된 자료를 의미한다. $\hat{\theta}_G$는 자료 $\boldsymbol{x}^*$를 이용하여 추정한(예를 들어, MLE 방식으로 추정한) 구조 G의 모수들을 의미한다. $size(G)$는 구조의 크기를 나타내는 것으로, 구조 G의 모수의 수가 이에 해당한다. N은 관찰된 자료 $\boldsymbol{x}^*$의 표본수를 뜻한다. 식 (2.55)에서 $\frac{size(G)}{2} log_2(N)$는 복잡한 구조에 대해 일종의 페널티를 적용한 것이라고 볼 수 있다.

점수-기반 구조 학습에서는 크기가 매우 큰 탐색 공간의 모든 구조에 대해 BIC 점수를 계산하는 것이 현실적으로 불가능하므로 탐색 절차라는 것을 적용한다. 가장 대표적인 방법은 greedy hill-climbing 알고리즘으로, 이는 어떤 구조에서 출발하여 그 구조에 유향 변을 딱 하나만 추가하거나 삭제하거나 방향을 바꾼 경우의 구조들에 대해서만 BIC 점수를 계산한 뒤 점수를 가장 많이 증가시키는 구조를 다음 구조로 선택하고, 다음 구조에 대해 점수의 증가량이 일정 수준 이하가 될 때까지 위의 과정을 반복하여 구조를 찾는 절차이다(Nagarajan, Scutari, & Lèbre, 2013, p. 19).

BIC 점수의 식에 포함된 $\hat{\theta}_G$로부터 알 수 있듯이, BIC 점수를 계산하기 위해서는 후보 구조에 대한 모수 추정이 먼저 이루어져야 한다. 완전 자료가 주어진 경우에는 이 모수 추정이 간단하게 이루어지는 반면, 불완전 자료가 주어진 경우의 모수 추정은 앞의 2.3.1절에서 살펴봤듯이 EM 알고리즘을 적용한 반복에 의해 이루어지는 복잡한 과정을 거친다. 점수-기반 구조 학습의 과정에서는 탐색 공간의 여러 구조들에 대해 많은 횟수로 BIC 점수 계산을 해야 하기 때문에, 불완전 자료가 주어진 경우에는 복잡한 모수 추정의 과정을 여러 번 반복하는 것이 현실적으로

어려우므로 점수-기반 구조 학습 방법을 그대로 적용하는 것이 불가능하다(불완전 자료가 주어진 경우의 구조 학습 방법은 Ⅳ장에서 다룬다).

한편, BIC 점수는 score equivalence라는 특징을 갖는다(Koller & Friedman, 2009, p. 807). 이는 동형인(equivalent) 구조들은 모두 BIC 점수가 같다는 점을 의미한다. 여기서 동형 구조의 개념을 이해하려면 다음의 두 가지 개념을 먼저 살펴보아야 한다.

1) v-structure

베이지안 네트워크의 세 노드 X, Y, Z에 대해 $X \rightarrow Z \leftarrow Y$와 같이 서로 다른 두 노드가 동일한 한 노드로 향하는 구조를 v-structure라고 한다(Koller & Friedman, 2009, p. 71). 아래 [그림 Ⅱ-6]의 (a)는 v-structure의 예를 보여준다.

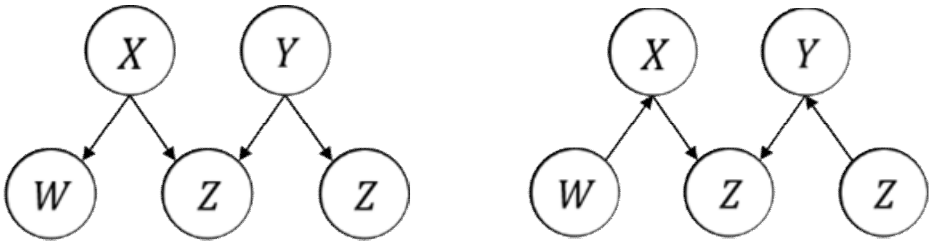
2) skeleton

베이지안 네트워크의 구조 G에 포함된 모든 유향 변의 방향을 제거한 무향 그래프(undirected graph)를 G의 skeleton이라고 한다(Jensen & Nielsen, 2007, p.231). 예를 들어, 다음 [그림 Ⅱ-6]의 (a)에 제시된 베이지안 네트워크 구조의 skeleton은 [그림 Ⅱ-6]의 (b)와 같다.

(a) v-structure　　(b) (a)에 제시된 구조의 skeleton

[그림 Ⅱ-6] v-structure와 skeleton의 예

이 때, 베이지안 네트워크의 두 구조 G_1, G_2에 대해, G_1과 G_2가 동일한 skeleton과 동일한 v-structure들의 집합을 가지는 경우, 두 구조는 **동형**이라고 한다(Koller & Friedman, 2009, p.77). 예를 들어, 다음 [그림 Ⅱ-7]에 제시된 두 구조는 동일한 skeleton과 동일한 v-structure의 집합 $\{X \rightarrow Z \leftarrow Y\}$을 가지므로 서로 동형이다.



[그림 Ⅱ-7] 동형인 베이지안 네트워크 구조의 예

그리고 BIC는 score equivalence라는 특징을 가지므로, 동일한 자료에 대해 [그림 Ⅱ-7]에 제시된 두 구조는 동일한 BIC 점수를 가진다.

3. 컴퓨터 적응검사

적응검사(adaptive testing)란 각 피험자에게 자신의 능력 수준에 맞는 문항을 선별적으로 제공함으로써 짧은 시간 안에 적은 수의 문항으로 피험자의 능력을 보다 정확하게 측정할 수 있다는 장점을 가진 검사 기법이다(Chang & Ying, 1996). 적응검사의 기본 원리는 검사의 각 단계에서 앞서 시행된 문항들에 대한 응답을 바탕으로 임시적으로 능력치를 추정하고 그 근방에서 정보를 많이 제공하는 문항을 다음 문항으로 선택하는 것이다(Wainer, 2000). 이와 같은 아이디어의 구현은 컴퓨터의 발전과 함께 가능해졌으며, 실제로 적응검사는 컴퓨터를 통해 이루어지므로 적응검사는 '컴퓨터 적응검사(CAT)'라는 명칭 아래 그 원리와 방법들이 발전되어 왔다.

컴퓨터 적응검사를 이루는 핵심 요소는 다음과 같다(Weiss & Kingsbury, 1984; Thompson, 2007)

1) 문항 반응 모형
2) 문제 은행
3) 검사 시작 기준
4) 문항 선택 방법
5) 검사 종료 기준

컴퓨터 적응검사의 문항 반응 모형은 일반적인 평가 모형의 발전에 따라 함께 변화하였다고 볼 수 있다. 초기에는 고전검사이론(classical test theory)을 모형으로 한 적응검사에서 출발하여(예를 들어, Linn, Rock, & Cleary, 1972), 문항반응이론(IRT)이 발전함에 따라 문항반응이론을 기반으로 한 컴퓨터 적응검사에 대한 연구가 이루어졌다(예를 들어, Reckase, 1983; Kingsbury & Weiss, 1983). 그리고 인지진단모형에 대한 관심이 높아짐에 따라 비교적 최근 들어 컴퓨터 적응검사에 인지진단모형을 적용한 인지진단 컴퓨터 적응검사(cognitive diagnostic CAT;

cheng, 2009)가 발전되기 시작하였다(예를 들어, C. Tatsuoka, 2002; C. Tatsuoka & Ferguson, 2003).

전통적인 컴퓨터 적응검사라 하면 일반적으로 IRT 기반의 컴퓨터 적응검사를 의미하며, 실제로 IRT 기반의 컴퓨터 적응검사에 대한 연구가 꽤 오랜 기간 활발히 이루어지고 있다. 따라서 이 절에서는 IRT 모형[4]을 측정모형으로 가정하고 앞서 나열한 핵심 요소를 중심으로 컴퓨터 적응검사의 원리를 살펴보고자 한다.

IRT 모형은 각 문항이 고유의 불변 특성을 가지는 것으로 가정하고 각 문항 특성에 기반하여 피험자의 능력을 추정하는 모형으로, 문항에 대한 피험자의 반응과 잠재적 특성 사이의 관계를 수리적 관계를 이용하여 분석한다(Wainer & Mislevy, 2000). IRT 모형은 일차원성(unidimensional) 가정과 지역 독립성(local independence) 가정 하에 문항모수의 불변성과 능력모수의 불변성이 확보된다는 장점을 가지며, 문항모수와 능력모수를 동일한 척도(θ)로 추정한다는 점 덕분에 컴퓨터 적응검사의 문항 반응 모형으로 사용되기에 적합하다.

문항반응이론에서는 피험자의 문항 반응에 대한 가정과 문항 특성에 대한 관점에 따라 1-모수(문항 난도(b_i)만 고려), 2-모수(문항 난도(b_i)와 변별도(a_i) 고려), 3-모수(문항 난도(b_i), 변별도(a_i), 추측도(c_i) 고려) IRT 모형으로 나뉜다. 세 모형에 대해 피험자 능력 θ에 따라 각 문항의 정답을 맞힐 확률을 나타내는 문항 반응 함수는 각각 다음과 같이 주어진다.

$$\text{1PL: } P_i(\theta) = \frac{1}{1+e^{-(\theta - b_i)}} \tag{2.56}$$

$$\text{2PL: } P_i(\theta) = \frac{1}{1+e^{-a_i(\theta - b_i)}} \tag{2.57}$$

$$\text{3PL: } P_i(\theta) = c_i + \frac{1-c_i}{1+e^{-a_i(\theta - b_i)}} \tag{2.58}$$

4) 단일차원(unidimensional) IRT 모형을 의미한다.

한편, 컴퓨터 적응검사의 실행을 위해서는 다양한 능력 수준(θ)을 측정할 수 있는 양질의 문항들이 개발되어 문항에 대한 다양한 정보들과 함께 컴퓨터에 저장되어 있어야 한다. 특히, 피험자 능력모수의 추정을 위해 컴퓨터 적응검사의 문제은행의 문항들에 대해서는 사전에 문항 특성(난이도, 변별도, 추측도 등의 문항모수치)에 대한 추정이 이루어져있어야 한다.

컴퓨터 적응검사의 시작 문항은 일반적으로 피험자 모집단의 능력 분포의 평균 $\hat{\theta}$에 맞추어지며, 모든 피험자들이 유사한 수준의 문항으로부터 출발하게 된다. 이 경우 초기 능력 수준에서 가장 정보를 많이 제공하는 문항이 모든 피험자들에게 실시된다면 그 문항의 노출 비율이 극단적으로 높아지므로 검사 보안과 관련한 문제가 발생할 수 있다(Wainer, 2000). 자격증 시험이나 피험자의 당락을 결정하기 위한 준거참조검사에서는 자격을 결정하는 기준에 해당하는 능력수준의 문항들로 검사를 시작할 수 있다(이혜옥, 2005).

컴퓨터 적응검사가 시작되고 피험자의 응답이 이루어지면 그 응답을 바탕으로 피험자의 능력 수준을 추정해야 한다. 문항 응답을 토대로 피험자의 능력을 추정하는 방법은 크게 MLE와 BME로 나뉜다.

피험자 i의 문항응답패턴 $\boldsymbol{x}_i$에 대한 능력모수 θ_i의 우도는 다음과 같이 정의된다.

$$l(\theta_i|\boldsymbol{x}_i) = P(\boldsymbol{x}_i|\theta_i, \boldsymbol{\beta}\} = \prod_j P_j(\theta_i)^{x_{ij}}[1-P_j(\theta_i)]^{1-x_{ij}} \qquad (2.59)$$

식 (2.59)에서 $\boldsymbol{\beta}$는 문항들에 대해 추정되어 있는 모수들의 벡터를 의미하며, 이는 각 문항반응함수 $P_j(\theta_i)$를 결정한다. 이때, 식 (2.59)의 우도를 최대로 만드는 $\hat{\theta}$을 찾아서 능력모수로 추정하는 방법이 MLE이다. 반면, BME는 능력모수에 대한 사전 분포 $P(\theta_i)$를 고려하여, 문항 응답이 주어졌을 때의 사후 분포

$$P(\theta_i|\boldsymbol{x}_i) \propto l(\theta_i|\boldsymbol{x}_i)P(\theta_i) \qquad (2.60)$$

를 최대로 만드는 $\hat{\theta}$을 능력모수로 추정하는 방식이다.

컴퓨터 적응검사에서 다음 문항을 선택하는 대표적인 방법으로는 최대정보문항(maximum item information) 선택 방법과 최대기대정확도(maximum expected precision) 문항 선택 방법이 있다(Thissen & Mislevy, 2000, p. 109). 이 중 보편적으로 사용되는 방법은 최대정보문항 선택 방법으로, 이는 능력모수 추정치 $\hat{\theta}_i$에서의 문항 정보

$$I_j(\hat{\theta}_i) = \frac{[P_j'(\hat{\theta}_i)]^2}{P_j(\hat{\theta}_i)[1 - P_j(\hat{\theta}_i)]} \tag{2.61}$$

가 최대가 되는 문항을 선택하는 방법이다. 식 (2.61)의 문항정보는 IRT에서 문항에 의한 피험자 능력 추정의 정확성을 나타내는 양으로, 능력수준과 문항 난이도가 일치할 때 그 값이 최대이다(한국교육평가학회, 2004).

컴퓨터 적응검사는 종료 기준에 따라 고정-길이 검사와 변동-길이 검사로 구분될 수 있다(Thissen & Mislevy, 2000, p. 112-113). 고정-길이 검사는 사전에 설정한 문항 수에 도달하였을 때 검사를 종료하는 방식으로, 실행이 용이하며 문항 사용 비율이 보다 정확하게 예측된다는 장점이 있다. 변동-길이 검사는 피험자 능력모수 추정의 오차가 사전에 설정한 기준 이하에 도달하면 검사를 종료하는 방법이다. 이 방법은 모든 피험자들의 능력모수가 동일한 수준의 정확도로 추정된다는 장점이 있다.

Ⅲ. BN-H: 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단모형

이 장에서는 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단모형인 BN-H 모형을 정의한다. BN-H 모형은 Ⅳ장에서 제시하는 위계 추정 방법과 Ⅴ장에서 제안하는 위계를 고려한 인지진단 컴퓨터 적응검사 방법의 토대가 되는 모형이다. 이하에서는 우선 인지요소 위계가 가정되지 않은 경우에 대한 일반적인 베이지안 네트워크 기반 인지진단모형을 검토한 뒤, 인지요소 위계를 가정한 BN-H 모형을 정의하고 Ⅱ장에서 다뤘던 기존의 인지요소 위계를 고려한 인지진단모형들과 어떤 차이점이 있는지 확인한다. 마지막으로 BN-H 모형을 이용한 인지진단 평가가 어떻게 이루어질 수 있는지 살펴본다.

1. 베이지안 네트워크 기반 인지진단모형

베이지안 네트워크 $B=(G,P)$는 여러 변수들의 관계와 복잡한 결합확률분포를 그래프 구조 G를 나타내는 DAG와 확률분포 P를 나타내는 CPT들로 표현하는 확률 그래프 모형으로, 베이지안 네트워크는 하나의 모형이라기보다는 여러 통계적 모형들을 나타낼 수 있는 일종의 프레임워크(framework)라 할 수 있다(Culbertson, 2016). 실제로 베이지안 네트워크를 이용하면 그 동안 제안되어온 여러 인지진단모형들을 표현할 수 있다. 이 절에서는 가장 일반적인 형태의 베이지안 네트워크 기반 인지진단 모형을 BN 모형이라고 칭하기로 하며, BN 모형의 정의와 특징을 살펴본다.

1.1. BN 모형

[BN 모형의 정의]

K개의 인지요소 $A_1, A_2, \cdots A_K$와 J개의 문항 $X_1, X_2, \cdots, X_J$으로 구성된 검사와 그에 대한 Q-행렬 $\boldsymbol{Q}=(q_{jk})$ $(j=1,2,\cdots,J;\ k=1,2,\cdots,K)$가 주어져있다고 하자. 이 검사를 나타내는 BN 모형은 다음과 같은 베이지안 네트워크 $B=(G,P)$로 구성된다.

- 그래프 구조 $G=(\boldsymbol{V},E)$의 노드 $\boldsymbol{V}$는 인지요소 변수 $\boldsymbol{A}=\{A_1, A_2, \cdots, A_K\}$과 문항 변수 $\boldsymbol{X}=\{X_1, X_2, \cdots, X_J\}$로 구성된다($\boldsymbol{V}=\boldsymbol{A}\cup\boldsymbol{X}$). 유향 변의 집합 E는 다음과 같이 구성된다.

  1) 인지요소 $A_k(k=1,2,\cdots,K)$와 문항 $X_j(j=1,2,\cdots,J)$에 대해, 문항 X_j가 인지요소 A_k를 필요로 하는 문항이면(즉, Q-행렬 $\boldsymbol{Q}$에서 $q_{jk}=1$이면) 유향 변 $A_k \rightarrow X_j$는 E의 원소가 된다.

  2) 만약, 서로 다른 두 인지요소 A_i, A_j에 대해 인지요소 A_i와 인지요소 A_j 사이에 특정 관계[5]를 가정하는 경우, 유향 변 $A_i \rightarrow A_j$ 또는 $A_j \rightarrow A_i$가 E의 원소가 된다.

- $\boldsymbol{V}=\{A_1, A_2, \cdots, A_K, X_1, X_2, \cdots, X_J\}$에 대한 확률분포 P는 다음과 같이 표현된다.

$$P(\boldsymbol{A},\boldsymbol{X}) = \underbrace{\prod_{k=1}^{K} P(A_k|Pa^G(A_k))}_{\text{구조모형}} \underbrace{\prod_{j=1}^{J} P(X_j|Pa^G(X_j))}_{\text{측정모형}} \tag{3.1}$$

식 (3.1)의 우변의 첫 번째 인수 $\prod_{k=1}^{K} P(A_k|Pa^G(A_k))$는 인지요소들 사이의 관계를 어떻게 설정하는지에 따라 결정되며, 이는 인지요소들 사이의 관계를 표현하는 부분이므로 잠재집단모형에서의 용어를 차용하여 **BN 모형의 구조모형**이라고 부르기로 한다. 만약 인지요소들 사이에 어떠한 관계도 가정하지 않는 경우 인지요소들은 서로 독립이라고 보

5) 인지요소 사이의 관계의 의미는 다음 절(Ⅲ장 1.2절)에서 다룬다.

며, 그 경우에는 다음의 식이 성립한다.

$$\prod_{k=1}^{K} P(A_k|Pa^G(A_k)) = \prod_{k=1}^{K} P(A_k) \tag{3.2}$$

식 (3.1)의 우변의 두 번째 인수 $\prod_{j=1}^{J} P(X_j|Pa^G(X_j))$에서 $Pa^G(X_j)$는 문항 X_j의 해결에 필요한 인지요소들의 집합이므로, $P(X_j|Pa^G(X_j))$는 결국 문항 X_j가 요구하는 인지요소들의 숙달 상태에 따른 문항 X_j의 응답 확률분포를 나타내는 것이다. 따라서 $P(X_j=1|Pa^G(X_j))$은 인지진단 모형의 문항반응함수를 나타내는 것이라고 볼 수 있다. 그런 의미에서 $\prod_{j=1}^{J} P(X_j|Pa^G(X_j))$를 **BN 모형의 측정모형**이라 부르기로 한다. ■

앞의 Ⅱ장에서 다뤘던 예시 검사를 BN 모형으로 표현해보도록 하자. Ⅱ장에서 다룬 예시 검사는 3개의 인지요소 A_1, A_2, A_3와 3개의 문항으로 이루어진 검사였으며, 예시 검사의 Q-행렬은 다음의 <표 Ⅲ-1>과 같았다(이 절에서는 문항 변수를 X_1, X_2, X_3로 표현하기로 한다).

<표 Ⅲ-1> 예시 검사의 Q-행렬

	인지요소 A_1	인지요소 A_2	인지요소 A_3
문항 X_1	0	1	0
문항 X_2	1	1	0
문항 X_3	0	1	1

세 인지요소 A_1, A_2, A_3 사이에 어떠한 관계도 가정되지 않은 경우의 BN 모형과, 인지요소 A_1이 인지요소 A_2, A_3와 각각 연관이 있다고 가정된 경우의 BN 모형은 각각 [그림 Ⅲ-1]과 [그림 Ⅲ-2]와 같다.

[그림 Ⅲ-1]과 [그림 Ⅲ-2]에서 인지요소들의 CPT는 BN 모형의 구조모형을 표현하며, 문항들의 CPT는 BN 모형의 측정모형을 나타낸다. 인지요소들 사이의 관계에 대한 가정이 추가되면 BN 모형의 구조모형은 변화하고 측정모형은 그대로 유지됨을 확인할 수 있다.

[그림 Ⅲ-1]에서 인지요소들의 CPT(즉, 구조모형)에 포함된 모수 $\gamma_k (k=1,2,3)$은 각 인지요소 A_k가 1일 확률을 나타낸다. 문항들의 CPT(측정모형)에 포함된 모수 π들의 경우, 첫 번째 아래 첨자는 문항을 나타내며 두 번째 아래 첨자는 그 문항을 맞히기 위해 필요한 인지요소들의 패턴이 어떠한 경우에 대한 조건부 확률을 의미하는 모수인지를 알려준다. [그림 Ⅲ-2]에서 $\lambda_{k,i} (k=2,3; i=0,1)$는 A_k의 부모 인지요소의 숙달 상태가 i일 때 A_k의 숙달 확률을 나타낸다.

인지요소 사이에 어떠한 관계도 없는 [그림 Ⅲ-1]의 BN 모형에서는 다음과 같은 연쇄법칙이 성립한다.

$$\begin{aligned} P(A_1, A_2, A_3, X_1, X_2, X_3) &= P(A_1, A_2, A_3)P(X_1, X_2, X_3|A_1, A_2, A_3) \\ &= P(A_1)P(A_2)P(A_3)P(X_1|A_2)P(X_2|A_1, A_2)P(X_3|A_2, A_3) \end{aligned} \tag{3.3}$$

[그림 Ⅲ-1]의 인지요소들의 CPT를 보면 각 인지요소별로 하나의 구조모수(γ)가 포함되어 있으며, 이와 같이 인지요소 사이의 관계가 가정되지 않은 BN 모형에서는 인지요소패턴의 분포 $P(A_1, A_2, A_3)$를 나타내는 구조모수의 수는 인지요소의 수와 동일하다.

반면 인지요소 사이에 특정 관계가 존재하는 [그림 Ⅲ-2]의 BN 모형에서는 다음과 같은 연쇄법칙이 성립한다.

$$\begin{aligned} P(A_1, A_2, A_3, X_1, X_2, X_3) &= P(A_1, A_2, A_3)P(X_1, X_2, X_3|A_1, A_2, A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1)P(X_1|A_2)P(X_2|A_1, A_2)P(X_3|A_2, A_3) \end{aligned} \tag{3.4}$$

[그림 Ⅲ-2]에서 알 수 있듯이, 인지요소 사이에 관계가 가정되면 그 관계에 따라 BN 모형의 구조모수(γ)의 수는 늘어난다. [그림 Ⅲ-2]의 경우에는 $P(A_1, A_2, A_3)$의 분포가 다섯 가지 모수 $\gamma_1, \gamma_{2,0}, \gamma_{2,1}\gamma_{3,0}, \gamma_{3,1}$에 의해 결정된다.

인지요소 A_1의 CPT

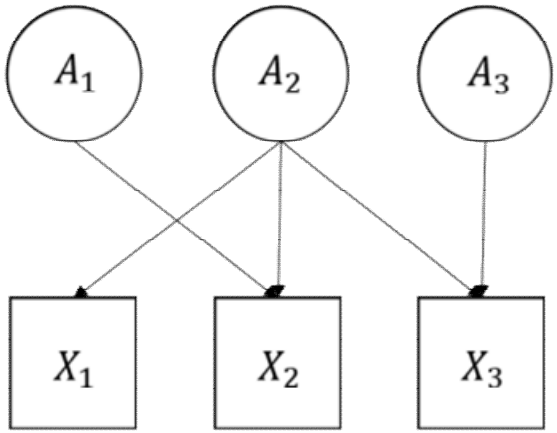
$P(A_1 = 0)$	$P(A_1 = 1)$
$1 - \gamma_1$	γ_1

인지요소 A_2의 CPT

$P(A_2 = 0)$	$P(A_2 = 1)$
$1 - \gamma_2$	γ_2

인지요소 A_3의 CPT

$P(A_3 = 0)$	$P(A_3 = 1)$
$1 - \gamma_3$	γ_3



문항 X_1의 CPT

$P(X_1 \mid A_2)$	$X_1 = 0$	$X_1 = 1$
$A_2 = 0$	$1 - \pi_{1,0}$	$\pi_{1,0}$
$A_2 = 1$	$1 - \pi_{1,1}$	$\pi_{1,1}$

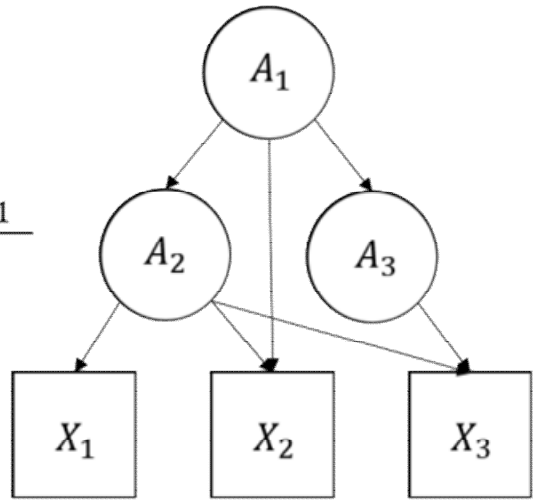
문항 X_2의 CPT

$P(X_2 \mid A_1, A_2)$	$X_2 = 0$	$X_2 = 1$
$A_1 = 0, A_2 = 0$	$1 - \pi_{2,(0,0)}$	$\pi_{2,(0,0)}$
$A_1 = 0, A_2 = 1$	$1 - \pi_{2,(0,1)}$	$\pi_{2,(0,1)}$
$A_1 = 1, A_2 = 0$	$1 - \pi_{2,(1,0)}$	$\pi_{2,(1,0)}$
$A_1 = 1, A_2 = 1$	$1 - \pi_{2,(1,1)}$	$\pi_{2,(1,1)}$

문항 X_3의 CPT

$P(X_3 \mid A_2, A_3)$	$X_3 = 0$	$X_3 = 1$
$A_2 = 0, A_3 = 0$	$1 - \pi_{3,(0,0)}$	$\pi_{3,(0,0)}$
$A_2 = 0, A_3 = 1$	$1 - \pi_{3,(0,1)}$	$\pi_{3,(0,1)}$
$A_2 = 1, A_3 = 0$	$1 - \pi_{3,(1,0)}$	$\pi_{3,(1,0)}$
$A_2 = 1, A_3 = 1$	$1 - \pi_{3,(1,1)}$	$\pi_{3,(1,1)}$

[그림 Ⅲ-1] 예시 검사에 대한 BN 모형(인지요소들이 서로 독립인 경우)



인지요소 A_1의 CPT

$P(A_1 = 0)$	$P(A_1 = 1)$
$1-\gamma_1$	γ_1

인지요소 A_2의 CPT

$P(A_2 \mid A_1)$	$A_2 = 0$	$A_2 = 1$
$A_1 = 0$	$1-\gamma_{2,0}$	$\gamma_{2,0}$
$A_1 = 1$	$1-\gamma_{2,1}$	$\gamma_{2,1}$

인지요소 A_3의 CPT

$P(A_3 \mid A_1)$	$A_3 = 0$	$A_3 = 1$
$A_1 = 0$	$1-\gamma_{3,0}$	$\gamma_{3,0}$
$A_1 = 1$	$1-\gamma_{3,1}$	$\gamma_{3,1}$

문항 X_1의 CPT

$P(X_1 \mid A_2)$	$X_1 = 0$	$X_1 = 1$
$A_2 = 0$	$1-\pi_{1,0}$	$\pi_{1,0}$
$A_2 = 1$	$1-\pi_{1,1}$	$\pi_{1,1}$

문항 X_2의 CPT

$P(X_2 \mid A_1, A_2)$	$X_2 = 0$	$X_2 = 1$
$A_1 = 0, A_2 = 0$	$1-\pi_{2,(0,0)}$	$\pi_{2,(0,0)}$
$A_1 = 0, A_2 = 1$	$1-\pi_{2,(0,1)}$	$\pi_{2,(0,1)}$
$A_1 = 1, A_2 = 0$	$1-\pi_{2,(1,0)}$	$\pi_{2,(1,0)}$
$A_1 = 1, A_2 = 1$	$1-\pi_{2,(1,1)}$	$\pi_{2,(1,1)}$

문항 X_3의 CPT

$P(X_3 \mid A_2, A_3)$	$X_3 = 0$	$X_3 = 1$
$A_2 = 0, A_3 = 0$	$1-\pi_{3,(0,0)}$	$\pi_{3,(0,0)}$
$A_2 = 0, A_3 = 1$	$1-\pi_{3,(0,1)}$	$\pi_{3,(0,1)}$
$A_2 = 1, A_3 = 0$	$1-\pi_{3,(1,0)}$	$\pi_{3,(1,0)}$
$A_2 = 1, A_3 = 1$	$1-\pi_{3,(1,1)}$	$\pi_{3,(1,1)}$

[그림 Ⅲ-2] 예시 검사에 대한 BN 모형(인지요소 사이의 관계가 가정된 경우)

한편, BN 모형의 문항모수 π에 대해 제약을 부여하면 앞서 살펴본 DINA 모형, DINO 모형, LCDM의 측정모형 등을 표현할 수 있다. 이와 관련한 내용은 2절에서 정의하는 BN-H 모형과 함께 다루도록 하겠다.

1.2. BN 모형에서 인지요소 사이의 관계의 의미

BN 모형은 베이지안 네트워크의 그래프 구조 G를 통해 변수들의 관계를 시각적으로 나타내는 모형이다. BN 모형의 그래프 구조에서 인지요소 노드로부터 문항 노드로 향하는 유향 변은 각 인지요소와 그 인지요소를 측정하는 문항의 관계를 나타내며, 이는 검사의 Q-행렬을 표현한 것이다. 반면, BN 모형의 그래프 구조에서 두 인지요소 노드 사이에 존재하는 유향 변은 인지요소 사이에 존재할 수 있는 다음과 같은 여러 관계들을 나타낼 수 있다(Almond et al., 2015).

1) 부분(part-of) 관계: 한 인지요소가 다른 인지요소의 하위 기능(subskill)인 경우를 의미한다. 부분 관계를 갖는 두 인지요소 사이의 유향 변은 일반적으로 복합 기능을 나타내는 인지요소에서 하위 기능을 나타내는 인지요소로 그 방향이 설정된다(Almond et al., 2015).

2) 선행(prerequisite) 관계: 한 인지요소의 숙달을 위해 다른 인지요소에 대한 일정 수준 이상의 숙달이 선행되어야 하는 경우를 뜻한다. 선행 관계를 갖는 두 인지요소 사이의 유향 변의 방향은 먼저 숙달되어야 하는 선행 인지요소에서 이후 숙달될 수 있는 후행 인지요소로 설정된다(Almond et al., 2015). 여기서 선행 관계는 선행 인지요소의 완전한 숙달이 이루어져야만 후행 인지요소의 숙달이 이루어지는 관계인 강한(hard) 선행관계와 선행 인지요소를 숙달하면 후행 인지요소를 숙달할 가능성이 높아지지만 선행 인지요소를 완전히 숙달하지 못했더라도 후행 인지요소의 숙달이 가능한 관계를 나타내는 약한(soft) 선행 관계를 모두 나타낼 수 있다. 본 연구에서 초점을 두고 있는 위계 관계는 강한 선행 관계만을 의미하는 것이므로 선행 관계는 위계 관계를 포괄하는 개념이라고 할 수 있다.

3) 상관(correlation) 관계: 두 인지요소 사이에 부분 관계나 선행 관계와 같은 인과 관계는 존재하지 않지만 상관이 존재하는 경우를 의미한다. 예를 들어 두 인지요소에 대한 학습이 교과과정상 항상 동일한 시기에 이루어진다면 두 인지요소 사이에 상관관계가 존재할 수 있다(Almond et al., 2015).

두 인지요소 A_i, A_j에 대해 유향 변 $A_i \rightarrow A_j$가 그래프 구조에 포함되어 있는 경우, BN 모형에서 인지요소 A_j의 숙달 확률은 인지요소 A_i의 숙달 여부에 따라 결정되며 이는 조건부 확률 $P(A_j|A_i)$로 표현된다. 이러한 의미에서 BN 모형에서 두 인지요소 A_i, A_j사이에 유향변이 존재하는 경우 A_i와 A_j 사이에 종속(dependency) 관계가 존재한다고 표현하기로 한다. 여기서 종속 관계는 A_j의 숙달 확률이 A_i의 상태에 따라 결정된다는 확률적 의미의 종속을 나타내는 표현이며, 종속 관계가 실제로 나타내는 관계는 부분 관계, 선행 관계, 상관 관계 중 하나가 된다. 즉, 두 인지요소의 관계가 부분, 선행 또는 상관 관계를 나타낼 때 두 인지요소 사이에는 종속 관계가 존재한다고 표현할 수 있다.

이상의 내용을 도식화하여 나타내면 [그림 Ⅲ-3]과 같다.

[그림 Ⅲ-3] 인지요소 사이의 관계들의 포함 관계

2. BN-H 모형

베이지안 네트워크는 그래프 기반의 모형으로 인지요소 위계를 그래프 구조 DAG에 반영하여 직접적으로 나타낼 수 있으므로 인지요소 위계를 고려한 인지진단모형으로 활용될 수 있다. 따라서 이 절에서는 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단모형인 BN-H 모형을 정의한다. BN-H 모형은 앞에서 살펴본 BN 모형에 인지요소 사이의 위계 관계가 추가로 가정된 모형이라고 볼 수 있다. 여기서 위계 관계란 앞의 1.3.1절에서 설명했던 것과 동일한 개념으로, 서로 다른 두 인지요소 A_i,A_j에 대해 인지요소 A_i가 인지요소 A_j의 숙달을 위해 필수적인 경우 A_i와 A_j는 위계 관계를 가진다고 하고 $A_i \rightarrow A_j$로 표현한다.

2.1. BN-H 모형

BN-H 모형을 정의하기에 앞서, BN-H 모형을 엄밀하게 정의하기 위한 목적으로 K. K. Tatsuoka(1995), Leighton 외(2004) 등에서 활용된 인접행렬(adjacency matrix)을 도입한다.

인접행렬 H는 인지요소들 사이의 위계 관계를 행렬로 나타낸 것으로 K개의 인지요소가 존재할 때 $K \times K$행렬로 표현된다. 인접행렬의 각 원소 $h_{rc}(r,c=1,2,\cdots,K)$는 인지요소 A_r이 A_c의 선행 인지요소인지의 여부, 즉 위계 관계 $A_r \rightarrow A_c$가 존재하는지의 여부를 나타내며, 위계 관계 $A_r \rightarrow A_c$이 존재하면 1, 그렇지 않으면 0의 값을 가진다.

$$H=(h_{rc})_{K\times K},\ h_{rc}=\begin{cases}1, & \text{위계 관계 } A_r \rightarrow A_c \text{가 존재하는 경우}\\ 0, & \text{그렇지 않은 경우}\end{cases} \tag{3.5}$$

예를 들어, 앞의 Ⅱ장에서 살펴본 [그림 Ⅱ-1]의 (a)와 같은 인지요소 위계($A_1 \rightarrow A_2, A_1 \rightarrow A_3$)에 대한 인접행렬은 $H=\begin{bmatrix}0&1&1\\0&0&0\\0&0&0\end{bmatrix}$ 이다.

[BN-H 모형의 정의]

K개의 인지요소 $A_1, A_2, \cdots A_K$와 J개의 문항 $X_1, X_2, \cdots, X_J$으로 구성된 검사와 그에 대한 Q-행렬 $\boldsymbol{Q}=(q_{jk})\,(j=1,2,\cdots,J;\, k=1,2,\cdots,K)$이 주어져 있고, 인지요소 위계가 존재하여 이를 나타내는 인접행렬 H가 주어져있다고 하자. 이 검사를 나타내는 BN-H 모형은 다음과 같은 베이지안 네트워크 $B=(G,P)$로 구성된다.

- 그래프 구조 $G=(\boldsymbol{V},E)$의 노드 $\boldsymbol{V}$는 인지요소 변수 $\boldsymbol{A}=\{A_1, A_2, \cdots, A_K\}$와 문항 변수 $\boldsymbol{X}=\{X_1, X_2, \cdots, X_J\}$로 구성된다($\boldsymbol{V}=\boldsymbol{A}\cup\boldsymbol{X}$). 유향 변의 집합 E는 다음과 같이 구성된다.

  1) 인지요소 $A_k\,(k=1,2,\cdots,K)$와 문항 $X_j\,(j=1,2,\cdots,J)$에 대해, 문항 X_j가 인지요소 A_k를 필요로 하는 문항이면(즉, Q-행렬 $\boldsymbol{Q}$에서 $q_{jk}=1$이면) 유향 변 $A_k \rightarrow X_j$는 E의 원소가 된다.
  2) 서로 다른 두 인지요소 A_r, A_c에 대해, 인접행렬 H의 r행 c열의 원소 h_{rc}가 1이면 유향 변 $A_r \rightarrow A_c$는 E의 원소가 된다.

- $\boldsymbol{V}=\{A_1, A_2, \cdots, A_K, X_1, X_2, \cdots, X_J\}$에 대한 확률분포 P는 다음의 식 (3.6)과 같이 표현된다.

$$P(\boldsymbol{A},\boldsymbol{X})=\underbrace{\prod_{k=1}^{K}P(A_k|Pa^G(A_k))}_{\text{구조모형}}\underbrace{\prod_{j=1}^{J}P(X_j|Pa^G(X_j))}_{\text{측정모형}} \tag{3.6}$$

여기서 $P(A_k=1|\prod Pa^G(A_k)=0)=0$이다. ■

BN-H 모형의 핵심적인 특징은 크게 두 가지로 볼 수 있다. 첫째, 인지요소 위계의 구조를 베이지안 네트워크의 DAG로 나타낸다. 둘째, 인지요소 위계 가정에 의해, 어떤 인지요소 A_k의 선행 인지요소($Pa^G(A_k)$) 중 하나라도 숙달되지 않은 경우($\prod Pa^G(A_k)=0$)에는 A_k의 숙달이 불가능 하므로, 조건부 숙달 확률 $P(A_k=1|\prod Pa^G(A_k)=0)$에 대한 구조모수를 0으로 제약한다.

앞에서 다룬 예시 검사를 이번에는 BN-H 모형으로 표현해보도록 하자. 세 인지요소 A_1, A_2, A_3에 대해서는 [그림 Ⅱ-1]의 (a)와 같은 인지요소 위계($A_1 \rightarrow A_2, A_1 \rightarrow A_3$)가 가정되었다고 하자. 이 경우에 대한 BN-H 모형은 [그림 Ⅲ-4]와 같다.

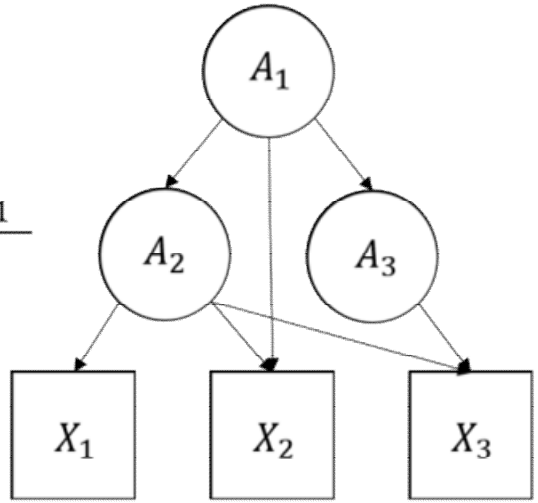
[그림 Ⅲ-4]의 BN-H 모형에서는 다음의 연쇄법칙이 성립한다.

$$\begin{aligned} P(A_1, A_2, A_3, X_1, X_2, X_3) &= P(A_1, A_2, A_3)P(X_1, X_2, X_3|A_1, A_2, A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1)P(X_1|A_2)P(X_2|A_1, A_2)P(X_3|A_2, A_3) \end{aligned} \tag{3.7}$$

앞의 BN 모형의 예시 중 인지요소 사이의 종속 관계의 구조를 현재 위계 구조와 동일하게 A_1의 숙달이 A_2와 A_3의 숙달에 영향을 미치는 것으로 설정하였었기 때문에([그림 Ⅲ-2]의 BN 모형), 그 경우에 해당하는 연쇄법칙의 식 (3.4)와 위의 식 (3.7)의 형태는 동일하다. 그러나 구조모수에 대한 제약 때문에 실제 식이 나타내는 확률분포는 다르다.

[그림 Ⅲ-2]의 종속 관계와 [그림 Ⅲ-4]의 위계 관계의 구조가 동일하게 설정되었으므로 두 경우의 DAG는 동일하며, [그림 Ⅲ-4]를 [그림 Ⅲ-2]와 비교하였을 때 인지요소 위계에 대한 가정이 추가됨에 따라 달라지는 부분은 선행 인지요소가 존재하는 인지요소들(A_2, A_3)의 CPT와 위계 관계를 가지는 한 쌍의 인지요소를 모두 측정하는 문항(X_2)의 CPT이다.

먼저, 인지요소 A_2와 A_3에 대해, 기존의 BN 모형에 각각 $\gamma_{2,0}, \gamma_{3,0}$로 표현되어 있던 모수가 BN-H 모형에서는 0으로 고정되었다. 이는 식 (3.7)에서 구조모형에 해당하는 $P(A_1, A_2, A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1)$의 확률분포에 인지요소 위계가 반영되도록 해주는 부분으로, 다음의 <표 Ⅲ-2>에 나타난 것과 같이 인지요소 위계 구조 하에서 불가능한 인지요소패턴의 확률이 0이 되도록 만들어준다.



인지요소 A_1의 CPT

$P(A_1 = 0)$	$P(A_1 = 1)$
$1-\gamma_1$	γ_1

인지요소 A_2의 CPT

$P(A_2 \mid A_1)$	$A_2 = 0$	$A_2 = 1$
$A_1 = 0$	**1**	**0**
$A_1 = 1$	$1-\gamma_{2,1}$	$\gamma_{2,1}$

인지요소 A_3의 CPT

$P(A_3 \mid A_1)$	$A_3 = 0$	$A_3 = 1$
$A_1 = 0$	**1**	**0**
$A_1 = 1$	$1-\gamma_{3,1}$	$\gamma_{3,1}$

문항 X_1의 CPT

$P(X_1 \mid A_2)$	$X_1 = 0$	$X_1 = 1$
$A_2 = 0$	$1-\pi_{1,0}$	$\pi_{1,0}$
$A_2 = 1$	$1-\pi_{1,1}$	$\pi_{1,1}$

문항 X_2의 CPT

$P(X_2 \mid A_1, A_2)$	$X_2 = 0$	$X_2 = 1$
$A_1 = 0, A_2 = 0$	$1-\pi_{2,(0,0)}$	$\pi_{2,(0,0)}$
$A_1 = 1, A_2 = 0$	$1-\pi_{2,(1,0)}$	$\pi_{2,(1,0)}$
$A_1 = 1, A_2 = 1$	$1-\pi_{2,(1,1)}$	$\pi_{2,(1,1)}$

문항 X_3의 CPT

$P(X_3 \mid A_2, A_3)$	$X_3 = 0$	$X_3 = 1$
$A_2 = 0, A_3 = 0$	$1-\pi_{3,(0,0)}$	$\pi_{3,(0,0)}$
$A_2 = 0, A_3 = 1$	$1-\pi_{3,(0,1)}$	$\pi_{3,(0,1)}$
$A_2 = 1, A_3 = 0$	$1-\pi_{3,(1,0)}$	$\pi_{3,(1,0)}$
$A_2 = 1, A_3 = 1$	$1-\pi_{3,(1,1)}$	$\pi_{3,(1,1)}$

[그림 Ⅲ-4] 예시 검사에 대한 BN-H 모형

<표 Ⅲ-2> 예시 검사에 대한 BN-H 모형 하에서의 $P(A_1, A_2, A_3)$

A_1	A_2	A_3	$P(A_1, A_2, A_3) = P(A_1)P(A_2\|A_1)P(A_3\|A_1)$
0	0	0	$(1-\gamma_1)\times 1\times 1$
0	0	1	$(1-\gamma_1)\times 1\times 0 = 0$
0	1	0	$(1-\gamma_1)\times 0\times 1 = 0$
0	1	1	$(1-\gamma_1)\times 0\times 0 = 0$
1	0	0	$\gamma_1\times(1-\gamma_{2,1})\times(1-\gamma_{3,1})$
1	0	1	$\gamma_1\times(1-\gamma_{2,1})\times\gamma_{3,1}$
1	1	0	$\gamma_1\times\gamma_{2,1}\times(1-\gamma_{3,1})$
1	1	1	$\gamma_1\times\gamma_{2,1}\times\gamma_{3,1}$

또한, [그림 Ⅲ-4]에서 확인할 수 있듯이, BN-H 모형의 구조모수의 수는 인지요소의 수와 동일해진다. 구조가 복잡해지더라도 각 인지요소에 대해서 설정되는 구조모수는 $\gamma_{k,1} = P(A_k = 1|\prod Pa^G(A_k) = 1)$로 유일하다.

다음으로 [그림 Ⅲ-4]에 제시된 BN-H 모형의 문항 X_2의 CPT를 [그림 Ⅲ-2]에 제시된 BN 모형의 CPT와 비교하여 살펴보면, $A_1 \rightarrow A_2$의 위계 관계로 인해 $(A_1 = 0, A_2 = 1)$은 불가능한 인지요소패턴이므로 그에 대한 문항모수를 나타내는 행이 삭제된 것을 확인할 수 있다. 즉, 위계 관계를 가지는 한 쌍의 인지요소를 모두 측정하는 문항의 경우, BN-H 모형에서는 그 위계 관계를 반영하여 문항모수의 수가 줄어들게 된다.

2.2. BN-H 모형과 기존 모형들의 관계

이 절에서는 BN-H 모형과 앞에서 다뤘던 기존의 인지요소 위계를 고려한 인지진단모형인 DINA-H 모형, DINO-H 모형, HDCM의 관계를 구조모형과 측정모형으로 나누어 살펴보고자 한다. 이를 위해 앞에서 다루었던 예시 검사에 대한 모수를 함께 비교해본다. 예시 검사에 대한 BN-H 모형의 모수는 <표 Ⅲ-3>과 같다.

<표 Ⅲ-3> 예시 검사에 대한 BN-H 모형의 모수

c	$\boldsymbol{\alpha}_c$	$P(\boldsymbol{\alpha}_c)$ (구조모수 λ)	문항 1	문항 2	문항 3
1	(0,0,0)	$1-\gamma_1$	$\pi_{1,0}$	$\pi_{2,(0,0)}$	$\pi_{3,(0,0)}$
2	(1,0,0)	$\gamma_1(1-\gamma_{2,1})(1-\gamma_{3,1})$	$\pi_{1,0}$	$\pi_{2,(1,0)}$	$\pi_{3,(0,0)}$
3	(1,0,1)	$\gamma_1(1-\gamma_{2,1})\gamma_{3,1}$	$\pi_{1,0}$	$\pi_{2,(1,0)}$	$\pi_{3,(0,1)}$
4	(1,1,0)	$\gamma_1\gamma_{2,1}(1-\gamma_{3,1})$	$\pi_{1,1}$	$\pi_{2,(1,1)}$	$\pi_{3,(1,0)}$
5	(1,1,1)	$\gamma_1\gamma_{2,1}\gamma_{3,1}$	$\pi_{1,1}$	$\pi_{2,(1,1)}$	$\pi_{3,(1,1)}$

구조모형의 경우, BN-H 모형이 나머지 DINA-H 모형, DINO-H 모형, HDCM보다 대체로 더 간명한 모형이다. DINA-H 모형, DINO-H 모형, HDCM은 모두 인지요소패턴 당 하나씩의 구조모수를 설정하므로, 가능한 인지요소패턴의 수가 R일 때 세 모형의 구조모수의 수는 $R-1$이었다. 반면, BN-H 모형은 인지요소 위계에 따른 제약으로 인해 인지요소 하나 당 하나의 모수가 설정되므로, 인지요소의 수가 K일 때 구조모수의 수도 K이다. K가 3 이상일 때 $K<R-1$이 성립하므로,[6] BN-H의 구조모형이 다른 모형들의 구조모형보다 대체로 더 간명하다고 볼 수 있다. 예시 검사의 경우에서 DINA-H 모형, DINO-H 모형, HDCM의 모수는 4개였던 반면, BN-H 모형의 모수는 3개이다(<표 Ⅱ-7>, <표 Ⅱ-8>, <표 Ⅱ-9>, <표 Ⅲ-3> 참고).

6) $K=2$인 경우, 두 인지요소 A_1, A_2에 대해 위계 관계 $A_1 \rightarrow A_2$가 존재하는 경우, 가능한 인지요소패턴은 $(0,0),(1,0),(1,1)$로 $R=3$이므로 $K=R-1$이다.

