한국통계학회 논문집 제 1권 1호
1994년 12월 pp. 192-205

문항반응이론에서의 추정방법과 대입학력고사의 문항분석

박정수, 조완현[1)]

요약

본 논문에서는 피험자의 능력과 검사문항에 정답할 확률과의 관계에 기초한 문항반응이론의 기본 가정과 통계적 모형을 소개하였다. 또한 검사의 목적상 필요한 피험자의 능력을 정확히 추정하는 방법과, 검사에 사용되는 각 문항을 특성지우는 문항모수의 여러가지 통계적 추정 방법에 대하여 정리하였다. 그 방법들은 결합 최우추정법, 조건부 최우추정법, 주변 최우추정법, 베이지안 추정법 및 이들의 혼합에 의한 방법이다. 문항반응이론의 적용의 한 예로서 93년도 대입학력고사의 수학 시험문항을 BILOG 라는 컴퓨터 프로그램을 이용하여 분석하였다.

1. 서론

한 피험자의 어떤 능력을 시험문제들로 이루어진 검사를 통하여 정확히 측정한다는 것은 어려운 일이지만 매우 중요한 일이다. 일반적으로 한 피험자의 능력은 그 피험자가 검사 내의 여러 문항들 중에서 맞춘 문항들의 수로서 측정되지만, 이는 문항 난이도 (item difficulty) 나 변별도 (discriminating power) 와 같은 문항 특성치를 고려하지 않은 방법이다. 즉, 어려운 (또는 변별도가 큰) 문항은 쉬운 (변별도가 작은) 문항에 비하여 가중치를 더 많이 주고 쉬운 (변별도가 작은) 문항은 상대적으로 가중치를 적게 주어서 그 가중합으로 한 피험자의 능력을 측정하여야 바람직하다. 그런데 한 문항이 쉬운 문항인지 어려운 문항인지, 즉 문항 난이도를 어떻게 측정할 것인가 하는 문제가 생긴다. 이에 대해 고전검사이론 (classical test theory) 에서는 피험자 총수 중에 그 문항을 못 맞춘 피험자의 수로서 그 문항의 난이도를 정의한다. 따라서 이 난이도는 피험자 집단의 특성에 따라 변하게 된다. 즉, 문항은 변함이 없는데도, 피험자 집단이 능력이 많은 집단이면 난이도는 낮게 측정되고 피험자 집단의 능력이 적은 집단이면 그 문항의 난이도는 높게 측정된다. 이는 문항의 난이도는 시험을 치르는 피험자 집단의 능력분포에 상관없이 그 문항을 특성지우는 하나의 고유적인 문항 특성 모수 (item characteristic parameter) 이여야 한다는 입장과 대치된다. 이러한 문항 난이도의 불변성과 이들 문항 특성치 들의 적용에 의한 피험자 능력의 측정 등은 본 논문에서 소개하는 문항반응이론 (Item Response Theory, IRT) 의 중요한 개념이다.

문항반응이론은 피험자의 능력수준과 각 문항에 대한 반응의 관계를 통계적 모형으로 기술하는 데서 부터 출발한다. 이는 몇가지 실제 자료에서 만족되기 어려운 가정을 하는 반면, 그 가정이 만족되기만 하면 검사의 제작과 분석 및 피험자의 능력 추정에 있어서 고전검사이론보다 더 적절하고 효율적이라고 한다. 예를들어, 고전검사이론에서 측정의 오차는 모든 피험자에 대해서 동일하다고 하는 받아들이기 어려운 가정을 문항반응이론에서는 정보함수의 개념을 도입하여 능력수준에 따라 상이하게 추정함으로서 해결하고 있다.

본 논문에서는 피험자의 능력과 문항 특성치 등에 관하여 통계적 모형을 도입한 문항반응이론

1) (500-757) 광주시 북구 용봉동 전남대학교 통계학과

에 바탕하여 피험자의 능력과 문항특성치를 추정하는 과정을 소개하고, 이를 대입학력고사 시험문항의 통계적 분석에 적용하였다. 본 논문의 제 2 장에서는 문항반응모형과 기본적 가정들을 소개하고, 제 3 장에서는 문항모수 및 피험자 능력의 통계적 추정 방법들을 기술한다. 제 4 장에서는 문항반응이론의 적용의 한 예로서 93년도 대입학력고사의 수학 시험문항을 분석하였다. 끝으로 제 5 장에서는 요약 및 문제점과 미래의 연구과제 등이 논의 되었다.