측정모형의 경우, BN-H 모형의 측정모형이 DINA-H, DINO-H, HDCM의 측정모형을 모두 표현할 수 있으므로 대체로 더 일반적이다. BN-H 모형에서 DINA-H 모형의 측정모형을 나타내고자 한다면, DINA-H 모형의 비보상적 가정을 나타내기 위해 문항 i의 해결에 필요한 인지요소들 중 하나라도 숙달하지 않은 경우에 문항을 맞힐 확률에 대한 모수들은 모두 동일하게 g_i로 설정하고, 문항 i의 해결에 필요한 인지요소들을 모두 숙달한 경우에 문항을 맞힐 확률에 대한 문항모수를 $1-s_i$로 설정하면 된다. 예를 들어, 예시 검사에 대한 BN-H 모형에서

$$\pi_{1,0} = g_1$$
$$\pi_{1,1} = 1-s_1$$
$$\pi_{2,(0,0)} = \pi_{2,(1,0)} = g_2$$
$$\pi_{2,(1,1)} = 1-s_2$$
$$\pi_{3,(0,0)} = \pi_{3,(0,1)} = \pi_{3,(1,0)} = g_3$$
$$\pi_{3,(1,1)} = 1-s_3$$

와 같이 제약하면 DINA-H 모형의 측정모형을 표현할 수 있게 된다 (<표 Ⅱ-7>, <표 Ⅲ-3> 참고).

HDCM의 경우 BN-H 모형과 문항모수의 수가 동일하며, BN-H 모형의 측정모형의 모수들에 대한 재매개화 과정을 거쳐 HDCM의 측정모형을 표현할 수 있다. 예시 검사에 대한 BN-H 모형에서

$$\pi_{1,0} = \lambda_{1,0}$$
$$\pi_{1,1} = \lambda_{1,0} + \lambda_{1,1,(2)}$$
$$\pi_{2,(0,0)} = \lambda_{2,0}$$
$$\pi_{2,(1,0)} = \lambda_{2,0} + \lambda_{2,1,(1)}$$
$$\pi_{2,(1,1)} = \lambda_{2,0} + \lambda_{2,1,(1)} + \lambda_{2,2,2(1)}$$
$$\pi_{3,(0,0)} = \lambda_{3,0}$$
$$\pi_{3,(0,1)} = \lambda_{3,0} + \lambda_{3,1,(3)}$$

$$\pi_{3,(1,0)} = \lambda_{3,0} + \lambda_{3,1,(2)}$$
$$\pi_{3,(1,1)} = \lambda_{3,0} + \lambda_{3,1,(2)} + \lambda_{3,2,2(3)}$$

와 같이 재매개화하면 HDCM의 측정모형을 표현할 수 있게 된다(<표 Ⅱ-9>, <표 Ⅲ-3> 참고).

2.3. BN-H 모형을 적용한 인지진단평가

BN-H 모형을 적용한 인지진단평가는 다음과 같은 과정에 따라 이루어질 수 있다.

① 검사의 인지요소와 인지요소 위계를 설정하고 인지요소들의 숙달을 측정하는 문항을 개발하여 검사를 제작한다. 인지요소 위계는 검사 영역의 내용 전문가들이 설정하거나, 동일한 인지요소들을 측정하는 다른 검사에 대해 수집된 자료가 있는 경우 그 자료를 이용하여 추정할 수 있다(위계를 추정하는 방법은 Ⅳ장에서 다룬다).

② 검사에 대한 자료를 수집하고 난 뒤, 수집된 자료를 이용하여 Q-행렬과 인지요소 위계에 대한 검증의 과정을 거친다. Q-행렬에 대한 검증은 기존에 개발된 Q-행렬 타당화 방법들을 이용하여 가능하다. 위계에 대한 검증은 이후 Ⅳ장에서 다루게 될 위계 추정의 방법을 이용하여 이루어질 수 있다. 인지요소 위계에 대한 타당성을 검증한 이후에 BN-H 모형을 적용하는 것이 바람직하다.

③ 인지요소 위계 구조에 따라 BN-H 모형의 DAG를 설정하고, CPT의 모수들을 추정한다. 모수 추정은 2.3절에서 설명한 베이지안 네트워크의 모수 학습 과정을 통해 이루어질 수 있다.

④ BN-H 모형의 DAG와 CPT의 모수들의 값이 모두 설정되고 나면, 각 피험자 i의 응답 $\boldsymbol{x}_i$을 BN-H 모형의 문항 변수에 증거(evidence)로 주입한 뒤 2.2절에서 설명한 베이지안 네트워크의 추론 과정을 통해 피험자 i의 인지요소패턴을 추정한다.

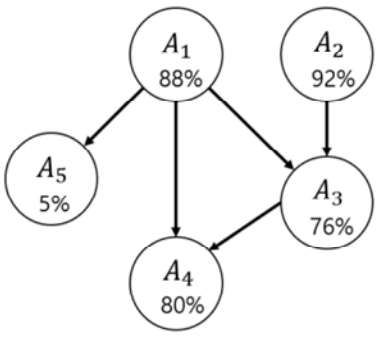
⑤ 각 피험자에게 문항응답 결과와 인지요소패턴 추정의 결과를 제공한다. 인지요소 위계를 고려하지 않은 모형을 활용한 인지진단평가에서는 개별 인지요소들의 숙달 상태만 결과로 제공된다. 반면, 인지요소 위계를 고려한 BN-H 모형이 적용된 인지진단평가에서는 다음의 [그림 Ⅲ-5]와 같이 인지요소들 사이의 관계를 결과표에 함께 제시함으로써, 학생들이 자신이 학습 과정상 어느 내용까지 숙달하였으며 다음으로 어떤 내용을 학습해야 하는지에 관한 정보를 제공받도록 할 수 있다.

문항 응답 결과 (정답=1, 오답=0)

문항1	문항2	문항3	문항4	문항5	문항6	문항7	문항8	문항9	문항10	문항11	문항12	문항13
0	0	1	1	1	1	1	0	1	1	1	1	1

인지요소 별 숙달 여부

A_1	분모가 같은 두 분수를 빼기	숙달
A_2	자연수 부분과 분수 부분을 분리하기	숙달
A_3	자연수 부분에서 분수 부분으로 1만큼 빌려오기	숙달
A_4	자연수를 분수로 변환하기	숙달
A_5	공통분모 찾기	비숙달



[그림 Ⅲ-5] BN-H 모형을 적용한 인지진단평가의 결과표의 예

3. 요약 및 논의

이 장에서는 인지요소 위계가 가정되지 않은 경우에 대한 일반적인 베이지안 네트워크 기반 인지진단모형을 BN 모형이라고 칭하여 그 정의와 특징을 살펴보고 BN 모형에 인지요소 위계에 대한 가정이 추가된 모형으로 BN-H 모형을 제시하였다. BN-H 모형의 수학적 정의에 포함된 두 가지 핵심적인 특징을 반영하면 BN-H 모형의 설명적 정의를 '인지요소 위계를 베이지안 네트워크의 DAG로 나타내고, 주어진 인지요소 위계에 따라 어떤 인지요소의 선행 인지요소들 중 적어도 하나를 숙달하지 않은 경우에 그 인지요소를 숙달할 확률을 나타내는 구조모수를 0으로 제약하는 모형'으로 내릴 수 있다.

기존의 인지요소 위계를 고려한 인지진단모형들과 BN-H 모형과의 관계를 구조모형과 측정모형의 관점에서 각각 살펴본 결과, 구조모형의 경우 BN-H 모형이 기존 모형들에 비해 더 간명한 모형임이 확인되었으며, 측정모형의 경우 BN-H 모형이 기존 모형들의 측정모형을 표현할 수 있는 더 일반적인 모형임이 밝혀졌다. 인지요소의 수가 늘어날수록 기존 모형들의 구조모수의 수는 기하급수적으로 늘어나는 반면 BN-H 모형의 구조모수의 수는 인지요소의 수와 동일하므로 선형적으로 늘어난다. 따라서 검사가 측정하고자 하는 인지요소의 수가 많아질수록 BN-H 모형이 기존 모형에 비해 구조모수의 수를 더 많이 줄일 수 있다. 그리고 일반적으로 추정해야 하는 모수의 수가 많아질수록 더 큰 표본이 필요하다는 점(Jackson, 2003)을 함께 고려해보면, 인지요소의 수가 많은 경우에 BN-H 모형의 구조모형을 가정하면 측정모형이 동일한 다른 모형들에 비해 더 적은 표본으로 동일한 수준의 추정 정확도를 얻을 수 있을 것이라 예상해볼 수 있다. 실제로 그러한 결과가 도출되는지에 대해서는 추후 모의실험을 통해 검증해볼 필요가 있을 것으로 보인다.

한편, BN-H 모형을 적용한 인지진단평가의 절차에서 확인하였듯이 인지요소 위계가 가정된 경우에는 인지요소 위계를 평가의 결과표에 함께 나타냄으로써 학생들이 학습 순서나 과정이 포함된 보다 구체적인 피

드백을 제공받을 수 있다는 장점이 있다. 또한, BN-H 모형에서 개별 학생이 아닌 전체 학생들의 응답 자료를 이용하여 추정된 베이지안 네트워크의 CPT 모수들은 학생 전체에 대한 인지요소 숙달 분포를 나타내므로, 이러한 정보가 교사에게 피드백으로 제공된다면 이는 교사의 추후 교수-학습 개선을 위한 자료로도 활용될 수 있을 것으로 보인다.

Ⅳ. SEM-H: BN-H 모형을 적용한 위계 추정 방법

인지요소 위계를 고려한 인지진단모형은 인지요소들이 이루는 참인 위계 관계는 이미 알려진 것으로 가정하고 주어진 인지요소 위계를 모델링한 인지진단모형이다. 따라서 인지요소 위계를 고려하여 인지진단검사를 제작하거나 검사 자료를 분석하기 위해서는 인지요소 위계를 사전에 설정해야 한다. 처음에는 주로 검사 영역의 내용 전문가들이 관련 개념이나 이론 등에 대한 질적 검토를 바탕으로 인지요소 위계를 설정하는 것이 일반적이나, 실제 검사 응답 자료가 수집된 이후에는 그 자료를 이용하여 역으로 사전에 설정된 인지요소 위계와 실제 응답이 부합하는지, 즉 사전에 인지요소 위계가 올바르게 설정되었는지에 대한 검증이 이루어져야 한다. 본 연구에서는 이처럼 실제 응답 자료를 이용하여 인지요소 위계를 확인하는 과정을 '위계 추정'이라 부르기로 한다.

앞의 Ⅲ장에서는 베이지안 네트워크를 이용하여 인지요소 위계를 모델링한 인지진단모형으로 BN-H 모형을 제안하였다. 이와 동일한 맥락에서, 이 장의 초점인 위계 추정을 베이지안 네트워크의 관점에서 보면 이는 베이지안 네트워크의 구조를 찾는 과정인 구조 학습의 과정과 유사하므로 Ⅱ장에서 다루었던 베이지안 네트워크의 구조 학습 방법을 위계 추정의 방법으로 활용하는 것을 고려해볼 수 있다. 그러나 앞의 Ⅱ장에서 설명했듯이 인지진단모형으로서 베이지안 네트워크가 활용되는 현 맥락에서는 BN 모형의 변수 중 관찰 불가능한 잠재변수인 인지요소 변수에 대한 자료를 얻을 수 없으므로, 완전 자료를 가정하고 발전된 구조 학습 방법들은 인지요소 위계 추정에 적용될 수 없다. 따라서 본 연구에서는 불완전 자료에 대한 베이지안 네트워크의 구조 학습 방법 중 하나인 SEM 알고리즘(Friedman, 1997)을 위계 추정에 적용하는 것을 시도하였으며, 구체적으로 인지요소 위계가 존재하는 맥락을 반영하기 위해 BN-H 모형의 개념을 적용한 위계 추정 방법인 SEM-H 알고리즘을 제안하고 기존에 개발된 방법과의 비교를 통해 SEM-H 방법의 위계 추정 능력을 검증하고자 하였다.

이하에서는 우선 SEM 알고리즘과 기존에 제안되었던 SEM 알고리즘 기반 위계 추정 방법인 COMMAND 방법을 검토하고 그 방법의 한계를 확인한다. 이어서, 기존 방법의 한계를 극복하도록 수정된 SEM-H 알고리즘을 제안하고, 모의실험을 통해 본 연구에서 제안하는 SEM-H 방법이 기존의 COMMAND 방법을 실제로 개선하는지를 확인한다. 마지막으로 실제 수학 검사 자료를 활용하여 SEM-H 방법을 적용한 위계 추정이 구체적으로 어떻게 이루어지는지를 살펴본다.

1. SEM 알고리즘과 위계 추정

1.1. SEM 알고리즘

완전 자료가 가정된 경우에 베이지안 네트워크(BN)의 구조 학습의 기본 아이디어는 탐색 공간 내의 후보 구조마다 각 구조가 자료와 얼마나 적합한지를 나타내는 점수를 각각 계산하여 그 값이 최대가 되는 구조를 찾는 것이었다. 이를 위해 구조의 자료 적합성을 나타내는 점수 함수(scoring function)가 정의되고, 대규모 탐색 공간에서 최대 점수를 갖는 구조를 효율적으로 찾기 위한 탐색 절차(search procedure)를 나타내는 다양한 알고리즘들이 제안되었다. 불완전 자료가 가정된 경우에는 BN의 구조 학습의 두 가지 핵심 요소인 점수 함수와 탐색 절차가 다소 복잡해지는데, Friedman(1998)은 EM 알고리즘을 변형한 방식을 이용하여 불완전 자료에서의 탐색을 완전 자료에서의 탐색으로 환원시킴으로써 구조를 보다 효율적으로 탐색하는 구조 학습 방법인 SEM 알고리즘을 제안하였다. SEM 알고리즘의 절차는 다음의 <표 Ⅳ-1>에 제시되어 있다.

<표 Ⅳ-1> 불완전 자료에서의 BN 구조 학습을 위한 SEM 알고리즘 (Friedman, 1998; Koller & Friedman, 2009)

SEM 알고리즘	
입력 { G_0	// 변수 $X_1, X_2, ..., X_n$에 대한 초기 BN 구조
Θ_0	// G_0에 대한 모수들의 초기 집합
D	// $X_1, X_2, ..., X_n$ 중 일부 변수에 대해 수집된 불완전 자료
}	
for $s = 0, 1, 2, ...$ until convergence {	
$\Theta_s^* \leftarrow$ Parametric-EM(G_s, Θ_s, D)	// E-step
$D^*_{G_s, \Theta_s^*} \leftarrow$ Inference(G_s, Θ_s^*, D)	// 완전 자료 생성
$G_{s+1} \leftarrow$ Learn-Structure($D^*_{G_s, \Theta_s^*}$)	// M-step
$\Theta_{s+1} \leftarrow$ Estimate-Parameters($D^*_{G_s, \Theta_s^*}, G_{s+1}$)	
}	
출력 { G_s, Θ_s	
}	

SEM 알고리즘은 초기 BN 구조인 G_0와 구조 G_0에 대한 초기 모수치 Θ_0를 가정하고 시작되며, 불완전 자료 D를 완전 자료 D^*로 변환하고 D^*를 토대로 구조와 모수를 차례로 갱신하는 EM 사이클을 반복하는 과정으로 이루어진다. s번째 EM 사이클의 E-step에서 이루어지는 Parametric-EM(G_s, Θ_s, D)은 현재 구조 G_s와 현재 모수치 Θ_s를 바탕으로 불완전 자료 D에 모수적 EM 추정방식을 적용하여 Θ_s를 Θ_s^*로 갱신하는 과정을 뜻한다. 모수가 Θ_s^*로 갱신되면, 현재 구조 G_s에 대한 모수로 Θ_s^*를 부여하여 BN을 구성한 뒤, 관찰된 변수들에 대한 증거(evidence)로서 불완전 자료 D를 주입하여 관찰되지 않은 변수들에 대한 추정값을 도출하는 BN의 추론 과정(Inference(G_s, Θ_s^*, D))을 거쳐 완전 자료 $D^*_{G_s, \Theta_s^*}$를 생성한다. 이렇게 생성된 완전 자료 $D^*_{G_s, \Theta_s^*}$를 이용하여 구조를 갱신하고 갱신된 구조에 대한 모수를 추정하는 과정이 EM 사이클의 M-step이다. M-step의 Learn-Structure($D^*_{G_s, \Theta_s^*}$)은 완전 자료 $D^*_{G_s, \Theta_s^*}$에 대해 BN의 구조 학습을 진행하는 과정으로, hill-climbing 알고리즘을 통해

BIC 점수를 증가시키는 구조를 찾아서 현재 구조를 G_{s+1}로 갱신한다. M-step의 Estimate-Parameters($D^*_{G_s,\theta_s^*}, G_{s+1}$)은 완전 자료 $D^*_{G_s,\theta_s^*}$를 이용하여 앞서 갱신된 구조 G_{s+1}에 대한 모수 Θ_{s+1}를 추정하는 과정이다. 이렇게 s번째 EM 사이클의 M-step의 결과로 도출된 G_{s+1}과 Θ_{s+1}가 다음 $s+1$번째 EM 사이클의 E-step이 시작될 때의 현재 구조와 현재 모수치가 되는 방식으로 EM 사이클이 반복된다. 직전 EM 사이클에서의 BIC 점수와 비교했을 때 점수가 더 이상 증가되지 않으면 (convergence) EM 사이클이 종료되고 최종적인 BN 구조와 그 구조에 대한 모수 추정치가 출력된다.

SEM 알고리즘은 불완전 자료에서의 BIC 점수 계산이 복잡하기 때문에 불완전 자료를 보강(augmentation)하여 완전 자료로 바꾼 다음 완전 자료에서 BIC 점수가 가장 많이 올라가는 구조를 찾는 방식을 취한다. 이와 같은 방식이 가능한 이유는 다음의 [정리]에 의해 어느 정도 설명된다(Friedman, 1998; Koller & Friedman, 2009).

[정리] G_0를 어떤 BN 구조라 하고, 불완전 자료 D가 주어져 있을 때 G_0에 대한 MAP 추정치를 Θ_0*라 하자. 그러면 임의의 BN 구조 G에 대해 다음이 성립한다:

$$\begin{aligned} &Score_{BIC}(G: D^*_{G_0,\theta_0^*}) - Score_{BIC}(G_0 : D^*_{G_0,\theta_0^*}) \\ &\le Score_{BIC}(G: D) - Score_{BIC}(G_0 : D) \end{aligned} \tag{4.1}$$

($D^*_{G_0,\theta_0^*}$: Inference(G_0, Θ_0^*, D)에 의해 생성된 완전 자료;
$Score_{BIC}$(G : D): 자료 D에 대한 구조 G의 BIC 점수)

이 정리는 완전 자료에 대해 BIC 점수가 증가하는 구조를 찾으면, 그 구조는 불완전 자료에 대해서도 BIC 점수가 증가하는 구조가 된다는 점을 보장한다. 하지만 최종 도출된 구조가 BIC 점수에 대한 극댓값(local maximum)을 갖는 구조라는 점까지는 보장해주지 않는다(Heckerman, Geiger, & Chickering, 1995).

1.2. COMMAND 알고리즘

SEM 알고리즘은 불완전 자료가 주어진 경우에 구조를 찾아주는 방법이므로, 문항(관찰변수)에 대한 응답 자료만 주어져 있는 상황에서 인지요소(관찰될 수 없는 잠재변수)들의 관계를 추정하고자 할 때 활용될 수 있는 구조 학습 방법이다. Chen, González-Brenes와 Tian(2016)은 SEM 알고리즘을 적용하여 지능형 튜터링 시스템(intelligent tutoring system)의 맥락에서 학습자가 숙달해야 하는 기능들의 선행-후행 관계를 나타내는 구조(skill prerequisite structure)를 탐색하는 방법인 COMMAND 알고리즘[7]을 제안하였다. COMMAND 알고리즘을 인지진단평가의 맥락에서 살펴보면, 학습자가 숙달해야 하는 기능은 '인지요소'에 해당하며, 기능들의 선행-후행 관계는 '위계 관계', 그 관계를 나타내는 구조는 '인지요소 위계 구조'라 할 수 있으므로 COMMAND 알고리즘을 위계 추정의 한 방법으로 볼 수 있다. Chen 외(2016)의 COMMAND 알고리즘의 절차는 <표 Ⅳ-2>에 제시되어 있다.

COMMAND 알고리즘은 세 단계로 구성되어 있다. 첫 단계는 Q-행렬을 이용하여 초기 BN 구조를 설정하는 단계이며, 두 번째 단계는 주어진 응답 자료에 대해 SEM 알고리즘을 적용하여 BN의 구조를 학습하는 단계이다. 그리고 세 번째 단계는 SEM 알고리즘을 이용한 구조 학습으로부터 도출된 구조와 동형인 구조들로 이루어진 동형 클래스(equivalence class)로부터 주어진 응답 자료가 나타내는 기능들의 선행-후행 관계가 가장 잘 반영된 특정 구조를 선택하는 단계로, 이 단계에서 일종의 위계 추정이 이루어진다고 볼 수 있다.

이하에서는 Chen 외(2016)에서 제시한 COMMAND 알고리즘의 각 단계를 조금 더 구체적으로 살펴본다. 용어의 통일을 위해 인지진단평가 맥락의 용어인 '인지요소'와 '위계 관계', '인지요소 위계 구조'라는 표현을 사용하여 COMMAND 알고리즘의 절차를 설명하도록 하겠다.

---

7) Chen 외(2016)에서 알고리즘의 명칭을 COMMAND라고 정한 이유는 밝혀져 있지 않다.

<표 Ⅳ-2> 위계 추정을 위한 COMMAND 알고리즘(Chen et al., 2016)

COMMAND 알고리즘

입력 { $\boldsymbol{Q}$ // Q-행렬
$\boldsymbol{D}$ } // 문항에 대한 응답 자료
$G_0 \leftarrow$ Initialize($\boldsymbol{S}, \boldsymbol{Q}$) // Q-행렬을 반영한 BN 구조 생성 ($\boldsymbol{S}$: skill들의 집합)

(1) 초기 설정

while 종료 기준에 도달하지 않음 **{**
$\boldsymbol{\Theta}_i^* \leftarrow$ Parametric-EM($G_i, \boldsymbol{D}$)
$\boldsymbol{D}_i^* \leftarrow$ Inference($G_i, \boldsymbol{\Theta}_i^*, \boldsymbol{D}$)
$(G_{i+1}, \boldsymbol{\Theta}_{i+1}) \leftarrow$ BN-Learning($G_i, \boldsymbol{D}_i^*$)
$i \leftarrow i+1$
}

(2) SEM

$RE \leftarrow$ Find-reversible-edges(G_i)
$EC \leftarrow$ Enumerate-equivalent-DAGs(G_i)
$DE \leftarrow \varnothing$
for all reversible edges $'S_i - S_j' \in RE$ **{**
$ratio \leftarrow \frac{P(S_j=0 \mid S_i=0)}{P(S_i=0 \mid S_j=0)}$
if $ratio \geq 1$ **then {**
$ratio^* = ratio$
$DE \leftarrow DE \cup \{S_i \rightarrow S_j\}$ **}**
else {
$ratio^* = 1/ratio$
$DE \leftarrow DE \cup \{S_i \leftarrow S_j\}$ **}**
}
sort(DE) by $ratio^*$에 대한 내림차순
while $DE \neq \varnothing$ **{**
$e \leftarrow$ Dequeue(DE)
if $\exists G \in EC,\ e \in G$ **then {**
$\forall G \in EC,$ remove G from EC if $e \notin G$ **}**
}

(3) 동형 구조들에 대한 판별

출력 { EC **}**

(1) 초기 설정

어떤 검사가 측정하는 인지요소들의 집합을 $\boldsymbol{S}$, 각 검사 문항이 어떤 기능을 측정하는지를 나타내는 Q-행렬을 $\boldsymbol{Q}$, 검사 문항에 대한 학생들의 응답 자료를 $\boldsymbol{D}$라고 하자. COMMAND 알고리즘의 첫 단계는 구조 학습 시작 전에 BN의 초기 구조 G_0를 설정하는 단계로($G_0 \leftarrow \mathsf{Initialize}(\boldsymbol{S}, \boldsymbol{Q})$), 주어진 Q-행렬($\boldsymbol{Q}$)을 따라서 각 문항과 그 문항이 측정하는 인지요소들을 유향 변으로 연결함으로써 초기 BN 구조가 설정된다(Chen et al., 2016).

(2) SEM

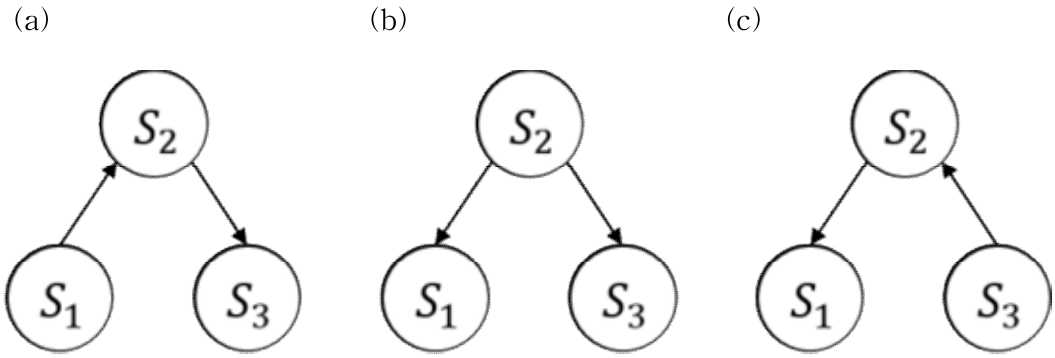
두 번째 단계는 주어진 응답 자료를 가장 잘 설명하는 구조, 즉 BIC 점수가 가장 높은 구조를 탐색하는 단계로, 앞의 1.1절에서 살펴본 SEM 알고리즘이 적용되는 단계이다. <표 Ⅳ-2>에 제시된 알고리즘에서는 드러나지 않지만, COMMAND 알고리즘은 SEM의 M-step에서 구조를 갱신할 때($\mathsf{BN\text{-}Learning}(G_i, \boldsymbol{D}_i^*)$) Q-행렬을 따르는 구조들만으로 이루어진 탐색 공간을 가정한다(Chen et al., 2016). COMMAND 알고리즘의 두 번째 단계의 결과로 구조 G_i와 그에 대한 모수 $\boldsymbol{\Theta}_i$가 도출된다.

(3) 동형 구조들에 대한 판별

SEM 알고리즘의 구조 학습에서 사용되는 BIC 점수는 동형인 BN 구조에 대해서 동일한 값을 도출하는 score equivalence[8]라는 특성을 가진다. 따라서 SEM의 결과로 도출되는 것은 엄밀히 말해 구조 G_i와 동형인 구조들로 이루어진 G_i의 동형 클래스(equivalence class)이며, 그 동형 클래스에 둘 이상의 구조가 포함되어 있는 경우에는 어떤 구조를 최종 구조로 볼 것인지 결정해야 한다. 예를 들어, 3개의 인지요소 S_1, S_2, S_3에 대해 SEM 알고리즘의 결과로 $S_1 \rightarrow S_2 \rightarrow S_3$라는 구조가 도출된 경우, 실제로는 그 구조와 동형인 두 가지 구조를 더 포함한 동형 클래스 $\{S_1 \rightarrow S_2 \rightarrow S_3, S_1 \leftarrow S_2 \rightarrow S_3, S_1 \leftarrow S_2 \leftarrow S_3\}$가 도출된 것과 같다([그림 Ⅳ-1]).

---

8) score equivalence의 뜻은 2.3.2절에 제시되어 있다.



[그림 Ⅳ-1] 동형인 BN 구조들로 이루어진 동형 클래스의 예

이 세 구조는 동일한 skeleton을 가지지만 변의 방향이 적어도 하나씩 다르기 때문에 서로 다른 구조를 나타낸다. 따라서 이들 중 최종 구조를 선택하는 과정이 필요하며, 이는 동형 클래스 내의 구조에 따라 가역 가능한 변(reversible edge)들의 방향을 결정하는 문제로 볼 수 있다. [그림 Ⅳ-1]의 예에서 가역 가능한 변은 $S_1 - S_2$과 $S_2 - S_3$이다(Chen et al., 2016).

COMMAND 알고리즘은 이 가역 가능한 변들의 방향을 결정하기 위해 우리가 Ⅲ장에서 이미 다루었던 위계의 개념을 활용한다. 예를 들어 가역 가능한 변 $S_1 - S_2$의 방향을 결정하는 상황에서, 인지요소 S_1이 S_2의 선행 인지요소라면 $S_1 = 0 \Rightarrow S_2 = 0$, 즉 $P(S_2 = 0 | S_1 = 0) = 1$일 것이므로, $P(S_2 = 0 | S_1 = 0) = 1$인지 $P(S_1 = 0 | S_2 = 0) = 1$인지를 확인하면 어느 방향이 위계 관계를 나타내는지 결정할 수 있다는 것이 기본 아이디어이다. 그러나 측정의 잡음(noisy) 등으로 인해, S_1이 S_2의 선행 인지요소일지라도 $P(S_2 = 0 | S_1 = 0)$가 반드시 1인 것은 아닐 수도 있다는 점을 받아들여, 다음과 같이 $P(S_2 = 0 | S_1 = 0)$과 $P(S_1 = 0 | S_2 = 0)$의 대소를 비교하여 변의 방향을 결정하는 규칙을 적용한다(Chen et al., 2016).

> $P(S_2 = 0 | S_1 = 0) \geq P(S_1 = 0 | S_2 = 0)$이면 $S_1 \rightarrow S_2$로 결정하고,
> 그렇지 않으면 $S_2 \rightarrow S_1$으로 결정한다.

위 규칙을 적용하면 기본적으로 모든 가역 가능한 변의 방향을 결정

할 수는 있으나, SEM의 결과로 도출된 동형 클래스 내에 있는 구조 중에서 하나를 선택해야 하는 상황에서는 각 변의 방향이 독립적으로 확정될 수 없다는 문제가 발생한다. 예를 들어 [그림 Ⅳ-1]과 같은 동형 클래스가 도출된 상황에서 $S_1 \rightarrow S_2$로 변의 방향을 결정하게 되면, 세 구조 중 (a)의 구조만 유향 변 $S_1 \rightarrow S_2$를 포함하고 있으므로 $S_2 - S_3$의 방향은 (a)의 구조를 따라 $S_2 \rightarrow S_3$로 결정될 수밖에 없다. 즉, 어느 한 변의 방향을 결정하는 것이 다른 가역 가능한 변의 방향을 제약하는 경우가 발생할 수 있는 것이다(Chen et al., 2016).

이와 같은 상황을 다루기 위해 COMMAND 알고리즘은 다음과 같은 방식을 취한다. 먼저, SEM의 결과로 도출된 동형 클래스로부터 가역 가능한 변들의 집합을 설정하고($RE \leftarrow$ Find-reversible-edges(G_i)), 도출된 동형 클래스 내의 모든 구조들을 나열한 후보 구조 집합을 설정한다($EC \leftarrow$ Enumerate-equivalent-DAGs(G_i)). 다음으로, 각 가역 가능한 변 $S_i - S_j$에 대해서 $ratio = \frac{P(S_j = 0 \mid S_i = 0)}{P(S_i = 0 \mid S_j = 0)}$를 확인하여 $ratio \geq 1$이면 $S_i \rightarrow S_j$로 방향을 결정하고 $ratio^* = ratio$로 놓고, 그렇지 않으면 $S_j \rightarrow S_i$로 방향을 결정하고 $ratio^* = 1/ratio$로 놓는다(**for**문). 이 $ratio^*$는 유향 변의 위계성의 정도를 알려주는 수치라고 볼 수 있다. 모든 가역 가능한 변에 대해 그 방향과 $ratio^*$가 정해지면, 결정된 유향 변들을 그에 대응하는 $ratio^*$에 대한 내림차순으로 나열한다(sort(DE) by $ratio^*$). 마지막으로, $ratio^*$가 큰 유향 변부터 차례로 확정하면서 어떤 유향 변 $S_i \rightarrow S_j$가 확정될 때마다 그 변을 포함하지 않은 BN 구조는 후보 구조 목록(EC)에서 제외해 나간다(**while**문). 이와 같은 과정을 거치면서 후보 구조 목록에는 한 구조만 남게 되며, 그 구조를 확인하여(**출력{** EC **}**) 최종 구조로 확정한다(Chen et al., 2016).

이상에서 살펴본 바와 같이, COMMAND 알고리즘은 SEM을 적용한 구조 학습으로부터 도출된 동형 구조들을 후보 구조로 한정하고 그 후보 구조들 중에서 자료가 나타내는 위계성을 가장 잘 반영하는 구조를 선택하는 방식으로 구조를 찾는 방법이다. 그러나 이와 같은 방식은 '위계 추정'이라는 관점에서 보면 추정의 정확성을 떨어뜨릴 위험이 있다. SEM을 적용한 구조 학습의 결과로 [그림 Ⅳ-1]과 같은 동형 클래스가 도출된 상황을 다시 예로 들어보자. COMMAND 알고리즘의 세 번째 단계에서 [그림 Ⅳ-1]의 세 구조는 후보 구조 EC의 원소가 되며, 가역 가능한 두 변 $S_1 - S_2$와 $S_2 - S_3$는 RE의 원소가 된다. 이때, 만약 두 변에 대해 $ratio$를 각각 확인해본 결과 $S_1 \rightarrow S_2$와 $S_3 \rightarrow S_2$로 각 방향이 결정되고 그에 대응하는 $ratio^*$의 값이 각각 8, 4로 도출되었다고 해보자. 이 경우 COMMAND 알고리즘은 $ratio^*$의 값이 큰 $S_1 \rightarrow S_2$를 두 인지요소 S_1과 S_2의 위계 관계로 먼저 확정하고, 변 $S_1 \rightarrow S_2$를 포함하지 않은 (b)와 (c)의 구조는 후보 구조에서 제외한다. 그렇게 제외하고 나면 후보 구조 집합에는 (a)의 구조만 남게 되므로 그 구조가 최종 구조로 확정된다.

[그림 Ⅳ-1]의 (a)의 구조에서 S_2와 S_3와 관계는 $S_3 \rightarrow S_2$가 아닌 $S_2 \rightarrow S_3$이다. 그러나 앞에서 변 $S_2 - S_3$에 대해 $ratio$를 확인하여 $S_3 \rightarrow S_2$로 관계의 방향이 결정되었다는 점은 S_2와 S_3 사이에 위계 관계가 존재한다면 그 관계는 $S_3 \rightarrow S_2$라는 점을 의미하므로, 위계 구조를 추정하는 입장에서는 $S_3 \rightarrow S_2$로 방향이 설정되는 것이 더 적절할 수 있다. 만약 실제로 참인 위계 구조가 $S_1 \rightarrow S_2 \leftarrow S_3$인데 SEM의 결과로 [그림 Ⅳ-1]과 같은 동형 구조 클래스가 도출되었다면, COMMAND 알고리즘은 참인 위계 구조를 정확히 추정할 수 없게 되는 것이다. 위계 추정의 관점에서는 BIC 점수가 조금 낮은 구조일지라도 그 구조가 각 위계 관계를 더 잘 반영하고 있다면 그 구조를 선택하는 것이 더 정확한 위계 추정이 될 수 있다.

2. SEM-H

앞서 살펴본 COMMAND 알고리즘의 위계 추정으로서의 한계를 개선하기 위해, 본 연구에서는 '위계 관계'의 판정에 보다 우선순위를 두고 SEM 알고리즘을 활용하는 위계 추정 방법인 SEM-H 알고리즘을 제안한다. 구조 학습의 종료 이후에 위계 관계에 대한 판정을 하는 COMMAND와 달리, SEM-H는 SEM을 이용한 구조 학습의 매 사이클마다 위계 관계를 반영하여 후보 구조를 제한하고 구조 학습을 진행한 뒤, 도출된 구조에서 위계 관계에 대한 최종 판정을 내리는 방식으로 이루어진다. 본 연구에서 제안하는 SEM-H 알고리즘의 절차는 <표 Ⅳ-3>에 제시되어 있다. 이하에서는 <표 Ⅳ-3>에 제시된 SEM-H 알고리즘의 절차를 차례로 살펴보도록 하겠다.

(1) 초기 설정

SEM-H 알고리즘 첫 단계에서는 COMMAND 알고리즘과 마찬가지로 Q-행렬을 반영한 초기 BN 구조를 설정하고($G_0 \leftarrow$ Initialize-DAG($\boldsymbol{Q}$)), 그 구조에 대한 초기 모수를 설정한다($\boldsymbol{\Theta}_0 \leftarrow$Initialize-Parameters($G_0$)). 이와 더불어, SEM-H 알고리즘에서는 이후 SEM 구조 학습의 제약에 사용될 화이트리스트와 블랙리스트를 설정한다($WL \leftarrow$Set-Whitelist($\boldsymbol{Q}$), $BL_0 \leftarrow$Set-Blacklist($\boldsymbol{Q}$)). 여기서 화이트리스트는 구조 학습 시 필수적으로 구조에 포함되어야 하는 유향 변들의 집합이며, 블랙리스트는 구조 학습 시 배제되어야 하는 유향 변들의 집합을 의미한다. 초기 설정 단계에서 화이트리스트는 Q-행렬에서 1로 표시된 인지요소-문항 관계를 포함해야 하는데, 구체적으로 어떤 인지요소 A_k를 문항 I_j가 측정하는 경우라면 A_k에서 I_j로 향하는 유향 변 $A_k \rightarrow I_j$을 나타내는 순서쌍 (A_k, I_j)가 화이트리스트의 원소가 된다. 반대로 블랙리스트는 Q-행렬에 0으로 표시된 인지요소-문항 관계를 포함하면 되는데, 만약 문항 $I_{j'}$이 인지요소 $A_{k'}$을 측정하지 않는 경우에는 우선 유향 변 $A_{k'} \rightarrow I_{j'}$을 나타내는 순서쌍 $(A_{k'}, I_{j'})$이 블랙리스트의 원소가 되며, 반대 방향에 해당하는 유향 변인

<표 Ⅳ-3> 본 연구에서 제안하는 위계 추정을 위한 SEM-H 알고리즘

SEM-H 알고리즘

인지요소 변수 집합 $\boldsymbol{A}=\{A_1, A_2, \cdots, A_K\}$; 문항 변수 집합 $\boldsymbol{I}=\{I_1, I_2, \cdots, I_J\}$;

입력 {$\boldsymbol{Q}$ // Q-행렬($J \times K$ 행렬)

$G_0 \leftarrow$ Initialize-DAG($\boldsymbol{Q}$) // Q-행렬을 반영한 BN 초기 구조

$WL \leftarrow$ Set-Whitelist($\boldsymbol{Q}$) // 필수적인 유향 변(순서쌍)들의 집합

$BL_0 \leftarrow$ Set-Blacklist($\boldsymbol{Q}$) // 배제해야 할 유향 변(순서쌍)들의 집합

$\boldsymbol{\Theta}_0 \leftarrow$ Initialize-Parameters(G_0) // G_0에 대한 모수들의 초기 집합

$\boldsymbol{D}$ } // 문항 $I_1, I_2, \cdots, I_J$에 대한 응답 자료

(1) 초기 설정

for $s = 0, 1, 2, \ldots$ **until convergence** {

$\boldsymbol{\Theta}_s^* \leftarrow$ Parametric-EM($G_s, \boldsymbol{\Theta}_s, \boldsymbol{D}$)

(2) SEM의 E-step

$\boldsymbol{D}^*_{G_s, \boldsymbol{\Theta}_s^*} \leftarrow$ Inference($G_s, \boldsymbol{\Theta}_s^*, \boldsymbol{D}$)

$CE \leftarrow \{(i,j) | 1 \le i, j \le K,\ i \ne j\}$ // 가능한 모든 인지요소 관계(순서쌍)의 집합

$HE \leftarrow \varnothing$

$BL \leftarrow \varnothing$

while $CE \ne \varnothing$ {

$p \leftarrow$ Dequeue(CE)

$p' \leftarrow (p[2], p[1])$

if $P(A_{p[2]}=1 | A_{p[1]}=0) < P(A_{p'[2]}=1 | A_{p'[1]}=0)$ **then** {

$HE \leftarrow HE \cup \{p\}$

$CE \leftarrow CE - \{p'\}$

$BL \leftarrow BL \cup \{p'\}$ }

else {

$HE \leftarrow HE \cup \{p'\}$

$CE \leftarrow CE - \{p\}$

$BL \leftarrow BL \cup \{p\}$ } //end if

} //end while

$BL \leftarrow BL \cup BL_0$

(3) 위계 관계 후보 판별

$G_{s+1} \leftarrow$ Learn-Structure($\boldsymbol{D}^*_{G_s, \boldsymbol{\Theta}_s^*}$: G_0, WL, BL, HE)

$\boldsymbol{\Theta}_{s+1} \leftarrow$ Estimate-Parameters($\boldsymbol{D}^*_{G_s, \boldsymbol{\Theta}_s^*}, G_{s+1}$)

(4) SEM의 M-step

} //end for

– SEM-H 구조 학습 종료 –

$cutline \leftarrow 0.01$

$G_s^* \leftarrow$Test-Hierarchy($G_s, \boldsymbol{\Theta}_s, cutline$) // G_s에 속한 각 유향 변 $A_i \rightarrow A_j$의 위계 관계 여부 판별

$\boldsymbol{\Theta}_s^* \leftarrow$ Parametric-EM($G_s^*, \boldsymbol{\Theta}_s, \boldsymbol{D}$) //

(5) 최종 위계 관계 판별

출력 { $G_s^*, \boldsymbol{\Theta}_s^*$ }

$I_{j'} \rightarrow A_{k'}$을 배제하지 않으면 구조 학습 시 선택될 가능성이 있으므로 그에 대응하는 순서쌍 $(I_{j'}, A_{k'})$도 블랙리스트에 추가된다.

[SEM 구조 학습(2~4단계)]

SEM-H 알고리즘의 2~4단계는 구조 학습이 이루어지는 단계로, SEM의 E-step(2단계)과 M-step(4단계)의 사이에 위계 관계를 판별하여 후보 구조를 제약하는 과정(3단계)이 추가되어 있다. 2~4단계는 BIC 점수가 더 이상 증가하지 않을 때까지 반복되는 과정이다.

(2) SEM의 E-step

이 단계에서는 현재 구조 G_s와 현재 모수치 $\boldsymbol{\Theta}_s$를 바탕으로 응답 자료 $\boldsymbol{D}$에 모수적 EM 추정방식을 적용하여 모수 $\boldsymbol{\Theta}_s$를 $\boldsymbol{\Theta}_s^*$로 갱신한다 ($\boldsymbol{\Theta}_s^* \leftarrow$ Parametric-EM$(G_s, \boldsymbol{\Theta}_s, \boldsymbol{D})$).

(3) 위계 관계 후보 판별

다음으로, E-step에서 도출된 모수를 바탕으로 BN의 추론 과정을 거쳐 불완전 자료를 완전 자료로 보강한다($\boldsymbol{D}^*_{G_s, \boldsymbol{\Theta}_s^*} \leftarrow$Inference$(G_s, \boldsymbol{\Theta}_s^*, \boldsymbol{D})$). 이 자료를 이용하여 M-step을 진행하기에 앞서, SEM-H 알고리즘은 이 단계에서 위계 관계의 후보를 판별하는 과정을 진행한다. 기본 아이디어는 가능한 모든 인지요소 관계 $A_i - A_j$에 대해 $A_i \rightarrow A_j$와 $A_j \rightarrow A_i$ 중 어느 방향이 위계 관계를 나타낼 가능성이 더 높은지 판별하여, 위계 관계를 나타낼 가능성이 높은 유향 변들로만 이루어진 구조들의 집합으로 탐색 공간을 한정하여 이후 M-step을 진행할 수 있도록 설정하는 것이다.

이를 위해, 우선, 가능한 모든 인지요소 관계들을 나타내는 순서쌍들의 집합을 설정하고($CE \leftarrow \{(i,j) | 1 \le i,j \le K, i \ne j\}$), 추후 이들을 판별한 뒤에 위계를 나타낼 가능성이 높은 관계들을 넣어둘 위계 관계 집합($HE \leftarrow \varnothing$)과 위계를 나타낼 가능성이 낮아서 구조 학습에서 배제할 관계들을 넣을 블랙리스트($BL \leftarrow \varnothing$)를 설정한다. 다음으로, 각 인지요소 관계 $A_i - A_j$에 대해 $P(A_j = 1 | A_i = 0)$과 $P(A_i = 1 | A_j = 0)$를 $\boldsymbol{D}^*_{G_s, \boldsymbol{\Theta}_s^*}$로부터 확인한 뒤 그 대소를 비교하여 $P(A_j = 1 | A_i = 0)$가 더 작은 경우 $A_i \rightarrow A_j$

를 HE에, $A_j \rightarrow A_i$는 BL에 넣고, 그렇지 않으면 $A_j \rightarrow A_i$를 HE에, $A_i \rightarrow A_j$를 BL에 넣는다(**while**문). 마지막으로, 1단계에서 설정한 Q-행렬을 반영한 블랙리스트 BL_0와 위계를 나타낼 가능성이 낮은 인지요소들의 관계로 이루어진 블랙리스트 BL을 합쳐서 M-step에서 적용할 최종적인 블랙리스트를 완성한다($BL \leftarrow BL \cup BL_0$).

(4) SEM의 M-step

SEM-H 알고리즘의 네 번째 단계에서는 SEM 구조 학습의 M-step을 진행하며, 구조 갱신($G_{s+1} \leftarrow \text{Learn-Structure}(\boldsymbol{D}^*_{G_s, \boldsymbol{\theta}_s^*} : G_0, WL, BL, HE)$)과 모수 갱신($\boldsymbol{\Theta}_{s+1} \leftarrow \text{Estimate-Parameters}(\boldsymbol{D}^*_{G_s, \boldsymbol{\theta}_s^*}, G_{s+1})$)의 두 가지 과정이 이루어진다. 구조 갱신에서는 앞서 설정한 화이트리스트(WL), 블랙리스트(BL), 위계 관계 집합(HE)이 제약으로 들어가며 구조 탐색의 시작점으로 초기 구조 G_0를 설정한다. 그러면 Learn-Structure는 구조 G_0에서 출발하여 HE에 포함된 유향 변들만 추가/삭제하면서 BIC 점수가 높아지는 구조를 탐색하여 다음 SEM 사이클을 위한 구조를 도출한다.

이와 같은 2~4단계의 구조 학습을 반복하여 최종적으로 도출된 구조는 이미 위계성을 반영하여 유향 변들의 방향이 설정되어 있는 구조이므로 가역 가능한 변이나 동형인 구조들을 고려할 필요가 없다. 그러나 위계 관계를 나타낼 가능성이 높은 유향 변들만 추가되도록 제약을 두고 구조 학습을 진행하였다 하더라도, 기본적으로 SEM 알고리즘의 BIC 점수 기반 구조 학습은 두 변수 사이에 유향 변이 추가되었을 때 점수가 높아지는지를 기준으로 구조를 찾는 방법일 뿐 추가된 유향 변이 반드시 위계 관계를 나타내는 것이라고 단정할 수는 없다. 만약 어떤 두 변수 간에 위계 관계가 존재한다면 Learn-Structure에 의해 그 관계가 구조에 추가될 것이나, 두 변수 간에 위계 관계가 아닌 종속 관계가 존재하면 그 관계 또한 추가될 수 있는 것이다. 따라서, SEM-H의 마지막 5단계에서는 구조 학습의 결과로 도출된 구조의 각 유향 변들이 실제로 위계 관계인지 아니면 위계 관계가 아닌 종속 관계를 나타낼 뿐인지에 대한 판정을 진행한다.

(5) 최종 위계 관계 판별

이 단계에서는 구조 학습의 결과로 도출된 G_s에 포함된 각 유향 변 $A_i \rightarrow A_j$의 위계 관계 여부를 판정한다. 기본 아이디어는 G_s에 대해 추정된 모수 집합 $\boldsymbol{\Theta}_s$에서 $\gamma_{j,0} = P(A_j = 1 | A_i = 0)$을 확인하여 그 값이 0에 가까운 값이면 $A_i \rightarrow A_j$가 위계 관계라고 판정하는 것이다. 모수 값이 0에 가까운 값임을 판정하는 기준으로 $cutline$을 설정하여($cutline \leftarrow 0.01$),[9] 각 관계 $A_i \rightarrow A_j$에 대해 $\gamma_{j,0} < cutline$인 경우 위계 관계로 판정하고, 그렇지 않은 경우 위계 관계가 아닌 종속 관계로 판정함으로써 최종 구조를 도출한다($G_s^* \leftarrow$Test-Hierarchy($G_s, \boldsymbol{\Theta}_s, cutline$)).[10] 마지막으로, 최종 구조에 대한 모수를 갱신한다($\boldsymbol{\Theta}_s^* \leftarrow$ Parametric-EM($G_s^*, \boldsymbol{\Theta}_s, \boldsymbol{D}$)).

9) 본 연구에서는 $cutline$을 0.01로 설정하였으나 약한 위계를 어느 정도 허용한다면 이 값을 더 높일 수도 있을 것이다.

10) 어떤 한 인지요소 변수 A_j의 부모 변수가 2개 이상인 경우에, 모수 $\gamma_{j,0}$는 두 부모 변수들 중 적어도 하나가 0일 때 A_j가 1일 확률 $P(A_j = 1 | \amalg Pa(A_j) = 0)$을 의미하므로 이 경우에는 위계 관계의 판정이 조금 더 복잡해진다. 이 과정에 대해서는 4절의 실제 검사 자료 분석을 통해 자세히 살펴볼 것이다.

3. 모의실험을 통한 SEM-H의 검증

이 절에서는 두 가지 모의실험을 통해 SEM-H의 위계 탐색 능력을 검증한 결과를 제시한다. 모의실험 1은 SEM-H 방법이 기존의 COMMAND 방법을 개선하는지 확인하기 위해 동일한 조건 하에서 두 방법의 위계 추정 능력을 비교하는 것에 초점을 둔다. 모의실험 2는 SEM-H의 객관적인 위계 추정 능력에 초점을 두고, 다양한 조건 하에서 SEM-H 방법의 위계 추정 정확도를 평가하는 것을 목적으로 한다. 모의실험에서는 오픈 소스 통계분석 소프트웨어인 R 3.4.2(R Core Team, 2017)를 이용하여 전반적인 코드를 작성하였으며, 특히 각 방법에서 후보 구조의 BIC 점수 계산과 베이지안 네트워크의 갱신을 위해 bnlearn 패키지(Scutari, 2010) 버전 4.1.1을 활용하였다.

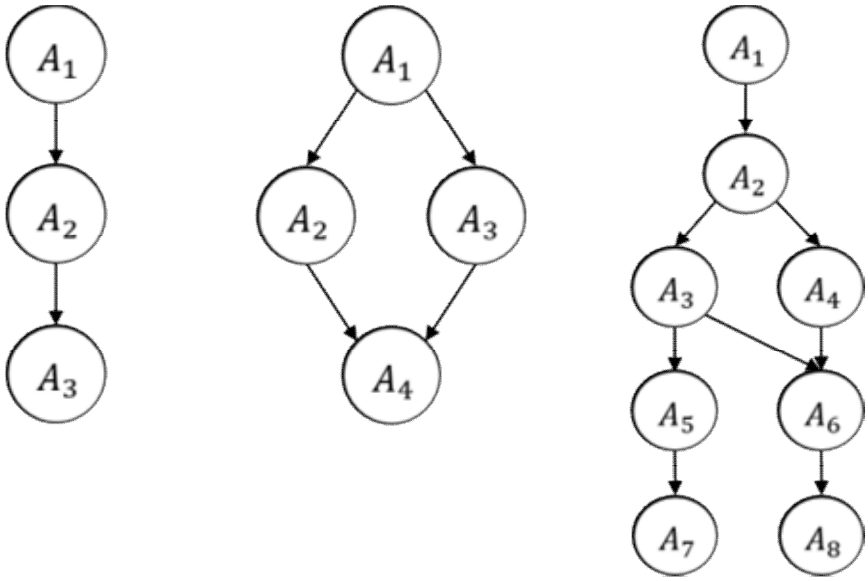
3.1. 모의실험 1

3.1.1. 모의실험 방법

3.1.1.1. 모의실험 설계

모의실험 1의 주목적은 기존의 COMMAND 방법과 본 연구에서 제안하는 SEM-H 방법의 비교이기 때문에, COMMAND 방법을 제시한 Chen 외(2016)의 모의실험에서 가정되었던 조건들을 포함하여 실험을 설계하고자 하였다. 이에 따라, 모의실험의 요인으로는 '인지요소 위계 구조', '표본 크기(피험자 수)', 'Q-행렬', '문항의 질'의 4가지 요인을 고려하였다.

먼저, ‘인지요소 위계 구조’ 요인의 경우, [그림 Ⅳ-2]와 같이 Chen 외(2016)에서 설정한 세 가지 구조를 동일하게 가정하였다. 구조 1은 3개의 인지요소와 2쌍의 위계 관계, 구조 2는 4개의 인지요소와 4쌍의 위계 관계로 이루어져 있으며, 구조 3은 8개의 인지요소와 8쌍의 위계 관계로 구성되어 있다.



(a) 구조 1 (b) 구조 2 (c) 구조 3

[그림 Ⅳ-2] 모의실험 1에서 가정한 인지요소 위계 구조

다음으로 고려한 요인은 '표본 크기(피험자 수)'이다. COMMAND 방법과 SEM-H 방법의 기반이 되는 SEM 구조 학습은 모수적 EM 방식으로 추정된 문항모수를 바탕으로 보강된 완전 자료를 이용하여 변수들 사이의 관계를 추정하는 방식이다. 따라서, 문항모수와 피험자모수 추정의 정확성에 영향을 미치는 것으로 알려져 있는 '표본 크기' 요인은 두 위계 추정 방법의 정확도에도 영향을 미칠 것으로 예상해볼 수 있다. de la Torre 외(2010)는 1000명 정도의 표본으로 DINA 모형의 문항모수와 피험자모수를 안정적으로 추정할 수 있음을 밝혔다. 그러나 현실적으로 교실이나 학교 단위 평가에서는 1000명의 피험자를 확보하는 것이 어려울 수 있다. 따라서 1000명보다 적은 피험자에 대해 수집된 자료에서도 어느 정도 안정적인 위계 추정 결과가 나오는지 확인해볼 필요가 있다. Chen 외(2016)는 1000명보다 적은 피험자 수를 나타내는 조건으로 200명과 500명을 설정하였고, 본 연구에서도 이를 따라 표본 크기를 200, 500, 1000으로 설정하였다.

Q-행렬'의 경우, Chen 외(2016)와 유사하게 단일 Q-행렬과 복합 Q-행렬 두 종류를 가정하였다. 단일 Q-행렬은 1개의 인지요소를 측정하는 문항들로만 검사가 이루어진 경우를, 복합 Q-행렬은 1개~3개의 인지요소를 측정하는 문항들로 검사가 이루어진 경우를 의미한다. Madison과 Bradshaw(2015)는 인지진단모형을 활용하고자 할 때 안정적인 추정을 위해서는 하나의 인지요소만을 정하는 문항이 인지요소별로 적어도 하나씩은 포함되어야 하며 되도록 두 문항 이상씩 포함되도록 검사를 구성할 것을 권고하였다. 이러한 점을 고려하여 본 실험에서 단일 Q-행렬은 인지요소 1개를 측정하는 문항이 인지요소별로 세 문항씩 포함되도록 구성하였으며, 복합 Q-행렬은 하나의 인지요소만을 측정하는 문항이 인지요소별로 두 문항씩 포함되도록 하고, 2개, 3개의 인지요소를 측정하는 문항은 가능한 모든 인지요소 조합별로 한 문항씩 포함되도록 설정하고자 하였다. 그러나 인지요소 수가 8개인 구조 3이 가정된 경우, 2개 또는 3개 인지요소를 측정하는 문항의 q-벡터로 가능한 인지요소 조합의 수가

82(=${}_8C_2+{}_8C_3$)개로 지나치게 많다. 따라서 구조 3이 가정된 경우에는 예외적으로 위계 구조의 그래프 상에서 인접한 2개 혹은 3개의 인지요소들을 측정하는 문항들만 검사에 포함시켰다. 이는 Liu, Huggins-Manley와 Bradshaw(2017)에서 위계 구조가 존재할 때 가능한 Q-행렬의 한 종류로 제시되었던 인접 Q-행렬의 개념을 반영한 것이다. 이와 같은 방식에 따라 모의실험 1에서 설정된 단일 Q-행렬과 복합 Q-행렬은 각각 [그림 Ⅳ-3]과 [그림 Ⅳ-4]와 같다.

[그림 Ⅳ-3]의 단일 Q-행렬과 [그림 Ⅳ-4]의 복합 Q-행렬을 비교해서 살펴보면, 복합 Q-행렬에는 단일 Q-행렬에 비해 한 인지요소를 측정하는 문항의 수가 더 많이 포함되어 있음을 확인할 수 있다. 단일 Q-행렬에는 한 인지요소를 측정하는 문항이 인지요소별로 세 문항씩 포함되어 있는 반면, 복합 Q-행렬에는 한 인지요소를 측정하는 문항이 위계 구조 1의 경우 다섯 문항씩, 위계 구조 2의 경우 여덟 문항씩, 위계 구조 3의 경우 인지요소에 따라 다섯에서 열 문항이 포함되어 있다.

'문항의 질' 요인의 경우 높음과 낮음의 두 수준을 가정하였다. DINA 모형을 측정모형으로 가정하는 경우 문항의 부주의 모수 s_j와 추측 모수 g_j의 값이 작으면 문항의 질이 높은 것으로, 값이 크면 문항의 질이 낮은 것으로 생각할 수 있다. de la Torre 외(2010)는 질이 높은 문항의 모수를 0.05~0.15로, 질이 낮은 문항의 모수를 0.2~0.3으로 조건화하였으며 Chen 외(2016)는 질이 높은 문항의 모수와 낮은 문항의 모수의 범위를 0.05~0.25와 0.25~0.45로 각각 설정하였다. 박세진과 이현숙(2017)은 DINA 모형의 모수가 0.2인 경우와 0.4인 경우를 각각 질이 높은 문항과 낮은 문항으로 정의하였다. 본 실험에서는 질이 높은 문항의 s_j와 g_j는 $U(0.1, 0.25)$에서, 질이 낮은 문항의 s_j와 g_j는 $U(0.15, 0.3)$에서 무선 추출하여 응답 자료 생성을 위한 진(true) 문항모수로 사용하였다.

	A_1	A_2	A_3
I1	1	0	0
I2	0	1	0
I3	0	0	1
I4	1	0	0
I5	0	1	0
I6	0	0	1
I7	1	0	0
I8	0	1	0
I9	0	0	1

	A_1	A_2	A_3	A_4
I1	1	0	0	0
I2	0	1	0	0
I3	0	0	1	0
I4	0	0	0	1
I5	1	0	0	0
I6	0	1	0	0
I7	0	0	1	0
I8	0	0	0	1
I9	1	0	0	0
I10	0	1	0	0
I11	0	0	1	0
I12	0	0	0	1

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
I1	1	0	0	0	0	0	0	0
I2	0	1	0	0	0	0	0	0
I3	0	0	1	0	0	0	0	0
I4	0	0	0	1	0	0	0	0
I5	0	0	0	0	1	0	0	0
I6	0	0	0	0	0	1	0	0
I7	0	0	0	0	0	0	1	0
I8	0	0	0	0	0	0	0	1
I9	1	0	0	0	0	0	0	0
I10	0	1	0	0	0	0	0	0
I11	0	0	1	0	0	0	0	0
I12	0	0	0	1	0	0	0	0
I13	0	0	0	0	1	0	0	0
I14	0	0	0	0	0	1	0	0
I15	0	0	0	0	0	0	1	0
I16	0	0	0	0	0	0	0	1
I17	1	0	0	0	0	0	0	0
I18	0	1	0	0	0	0	0	0
I19	0	0	1	0	0	0	0	0
I20	0	0	0	1	0	0	0	0
I21	0	0	0	0	1	0	0	0
I22	0	0	0	0	0	1	0	0
I23	0	0	0	0	0	0	1	0
I24	0	0	0	0	0	0	0	1

[그림 Ⅳ-3] 모의실험 1에서 가정한 단일 Q-행렬

	A_1	A_2	A_3
I1	1	0	0
I2	0	1	0
I3	0	0	1
I4	1	0	0
I5	0	1	0
I6	0	0	1
I7	1	1	0
I8	1	0	1
I9	0	1	1
I10	1	1	1

	A_1	A_2	A_3	A_4
I1	1	0	0	0
I2	0	1	0	0
I3	0	0	1	0
I4	0	0	0	1
I5	1	0	0	0
I6	0	1	0	0
I7	0	0	1	0
I8	0	0	0	1
I9	1	1	0	0
I10	1	0	1	0
I11	1	0	0	1
I12	0	1	1	0
I13	0	1	0	1
I14	0	0	1	1
I15	1	1	1	0
I16	1	1	0	1
I17	1	0	1	1
I18	0	1	1	1

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
I1	1	0	0	0	0	0	0	0
I2	0	1	0	0	0	0	0	0
I3	0	0	1	0	0	0	0	0
I4	0	0	0	1	0	0	0	0
I5	0	0	0	0	1	0	0	0
I6	0	0	0	0	0	1	0	0
I7	0	0	0	0	0	0	1	0
I8	0	0	0	0	0	0	0	1
I9	1	0	0	0	0	0	0	0
I10	0	1	0	0	0	0	0	0
I11	0	0	1	0	0	0	0	0
I12	0	0	0	1	0	0	0	0
I13	0	0	0	0	1	0	0	0
I14	0	0	0	0	0	1	0	0
I15	0	0	0	0	0	0	1	0
I16	0	0	0	0	0	0	0	1
I17	1	1	0	0	0	0	0	0
I18	0	1	1	0	0	0	0	0
I19	0	1	0	1	0	0	0	0
I20	0	0	1	0	1	0	0	0
I21	0	0	1	0	0	1	0	0
I22	0	0	0	1	0	1	0	0
I23	0	0	0	0	1	0	1	0
I24	0	0	0	0	0	1	0	1
I25	1	1	1	0	0	0	0	0
I26	1	1	0	1	0	0	0	0
I27	0	1	1	0	1	0	0	0
I28	0	1	1	0	0	1	0	0
I29	0	1	0	1	0	1	0	0
I30	0	0	1	0	1	0	1	0
I31	0	0	1	0	0	1	0	1
I32	0	0	0	1	0	1	0	1

[그림 Ⅳ-4] 모의실험 1에서 가정한 복합 Q-행렬

이상에서 설명한 모의실험 1의 요인들을 정리하면 <표 Ⅳ-4>와 같다.

<표 Ⅳ-4> 모의실험 1의 요인

요인	수	요인 상세
인지요소 위계 구조	3	구조 1: 인지요소 3개 구조 2: 인지요소 4개 구조 3: 인지요소 8개
표본 크기 (피험자 수)	3	200, 500, 1000
Q-행렬	2	단일: 각 문항이 1개 인지요소 측정 복합: 각 문항이 1~3개 인지요소 측정
문항의 질	2	높음: $s_j, g_j \in [0.1, 0.25]$ 낮음: $s_j, g_j \in [0.15, 0.3]$

결과적으로, 모의실험 1에서는 총 36개의 조건('인지요소 위계 구조' 3 × '표본 크기(피험자 수)' 3 × 'Q-행렬' 2 × '문항의 질' 2)을 고려하였다.

3.1.2. 모의실험 자료의 생성

앞에서 설명한 각 모의실험 조건 하에서 500개의 반복 자료(replication datasets)를 생성하였다. 각 자료는 피험자들의 진 인지요소 패턴과 문항 응답으로 구성되었으며, 자료 생성 과정은 다음과 같다.

① 피험자집단의 인지요소패턴 분포를 나타내는 베이지안 네트워크를 설정한다. 이때 베이지안 네트워크의 DAG는 '인지요소 위계 구조' 조건에 의해 설정되며, CPT의 값들은 위계 구조에 따라 축소된 인지요소패턴들을 피험자들이 고르게 가진다는 가정을 충족하도록 설정하되, Chen 외(2016)와 동일하게 각 위계 관계 쌍 $A_i \rightarrow A_j$에 대해 $P(A_j = 1 | A_i = 0)$의 값은 [0, 0.1]에서 무선 추출하였다. 이는 강한 위계 관계를 따르지 않는 패턴들을 낮은 확률로 허용한다는 점을 의미한다.

② ①에서 구성한 베이지안 네트워크를 기반으로 각 피험자별 진 인지요소패턴을 다음과 같이 생성한다. 한 피험자에 대해, 인지요소 위계 구조의 최상단에 있는 최하위 수준 인지요소인 A_1의 숙달 여부를 가장 먼저 결정한다. 이때 베이지안 네트워크의 CPT에서 $P(A_1 = 1)$의 값이 γ_1이라면 A_1의 숙달 여부는 $Bernoulli(\gamma_1)$에서 무선 추출된다. A_1의 숙달 여부가 0 또는 1로 결정되고 나면, A_1의 숙달 여부에 따라 A_1의 후행 인지요소의 숙달 여부를 결정한다. 예를 들어, 두 인지요소 A_1, A_2 사이에 위계 관계 $A_1 \rightarrow A_2$가 존재하고 베이지안 네트워크의 CPT에서 $P(A_2 = 1 | A_1 = 0)$의 값이 $\gamma_{2,0}$, $P(A_2 = 1 | A_1 = 1)$의 값이 $\gamma_{2,1}$이라면, A_2의 숙달 여부는 $A_1 = 0$인 경우에는 $Bernoulli(\gamma_{2,0})$에서, $A_1 = 1$인 경우에는 $Bernoulli(\gamma_{2,1})$에서 무선 추출된다. 이와 같은 방식으로 위계 구조의 상단에서 하단으로 내려가면서 각 인지요소들의 숙달 여부를 결정하였다.

③ ②에서 생성된 인지요소패턴을 기반으로 각 피험자의 문항 응답을 다음과 같은 방식으로 생성한다. 어떤 문항 j의 부주의 모수의 값이 s_j, 추측 모수의 값이 g_j일 때, 한 피험자의 문항 j에 대한 응답은 그 피험자가 문항 j가 측정하는 모든 인지요소들을 숙달하였으면 $Bernoulli(1-s_j)$에서, 문항 j가 측정하는 인지요소 중 하나라도 숙달하지 않았으면 $Bernoulli(g_j)$에서 무선 추출된다.

이렇게 생성된 자료에서 피험자들의 진 인지요소패턴은 실제로는 관찰될 수 없는 자료이기 때문에, 피험자들의 응답 자료만 가지고 위계 추정 방법을 적용하여 위계 추정 결과를 도출한다.

3.1.3. 평가 지표

각 방법에 의해 추정된 위계 구조가 진(true) 위계 구조와 얼마나 일치하는지를 평가하는 기준으로 크게 다음의 두 지표를 고려하였다.

$$\text{TPER} = \frac{\text{추정된 위계 구조의 진 위계 관계 수}}{\text{진 위계 구조의 총 위계 관계 수}} \tag{4.2}$$

$$\text{TDER} = \frac{\text{추정된 위계 구조의 진 위계 관계 수}}{\text{추정된 위계 구조의 총 위계 관계 수}} \tag{4.3}$$

TPER은 진 위계 구조의 위계 관계를 얼마나 많이 추정하였는지를 나타내는 지표이며, TDER은 추정된 위계 구조에 진 위계 관계가 얼마나 포함되었는지를 나타내는 지표이다. 이 중 TDER의 경우, 진 위계 구조의 간접적인 위계 관계를 추정한 경우는 옳게 추정한 것으로 계산되도록 수정한 지표인

$$\text{r-TDER} = \frac{\text{진 위계 관계 수} + \text{진 간접 위계 관계 수}}{\text{추정된 위계 구조의 총 위계 관계 수}} \tag{4.4}$$

도 평가 지표로 함께 사용하였다.

예를 들어, [그림 Ⅳ-2]의 (a)에 제시된 구조 1이 진 위계 구조인 경우, 진 위계 구조의 총 위계 관계 수는 2쌍($A_1 \rightarrow A_2$, $A_2 \rightarrow A_3$)이고, 진 위계 구조에서 이 2쌍의 위계 관계로 인해 생기는 간접적인 위계 관계 $A_1 \rightarrow A_3$가 진 간접 위계 관계가 된다.[11] 따라서 만약 구조 1에 대해 추정된 위계 구조가 $A_1 \rightarrow A_2$, $A_2 \rightarrow A_3$, $A_1 \rightarrow A_3$의 3쌍의 위계 관계로 이루어져 있는 경우, 각 평가지표는 다음과 같이 계산된다.

$$\text{TPER} = \frac{\text{추정된 위계 구조의 진 위계 관계 수}}{\text{진 위계 구조의 총 위계 관계 수}} = \frac{2}{2} = 1$$

$$\text{TDER} = \frac{\text{추정된 위계 구조의 진 위계 관계 수}}{\text{추정된 위계 구조의 총 위계 관계 수}} = \frac{2}{3} \fallingdotseq 0.667$$

$$\text{r-TDER} = \frac{\text{진 위계 관계 수 + 진 간접 위계 관계 수}}{\text{추정된 위계 구조의 총 위계 관계 수}} = \frac{2+1}{3} = 1$$

TPER은 참인 위계 관계들을 얼마나 잘 찾았는지를 나타내는 지표로, TPER이 높을수록 진 위계 구조에 포함된 위계 관계들을 더 잘 추정한 것이라고 해석할 수 있다. TDER은 추정된 위계 구조에 있는 위계 관계들 중 참인 관계가 얼마나 있는지를 나타내는 지표로, TDER이 높을수록 참이 아닌 위계 관계를 추정하는 경우가 적다고 해석할 수 있다. 동일한 조건 하에서 TDER에 비해 r-TDER이 높아진다면 진 간접 위계 관계를 추정한 경우가 있는 것으로 해석할 수 있다.

11) 구조 2에 포함된 진 간접 위계 관계는 $A_1 \rightarrow A_4$의 1쌍이며, 구조 3에 포함된 진 간접 위계 관계는 $A_1 \rightarrow A_3$, $A_1 \rightarrow A_4$, $A_1 \rightarrow A_5$, $A_1 \rightarrow A_6$, $A_1 \rightarrow A_7$, $A_1 \rightarrow A_8$, $A_2 \rightarrow A_5$, $A_2 \rightarrow A_6$, $A_2 \rightarrow A_7$, $A_2 \rightarrow A_8$, $A_3 \rightarrow A_7$, $A_3 \rightarrow A_8$, $A_4 \rightarrow A_8$의 총 13쌍이다.

3.2. 모의실험 결과

모의실험의 각 조건 하에서 위계 추정 결과는 <표 Ⅳ-5>와 같다.

<표 Ⅳ-5> 모의실험 1의 위계 추정 결과

위계 구조	표본 크기	Q-행렬	문항 질	TPER		TDER		r-TDER	
				CMD	SEM-H	CMD	SEM-H	CMD	SEM-H
1	200	단일	높음	0.970	0.981	0.970	0.981	0.988	1.000
			낮음	0.874	0.892	0.947	0.966	0.985	0.998
		복합	높음	0.729	0.879	0.763	0.949	0.851	1.000
			낮음	0.702	0.746	0.814	0.918	0.892	0.998
	500	단일	높음	0.999	1.000	0.992	0.993	0.999	1.000
			낮음	0.987	0.996	0.978	0.986	0.994	1.000
		복합	높음	0.798	0.973	0.722	0.865	0.868	1.000
			낮음	0.809	0.918	0.770	0.902	0.885	1.000
	1000	단일	높음	1.000	1.000	0.980	0.980	1.000	1.000
			낮음	1.000	1.000	0.973	0.973	1.000	1.000
		복합	높음	0.893	0.998	0.679	0.747	0.916	1.000
			낮음	0.863	0.984	0.715	0.812	0.904	1.000
2	200	단일	높음	0.694	0.771	0.872	0.973	0.881	0.980
			낮음	0.510	0.548	0.873	0.952	0.891	0.967
		복합	높음	0.728	0.858	0.823	0.978	0.831	0.980
			낮음	0.563	0.683	0.738	0.949	0.750	0.956
	500	단일	높음	0.880	0.960	0.900	0.982	0.901	0.983
			낮음	0.712	0.797	0.872	0.971	0.879	0.976
		복합	높음	0.936	0.987	0.932	0.985	0.935	0.985
			낮음	0.812	0.906	0.853	0.956	0.858	0.957
	1000	단일	높음	0.942	0.999	0.900	0.943	0.900	0.944
			낮음	0.839	0.949	0.869	0.976	0.870	0.977
		복합	높음	0.978	0.999	0.943	0.963	0.946	0.963
			낮음	0.933	0.980	0.908	0.953	0.914	0.953
3	200	단일	높음	0.749	0.861	0.779	0.909	0.853	0.962
			낮음	0.577	0.669	0.692	0.857	0.801	0.951
		복합	높음	0.658	0.820	0.691	0.880	0.801	0.966
			낮음	0.510	0.650	0.610	0.843	0.734	0.945
	500	단일	높음	0.882	0.983	0.815	0.911	0.871	0.968
			낮음	0.765	0.895	0.764	0.908	0.835	0.964
		복합	높음	0.792	0.897	0.709	0.828	0.861	0.971
			낮음	0.646	0.805	0.662	0.842	0.791	0.953
	1000	단일	높음	0.914	0.998	0.778	0.838	0.868	0.945
			낮음	0.851	0.976	0.762	0.863	0.843	0.955
		복합	높음	0.881	0.919	0.679	0.736	0.906	0.965
			낮음	0.739	0.846	0.648	0.768	0.839	0.950

<표 Ⅳ-5>의 결과를 보면, 두 위계 추정 방법에서 각 평가 지표의 값들이 동일하게 나타난 일부 조건(위계 구조 1, 표본 크기 1000, 단일 Q-행렬의 경우)을 제외하고는, TPER, TDER, r-TDER의 세 가지 평가 지표 모두에 대해 COMMAND 방법의 결과 값보다 SEM-H 방법의 결과 값이 더 높게 나타났다. 이러한 결과는 COMMAND 방법보다 SEM-H 방법의 위계 추정 능력이 전반적으로 더 뛰어남을 보여준다. 따라서 본 연구에서 제안한 SEM-H 방법이 기존의 COMMAND 방법의 위계 추정 정확도를 실질적으로 개선하는 방법이라고 할 수 있다.

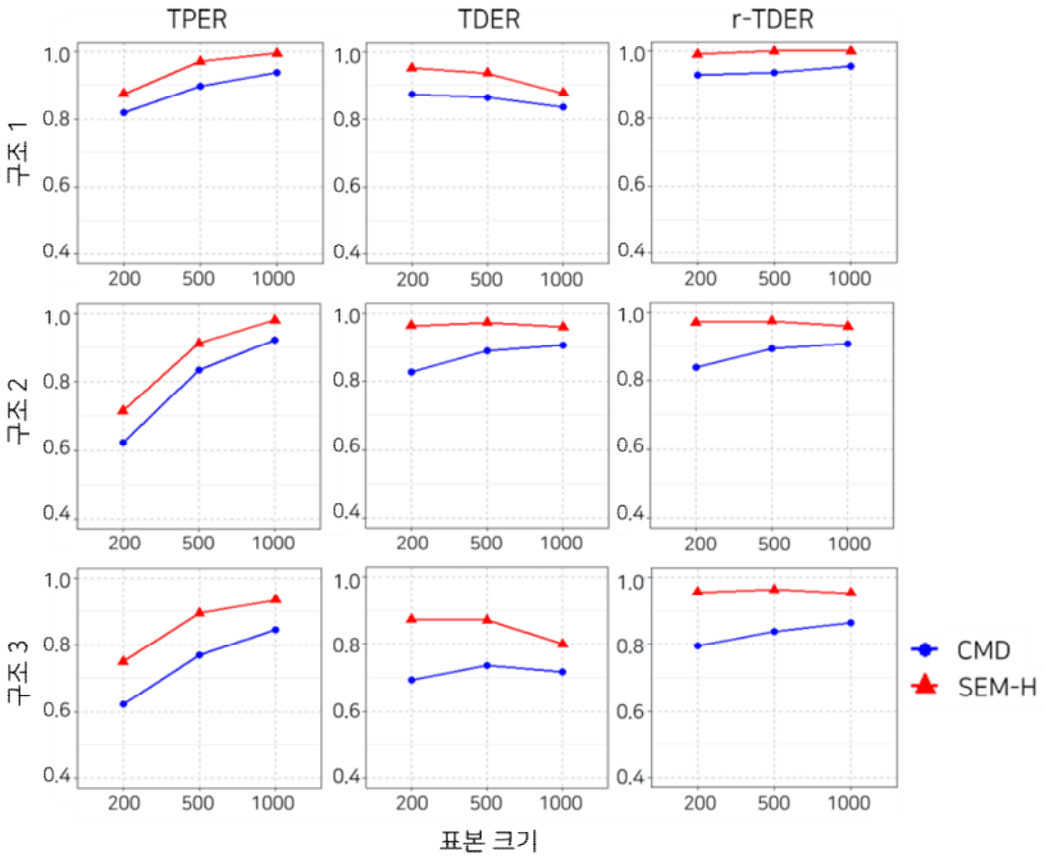
다음으로, <표 Ⅳ-5>에서 인지요소 위계 구조별로 각 요인(표본 크기, Q-행렬, 문항의 질)이 동일한 조건들에 대해 각 평가 지표들의 평균을 낸 결과를 토대로 각 요인이 변함에 따라 COMMAND 방법과 SEM-H 방법 간의 차이가 어떻게 변하는지 구체적으로 살펴본다.

3.2.1. 표본 크기에 따른 결과 비교

다음의 <표 Ⅳ-6>은 인지요소 위계 구조별로 각 표본 크기에 따른 평균적인 위계 추정 결과 값들을 나타낸 표이며, 그 결과를 그래프로 나타내면 [그림 Ⅳ-5]와 같다.

<표 Ⅳ-6> 인지요소 위계 구조별 각 표본 크기에 대한 위계 추정 결과 평균

위계 구조	표본 크기	TPER		TDER		r-TDER	
		CMD	SEM-H	CMD	SEM-H	CMD	SEM-H
1	200	0.819	0.875	0.874	0.953	0.929	0.999
	500	0.898	0.972	0.866	0.937	0.937	1.000
	1000	0.939	0.996	0.837	0.878	0.955	1.000
2	200	0.623	0.715	0.827	0.963	0.838	0.971
	500	0.835	0.912	0.889	0.973	0.893	0.975
	1000	0.923	0.981	0.905	0.959	0.908	0.959
3	200	0.623	0.750	0.693	0.873	0.797	0.956
	500	0.771	0.895	0.737	0.872	0.839	0.964
	1000	0.847	0.935	0.717	0.801	0.864	0.954



[그림 Ⅳ-5] 표본 크기에 따른 위계 추정 결과의 그래프

위에 제시된 결과를 각 평가 지표별로 살펴보자. 먼저, 진 위계 구조의 위계 관계를 얼마나 잘 추정하였는지를 나타내는 TPER의 경우, 각 위계 구조 하에서 두 방법 모두 표본 크기가 커짐에 따라 그 값이 일관되게 증가하고 있으나 두 방법 간 TPER 값의 차이는 표본 크기가 커지더라도 크게 변화하지 않았다.

추정된 위계 구조에 진 위계 관계가 얼마나 포함되어 있는지를 나타내는 TDER과 추정된 위계 구조에 진 위계 관계 및 진 간접 위계 관계가 얼마나 포함되어 있는지를 나타내는 r-TDER의 경우, 표본 크기가 커짐에 따라 두 방법의 차이가 대체로 줄어드는 결과를 나타낸다. 이는 표본 크기가 작은 경우에 SEM-H 방법이 COMMAND 방법에 비해 참이 아닌 위계 관계를 추정할 가능성이 특히 더 낮다는 점을 시사한다.

한편, 구조 1과 구조 3의 TDER 결과에 주목해보면, TPER과는 달리 표본 크기가 커짐에 따라 그 값이 대체로 내려가는 결과가 나타난다. 이와 같은 결과가 나타난 이유는 TDER과 r-TDER을 함께 비교함으로써 확인해볼 수 있다. 구조 1과 구조 3의 r-TDER은 TDER 보다 상당히 높게 나타난다. r-TDER은 진 간접 위계를 찾은 경우에는 위계 관계를 정확하게 추정한 것으로 보는 지표이므로, TDER에 비해 r-TDER이 높게 나타나는 결과는 진 간접 위계 관계를 추정한 경우가 다수 있었다는 점을 시사한다. 따라서 구조 1과 3에서는 표본 크기가 커짐에 따라 진 간접 위계 관계를 추정하는 경우가 늘어나면서 TDER이 낮게 나타난 것이었다고 볼 수 있다.

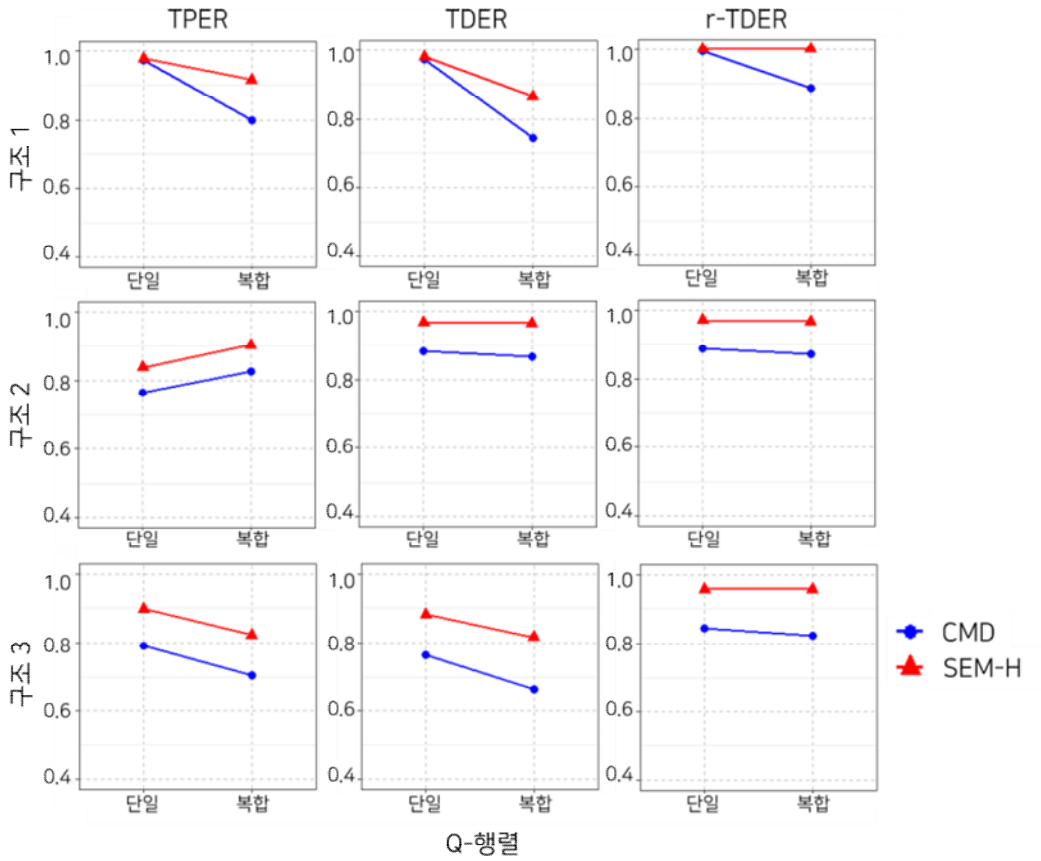
3.2.2. Q-행렬에 따른 결과 비교

다음의 <표 Ⅳ-7>은 인지요소 위계 구조별로 각 Q-행렬에 따른 평균적인 위계 추정 결과 값들을 나타낸 표이며, 그 결과를 그래프로 나타내면 [그림 Ⅳ-6]과 같다.

<표 Ⅳ-7> 인지요소 위계 구조별 각 Q-행렬에 대한 위계 추정 결과 평균

위계 구조	Q-행렬	TPER		TDER		r-TDER	
		CMD	SEM-H	CMD	SEM-H	CMD	SEM-H
1	단일	0.972	0.978	0.974	0.980	0.994	1.000
	복합	0.799	0.916	0.744	0.865	0.886	1.000
2	단일	0.763	0.837	0.881	0.966	0.887	0.971
	복합	0.825	0.902	0.866	0.964	0.872	0.966
3	단일	0.790	0.897	0.765	0.881	0.845	0.958
	복합	0.704	0.823	0.666	0.816	0.822	0.958

결과에 따르면 일단 두 위계 추정 방법 모두 구조 2의 TPER 결과를 제외하고는 단일 Q-행렬보다 복합 Q-행렬이 사용되는 경우에 위계 추정의 정확도들이 대체로 낮아지는 양상을 보인다. 복합 Q-행렬에 포함된 2개 이상의 인지요소를 동시에 측정하는 문항들은 인지요소패턴 추



[그림 Ⅳ-6] Q-행렬에 따른 위계 추정 결과의 그래프

정의 정확성을 낮출 수 있기 때문에(Madison & Bradshaw, 2015), 추정된 인지요소패턴을 기반으로 위계를 추정하는 두 방법의 정확도가 함께 낮아진 것이라 해석해볼 수 있다.

한편, 단일 Q-행렬이 아닌 복합 Q-행렬이 사용됨에 따라 정확도가 낮아지는 정도는 SEM-H 방법보다 COMMAND 방법에서 더 크게 나타났다.[12] 즉, 단일 Q-행렬보다는 복합 Q-행렬이 사용되는 경우에 SEM-H 방법이 COMMAND 방법의 위계 추정 결과를 더 크게 개선할 수 있음이 확인된다.

12) 위계 구조 1이 가정된 경우 복합 Q-행렬이 사용된 경우에 TPER 값이 단일 Q-행렬에 비해 얼마나 낮아지는지를 <표 Ⅳ-7>로부터 확인해보면, COMMAND 방법의 경우 0.173(=0.972−0.799) 만큼 낮아진 반면 SEM-H 방법은 0.062(=0.978−0.916) 낮아진 결과를 보이는 것을 확인할 수 있다.

3.2.3. 문항의 질에 따른 결과 비교

다음의 <표 Ⅳ-8>은 인지요소 위계 구조별로 각 문항의 질에 따른 평균적인 위계 추정 결과 값들을 나타낸 표이며, 그 결과를 그래프로 나타내면 [그림 Ⅳ-7]과 같다.

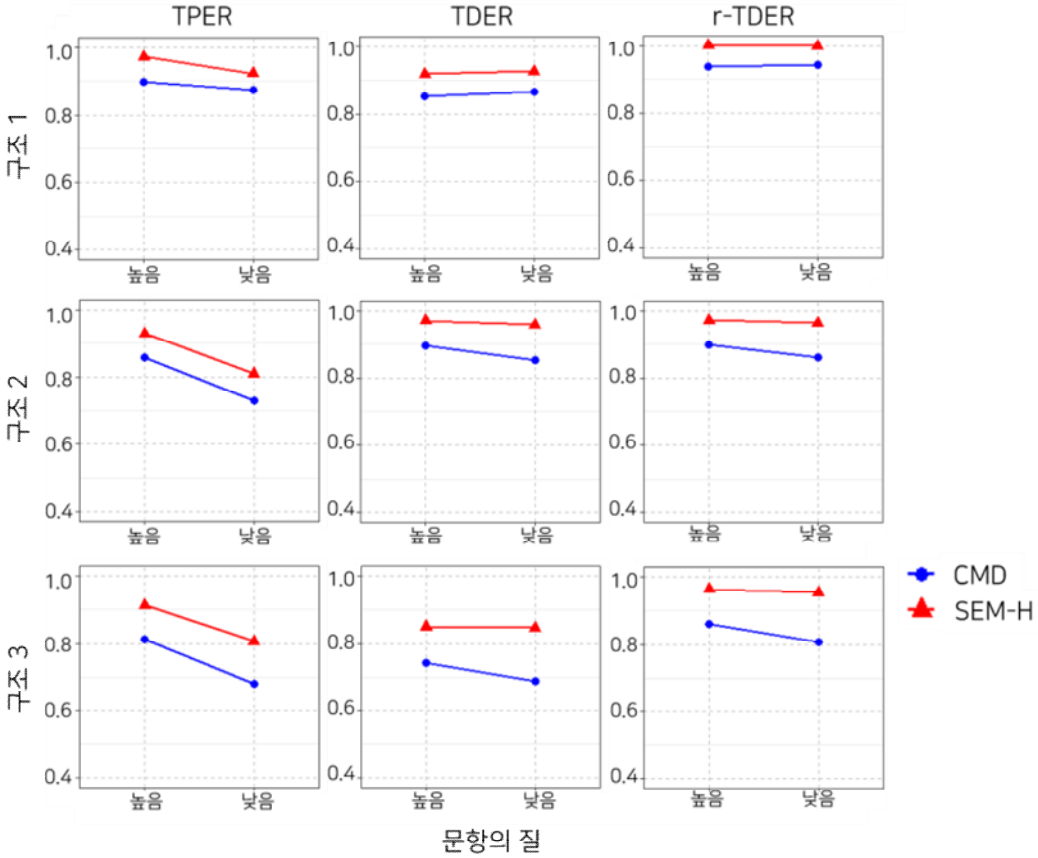
<표 Ⅳ-8> 인지요소 위계 구조별 각 문항 질에 대한 위계 추정 결과 평균

위계 구조	문항 질	TPER		TDER		r-TDER	
		CMD	SEM-H	CMD	SEM-H	CMD	SEM-H
1	높음	0.898	0.972	0.851	0.919	0.937	1.000
	낮음	0.873	0.923	0.866	0.926	0.943	0.999
2	높음	0.859	0.929	0.895	0.971	0.899	0.973
	낮음	0.728	0.810	0.852	0.959	0.860	0.964
3	높음	0.813	0.913	0.742	0.850	0.860	0.963
	낮음	0.681	0.807	0.690	0.847	0.807	0.953

<표 Ⅳ-8>과 [그림 Ⅳ-7]에 제시된 결과에 따르면, 구조 1의 일부 경우를 제외한 대부분의 경우에서 문항의 질이 낮아짐에 따라 위계 추정의 정확도도 낮아지는 양상이 나타났으며, 위계 구조가 복잡해질수록 문항의 질이 낮을 때 SEM-H 방법을 사용함으로써 COMMAND 방법의 위계 추정 결과를 더 크게 개선할 수 있음이 확인된다.[13)]

[그림 Ⅳ-7]에서 SEM-H 방법의 결과를 나타내는 그래프에 주목해보면, 문항의 질이 낮아짐에 따라 SEM-H 방법의 TPER은 상당히 낮아지는 반면, TDER은 크게 낮아지지 않는다는 점을 알 수 있다. 이는 문항의 질이 낮아지면 SEM-H 방법이 진 위계 구조의 위계 관계들

13) 각 위계 구조 하에서 문항의 질이 낮을 때 COMMAND 방법 대신 SEM-H 방법을 사용함으로써 각 평가 지표의 값이 얼마나 향상되는지를 살펴보면, 구조 1에서 TPER은 $0.05(=0.923-0.873)$, TDER은 $0.06(=0.926-0.866)$, r-TDER은 $0.057(=0.999-0.943)$만큼 향상된 반면, 위계 구조 2에서는 $0.082, 0.107, 0.104$만큼, 위계 구조 3에서는 $0.126, 0.157, 0.146$만큼 향상되는 등, 구조가 복잡해질수록 SEM-H 방법이 COMMAND 방법의 위계 추정 정확도를 더 크게 향상시키고 있음을 확인할 수 있다.



[그림 Ⅳ-7] 문항의 질에 따른 위계 추정 결과의 그래프

을 찾아내는 정도는 낮아지지만 그렇다고 하여 참이 아닌 위계 관계를 추정해내지는 않는다는 점을 시사한다.

한편, 위계 구조 1이 가정된 경우에 COMMAND 방법과 SEM-H의 TDER, 그리고 COMMAND 방법의 r-TDER에서 문항의 질이 낮아질 때 그 값이 올라가는 예외적인 결과가 나타났다. 앞서 제시한 <표 Ⅳ-5>의 결과를 함께 살펴보면 이러한 결과는 표본 크기 500과 1000에서 복합 Q-행렬이 사용된 경우에 나타났던 것으로 확인된다. 예를 들어, (위계 구조 1, 표본 크기 500, 복합 Q-행렬)의 경우에서 COMMAND의 TDER은 문항 질이 높을 때 0.722, 낮을 때 0.770이었으며, SEM-H의 TDER은 문항 질이 높을 때 0.865, 낮을 때 0.902로 나타났다. 여러 연구들을 통해 문항의 질이 높아질수록 추정의 정확도는 일반적으로 올라간

다는 점이 밝혀져 있는 만큼(예를 들어, de la Torre, Hong, & Deng, 2010; Huebner & Wang, 2011; 박세진, 이현숙, 2017; 박세진, 이현숙, 2018) 이러한 예외적인 결과가 실제로 유의미한 차이를 나타내는 것인지에 대해서는 추가적인 분석이 필요할 것으로 보인다.

3.2. 모의실험 2

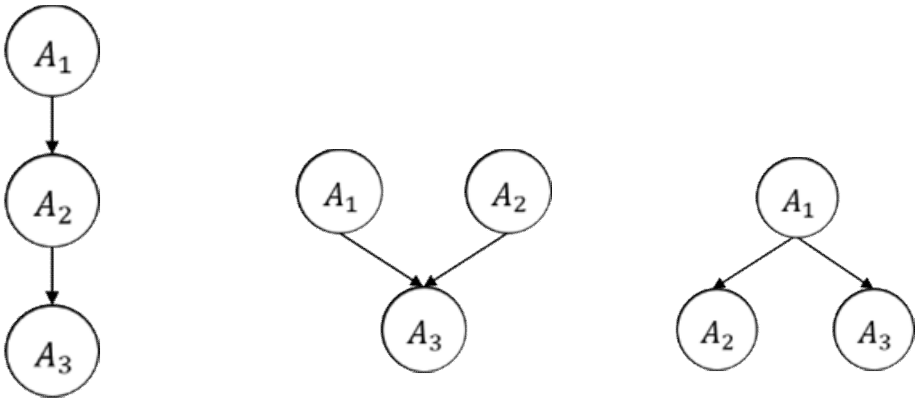
3.2.1. 모의실험 방법

모의실험 1은 기존의 COMMAND 방법과 SEM-H 방법의 비교에 초점을 두었던 반면, 모의실험 2는 다양한 조건 하에서 SEM-H 방법의 위계 추정 능력을 평가하는 것에 목적을 둔다. 모의실험 2에서는 '인지요소 위계 구조', '문항의 질', '표본 크기(피험자 수)', '문항 수'를 실험 요인으로 고려하였다.

'인지요소 위계 구조' 요인은 위계 구조의 유형에 따라 위계 추정의 결과가 어떻게 달라지는지를 확인하는 것을 목적으로 설정된 요인이다. 이를 위해 본 연구에서는 우선 Leighton 외(2004), Liu(2018)등에서 제시한 위계 구조의 종류를 참고로 [그림 Ⅳ-8]과 같이 인지요소 위계 구조의 유형을 세 가지로 분류하였다.

다음으로 각 위계 구조 유형별로 위계 구조에 인지요소와 유향 변을 하나씩 추가해나가는 방식으로 위계 구조 조건들을 설정하였다. 인지요소의 수가 많아짐에 따라 가능한 구조의 수가 기하급수적으로 늘어나기 때문에 모든 인지요소를 고려하지는 못하였으며, 다음과 같은 방식으로 구조를 설정하였다.

우선 선형 구조의 경우 인지요소가 늘어나도 선형이 되도록 구조의 아래쪽으로만 인지요소와 유향 변을 추가하는 방식으로 구조를 설정하였다. 반면, 수렴형과 발산형 구조의 경우 구조의 위쪽 또는 아래쪽에 인지요소와 유향 변을 추가하는 경우와 옆쪽으로 유향 변과 인지요소를 추가

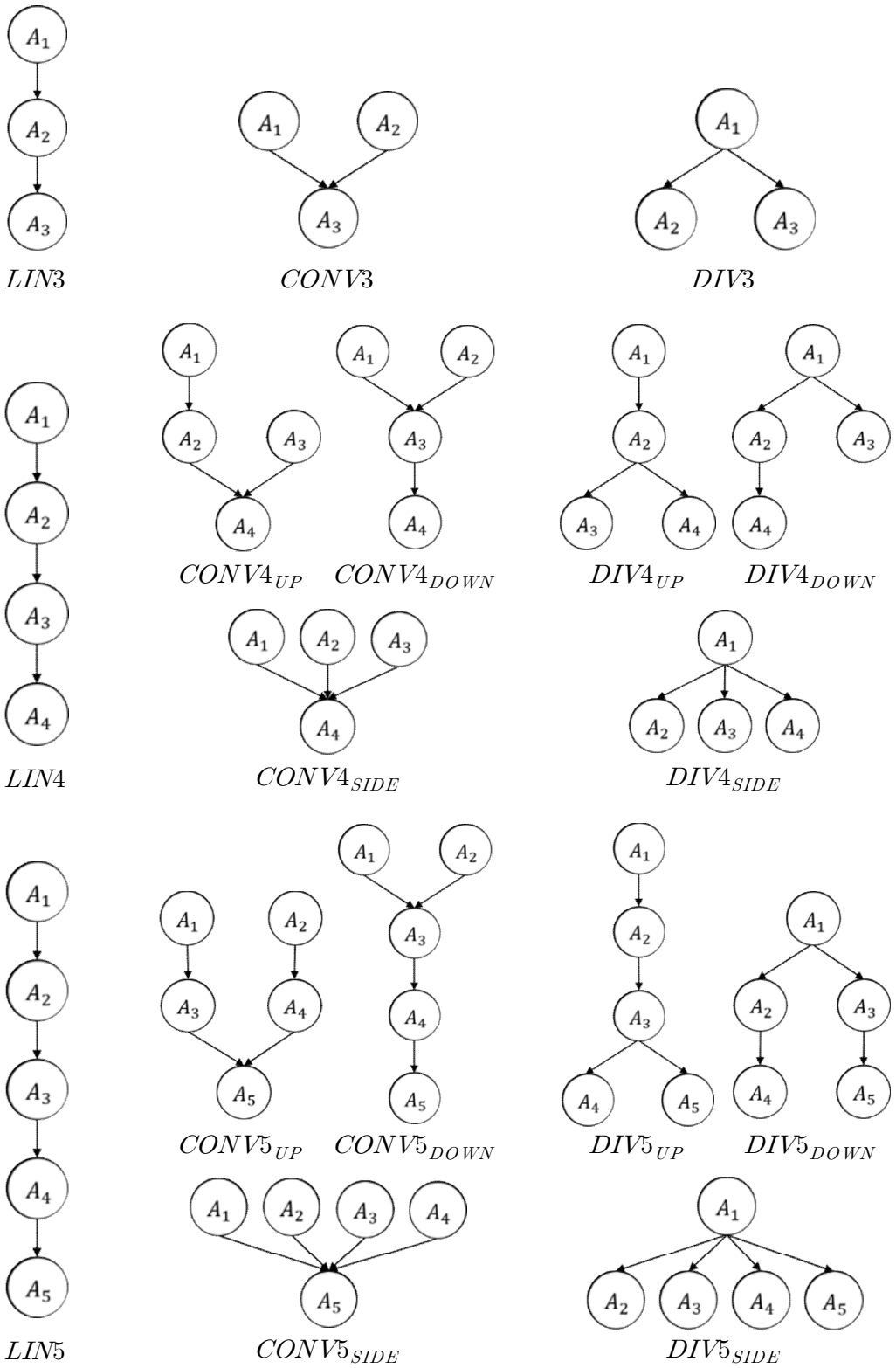


(a) 선형 (linear) (b) 수렴형 (convergent) (c) 발산형 (divergent)

[그림 Ⅳ-8] 모의실험 2에서 고려한 인지요소 위계 구조의 유형

하는 경우를 모두 고려하였다. 인지요소의 수는 3개, 4개, 5개인 경우를 가정하였으며, 인지요소의 수가 동일한 구조들은 유향 변의 수도 동일하다. 최종적으로 본 모의실험에서 설정된 인지요소 위계 구조는 [그림 Ⅳ-9]와 같다. 설정된 위계 구조는 총 17개로, 인지요소의 수가 3인 위계 구조 3개, 인지요소의 수가 4와 5인 위계 구조가 각 7개씩이다.