문항반응이론은 교육평가 및 측정분야에 통계학을 잘 적용한 경우로서 국내외의 교육평가학자들에 의해서 최근 활발히 연구되고 있다. 문항반응이론에 대한 대표적 참고서적으로는 국내에서는 성태제 (1991), 이종성 (1990)이 있고, 국외에서는 Lord (1980), Hambleton & Swaminathan (1985), Hulin, Drasgow & Parsons (1983) 과 Baker (1992) 등을 들 수 있다. 본 논문에서 요약된 형태로 기술되는 주요 내용은 이들 책에서 그 상세한 설명을 볼 수 있다.

2. 문항반응이론의 기본가정과 문항반응함수

2.1 기본 모형과 가정

대부분의 검사는 피험자의 능력수준을 정확히 그리고 신뢰성 높게 추정하는 일이 주요 과제이다. 피험자의 능력을 잘 추정하려면, 피험자의 능력이나 기능이 문항에 어떠한 반응을 하는지, 그 관계를 규명하여야 할 것이다. 따라서 문항반응이론은 피험자의 능력수준과 각 문항에 대한 반응의 관계를 통계적 모형으로 기술하는데서 부터 출발한다. 즉 문항반응이론의 중심은 어떤 피험자가 어떤 문항에 대해 어떤 반응을 할 것인가를 확률함수로 표현하는 데 있다. 이 때의 확률함수를 문항반응함수 (Item Response Function, IRF) 라고 부르며,

이는 주어진 능력 $\underline{\theta}$ 를 가진 피가 j 번째 문항을 맞출 확률로서 나타내 진다. 즉

$$P_j(\underline{\theta}) = P_r\,(U_j=1 \mid \underline{\boldsymbol{\theta}}=\underline{\theta}) \tag{2.1}$$

여기서 확률벡터 $\underline{\boldsymbol{\theta}}$ 는 임의로 선택된 피험자의 직접 관찰 가능하지 않은 잠재능력 (latent ability)을 나타내고, 확률변수 U_j 는 j 번째 문항에 대한 응답으로서 그 값이 1 이면 정답한 경우이고 0 이면 오답한 경우를 나타낸다. 이때 $\underline{\theta}$ 는 확률벡터 $\underline{\boldsymbol{\theta}}$ 가 갖는 특정값을 표시하고 u_j 는 U_j 의 실제 관측치를 나타낸다. 하나의 검사에 J 개의 문항이 있고 N 명의 피험자가 그 검사를 치뤘다면 그 검사는 $N \times J$ 의 0 과 1 로 이루어진 문항반응 행렬을 형성한다.

위의 문항반응함수 (2.1)을 이용한 문항반응이론의 전개를 위하여 일반적으로 세가지의 기본가정을 세우는데, 이는 조건부 독립성, 단조성과 일차원성이다.

첫째는 조건부 독립성 (conditional independence) 또는 국소적 독립성 (local independence) 으로, 이는 모든 $\underline{U}$ 와 $\underline{\boldsymbol{\theta}}$ 에 대하여

$$P(\underline{U}=\underline{u} \mid \underline{\boldsymbol{\theta}}=\underline{\theta}) = \prod_{i=1}^{N} P(U_i=u_i \mid \underline{\theta}) = \prod_{i=1}^{N} P_i(\underline{\theta})^{u_i}\,(1-P_i(\underline{\theta}))^{1-u_i} \tag{2.2}$$

이 만족된다는 것이다. 여기서 확률벡터 $\underline{U}=(U_1, U_2, \ldots, U_N)$ 이고 $\underline{u}=(u_1, u_2, \ldots, u_N)$ 이다. 이 가정은 어떤 특정 능력을 가진 피험자에게 있어서 각 문항들은 서로 독립적이다는 것이다. 다시말해 어떤 특정 능력의 피험자에게 있어서 어떤 한 문항을 맞추고 못 맞추고가 다른 문항을

맞추고 못 맞추고에 아무런 영향을 미치지 않게끔 검사가 작성되었다는 가정이다. 또한 문항의 배치 순서가 각 피험자의 각각의 문항을 맞출 확률에 영향을 미치지 않는다는 가정으로,바람직한 검사의 한 요건이기도 하다.

두번째 가정으로 단조성 (monotonocity) 인데, 이는 각 문항의 문항반응함수 $P_j(\underline{\theta})$ 가 $\underline{\theta}$ 에 대해 단조증가 (monotone nondecreasing) 하는 함수라는 가정이다. 다시말해 $\underline{\theta}_1 < \underline{\theta}_2$ 이면 모든 문항 U_j 에 대해서

$$P(U_j=1 \mid \underline{\theta}_1) \le P(U_j=1 \mid \underline{\theta}_2)$$

이 만족된다는 가정이다. 이는 능력이 많을수록 문항을 맞출 확률이 증가한다는 가정으로서, 잘 만들어진 문항은 이 가정을 당연히 만족시킬 것이다. 현재 사용되고 있는 로지스틱 (logistic) 모형 또는 정규 오자이브 (Normal ogive) 모형등 대부분의 문항반응함수는 이 단조성을 만족하는 함수들 중의 일부이다.

세번째 가정은 일차원성 (unidimensionality) 인데, 이는 식 (2.1) 과 (2.2) 에서 $\underline{\theta}$ 가 1 차원의 확률변수라는 가정이다. 이를 좀더 자세히 잘 이해하기 위해서 문항반응이론에서 사용하고 있는 차원 (dimension) 의 정의부터 파악하자.

먼저 인지적인 측면에서의 심리학적인 정의로는, 한 검사의 모든 문항을 올바르게 응답하는데 필요로하는 인간 능력의 수 (the number of human abilities) 를 그 검사의 차원이라 한다. 이는 어떤 검사가 k 개의 잠재능력을 측정하려고 제작되었다면, 피험자들이 그 검사의 모든 문항을 올바르게 응답하는데 그 k 개의 잠재능력 이외의 능력이 필요치 않아야 한다. 여기서 검사의 일차원성이란 한 검사가 단지 하나 만의 잠재능력을 측정하도록 제작되었음을 뜻한다.

한편 통계학적인 정의로는, 최소한 k 개의 잠재능력이 있어야 조건부 독립성을 만족시키는 문항반응함수 (2.1)을 세울 수 있을때 그 검사의 차원은 k 차원이라고 한다 (Hattie, 1985). 이는 k 차원의 검사의 자료에 대해서 조건부 독립성을 만족시키는 문항반응함수를 세울 수 있기 위해서는 최소한 k 개의 $\underline{\theta}$ 의 요소 (components)가 필요하다는 뜻이다. 특히 k=1 인 경우의 검사를 일차원성을 만족하는 검사라고 한다. 따라서 검사가 일차원성을 만족한다 함은 한 개의 (즉, 일차원의) 잠재능력 만으로도 그 검사의 자료에 대해서 조건부 독립성을 만족시키는 문항반응함수를 세울수 있음을 뜻한다. 실제로 널리 사용되고 있는 LOGIST 나 BILOG 와 같은 컴퓨터 프로그램도 일차원성의 가정에서만 사용될 수 있다. 일차원성에 관한 좀더 상세한 논의 및 여러가지 가설감정 방법과 특히 Stout (1987)의 Dimtest 라는 검정통계량에 관해서는 박정수 (1992)를 참조하기 바란다.