[그림 Ⅳ-9]의 위계 구조의 명칭에서 *LIN*은 선형 위계 구조를, *CONV*는 수렴형 위계 구조를, *DIV*는 발산형 위계 구조를 각각 나타내고, 그 옆에 인지요소의 수를 붙여서 나타내었다. 또한 UP, DOWN, SIDE와 같은 첨자는 어느 방향으로 유향 변을 추가한 것인지를 나타내는 표기로, 예를 들어 $CONV4_{UP}$은 *CONV*3 구조의 위쪽으로 인지요소와 유향 변을 하나 추가한 구조임을 나타낸다.



[그림 Ⅳ-9] 모의실험 2에서 가정한 인지요소 위계 구조

모의실험 2에서 Q-행렬은 복합 Q-행렬로 고정하였다. 모의실험 1에서와 마찬가지 방식으로 Madison과 Bradshaw(2015)의 권고를 따라 인지요소 1개를 측정하는 문항은 인지요소별로 두 문항씩 포함되도록 하고, 인지요소 2개와 3개를 측정하는 문항은 가능한 모든 인지요소 조합이 한 번씩 포함되도록 복합 Q-행렬을 구성하였다. 모의실험 2에서 가정한 Q-행렬은 [그림 Ⅳ-10]과 같다.

	A_1	A_2	A_3
I1	1	0	0
I2	0	1	0
I3	0	0	1
I4	1	0	0
I5	0	1	0
I6	0	0	1
I7	1	1	0
I8	1	0	1
I9	0	1	1
I10	1	1	1

	A_1	A_2	A_3	A_4
I1	1	0	0	0
I2	0	1	0	0
I3	0	0	1	0
I4	0	0	0	1
I5	1	0	0	0
I6	0	1	0	0
I7	0	0	1	0
I8	0	0	0	1
I9	1	1	0	0
I10	1	0	1	0
I11	1	0	0	1
I12	0	1	1	0
I13	0	1	0	1
I14	0	0	1	1
I15	1	1	1	0
I16	1	1	0	1
I17	1	0	1	1
I18	0	1	1	1

	A_1	A_2	A_3	A_4	A_5
I1	1	0	0	0	0
I2	0	1	0	0	0
I3	0	0	1	0	0
I4	0	0	0	1	0
I5	0	0	0	0	1
I6	1	0	0	0	0
I7	0	1	0	0	0
I8	0	0	1	0	0
I9	0	0	0	1	0
I10	0	0	0	0	1
I11	1	1	0	0	0
I12	1	0	1	0	0
I13	1	0	0	1	0
I14	1	0	0	0	1
I15	0	1	1	0	0
I16	0	1	0	1	0
I17	0	1	0	0	1
I18	0	0	1	1	0
I19	0	0	1	0	1
I20	0	0	0	1	1
I21	1	1	1	0	0
I22	1	1	0	1	0
I23	1	1	0	0	1
I24	1	0	1	1	0
I25	1	0	1	0	1
I26	1	0	0	1	1
I27	0	1	1	1	0
I28	0	1	1	0	1
I29	0	1	0	1	1
I30	0	0	1	1	1

[그림 Ⅳ-10] 모의실험 2에서 가정한 복합 Q-행렬

‘문항의 질’, ‘표본 크기(피험자 수)’, ‘문항 수’ 요인의 경우, 앞에서 설정된 다양한 위계 구조 하에서 일정 수준 이상의 위계 추정 정확도에 도달하기 위해 어느 정도의 문항 수와 표본 크기가 필요한지 확인하는 것을 목적으로 설정된 요인이다.

‘문항의 질’ 요인의 경우 높음과 낮음의 두 수준을 가정하였으며, 결과 해석의 명료화를 위해 문항모수의 변이성을 통제하고 문항의 질이 높은 경우는 모든 문항의 모수를 0.15로, 낮은 경우는 0.3으로 고정하였다.

‘표본 크기(피험자 수)’의 경우 모의실험 1에서 고려한 200, 500, 1000의 조건과 함께 2000인 경우도 실험 조건으로 고려하였다. 이는 문항의 질이 낮은 경우에는 표본 크기가 1000명 정도이더라도 일정 수준 이상의 위계 추정 정확도에 도달하지 못하는 경우가 있을 것이라는 예상 하에 설정된 수치이다.

‘문항 수’는 적음과 많음의 두 수준을 가정하였다. 문항의 수가 적은 경우로는 [그림 Ⅳ-10]의 Q-행렬을 가정하였으며, 이때 문항의 수는 인지요소가 3개, 4개, 5개인 경우에 각각 10문항, 18문항, 30문항이다. 그리고 이들을 각각 2배하여 문항의 수가 많은 경우를 설정하였다(20문항, 36문항, 60문항).

정리하면 총 모의실험 조건은 ‘인지요소 위계 구조’ 17 × ‘문항의 질’ 2 × 표본 크기(피험자 수)‘ 4 × ’문항 수‘ 2의 272가지이며, 각 조건별로 100번씩 반복하여 총 27200개의 모의 자료를 생성하였다. 이때 피험자의 진 인지요소패턴은 각 위계 구조 하에서 가능한 인지요소패턴들 중에서 무선 추출하였고, 생성된 인지요소패턴을 기반으로 문항 응답자료가 생성되었다. 이렇게 생성된 자료에 대해 SEM-H 방법을 적용하여 위계 추정 결과를 도출하였다.

평가 지표는 모의실험 1과 동일하게 TPER, TDER, r-TDER의 세 가지를 사용하였다.

3.2.2. 모의실험 결과

모의실험 2에서는 위계 구조의 유형에 따라 SEM-H 방법의 위계 추정 결과가 어떻게 달라지는지와 각 위계 구조 하에서 문항의 질이 높을 때와 낮을 때 각각 어느 정도의 문항 수와 표본 크기가 확보되어야 SEM-H 방법을 이용하여 일정 수준 이상의 위계 추정 정확도에 도달할 수 있는지에 초점을 맞추어 결과를 살펴보았다.

3.2.2.1. 인지요소의 수가 3개인 위계 구조에 대한 위계 추정 결과

<표 Ⅳ-9>는 인지요소의 수가 3개인 위계 구조들에 대한 위계 추정 결과이며, 그 결과를 그래프로 나타내면 [그림 Ⅳ-11]과 같다.

진 위계 구조의 위계 관계를 얼마나 잘 추정하였는지를 나타내는 TPER의 경우, 문항의 질이 높을 때($s, j = 0.15$)는 문항 수와 표본 크기에 관계없이 모든 위계 구조에 대해 0.9 이상으로 높게 나타나는 반면, 문항의 질이 낮을 때($s, j = 0.3$)는 위계 구조의 유형에 따라 서로 다른 결과가 나타난다.

문항의 질이 낮은 조건에서, 선형 구조 *LIN*3의 경우 문항의 수가 10문항으로 적을 때는 표본 크기가 500 정도 되어야 0.8 이상의 TPER이 도출되고 표본 크기가 1000 이상이면 0.9 이상의 값을 나타내며, 문항의 수가 20문항으로 많을 때는 표본 크기가 200 정도만 되어도 0.8 이상의 값을 얻을 수 있는 것으로 나타난다. 수렴형 구조 *CONV*3에 대해서는 문항의 수가 10문항으로 적을 때는 표본 크기가 2000 정도 되어야 0.8 이상의 TPER이 나타나지만, 문항의 수가 20문항으로 많아지면 표본 크기가 500 정도만 되어도 0.9 이상의 값을 얻을 수 있다. 한편, 발산형 구조에 대한 TPER은 다른 두 구조의 경우와 다른 경향을 나타낸다. 발산형 구조 *DIV*3의 경우 문항의 수가 10문항으로 적을 때는 표본 크기가 1000인 경우 TPER이 0.67로 나타나며, 표본 크기가 2000으로 커져도 그 값이 더 이상 증가하지 않는다. 문항의 수가 20문항으로 많을 때에는 표본 크기가 커짐에 따라 TPER이 증가하며, 표본 크기가 1000인 경우 0.8

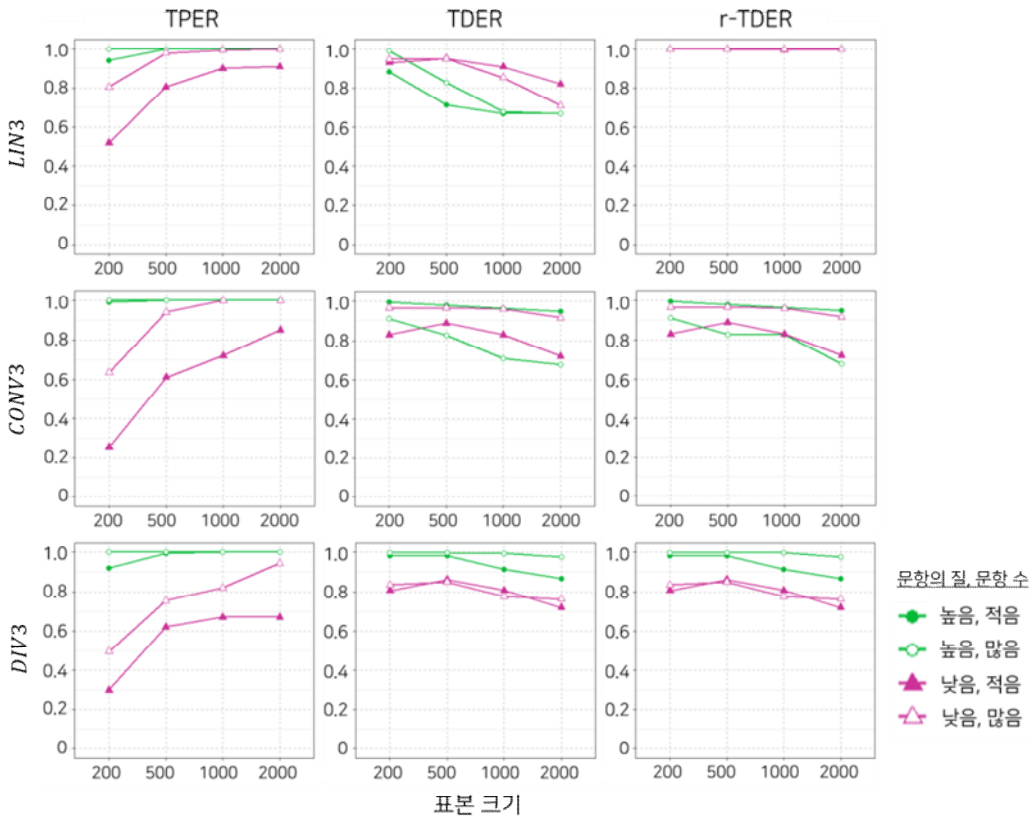
이상의 값을 나타낸다. 이러한 결과는 발산형 구조에 대한 위계 추정에서는 일정 수 이상의 문항이 확보되지 않으면 표본 크기가 크더라도 높은 TPER을 얻지 못할 수 있음을 보여준다.

추정된 위계 구조에 진 위계 관계가 얼마나 포함되어 있는지를 나타내는 TDER은 TPER과 달리 대부분의 조건에서 대체로 표본 크기가 커짐에 따라 오히려 그 값이 낮아지는 양상을 보인다. 진 간접 위계 관계를 포함하고 있는 선형 구조 *LIN*3의 경우 r-TDER이 대부분 1을 나타낸다는 점에 비추어볼 때 표본 크기가 커짐에 따라 진 간접 위계 관계가 추정되는 경우가 많아지면서 TDER이 낮아지는 결과가 나타난 것이라 할 수 있다. *CONV*3 구조의 경우, 문항의 질이 높고 문항의 수가 많은 경우에 표본 크기가 1000 이상이 되면 오히려 TDER이 0.8 미만으로 떨어지며, 문항의 질이 낮고 문항의 수가 적은 경우에는 표본 크기가 2000일 때 TDER이 0.72로 표본 크기가 작을 때보다도 낮게 나타난다. 반면, 문항의 질이 높고 문항 수가 10문항으로 적은 경우와 문항의 질이 낮고 문항 수가 20문항으로 많은 경우에는 표본 크기가 커짐에 따라 TDER이 낮아지기는 하지만 모두 0.9 이상으로 높은 값을 나타낸다. 이러한 결과는 수렴형 구조에서 문항의 질이 높아지거나 문항의 수가 많아진다고 하여 반드시 TDER이 높아지는 것은 아니라는 점을 보여준다. *DIV*3 구조의 경우, 문항의 수 보다는 문항의 질에 따라 TDER의 변화가 크게 나타난다. <표 Ⅳ-9>에서 동일한 문항 수 조건 하에서 문항의 질이 높은 경우와 낮은 경우의 *DIV*3 구조에 대한 TDER 결과를 비교해보면 문항의 질이 낮아짐에 따라 TDER이 약 0.1~0.2 정도 낮아짐을 확인할 수 있다. 따라서 발산형 구조에 대한 위계 추정에서는 보다 높은 TDER을 얻기 위해서는 양질의 문항이 확보되어야 할 것으로 보인다.

한편, 동일한 조건 하에서 세 위계 구조 유형의 결과를 비교해보면, 대체로 선형 구조의 위계 추정 결과가 나머지 두 구조에 비해 높게 나타나며, 수렴형 구조와 발산형 구조 중에는 수렴형 구조의 위계 추정 정확도가 높게 나타나는 경우가 더 많은 것으로 보인다.

<표 Ⅳ-9> 인지요소 3개인 위계 구조에 대한 위계 추정 결과

평가 지표	위계 구조	문항의 질 '높음' ($s, g = 0.15$)								문항의 질 '낮음' ($s, g = 0.30$)							
		문항 수 '적음' ($J = 10$)				문항 수 '많음' ($J = 20$)				문항 수 '적음' ($J = 10$)				문항 수 '많음' ($J = 20$)			
		200	500	1000	2000	200	500	1000	2000	200	500	1000	2000	200	500	1000	2000
TPER	*LIN*3	0.940	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.520	0.805	0.900	0.910	0.805	0.980	0.995	1.000
	*CONV*3	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.255	0.610	0.720	0.850	0.635	0.940	1.000	1.000
	*DIV*3	0.920	0.995	1.000	1.000	1.000	1.000	1.000	1.000	0.300	0.620	0.670	0.670	0.495	0.755	0.820	0.945
TDER	*LIN*3	0.882	0.713	0.667	0.667	0.990	0.827	0.677	0.667	0.930	0.952	0.908	0.820	0.948	0.950	0.852	0.707
	*CONV*3	0.993	0.980	0.963	0.947	0.910	0.827	0.707	0.677	0.830	0.888	0.830	0.720	0.964	0.965	0.960	0.917
	*DIV*3	0.982	0.983	0.913	0.867	0.997	0.997	0.993	0.977	0.806	0.862	0.808	0.719	0.836	0.849	0.777	0.762
r-TDER	*LIN*3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
	*CONV*3	0.993	0.980	0.963	0.947	0.910	0.827	0.707	0.677	0.830	0.888	0.830	0.720	0.964	0.965	0.960	0.917
	*DIV*3	0.982	0.983	0.913	0.867	0.997	0.997	0.993	0.977	0.806	0.862	0.808	0.719	0.836	0.849	0.777	0.762



[그림 Ⅳ-11] 인지요소 3개인 위계 구조에 대한 위계 추정 결과의 그래프

3.2.2.2. 인지요소의 수가 4개인 위계 구조에 대한 위계 추정 결과

<표 Ⅳ-10>은 인지요소의 수가 4개인 위계 구조들에 대한 위계 추정 결과이며, 그 결과를 그래프로 나타내면 [그림 Ⅳ-12]와 같다.

TPER의 경우 문항의 질이 높을 때($s, j = 0.15$)는 $CONV4_{SIDE}$ 구조를 제외한 모든 경우에서 문항 수와 표본 크기에 관계없이 0.9 이상의 높은 값이 나타났다. $CONV4_{SIDE}$ 구조는 예외적으로 표본 크기가 200일 때 문항 수가 적은 경우와 많은 경우의 TPER이 각각 0.747, 0.807로 다소 낮게 나타나지만, 표본 크기가 500 이상이 되면 다른 구조들과 유사하게 1에 가까운 값을 보인다. 문항의 질이 낮은 경우($s, j = 0.3$)에는 위계 구조의 유형에 따라 결과가 다르게 나타난다.

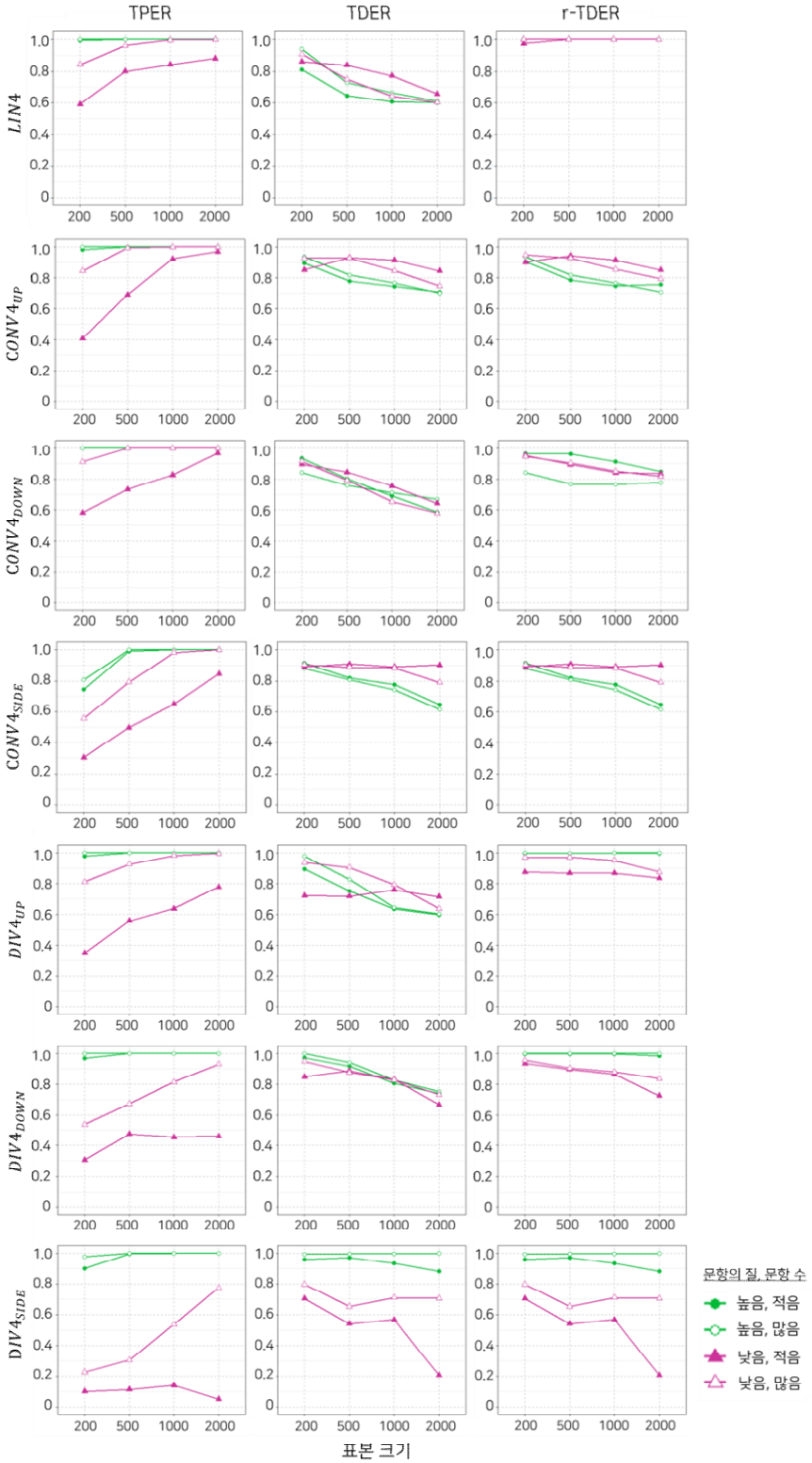
문항의 질이 낮은 조건에서, 선형 구조 LIN_4의 경우 문항의 수가 18문항으로 적은 경우에는 표본 크기가 500 정도 되어야 0.8 이상의 TPER을 보이는 반면, 문항의 수가 36문항이 되면 표본 크기가 200 정도만 되어도 0.8 이상의 TPER을 얻을 수 있는 것으로 나타난다. 수렴형 구조의 경우, $CONV3$의 위쪽과 아래쪽에 유향 변이 각각 추가된 구조인 $CONV4_{UP}$과 $CONV4_{DOWN}$에 대해서는 문항의 수가 적을 때 표본 크기가 1000 정도 되면 0.8 이상의 TPER을 얻을 수 있는 것으로 나타난 반면, $CONV3$의 옆쪽에 유향 변이 추가된 $CONV4_{SIDE}$구조에 대해서는 표본 크기가 2000 정도 되어야 0.8 이상의 TPER을 얻을 수 있다. 문항의 수가 많은 경우에도 $CONV4_{SIDE}$ 구조가 $CONV4_{UP}$과 $CONV4_{DOWN}$ 구조에 비해 동일한 수준의 TPER에 도달하기 위해 더 많은 크기를 필요로 하는 것으로 나타난다. 발산형 구조는 선형, 수렴형 구조에 비해 상대적으로 낮은 TPER을 나타내며, 문항의 수가 적은 경우에는 표본 크기가 2000 정도 되어도 TPER이 0.8을 넘지 못한다. 특히, $DIV3$ 구조의 옆쪽에 유향 변이 추가된 $DIV4_{SIDE}$ 구조는 문항의 질이 높을 때는 0.9 이상의 TPER을 나타내는 반면 문항의 질이 낮아지면서 TPER이 현저히 떨어지는 결과를 보이며, 문항의 수가 적은 경우에는 심지어 거의 0에 가까운 TPER을 보인다. 이러한 결과는 $DIV4_{SIDE}$ 구조로부터 생성된 자료

에 대해 SEM-H 방법을 적용하여 위계 추정을 실시했을 때 구조 학습의 첫 단계에서 인지요소 사이의 관계를 단 하나도 추가하지 않은 경우의 BIC 점수가 가장 높게 계산되어 위계 관계를 전혀 추정하지 못하는 현상이 발생하기 때문인 것으로 확인된다. 하지만 문항의 수를 2배로 늘리면 표본 크기가 커짐에 따라 $DIV4_{SIDE}$의 TPER이 높아지는 결과가 나타나며 표본 크기 2000일 때 그 값은 0.773이었다. 이러한 결과는 $DIV4_{SIDE}$ 구조와 같이 한 인지요소의 후행 인지요소가 여러 개인 위계 구조를 잘 추정해내기 위해서는 문항의 질이 높아야 하며, 문항의 질이 낮을 때는 다른 구조들보다 더 많은 문항 수와 표본 크기가 확보되어야 함을 시사한다.

TDER의 경우 앞에서 살펴본 인지요소 수가 3개인 위계 구조의 경우에서와 유사하게 표본 크기가 커짐에 따라 그 값이 낮아지는 결과가 나타나며, 위계 구조 유형별 경향성도 인지요소 수가 3개인 위계 구조에서와 유사하게 나타난다. 선형 구조 $LIN4$의 경우 표본 크기가 커짐에 따라 TDER이 낮아지는 결과가 나타나지만 r-TDER이 대부분의 조건에서 1인 것으로 볼 때 선형 구조에서는 진 위계 관계와 더불어 진 간접 위계 관계만을 함께 추정하는 것을 알 수 있다. 다시 말해, 선형 구조에 대해서는 위계 관계의 방향을 반대로 추정하는 경우는 거의 나타나지 않는 것이다. 수렴형 구조들에서는 앞에서와 마찬가지로 문항의 질이 낮아질 때 TDER이 높아지는 결과가 나타나기도 하며(문항 수가 36문항으로 많을 때 $CONV4_{DOWN}$와 $CONV4_{SIDE}$구조의 경우), 문항의 수가 많아질 때 TDER이 낮아지기도 한다(문항의 질이 높을 때 $CONV4_{DOWN}$와 $CONV4_{SIDE}$ 구조의 경우, 문항의 질이 낮을 때 표본 크기 500 이상인 경우). 발산형 구조에서는 문항의 질이 낮을 때 $DIV4_{SIDE}$ 구조의 TDER이 다른 두 발산형 구조에 비해 낮게 나타나며, 문항의 수가 적은 경우에는 심지어 그 값들이 0.6 이하로 떨어진다. 이러한 결과는 $DIV4_{SIDE}$의 위계 추정에서 참이 아닌 위계 구조를 도출하는 경우를 줄이기 위해서는 문항의 질이 높아야 한다는 점을 시사한다.

<표 Ⅳ-10> 인지요소 4개인 위계 구조에 대한 위계 추정 결과

평가 지표	위계 구조	문항의 질 '높음' ($s, g = 0.15$)								문항의 질 '낮음' ($s, g = 0.30$)							
		문항 수 '적음' ($J = 18$)				문항 수 '많음' ($J = 36$)				문항 수 '적음' ($J = 18$)				문항 수 '많음' ($J = 36$)			
		200	500	1000	2000	200	500	1000	2000	200	500	1000	2000	200	500	1000	2000
TPER	*LIN*4	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.590	0.800	0.843	0.883	0.843	0.963	0.997	1.000
	$CONV4_{UP}$	0.983	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.410	0.690	0.923	0.970	0.843	0.993	1.000	1.000
	$CONV4_{DOWN}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.580	0.733	0.827	0.970	0.913	1.000	1.000	1.000
	$CONV4_{SIDE}$	0.747	0.990	1.000	1.000	0.807	1.000	1.000	1.000	0.303	0.497	0.650	0.847	0.557	0.793	0.983	1.000
	$DIV4_{UP}$	0.980	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.350	0.557	0.637	0.777	0.810	0.927	0.983	0.997
	$DIV4_{DOWN}$	0.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.307	0.473	0.453	0.460	0.537	0.673	0.813	0.930
	$DIV4_{SIDE}$	0.900	0.997	1.000	1.000	0.980	1.000	1.000	1.000	0.100	0.113	0.140	0.050	0.227	0.307	0.537	0.773
TDER	*LIN*4	0.811	0.643	0.605	0.600	0.943	0.728	0.663	0.608	0.862	0.840	0.772	0.655	0.910	0.748	0.641	0.601
	$CONV4_{UP}$	0.894	0.776	0.742	0.705	0.936	0.817	0.764	0.698	0.850	0.930	0.914	0.843	0.925	0.927	0.844	0.745
	$CONV4_{DOWN}$	0.936	0.803	0.691	0.584	0.843	0.759	0.713	0.669	0.896	0.845	0.754	0.641	0.915	0.791	0.653	0.578
	$CONV4_{SIDE}$	0.918	0.819	0.776	0.645	0.883	0.807	0.744	0.615	0.892	0.909	0.893	0.904	0.905	0.890	0.886	0.791
	$DIV4_{UP}$	0.894	0.748	0.633	0.598	0.978	0.826	0.644	0.602	0.721	0.716	0.756	0.713	0.938	0.905	0.792	0.636
	$DIV4_{DOWN}$	0.972	0.918	0.804	0.740	1.000	0.943	0.825	0.753	0.848	0.887	0.831	0.668	0.949	0.875	0.827	0.734
	$DIV4_{SIDE}$	0.960	0.973	0.936	0.881	0.993	0.998	0.998	1.000	0.703	0.542	0.568	0.208	0.796	0.653	0.709	0.704
r-TDER	*LIN*4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$CONV4_{UP}$	0.903	0.782	0.746	0.754	0.936	0.817	0.764	0.706	0.901	0.943	0.914	0.849	0.948	0.927	0.853	0.793
	$CONV4_{DOWN}$	0.966	0.962	0.913	0.847	0.843	0.767	0.763	0.777	0.955	0.894	0.842	0.835	0.947	0.904	0.851	0.816
	$CONV4_{SIDE}$	0.918	0.819	0.776	0.645	0.883	0.807	0.744	0.615	0.892	0.909	0.893	0.904	0.905	0.890	0.886	0.791
	$DIV4_{UP}$	0.996	0.998	1.000	0.997	1.000	0.998	1.000	1.000	0.875	0.868	0.868	0.835	0.971	0.973	0.952	0.876
	$DIV4_{DOWN}$	0.998	0.998	0.996	0.986	1.000	1.000	1.000	1.000	0.935	0.896	0.864	0.725	0.957	0.904	0.879	0.834
	$DIV4_{SIDE}$	0.960	0.973	0.936	0.881	0.993	0.998	0.998	1.000	0.703	0.542	0.568	0.208	0.796	0.653	0.709	0.704



[그림 Ⅳ-12] 인지요소 4개인 위계 구조에 대한 위계 추정 결과의 그래프

3.2.2.3. 인지요소의 수가 5개인 위계 구조에 대한 위계 추정 결과

<표 Ⅳ-11>은 인지요소의 수가 5개인 위계 구조들에 대한 위계 추정 결과이며, 그 결과를 그래프로 나타내면 [그림 Ⅳ-13]과 같다.

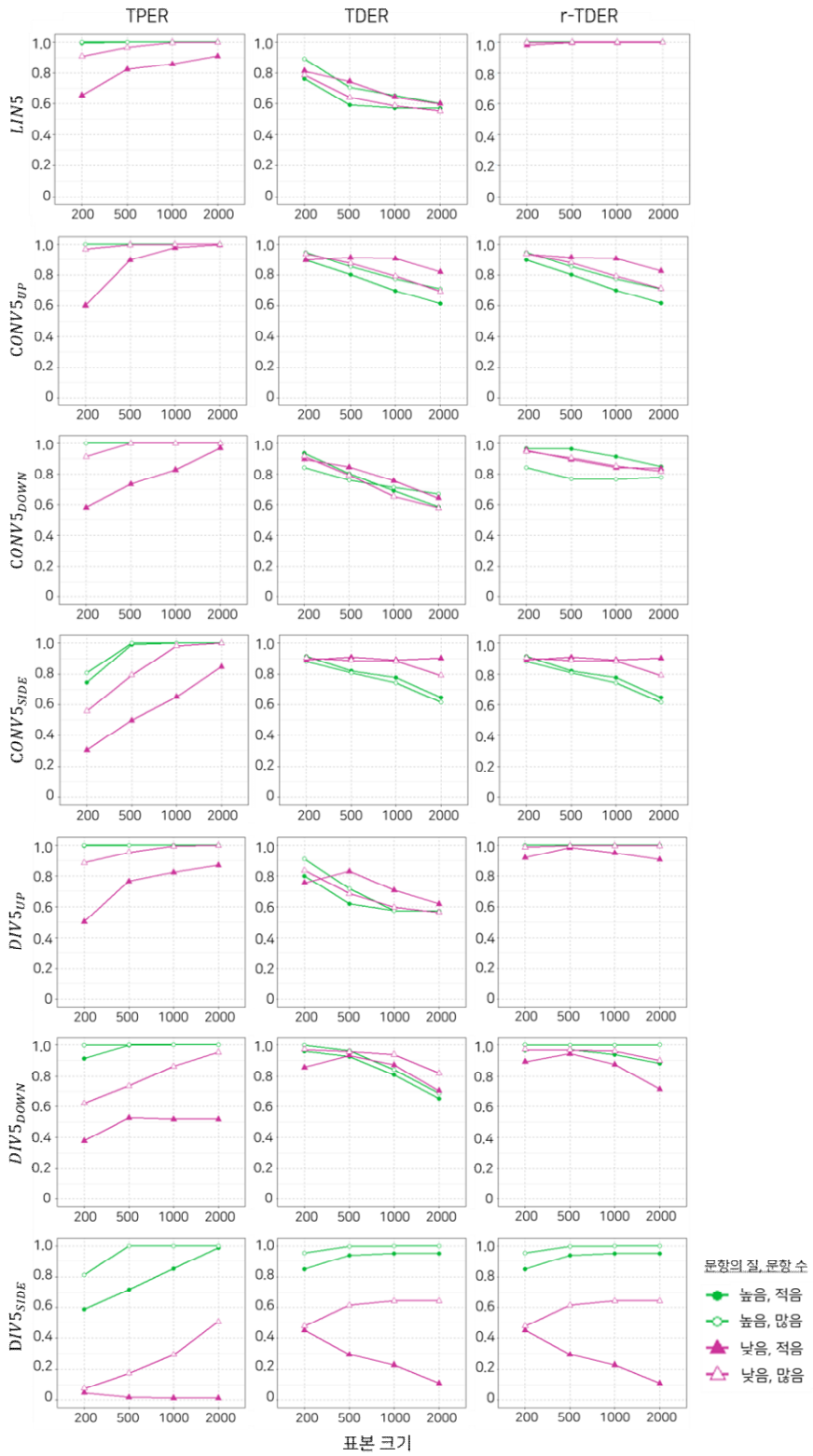
TPER의 경우, 문항의 질이 높을 때($s, j = 0.15$) $CONV5_{SIDE}$와 $DIV5_{SIDE}$ 구조에 대해서는 표본 크기가 1000 정도 확보되어야 0.8 이상의 값을 얻을 수 있는 반면, 나머지 구조에 대해서는 표본 크기 200 정도만으로도 0.9 이상의 값을 얻을 수 있는 것으로 보인다. 앞에서 살펴본 인지요소의 수가 4개인 위계 구조의 경우에서도 문항의 질이 높고 표본 크기가 작을 때 $CONV4_{SIDE}$와 $DIV4_{SIDE}$의 TPER이 다른 구조에 비해 살짝 떨어지는 결과가 나타났었는데(<표 Ⅳ-10>과 [그림 Ⅳ-12] 참고), 그 두 구조에 각각 유향변이 하나씩 더 추가됨에 따라 동일한 조건 하에서 $CONV5_{SIDE}$와 $DIV5_{SIDE}$의 TPER이 더 낮아진 것으로 보인다. 문항의 질이 낮은 경우($s, j = 0.3$)에는 앞에서와 마찬가지로 위계 구조 유형별로 TPER 결과에 차이가 나타난다. 문항이 질이 낮은 조건에서도 $CONV5_{SIDE}$와 $DIV5_{SIDE}$이 동일한 유형 내의 다른 구조들에 비해 TPER이 낮게 나타나는 것으로 보인다. $CONV5_{SIDE}$ 구조의 경우 문항의 수가 많을 때는 표본 크기가 2000에 도달하면 1에 가까운 TPER을 얻을 수 있으나, $DIV5_{SIDE}$ 구조의 경우 문항의 수가 많아지더라도 표본 크기 2000에서 TPER이 0.505로 낮게 나타난다. 이러한 결과를 앞에서 살펴본 $CONV3$, $DIV3$의 결과 및 $CONV4_{SIDE,}$ $DIV4_{SIDE}$의 결과와 연결해보면, 수렴형 구조나 발산형 구조에서 한 인지요소가 갖는 선행 인지요소나 후행 인지요소의 수가 많아질수록 SEM-H 방법으로 위계 관계를 잘 추정하기 위해 더 많은 수의 문항과 표본 크기가 요구됨을 알 수 있다.

TDER과 r-TDER의 경우, 표본 크기의 변화에 따른 경향성은 앞에서 살펴본 바와 동일하게 나타난다. 구조 유형에 따른 결과를 살펴보면, 선형 구조는 앞에서 살펴본 인지요소 3개, 4개인 경우와 유사한 결과가 도출되며, 수렴형 구조와 발산형 구조 중에서는 $CONV5_{DOWN}$과 $DIV5_{UP}$ 구조의 경우 TDER이 상대적으로 낮게 나타났으나 r-TDER을 함께 살펴

보면 진 간접 위계 관계를 추정한 경우가 있었던 것으로 보인다. $CONV5_{DOWN}$와 $DIV5_{UP}$ 구조의 경우 $CONV4_{DOWN}$과 $DIV4_{UP}$ 구조에 유향 변이 하나씩 각각 추가되면서 진 간접 위계 관계의 수가 3개씩 늘어나기 때문에([그림 Ⅳ-9] 참고) 이러한 결과가 나타난 것으로 보인다. 또한, 문항의 질이 낮을 때, $DIV5_{SIDE}$ 구조는 인지요소가 4개인 경우인 $DIV4_{SIDE}$ 구조와 마찬가지로 다른 구조들에 비해 현저히 낮은 TDER(r-TDER)을 나타난다. 문항의 수가 30문항으로 적은 경우에는 모든 표본 크기 조건에서 TDER(r-TDER)이 0.5를 넘지 못하며, 문항의 수가 60문항으로 많아지면 그 값이 증가하기는 하나 표본 크기가 2000이 되어도 그 값이 0.642로 다른 구조들의 r-TDER보다 낮게 나타난다. $DIV5_{SIDE}$와 같이 한 인지요소를 선행 인지요소로 갖는 여러 개의 후행 인지요소가 존재하는 구조의 위계 추정을 위해 높은 질의 문항이 요구된다는 점이 여기서도 다시 한 번 확인된다.

<표 Ⅳ-11> 인지요소 5개인 위계 구조에 대한 위계 추정 결과

평가 지표	위계 구조	문항의 질 '높음' ($s, g = 0.15$)								문항의 질 '낮음' ($s, g = 0.30$)							
		문항 수 '적음' ($J = 30$)				문항 수 '많음' ($J = 60$)				문항 수 '적음' ($J = 30$)				문항 수 '많음' ($J = 60$)			
		200	500	1000	2000	200	500	1000	2000	200	500	1000	2000	200	500	1000	2000
TPER	$LIN5$	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.653	0.823	0.858	0.910	0.908	0.965	0.998	1.000
	$CONV5_{UP}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.600	0.898	0.978	0.998	0.968	0.998	1.000	1.000
	$CONV5_{DOWN}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.638	0.793	0.900	0.975	0.935	1.000	1.000	1.000
	$CONV5_{SIDE}$	0.460	0.728	0.890	1.000	0.468	0.730	0.920	1.000	0.215	0.388	0.525	0.695	0.375	0.578	0.763	0.980
	$DIV5_{UP}$	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.503	0.763	0.823	0.870	0.885	0.955	0.993	1.000
	$DIV5_{DOWN}$	0.915	0.998	1.000	1.000	0.998	1.000	1.000	1.000	0.378	0.528	0.518	0.518	0.620	0.735	0.863	0.953
	$DIV5_{SIDE}$	0.588	0.718	0.853	0.988	0.810	1.000	1.000	1.000	0.048	0.018	0.013	0.013	0.075	0.175	0.293	0.505
TDER	$LIN5$	0.762	0.590	0.573	0.568	0.891	0.704	0.651	0.601	0.812	0.744	0.641	0.598	0.786	0.639	0.585	0.553
	$CONV5_{UP}$	0.899	0.802	0.696	0.613	0.947	0.853	0.773	0.708	0.897	0.914	0.906	0.820	0.937	0.876	0.791	0.693
	$CONV5_{DOWN}$	0.819	0.613	0.567	0.519	0.819	0.727	0.674	0.607	0.824	0.749	0.625	0.599	0.774	0.625	0.556	0.501
	$CONV5_{SIDE}$	0.828	0.837	0.741	0.655	0.896	0.901	0.831	0.767	0.764	0.810	0.759	0.658	0.869	0.842	0.801	0.719
	$DIV5_{UP}$	0.798	0.619	0.576	0.571	0.912	0.714	0.576	0.571	0.753	0.829	0.705	0.618	0.835	0.682	0.595	0.563
	$DIV5_{DOWN}$	0.959	0.926	0.809	0.651	0.998	0.963	0.842	0.686	0.857	0.933	0.872	0.703	0.971	0.956	0.939	0.820
	$DIV5_{SIDE}$	0.847	0.939	0.950	0.950	0.954	0.998	1.000	1.000	0.450	0.295	0.227	0.110	0.476	0.615	0.643	0.642
r-TDER	$LIN5$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.981	0.998	0.998	1.000	0.998	1.000	1.000	1.000
	$CONV5_{UP}$	0.899	0.802	0.698	0.616	0.947	0.853	0.773	0.708	0.936	0.915	0.906	0.825	0.939	0.880	0.791	0.712
	$CONV5_{DOWN}$	0.938	0.886	0.865	0.870	0.837	0.818	0.832	0.848	0.968	0.936	0.924	0.906	0.957	0.892	0.874	0.875
	$CONV5_{SIDE}$	0.828	0.837	0.741	0.655	0.896	0.901	0.831	0.767	0.764	0.810	0.759	0.658	0.869	0.842	0.801	0.719
	$DIV5_{UP}$	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	0.920	0.984	0.950	0.906	0.987	0.996	0.996	0.998
	$DIV5_{DOWN}$	0.966	0.969	0.939	0.884	1.000	0.998	0.998	1.000	0.893	0.945	0.875	0.715	0.971	0.967	0.960	0.901
	$DIV5_{SIDE}$	0.847	0.939	0.950	0.950	0.954	0.998	1.000	1.000	0.450	0.295	0.227	0.110	0.476	0.615	0.643	0.642



[그림 Ⅳ-13] 인지요소 5개인 위계 구조에 대한 위계 추정 결과의 그래프

4. 수학 검사 자료에 대한 SEM-H의 적용

이 절에서는 실제 수학 검사 자료에 SEM-H 방법을 적용한 위계 추정이 구체적으로 어떠한 방식으로 이루어질 수 있는지 확인하는 것을 목적으로 한다. 위계 추정에 사용된 자료는 대분수 뺄셈 검사 자료와 일차방정식과 일차부등식 검사 자료로, 대분수 뺄셈 검사 자료는 K. K. Tatsuoka(1984)에서 수집된 이래 여러 인지진단평가 관련 연구들에서 다양한 분석에 활용되어온 고전적인 검사 자료이며, 일차방정식과 일차부등식 검사 자료는 윤지영(2015)에서 우리나라 중학교 7~9학년을 대상으로 비교적 최근에 수집된 검사 자료이다. 이하에서는 각 자료를 소개하고, 각 자료에 SEM-H 방법을 적용하여 위계를 추정하는 구체적인 과정과 결과를 살펴본다.

4.1. 분수 뺄셈 검사

4.1.1. 분석 대상 자료

분수 뺄셈 검사 자료는 K. K. Tatsuoka(1984)에서 수집된 자료의 일부로 C. Tatsuoka(2002)에서 공개된 자료이다. K. K. Tatsuoka(1984)에서 원 자료가 수집될 당시 연구의 목적은 대분수의 덧셈과 뺄셈 문제에서 중학생들이 보이는 다양한 오개념들을 분석하는 것이었다. 이에 따라 수집된 원 자료는 대분수 덧셈에 관한 48문항과 대분수 뺄셈에 관한 42문항에 대한 7~9학년 학생 총 592명의 응답 자료로, 모든 문항은 두 대분수의 덧셈 또는 뺄셈의 결과를 묻는 단답형 문항들이었다. K. K. Tatsuoka(1990)는 원 자료 중 뺄셈에 대한 40문항에 관한 응답 자료만 추출하여 규칙장 모형을 이용한 분석에 사용하였으며, C. Tatsuoka(2002)는 K. K. Tatsuoka(1990)의 자료 중 일부인 20문항에 대한 응답 자료만을 추출하여 분석에 사용하고 자료를 공개하였다. 이후

그 자료가 “분수 뺄셈 검사 자료(fraction subtraction data)”로 불리면서 인지진단평가 관련 다양한 연구에서 활발하게 사용되게 되었다. 분수 뺄셈 검사 자료는 두 대분수의 뺄셈을 묻는 단답형 20문항에 대한 7~9학년 학생 총 536명의 응답이 각 학생이 각 문항의 정답을 맞힌 경우 1로, 오답한 경우 0으로 표기된 자료이다.

두 대분수의 뺄셈 방법은 크게 다음의 두 가지로 구분된다(Klein, Birnbaum, Standiford, & Tatsuoka, 1981):

방법 A: 대분수를 가분수로 변환하여 뺄셈을 한다.
방법 B: 대분수에서 자연수 부분과 분수 부분을 분리한 뒤 자연수는 자연수끼리, 분수는 분수끼리 뺄셈을 한다.

방법 A의 경우 대분수에 있는 자연수 부분을 모두 분수로 바꾸어 대분수를 가분수로 만드는 단계, 가분수끼리 뺄셈하는 단계, 그리고 최종 결과가 가분수이면 다시 대분수로 바꾸거나, 약분이 필요하면 약분을 하는 단계를 거쳐 분수의 뺄셈이 이루어지는 방법이며, 방법 B의 경우 대분수의 분수 부분끼리 뺄셈이 가능한지 확인하여 피감수의 분자가 더 작은 경우 자연수에서 1만큼을 분수 부분으로 빌려오는 단계, 대분수를 자연수와 분수로 분리하는 단계, 자연수는 자연수끼리 분수는 분수끼리 뺄셈하는 단계, 약분이 필요하면 약분하는 단계를 거쳐 분수의 뺄셈이 이루어지는 방법이다. 예를 들어, $4\frac{4}{12}-2\frac{7}{12}$와 같은 뺄셈의 경우, 방법 A로 풀이하면

$$4\frac{4}{12}-2\frac{7}{12}=\frac{52}{12}-\frac{31}{12}=\frac{21}{12}=1\frac{9}{12}=1\frac{3}{4}$$

와 같은 순서로 계산되며, 방법 B로 풀이하면,

$$4\frac{4}{12}-2\frac{7}{12}=3\frac{16}{12}-2\frac{7}{12}=(3-2)+(\frac{16}{12}-\frac{7}{12})=1+\frac{9}{12}=1\frac{3}{4}$$

와 같은 계산 순서에 따라 답을 얻는다.

K. K. Tatsuoka(1990)는 분수 뺄셈 검사를 위의 두 방법으로 풀이하는 경우를 각각 고려하여 두 가지 버전의 인지요소와 Q-행렬을 제시하

고 있으나, 분수 뺄셈 검사 자료를 활용한 대다수의 연구들은 방법 B에 대한 인지요소와 Q-행렬을 이용하여 분석을 진행하였다(Mislevy, 1995; de la Torre & Douglas, 2004; de la Torre, 2008; de la Torre, 2009; Henson, et al., 2009; de la Torre, 2011; DeCarlo, 2011; Chiu, 2013; Chen, Liu, Xu, & Ying, 2015; de la Torre & Chiu, 2016). 이에 따라 본 연구에서도 방법 B로 분수 뺄셈 검사를 풀이하는 경우에 대한 인지요소와 Q-행렬을 바탕으로 분석을 진행하고자 하였다. K. K. Tatsuoka(1990)에 제시된 분수 뺄셈 검사의 방법 B에 대한 인지요소는 [그림 Ⅳ-14]와 같으며, 검사 문항[14]과 Q-행렬은 <표 Ⅳ-12>와 같다.

A_1: 자연수를 분수로 변환하기
(Convert a whole number to fraction or mixed number)

A_2: 자연수를 분수로부터 분리하기
(Separate whole number from fraction)

A_3: 뺄셈 전에 분수를 간단하게 만들기
(Simplify before getting final number)

A_4: 공통분모 찾기
(Find the common denominator)

A_5: 자연수 부분에서 1만큼 빌려오기
(Borrow 1 from whole number part)

A_6: 첫 번째 분자를 두 번째 분자에서 빼기 위해 자연수에서 빌려오기
(Column borrow to subtract the second numerator from the first)

A_7: 답을 가장 간단한 형태로 나타내기
(Reduce answer to simplest form)

[그림 Ⅳ-14] 대분수 뺄셈 검사의 방법 B에 대한 인지요소
(K. K. Tatsuoka, 1990)

14) 검사 문항은 Chiu(2013)에서 발췌하였으며, 이 중 12번 문항의 경우 Chiu(2013)에는 $1\frac{1}{8}-\frac{1}{8}$로 제시되어 있으나, K. K. Tatsuoka(1984), K. K. Tatsuoka(1990) 등을 검토한 결과 $\frac{11}{8}-\frac{1}{8}$인 것으로 확인되었다.