2.2 문항반응 모형

일차원성과 단조성의 가정을 만족하는 문항반응함수는 한 개 또는 그 이상의 문항을 특성지우는 모수 (parameter) 들에 의존한다. 현재 가장 널리 사용되고 있는 일차원의 문항반응함수들은 로지스틱 (logistic) 모형에 기초하고 있는데, 이들을 살펴보면 다음과 같이 모수의 갯수에 따라 1-모수, 2-모수, 3-모수 로지스틱 모형이라고 불린다. (이들의 구체적이고도 자세한 내용은 성태제 (1991) 의 책, 제 1 장과 제 2 장을 보기 바람.) 이때 θ 는 벡터가 아니고 일차원의 스칼라 값 임을 주시한다 (즉, 이 부분 이하에서는 모든 검사의 일차원성을 가정하기 때문이다).

① 1-모수 로지스틱 모형 (1-parameter logistic model, 1PL) 또는 Rasch 모형,

$$P_j(\theta) = \frac{1}{1+e^{-(\theta-b_j)}} \tag{2.3}$$

여기서 θ 는 능력을 나타내고, b_j 는 문항 j 의 난이도 (item difficulty) 에 해당하는 위치모수 (location parameter) 이다. 즉, 큰 b_j 값은 어려운 문항을 나타내고, 작은 b_j 값은 쉬운 문항을 뜻한다. 여기서 난이도 b_j 는 $P_j(\theta)$ 의 변곡점이며 동시에 $P_j(\theta)=1/2$ 이 만족되는 θ 의 값으로 정의되었음을 알 수 있다.

② 2-모수 로지스틱 모형 (2-parameter logistic model, 2PL),

$$P_j(\theta) = \frac{1}{1+e^{-a_j(\theta-b_j)}} \tag{2.4}$$

여기서 a_j 는 문항 j 의 변별도 (item discriminating power) 을 나타내고 b_j 는 1-모수 모형에서와 같은 문항 난이도를 나타낸다. 변별도 a_j 는 $\theta=b_j$ 에서의 $P_j(\theta)$ 의 기울기를 뜻하며 a_j 가 클수록 변별도가 커지며 a_j 가 작을수록 변별도도 작아진다. 여기서 변별도가 큰 문항이란 능력이 많은 피험자와 능력이 적은 피험자 간에 (특히 난이도 근처에서) 구분을 분명히 해주는 문항을 말한다.

③ 3-모수 로지스틱 모형 (3-parameter logistic model, 3PL) 또는 추측 모형,

$$P_j(\theta) \;=\; g_j \;+\; \frac{1-g_j}{e^{-a_j(\theta-b_j)}} \tag{2.5}$$

여기서 g_j 는 추측모수 (guessing parameter) 로서 다항선택의 문항에서 단지 추측의 결과로서 정답을 할 확률을 말한다. 만약 정답들이 각 문항에 임의로 퍼져있고 추측하는 모든 피험자들이 임의로 추측한다면 g_j 값은 1/A 이 될 것이다 (여기서 A 는 각 문항에서의 선택의 수이다). 다른 모수들의 의미는 2-모수모형의 경우와 같으나 난이도 b_j 는 $P_j(\theta)$ 의 변곡점, 즉, $P_j(\theta)=(1.+g_j)/2$ 를 만족하는 θ 의 값으로 정해진다. 추측모수 g_j 는 문항반응함수의 하한 (lower asymptote) 을 나타내는데, 만약 g_j 가 0 이면 3PL 은 2PL 과 같아진다.

로지스틱 함수와도 관련된 또 다른 문항반응함수로서 정규 오자이브 (Normal Ogive) 모형이 있는데, 이는 다음과 같다.

$$P_j(\theta) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\theta-b_j}{\sigma_j}} e^{-\frac{1}{2}t^2}\, dt \tag{2.6}$$

여기서 σ_j 는 문항 분산 (item dispersion) 이라고 불리는데 로지스틱 모형에서 $1/a_j$ 에 해당되는 값이다. 식 (2.6) 은 평균이 b_j 이고 분산이 σ_j^2 인 정규분포의 누적분포함수이다. 이러한 문항반응함수들이 실제 자료에 얼마나 잘 부합되는지는 자료를 이 모형에 적합시킨 다음 잔차분석등을 통하여 검정할 수 있겠다. 일반적으로 실제문제를 다루는데는 정규오자이브 함수보다는 로지스틱 함수가 선호되고 있으며, 특히 3-모수 로지스틱 함수가 널리 쓰이고 있다.

3. 문항 모수 및 피험자 능력의 추정 방법

3.1 추정의 어려움

문항모수 ($\underline{a}$, $\underline{b}$, $\underline{g}$) 또는 능력모수 ($\underline{\theta}$) 어느 한쪽을 알고 있을 때, 다른 쪽의 모수들을 추정하기는 비교적 어렵지 않다. 이 경우에, 다음에서 기술하는 여러 가지 추정법과 기본적으로 비슷한 원리를 사용하지만 그 추정치를 구하기가 상대적으로 용이하다는 뜻이다. 그런데 현실적으로 능력모수를 알고 있는 경우는 드물 것이므로 이러한 논의는 여기서는 생략한다. (관심있는 독자들은 Hambletone and Swaminathan (1985, Chapter 5) 을 참조하기 바란다.) 이 장에서 우리는 두 가지의 모수들을 모두 모르는 경우 만을 중점으로 다룰 것이다. 그러나 이 과정에서 어느 한쪽의 모수들을 먼저 추정하고 이를 바탕으로 또 다른 쪽의 모수들을 추정하므로 위의 방법을 이용하게 된다. 검사의 반응치를 바탕으로 능력을 추정 (test scoring) 하기 위해서 먼저 문항모수를 추정해야 하는데, 문항모수를 추정하고 모형의 적합성을 검증하는 일을 item calibration 이라고 한다.

추정상의 문제는 문항의 수 또는 피험자의 수가 증가할수록 추정해야할 모수의 수가 증가하게 되므로 모수추정을 어렵게 만든다. 즉, 추정해야 할 모수의 수는 J 개의 문항에 대한 문항모수의 수와 N 명의 피험자의 능력모수의 수의 합에 비례한다. 2-모수 로지스틱 모형을 사용하는 경우, 추정되어야 할 총 모수의 수는 N+2J-2 이다. 여기서 2 만큼 빼준 것은 모수의 무결정성 (indeterminacy) 때문이다. 이 무결정성을 피하기 위해 보통 θ(또는 b) 의 평균이 0 이고 분산이 1 이라고 고정시킨다. 그러면 θ 및 b 는 대략 -4 에서 4 사이의 값을 취하게 된다.

이 장에서는 모수들에 대한 여러 가지 통계적 추정방법에 대해 다루고자 하는데, 이 방법들은 최우추정법과 베이지안 추정법을 중심으로 하고 있다.

3.2 결합 최우추정법 (Joint Maximum Likelihood Estimation, JMLE)

U_{ij} 를 i 번째 피험자의 j 번째 문항에 대한 정오의 확률변수라 할 때, 이는 베르누이분포를 따를 것이다. 각 피험자들은 서로에게 아무런 영향을 미치지 않는다는 가정하에서 검사가 치뤄지므로, 조건부 독립성의 가정과 함께 식 (2.2) 를 이용하면, 결합우도함수는 다음과 같아진다.

$$L(\underline{u} \mid \underline{\theta}, \underline{a}, \underline{b}, \underline{g}) = \prod_{i=1}^{N} \prod_{j=1}^{J} P_j(\theta_i)^{u_{ij}} (1-P_j(\theta_i))^{1-u_{ij}} \tag{3.1}$$

여기서 $P_j(\theta_i)$ 는 i 번째 피험자의 j 번째 문항에 대한 문항반응모형을 나타내는데, 이는 예를 들어 3PL 의 경우 a_j, b_j 와 g_j 에 의존하고 있음을 상기한다. 따라서 능력 및 문항모수의 JMLE 는 식 (3.1) 을 최대로 하는 모수들을 찾음으로서 얻어진다. 이 최대값은 각 모수들에 대한 식 (3.1) 의 1차 미분치가 0 이 되는 점에서 찾을 수 있는데, 이러한 일련의 방정식을 풀기 위해서는 Gauss-Newton 의 최적화 방법과 같은 수치적인 연쇄계산 (numerical iteration) 이 필요하다. 이러한 계산은 LOGIST 에서 수행되는데, 실제의 계산을 살펴보면, 먼저 문항모수 ($\underline{a}$, $\underline{b}$, $\underline{g}$) 의 초기치를 주고 이를 고정시킨 다음 (3.1)을 능력모수 $\underline{\theta}$ 에 대하여 최대화 함으로서 $\underline{\theta}$ 를 추정한다. 이렇게 구해진 $\underline{\theta}$ 를 이제 고정시키고 (3.1)을 최대로 하는 문항모수를 추정한다. 이러한 2-단계 과정 (ping-pong process) 은 수렴될 때까지 계속된다.