<표 Ⅳ-12> 대분수 뺄셈 검사의 문항과 Q-행렬(K. K. Tatsuoka, 1990)

#	문항	A_1	A_2	A_3	A_4	A_5	A_6	A_7
1	$\frac{5}{3}-\frac{3}{4}$	0	0	0	1	0	1	0
2	$\frac{3}{4}-\frac{3}{8}$	0	0	0	1	0	0	0
3	$\frac{5}{6}-\frac{1}{9}$	0	0	0	1	0	0	0
4	$3\frac{1}{2}-2\frac{3}{2}$	0	1	1	0	1	0	0
5	$4\frac{3}{5}-3\frac{4}{10}$	0	1	1	1	0	0	1
6	$\frac{6}{7}-\frac{4}{7}$	0	0	0	0	0	0	0
7	$3-2\frac{1}{5}$	1	1	0	0	0	0	0
8	$\frac{2}{3}-\frac{2}{3}$	0	0	0	0	0	0	0
9	$3\frac{7}{8}-2$	0	1	0	0	0	0	0
10	$4\frac{4}{12}-2\frac{7}{12}$	0	1	1	0	1	1	1
11	$4\frac{1}{3}-2\frac{4}{3}$	0	1	1	0	1	0	0
12	$\frac{11}{8}-\frac{1}{8}$	0	0	1	0	0	0	1
13	$3\frac{3}{8}-2\frac{5}{6}$	0	1	0	1	1	0	0
14	$3\frac{4}{5}-3\frac{2}{5}$	0	1	0	0	0	0	0
15	$2-\frac{1}{3}$	1	1	0	0	0	0	0
16	$4\frac{5}{7}-1\frac{4}{7}$	0	1	0	0	0	0	0
17	$7\frac{3}{5}-\frac{4}{5}$	0	1	0	0	1	0	0
18	$4\frac{1}{10}-2\frac{8}{10}$	0	1	1	0	1	1	0
19	$4-1\frac{4}{3}$	1	1	1	0	1	0	0
20	$4\frac{1}{3}-1\frac{5}{3}$	0	1	1	0	1	0	0

본 연구의 연구자는 K. K. Tatsuoka(1990)의 인지요소와 Q-행렬을 일차적으로 검토한 결과, 일부 의미가 분명하지 않은 인지요소와 Q-행렬을 수정할 필요성이 있음을 확인하였다. 실제로 분수 뺄셈 검사를 활용한 여러 선행연구들도 K. K. Tatsuoka(1990)의 인지요소와 Q-행렬을 그대로 사용하지 않고 연구 목적에 따라 일부 수정하여 활용하고 있다. 이에 따라 본 연구에서는 분수 뺄셈 검사를 활용한 선행 연구들의 인지요소와 Q-행렬들과 각 문항에 대한 검토를 거쳐 인지요소와 Q-행렬을 재설정한 뒤 위계 추정을 진행하고자 하였다.

4.1.2. 분석 절차 및 방법

4.1.2.1. 인지요소 및 Q-행렬의 재설정

본 연구에서는 [그림 Ⅳ-14]와 <표 Ⅳ-12>에 제시된 K. K. Tatsuoka(1990)의 인지요소 및 Q-행렬에 대한 검토를 통해 수정이 필요한 부분을 다음과 같이 확인하였다.

첫째, 분모가 같은 분수끼리의 뺄셈(분모는 그대로 두고 분자끼리 빼기)에 대한 인지요소가 필요하다. 이는 분수 뺄셈 검사의 모든 문항의 해결에 필수적인 능력이므로 하나의 인지요소로 설정되어야 하며, 모든 문항이 이 인지요소를 측정하는 것으로 Q-행렬 또한 수정되어야 한다. <표 Ⅳ-12>의 Q-행렬에서 문항 6번이 측정하는 인지요소가 하나도 없는 것으로 표현되어 있다는 문제점도 발견되는데, 이 또한 분모가 같은 분수 뺄셈에 관한 인지요소가 부재하기 때문인 것으로 볼 수 있다. 선행연구들을 검토한 결과 Mislevy(1995), de la Torre(2008; 2009; 2011) 등에 'basic fraction subtraction'이라는 이름으로 분모가 같은 분수끼리의 뺄셈에 관한 인지요소가 설정되어 있는 것을 확인할 수 있었다.

둘째, 인지요소 A_3와 A_7은 유사한 능력을 나타내는 것으로 볼 수 있다. 인지요소 A_3의 원래 이름에 포함된 'simplify'와 A_7에 포함된 'reduce'는 분수 뺄셈의 맥락에서 둘 다 기약분수가 아닌 분수를 약분하여 기약분수로 변환하거나 가분수를 대분수로 변환하는 과정을 의미한다. 따라서 인지요소 A_3와 A_7은 simplify와 reduce 뒤의 표현에 의해 구분이 되는데, A_3의 이름 'simplify before getting final number'에서 final number는 분수 뺄셈의 결과로 나온 수를 의미하므로 A_3은 뺄셈을 하기 전에 문항에 포함된 분수를 약분하거나 대분수로 변환하는 과정을 나타내는 인지요소이며, 반면 A_7의 이름 'reduce answer to simplest form'에서 answer는 문항의 답을 뜻하므로 A_7은 뺄셈의 결과로 나온 수를 약분하거나 대분수로 만드는 과정을 뜻하는 인지요소로 볼 수 있다. 이 두 인지요소는 결국 뺄셈 문제를 해결하는 과정의 어느 단계에서 분수를 변환하는지에 따라 구분되는 인지요소인데, 두 인지요소 모두 '어떤 한 분수를 보고 변환이 필요함을 인지하고 올바르게 변환하기'와 같은 분수에 대한 조작 기능을 의미한다는 점에서는 유사한 능력을 나타낸다고도 볼 수 있다. 실제로 Mislevy(1995)와 de la Torre(2008; 2009; 2011)는 이 둘을 묶어서 'simplify/reduce'라는 하나의 인지요소로 설정하였음을 확인할 수 있었다.

셋째, 인지요소 A_5와 A_6의 차이가 불분명하다. <표 Ⅳ-12>의 Q-행렬에서 A_5나 A_6를 측정하는 것으로 표시된 문항들(1, 4, 10, 11, 13, 17, 18, 19, 20번)은 모두 피제수(첫 번째 수)의 분수 부분보다 제수(두 번째 수)의 분수 부분이 더 큰 문항들이라는 공통점이 있다. 그런데 그 문항들을 A_5만 측정하는 문항(4, 11, 13, 17, 19, 20번), A_6만 측정하는 문항(1번), A_5와 A_6를 모두 측정하는 문항(10, 18번)로 분류하여 서로 비교해 보았을 때, 인지요소 A_5와 A_6를 구분할만한 특징을 발견할 수 없었다. 즉, 두 번째 분자에서 첫 번째 분자를 빼기 위해 column borrow를 하는 것으로 표현된 인지요소 A_6와 자연수 부분에서 1을 빌려오는 것으로 표현된 A_5는 동일한 기능을 나타내는 것으로 보인다. 본 연구에서 검토한 선

행 연구들은 인지요소 A_6를 목록에서 제외하고 A_5만 인지요소로 설정하여 분석을 진행하였음이 확인되었다.

마지막으로, 분수 뺄셈 검사의 문항들 중 방법 B를 이용한 풀이를 하더라도 두 가지 이상의 풀이로 해결 가능한 문항들이 확인되었다. 예를 들어, 5번 문항($4\frac{3}{5}-3\frac{4}{10}$)의 풀이는 두 번째 분수를 약분한 뒤에 뺄셈을 하는 과정으로 해결하는 풀이 1과 공통분모를 찾아서 분모를 통일시킨 다음 뺄셈을 한 뒤 다시 약분하는 과정으로 해결하는 풀이 2로 나뉠 수 있다:

$$\text{풀이 1: } 4\frac{3}{5}-3\frac{4}{10}=4\frac{3}{5}-3\frac{2}{5}=(4-3)+(\frac{3}{5}-\frac{2}{5})=1\frac{1}{5}$$

$$\text{풀이 2: } 4\frac{3}{5}-3\frac{4}{10}=4\frac{6}{10}-3\frac{4}{10}=(4-3)+(\frac{6}{10}-\frac{4}{10})=1\frac{2}{10}=1\frac{1}{5}$$

풀이 1로 문항을 해결하기 위해서는 인지요소 A_3는 필요하지만 A_4와 A_7은 필요하지 않은 반면, 풀이 2로 문항을 해결하기 위해서는 인지요소 A_4와 A_7은 필요하지만 A_3는 필요하지 않다. 즉 어떤 방식으로 풀이하는지에 따라 문항 해결을 위해 필요한 인지요소가 달라진다. 5번 문항 이외에도, 가분수를 포함하고 있는 문항들(1, 4, 11, 12, 19, 20번)의 경우 가분수를 대분수로 바꾼 뒤 뺄셈을 하는 경우와 바꾸지 않고 뺄셈을 하는 경우에 따라 문항 해결을 위해 필요한 인지요소가 달라진다. 이러한 문항들을 분석에 포함시키려면 BN-H 모형의 측정모형으로 DINA 모형보다 더 일반적인 모형을 활용해야 하는데, 이는 본 연구에서 한정한 SEM-H 방법의 적용 범위를 벗어난다.

이와 같은 점들을 고려하여, 본 연구에서는 다음과 같은 순서로 문항과 인지요소를 재설정하였다.

1) 두 가지 이상의 풀이가 가능한 1, 4, 5, 11, 12, 19, 20번 문항을 제외하였다.
2) 인지요소 A_3와 A_7을 하나의 인지요소로 묶고자 하였으나, 1)에서 7개 문항을 제외하고 나면 A_3나 A_7을 필요로 하는 문항이 1개 밖에 남지 않아서 A_3와 A_7은 인지요소 목록에서 아예 삭제하였다.
3) 분모가 같은 진분수끼리의 뺄셈에 관한 인지요소를 추가하고, 인지요소 A_5와의 구분이 어려웠던 A_6를 인지요소 목록에서 삭제하였다.

이상의 과정을 거쳐 재구성된 분수 뺄셈 검사의 인지요소의 수는 5개, 문항은 총 13문항이며, 분석 자료 또한 13문항에 대한 응답 자료로 축소되었다. 다음으로, 분수 뺄셈 검사를 활용했던 선행 연구들에 대한 검토를 바탕으로 각 문항의 해결을 위해 필요한 인지요소들을 재점검하여 검사의 Q-행렬을 작성하였다. 다음으로 타당화 방법 중 경험적 δ 방법(de la Torre, 2008)을 적용하여[15] Q-행렬에 대한 수정 가능성을 검토하는 과정을 거쳐 최종적인 Q-행렬을 확정하였다. 최종적으로 분석에 사용된 Q-행렬과 인지요소는 <표 Ⅳ-13>과 같다(참고로 각 문항의 정답률을 <표 Ⅳ-13>의 마지막 열에 제시하였다).

15) 본 연구에서 경험적 δ 방법을 통한 Q-행렬 타당화는 R의 CDM 패키지 버전 6.1-10(Robitzsch, Kiefer, George, & Ünlü, 2018)을 이용하여 이루어졌다.

<표 Ⅳ-13> 본 연구에서 활용한 분수 뺄셈 검사의 인지요소, 문항 및 Q-행렬

#	문항	A_1	A_2	A_3	A_4	A_5	정답률
1	$\frac{3}{4}-\frac{3}{8}$	1	0	0	0	1	0.576
2	$\frac{5}{6}-\frac{1}{9}$	1	0	0	0	1	0.515
3	$\frac{6}{7}-\frac{4}{7}$	1	0	0	0	0	0.795
4	$3-2\frac{1}{5}$	1	1	0	1	0	0.399
5	$\frac{2}{3}-\frac{2}{3}$	1	0	0	0	0	0.748
6	$3\frac{7}{8}-2$	0	1	0	0	0	0.647
7	$4\frac{4}{12}-2\frac{7}{12}$	1	1	1	0	0	0.382
8	$3\frac{3}{8}-2\frac{5}{6}$	1	1	1	0	1	0.295
9	$3\frac{4}{5}-3\frac{2}{5}$	1	1	0	0	0	0.722
10	$2-\frac{1}{3}$	1	0	0	1	0	0.451
11	$4\frac{5}{7}-1\frac{4}{7}$	1	1	0	0	0	0.696
12	$7\frac{3}{5}-\frac{4}{5}$	1	1	1	0	0	0.433
13	$4\frac{1}{10}-2\frac{8}{10}$	1	1	1	0	0	0.468

A_1: 분모가 같은 두 분수를 빼기(분자끼리 빼기)
A_2: 자연수 부분과 분수 부분을 분리하기
A_3: 자연수 부분에서 분수 부분으로 1만큼 빌려오기
A_4: 자연수를 분수로 변환하기
A_5: 공통분모 찾기

4.1.2.2. SEM-H 방법의 적용

앞서 설정한 분수 뺄셈 검사의 5개 인지요소들이 위계 관계를 나타내는지를 확인하기 위해 분수 뺄셈 검사 자료에 SEM-H 방법을 적용하여 위계 추정을 시도하였다.

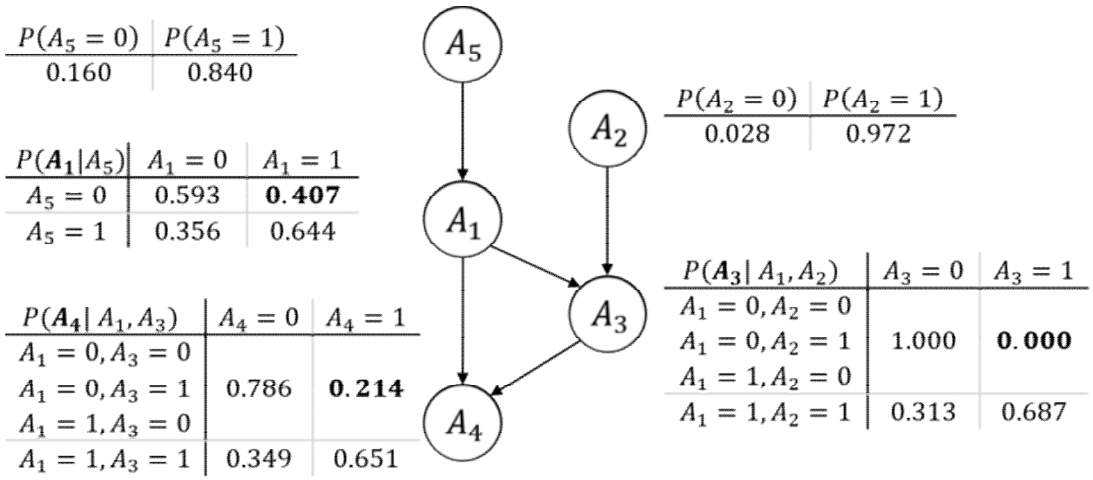
SEM-H는 BN의 구조 학습을 응용한 위계 추정 방법이다. 앞에서 설명하였듯이 BN의 구조 학습은 기본적으로 변수들 간의 종속성을 판단하여 구조를 찾는 방식을 따르기 때문에 SEM-H방법의 구조 학습(4단계까지)의 결과로 도출되는 구조는 반드시 위계 구조를 나타낸다고 볼 수는 없다. 도출된 구조에 포함된 종속 관계들이 실제로 위계 관계를 나타내는지를 판단하기 위해서는, 도출된 구조의 CPT에서 $P(A_j = 1 | \amalg Pa(A_j) = 0)$에 해당하는 모수 $\lambda_{j,0}$의 최종 추정치[16]가 설정된 $cutline$보다 작은 값을 나타내는지를 함께 확인해야 한다. 그런데 만약 A_j의 부모 인지요소가 둘 이상인 경우($|Pa(A_j)| \geq 2$) $\lambda_{j,0}$가 $cutline$을 넘는다고 하여 A_j로 향하는 모든 관계들이 위계 관계가 아니라고 판정할 수 없다. 이 경우에는 조금 더 복잡한 절차를 거쳐서 각각의 관계가 위계 관계인지의 여부를 확인해야 한다. 그 과정은 이 절의 실제 수학 검사 자료에 대한 분석에서 자세히 살펴볼 것이다.

또한, 모의자료를 이용한 위계 추정과 달리 실제 검사 자료를 이용하여 위계 추정을 할 때에는 인지요소들이 이루는 진 위계 관계를 알 수 없으므로, 검사의 내용 영역과 관련한 교육 과정이나 교과서, 피험자들의 실제 학습 과정 등을 바탕으로 위계 관계 여부를 판단할 필요가 있다. 분수 뺄셈 검사 자료에 대한 SEM-H 방법의 적용에서는 분수 뺄셈의 학습 순서와 연관지어 위계 추정의 결과를 살펴보고자 하였다.

16) 모수의 최종 추정치는 그 모수에 대한 사전 분포를 어떻게 설정하는지에 따라 달라질 수 있는데, 여기서는 인지요소들의 관계에 대한 어떠한 사전 정보도 없는 상태를 가정하고 균등분포를 부여하였다. 만약 위계에 관한 사전 정보가 있는 경우에는 그 정보를 반영한 베타 분포를 사전 분포로 설정할 수 있다.

4.1.3. 분석 결과

분수 뺄셈 검사 자료에 대해 SEM-H 방법을 적용하여 구조 학습(4단계)까지 진행한 결과, 총 6번의 SEM 사이클을 거쳐 다음의 [그림 Ⅳ-15]와 같은 구조와 모수들이 추정되었다.



[그림 Ⅳ-15] 분수 뺄셈 검사 자료에 대한 SEM-H의 구조 학습 결과

분수 뺄셈 검사에 대해 SEM-H 방법의 구조 학습 결과, 총 5쌍의 인지요소 관계($A_5 \rightarrow A_1$, $A_1 \rightarrow A_3$, $A_2 \rightarrow A_3$, $A_1 \rightarrow A_4$, $A_3 \rightarrow A_4$)를 갖는 구조가 도출되었다. SEM-H의 마지막 5단계에서는 도출된 위계 구조에서 자식 변수인 A_1, A_3, A_4의 CPT의 모수 추정치를 바탕으로 실제로 각 인지요소 관계들이 위계 관계를 나타내는지의 여부를 판정해야 한다. 그 과정이 실제로 어떻게 이루어지는지 구체적으로 살펴보도록 하자.

(1) $A_5 \rightarrow A_1$의 위계 관계 여부

$A_5 \rightarrow A_1$이 위계 관계라면 A_5의 숙달 없이는 A_1의 숙달이 불가능해야 하므로 $P(A_1=1|A_5=0)$의 값이 $cutline=0.01$보다 작게 추정되어야 한다. 그러나 [그림 Ⅳ-15]에 제시된 A_1의 CPT(왼쪽 가운데 표)에 따르면 그 값은 0.407로 추정되었다. 이는 0에 가까운 값이라고 보기 어

려우며, 따라서 $A_5 \rightarrow A_1$**은 위계 관계가 아닌 종속 관계를 나타낸다**고 판정할 수 있다. 그리고 Ⅲ장에서 살펴보았듯이 이 종속 관계는 부분 관계, 약한 선행관계, 상관 관계 중 하나를 나타내는 것이라 볼 수 있으며, 이 중 어느 관계를 나타내는지 결정하기 위해서는 두 인지요소의 의미를 함께 살펴볼 필요가 있다.

A_1은 '분모가 같은 두 분수를 빼기(분자끼리 빼기)'를 나타내는 인지요소이며, A_5는 '공통분모 찾기'를 나타내는 인지요소이다. 일단 우리나라 교육과정과 교과서를 기준으로 분수의 뺄셈에 대한 학습 순서를 살펴보면 분모가 같은 두 분수의 뺄셈을 먼저 학습한 뒤 분모가 다른 두 분수의 뺄셈을 위해 공통분모를 찾는 방법을 학습하는 순으로 이루어진다. 따라서 만약 A_1과 A_5 사이에 선행 관계가 있다면 학습의 순서를 고려하였을 때 선행 관계의 방향이 $A_5 \rightarrow A_1$이 아니라 $A_1 \rightarrow A_5$라고 보는 것이 더 적절할 수 있다. 그러나 SEM-H 방법의 4단계까지 적용한 결과로 $A_5 \rightarrow A_1$이 도출되었다는 점은 $P(A_1 = 1 | A_5 = 0)$이 $P(A_5 = 1 | A_1 = 0)$보다 작았다는 점을 함의하며, 이는 통계적으로는 $A_1 \rightarrow A_5$보다 $A_5 \rightarrow A_1$가 선행 관계를 나타낼 가능성이 높다는 점을 뜻한다. 인지요소들의 의미와 교육과정을 고려하였을 때 두 인지요소 사이의 관계의 방향과 자료에 기반한 통계적 분석의 결과로 도출된 관계의 방향이 서로 반대이므로 이 종속 관계를 선행 관계로 결정하기는 어렵다. 또한, 인지요소의 의미상 '분수가 같은 두 분수를 빼기'가 공통분모 찾기'의 하위 기능(subskill)이라고 보기는 어려우므로, 두 인지요소 사이의 관계는 부분 관계라고도 볼 수 없다. 따라서 **두 인지요소 A_1과 A_5 사이의 종속 관계는 상관 관계를 나타낸다**고 판정할 수 있으며, 이 상관 관계는 두 인지요소의 본질적인 특성이라기보다는 응답 자료로부터 추정된 인지요소패턴 자료들이 갖는 통계적 특성으로 보는 것이 적절하다.

A_1과 A_5이 상관 관계를 나타낸다고 할 때, 이 상관 관계가 인과 관계를 나타내는 것은 아니므로 사실 두 변수 사이에는 방향을 설정하지 않는 것이 더 적절할 수 있다. 그러나 베이지안 네트워크는 유향 비순

환 그래프를 가정하기 때문에 두 변수 사이에는 어느 방향으로든 방향이 설정되어야만 하며, 따라서 그 방향에 대한 결정이 필요하다. SEM-H의 4단계까지 적용한 결과 도출된 종속 관계는 $A_5 \rightarrow A_1$이었지만, 사실 이 방향의 관계가 도출된 이유는 SEM-H 방법이 '만약 둘 사이의 위계가 있는 경우 어느 방향이 더 적절할지'를 $P(A_1 = 1|A_5 = 0)$와 $P(A_5 = 1|A_1 = 0)$의 크기 비교를 통해 판정하여 $A_5 \rightarrow A_1$만 설정 가능한 것으로 제약했기 때문이라는 점에 있다. 그러나 둘 사이의 관계를 상관 관계로 결정한 경우, 그 관계의 방향은 $A_5 \rightarrow A_1$과 $A_1 \rightarrow A_5$ 중 어느 것으로 설정하여도 통계적으로는 문제가 없다. 따라서, 만약 더 많은 학습자들이 숙달한 것으로 나타난 인지요소를 구조의 상단에 표시하고 싶다면 처음에 도출된 $A_5 \rightarrow A_1$으로 관계의 방향을 나타낼 수 있으며, 학습 순서 상 먼저 배우는 내용에 대한 인지요소를 구조의 상단에 표시하고 싶다면 $A_1 \rightarrow A_5$로 방향을 설정할 수 있다. 본 연구의 연구자는 분수 뺄셈의 학습 순서를 반영하여 A_1과 A_5의 종속 관계를 $A_1 \rightarrow A_5$로 나타내기로 결정하였다.

(2) $A_1 \rightarrow A_3$, $A_2 \rightarrow A_3$의 위계 관계 여부

[그림 Ⅳ-15]에서 A_3의 CPT를 보면 인지요소 A_1이나 A_2를 숙달하지 않은 경우에 A_3를 숙달하였을 확률 $P(A_3 = 1|A_1A_2 = 0)$의 값이 0으로 추정되었으므로, SEM-H 방법의 적용 결과로 도출된 구조에서 **$A_1 \rightarrow A_3$와 $A_2 \rightarrow A_3$는 모두 위계 관계를 나타낸다**고 판정할 수 있다.

인지요소 A_1, A_2, A_3는 각각 '분모가 같은 두 분수를 빼기(분자끼리 빼기)', '자연수 부분과 분수 부분을 분리하기', '자연수 부분에서 분수 부분으로 1만큼 빌려오기'를 나타내는 인지요소이다. 먼저 인지요소 A_2와 A_3의 의미를 연관지어 살펴보면, '자연수 부분에서 분수 부분으로 1만큼 빌려오기(A_3)'의 목적이 '자연수 부분과 분수 부분을 분리하여(A_2)' 뺄셈을 하기 위함이라고 볼 수 있다. 따라서 A_3를 숙달한 학생은 A_2를 숙달하였다고 보는 것이 타당하다. 따라서 $A_2 \rightarrow A_3$는 인지요소들의 의미를 고려해봤을 때에도 위계 관계라고 판정하는 것이 타

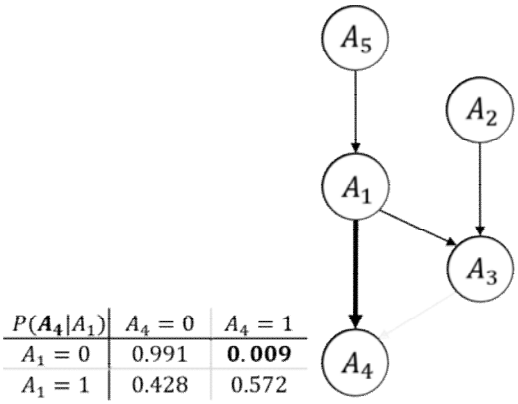
당하다고 볼 수 있다. 인지요소 A_1과 A_3의 경우도 마찬가지로, '자연수 부분에서 분수 부분으로 1만큼 빌려오기(A_3)'는 두 대분수의 뺄셈에서 첫 번째 분수(피감수)의 분수부분의 분자가 두 번째 분수(감수)의 분수 부분의 분자보다 작다는 것을 인지하고 분자끼리 빼기 위해 자연수 부분에서 1을 빌려오는 과정이므로, '분모가 같은 두 분수의 분자끼리 빼는(A_1)' 뺄셈을 숙달이 먼저 이루어진 경우에 '자연수 부분에서 분수 부분으로 1만큼 빌려오기(A_3)'를 숙달할 수 있다고 보는 것이 적절하다. 따라서 $A_1 \rightarrow A_3$의 위계 관계 또한 타당하다고 볼 수 있다.

(3) $A_1 \rightarrow A_4$, $A_3 \rightarrow A_4$의 위계 관계 여부

[그림 Ⅳ-15]의 A_4의 CPT에서 인지요소 A_1이나 A_3를 숙달하지 않은 경우에 A_4를 숙달하였을 확률 $P(A_4 = 1 | A_1 A_3 = 0)$은 0.214로, $cutline (= 0.01)$을 넘는 값으로 추정되었다. 그러나 이 확률은 두 관계 $A_1 \rightarrow A_4$과 $A_3 \rightarrow A_4$가 결합되어 추정된 값이므로, 그 값이 0에 가깝지 않다고 하여 두 관계가 모두 위계 관계가 아니라고 판정을 내릴 수는 없다. 실제로 두 관계 중 하나는 위계 관계이고 하나는 종속 관계인 경우가 있을 수 있기 때문이다. 따라서 이와 같은 경우에는 $A_1 \rightarrow A_4$와 $A_3 \rightarrow A_4$각각에 대해 위계 관계 여부를 판정해야 하며, 이를 위해 추정된 구조에서 $A_1 \rightarrow A_4$와 $A_3 \rightarrow A_4$중 위계 관계 여부를 판정하고자 하는 관계는 남겨두고 나머지 한 관계는 삭제한 뒤 다시 모수 추정을 실시하였다.

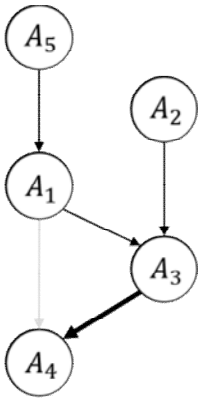
그 결과, 관계 $A_1 \rightarrow A_4$의 경우([그림 Ⅳ-16]의 (a) 참조) A_1을 숙달하지 않았을 때 A_4를 숙달할 확률 $P(A_4 = 1 | A_1 = 0)$의 추정값이 0.009로 $cutline (= 0.01)$을 넘지 않으므로 위계 관계라고 판정할 수 있는 반면, 관계 $A_3 \rightarrow A_4$의 경우([그림 Ⅳ-16]의 (b) 참조) A_3를 숙달하지 않았을 때 A_4를 숙달할 확률 $P(A_4 = 1 | A_3 = 0)$의 추정값이 0.149로 $cutline (= 0.01)$을 넘는 값이므로 위계 관계라고 보기 어렵다. 따라서 **$A_1 \rightarrow A_4$는 위계 관계, $A_3 \rightarrow A_4$는 위계 관계가 아닌 종속 관계를 나타낸다**고 판정할 수 있다.

(a) $A_1 \rightarrow A_4$의 위계 관계 여부 판정



$P(A_4 \mid A_1)$	$A_4 = 0$	$A_4 = 1$
$A_1 = 0$	0.991	**0.009**
$A_1 = 1$	0.428	0.572

(b) $A_3 \rightarrow A_4$의 위계 관계 여부 판정



$P(A_4 \mid A_3)$	$A_4 = 0$	$A_4 = 1$
$A_3 = 0$	0.851	**0.149**
$A_3 = 1$	0.349	0.651

[그림 Ⅳ-16] $A_1 \rightarrow A_4$와 $A_3 \rightarrow A_4$의 위계 관계 판정을 위한 모수 추정 결과

앞에서와 마찬가지로 인지요소들의 의미를 살펴보면, 인지요소 A_1과 A_4의 경우, A_4에 해당하는 '자연수를 분수로 변환하기'의 목적이 A_1에 해당하는 '분모가 같은 두 분수의 뺄셈(A_1)'을 하기 위함이므로 A_4를 숙달한 학생은 A_1을 이미 숙달하였을 것이라 볼 수 있다. 따라서 $A_1 \rightarrow A_4$를 위계 관계로 보는 것은 타당하다. 한편, 인지요소 A_3인 '자연수 부분에서 분수 부분으로 1만큼 빌려오기'은 대분수의 자연수 부분에서 1을 분수 부분으로 가져와서 분수로 변환하는 과정까지를 의미하므로 '자연수를 분수로 변환하기(A_4)'의 과정을 내포하고 있는 것처럼 보일 수 있다. 그러나 A_3와 A_4의 차이는 두 인지요소가 서로 다른 유형의 문항의 해결에 필요한 것이라는 점으로, A_3는 (대분수)-(대분수) 형태의 뺄셈 문항에, A_4는 (자연수)-(대분수) 형태의 뺄셈 문항에 각각 필요한 인지요소이다. 우리나라 2015 개정 교육과정에 따른 초등 교과서에서는 두 형태의 뺄셈을 동일한 학기에 학습하도록 되어 있으므로 둘 사이의 관계는 상관 관계로 보는 것이 적절하다. 따라서, 인지요소의 의미를 고려했을 때 **두 인지요소 A_3와 A_4 사이의 종속 관계는 상관 관계를 나타낸다**고 판정할 수 있다. 동일한 시기에 학습되는 두 인지요소의 학습 순서를 결정하기는 어려우므로, 두 인지요소의 관계의 방향은 처음에 도출된 대로 $A_3 \rightarrow A_4$으로 나타내기로 한다.

이상의 논의에 따라 처음에 SEM-H의 구조 학습 결과로 도출되었던 총 5쌍의 인지요소 관계($A_5 \rightarrow A_1$, $A_1 \rightarrow A_3$, $A_2 \rightarrow A_3$, $A_1 \rightarrow A_4$, $A_3 \rightarrow A_4$)에 대한 판정이 어떻게 이루어졌는지 정리해보면 다음과 같다:

$A_5 \rightarrow A_1$: 위계 관계가 아닌 종속 관계(상관 관계); $A_1 \rightarrow A_5$로 나타냄.

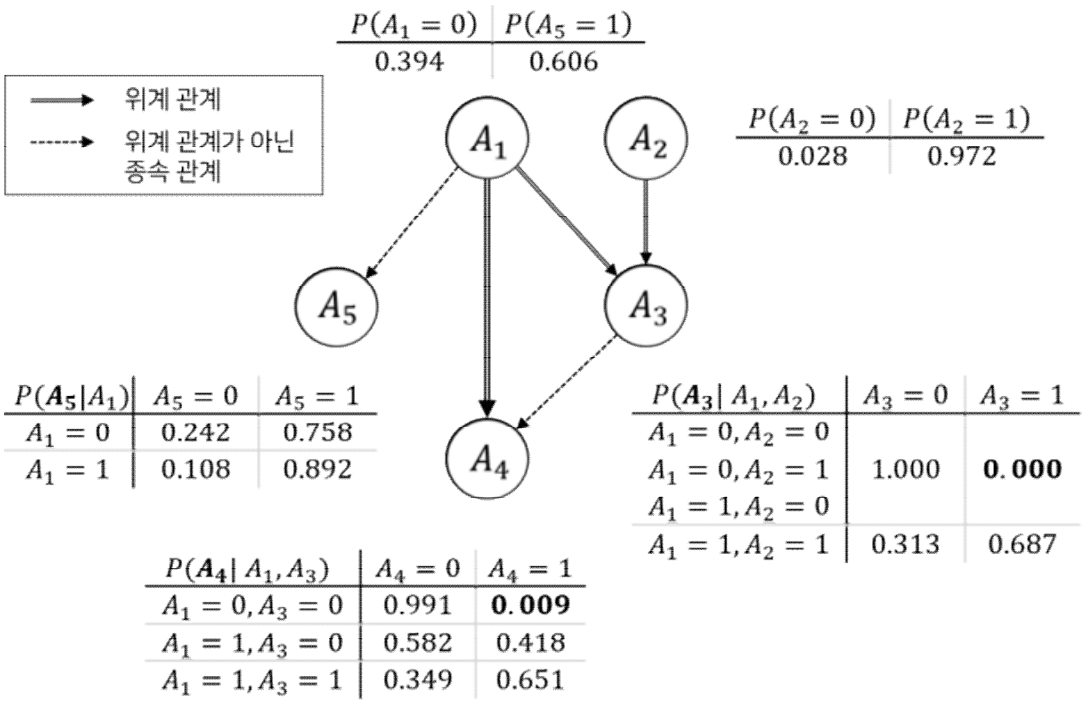
$A_1 \rightarrow A_3$: 위계 관계

$A_2 \rightarrow A_3$: 위계 관계

$A_1 \rightarrow A_4$: 위계 관계

$A_3 \rightarrow A_4$: 위계 관계가 아닌 종속 관계(상관 관계)

이 5쌍의 관계로 이루어진 최종적인 인지요소 위계 구조의 DAG는 [그림 Ⅳ-17]에 제시되어 있으며, 최종 구조에 대해 추정된 모수를 포함한 CPT도 함께 나타내었다.



$P(A_1 = 0)$	$P(A_5 = 1)$
0.394	0.606

$P(A_2 = 0)$	$P(A_2 = 1)$
0.028	0.972

$P(\boldsymbol{A_5}\|A_1)$	$A_5 = 0$	$A_5 = 1$
$A_1 = 0$	0.242	0.758
$A_1 = 1$	0.108	0.892

$P(\boldsymbol{A_3}\| A_1, A_2)$	$A_3 = 0$	$A_3 = 1$
$A_1 = 0, A_2 = 0$		
$A_1 = 0, A_2 = 1$	1.000	**0.000**
$A_1 = 1, A_2 = 0$		
$A_1 = 1, A_2 = 1$	0.313	0.687

$P(\boldsymbol{A_4}\| A_1, A_3)$	$A_4 = 0$	$A_4 = 1$
$A_1 = 0, A_3 = 0$	0.991	**0.009**
$A_1 = 1, A_3 = 0$	0.582	0.418
$A_1 = 1, A_3 = 1$	0.349	0.651

A_1: 분모가 같은 두 분수를 빼기(분자끼리 빼기)
A_2: 자연수 부분과 분수 부분을 분리하기
A_3: 자연수 부분에서 분수 부분으로 1만큼 빌려오기
A_4: 자연수를 분수로 변환하기
A_5: 공통분모 찾기

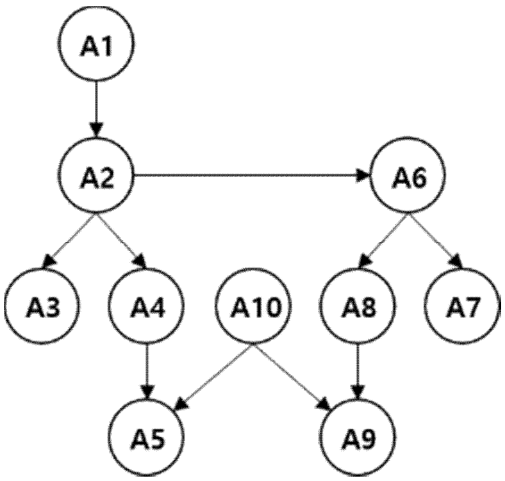
[그림 Ⅳ-17] 분수 뺄셈 검사에 대해 SEM-H 방법을 적용하여 최종적으로 추정된 인지요소 구조

4.2. 일차방정식과 일차부등식 검사

4.2.1. 분석 대상 자료

일차방정식과 일차부등식 검사 자료는 윤지영(2015)에서 우리나라 중학교 2~3학년 총 936명을 대상으로 수집된 검사자료로, 일차방정식과 일차부등식에 대한 5지 선다형 25문항에 대한 936명의 응답이 정답을 맞힌 경우 1로, 오답한 경우 0으로 표기된 자료이다. 윤지영(2015)은 검사 설계 단계에서부터 인지요소 간 위계 관계를 가정하고 이를 반영하여 검사를 제작하고 수집된 자료를 AHM으로 분석하였다. 윤지영(2015)에서 설정한 일차방정식과 일차부등식 검사의 10가지 인지요소와 인지요소 위계 구조는 [그림 Ⅳ-18]과 같으며, 검사의 Q-행렬은 <표 Ⅳ-14>에 제시되어 있다.

A1: 사칙연산 및 식의 이해
A2: 등식의 이해
A3: 일차방정식의 해의 의미
A4: 일차방정식 풀기
A5: 일차방정식 활용
A6: 부등식의 이해
A7: 일차부등식의 해의 의미
A8: 일차부등식 풀기
A9: 일차부등식 활용
A10: 대수적 모델링



[그림 Ⅳ-18] 일차방정식과 일차부등식 검사의 인지요소와 위계 구조(윤지영, 2015)

<표 Ⅳ-14> 일차방정식과 일차부등식 검사의 Q-행렬(윤지영, 2015)

문항	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
1	1	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	1	1	1	0	0	0	0	0	0	0
6	1	1	1	0	0	0	0	0	0	0
7	1	1	0	1	0	0	0	0	0	0
8	1	1	1	1	0	0	0	0	0	0
9	1	1	1	0	0	0	0	0	0	1
10	1	0	0	0	0	0	0	0	0	1
11	1	1	0	1	1	0	0	0	0	1
12	1	1	0	1	1	0	0	0	0	1
13	1	1	0	0	0	1	0	0	0	0
14	1	1	0	0	0	1	0	0	0	0
15	1	1	0	0	0	1	0	1	0	0
16	1	1	0	0	0	1	1	0	0	0
17	1	1	0	0	0	1	1	1	0	0
18	1	1	0	0	0	1	0	1	0	0
19	1	1	0	0	0	1	0	1	0	0
20	1	1	0	0	0	1	0	1	0	0
21	1	1	0	0	0	1	0	1	1	1
22	1	1	0	0	0	1	0	1	1	1
23	1	1	0	0	0	1	0	1	1	1
24	1	1	0	1	1	0	0	0	0	1
25	1	1	0	0	0	1	0	1	1	1

4.2.2. 분석 절차 및 방법

본 연구에서는 윤지영(2015)에서 설정한 인지요소와 Q-행렬을 그대로 활용하고자 하였으나, SEM-H의 적용을 어렵게 만드는 문제점이 확인되어 그에 따라 인지요소와 Q-행렬이 재설정되었다.

윤지영(2015)에서는 인지요소 위계 구조를 설정한 뒤 그 위계 구조를 반영하여 Q-행렬을 설정하는 방식을 취하였다. 이는 인지요소 위계라는 개념을 처음 정의하고 AHM을 제안한 Leighton 외(2004)에서 사용한 방식으로, 예를 들어 인지요소가 세 인지요소 A_1, A_2, A_3를 측정하는 검사에서 $A_1 \rightarrow A_2 \rightarrow A_3$라는 위계 관계가 있는 경우, 만약 어떤 문항이 인지요소 A_3만 측정하는 문항이라 할지라도 그 문항의 해결을 위해서 A_3뿐만 아니라 그 선행 인지요소인 A_1과 A_2도 필요하다고 보고 그 문항의 Q-벡터를 (0,0,1)이 아니라 (1,1,1)로 설정하는 방식이다. 그런데 이와 같이 이미 위계 구조가 반영되어 있는 Q-행렬을 이용하여 위계 구조를 추정하면 편향된 위계 추정 결과가 도출될 수 있다는 문제점이 있다. SEM-H 방법을 적용한 위계 추정은 인지요소들 간의 위계 관계에 대한 가정이 없는 상태에서 출발하여 이를 찾아나가는 과정이므로, 보다 정확한 위계 관계 추정을 위해서는 각 문항이 직접적으로 측정하는 인지요소만 1로 표시되어 있는 Q-행렬을 사용하여야 한다. 따라서 본 연구에서는 검사의 원 제작자를 포함한 수학교육 연구자들로 구성된 전문가 회의를 거쳐 전체 문항을 재검토하고 각 문항이 직접적으로 측정하는 인지요소만 나타내도록 Q-행렬을 재설정하였다. 또한, 그 과정에서 표본 크기에 비해 인지요소의 수가 많다는 점과 일부 인지요소의 의미가 분명하지 않다는 점이 문제점으로 제기되어 인지요소를 재설정하였으며, 삭제된 인지요소에 대한 문항과 두 가지 이상의 방법으로 풀이가 가능한 일부 문항들이 삭제되었다. 이러한 과정에 따라 일차방정식 및 일차부등식과 관련한 6개 인지요소와 이들을 측정하는 문항 15개로 이루어진 검사가 구성되었고, 검사에 대한 Q-행렬 초안을 작성한 뒤 경험적 δ 방법을 통한 Q-행렬 타당화 과정을 거쳐 최종 Q-행렬을 확정하였다. 분석에 활용된 인지

요소와 Q-행렬은 <표 Ⅳ-15>와 같고, 검사 문항은 부록에 제시되어 있다.

재구성된 검사와 Q-행렬을 바탕으로 일차방정식과 일차부등식 검사 자료에 SEM-H 방법의 구조 학습을 적용하여 구조를 도출하였으며, 모수 추정치를 이용하여 위계 관계 여부를 확인하였다.

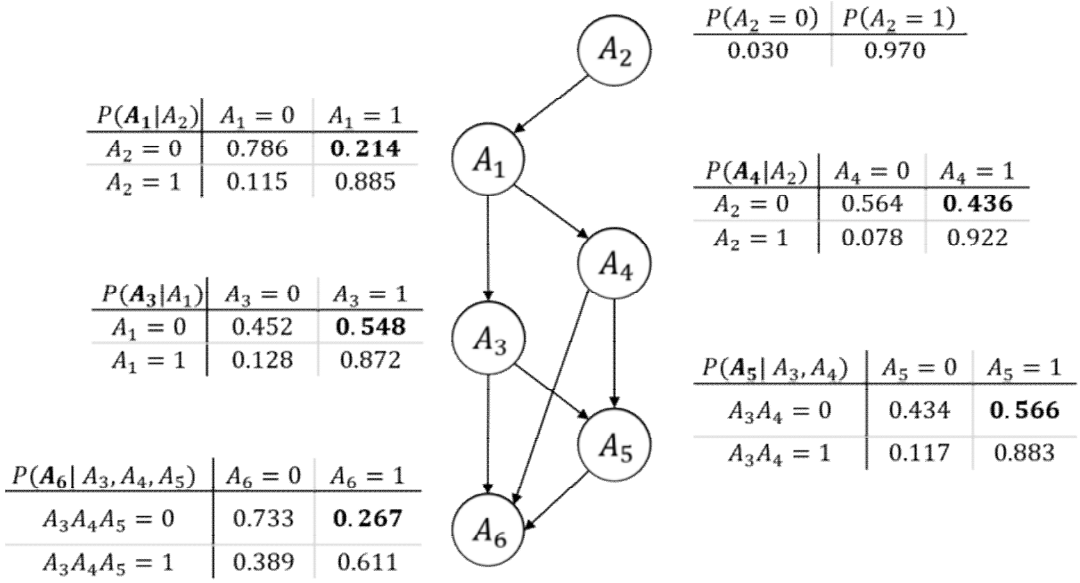
<표 Ⅳ-15> 본 연구에서 활용한 일차방정식과 일차부등식 검사의 인지요소와 Q-행렬

문항	(기존 문항)	A_1	A_2	A_3	A_4	A_5	A_6	정답률
1	(1)	1	0	0	0	0	0	0.822
2	(2)	1	0	0	0	0	0	0.874
3	(3)	0	1	0	0	0	0	0.923
4	(4)	0	0	1	0	0	0	0.708
5	(7)	0	0	0	1	0	0	0.833
6	(11)	0	0	0	1	0	1	0.682
7	(12)	0	0	0	1	0	1	0.776
8	(13)	0	1	0	0	0	0	0.888
9	(14)	0	0	1	0	0	0	0.557
10	(18)	0	0	0	0	1	0	0.754
11	(19)	1	0	0	0	1	0	0.678
12	(22)	0	0	0	0	1	1	0.368
13	(23)	0	0	0	0	1	1	0.376
14	(24)	0	0	0	1	0	1	0.673
15	(25)	0	0	0	0	1	1	0.501

A_1: 사칙연산 및 식의 계산(분수와 소수를 포함한 식의 계산)
A_2: 등식과 부등식의 이해(정수의 대소 비교하기)
A_3: 간단한 관계를 표현한 문장을 식으로 나타내기
A_4: 일차방정식 풀기
A_5: 일차부등식 풀기
A_6: 대수적 모델링(문장제의 문제 상황을 식으로 나타내기)

4.2.3. 분석 결과

일차방정식과 일차부등식 검사 자료에 대해 SEM-H 방법을 적용하여 위계를 추정한 결과 총 4번의 SEM 사이클을 거쳐 다음의 [그림 Ⅳ-19]와 같은 구조와 모수들이 추정되었다.



$P(A_2=0)$	$P(A_2=1)$
0.030	0.970

$P(\boldsymbol{A_1}\|A_2)$	$A_1=0$	$A_1=1$
$A_2=0$	0.786	**0.214**
$A_2=1$	0.115	0.885

$P(\boldsymbol{A_4}\|A_2)$	$A_4=0$	$A_4=1$
$A_2=0$	0.564	**0.436**
$A_2=1$	0.078	0.922

$P(\boldsymbol{A_3}\|A_1)$	$A_3=0$	$A_3=1$
$A_1=0$	0.452	**0.548**
$A_1=1$	0.128	0.872

$P(\boldsymbol{A_5}\|A_3,A_4)$	$A_5=0$	$A_5=1$
$A_3A_4=0$	0.434	**0.566**
$A_3A_4=1$	0.117	0.883

$P(\boldsymbol{A_6}\|A_3,A_4,A_5)$	$A_6=0$	$A_6=1$
$A_3A_4A_5=0$	0.733	**0.267**
$A_3A_4A_5=1$	0.389	0.611

A_1: 사칙연산 및 식의 계산(분수와 소수를 포함한 식의 계산)
A_2: 등식과 부등식의 이해(정수의 대소 비교하기)
A_3: 간단한 관계를 표현한 문장을 식으로 나타내기
A_4: 일차방정식 풀기
A_5: 일차부등식 풀기
A_6: 대수적 모델링(문장제의 문제 상황을 식으로 나타내기)

[그림 Ⅳ-19] 일차방정식과 일차부등식 검사에 대한 SEM-H 적용 결과

[그림 Ⅳ-19]에서 확인할 수 있듯이, 일차방정식과 일차부등식 검사에 대한 SEM-H 적용 결과로 총 8쌍의 인지요소 관계($A_2 \rightarrow A_1$, $A_1 \rightarrow A_3$, $A_1 \rightarrow A_4$, $A_3 \rightarrow A_5$, $A_4 \rightarrow A_5$, $A_3 \rightarrow A_6$, $A_4 \rightarrow A_6$, $A_5 \rightarrow A_6$)가 도출되었다. 각각의 관계들이 위계 관계를 나타내는지의 여부는 분수 뺄셈 검사에서와 마

찬가지로 도출된 위계 구조에서 CPT의 모수 $P(A_j=1|A_i=0)$의 추정치를 바탕으로 판정하였으며, 서로 결합적인 관계에 있는 인지요소 관계들 $((A_3\rightarrow A_5, A_4\rightarrow A_5)$, $(A_3\rightarrow A_6, A_4\rightarrow A_6, A_5\rightarrow A_6))$의 경우에는 각 관계가 하나씩만 존재하는 구조에서 모수를 추정하여 위계 관계 여부를 판정하였다.