이 JMLE 방법은 개념상으로 적절하며 LOGIST (Wingersky et. al., 1982) 의 보급과 함께 널

리 사용되고 있으나, 다음과 같은 문제점을 가지고 있다. 먼저 문항수가 고정되어 있고 피험자 수가 무한히 증가할 때라도 문항모수의 추정치가 일치추정량이 되지 못하고, 또 피험자수가 고정되고 문항수가 무한히 증가할지라도 능력모수의 추정치가 일치추정량이 되지 못한다는 이론적인 난점을 갖는다. 두번째 문제점으로는, 피험자 수가 늘어남에 따라 계산상의 어려움이 크게 증가하고, 모든 문항을 다 맞춘 피험자의 능력과 한 문항도 못 맞춘 피험자의 능력은 추정할 수 없다는 점이다. 한편 다음에서 다루는 주변 최대우도 추정법이나 베이지안 추정법에서는 이런 피험자의 능력을 추정할 수 있다.

3.3 조건부 최우추정법 (Conditional MLE, CMLE)

Andersen (1970) 은 Rasch 모형에서 i 번째 피험자의 정답수 $r_i = \sum_{j=1}^{J} u_{ij}$ 은 그 피험자의 능력 모수 θ_i 에 대한 충분통계량 임을 보였다. 즉, 조건부 우도함수 $L(\underline{U} \mid \underline{r}, \underline{b}, \underline{\theta})$ 는 $\underline{\theta}$ 에 무관한 함수가 되고, 따라서 이를 먼저 $\underline{b}$ 에 대하여 최대화 함으로서 문항모수의 조건부 최우추정치를 구한다. 능력모수는 문항모수의 추정치를 (3.1)에 대입하여 이를 최대로 함으로서 구해진다. 결국 이 능력의 추정치는 정답수 r 의 함수로서 구해지는데, 이는 고전검사이론에서 능력을 정답수로 추정하는 경우와 유사하다. Andersen (1970) 이 지적 한대로 CMLE 는 여러 가지의 좋은 특성 (일치추정량) 을 가지고 있으나 그 계산상의 어려운 점이 (특히 많은 문항수에 대해서) 있고, 오직 Rasch 모형에만 적용할 수 있는 한계를 가지고 있다.

3.4 주변 최우추정법 (Marginal MLE, MMLE)

위의 CMLE 에서 본 바와 같이 능력 θ 의 충분통계량이 존재하면 조건부 우도함수는 문항모수들만의 함수가 되어 이를 최대로 함으로서 좋는 모수추정치를 구할 수 있는데, 2 모수 이상의 로지스틱 모형과 모든 정규 오자이브 모형에서는 불행하게도 θ 의 충분통계량이 존재하지 않음으로서 위의 과정을 따를 수 없다. 그런데 여기서 θ 에 무관한 우도함수를 유도해 내는게 중요하므로, 베이지안 방법에서처럼 θ 를 하나의 확률변수로 간주하여 θ 의 사전확률 밀도함수 (사전 pdf) 를 이용하여 적분 (integrate out) 을 시킴으로서 θ 에 무관한 주변 우도함수 (marginal likelihood function) 를 유도해낼 수 있다. θ 를 확률변수로 보는 것은 어떤 능력을 가진 피험자들이 어떤 능력의 분포를 이루는 전체 피험자 집단으로 부터의 임의 추출되었다고 해석된다. 이 경우에 문항에 대한 주변 확률 (marginal probability) 은 다음으로 계산된다.

$$P(u) = \int P(u \mid \theta)\, g(\theta)\, d\theta. \tag{3.2}$$

여기서 $g(\theta)$ 는 θ 의 사전 pdf 로서 그 자체에 모르는 모수 (hyperparameter) 를 포함하고 있다. 이에 대응하는 주변 우도함수는 다음과 같다.

$$L(\underline{u} \mid \underline{a}, \underline{b}, \underline{g}) = \prod_{i=1}^{N} \int \prod_{j=1}^{J} P_j(\theta)^{u_{ij}} (1-P_j(\theta))^{1-u_{ij}} \cdot g(\theta)\, d\theta. \tag{3.3}$$

이 식은 피험자 수가 증가하더라도 변하지 않는 문항모수들과 $g(\theta)$ 내의 모르는 모수들에만 의존하고 있다. 따라서 문항모수의 MMLE 는 (3.3) 을 문항모수에 대하여 최대로 함으로서 구해진

다. 피험자수가 증가할 때 문항모수의 MMLE 는 일치추정량이 된다. 일단 문항모수가 추정되면 그것을 대입한 식 (3.1)을 능력모수에 대하여 최대로 함으로서 능력모수를 추정할 수 있다. 또는 θ 의 사전 pdf $g(\theta)$ 를 이용하여 베이지안 방법으로 추정할 수 있다. 즉, 문항모수가 추정되어서 고정된 상태에서의 $\underline{\theta}$ 의 사후 pdf 는

$$f(\underline{\theta} \mid \underline{u}) \quad \propto \quad L(\underline{u} \mid \underline{\theta})\, g(\underline{\theta}) \tag{3.4}$$

이므로 이를 $\underline{\theta}$ 에 대하여 최대로 함으로서 능력모수의 베이지안 추정치 (maximum a posterior estimator) 를 구할수 있다. 또는 $\underline{\theta}$ 의 사후 기대치 (expectation a posterior) 를 구할수 있다.

실제에 있어서 식 (3.3) 의 적분은 쉽게 계산되지 않는다. 따라서 수치해석적인 방법을 이용하여 작은 오차 내에서 근사적으로 구해진다. 이는 가우시안 구적 공식 (Gaussian quadrature formula) 을 이용 함으로서 수행된다. 능력모수의 사전 pdf $g(\theta)$ 의 선택이 중요한데, 만약 θ 에 대한 사전 정보가 없다면 일반적으로 정규분포가 많이 사용된다. 정규분포를 사용하기 위해서는 평균과 분산을 알아야 되는데 이는 미리 주어지거나 응답 자료로부터 추정하여 사용된다.

원래 MML 방법은 2PL 의 경우 Bock and Lieberman (1970) 에 의해서 제안 되었다. 그들은 θ 의 사전분포로 표준정규분포를 사용하였다. 이때 가장 큰 문제는 2^J 만큼의 가능한 응답형태에 대해 수치해석적인 적분을 실시해야 한다는 계산상의 문제점이였다. 이런 계산상의 난점은 Bock and Aitkin (1981) 에 의해 θ 의 사전 pdf 를 경험적으로 (empirical Bayes) 추정하고, 수정된 EM 알고리즘을 적용함으로서 상당히 개선되었다. 이는 MMLE/EM 방법이라고도 불리는데, 이것이 컴퓨터 프로그램 BILOG 에 구현되었다. 자세한 사항은 Bock and Aitkin (1981) 및 Harwell, Baker and Zwarts (1988) 에서 찾아 볼 수 있다. 특히 마지막의 참고문헌은 MMLE/EM 방법의 중요한 수학적 세목과 그것이 모수 추정을 위해서 어떻게 이용되고 있는지 상세히 보여주고 있다. 이 MMLE/EM 방법은 다음에서 다루는 베이지안 추정법과 혼합되어 사용되기도 한다 (marginal maximum a posterior, MMAP 추정법).