8쌍의 인지요소 관계에 대해 위계 관계를 판정한 결과는 다음의 <표 Ⅳ-16>과 같다. 일차방정식과 일차부등식 검사의 경우 분수 뺄셈 검사에서와는 달리 **모든 인지요소 관계가 위계 관계가 아닌 종속 관계**인 것으로 판정되었다.

<표 Ⅳ-16> 일차방정식과 일차부등식 검사의 모수에 대한 위계 관계 판정 결과

인지요소 관계 $(A_i\rightarrow A_j)$	모수 λ_{j1}의 추정치		관계 판정
	$P(A_j=1\mid \Pi\,[pa(A_j)]=0)$	$P(A_j=1\mid A_i=0)$	
$A_2\rightarrow A_1$	**0.214**	좌동	종속
$A_1\rightarrow A_3$	**0.548**	좌동	종속
$A_1\rightarrow A_4$	**0.436**	좌동	종속
$A_3\rightarrow A_5$	0.566	**0.669**	종속
$A_4\rightarrow A_5$		**0.470**	종속
$A_3\rightarrow A_6$	0.267	**0.214**	종속
$A_4\rightarrow A_6$		**0.321**	종속
$A_5\rightarrow A_6$		**0.344**	종속

※여기서 '종속'은 위계 관계가 아닌 종속 관계를 의미한다.

비록 여기서 도출된 구조가 위계 구조를 나타내는 것은 아니지만, SEM-H 방법으로 도출된 구조에서 관계의 방향은 기본적으로 선행 관계가 될 가능성이 높은 방향을 나타내는 것이므로 윤지영(2015)에서 설정했던 위계 구조와 관계의 방향이 다르게 나온 결과에 대해 살펴보고 넘어가고자 한다.

'사칙연산 및 식의 계산'을 나타내는 인지요소 A_1과 '등식과 부등식의 이해'를 나타내는 인지요소 A_2의 경우, 윤지영(2015)은 교육과정 순서를 고려하여 $A_1\rightarrow A_2$로 관계의 방향을 설정하였으나, SEM-H 방법을 통해

자료를 분석한 결과 통계적으로는 $A_2 \rightarrow A_1$이 선행 관계를 나타낼 가능성이 더 높은 것으로 나타났다. 이러한 결과가 나타난 원인은 검사 문항과 정답률을 살펴봄으로써 확인해볼 수 있다. 일차방정식과 일차부등식 검사에서 인지요소 A_1을 측정하는 문항들(문항 1, 2, 11)은 분수 또는 소수를 포함한 식의 계산을 포함하고 있으며, 반면 인지요소 A_2를 측정하는 문항들(문항 3, 4)은 단순히 등호와 부등호를 옳게 나타낸 것을 찾는 문항으로 이루어져 있다(부록 참고). 사실 문항 3과 문항 4는 등식과 부등식의 이해를 측정한다기보다는 정수의 대소를 비교할 수 있는지를 묻는 문항에 가까운 것으로 보이며, 이는 분수나 소수를 포함한 식의 계산보다 먼저 학습이 이루어지는 내용이다. 실제로 <표 Ⅳ-15>에서 문항의 정답률을 살펴보면 A_2를 측정하는 문항 3과 문항 4의 정답률이 인지요소 A_1을 측정하는 문항 1, 문항 2, 문항 11의 정답률보다 높게 나타났다. 이러한 응답 결과로 인해 인지요소 A_1보다 A_2를 숙달한 피험자가 더 많은 것으로 나타나고 그에 따라 두 인지요소 관계의 방향이 $A_2 \rightarrow A_1$으로 도출된 것으로 보인다.

한편, 윤지영(2015)에서는 '대수적 모델링'을 '일차방정식 풀기', '일차부등식 풀기'와 서로 관련이 없는 것으로 설정하였던 반면, SEM-H 방법을 통해 자료를 분석한 결과 이들이 종속 관계를 가지는 것으로 확인되었다. 이는 본 연구에서 분석에 활용한 검사 문항 중 '대수적 모델링(A_6)'을 독립적으로 측정하는 문항이 없었기 때문에 도출된 결과인 것으로 보인다.[17] 또한, 본 연구의 분석에서는 '일차방정식 풀기(A_4)', '일차부등식 풀기(A_5)', '대수적 모델링(A_6)' 인지요소들 사이의 종속 관계의 방향이 $A_4 \rightarrow A_5$, $A_4 \rightarrow A_6$, $A_5 \rightarrow A_6$으로 도출되었는데, 이는 일차방정식 풀이를 학습한 이후에 일차부등식 풀이를 학습하고 일차방정식/일차부등식의 풀이 이후에 이들을 활용하는 문장제(word problem)를 다루도록 되어있는 우리나라 교과서의 학습 순서가 반영된 결과인 것으로 보인다.

17) <표 Ⅳ-15>의 Q-행렬을 살펴보면 '대수적 모델링(A_6)'을 필요로 하는 문항들은 모두 '일차방정식 풀기(A_4)' 또는 '일차부등식 풀기(A_5)' 인지요소를 포함하고 있음을 확인할 수 있다.

5. 요약 및 논의

이 장에서는 베이지안 네트워크의 구조 학습의 한 방법인 SEM 알고리즘과 Ⅲ장에서 도입한 BN-H 모형의 개념을 결합한 새로운 위계 추정 방법인 SEM-H를 제안하였다. SEM-H 방법은 기존에 제안되었던 위계 추정 방법인 COMMAND 방법을 개선할 수 있도록 변형된 것으로, 인지요소들 간의 단순 종속 관계가 아닌 '위계 관계'의 추정이라는 목표를 보다 전면에 내세우고 구조를 탐색하는 방식을 취한다. 또한 구조 학습의 결과로 도출된 구조에 대해 BN-H 모형의 모수 제약의 개념을 활용한 판정 방식을 적용하여 종속 관계와 위계 관계를 판별해낸다.

모의실험 1에서 SEM-H 방법과 기존의 COMMAND 방법을 모의실험을 통해 비교한 결과, SEM-H 방법이 모든 경우에 있어 COMMAND 방법보다 더 정확한 위계 추정 결과를 도출하는 것을 확인할 수 있었다. 또한, 여러 요인들이 위계 추정 결과에 어떤 영향을 미치는지 확인해 본 결과, 표본 크기가 작을수록, 복합 Q-행렬일수록, 또 문항의 질이 떨어지고 위계 구조가 복잡할수록, COMMAND 방법보다 SEM-H 방법이 대체로 더 뛰어난 위계 추정 정확도를 보이는 것을 알 수 있었다. 모의실험 2에서는 다양한 위계 구조를 가정하고 특정 조건 하에서 일정 수준 이상의 위계 추정 정확도를 얻기 위해 어느 정도의 문항과 어느 정도의 표본 크기가 필요한지를 확인하였다. 모의실험 2에서 크게 두드러졌던 결과는 한 인지요소에 선행 인지요소나 후행 인지요소가 여러 개(특히 3개 이상) 이상 존재하는 구조의 경우 그렇지 않은 구조에 비해 위계 추정의 정확도가 낮게 나타난다는 점이었다. 특히 한 인지요소에 대해 후행 인지요소가 늘어나는 발산형 구조에 대해서는 문항의 질이 높아야 다른 구조들과 비슷한 수준의 위계 추정 정확도를 얻을 수 있으며, 문항의 질이 낮은 경우에는 다른 구조들에 비해 더 많은 수의 문항과 표본이 요구된다는 점이 확인되었다.

모의실험 결과에서 추가로 확인할 수 있는 점은, 모의실험의 각 요인들이 SEM-H 방법의 위계 추정 정확도에 미치는 영향을 살펴보았을 때 각 요인들이 인지요소 추정의 정확도에 미치는 영향에 대해 밝혀진 바와 유사한 경향성을 보였다는 점이다. 그 이유는 SEM-H 방법이 SEM 알고리즘 기반의 방법이라는 점에서 찾을 수 있을 것으로 보인다. SEM의 구조 학습은 인지요소 추정치를 기반으로 불완전 자료를 완전 자료로 만든 다음 그 자료를 바탕으로 관계를 탐색하는 방법이기 때문에 SEM-H 방법의 위계 추정의 정확도는 인지요소 추정 정확도의 영향을 받을 수밖에 없는 것이다. 따라서 SEM-H 방법을 이용한 위계 추정에서 보다 높은 정확도를 얻기 위해서는 인지요소 추정의 정확도를 높이는 것으로 알려진 요인들에 대해 고려할 필요가 있다.

이 장에서는 두 가지 실제 수학 검사 자료인 '분수 뺄셈 검사 자료'와 '일차방정식과 일차부등식 검사 자료'에 SEM-H 방법을 적용하여 위계 추정을 실시하였다. 그 결과 '분수 뺄셈 검사 자료'의 인지요소들에 대해서는 위계 관계와 위계 관계가 아닌 종속 관계가 동시에 존재하는 구조가 도출되었으며, '일차방정식과 일차부등식 검사 자료'에 대해서는 위계 관계가 존재하지 않는 종속 관계로만 이루어진 구조가 도출되었다. '일차방정식과 일차부등식 검사'의 경우 원 자료가 수집될 당시부터 인지요소 위계에 대한 가정을 받아들이고 분석이 이루어졌었다. 그러나 실제로 본 연구의 SEM-H 방법을 이용하여 위계 추정을 진행해본 결과, '일차방정식과 일차부등식 검사'에 대해 수집된 자료는 인지요소 위계에 대한 충분한 근거를 제공하지 못하고 있음을 확인할 수 있었다.

인지요소 위계를 고려한 인지진단모형을 활용하여 평가를 개발하고 분석하고자 할 때, 설정된 위계 구조가 타당하지 않으면 왜곡된 검사 결과가 도출될 가능성이 있다. 따라서 위계 구조의 타당성에 대한 검증은 인지요소 위계를 고려한 인지진단모형을 활용한 평가를 진행함에 있어 필수적으로 거쳐야 하는 과정이다. Templin과 Bradshaw(2014)는 HDCM을 이용하여 위계 관계를 검사하는 방안을 제안한 바 있으나, 그들의 방법은 HDCM으로부터 추정된 전체 피험자의 인지요소패턴의 분

포를 분석하여 위계 관계가 의심되는 부분을 찾고 가설검정의 과정을 통해 그 위계 관계 여부를 일일이 판정하는 다소 수동적인 절차로 이루어진다. 반면, 본 연구에서 제안하는 SEM-H는 주어진 검사자료에 대한 구조 학습을 거쳐 자동으로 일차적인 구조를 도출해주며 구조 학습 과정에서 함께 추정되는 모수에 대한 판정을 통해 위계 관계를 추정하는 보다 효율적인 방식을 따른다.

한편, 본 연구에서는 Leighton 외(2004)의 정의를 따라 '인지요소 위계'를 인지요소들의 모든 관계가 위계 관계인 구조를 의미하는 것으로 한정하고 있으나, 실제 수학 검사 자료 분석에서 확인할 수 있었듯이 어떤 한 검사의 인지요소들 사이의 관계가 반드시 위계 관계로만 이루어지는 것은 아닐 수 있다. 사실 '인지요소 위계'는 인지요소들 사이의 관계에 대한 굉장히 강력한 가정이며, 실제로 완벽한 인지요소 위계를 나타내는 검사 자료는 많지 않을 수 있다. 그렇다 하더라도, SEM-H 방법은 일차적으로는 종속 관계를 나타내는 구조를 찾아주고 그 구조로부터 추가로 위계 관계에 대한 판정을 내리는 방식을 취하므로 인지요소들의 종속 관계와 위계 관계를 모두 포함한 구조에 대한 추정이 가능하다는 점에서 실제적인 활용 가능성이 높다고 볼 수 있다. 또한 SEM-H 방법에서 위계 관계에 대한 판정은 $cutline$ 값을 기준으로 이루어지므로, 어느 정도의 엄밀함으로 위계를 판정할 것인지에 따라[18] $cutline$을 조정함으로써 보다 유연한 위계 추정이 가능하다. 따라서 SEM-H는 향후 실제 여러 검사 자료에 대한 위계 추정의 방법으로서 그 활용 가치가 높을 것으로 기대된다.

18) 예를 들어, Gágne 외(2005)는 지적 기능들의 위계와 관련한 논의에서, 어떤 지적기능의 선행 기능은 필수적인(essential) 선행 기능과 보조적인(supportive) 선행 기능이 모두 가능하다고 하였다. 이들의 용어를 차용하면, 한 쌍의 인지요소에 대해, 한 인지요소가 다른 인지요소의 필수적인 선행 인지요소인 경우에만 그 둘의 관계를 위계 관계로 볼 것인지, 보조적인 선행 인지요소인 경우도 위계 관계로 볼 것인지에 따라 위계 관계에 대한 판정이 달라질 수 있다.

Ⅴ. BN-H 모형을 적용한 인지진단 컴퓨터 적응검사

인지진단 컴퓨터 적응검사는 기존의 컴퓨터 적응검사에 인지진단모형을 결합한 것으로, 개별 학생의 인지요소 숙달 수준에 맞는 문항을 컴퓨터가 실시간으로 제시함에 따라 지필검사에 비해 적은 문항으로 보다 정확하게 각 학생의 인지요소패턴을 추정할 수 있는 효율적인 검사 방식이다. 그 동안의 인지진단 컴퓨터 적응검사들은 인지요소 간 위계 관계를 고려하지 않고 가능한 모든 인지요소패턴을 가정하고 각 학생의 숙달 프로파일을 추정해왔으나, 만약 인지요소 위계가 존재한다면 이를 고려하여 축소된 인지요소패턴 공간에서 각 학생의 숙달 프로파일을 추정함으로써 한층 더 효율적인 인지진단 컴퓨터 적응검사가 이루어질 것이라 예상해볼 수 있다. 따라서 이 장에서는 Ⅲ장에서 제시한 BN-H 모형을 이용하여 인지요소 위계를 고려한 인지진단 컴퓨터 적응검사인 BN-H-CAT을 제안하고, 실제로 인지요소 위계가 존재하는 상황에서 이를 고려한 BN-H 기반 인지진단 컴퓨터 적응검사가 인지요소 위계를 고려하지 않은 인지진단 컴퓨터 적응검사에 비해 높은 효율성을 보이는지 검증해보고자 한다.

이하에서는 먼저 기존 인지진단 컴퓨터 적응검사의 문항 선택 알고리즘과 검사 종료 규칙을 검토한 뒤, BN-H 모형을 적용한 인지진단 컴퓨터 적응검사를 정의하고 그 특징을 논한다. 마지막으로, 인지요소 위계를 고려함으로써 인지진단 컴퓨터 적응검사의 효율성이 얼마나 향상되는지를 일련의 모의실험을 통해 살펴본다.

1. 인지진단 컴퓨터 적응검사

앞의 Ⅱ장 3절에서는 IRT 모형 기반 컴퓨터 적응검사의 핵심 요소를 따라 그 원리와 세부적인 특징을 살펴보았다. 그러나 이 장의 초점인 인지진단 컴퓨터 적응 검사는 IRT 모형 대신 인지진단모형을 활용하는 컴퓨터 적응검사로, 피험자의 능력 수준의 척도가 연속형 척도에서 범주형 척도로 변화함에 따라 기존에 IRT 모형의 능력모수 θ를 중심으로 이루어졌던 문항 선택 방법과 검사 종료 규칙이 인지요소패턴 $\boldsymbol{\alpha}$를 중심으로 이루어지도록 변화하게 된다. 따라서 이 절에서는 그동안 제안되어온 인지진단 컴퓨터 적응검사의 문항 선택 방법과 검사 종료 규칙을 살펴보도록 한다.

1.1. 문항 선택 방법

인지진단 컴퓨터 적응검사의 문항 선택 방법은 검사의 각 t단계에서 개별 피험자 i의 이전 $t-1$단계까지의 문항 응답 $\mathbf{x}_{\mathrm{i,t-1}}$이나 이를 바탕으로 추정된 인지요소패턴[19] $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$을 이용하여 피험자 i에게 다음 문항으로 제시되기에 가장 적합한 문항을 찾는 방법을 의미하며, 어떤 기준을 충족하는 문항을 가장 적합한 문항으로 볼 것인가에 따라 다양한 문항 선택 지수(index)들이 제안되어왔다. 컴퓨터 적응검사의 문항 선택은 후보문항들에 대해 특정 문항 선택 지수 값들을 모두 계산한 뒤 그 값이 가장 높거나 가장 낮은 문항을 선택하는 방식으로 이루어진다.

---

19) 인지진단 컴퓨터 적응검사에서는, 피험자 i의 t단계에서의 인지요소패턴의 추정치 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$는 이전 $t-1$단계까지의 문항 응답이 주어졌을 때 패턴들의 사후 확률 $\pi_i(\boldsymbol{\alpha}|\mathbf{x}_{i,t-1})$이 가장 큰 패턴으로 결정된다:

$$\hat{\boldsymbol{\alpha}}_i^{(t-1)} = \boldsymbol{\alpha}_{\hat{c},} \quad \hat{c} = \arg max_c \{\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1})\} \ (c = 1,2,...,2^K)$$

이와 같은 추정 방법을 MAP 추정이라고 한다.

인지진단 컴퓨터 적응검사의 문항 선택 지수들은 크게 두 부류로 구분될 수 있다. 한 부류는 Kullack-Leibler(이하 KL) 정보로부터 파생되어 발전한 지수들이고, 다른 한 부류는 Shannon entropy를 기반으로 발전된 지수들이다.

1.1.1. KL 정보 기반의 문항 선택 지수

1.1.1.1. KL 지수

인지진단 컴퓨터 적응검사의 문항 선택 지수로서 처음에 제안된 KL 지수(Xu, Chang, & Douglas, 2003)는 피험자 i에 대해 $t-1$번째 검사 단계에서 추정된 인지요소패턴 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$가 주어져 있을 때 어떤 문항 j에 대한 피험자 i의 응답의 분포 $P(X_{ij}=x|\hat{\boldsymbol{\alpha}}_i^{(t-1)})$와, 가능한 인지요소패턴 $\boldsymbol{\alpha}_c$ ($c=1,2,...,2^K$; K:인지요소 수)가 주어졌을 때 문항 j에 대한 응답 분포 $P(X_{ij}=x|\boldsymbol{\alpha}_c)$들 사이의 KL 거리(distance)의 합으로 정의된다. KL 지수를 식으로 나타내면

$$KL_j(\hat{\boldsymbol{\alpha}}_i^{(t-1)})=\sum_{c=1}^{2^K}\left[\sum_{x=0}^{1}\log\left(\frac{P(X_{ij}=x|\hat{\boldsymbol{\alpha}}_i^{(t-1)})}{P(X_{ij}=x|\boldsymbol{\alpha}_c)}\right)P(X_{ij}=x|\hat{\boldsymbol{\alpha}}_i^{(t-1)})\right] \quad (5.1)$$

와 같으며, KL 지수 기반의 문항 선택 방법은 KL 지수의 값이 가장 높은 문항을 피험자 i에 대한 t번째 문항으로 선택한다.

KL 지수는 각 단계에서 인지요소패턴 추정치 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$만 결정되면 비교적 간단하게 계산이 된다는 장점이 있으나, 실제로 KL 지수 기반의 문항 선택 방법에 따라 적응검사를 진행했을 때 피험자의 인지요소패턴 추정의 정확도가 낮게 나타나는 문제가 확인되었으며(Xu et al., 2003; Cheng, 2009), 이러한 문제를 해결하도록 KL 지수를 보완한 다른 지수들이 제안되었다.

1.1.1.2. PWKL 지수

Cheng(2009)은 KL 지수에 포함된 각 KL 거리에 대해 각 인지요소패턴 $\boldsymbol{\alpha}_c$의 사후 분포 $\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1})$를 가중치로 곱해주는 방식으로 기존의 KL지수를 수정한 posterior-weighted KL(이하 PWKL) 지수를 제안하였다. PWKL 지수는

$$PWKL_j(\hat{\boldsymbol{\alpha}}_i^{(t-1)}) = \sum_{c=1}^{2^K}\left[\sum_{x=0}^{1}\log\left(\frac{P(X_{ij}=x|\hat{\boldsymbol{\alpha}}_i^{(t-1)})}{P(X_{ij}=x|\boldsymbol{\alpha}_c)}\right)P(X_{ij}=x|\hat{\boldsymbol{\alpha}}_i^{(t-1)})\,\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1})\right] \quad (5.2)$$

(KL distance, weight)

와 같으며, 마지막에 가중치로 곱해진 각 인지요소패턴 $\boldsymbol{\alpha}_c$의 사후 분포 $\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1})$는 베이즈 정리에 의해

$$\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}) \propto p(\boldsymbol{\alpha}_c)\prod_{j=1}^{t-1}P(X_{ij}=1|\boldsymbol{\alpha}_c)^{x_{ij}}[1-P(X_{ij}=1|\boldsymbol{\alpha}_c)]^{1-x_{ij}} \quad (5.3)$$

($p(\boldsymbol{\alpha}_c)$: $\boldsymbol{\alpha}_c$의 사전 분포. 특별한 정보가 없으면 균등분포를 가정함.)

이다. PWKL 지수 기반의 문항 선택 방법은 KL 지수 기반의 문항 선택 방법과 동일하게 PWKL 지수의 값이 가장 높은 문항을 다음 문항으로 선택한다.

PWKL 지수는 KL 지수와 마찬가지로 각 단계에서 추정된 인지요소패턴 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$에 따라 결정된다. 그러나 적응검사의 초기 단계에는 추정치 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$가 피험자 i의 진 인지요소패턴과 다를 가능성이 높으므로, PWKL 지수 기반의 문항 선택 방법은 검사 초기 단계에 적절한 문항을 선택하지 못할 수 있다는 문제가 있다.

1.1.1.3. MPWKL 지수

Kaplan, de la Torre와 Barrada(2015)는 PWKL이 각 단계에서의 인지요소패턴 추정치 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$에 의해 결정됨으로 인해 발생하는 문제점을 해결하도록 PWKL 지수를 수정한 modified PWKL(이하 MPWKL) 지수를 제안하였다. MPWKL 지수는

$$MPWKL_{ij}^{(t)} = \sum_{d=1}^{2^K}\left[\underbrace{\sum_{c=1}^{2^K}\left[\sum_{x=0}^{1}\log\left(\frac{P(X_{ij}=x|\boldsymbol{\alpha}_d)}{P(X_{ij}=x|\boldsymbol{\alpha}_c)}\right)P(X_{ij}=x|\boldsymbol{\alpha}_d)\pi_i(\boldsymbol{\alpha}_c|\boldsymbol{x}_{i,t-1})\right]}_{PWKL_j(\boldsymbol{\alpha}_d)}\underbrace{\pi_i(\boldsymbol{\alpha}_d|\boldsymbol{x}_{i,t-1})}_{\text{weight}}\right] \quad (5.4)$$

와 같이 주어진다. 이는 모든 인지요소패턴 $\boldsymbol{\alpha}_d\,(d=1,2,...,2^K)$가 피험자 i의 진 인지요소패턴이 될 수 있다는 점을 반영하여, 각 $\boldsymbol{\alpha}_d$가 진 인지요소패턴일 경우의 PWKL 지수의 값 $PWKL_j(\boldsymbol{\alpha}_d)$에 $\boldsymbol{\alpha}_d$가 진 인지요소패턴일 확률 $\pi_i(\boldsymbol{\alpha}_d|\mathbf{x}_{i,t-1})$을 가중치로 곱한 값을 모든 인지요소패턴 $\boldsymbol{\alpha}_d\,(d=1,2,...,2^K)$들에 대해 더해준 값이다. MPWKL 지수 기반의 문항선택 방법 역시 MPWKL 지수의 값이 가장 높은 문항을 다음 문항으로 선택한다.

MPWKL 지수는 인지요소패턴 추정치에 영향을 받지 않는 값으로, Kaplan 외(2015)는 MPWKL이 PWKL에 비해 향상된 결과를 나타내는 것을 확인하였다.

1.1.1.4. PWCDI, PWACDI 지수

Zheng과 Chang(2016)은 Henson과 Douglas(2005)가 KL 정보를 이용하여 정의한 지필 인지진단평가의 문항 변별도 지수인 CDM information index(이하 CDI)를 수정하여 적응검사의 문항 선택 지수인 posterior-weighted CDI(이하 PWCDI)와 posterior-weighted attribute-level CDI(이하 PWACDI)를 제안하였다. PWCDI 지수는

$$PWCDI_{ij}^{(t)} = \frac{1}{\sum_{(c,d)\ \text{for}\ c \neq d} h(\boldsymbol{\alpha}_c, \boldsymbol{\alpha}_d)^{-1}} \times$$

$$\sum_{(c,d)\ \text{for}\ c \neq d} \left[h(\boldsymbol{\alpha}_c, \boldsymbol{\alpha}_d)^{-1} \pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}) \pi_i(\boldsymbol{\alpha}_d | \mathbf{x}_{i,t-1}) \left[\sum_{x=0}^{1} \log\left(\frac{P(X_{ij} = x | \boldsymbol{\alpha}_d)}{P(X_{ij} = x | \boldsymbol{\alpha}_c)} \right) P(X_{ij} = x | \boldsymbol{\alpha}_d) \right] \right] \quad (5.5)$$

이며, 여기서 $h(\boldsymbol{\alpha}_c, \boldsymbol{\alpha}_d)$ 두 인지요소패턴 $\boldsymbol{\alpha}_c$와 $\boldsymbol{\alpha}_d$ 사이의 Hamming 거리 $\sum_{k=1}^{K} |\boldsymbol{\alpha}_{ck} - \boldsymbol{\alpha}_{dk}|$를 의미한다. 한편, PWACDI 지수는

$$PWACDI_{ij}^{(t)} = \frac{1}{2^K} \times$$

$$\sum_{\substack{(c,d)\ \text{for} \\ h(\boldsymbol{\alpha}_c, \boldsymbol{\alpha}_d) = 1}} \left[\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}) \pi_i(\boldsymbol{\alpha}_d | \mathbf{x}_{i,t-1}) \left[\sum_{x=0}^{1} \log\left(\frac{P(X_{ij} = x | \boldsymbol{\alpha}_d)}{P(X_{ij} = x | \boldsymbol{\alpha}_c)} \right) P(X_{ij} = x | \boldsymbol{\alpha}_d) \right] \right] \quad (5.6)$$

와 같이 정의된다. PWCDI 지수와 PWACDI 지수 기반의 문항 선택 방법 또한 각 지수의 값이 가장 높은 문항을 다음 문항으로 선택하는 방법이다.

PWCDI 지수는 Hamming 거리가 가까운 인지요소패턴들을 더 잘 변별하는 문항이 보다 적합한 문항이라는 관점이 반영된 지수이며, PWACDI 지수는 특히 Hamming 거리가 1인, 다시 말해 오직 하나의 인지요소 숙달 여부에서만 차이가 있는 두 인지요소패턴의 쌍들을 잘 변별하는 문항을 적합한 문항으로 보는 지수이다.

1.1.2. Shannon entropy 기반의 문항 선택 지수

1.1.2.1. SHE 지수

Shannon entropy는 정보이론에서 파생된 개념으로, 확률변수의 불확실성의 정도를 나타낸다(Rényi, 1961). 확률변수 X의 확률질량함수가 $p(x)$일 때 X의 shannon entropy는

$$sh[X] = sh(p(x)) = \sum_{x} p(x) \log \frac{1}{p(x)} \tag{5.7}$$

와 같이 정의된다. Shannon entropy는 확률변수의 분포가 균등분포일 때 최대이며, 확률변수의 확률 질량이 한 값에 응축되어 있을 때(즉, 확률변수가 어느 한 값을 가질 확률이 1일 때) 최소이다.

C. Tatsuoka(2002), C. Tatsuoka와 Ferguson(2003)는 인지진단 컴퓨터 적응검사의 문항 선택 지수로서 Shannon entropy(이하 SHE) 지수를 제안하였다. SHE 지수는 인지요소패턴 추정치의 사후 분포 $\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}, X_{ij}=x)$의 expected Shannon entropy로 다음과 같이 정의된다.

$$\begin{aligned} SHE_{ij}^{(t)} &= \sum_{x=0}^{1} \left[\sum_{c=1}^{2^K} \pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}, X_{ij}=x) \log\left(\frac{1}{\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}, X_{ij}=x)} \right) \right] \left[\sum_{c=1}^{2^K} P(X_{ij}=x|\boldsymbol{\alpha}_c) \pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}) \right] \\ &= \sum_{x=0}^{1} P(X_{ij}=x|\mathbf{x}_{i,t-1}) \left[\sum_{c=1}^{2^K} \pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}, X_{ij}=x) \log\left(\frac{1}{\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1}, X_{ij}=x)} \right) \right] \end{aligned} \tag{5.8}$$

위 식에서 $P(X_{ij}=x|\mathbf{x}_{i,t-1})$는 피험자 i의 이전 $t-1$번째 문항까지에 대한 응답 $\mathbf{x}_{i,t-1}$가 주어져 있을 때 문항 j에 대한 피험자 i의 응답의 조건부 확률분포를 나타낸다.

어떤 문항의 SHE 지수 값은 그 문항에 대한 응답이 주어졌을 때 인지요소패턴 추정치의 사후 분포의 불확실성의 정도에 대한 일종의 기댓값을 나타내는 것으로, 불확실성의 정도가 낮을수록 특정 인지요소패턴이 진 인지요소패턴일 확률이 높으며 이는 곧 인지요소패턴 추정의 정확도가 높다는 의미이다. 따라서 SHE 지수 기반의 문항 선택 방법은 SHE 지수의 값이 가장 낮은 문항을 다음 문항으로 선택한다.

1.1.2.2. MI 지수

Wang(2013)은 인지진단 컴퓨터 적응검사의 문항 선택 지수로 mutual information(이하 MI) 지수를 제안하였다. MI 지수는

$$MI_{ij}^{(t)} = \sum_{x=0}^{1} P(X_{ij} = x | \mathbf{x}_{i,t-1}) \left[\sum_{c=1}^{2^K} \pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x) \log \left(\frac{\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x)}{\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1})} \right) \right] \quad (5.9)$$

와 같이 정의되는데, 이 지수는 인지요소패턴 추정치의 $t-1$단계에서의 사후 분포 $\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1})$의 Shannon entropy와 t단계에서 문항 j에 대한 응답이 주어졌을 때의 사후 분포 $\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x)$의 Shannon entropy의 차이에 대한 기댓값으로 다시 표현될 수 있다.

$$\begin{aligned} MI_{ij}^{(t)} &= \sum_{x=0}^{1} P(X_{ij} = x | \mathbf{x}_{i,t-1}) \\ &\left[\sum_{c=1}^{2^K} \pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x) \log \left(\frac{1}{\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1})} \right) - \sum_{c=1}^{2^K} \pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x) \log \left(\frac{1}{\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x)} \right) \right] \\ &= E_{X_{ij}}[sh(\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}))] - E_{X_{ij}}[sh(\pi_i(\boldsymbol{\alpha}_c | \mathbf{x}_{i,t-1}, X_{ij} = x))] \end{aligned} \quad (5.10)$$

이 식은 MI 지수 값이 문항 j가 다음 문항으로 선택되기 전과 후의 인지요소패턴 추정치의 사후 분포의 불확실성의 정도의 차이를 나타냄을 보여주며, 이 값이 큰 문항일수록 다음 문항으로 선택되었을 때 인지요소패턴 추정치의 정확도를 높일 것이라 기대할 수 있다. 따라서 MI 지수 기반의 문항 선택 방법은 MI 지수의 값이 가장 큰 문항을 다음 문항으로 선택한다.

1.2. 검사 종료 규칙

컴퓨터 적응검사의 종료 규칙은 크게 두 가지, '고정-길이 규칙'과 '고정-정확도 규칙'으로 구분된다. '고정-길이 규칙'은 사전에 정해놓은 검사 길이에 도달하는 경우 검사를 종료하는 규칙으로, 이 규칙이 사용되는 경우 모든 피험자가 동일한 길이의 검사를 받게 된다. '고정-정확도 규칙'은 각 피험자의 능력 추정치의 정확도가 사전에 정해놓은 값에 도달했을 때 검사를 종료하는 규칙으로, 이 규칙이 사용되는 경우 특정 정확도에 도달할 때까지 제시되는 문항의 수가 피험자마다 달라지기 때문에 피험자들이 서로 다른 길이의 검사를 받게 된다. 이러한 점으로 인해 '고정-정확도 규칙'은 '변동-길이 규칙'으로 불리기도 한다(Hsu, Wang, & Chen, 2013).

인지진단 컴퓨터 적응검사는 각 피험자의 인지요소패턴을 정확하게 추정하는 것을 목표로 하기 때문에, 추정된 인지요소패턴의 사후 확률 $\pi_i(\hat{\boldsymbol{\alpha}}_i^{(t)}|\boldsymbol{x}_{i,t-1}, X_{ij}=x)$의 값이 추정의 정확도를 나타낸다고 볼 수 있다. Hsu 외(2013)는 인지요소패턴들의 사후 확률을 PPLS(posterior probability of each latent class)라 칭하고 이를 이용하여 인지진단 컴퓨터 적응검사의 변동-길이 종료 규칙 두 가지를 다음과 같이 제안하였다:

규칙1: 가장 큰 PPLS의 값이 사전에 정한 값(예를 들어, 0.7)에 도달했을 때 검사를 종료한다.

규칙2: 가장 큰 PPLS의 값이 사전에 정한 값(예를 들어, 0.7)에 도달할 뿐 아니라, 두 번째로 큰 PPLS의 값이 특정 값(예를 들어, 0.1)을 넘지 않을 때 검사를 종료한다.

인지진단 컴퓨터 적응검사에서는 사후 확률이 가장 높은 패턴을 피험자의 인지요소패턴 추정치로 결정하기 때문에, 위의 규칙들에서 가장 큰 PPLS의 값이란 곧 추정된 인지요소패턴의 사후 확률을 의미한다.

2. BN-H-CAT

이 절에서는 인지요소 위계를 고려한 BN-H 모형 기반의 인지진단 컴퓨터 적응검사 방법인 BN-H-CAT을 제안한다. BN-H-CAT은 인지진단 컴퓨터 적응검사의 일종이므로 기본적으로 인지진단 컴퓨터 적응검사의 절차를 따르지만, 인지요소 위계를 고려한다는 점과 베이지안 네트워크 기반의 모형을 인지진단모형으로 사용한다는 점에서 기존의 인지진단 컴퓨터 적응검사와는 다른 몇 가지 고유한 특징을 지닌다. 이하에서는 BN-H-CAT의 절차를 설계와 실행의 두 단계로 구분하고, 각 단계별로 BN-H-CAT의 고유한 특징을 포함한 방법을 제시한다.

2.1. BN-H-CAT의 설계

BN-H-CAT의 설계 단계에서는 인지요소 위계를 BN-H 모형으로 설정하고 실제 검사에서 사용될 문항들을 설계하여 문제은행을 구축하는 과정이 이루어져야 한다.

2.1.1. BN-H 모형 설정

BN-H-CAT은 인지요소 사이에 존재하는 위계 관계를 BN-H 모형으로 모델링하고 그 모형을 기반으로 진행하는 인지진단 컴퓨터 적응검사이다. 검사의 내용과 목적에 따라 설정된 인지요소들 간의 위계 관계가 구체화되면, BN-H-CAT은 그 위계 구조를 BN-H 모형으로 모델링하는 것에서부터 출발한다. 앞의 Ⅲ장에서 설명했듯이 인지요소 위계는 BN-H 모형의 구조모형으로 모델링된다. 이는 인지요소 위계를 베이지안 네트워크의 DAG로 시각화하여 나타내는 과정과 DAG의 각 인지요소 노드들에 대응하는 CPT 내의 모수들의 초기값을 설정하는 과정으로 이루어진다.

어떤 검사에서 측정하고자 하는 인지요소가 총 5개(A_1, A_2, A_3, A_4, A_5)이고 $A_1 \rightarrow A_2, A_1 \rightarrow A_3, A_2 \rightarrow A_4, A_3 \rightarrow A_5$와 같은 위계 관계가 확인되었다고 하자. 그러한 관계들로 이루어진 인지요소 위계는 [그림 Ⅴ-1]의 (a)와 같이 베이지안 네트워크의 DAG로 표현될 수 있으며, DAG의 각 인지요소 노드에 대응하는 CPT는 [그림 Ⅴ-1]의 (b)와 같이 표현된다. [그림 Ⅴ-1] (b)의 CPT에 포함된 모수들($\gamma_1, \gamma_{2,1}, \gamma_{3,1}, \gamma_{4,1}, \gamma_{5,1}$)은 BN-H 모형의 구조모형을 결정하며, BN-H-CAT에서는 구조모형 모수들의 초기값 $\gamma_1^{(0)}, \gamma_{2,1}^{(0)}, \gamma_{3,1}^{(0)}, \gamma_{4,1}^{(0)}, \gamma_{5,1}^{(0)}$이 검사 진행 이전에 설정되어야 한다.[20]

구조모형 모수들의 초기값은 피험자 집단의 각 인지요소 숙달 비율에 대한 사전 정보를 나타낸다고 볼 수 있다. 예를 들어, 인지요소 A_1에 대응하는 모수 $\gamma_1 = P(A_1 = 1)$의 초기값 $\gamma_1^{(0)}$는 검사 진행 전 인지요소 A_1을 숙달했을 것으로 예상되는 피험자의 비율을 의미하며, 인지요소 A_2에 대응하는 모수 $\gamma_{2,1} = P(A_2 = 1 | A_1 = 1)$의 초기값 $\gamma_{2,1}^{(0)}$은 인지요소 A_1을 숙달했을 것으로 예상되는 피험자들 중 인지요소 A_2도 숙달했을 것으로 예상되는 피험자의 비율을 의미한다. 이와 같은 비율에 대한 사전 정보가 있는 경우 이를 직접적으로 각 모수의 초기값으로 부여할 수 있다. 만약 그러한 사전 정보가 없는 경우라면 피험자 집단에서 각 인지요소패턴 $\boldsymbol{\alpha}_c$의 비율이 동일하다고 보는 이산 균등분포를 가정하고 모수들의 초기값을 부여할 수 있다. 이를 위해서는 [그림 Ⅴ-1] (c)의 표에 제시된 10개의 가능한 인지요소패턴들의 사전 숙달확률 $P(\boldsymbol{\alpha}_c)(c = 1,2,...,10)$들이 모두 동일한 값, 즉 $\frac{1}{10}(=0.1)$을 갖도록 모수들의 초기값 $\gamma_1^{(0)}, \gamma_{2,1}^{(0)}, \gamma_{3,1}^{(0)}, \gamma_{4,1}^{(0)}, \gamma_{5,1}^{(0)}$을 설정하면 되는데, 그 값은 다음과 같이 10개의 가능한 인지요소패턴 중에서 각 모수가 나타내는 확률에 대응하는 패턴의

20) 문항이 이미 정해져 있는 지필검사 형태의 인지진단검사에 BN-H 모형이 사용되는 경우에는 검사 자료가 수집된 이후 그 자료를 이용하여 구조모형의 모수들을 추정하고 추정된 모수 값들을 이용하여 진단이 이루어질 수 있지만, 실시간으로 피험자의 인지요소패턴을 진단하여 그에 따라 다음 문항이 유동적으로 제시되는 적응검사에 BN-H 모형이 적용되는 BN-H-CAT에서는 피험자의 문항 응답이 주어질 때마다 구조모형의 모수 값들을 실시간으로 업데이트 하면서 검사가 이루어지므로 검사 진행 전에 각 모수의 초기값이 설정되어야 한다.

비율을 계산함으로써 구할 수 있다:

$$\gamma_1^{(0)} = P(A_1 = 1) = \frac{A_1 = 1\text{인 인지요소 패턴의 수}}{\text{가능한 총 인지요소 패턴의 수}} = \frac{9}{10}$$

$$\gamma_{2,1}^{(0)} = P(A_2 = 1 | A_1 = 1) = \frac{A_1 = 1, A_2 = 1\text{인 인지요소 패턴의 수}}{A_1 = 1\text{인 인지요소 패턴의 수}} = \frac{6}{9} = \frac{2}{3}$$

$$\gamma_{3,1}^{(0)} = P(A_3 = 1 | A_1 = 1) = \frac{A_1 = 1, A_3 = 1\text{인 인지요소 패턴의 수}}{A_1 = 1\text{인 인지요소 패턴의 수}} = \frac{6}{9} = \frac{2}{3}$$

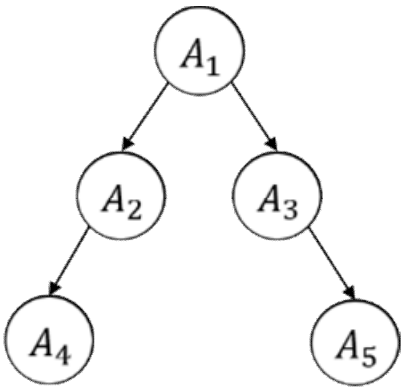
$$\gamma_{4,1}^{(0)} = P(A_4 = 1 | A_2 = 1) = \frac{A_2 = 1, A_4 = 1\text{인 인지요소 패턴의 수}}{A_2 = 1\text{인 인지요소 패턴의 수}} = \frac{3}{6} = \frac{1}{2}$$

$$\gamma_{5,1}^{(0)} = P(A_5 = 1 | A_3 = 1) = \frac{A_3 = 1, A_5 = 1\text{인 인지요소 패턴의 수}}{A_3 = 1\text{인 인지요소 패턴의 수}} = \frac{3}{6} = \frac{1}{2}$$

실제로 이 값들을 [그림 Ⅴ-1] (c)의 표의 세 번째 열에 제시된 각 인지요소패턴의 사전 숙달확률($P(\boldsymbol{\alpha}_c)$) 식에 대입해보면 모두 $\frac{1}{10}$이 됨을 확인할 수 있다.

한편, BN-H 모형의 측정모형은 특정 인지요소패턴 $\boldsymbol{\alpha}_c$를 가진 피험자 i가 각 문항 j을 맞힐 확률 $P(X_{ij} = 1 | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c)$을 나타내는 모형으로, 이는 문제은행에 포함될 문항들의 특성을 고려해서 결정해야 한다. 예를 들어, 각 문항의 정답을 맞히기 위해 문항이 측정하는 모든 인지요소들을 반드시 숙달해야만 하는 경우 DINA 모형을, 문항이 측정하는 인지요소 중 하나라도 숙달하면 문항의 정답을 맞힐 수 있는 경우라면 DINO 모형을 측정모형으로 결정할 수 있다. 만약 문제은행에 두 가지 유형의 문항이 모두 포함될 수 있는 상황이면 두 모형을 모두 표현할 수 있는 LCDM과 같은 보다 일반적인 모형을 측정모형으로 결정하면 된다.

(a) 인지요소 위계를 나타낸 DAG



(b) 각 인지요소 노드의 CPT

$P(A_1=0)$	$P(A_1=1)$
$1-\gamma_1$	γ_1

$P(A_2\|A_1)$	$A_2=0$	$A_2=1$
$A_1=0$	1	0
$A_1=1$	$1-\gamma_{2,1}$	$\gamma_{2,1}$

$P(A_3\|A_1)$	$A_3=0$	$A_3=1$
$A_1=0$	1	0
$A_1=1$	$1-\gamma_{3,1}$	$\gamma_{3,1}$

$P(A_4\|A_2)$	$A_4=0$	$A_4=1$
$A_2=0$	1	0
$A_2=1$	$1-\gamma_{4,1}$	$\gamma_{4,1}$

$P(A_5\|A_3)$	$A_5=0$	$A_5=1$
$A_3=0$	1	0
$A_3=1$	$1-\gamma_{5,1}$	$\gamma_{5,1}$

(c) 주어진 인지요소 위계 하에서 가능한 인지요소패턴 $\boldsymbol{\alpha}_c$와 각 인지요소 패턴의 사전 숙달확률 $P(\boldsymbol{\alpha}_c)$ ($c=1,2,...,10$)

c	$\boldsymbol{\alpha}_c=(\alpha_{c1},\alpha_{c2},\alpha_{c3},\alpha_{c4},\alpha_{c5})$	$P(\boldsymbol{\alpha}_c)=P(A_1=\alpha_{c1})P(A_2=\alpha_{c2}\|A_1=\alpha_{c1})P(A_3=\alpha_{c3}\|A_1=\alpha_{c1})$ $\times P(A_4=\alpha_{c4}\|A_2=\alpha_{c2})P(A_5=\alpha_{c5}\|A_3=\alpha_{c3})$
1	(0,0,0,0,0)	$(1-\gamma_1)\cdot 1\cdot 1\cdot 1\cdot 1$
2	(1,0,0,0,0)	$\gamma_1\cdot(1-\gamma_{2,1})\cdot(1-\gamma_{3,1})\cdot 1\cdot 1$
3	(1,1,0,0,0)	$\gamma_1\cdot\gamma_{2,1}\cdot(1-\gamma_{3,1})\cdot(1-\gamma_{4,1})\cdot 1$
4	(1,0,1,0,0)	$\gamma_1\cdot(1-\gamma_{2,1})\cdot\gamma_{3,1}\cdot 1\cdot(1-\gamma_{5,1})$
5	(1,1,1,0,0)	$\gamma_1\cdot\gamma_{2,1}\cdot\gamma_{3,1}\cdot(1-\gamma_{4,1})\cdot(1-\gamma_{5,1})$
6	(1,1,0,1,0)	$\gamma_1\cdot\gamma_{2,1}\cdot(1-\gamma_{3,1})\cdot\gamma_{4,1}\cdot 1$
7	(1,1,1,1,0)	$\gamma_1\cdot\gamma_{2,1}\cdot\gamma_{3,1}\cdot\gamma_{4,1}\cdot(1-\gamma_{5,1})$
8	(1,0,1,0,1)	$\gamma_1\cdot(1-\gamma_{2,1})\cdot\gamma_{3,1}\cdot 1\cdot\gamma_{5,1}$
9	(1,1,1,0,1)	$\gamma_1\cdot\gamma_{2,1}\cdot\gamma_{3,1}\cdot(1-\gamma_{4,1})\cdot\gamma_{5,1}$
10	(1,1,1,1,1)	$\gamma_1\cdot\gamma_{2,1}\cdot\gamma_{3,1}\cdot\gamma_{4,1}\cdot\gamma_{5,1}$

[그림 Ⅴ-1] BN-H 모형의 구조모형

2.1.2. 문제은행 구축

인지요소 위계를 고려하여 BN-H 모형의 구조모형이 설정되고 어떤 모형을 측정모형으로 가정할지 결정되고 나면 실제 검사 진행에 필요한 문항들을 작성 · 수집하여 문제은행을 구축해야 한다. 문제은행에는 각 문항의 내용뿐만 아니라 문항의 형태, 문항의 측정학적 정보 등이 저장되어야 한다. 또한, IRT 기반 컴퓨터 적응검사에서 검사의 효율성과 정확성을 높이기 위해 다양한 능력수준의 문항을 확보해야 하는 것과 마찬가지로, BN-H-CAT과 같은 인지진단 컴퓨터 적응검사에서는 다양한 인지요소 조합을 측정하는 양질의 문항들이 문제은행에 다수 포함되어 있어야 검사의 효율성과 정확성을 높일 수 있다. BN-H-CAT의 진행을 위해 문제은행에 필수적으로 저장되어야 할 측정학적 정보는 다음과 같다.

- 문항의 Q벡터: 문항이 각 인지요소를 측정하는지의 여부를 0/1로 표시한 벡터. 예를 들어, 어떤 문항이 $A_1 \sim A_5$까지 총 5개의 인지요소 중 A_1과 A_3를 측정하는 문항인 경우 그 문항의 Q벡터는 (1,0,1,0,0)이 된다.
- 문항모수: 문항이 측정하는 인지요소들의 숙달 여부에 따라 문항을 맞힐 확률을 나타내는 모수로, 측정모형의 종류에 따라 달라질 수 있다. 예를 들어, DINA 모형이나 DINO 모형이 측정모형으로 사용되는 경우 문항모수는 추측 모수와 부주의 모수 두 가지가 될 것이며, LCDM이 측정모형으로 사용되는 되는 경우 문항모수는 절편 모수와 주효과 모수, 상호작용효과 모수로, 각 문항이 측정하는 인지요소의 수에 따라 문항모수의 수도 달라진다.

2.2. BN-H-CAT의 실행

BN-H-CAT은 기본적으로 인지진단 컴퓨터 적응검사의 절차를 따라 다음과 같은 과정으로 진행된다:

각 피험자에 대해
(1) 첫 번째 문항을 선택하여 제시한다.
(2) 제시된 문항에 대한 피험자의 응답을 토대로 피험자의 인지요소패턴을 추정/갱신한다.
(3) 추정/갱신된 인지요소패턴을 토대로 적절한 다음 문항을 선택하여 제시한다.
(4) (2)와 (3)의 과정을 검사 종료 기준에 도달할 때까지 반복한다.
(5) 마지막 단계에서 갱신된 인지요소패턴을 피험자의 최종 인지요소패턴으로 확정하고 검사 결과를 피험자에게 제공한다.

(1)에서 첫 번째 문항을 선택할 때에는, 각 피험자마다 초기(검사 시작 전) 인지요소패턴 추정치가 있다면 그 수준에 맞는 문항을 선택할 수 있다. 그러나 인지요소패턴에 동일한 사전 확률을 부여한 경우에는 피험자의 초기 인지요소패턴 추정치가 존재하지 않으므로 첫 문항은 임의로 선택할 수도 있다. (2)에서 인지요소패턴을 추정/갱신할 때에는 BN-H 모형을 활용하기 때문에 BN의 추론을 통한 추정이 이루어진다. (3)에서 다음 문항을 선택할 때에는 1절에서 설명한 문항 선택 방법들을 이용하나, BN-H-CAT에서는 인지요소 위계를 고려하기 때문에 이를 반영하여 문항 선택 지수들에 약간의 변형이 필요하다. (4)의 검사 종료 기준은 1절에서 설명한 규칙들 중 하나를 이용하면 되며, 마지막으로 검사 결과를 보고하는 (5)에서는 각 피험자에게 제시되었던 문항 및 그에 대한 응답, 문항의 실제 정답과 정오 여부 등과 더불어, 검사 결과 최종적으로 추정된 인지요소패턴은 무엇인지, 개별 인지요소를 각각 어느 정도 숙달하였다고 평가되었는지에 대한 정보를 제공하여야 한다.

이하의 2.2.1절과 2.2.2절에서는 위의 다섯 가지 과정 중 일반적인 인지진단 컴퓨터 적응검사와는 다른 BN-H-CAT만의 고유한 특징을 포함하고 있는 (2)와 (3)의 과정을 각각 살펴봄으로써 BN-H-CAT의 실행 방법에 대한 설명을 마무리한다.

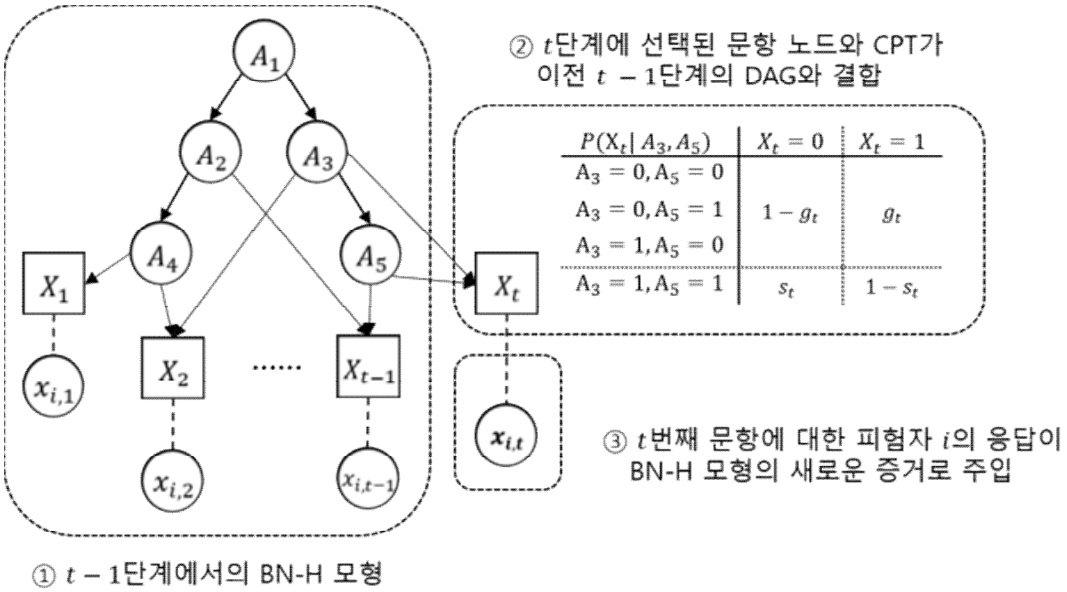
2.2.1. 베이지안 네트워크의 추론을 통한 인지요소패턴 추정

BN-H-CAT은 BN-H 모형을 인지진단모형으로 활용하는 적응검사이므로, BN의 추론을 통해 인지요소패턴 추정이 이루어진다. BN의 추론이란 모형의 일부 변수들에 대한 관찰값(observation)이 주어져 있을 때 관심의 대상이 되는 변수들에 대한 사후 확률분포를 계산하는 과정을 의미한다(Nadkarni & Shenoy, 2001). BN-H-CAT과 같은 인지진단 컴퓨터 적응검사의 맥락에서 그 과정은 다음과 같이 구체화된다.

① 적응검사는 검사가 진행되는 과정에서 실시간으로 문항이 결정되므로, BN-H-CAT의 인지진단모형인 BN-H 모형은 문항 노드 없이 인지요소 노드로만 이루어진 DAG와 그 CPT들로 구성된 구조모형만 설정된 상태로 검사가 시작된다. 이 구조모형은 각 인지요소패턴의 사전 확률($P(\boldsymbol{\alpha}_c)$, $c=1,...,C$, C:가능한 인지요소패턴의 수)을 나타낸다(2.1절의 [그림 Ⅴ-1] 참고).

② 피험자 i에 대해 검사가 진행되는 동안, 검사의 각 t단계에서 제시될 문항이 선택되면, 선택된 문항의 노드를 이전 $t-1$단계의 DAG에 추가되며 그 문항 노드에 대응하는 CPT에는 선택된 문항의 문항모수 값이 주어진다([그림 Ⅴ-2] 참고).

③ 선택된 문항에 대한 피험자의 응답 $x_{i,t}$이 해당 문항 노드에 대한 새로운 증거(evidence)로서 BN-H 모형에 주입되면서 BN-H 모형이 갱신된다([그림 Ⅴ-2] 참고).



[그림 Ⅴ-2] BN-H-CAT에서 BN-H 모형의 갱신

④ 새로 주입된 증거를 바탕으로 그 피험자의 각 인지요소패턴의 사후 확률이 다음과 같이 계산된다:

$$P(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t}) \propto P(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1})P(x_{i,t}|\boldsymbol{\alpha}_c) = P(\boldsymbol{\alpha}_c)\prod_{n=1}^{t} P(x_{i,n}|\boldsymbol{\alpha}_c) \tag{5.11}$$

여기서, $\mathbf{x}_{i,t} = (x_{i,1}, x_{i,2}, \cdots, x_{i,t})$는 t단계까지 제시된 모든 문항에 대한 피험자 i의 문항응답을 나타내는 벡터이다.

⑤ 사후 확률이 가장 높은 인지요소패턴을 피험자 i의 t단계에서의 인지요소패턴 추정치로 결정한다.

$$\hat{\boldsymbol{\alpha}}_i^{(t)} = \boldsymbol{\alpha}_{\hat{c}}, \quad \hat{c} = \arg max_c \{P(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t})\} \tag{5.12}$$

2.2.2. 축소된 인지요소패턴 공간에 기반한 문항 선택

BN-H-CAT에서는 인지요소 위계를 고려하기 때문에 피험자들이 가질 수 있는 인지요소패턴의 수가 줄어든다. 예를 들어, 인지요소가 5개인 검사에서 인지요소 위계가 설정되지 않았을 경우 가능한 인지요소패턴의 수 $2^5 = 32$이지만, [그림 Ⅴ-1]의 (a)와 같은 구조를 갖는 인지요소 위계가 설정된 경우 가능한 인지요소패턴의 수는 10개로 줄어든다.

BN-H-CAT은 인지진단 컴퓨터 적응검사의 일종이므로 문항 선택을 위해 앞의 1절에서 다룬 문항 선택 지수들을 활용할 수 있으나, 인지요소 위계로 인해 가능한 인지요소패턴의 수가 줄어들었기 때문에 문항 선택 지수들이 포함하고 있는 합(summation)의 범위가 달라진다. 구체적으로, 2^K개(K: 인지요소의 수)의 인지요소패턴에 대한 합인 $\sum_{c=1}^{2^K} f(\boldsymbol{\alpha}_c)$ 또는 $\sum_{d=1}^{2^K} f(\boldsymbol{\alpha}_d)$ ($f(\cdot)$는 인지요소패턴에 관한 식을 의미)를 포함하고 있는 문항 선택 지수인 KL, PWKL, MPWKL, SHE, MI의 경우, BN-H-CAT에서는 그 합이 $\sum_{c=1}^{R} f(\boldsymbol{\alpha}_c)$ 또는 $\sum_{d=1}^{R} f(\boldsymbol{\alpha}_d)$ (R은 인지요소 위계에 의해 줄어든 인지요소패턴의 수)의 형태로 변형되어야 한다. 예를 들어, BN-H-CAT에서 MPWKL 지수는 다음과 같이 줄어든 수의 인지요소패턴들에 대해 더해진 값으로 계산되어야 한다.

$$MPWKL_{ij}^{(t)} = \sum_{d=1}^{R}\left[\sum_{c=1}^{R}\left[\sum_{x=0}^{1}\log\left(\frac{P(X_{ij}=x|\boldsymbol{\alpha}_d)}{P(X_{ij}=x|\boldsymbol{\alpha}_c)}\right)P(X_{ij}=x|\boldsymbol{\alpha}_d)\pi_i(\boldsymbol{\alpha}_c|\mathbf{x}_{i,t-1})\right]\pi_i(\boldsymbol{\alpha}_d|\mathbf{x}_{i,t-1})\right] \quad (5.13)$$

(R: 인지요소 위계에 의해 줄어든 인지요소패턴의 수)

BN-H-CAT에서는 PWCDI, PWACDI 지수도 마찬가지로 합의 범위가 인지요소 위계에 따라 축소된 인지요소패턴 공간으로 한정되며, PWACDI 지수의 식에 포함된 2^K 역시 R로 수정된다(1.1절의 문항 선택 방법 참고).

3. 모의실험

본 모의실험의 목적은 실제로 인지요소 위계가 존재하는 상황에서 BN-H 모형을 통해 인지요소 위계를 고려한 인지진단 컴퓨터 적응검사가 인지요소 위계를 고려하지 않은 검사와 비교하여 효율성을 향상시킬 수 있는지를 다양한 조건 하에서 검증하는 것이다. 이를 위해 본 실험에서는 인지진단모형의 측정모형은 DINA 모형으로 가정하고, BN-H모형을 통해 인지요소 위계를 모델링한 인지진단모형(이하 BN-H-DINA)과 인지요소 위계를 모델링하지 않은 인지진단모형(이하 Non-hierarchical DINA)을 각각 적용한 인지진단 컴퓨터 적응검사의 효율성을 비교하였다. 여기서 인지진단 컴퓨터 적응검사의 효율성은 검사 종료 규칙에 따라 그 의미가 달라지는데, 예를 들어 고정-길이 규칙이 사용되는 경우 인지요소패턴 추정의 정확도가 높을수록 검사의 효율성이 높으며, 고정-정확도 규칙이 사용되는 경우 검사에 사용된 문항 수가 적을수록, 즉 검사 길이가 짧을수록 검사의 효율성이 높다고 할 수 있다. 본 모의실험에서는 두 가지 검사 종료 규칙을 모두 실험 조건으로 고려하였다.

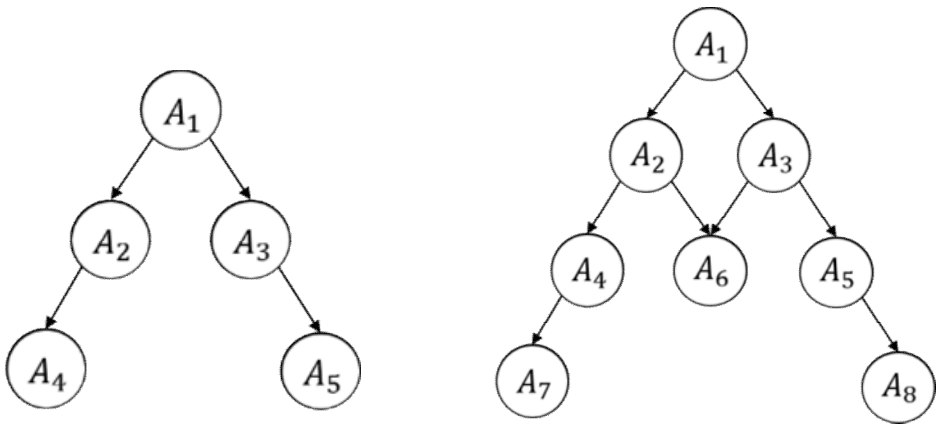
한편, 이 장의 모의실험은 이전 실험들과 마찬가지로 통계 분석 소프트웨어 R 3.4.2 버전(R Core Team, 2017)을 이용하여 실행되었으며, 특히 BN-H 기반 인지진단 컴퓨터 적응검사의 실행에는 bnlearn 패키지(Scutari, 2010)와 gRain 패키지(Højsgaard, 2012)가 사용되었다.

3.1. 모의실험 설계

본 모의실험에서는 '인지요소 위계 구조', '문제은행(문항)의 질', '문항 선택 방법', '검사 종료 규칙'의 4가지 요인을 고려하였다. 이하에서는 모의실험의 절차와 함께 각 요인을 어떻게 설정하였는지 설명한다.

3.1.1. 인지요소 위계 구조

‘인지요소 위계 구조’는 인지요소의 수가 서로 다른 구조를 설정하여 인지요소의 수가 많아짐에 따라 효율성에 어떠한 변화가 나타나는지를 확인하기 위해 고려한 요인이다. Xu 외(2013), Cheng(2009), Wang(2013) 등을 비롯한 여러 선행 연구들은 대체로 5~8개의 인지요소를 가정하였다. 따라서 본 연구에서는 인지요소의 수를 5개와 8개로 설정하고 [그림 Ⅴ-3]과 같은 두 가지 위계 구조를 가정하였다. 인지요소 5개로 이루어진 위계 구조가 인지요소 8개로 이루어진 위계 구조의 부분 그래프가 되도록 설정하여 구조의 형태에 따른 결과 차이를 줄이고자 하였다.



(a) 구조 1: 인지요소 5개 위계 구조 (b) 구조 2: 인지요소 8개 위계 구조

[그림 Ⅴ-3] 모의실험에서 가정한 인지요소 위계 구조

3.1.2. 모의피험자의 진 인지요소패턴의 생성

모의피험자의 수는 1000명으로 가정하였으며, [그림 Ⅴ-3]의 각 인지요소 위계 구조 하에서 가능한 인지요소패턴들 중에서 임의로 한 패턴을 선택하여 각 모의피험자의 진 인지요소패턴으로 지정하는 방식으로 모의피험자들의 진 인지요소패턴을 생성하였다. 가능한 인지요소패턴의 수는

구조 1([그림 Ⅴ-3]의 (a))의 경우 10가지, 구조 2([그림 Ⅴ-3]의(b))의 경우 26가지이다. 이러한 생성 방식은 모의피험자들의 진 인지요소패턴의 분포를 균등분포로 가정한 것이라 볼 수 있다. 각 인지요소 위계 구조 하에서 실제 생성된 모의피험자들의 진 인지요소패턴의 분포는 [그림 Ⅴ-4]에 제시되어 있다.

(a) 구조 1(인지요소 5개)

#	인지요소패턴	모의피험자 수
1	(0,0,0,0,0)	104
2	(1,0,0,0,0)	110
3	(1,1,0,0,0)	95
4	(1,0,1,0,0)	76
5	(1,1,1,0,0)	107
6	(1,1,0,1,0)	88
7	(1,1,1,1,0)	108
8	(1,0,1,0,1)	91
9	(1,1,1,0,1)	115
10	(1,1,1,1,1)	106
	합	1000

(b) 구조 2(인지요소 8개)

#	인지요소패턴	모의피험자 수
1	(0,0,0,0,0,0,0,0)	40
2	(1,0,0,0,0,0,0,0)	52
3	(1,1,0,0,0,0,0,0)	32
4	(1,0,1,0,0,0,0,0)	36
5	(1,1,1,0,0,0,0,0)	33
6	(1,1,0,1,0,0,0,0)	45
7	(1,1,1,1,0,0,0,0)	41
8	(1,0,1,0,1,0,0,0)	31
9	(1,1,1,0,1,0,0,0)	29
10	(1,1,1,1,1,0,0,0)	34
11	(1,1,1,0,0,1,0,0)	34
12	(1,1,1,1,0,1,0,0)	52
13	(1,1,1,0,1,1,0,0)	43
14	(1,1,1,1,1,1,0,0)	39
15	(1,1,0,1,0,0,1,0)	36
16	(1,1,1,1,0,0,1,0)	32
17	(1,1,1,1,1,0,1,0)	43
18	(1,1,1,1,0,1,1,0)	36
19	(1,1,1,1,1,1,1,0)	37
20	(1,0,1,0,1,0,0,1)	43
21	(1,1,1,0,1,0,0,1)	46
22	(1,1,1,1,1,0,0,1)	34
23	(1,1,1,0,1,1,0,1)	35
24	(1,1,1,1,1,1,0,1)	43
25	(1,1,1,1,1,0,1,1)	33
26	(1,1,1,1,1,1,1,1)	41
	합	1000

[그림 Ⅴ-4] 모의피험자들의 진 인지요소패턴 분포

3.1.3. 문제은행

문제은행의 크기의 경우, 인지요소 5개와 8개를 가정하고 문제은행의 크기를 350문항으로 설정하였던 Wang(2013)의 연구와 유사하게 400문항으로 설정하였다. '문제은행의 질' 요인은 높음과 낮음의 두 수준을 가정하였다. 이에 따라, 총 4개('인지요소 위계 구조' 2 × '문제은행의 질' 2)의 문제은행을 다음의 과정에 따라 생성하였다:

① 두 인지요소 위계 구조 하에서 400문항의 Q-행렬을 각각 생성한다. 본 실험에서는 Zheng과 Chang(2016)에서 Q-행렬을 생성한 방식과 동일하게 각 문항이 각 인지요소를 측정하는지의 여부(1/0)를 베르누이 분포 Bernoulli(0.3)에서 무선 추출한 값으로 결정하였으며, 이때 각 문항은 적어도 하나의 인지요소는 측정하도록 만들었다.

② ①에서 생성된 Q-행렬을 고정하고 질이 높은 문제은행과 질이 낮은 문제은행의 문항모수들을 각각 생성한다. 본 실험에서는 측정모형으로 DINA 모형을 가정하였으며, 이때 문항모수는 부주의 모수와 추측 모수 두 가지이다. DINA 모형을 가정하고 문제은행의 질을 모의실험 요인으로 고려했던 Wang(2013), Kaplan 외(2015), Zheng과 Chang(2016)등의 연구에서는 질이 높은 문제은행의 문항모수와 질이 낮은 문제은행의 문항모수를 범위가 다른 두 균등분포에서 각각 추출하는 방식으로 문항모수들을 생성하였다. 예를 들어 Zheng과 Chang(2016)은 질이 높은 문항들의 부주의 모수와 추측 모수는 균등분포 $U(0.05, 0.25)$에서, 질이 낮은 문항들의 모수들은 $U(0.1, 0.3)$에서 무선 추출하였다. 따라서 본 실험에서도 이와 유사한 방식을 따라 질이 높은 문제은행의 문항모수들은 $U(0.1, 0.25)$에서, 질이 낮은 문제은행의 문항모수들은 $U(0.15, 0.3)$에서 각각 무선 추출하였다. 적응검사가 진행되는 동안 각 문항의 모수는 알려진(known) 것으로 간주되어 확률 갱신에 사용되었다.

3.1.4. 문항 선택 방법

‘문항 선택 방법’ 요인의 경우, 그 동안 여러 선행 연구를 통해 인지진단 컴퓨터 적응검사의 문항 선택 방법으로 제안되어온 다음의 6가지 문항 선택 지수들을 조건으로 고려하였다: PWKL, MPWKL, PWCDI, PWACDI, SHE, MI. 또한, 고정-길이 규칙이 종료 규칙으로 사용되는 경우, 일종의 비교군으로서 다음 문항을 임의로 선택하는 RANDOM 방법도 요인으로 고려하였다.[21)]

한편, 각 모의피험자에 대해, 첫 번째 문항은 임의로 선택하고 그 문항을 모든 문항 선택 방법에 대해서 동일하게 고정하였다. 이는 첫 문항의 차이로 인해 문항 선택 방법 간 결과의 차이가 발생하는 것을 방지하기 위함이었다.

3.1.5. 모의피험자 인지요소패턴의 추정

본 모의실험에서 인지진단 컴퓨터 적응검사가 진행되는 동안 각 모의피험자의 인지요소패턴은 MAP 추정 방식으로 추정되었으며, 이때 사전분포로는 균등분포를 가정하였다(베이지안 네트워크의 CPT에서 인지요소 변수들에 대한 모수인 γ들의 사전 확률 또한 균등분포를 나타내도록 설정되었다). 구체적으로, Non-hierarchical DINA 모형 기반 검사에서는 2^K (K: 인지요소 수)개의 가능한 인지요소패턴들 중에서 사후 확률이 가장 높은 패턴이 추정치로 선택되었으며, BN-H-DINA 모형 기반 검사에서는 위계 구조에 따라 줄어든 수(구조 1에서는 10개, 구조 2에서는 26개)의 인지요소패턴들 중에서 사후 확률이 가장 높은 패턴이 추정치로 선택되었다.

---

21) 종료 규칙이 고정-정확도 규칙인 경우에는 RANDOM 방법으로 문항을 선택하면 사전에 설정한 정확도에 도달하기까지 지나치게 많은 문항이 필요한 상황이 발생하기 때문에 요인에서 제외하였다.

3.1.6. 검사 종료 규칙

검사 종료 규칙으로는 고정-길이 규칙과 고정-정확도 규칙을 모두 요인으로 고려하였다. 고정-길이 규칙으로는 Wang(2013), Zheng과 Chang(2016)에서와 동일하게 검사 길이가 5문항, 10문항, 20문항일 때 검사를 종료하는 조건을 설정하였다. 고정-정확도 규칙으로는 Zheng과 Chang(2016)에서와 같이 Hsu 외(2013)의 규칙1을 고려하여 최대 사후확률이 0.7, 0.8, 0.9 이상일 때 검사를 종료하는 조건을 설정하였다.

이하에서는 서술의 편의를 위해 고정-길이 규칙이 적용된 실험을 '고정-길이 검사 실험'으로, 고정-정확도 규칙이 적용된 실험을 '변동-길이 검사 실험'으로 부르기로 하겠다.

3.1.7. 평가 지표

고정-길이 검사 실험에서 적응검사의 효율성은 추정된 인지요소패턴의 정확도에 의해 결정된다. Xu 외(2003), Wang 외(2011), Kaplan 외(2015), Wang(2013), Zheng과 Chang(2016) 등은 개별 인지요소 수준에서 추정의 정확도와 인지요소패턴 수준에서 추정의 정확도를 고정-길이 검사 실험의 평가 기준으로 고려하였다. 따라서 본 실험에서도 선행 연구들과 동일하게 다음의 두 평가 지표를 설정하였다.

1) Attribute correct classification rate (ACCR)

개별 인지요소 수준에서 각 인지요소가 얼마나 정확하게 추정되었는지를 나타내는 지표이다. 각 k번째 인지요소 α_k에 대해 α_k의 숙달여부가 옳게 추정된 모의피험자의 비율로 계산되며 식으로 표현하면 다음과 같다:

$$ACCR_k = \sum_{i=1}^{1000} I(\alpha_{ik} = \hat{\alpha}_{ik})/1000,\ k = 1,2,...,K \qquad (5.14)$$

여기서 I는 지시함수로, $I(\alpha_{ik}=\hat{\alpha}_{ik})$는 피험자 i의 k번째 인지요소에 대한 진 숙달 여부 α_{ik}와 최종적으로 추정된 숙달 여부 $\hat{\alpha}_{ik}$이 일치하면 1, 일치하지 않으면 0이다.

2) Pattern correct classification rate (PCCR)

인지요소패턴이 얼마나 정확하게 추정되었는지를 나타내는 지표로, 인지요소패턴이 옳게 추정된 모의피험자의 비율로 계산되며 식으로 표현하면 다음과 같다:

$$PCCR=\sum_{i=1}^{1000} I(\boldsymbol{\alpha}_i=\hat{\boldsymbol{\alpha}}_i)/1000 \tag{5.15}$$

반면, 변동-길이 검사 실험에서 적응검사의 효율성은 검사 길이에 의해 결정된다. 본 실험에서는 변동-길이 검사를 고려했던 Hsu 외(2013), Kaplan 외(2015) 등의 연구와 유사하게 피험자 i가 받은 검사의 길이를 l_i라 할 때 전체 모의피험자들에 대한 평균 검사 길이 $Mean(l_i)=\sum_{i=1}^{1000} l_i/1000$를 주된 평가 지표로 사용하였으며, 검사 길이의 최솟값, 최댓값, 표준편차도 함께 확인하였다.

3.2. 모의실험 결과

3.2.1. 고정-길이 검사

<표 Ⅴ-1>부터 <표 Ⅴ-4>에는 총 4개('인지요소 위계 구조' 2 × '문제은행의 질' 2)의 조건 하에서 고정-길이 검사 실험의 ACCR 결과가 제시되어 있다.

먼저, 동일한 조건 하에서 인지요소 위계를 모델링함으로써 개별 인지요소 수준에서의 추정의 정확도인 ACCR이 실제로 향상되었는지를 살펴보자. <표 Ⅴ-1>부터 <표 Ⅴ-4>까지의 모든 결과는 인지요소 위계를 모델링하지 않은 경우(non-hierarchical DINA) 보다 인지요소 위계를 모델링한 경우(BN-H-DINA)에 ACCR이 더 높게 나타남을 보여주고 있다. 예를 들어, 인지요소 5개인 위계 구조(구조 1)에서 질이 높은 문제은행을 사용한 실험의 결과인 <표 Ⅴ-1>에서 검사 길이가 5문항인 경우에 주목해보면, 위계를 모델링하지 않은 non-hierarchical DINA 모형이 사용되었을 때 ACCR은 약 0.70~0.94의 값을 보인 반면, 위계를 모델링한 BN-H-DINA 모형이 사용되었을 때에는 그 값들이 약 0.93~0.97까지로 더 높은 범위에 분포하였음을 확인할 수 있다.[22] 이러한 경향은 <표 Ⅴ-2>, <표 Ⅴ-3>, <표 Ⅴ-4>에서도 유사하게 나타난다. 이와 같은 결과는 실제로 인지요소 간 위계 관계가 존재할 때 이를 인지진단모형에 직접적으로 반영하여 모델링하고 이를 기반으로 적응검사를 진행함으로써 피험자들의 개별 인지요소 숙달 여부를 더 정확하게 진단할 수 있음을 보여준다.

22) RANDOM 문항 선택 방법은 다른 문항 선택 방법들과의 비교를 위한 비교군으로 포함된 방법이었기 때문에 본 결과 서술에는 포함시키지 않았으며, ACCR과 PCCR 값은 소수점 셋째자리에서 반올림한 값을 기준으로 서술하였다.

한편, BN-H-DINA 모형으로 인지요소 위계가 모델링된 경우, 특히 검사 길이가 5문항으로 짧을 때 인지요소 위계 구조 상 가장 낮은 수준의 인지요소인 A_1의 ACCR 값들이 다른 상위 수준의 인지요소들의 ACCR 값들보다 높게 나타난다. 예를 들어, <표 Ⅴ-2>에서 BN-H-DINA 모형 하에서 검사 길이가 5문항으로 짧은 경우의 결과를 보면, A_1의 ACCR은 약 0.94~0.95의 값을 보이는 반면 다른 인지요소들(A_2, A_3, A_4, A_5)의 ACCR은 0.88~0.91 정도의 값을 나타낸다. 이러한 경향은 <표 Ⅴ-1>, <표 Ⅴ-3>, <표 Ⅴ-4>의 결과에서도 유사하게 확인되며, 문제은행의 질이 낮을수록, 그리고 인지요소의 수가 많을수록 그 차이가 더 크게 나타난다. 그러나 인지요소 위계가 모델링되지 않은 Non-hierarchical DINA 모형 하에서의 결과를 확인해보면 이와 같은 경향성이 전혀 나타나지 않는다. 이는 인지요소 위계를 모델링한 인지진단모형을 사용하면 검사 초기에 최하위 수준 인지요소의 숙달 여부는 다른 상위 인지요소들의 숙달 여부보다 상대적으로 잘 진단해낼 수 있음을 보여주는 결과라고 할 수 있다.

<표 Ⅴ-1> 고정-길이 검사의 ACCR 결과: 구조 1, 문제은행의 질이 높은 경우

검사 길이	문항선택 방법	Non-hierarchical DINA					BN-H-DINA				
		A_1	A_2	A_3	A_4	A_5	A_1	A_2	A_3	A_4	A_5
5	RANDOM	0.693	0.737	0.741	0.748	0.755	0.904	0.773	0.786	0.829	0.853
	PWKL	0.748	0.856	0.804	0.848	0.856	0.960	0.942	0.936	0.952	0.933
	MPWKL	0.763	0.842	0.815	0.819	0.852	0.965	0.942	0.938	0.948	0.929
	PWCDI	0.768	0.935	0.900	0.937	0.930	0.956	0.956	0.931	0.954	0.942
	PWACDI	0.777	0.929	0.872	0.928	0.906	0.957	0.940	0.926	0.938	0.944
	SHE	0.719	0.926	0.906	0.924	0.941	0.972	0.950	0.931	0.944	0.942
	MI	0.702	0.921	0.899	0.928	0.916	0.966	0.954	0.945	0.960	0.943
10	RANDOM	0.762	0.896	0.824	0.921	0.886	0.966	0.946	0.912	0.955	0.943
	PWKL	0.757	0.810	0.819	0.837	0.850	0.912	0.862	0.881	0.924	0.939
	MPWKL	0.904	0.963	0.963	0.970	0.958	0.986	0.989	0.993	0.994	0.991
	PWCDI	0.948	0.977	0.976	0.984	0.979	0.991	0.994	0.994	0.989	0.995
	PWACDI	0.967	0.987	0.981	0.984	0.985	0.995	0.994	0.994	0.990	0.993
	SHE	0.908	0.968	0.949	0.956	0.965	0.990	0.987	0.990	0.992	0.990
	MI	0.947	0.976	0.981	0.983	0.985	0.991	0.993	0.991	0.992	0.989
20	RANDOM	0.954	0.982	0.972	0.976	0.981	0.991	0.992	0.990	0.992	0.994
	PWKL	0.932	0.966	0.952	0.960	0.961	0.993	0.987	0.984	0.993	0.995
	MPWKL	0.856	0.917	0.935	0.929	0.939	0.931	0.946	0.941	0.979	0.983
	PWCDI	0.999	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	PWACDI	0.994	0.999	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000
	SHE	0.993	1.000	0.997	0.998	0.999	0.998	1.000	1.000	1.000	1.000
	MI	0.994	0.994	0.996	0.998	0.997	1.000	1.000	1.000	1.000	1.000

<표 Ⅴ-2> 고정-길이 검사의 ACCR 결과: 구조 1, 문제은행의 질이 낮은 경우

검사 길이	문항선택 방법	Non-hierarchical DINA					BN-H-DINA				
		A_1	A_2	A_3	A_4	A_5	A_1	A_2	A_3	A_4	A_5
5	RANDOM	0.680	0.699	0.672	0.673	0.714	0.900	0.741	0.709	0.790	0.801
	PWKL	0.757	0.720	0.786	0.783	0.756	0.940	0.887	0.878	0.902	0.896
	MPWKL	0.859	0.744	0.864	0.919	0.860	0.954	0.879	0.894	0.900	0.934
	PWCDI	0.853	0.730	0.854	0.888	0.838	0.937	0.895	0.885	0.910	0.892
	PWACDI	0.787	0.742	0.817	0.864	0.822	0.943	0.849	0.905	0.913	0.894
	SHE	0.849	0.709	0.835	0.865	0.833	0.949	0.889	0.888	0.915	0.910
	MI	0.850	0.761	0.866	0.905	0.846	0.944	0.886	0.903	0.914	0.890
10	RANDOM	0.759	0.796	0.769	0.770	0.805	0.911	0.847	0.810	0.887	0.872
	PWKL	0.862	0.875	0.904	0.916	0.913	0.982	0.962	0.960	0.972	0.967
	MPWKL	0.901	0.948	0.929	0.973	0.954	0.985	0.967	0.978	0.977	0.984
	PWCDI	0.902	0.940	0.934	0.948	0.938	0.980	0.973	0.971	0.978	0.968
	PWACDI	0.880	0.926	0.921	0.928	0.928	0.983	0.964	0.957	0.987	0.969
	SHE	0.902	0.937	0.936	0.951	0.931	0.983	0.969	0.971	0.975	0.980
	MI	0.919	0.926	0.932	0.963	0.953	0.979	0.972	0.970	0.979	0.973
20	RANDOM	0.835	0.866	0.855	0.874	0.899	0.935	0.911	0.899	0.944	0.953
	PWKL	0.978	0.988	0.985	0.991	0.994	0.998	0.998	0.996	0.998	0.996
	MPWKL	0.981	0.997	0.993	0.997	0.999	0.998	0.996	0.997	1.000	1.000
	PWCDI	0.980	0.994	0.989	0.993	0.988	0.997	0.998	0.996	0.997	0.998
	PWACDI	0.975	0.985	0.982	0.986	0.990	0.999	0.996	0.997	0.999	0.997
	SHE	0.991	0.991	0.992	0.994	0.991	0.998	0.995	0.996	0.998	0.997
	MI	0.985	0.991	0.986	0.999	0.992	0.997	0.995	0.997	0.997	0.998

<표 Ⅴ-3> 고정-길이 검사의 ACCR 결과: 구조 2, 문제은행의 질이 높은 경우

검사 길이	문항선택 방법	Non-hierarchical DINA								BN-H-DINA							
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
5	RANDOM	0.651	0.689	0.642	0.678	0.655	0.594	0.612	0.643	0.957	0.751	0.757	0.701	0.752	0.731	0.788	0.796
	PWKL	0.698	0.666	0.682	0.707	0.681	0.648	0.713	0.685	0.991	0.932	0.945	0.901	0.900	0.768	0.852	0.823
	MPWKL	0.679	0.678	0.687	0.717	0.720	0.663	0.676	0.645	0.983	0.939	0.927	0.892	0.899	0.765	0.853	0.823
	PWCDI	0.885	0.696	0.550	0.947	0.765	0.752	0.607	0.711	0.965	0.931	0.882	0.925	0.906	0.914	0.880	0.875
	PWACDI	0.878	0.629	0.585	0.929	0.737	0.691	0.658	0.703	0.954	0.932	0.894	0.904	0.870	0.906	0.898	0.879
	SHE	0.889	0.637	0.654	0.942	0.748	0.732	0.627	0.720	0.955	0.940	0.914	0.916	0.892	0.930	0.883	0.882
	MI	0.892	0.705	0.534	0.931	0.739	0.720	0.591	0.703	0.956	0.942	0.907	0.926	0.909	0.922	0.898	0.878
10	RANDOM	0.808	0.674	0.635	0.849	0.679	0.714	0.675	0.711	0.971	0.886	0.916	0.878	0.870	0.906	0.873	0.872
	PWKL	0.740	0.765	0.735	0.747	0.718	0.694	0.721	0.715	0.965	0.825	0.830	0.807	0.836	0.800	0.848	0.871
	MPWKL	0.821	0.782	0.784	0.865	0.824	0.837	0.824	0.825	0.996	0.987	0.975	0.977	0.968	0.973	0.957	0.957
	PWCDI	0.915	0.880	0.876	0.947	0.898	0.925	0.893	0.907	0.995	0.982	0.975	0.984	0.982	0.977	0.980	0.982
	PWACDI	0.926	0.900	0.898	0.957	0.912	0.923	0.904	0.921	0.992	0.987	0.983	0.979	0.974	0.982	0.982	0.970
	SHE	0.806	0.826	0.785	0.829	0.826	0.810	0.814	0.792	0.994	0.985	0.973	0.975	0.968	0.957	0.972	0.951
	MI	0.913	0.911	0.909	0.951	0.908	0.937	0.902	0.929	0.994	0.984	0.977	0.984	0.983	0.977	0.974	0.981
20	RANDOM	0.925	0.891	0.898	0.949	0.877	0.907	0.894	0.918	0.989	0.984	0.975	0.978	0.968	0.985	0.979	0.970
	PWKL	0.898	0.883	0.892	0.918	0.883	0.912	0.869	0.903	0.986	0.967	0.956	0.971	0.954	0.972	0.970	0.971
	MPWKL	0.833	0.843	0.826	0.821	0.840	0.810	0.812	0.830	0.967	0.893	0.894	0.918	0.915	0.891	0.940	0.937
	PWCDI	0.986	0.992	0.989	0.993	0.992	0.990	0.989	0.988	1.000	1.000	0.998	0.999	0.999	1.000	0.997	0.998
	PWACDI	0.982	0.989	0.985	0.989	0.991	0.991	0.986	0.993	0.998	0.997	0.997	1.000	0.996	0.998	0.998	0.999
	SHE	0.948	0.957	0.948	0.977	0.969	0.963	0.954	0.964	1.000	0.999	0.996	0.998	1.000	0.999	0.999	0.999
	MI	0.929	0.949	0.940	0.967	0.957	0.949	0.948	0.961	1.000	1.000	0.997	0.999	0.999	1.000	0.998	1.000

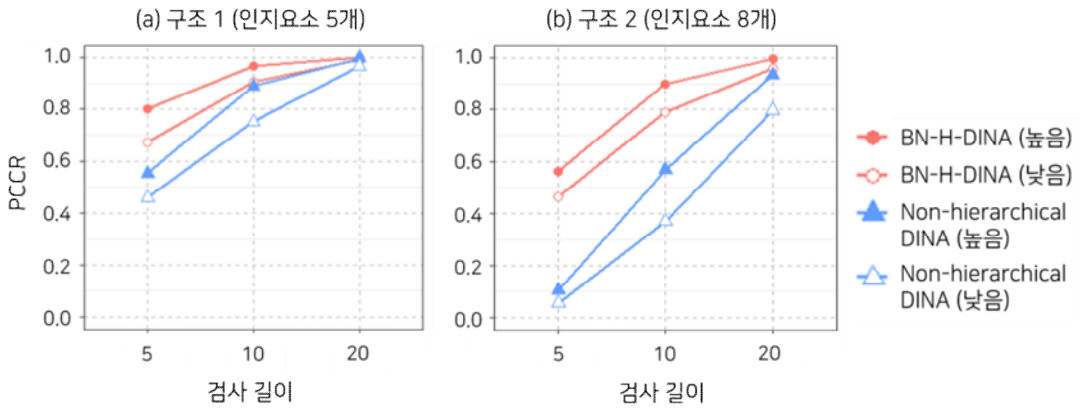
<표 Ⅴ-4> 고정-길이 검사의 ACCR 결과: 구조 2, 문제은행의 질이 낮은 경우

검사 길이	문항선택 방법	Non-hierarchical DINA								BN-H-DINA							
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
5	RANDOM	0.652	0.654	0.633	0.629	0.621	0.598	0.549	0.610	0.948	0.727	0.766	0.682	0.713	0.706	0.752	0.758
	PWKL	0.697	0.709	0.651	0.666	0.657	0.694	0.647	0.665	0.966	0.888	0.893	0.790	0.831	0.773	0.838	0.838
	MPWKL	0.691	0.712	0.662	0.653	0.646	0.639	0.625	0.646	0.968	0.904	0.895	0.799	0.828	0.757	0.830	0.847
	PWCDI	0.717	0.855	0.783	0.617	0.669	0.734	0.614	0.751	0.953	0.866	0.897	0.828	0.856	0.864	0.884	0.879
	PWACDI	0.740	0.850	0.753	0.636	0.627	0.742	0.588	0.701	0.957	0.879	0.899	0.811	0.803	0.890	0.862	0.855
	SHE	0.726	0.849	0.767	0.633	0.648	0.715	0.589	0.672	0.955	0.888	0.899	0.817	0.856	0.890	0.878	0.915
	MI	0.743	0.849	0.765	0.594	0.652	0.712	0.586	0.705	0.949	0.876	0.883	0.802	0.848	0.891	0.869	0.898
10	RANDOM	0.736	0.793	0.632	0.665	0.635	0.757	0.681	0.674	0.964	0.884	0.846	0.815	0.771	0.879	0.857	0.846
	PWKL	0.692	0.717	0.703	0.681	0.704	0.643	0.669	0.688	0.946	0.793	0.835	0.760	0.788	0.760	0.808	0.825
	MPWKL	0.781	0.838	0.763	0.772	0.784	0.795	0.801	0.806	0.988	0.962	0.973	0.925	0.928	0.938	0.935	0.940
	PWCDI	0.827	0.877	0.871	0.842	0.865	0.860	0.882	0.896	0.976	0.964	0.957	0.933	0.949	0.949	0.955	0.966
	PWACDI	0.826	0.885	0.867	0.839	0.856	0.873	0.879	0.905	0.977	0.965	0.958	0.944	0.954	0.950	0.963	0.965
	SHE	0.786	0.827	0.775	0.778	0.784	0.789	0.774	0.800	0.982	0.971	0.965	0.936	0.936	0.936	0.939	0.940
	MI	0.839	0.888	0.863	0.847	0.871	0.867	0.893	0.877	0.983	0.968	0.954	0.941	0.958	0.945	0.967	0.974
20	RANDOM	0.809	0.892	0.872	0.828	0.847	0.866	0.863	0.870	0.981	0.959	0.958	0.937	0.955	0.960	0.955	0.959
	PWKL	0.834	0.910	0.823	0.816	0.845	0.868	0.871	0.859	0.976	0.933	0.929	0.923	0.917	0.928	0.953	0.954
	MPWKL	0.783	0.813	0.793	0.807	0.772	0.721	0.730	0.819	0.955	0.853	0.879	0.851	0.844	0.854	0.900	0.911
	PWCDI	0.939	0.971	0.960	0.974	0.964	0.959	0.977	0.976	0.994	0.994	0.990	0.992	0.993	0.990	0.992	0.999
	PWACDI	0.935	0.974	0.965	0.975	0.949	0.966	0.970	0.974	0.996	0.997	0.988	0.991	0.990	0.995	0.991	0.995
	SHE	0.905	0.948	0.933	0.935	0.920	0.925	0.940	0.935	0.999	0.995	0.995	0.992	0.990	0.997	0.996	0.993
	MI	0.879	0.947	0.922	0.918	0.926	0.938	0.926	0.948	0.993	0.997	0.996	0.994	0.991	0.995	0.993	0.996

<표 Ⅴ-5> 고정-길이 검사의 PCCR 결과

검사 길이	문항선택 방법	Non-hierarchical DINA				BN-H-DINA			
		구조 1		구조 2		구조 1		구조 2	
		높음	낮음	높음	낮음	높음	낮음	높음	낮음
5	RANDOM	0.238	0.188	0.072	0.047	0.449	0.350	0.188	0.159
	PWKL	0.487	0.350	0.145	0.105	0.802	0.664	0.460	0.364
	MPWKL	0.554	0.461	0.102	0.054	0.800	0.670	0.561	0.465
	PWCDI	0.563	0.440	0.087	0.053	0.793	0.652	0.531	0.428
	PWACDI	0.529	0.405	0.123	0.075	0.798	0.659	0.519	0.416
	SHE	0.523	0.396	0.066	0.052	0.815	0.664	0.558	0.443
	MI	0.579	0.447	0.072	0.058	0.796	0.666	0.536	0.450
10	RANDOM	0.402	0.347	0.173	0.106	0.643	0.540	0.345	0.288
	PWKL	0.832	0.663	0.411	0.338	0.960	0.878	0.859	0.744
	MPWKL	0.890	0.753	0.586	0.369	0.966	0.908	0.898	0.788
	PWCDI	0.893	0.741	0.592	0.364	0.967	0.896	0.873	0.778
	PWACDI	0.844	0.739	0.553	0.408	0.960	0.892	0.857	0.728
	SHE	0.891	0.726	0.544	0.382	0.968	0.899	0.892	0.769
	MI	0.919	0.741	0.578	0.357	0.972	0.895	0.895	0.778
20	RANDOM	0.668	0.546	0.362	0.252	0.819	0.727	0.578	0.476
	PWKL	0.988	0.945	0.820	0.708	0.998	0.989	0.993	0.968
	MPWKL	0.997	0.968	0.934	0.798	1.000	0.992	0.994	0.959
	PWCDI	0.996	0.957	0.921	0.775	0.999	0.988	0.992	0.972
	PWACDI	0.989	0.943	0.890	0.753	0.999	0.990	0.986	0.947
	SHE	0.991	0.962	0.929	0.780	1.000	0.985	0.987	0.955
	MI	0.995	0.958	0.926	0.804	1.000	0.985	0.997	0.970

다음으로, 인지요소패턴 수준에서의 추정의 정확도인 PCCR의 결과를 살펴보자. 모든 조건에서의 PCCR 결과가 위의 <표 Ⅴ-5>에 제시되어 있다. 동일한 조건 하에서 인지요소 위계를 모델링함으로써 PCCR이 실제로 향상되었는지를 확인해보면, 앞의 ACCR 결과에서도 예상할 수 있듯이 인지요소 위계를 모델링하지 않은 non-hierarchical DINA 모형에서 보다 인지요소 위계를 모델링한 BN-H-DINA 모형 하에서 PCCR이 더 높게 나타남을 알 수 있다.



[그림 Ⅴ-5] 고정-길이 검사 실험에서 MPWKL 방법의 PCCR 결과

다른 실험 요인들이 PCCR 값에 어떤 영향을 미치는지를 함께 확인하기 위해 각 인지요소 위계 구조 하에서 MPWKL 문항 선택 방법이 사용된 경우에 PCCR 값이 검사 길이에 따라 어떻게 변하는지를 그래프로 나타내면 [그림 Ⅴ-5]와 같다. 인지요소가 5개인 구조 1의 경우([그림 Ⅴ-5]의 (a) 참고), 검사 길이가 5문항으로 매우 짧은 경우에는 문제은행의 질에 관계없이 non-hierarchical DINA 모형 하에서의 PCCR 값보다 BN-H-DINA 모형 하에서의 PCCR 값이 더 높게 나타난다. 두 값의 차이는 인지요소 위계 구조를 모델링함으로써 패턴 추정의 정확도가 향상되는 정도를 나타낸다. 그러나 검사 길이가 길어짐에 따라 non-hierarchical DINA 모형 하에서 PCCR 값이 빠르게 높아지면서 그 차이는 점차 줄어들며, 검사 길이가 20문항이 되면 두 모형 하에서 PCCR은 모두 1에 가까운 값에 도달하면서 그 차이가 거의 사라진다. 반면, 인지요소가 8개인 구조 2의 경우([그림 Ⅴ-5]의 (b) 참고), 검사 길이가 늘어날수록 두 모형 하에서의 PCCR 차이가 줄어드는 결과는 동일하게 나타났으나, 문제은행의 질이 낮은 경우에는 검사 길이가 20문항까지 늘어나도 non-hierarchical DINA 모형 하에서의 PCCR이 BN-H-DINA 모형의 수준에는 도달하지 못하였다. 이러한 점에 비추어보면, 어떤 검사가 측정하고자 하는 인지요소의 수는 많지만 현실적으로 시간상의 제약이나 문제은행 관리 등의 이유로 인지요소의 수에 비해 검사의 길이는

짧게 유지해야 하는 상황이라면 인지요소 위계를 모델링함으로써 보다 정확한 인지진단 결과를 얻을 수 있을 것이다.

마지막으로 문항 선택 방법에 따라 PCCR 결과가 어떻게 달라지는지를 확인해보자. 우선 <표 Ⅴ-5>에서 동일한 조건하에서 문항 선택 방법들 간의 PCCR 값의 차이를 비교해보면 일단 문항을 특정 기준 없이 임의로 선택하는 RANDOM 방법이 다른 방법들에 비해 낮은 PCCR을 보임을 확인할 수 있다. RANDOM을 제외한 나머지 방법들 중에서는, 검사 길이가 짧은 경우에 PWKL 지수 기반의 문항 선택 방법을 사용했을 때 PCCR이 다른 방법들에 비해 약간 낮은 값을 나타낸다. 이러한 결과는 선행 연구들(Wang, 2013; Kaplan et al., 2015; Zheng & Chang, 2016)에서 확인된 결과와도 일치하며, 그 이유는 1절에서도 언급했듯이 PWKL이 각 단계에서 잠정적으로 추정된 인지요소패턴 $\hat{\boldsymbol{\alpha}}_i^{(t-1)}$에 기초하여 문항을 선택하지만 검사 초기에는 그 추정치의 정확도가 떨어지므로 적절한 문항을 선택하지 못했기 때문이었을 것으로 판단된다.

3.2.2. 변동-길이 검사

<표 Ⅴ-6>과 <표 Ⅴ-7>에는 각 인지요소 위계 구조 하에서 변동-길이 검사 실험의 검사 길이의 기술 통계량(최솟값, 최댓값, 평균, 표준편차)과 인지요소패턴 수준에서의 추정 정확도인 PCCR이 함께 제시되어 있다.

<표 Ⅴ-6>과 <표 Ⅴ-7>에서 non-hierarchical DINA 모형의 결과와 BN-H-DINA 모형의 결과를 비교해보면, 변동-길이 검사에서도 고정-길이 검사에서와 유사하게 인지요소 위계를 모델링함으로써 효율성이 향상되었음을 확인할 수 있다. 검사 길이의 측면에서는 모든 조건 하에서 인지진단모형으로 non-hierarchical DINA 모형을 사용한 경우에 비해 BN-H-DINA 모형을 사용한 경우에 검사 길이의 평균이 대략 30~50% 가량 줄어들었다. 예를 들어, 인지요소 8개인 위계 구조(구조 2)에서 질이 낮은 문제은행을 사용한 실험의 결과인 <표 Ⅴ-7>에서 검사 종료 규칙으로 최대 사후 확률값 0.9가 사용된 경우에, non-hierarchical DINA 모형 하에서 검사 길이의 평균은 대략 21문항 정도였던 반면, BN-H-DINA 모형을 사용하면 검사 길이의 평균이 약 12문항으로 줄어드는 것을 확인할 수 있다. 패턴 수준의 추정 정확도 측면에서는 거의 모든 조건 하에서[23] 인지진단모형으로 non-hierarchical DINA 모형을 사용한 경우에 비해 BN-H-DINA 모형을 사용한 경우에 더 높은 PCCR 값이 나타난다. 두 측면의 결과를 종합해보면, 변동-길이 검사에서는 인지요소 위계를 모델링함으로써 보다 적은 문항으로 더 높은 정확도의 인지진단 결과를 얻을 수 있다고 할 수 있다.

---

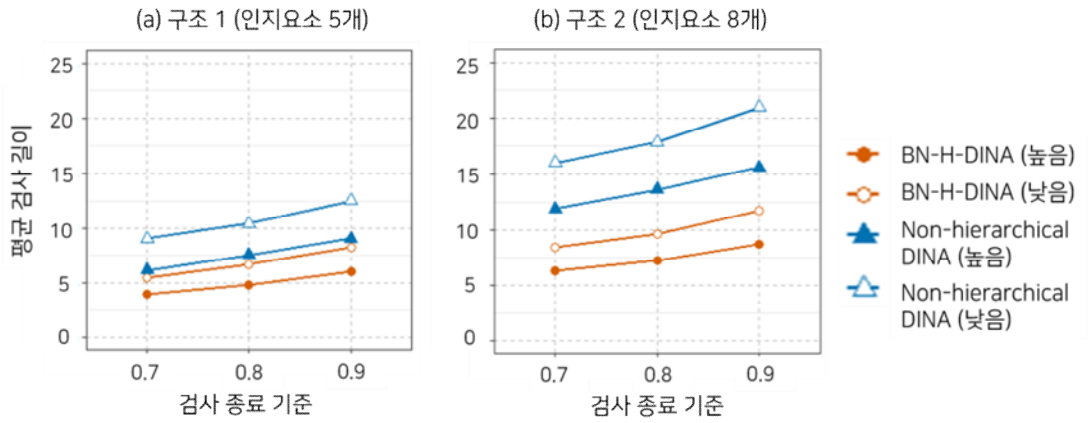
23) 예외적으로, 구조 1에서 검사 종료 규칙으로 최대 사후 확률값 0.7, 문항 선택 기준으로 MPWKL 방법과 질이 높은 문제은행이 사용된 경우에는 non-hierarchical DINA 모형의 PCCR 값이 0.802로 BN-H-DINA 모형의 PCCR 값인 0.798보다 살짝 높게 나타났다(<표 Ⅴ-6> 참고).

<표 Ⅴ-6> 변동-길이 검사의 검사 길이에 대한 기술통계량 및 PCCR 결과: 구조 1의 경우

검사 종료규칙	문항선택 방법	Non-hierarchical DINA										BN-H-DINA									
		높음					낮음					높음					낮음				
		Min	Max	Mean	SD	PCCR	Min	Max	Mean	SD	PCCR	Min	Max	Mean	SD	PCCR	Min	Max	Mean	SD	PCCR
0.7	PWKL	2	18	7.1	2.7	0.762	3	29	9.7	4.1	0.755	2	10	4.1	1.0	0.814	2	17	5.5	2.1	0.776
	MPWKL	3	14	6.1	1.3	0.802	4	26	9.0	3.0	0.732	2	9	4.0	0.9	0.798	2	16	5.4	1.9	0.805
	PWCDI	3	15	6.1	1.4	0.765	3	23	8.7	3.0	0.755	2	8	4.0	0.9	0.789	2	15	5.5	2.0	0.805
	PWACDI	2	20	6.2	1.7	0.730	3	33	9.1	3.8	0.717	2	8	4.1	1.1	0.802	2	17	5.6	2.2	0.803
	SHE	3	16	6.1	1.3	0.778	4	22	9.0	2.9	0.723	2	8	4.0	0.8	0.819	2	14	5.4	1.9	0.757
	MI	3	13	6.1	1.3	0.783	4	30	9.0	3.1	0.726	2	10	4.0	0.9	0.804	2	14	5.3	1.8	0.783
0.8	PWKL	3	23	8.3	3.3	0.841	3	39	11.2	4.8	0.850	2	14	4.9	1.5	0.855	2	24	6.8	2.7	0.851
	MPWKL	3	19	7.5	2.1	0.818	4	25	10.4	3.5	0.830	2	11	4.8	1.4	0.848	2	20	6.6	2.7	0.869
	PWCDI	3	20	7.4	2.1	0.818	4	25	10.5	3.5	0.817	2	10	4.7	1.4	0.863	2	17	6.6	2.6	0.861
	PWACDI	3	29	7.4	2.4	0.819	3	38	10.8	4.5	0.843	2	12	4.8	1.5	0.854	2	24	7.1	3.2	0.865
	SHE	3	17	7.5	2.0	0.832	4	27	10.4	3.3	0.792	2	12	4.7	1.3	0.851	2	20	6.4	2.5	0.857
	MI	4	21	7.6	2.3	0.814	4	30	10.6	3.6	0.823	2	13	4.8	1.4	0.869	2	21	6.5	2.5	0.851
0.9	PWKL	3	27	9.8	3.9	0.939	4	36	12.9	5.4	0.920	2	18	6.1	2.1	0.942	3	20	8.2	3.0	0.938
	MPWKL	4	21	9.0	2.6	0.926	5	31	12.5	4.0	0.926	2	22	6.0	2.1	0.936	3	23	8.2	3.2	0.933
	PWCDI	4	26	9.2	3.0	0.918	4	37	12.4	4.2	0.930	2	17	6.0	2.0	0.941	3	24	8.2	3.3	0.943
	PWACDI	3	26	9.1	3.3	0.922	4	44	12.5	5.2	0.924	2	18	6.4	2.4	0.956	3	29	8.5	3.5	0.932
	SHE	4	21	9.1	2.7	0.920	5	35	12.5	4.4	0.904	2	16	6.1	2.1	0.934	3	24	8.0	3.1	0.934
	MI	4	25	9.0	2.8	0.905	5	29	12.4	4.1	0.915	2	16	6.0	2.0	0.951	3	28	8.1	3.3	0.926

<표 Ⅴ-7> 변동-길이 검사의 검사 길이에 대한 기술통계량 및 PCCR 결과: 구조 2의 경우

검사 종료규칙	문항선택 방법	Non-hierarchical DINA										BN-H-DINA									
		높음					낮음					높음					낮음				
		Min	Max	Mean	SD	PCCR	Min	Max	Mean	SD	PCCR	Min	Max	Mean	SD	PCCR	Min	Max	Mean	SD	PCCR
0.7	PWKL	4	46	14.7	7.7	0.727	4	51	18.3	8.9	0.728	2	16	7.0	2.1	0.771	3	26	8.9	3.1	0.767
	MPWKL	6	28	11.9	3.9	0.717	7	53	16.0	5.7	0.703	3	16	6.3	1.7	0.785	4	22	8.4	3.0	0.755
	PWCDI	5	36	11.7	3.9	0.720	6	40	15.7	5.8	0.719	3	15	6.4	1.8	0.777	4	22	8.3	3.0	0.752
	PWACDI	4	42	12.2	5.1	0.747	5	57	16.5	7.7	0.753	2	19	6.7	2.3	0.789	3	25	8.8	3.5	0.747
	SHE	6	34	11.9	3.9	0.707	7	44	15.8	5.7	0.688	3	17	6.3	1.7	0.775	4	24	8.1	2.8	0.773
	MI	6	36	11.8	4.1	0.754	7	49	15.9	5.8	0.705	3	17	6.4	1.8	0.784	4	25	8.1	2.7	0.765
0.8	PWKL	4	46	16.3	8.0	0.832	5	72	20.2	9.7	0.814	3	20	8.1	2.6	0.848	3	29	10.3	3.6	0.848
	MPWKL	6	38	13.6	4.5	0.809	7	59	17.9	6.7	0.810	3	18	7.2	2.1	0.847	4	29	9.6	3.5	0.836
	PWCDI	5	34	13.4	4.6	0.805	7	59	17.7	6.8	0.819	3	20	7.4	2.3	0.827	4	27	9.8	3.5	0.852
	PWACDI	5	45	13.9	6.1	0.819	6	57	18.5	8.7	0.834	3	21	7.8	2.7	0.852	3	36	10.4	4.2	0.848
	SHE	6	33	13.2	4.2	0.810	7	53	18.2	6.8	0.817	3	20	7.3	2.3	0.843	4	25	9.6	3.3	0.826
	MI	6	63	13.6	4.8	0.797	7	63	17.7	6.5	0.816	4	21	7.3	2.3	0.868	4	29	9.7	3.5	0.845
0.9	PWKL	5	54	18.1	8.6	0.921	6	71	22.8	11.3	0.909	3	24	9.6	3.0	0.944	4	32	12.2	4.6	0.928
	MPWKL	6	42	15.6	5.1	0.921	7	57	21.0	8.0	0.899	3	27	8.8	2.9	0.939	5	34	11.7	4.4	0.935
	PWCDI	6	44	15.8	5.6	0.916	7	64	20.8	8.5	0.918	3	27	8.9	3.1	0.927	5	42	11.8	4.7	0.930
	PWACDI	5	58	15.9	7.2	0.927	6	80	21.6	10.6	0.904	3	29	9.3	3.4	0.950	4	40	12.0	4.9	0.917
	SHE	6	37	15.7	5.1	0.902	8	61	20.7	8.0	0.901	3	22	8.7	2.6	0.931	5	38	11.7	4.4	0.917
	MI	6	39	15.6	5.0	0.926	8	57	20.9	8.0	0.880	3	25	8.9	3.0	0.936	5	35	11.6	4.3	0.943



[그림 Ⅴ-6] 변동-길이 검사 실험에서 MPWKL 방법의 평균 검사 길이

다른 실험 요인들이 평균 검사 길이에 어떤 영향을 미치는지 확인하기 위해 각 인지요소 위계 구조 하에서 MPWKL 문항 선택 방법이 사용된 경우에 평균 검사 길이가 검사 종료 기준에 따라 어떻게 변하는지를 그래프로 나타내면 [그림 Ⅴ-6]과 같다. 우선, 동일한 조건 하에서 인지요소가 5개인 구조 1보다 인지요소가 8개인 구조 2의 경우에서 평균 검사 길이가 더 길게 나타나고, 문제은행의 질이 높은 경우와 낮은 경우의 평균 검사 길이의 차이 또한 구조 1보다 구조 2에서 크게 나타남을 알 수 있다. 또한, 두 위계 구조 하에서 인지요소 위계를 모델링함으로써 줄어드는 평균 검사 길이의 비율은 검사 종료 기준에 관계없이 대체로 유사함을 확인할 수 있다.

마지막으로 변동-길이 검사에서 문항 선택 방법에 따라 결과가 어떻게 달라지는지 비교해보면, 일단 PWKL 지수 기반의 문항 선택 방법이 다른 방법에 비해 가장 긴 평균 검사 길이를 나타냄을 확인할 수 있다. 이는 고정-길이 검사에서의 결과와도 일맥상통한다. PWKL을 제외한 다섯 가지 방법들(MPWKL, PWCDI, PWACDI, SHE, MI) 중에서는 PWACDI가 나머지 네 가지 방법에 비해 평균 검사 길이가 약간 더 길게 나타난다. 이러한 결과가 나타난 이유에 대해서는 다음과 같이 부분적으로 설명이 가능할 것으로 판단된다. KL 지수를 기반으로 발전된 MPWKL, PWCDI, PWACDI의 지수 식을 비교해보면, PWACDI는 오직

Hamming 거리가 1인 패턴들만 고려하고 있기 때문에 거의 모든 인지요소패턴을 고려하는 MPWKL이나 PWCDI에 비해 동일한 정도의 추정 정확도에 도달하기 위해 더 많은 문항이 필요할 수 있다.

4. 요약 및 논의

이 장에서는 인지요소 위계를 고려한 BN-H 모형 기반의 인지진단 컴퓨터 적응검사 방법인 BN-H-CAT을 제안하고, 실제로 인지요소 간 위계 관계가 존재함을 가정했을 때 인지요소 위계를 고려한 인지진단모형으로 사용함으로써 인지진단 컴퓨터 적응검사의 효율성이 어떻게 변화하는지를 모의실험을 통해 다양한 조건하에서 확인하고자 하였다.

이를 위해 본 연구에서는 인지요소 위계를 모델링한 모형이 사용된 경우와 인지요소 위계를 모델링하지 않은 모형이 사용된 경우의 인지진단 컴퓨터 적응검사의 효율성을 고정-길이 검사의 추정 정확도 측면과 변동-길이 검사의 검사 길이 측면에서 각각 비교하였다. 그 결과 인지요소 위계를 모델링함으로써, 고정-길이 검사 실험에서는 특히 검사 길이가 짧은 검사 초기에 더 높은 추정 정확도를 얻을 수 있고, 변동-길이 검사 실험에서는 모든 조건 하에서 검사 종료 기준에 도달하기까지 필요한 문항의 수를 상당히 줄이면서 더 높은 추정 정확도를 얻을 수 있음이 확인되었다. 또한, 한 가지 주목할 만한 결과는 고정-길이 검사에서 인지요소 위계를 모델링했을 때 최하위 인지요소(A_1)의 추정 정확도가 나머지 상위 인지요소들의 추정 정확도보다 높게 나왔다는 점이었다. 이는 인지요소 위계 구조에서 최하위 인지요소는 나머지 상위 인지요소들의 선행 인지요소이기 때문에, 상위 인지요소들을 측정하는 문항에 대한 피험자의 응답이 하위 인지요소의 숙달여부에 대해서도 정보를 제공하기 때문이었을 것으로 판단된다.

이 장의 모의실험은 인지요소들이 이루는 진 위계 구조를 알고 있는 것으로 가정하고 진행되었다. 그러나 실제로는 학생들의 학습 과정에 내재된 정확한 위계 관계를 밝히는 것이 선행되어야 하며, 이는 교육과정의 내용 지식과 학습에 대한 인지이론 등에 대한 철저한 조사 및 검토와 더불어 경험 연구를 통한 검증의 과정을 필요로 하는 다소 복잡하고 어려운 작업일 수 있다. 게다가 인지요소들 간의 관계가 본 연구에서 가정

한 강한 위계가 아니라 약한 위계를 나타낼 가능성도 있다. 즉, 선행 인지요소를 숙달하면 후행 인지요소를 숙달할 가능성이 높아지지만 선행 인지요소를 완전히 숙달하지 못했더라도 후행 인지요소의 숙달이 가능한 경우가 있을 수 있다. 이와 같은 약한 위계 관계가 만약 강한 위계 관계로 모델링되는 경우에는 과소추정(underestimation) 문제로 인해 추정의 정확도가 낮아질 수 있다. 실제로 최근 연구들에서 인지요소 위계의 오지정(misspecification)이 진단의 정확도에 영향을 미친다는 점이 확인되기도 하였다(Templin et al., 2008; Liu, 2018). 이러한 점에 비추어 볼 때 추후 연구에서는 강한 위계와 약한 위계가 모두 존재하는 다양한 위계 구조를 가정하고 인지요소 위계를 오지정하였을 때 BN-H-CAT의 효율성이 어떻게 변하는지에 대해서도 경험적인 연구가 행해져야 할 것이다.

본 연구는 인지진단 컴퓨터 적응검사를 위해 그동안 개발되어온 문항 선택 방법들을 활용하여 모의실험을 진행하였으며, 적응검사의 실제적인 실행과 관련되는 문항 노출 통제(item exposure control)나 내용 제약(content constraints) 등을 고려한 문항 선택 방법들을 다루지 않았다는 점은 제한점이라고 볼 수 있다. 향후 연구에서는 보다 다양한 문항 선택 방법들을 이용한 확장된 모의실험을 통해 인지요소 위계가 존재하는 경우의 적응검사에 대한 보다 풍부한 논의를 이끌어야 할 것이다. 또한 본 연구에서 베이지안 네트워크를 활용한 인지진단 컴퓨터 적응검사인 BN-H-CAT을 제안한 만큼 베이지안 네트워크 고유의 특징을 반영한 새로운 문항 선택 방법을 개발하는 연구도 필요할 것으로 보이며, 인지요소 위계의 특징을 다른 방식으로 반영한 새로운 문항 선택 방법의 개발도 흥미로운 연구 주제가 될 수 있을 것이다.

Ⅵ. 요약 및 결론

1. 요약

본 연구는 베이지안 네트워크라는 프레임워크 안에서 인지요소 위계를 고려한 인지진단평가에 대한 일련의 방법론을 개발하여 구체화하고자 하였다. 이를 위해 Ⅱ장에서는 먼저 방법론 연구의 토대가 되는 관련 개념들을 검토하였다. Ⅱ장에서 수행한 주요 내용은 다음과 같다. 첫째, 인지진단이론과 관련된 여러 개념 및 원리를 검토하였다. 구체적으로, Q-행렬, 인지요소 등의 개념을 확인한 뒤, 인지진단모형을 잠재집단모형의 제약 모형으로 보는 관점을 취하여 구조모형과 측정 모형이라는 두 가지 틀을 이용하여 여러 인지진단모형의 특징을 파악하고자 하였다. 또한, 인지요소 위계의 정의를 살펴보고 인지요소 위계를 고려한 선행 인지진단모형들을 검토하였으며, 인지진단평가에서 인지요소패턴을 추정하는 대표적인 방법들의 원리를 검토하였다 둘째, 베이지안 네트워크의 수학적 정의를 검토하고, 베이지안 네트워크의 활용을 위해 반드시 이해해야 하는 과정이라고 할 수 있는 추론과 학습의 원리를 살펴보았다. 추론과 학습의 방법은 이후 BN-H 모형을 이용한 인지진단평가에서 모형의 모수 추정과 인지요소패턴의 추정을 위해 직접적으로 활용된다. 셋째, IRT 모형에 기반한 컴퓨터 적응검사의 기본 원리와 절차를 검토하였다.

Ⅲ장에서는 본 연구의 방법론을 관통하는 핵심 모형인 BN-H 모형을 엄밀하게 정의하는 것을 목적으로 하였다. 이를 위해 먼저 일반적인 베이지안 네트워크 기반 인지진단모형인 BN 모형을 정의하고 예시를 통해 BN 모형의 특징을 확인하였다. 그리고 BN 모형에 대해 위계 관계에 대한 가정을 추가한 모형으로서 BN-H 모형을 도입하였으며 수학적으로 엄밀하게 정의하고자 하였다. BN-H 모형에 대해서도 예시를 제시하여 그 특징을 구체적으로 확인하였으며, 앞서 Ⅱ장에서 살펴본 기존의 인지

요소 위계를 고려한 인지진단모형인 DINA-H 모형, DINO-H 모형, HDCM과 BN-H 모형 사이에는 어떤 차이가 있는지 그 관계를 구조모형과 측정모형으로 나누어 살펴보았다. 마지막으로 BN-H 모형을 적용한 인지진단평가의 원리와 절차를 제시하였다.

Ⅳ장에서는 인지요소 위계를 가정한 인지진단모형을 활용함에 있어 고려해야 하는 위계 타당화의 과정인 위계 추정과 관련된 문제를 다루었다. 이를 위해 먼저 베이지안 네트워크의 구조 학습 방법인 SEM 알고리즘을 검토하고, 이를 활용하여 위계 추정을 실시한 선행 연구의 방법을 검토하여 한계점을 확인하였다. 다음으로, SEM 알고리즘을 이용하여 자료로부터 위계를 추정하는 새로운 위계 추정 방법인 SEM-H의 알고리즘을 제시하고 그 절차를 자세히 서술하였다. 모의실험에서는 기존의 위계 추정 방법과의 비교를 통해 SEM-H 방법의 위계 추정 성능을 검증하고, 특정 조건 하에서 일정 수준 이상의 위계 추정 정확도를 얻기 위해 어느 정도의 문항과 어느 정도의 표본 크기가 필요한지를 확인함으로써 향후 SEM-H 방법을 실제로 활용하고자 하는 목적으로 검사를 설계함에 있어 가이드라인이 될 만한 시사점들을 도출하였다. 또한, SEM-H 방법의 실제적인 활용 가능성을 경험적으로 검토하기 위해 두 가지 실제 수학 검사 자료에 SEM-H 방법을 적용하여 위계 추정을 실시하고 그 결과를 분석하였다. 이로부터 SEM-H 방법이 실제 수학 검사 자료에서 인지요소들 간의 위계 관계를 포함한 종속 관계를 추정해주는 유용한 방법이 될 수 있음을 확인하였다.

Ⅴ장에서는 인지진단 컴퓨터 적응검사에 BN-H 모형을 적용한 BN-H-CAT을 제안하였다. 이를 위해 먼저 인지진단 컴퓨터 적응검사의 문항선택방법과 검사종료기준을 검토하였다. 다음으로 BN-H-CAT의 구체적인 절차를 설계와 실행의 두 측면으로 나누어 살펴보았다. 모의실험에서는 실제로 인지요소 위계가 존재하는 경우에 BN-H 모형을 통해 이를 직접적으로 모델링함으로써 인지진단 컴퓨터 적응검사의 검사 길이와 인지요소 추정 정확도의 두 측면에서 효율성을 높일 수 있다는 점을 확인하였다.

2. 결론 및 제언

위계를 고려한 인지진단평가에 대한 베이지안 네트워크 기반의 방법론을 구체화한 본 연구의 결과로부터 도출된 결론은 다음과 같다.

첫째, 인지요소 위계를 고려한 베이지안 네트워크 기반 인지진단모형인 BN-H 모형은 일반적인 베이지안 네트워크 기반 인지진단모형인 BN 모형에 인지요소 위계에 대한 가정을 추가한 모형으로, 인지요소 위계 구조는 베이지안 네트워크의 그래프 구조 DAG로 표현되며 위계 관계에 대한 가정에 따라 어떤 인지요소의 선행 인지요소를 미숙달하였을 때 그 인지요소를 숙달할 확률에 대한 모수를 0으로 제약하는 모형으로 정의된다. 이는 Templin과 Bradshaw(2014)가 LCDM에 대한 제약 모형으로 HDCM을 정의한 것과 유사한 특징이라고 볼 수 있다. 본 연구에서는 선행 연구에서 정의되었던 인지요소 위계의 개념을 BN-H 모형을 통해 그대로 모델링하는 것을 목표로 하였기 때문에, BN-H 모형은 DAG 구조의 모든 유향 변들이 오직 위계 관계를 나타낸다는 가정이 있다. 하지만 베이지안 네트워크는 위계 관계뿐 아니라 종속 관계도 얼마든지 나타낼 수 있는 모형이므로, BN-H 모형의 구조에 대한 정의에 '종속 관계를 유향 변으로 설정한다'는 기존 BN 모형의 정의를 추가함으로써, BN-H 모형은 위계 관계뿐만 아니라 위계 관계가 아닌 종속 관계도 나타내는 모형의 역할도 할 수 있을 것으로 기대된다.

둘째, 베이지안 네트워크 구조의 국지 독립성 가정으로 인해 BN-H 모형의 구조모형을 이루는 인지요소패턴의 확률은 각 인지요소의 숙달에 대한 조건부 확률의 곱으로 표현되며, 이에 따라 BN-H 모형은 구조모수의 수가 인지요소의 수와 동일하다는 특징을 드러낸다. 따라서 BN-H 모형의 구조모형은 기존의 DINA-H 모형, DINO-H 모형 및 HCDM의 구조모형과 비교하여 더 간명한 모형이라고 할 수 있다. 그리고 인지요소의 수가 많아질수록 구조모수의 수의 차이가 더 커지기 때문에 BN-H 모형은 특히 인지요소의 수가 많은 경우에 다른 모형들에 비해 인지요소 위계에 대한 보다 간결한 모델링을 가능하게 해주는 유용한 모형이 될

수 있을 것으로 보인다. 또한, BN-H 모형은 베이지안 네트워크의 추론과 학습 등의 과정을 활용하여 문항모수와 인지요소패턴의 추정이 가능하므로, 실제로 인지요소 위계가 존재하는 경우에 이를 모델링하는 인지진단평가의 모형으로 충분히 활용될 수 있다는 점에서 의의가 있다.

셋째, 본 연구에서는 베이지안 네트워크의 구조 학습 방법인 SEM 알고리즘과 BN-H 모형의 개념을 결합한 위계 추정 방법인 SEM-H를 제안하였으며, 모의실험과 실제 수학 검사 자료의 분석을 통해 실제적으로 활용 가능한 위계 추정 방법임을 밝혔다. 위계 관계의 방향을 반대로 찾을 가능성을 열어두는 기존 COMMAND 방법의 문제점을 개선하기 위해, SEM-H 방법은 SEM 알고리즘의 구조 학습 절차의 중간에 위계 관계의 방향에 대한 판단을 내리는 과정을 추가함으로써 위계 관계의 방향에 대한 올바른 추정의 가능성을 높이고자 하였으며, 모의실험을 통해 다양한 조건 하에서 SEM-H 방법이 기존의 COMMAND 방법을 개선하는 위계 추정 방법임이 검증되었다. 이와 더불어 위계 추정 방법으로서 SEM-H의 실제적인 활용성을 검토하기 위해 두 가지 실제 수학 검사 자료에 SEM-H 방법을 적용하여 위계 추정을 실시한 결과, SEM-H 방법은 일차적으로 구조 학습의 과정을 통해 인지요소들 간의 종속 관계로 이루어진 구조를 도출해주며, 구조와 함께 추정되는 모수에 대한 판정을 통해 구조를 이루고 있는 각 관계가 위계 관계인지 위계 관계가 아닌 종속 관계인지를 판별해주는 유용한 방법임이 확인되었다. 이는 SEM-H 방법이 위계 관계뿐 아니라 인지요소들 사이에 존재할 수 있는 일련의 종속 관계들에 대한 추정도 가능하게 해준다는 점을 시사한다. 또한, SEM-H 방법의 구조 학습에 대한 일련의 제약을 설정하는 집합인 블랙리스트와 화이트리스트, 그리고 위계 관계 여부를 판정하는 기준인 $cutline$은 인지요소 관계에 대한 추가적인 가정이나 위계에 대한 관점을 반영하여 얼마든지 수정될 수 있으므로, SEM-H는 다양한 상황에서 인지요소 관계의 탐색을 유연하게 실행할 수 있는 방법이라고 할 수 있다.

넷째, 본 연구에서는 인지요소 위계를 모델링한 인지진단 컴퓨터 적응검사인 BN-H-CAT이 실제 인지요소 위계가 존재하는 경우에 인지진단

컴퓨터 적응검사의 효율성을 향상시켰다는 점을 확인하였다. 이는 인지요소 위계가 제공하는 정보가 개별 인지요소에 대한 추정의 정확도를 높여주기 때문인 것으로 보인다. 물론 인지요소가 실제로 존재한다는 다소 강한 가정을 기반으로 하나, 그에 따라 인지진단 컴퓨터 적응검사의 검사 길이를 줄일 수 있고 인지요소패턴 추정의 정확도를 높일 수 있다는 점은 충분히 의미가 있는 결과이다. 컴퓨터 적응검사는 본디 개별 피험자의 수준에 맞는 문항을 실시간으로 제공함으로써 검사의 효율성을 높이는 것을 목적으로 한다. 따라서 인지요소 위계가 존재하는 경우에는 이를 직접적으로 모델링함으로써 컴퓨터 적응검사의 본 목적의 달성을 한결 더 수월하게 해줄 수 있으며, 이를 위해 인지요소 위계를 나타내는 모형으로 BN-H 모형을 적용한 BN-H-CAT이 이루어질 수 있다. 한편, BN-H-CAT의 실제적인 실행을 위해서는 BN-H-CAT의 절차를 구현한 컴퓨터 기반의 평가 시스템이 구축되어야 하며 충분한 수의 문항이 확보되어야 하며 이는 많은 시간과 비용을 필요로 하는 작업이다. 이와 같은 현실적인 이유로 본 연구에서 BN-H-CAT을 실제로 진행해보지는 못한 점은 제한점으로 남는다.

본 연구는 위계를 고려한 베이지안 네트워크 기반 인지진단모형인 BN-H 모형을 중심으로 한 인지진단평가와 관련한 일련의 방법론들을 구체화하였다. 첫째로, BN-H 모형을 적용한 인지진단평가가 어떻게 이루어질 수 있는지 그 과정을 밝혔으며, 둘째로, 베이지안 네트워크의 구조 학습의 아이디어에 BN-H 모형의 개념을 적용한 위계 추정 방법의 절차도 구체적으로 제시하였다. 이 두 방법론은 수학 인지진단평가를 계획하고 실시함에 있어 직접적으로 활용될 수 있을 것이라 판단된다. 실제로 위계를 고려한 인지진단평가를 실시함에 있어 처음부터 인지요소들이 가지는 위계를 설정하고 검사를 시작하는 방식을 취할 수도 있고, 위계에 대한 가정 없이 검사를 설계하되 자료가 수집된 이후에 위계에 대한 판단을 내리는 방식을 취할 수도 있다. 이 두 경우 모두에 대해 본 연구에서 제안한 위계 추정 방법인 SEM-H 방법이 활용될 수 있으며, 일련의 위계 관계가 도출된 이후에는 이를 BN-H 모형으로 설정하고 그

에 따른 분석을 진행하여 인지진단 결과를 얻을 수 있을 것이다.

한편, 본 연구는 제안한 방법론을 기 수집된 실제 수학 검사 자료에 적용하는 사후 분석(retrofitting)은 진행하였지만 실제 검사의 설계 단계에서부터 결과를 도출하기까지의 전반적인 과정에 직접 적용해보지 못하였다는 제한점이 있다. 따라서 BN-H 모형을 중심으로 한 수학 인지진단평가를 실제로 설계하고 자료를 수집하여 분석하면서 그 과정에서 나타나는 특성이나 피험자들의 반응 등에 관한 연구가 후속 연구로 진행될 필요가 있다. 이와 같은 후속 연구들을 통해 위계를 고려한 베이지안 네트워크 기반 인지진단평가의 특징들이 더욱 면밀히 밝혀지고, 이를 통해 본 연구에서 제안한 방법론이 수학 인지진단평가에서 널리 활용될 수 있기를 기대하는 바이다.

참 고 문 헌

김명연, 강태훈(2016). 보상성 및 비보상성 가정을 평가하기 위한 모형선택지수의 활용가능성 탐색: DINA와 DINO 모형 비교를 중심으로. **교육평가연구, 29**(3), 383-404.

김선희, 김수진, 송미영(2008). 수학 평가 결과의 분석을 위한 인지 진단 이론의 활용. **학교수학, 10**(2), 259-277.

김성은(2015). 인지진단모형 Q행렬 타당화 방법 간의 회복률 비교. **교육평가연구, 28**(3), 731-751.

김성은, 박윤수, 이영선(2012). 인지진단모형을 적용한 잠재집단별 분수뺄셈검사 문제해결 전략의 차이 분석. **교육평가연구, 25**(1), 49-68.

김희경, 김부미(2013). 인지진단모형을 활용한 수학 학업성취 결과 분석. **학교수학, 15**(2), 289-314.

박세진, 이현숙(2017). DINA와 DINO 모형의 문항 모수 및 피험자 모수 추정 정확성에 영향을 주는 요인 탐색. **교육평가연구, 31**(3), 415-444.

박세진, 이현숙(2018). 인지진단이론의 학교 현장 적용 가능성 탐색: 모수 및 비모수적 접근 방법의 비교. **교육평가연구, 31**(1), 225-254.

박지현, 김수진(2015). 수학 인지적 속성에 따른 TIMSS 2011 8학년 성취도 상위국 특성 비교. **수학교육학연구, 25**(3), 303-321.

성태제(2018). 인지진단모형에 의한 중1 수학 '수와 연산'의 형성평가 도구 개발과 분석. **교육평가연구, 31**(1), 29-53.

윤지영(2015). **요소 간 위계 방식과 인공신경망을 적용한 수학 인지진단 평가 연구.** 서울대학교 대학원 석사학위논문.

이연숙(2013). 인지 진단 평가 모델 비교 연구: 언어시험 데이터 적용의 경우. **교육평가연구, 26**(3), 623-646.

이현숙, 고호경(2014). 인지진단모형을 적용한 TIMSS 8학년 수학 기하

영역의 성차 분석. **학교수학, 16**(2), 387-407.

이혜옥(2005). **적응검사**. 서울: 교육과학사.

한국교육평가학회(2004). **교육평가 용어사전**. 서울: 학지사.

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.

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement, 23*(3), 223-237.

Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *Bayesian networks in educational assessment*. New York, NY: Springer.

Baroody, A. J., Cibulskis, M., Lai, M. L., & Li, X. (2004). Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning, 6*(2), 227-260.

Bertelé, U. & Brioschi, F. (1972). *Nonserial dynamic programming*. New York, NY: Academic Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice, 5*(1), 7-74.

Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement, 20*(3), 213-229.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association, 110*(510), 850-866.

Chen, Y., González-Brenes, J. P., & Tian, J. (2016). Joint discovery of skill prerequisite graphs and student models. In T. Barnes, M. Chi

& M. Feng (Eds.), *Proceedings of the ninth international conference on educational data mining* (pp. 46-53). Raleigh, NC: International Educational Data Mining Society.

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika, 74*(4), 619-632.

Cheng, Y. (2010). Improving cognitive diagnostic computerized adaptive testing by balancing attribute coverage: The modified maximum global discrimination index method. *Educational and Psychological Measurement, 70*(6), 902-913.

Chiu, C. Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement, 37*(8), 598-618.

Choi, K. M., Lee, Y. S., & Park, Y. S. (2015). What CDM can tell about what students have learned: An analysis of TIMSS eighth grade mathematics. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(6), 1563-1577.

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.

Culbertson, M. J. (2016). Bayesian networks in educational assessment: The state of the field. *Applied Psychological Measurement*, *40*(1), 3-21.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-Matrix. *Applied Psychological Measurement, 35*(1), 8-26.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343-362.

de la Torre, J. (2009). DINA model and parameter estimation: A

didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115-130.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179-199.

de la Torre, J., & Chiu, C. Y. (2016). General method of empirical Q-matrix validation. *Psychometrika, 81*(2), 253-273.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333-353.

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. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Erlbaum.

Friedman, N. (1997). Learning belief networks in the presence of missing values and hidden variables. In D. Fisher (Ed.), *Fourteenth international conference on machine learning* (pp. 125-133). San Francisco, CA: Morgan Kaufmann.

Gágne, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design*. Belmont, CA: Wadsworth/ Thomson Learning.

Gierl, M. J., Cui, Y., & Hunka, S. (2008). Using connectionist models to evaluate examinees' response patterns to achievement tests. *Journal of Modern Applied Statistical Methods, 7*(1), 234-245.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational*

Measurement, 26(4), 301-321.

Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practice*. Doctoral dissertation, University of Illinois at Urbana Champaign.

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning, 20*(3), 197-243.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement, 29*(4), 262-277.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191.

Højsgaard, S. (2012). Graphical independence networks with the gRain package for R. *Journal of Statistical Software, 46*(10), 1-26.

Hsu, C. L., Wang, W. C., & Chen, S. Y. (2013). Variable-length computerized adaptive testing based on cognitive diagnosis models. *Applied Psychological Measurement, 37*(7), 563-582.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation, 15*(3), 1-7.

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement, 71*(2), 407-419.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10*(1), 128-141.

Jensen, F. V., & Nielsen, T. D. (2009). *Bayesian networks and decision graphs*. New York, NY: Springer.

Jensen, F. V., Olesen, K. G., & Andersen, S. K. (1990). An algebra of Bayesian belief universes for knowledge-based systems. *Networks 20*(5), 637-659.

Junker, B., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258-272.

Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*(3), 167-188.

Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait theory and computerized adaptive testing* (pp. 237-254). New York, NY: Academic Press.

Klein, M. F., Birnbaum, M., Standiford, S. N., & Tatsuoka, K. K. (1981). *Logical error analysis and construction of tests to diagnose student "bugs" in addition and subtraction of fractions* (Research Report No. 81-6). Urbana, IL: University of Illinois, Computer-Based Education Research Laboratory.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

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.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on

Tatsuoka's rule space approach. *Journal of Educational Measurement*, *41*(3), 205-237.

Leighton, J. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment? In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3-18). Cambridge, UK: Cambridge University Press.

Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling*. Boca Raton, FL: Chapman and Hall/CRC.

Linn, R. L., Rock, D. A., & Cleary, T. A. (1972). Sequential testing for dichotomous decisions. *Educational and Psychological Measurement, 32*(1), 85-95.

Liu, R. (2018). Misspecification of attribute structure in diagnostic measurement. *Educational and Psychological Measurement. Educational and Psychological Measurement, 78*(4), 605-634.

Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2017). The impact of Q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. *Educational and Psychological Measurement, 77*(2), 220-240.

Madison, M. J., & Bradshaw, L. P. (2015). The effects of Q-matrix design on classification accuracy in the log-linear cognitive diagnosis model. *Educational and Psychological Measurement, 75*(3), 491-511.

Mao, X., & Xin, T. (2013). The application of the Monte Carlo approach to cognitive diagnostic computerized adaptive testing with content constraints. *Applied Psychological Measurement, 37*(6), 482-496.

McMillan, J. H. (2008). *Assessment essentials for standards-based education.* Thousand Oaks, CA: Corwin Press.

Millán, E., Pérez-de-la-Cruz, J. L., & Suárez, E. (2000). Adaptive Bayesian networks for multilevel student modelling. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent tutoring systems: 5th international conference* (pp. 534-543). New York, NY: Springer.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*(4), 439-483.

Mislevy, R. J. (1995). Probability-based inference in cognitive diagnosis. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 43-71). Hillsdale, NJ: Erlbaum.

Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (1999). Bayes nets in educational assessment: Where do the numbers come from. In K. B. Laskey & H. Prade (Eds.), *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 437 - 446). San Francisco, CA: Morgan Kaufmann.

Nadkarni, S., & Shenoy, P. P. (2001). A Bayesian network approach to making inferences in causal maps. *European Journal of Operational Research, 128*, 479-498.

Nagarajan, R., Scutari, M, & Lèbre, S. (2013). *Bayesian networks in R: With applications in systems biology*. New York, NY: Springer.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufmann.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reckase, M. D. (1983). A procedure for decision making using tailored testing. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait theory and computerized adaptive testing* (pp. 237-254). New York: Academic Press

Rényi, A. (1961). On measures of entropy and information. In M. Eden (Ed.), *Proceeding of the fourth berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics* (pp. 547-561). The Regents of the University of California.

Robitzsch, A., Kiefer, T., George, A. C., & Ünlü, A. (2018). *CDM: Cognitive diagnosis modeling*. R Package version, 6.1-10. Retrieved from https://cran.r-project.org/package=CDM

Rupp, A. A., & Templin, J. (2008a). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68*(1), 78-96.

Rupp, A. A., & Templin, J. L. (2008b). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement, 6*, 219-262.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics,* 6, 461-464.

Scutari, M. (2010). Learning Bayesian networks with the bnlearn R

package. *Journal of Statistical Software, 35(3),* 1-22.

Shafer, G., & Shenoy, P. (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence 2,* 327-352.

Su, Y. L. (2013). *Cognitive diagnostic analysis using hierarchically structured skills*. Unpublished doctoral dissertation. University of Iowa.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *51*(3), 337-350.

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 143-157.

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. (1984). *Analysis of errors in fraction addition and subtraction problems* (Report No. NIE-G-81-0002). Urbana: Computer-based Education Research Laboratory, University of Illinois.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics, 10*(1), 55-73.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (1993). Item construction and psychometric models appropriate for constructed responses. In R. E. Bennett & W. C.

Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 107-133). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327-359). Hillsdale, NJ: Erlbaum.

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, *79*(2), 317-339.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods, 11*(3), 287.

Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement, 32*(7), 559-574.

Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer(Ed.), *Computerized adaptive testing: A primer* (pp. 101-133). Mahwah, NJ: Erlbaum.

Thompson, N. A. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment Research & Evaluation, 12*(1), 1-13.

Wainer, H. (2000). Introduction and history. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (pp. 1-21). Mahwah, NJ: Erlbaum.

Wainer, H., & Mislevy, R. J. (2000). Item response theory, item calibration, and proficiency estimation. In H. Wainer (Ed.),

Computerized adaptive testing: A primer (pp. 61-100). Mahwah, NJ: Erlbaum.

Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Educational and Psychological Measurement*, *73*(6), 1017-1035.

Wang, C., & Chang, H. H. (2018). Cognitive diagnostic assessment. In W. J. van der Linden (Ed.), *Handbook of item response theory, volume 3: Applications* (pp. 329-347). Boca Raton, FL: CRC Press.

Weiss, D., & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*(4), 361-374.

Xu, X., Chang, H., & Douglas, J. (2003, April). *A simulation study to compare CAT strategies for cognitive diagnosis*. Paper presented at the annual meeting of National Council on Measurement in Education, Chicago, IL.

Zheng, C., & Chang, H. H. (2016). High-efficiency response distribution - based item selection algorithms for short-length cognitive diagnostic computerized adaptive testing. *Applied Psychological Measurement*, *40*, 608-624.

Zheng, C., & Wang, C. (2017). Application of binary searching for item exposure control in cognitive diagnostic computerized adaptive testing. *Applied Psychological Measurement, 41*(7), 561-576.

부록: 일차방정식과 일차부등식 검사

윤지영(2015)에서 저자의 동의하에 일부 발췌함.

1. $\frac{x-1}{2}+\frac{3x}{4}$ 를 간단히 나타낸 것은?

① $\frac{4x-1}{6}$

② $4x-\frac{1}{6}$

③ $4x-\frac{1}{2}$

④ $\frac{5x-1}{4}$

⑤ $\frac{5x-2}{4}$

2. $\frac{3a-5b}{2}-\frac{4a-6b}{3}$을 간단히 나타낸 것은?

① $a-b$

② $\frac{-a+b}{5}$

③ $\frac{-a+b}{6}$

④ $\frac{7a-11b}{5}$

⑤ $\frac{a-3b}{6}$

3. 다음 중 등호 '='를 바르게 사용한 것은?

① $3=5$

② $2+2=2$

③ $0-1=1$

④ $-3+2=0+1$

⑤ $(-1)+2=5-4$

4. 다음 중 등호 '='를 사용하여 나타낼 수 있는 것은?

① x의 2배에 3를 뺀다.

② x에서 1을 더하면 양수이다.

③ x에 5를 더하면 2와 같지 않다.

④ x는 y보다 1만큼 작다.

⑤ x의 2배는 y보다 크지 않다.

5. 일차방정식 $3(2x-1)+2x=21$을 만족하는 x의 값은?

① $-\frac{11}{4}$

② $\frac{11}{4}$

③ -3

④ 3

⑤ $\frac{9}{4}$

6. 길이가 40cm인 나무 막대를 세 조각으로 잘랐더니 각각의 길이(cm)가 다음과 같다.

$$2x-5$$
$$x+7$$
$$x+6$$

가장 긴 막대의 길이는 얼마인가?

① 8 cm ② 14 cm

③ 15 cm ④ 19 cm

⑤ 20 cm

7 삼 남매의 나이가 각각 세 살씩 차이가 나고, 맏이의 나이는 막내의 나이의 2배보다 세 살 적다고 한다. 맏이의 나이를 구하여라.

① 11살 ② 13살
③ 15살 ④ 17살
⑤ 19살

8. 다음 중 두 수의 대소 관계를 바르게 나타낸 것은?

① $5 < 3$
② $1 < -2$
③ $-3 > -1$
④ $1+3 < 4$
⑤ $2-3 \leq 0$

9. 다음 중 'a는 3보다 크지 않다'를 부등호를 사용하여 바르게 나타낸 것은?

① $a < 3$
② $a > 3$
③ $a \leq 3$
④ $a \geq 3$
⑤ $a \neq 3$

10. 일차부등식 $3x-2 < 1$을 풀면?

① $x < 4$
② $x < 1$
③ $x > 1$
④ $x < 0$
⑤ $x > 0$

11. 부등식 $0.2x+1 < \frac{1}{5}(2x-1)$을 풀면?

① $x = 6$
② $x > 6$
③ $x \geq 6$
④ $x \leq -6$
⑤ $x > -6$

12. 사진을 인화하는데 기본 10장은 5000원이고, 한 장 더 추가 시 250원을 받는다고 한다. 최소 몇 장 이상 뽑아야 1장의 가격이 300원 이하가 되는가?

① 40장
② 50장
③ 60장
④ 70장
⑤ 80장

13. 현지네 반 학생들이 미술관에 갔더니 1인당 입장료가 7000원이고 20명 이상인 경우 10% 할인, 30명 이상인 경우 20% 할인 된 가격의 단체권을 구입하여 입장할 수 있다고 한다. 20명 이상 30명 미만의 단체는 최소 몇 명 이상일 때 30명의 단체권을 구입하여 입장하는 것이 유리한지 구하면?

① 25명 이상
② 26명 이상
③ 27명 이상
④ 28명 이상
⑤ 29명 이상

[14~15] 태수는 집 앞에 있는 DVD 대여점에서 종종 DVD를 빌려본다. 이 가게의 연회비는 1만원이다. 다음 표와 같이 회원의 DVD 대여요금은 비회원보다 싸다.

비회원의 DVD 1장당 대여요금	회원의 DVD 1장당 대여요금
3200원	2500원

14. 태수는 작년에 이 DVD 대여점의 회원이었다. 태수는 작년에 연회비를 포함하여 총 52500원을 썼다. 만약 태수가 회원 가입을 하지 않았다면, 같은 수의 DVD를 대여하는 데 쓴 비용은 얼마였겠는가?

① 64000원
② 54400원
③ 49300원
④ 42500원
⑤ 33200원

15. 연회비를 낸 것이 손해가 되지 않으려면 DVD를 최소 몇 장 이상 대여해야 하는가?

① 9장 이상
② 11장 이상
③ 13장 이상
④ 15장 이상
⑤ 17장 이상

Abstract

Cognitive Diagnosis Model with Attribute Hierarchy based on Bayesian Network

Soo-Yun Han

Department of Mathematics Education

The Graduate School

Seoul National University

Cognitive diagnostic assessment (CDA) aims to determine the students' mastery status for cognitive attributes representing specific knowledge, skills or ability necessary for learning in a particular domain. CDA has received considerable attention because it provides detailed diagnostic feedback on the cognitive strengths and weaknesses of students, which can be valuable information for classroom instruction and learning. The statistical models constructed to perform CDA are called cognitive diagnosis model (CDM). Several different CDMs have been proposed under various assumptions about the

relationship among cognitive attributes and/or between cognitive attributes and item responses. Some models among them assume ‘attribute hierarchy’ which implies hierarchical relationships between cognitive attributes. The CDMs with attribute hierarchy have potentials to be used for CDA reflecting the systematic nature of mathematics.

The Bayesian network, a probabilistic graphical model, can be used for CDM with attribute hierarchy in that attribute hierarchy is commonly represented as a graph structure. However, previous studies considering Bayesian network for CDMs did not focus on modeling attribute hierarchy in a strict sense. Thus, this study aims to define CDM with attribute hierarchy based on Bayesian network, called BN-H model, and to devise the CDA methodology considering attribute hierarchy centered on BN-H model. Specifically, a method of estimating attribute hierarchy, which should be addressed when CDM with attribute hierarchy is used for assessment, and a method of cognitive diagnostic computerized adaptive testing (CD-CAT) with attribute hierarchy are proposed.

In this study, BN-H model was defined as ‘the model that express an attribute hierarchy using the graph structure of Bayesian network and constrain certain parameters of the general Bayesian network model to zero according to the attribute hierarchy.’ BN-H model assumes more parsimonious structural model and more general measurement model than the existing CDMs with attribute hierarchy.

As a method of hierarchy estimation based on the BN-H model, SEM-H method was suggested to improve the existing hierarchy estimation method based on the structure learning algorithm of Bayesian network. In the first simulation experiment, SEM-H method performed better in estimating correct attribute hierarchies compared to the existing method. In the second simulation experiment, performance of SEM-H method for estimating attribute hierarchy was evaluated under various simulation conditions. In addition, the practical applicability of SEM-H method was examined by applying SEM-H method to two actual math test data sets.

Next, CD-CAT based on the BN-H model, which is called BN-H-CAT, was introduced and the procedures of BN-H-CAT were presented. BN-H-CAT is distinguished from the conventional CD-CAT in that it considers attribute hierarchy and uses CDM based on Bayesian network. The results of simulation experiments comparing BN-H-CAT with the conventional CD-CAT without assuming attribute hierarchy show that BN-H-CAT can improve the efficiency of CD-CAT by modeling attribute hierarchy via BN-H model in the presence of attribute hierarchy.

Finally, the implications of this study were discussed and the directions for future research based on the CDA methodology proposed in this study were suggested.

Keywords: **cognitive diagnostic assessment, cognitive diagnosis model, attribute hierarchy, Bayesian network, hierarchy estimation, cognitive diagnostic computerized adaptive testing**

***Student Number*: 2011-21567**