3.5 베이지안 추정법

문항모수에 대한 사전정보를 가지고 있을 때, 또는 JMLE 나 MMLE 에서 문항모수의 추정이 실패하였거나 엉뚱한 추정치를 얻었을 때, 문항모수의 사전 pdf 를 이용한 베이지안 추정법을 생각할 수 있다. 이 경우 ($\underline{\theta}$, $\underline{a}$, $\underline{b}$, $\underline{g}$) 의 사후 pdf 는 다음과 같다.

$$p(\underline{\theta},\ \underline{a},\ \underline{b},\ \underline{g} \mid \underline{u}) = L(\underline{u} \mid \underline{\theta},\ \underline{a},\ \underline{b},\ \underline{g}) \left\{ \prod_{j=1}^{J} h(a_j)\ h(b_j)\ h(g_j) \right\} \cdot \prod_{i=1}^{N} g(\theta_i) \tag{3.5}$$

여기서 $h(\cdot)$ 는 각 문항모수의 사전 pdf 이다. 베이지안 추정치는 위의 사후 pdf 를 최대로 하는 문항모수 및 능력모수이다. 또 다른 추정법으로는 주변 최대우도 추정법에서 처럼 먼저 $\underline{\theta}$ 를 적분 (integrate out) 한 뒤, 주변 사후 pdf 를 문항모수에 대하여 최대로 함으로서 문항모수를 추정할 수 있다. 이를 주변 최대 사후 추정법 (marginal maximum a posterior estimation, MMAP) 이라고 하며 BILOG 에 구현되어 있다.

보통 $\underline{a}$ 는 대수정규 (log Normal) 분포가 가정되고, $\underline{b}$ 와 $\underline{\theta}$ 는 정규분포를 가정한다. 사전분포의 모수는 사전 정보에 의해서 주어질 수도 있고 또는 자료로부터 추정될 수도 있다. 3-모수 모

형에서도 추측모수 g_j 가 0 과 1 사이에 있도록 하기 위하여, 각 문항에 대해서 α 와 β 두개의 모수를 갖으며 0 과 1 사이의 구간에서 정의되는 베타 분포가 적절할 것이다. 베이지안 추정법의 자세한 내용은 Mislevy (1986) 을 보기 바란다.

이런 베이지안 방법은 다른 방법에서 가끔 일어나는 경우, 즉 추정치가 엉뚱하게도 어떤 영역 밖의 값이 되는 경우를 사전 분포를 이용하여 자연스럽고 효과적으로 방지해 준다는 이점을 가지고 있다. 이런 이점에도 불구하고 베이지안 방법은 아직 좀더 많은 연구가 요구되고 있다.

4. 93년도 대입학력고사에의 응용

위에서 다룬 이론 및 방법을 93년도 대입학력고사 중 수학 II 문제중 객관식 (4지 선다형) 26 문항으로 적용하였다. 피험자는 전남대학교 자연계열에 지원한 응시자 중 4404 명을 대상으로 하였다. BILOG 를 이용하여 각 문항을 3 모수 로지스틱 모형 (3-PL) 에 적합시켰다. 여기서 구해지는 문항모수의 추정치는 전국의 응시자 모두를 대상으로 한다고 해도 (적어도 이론상으로는) 큰 차이를 보이지 않을 것으로 기대된다. 왜냐하면 문항모수의 불변성 원리 때문이다. 문항반응이론의 가장 큰 특성은 문항모수의 불변성과 능력모수의 불변성인데, 첫번째는 주어진 문항에 대하여 어떤 피험자 집단이 시험을 치르드라도 그 문항의 특성모수는 고유의 값으로 불변한다는 원리이다. (능력모수의 불변성이란 어떤 문항들로 시험을 치르더라도 한 피험자의 능력은 고유의 값으로 일정하다는 원리이다.) 따라서 조금의 편차는 있겠지만 (실제의 Bilog 출력에서는 기대보다 큰 차이를 낼 수도 있지만), 능력이 우수한 피험자 집단을 대상으로하여 추정한 문항모수 추정치와 능력이 우수하지 못한 피험자 집단에서의 추정치 간에는 근본적인 차이는 없다는 것이다. 다만, 능력의 사전분포 $g(\theta)$ 또는 사전분포의 모수 (hyperparameters) 를 자료로부터 추정하는 경우에 피험자 집단의 특성에 따라 MMLE 나 베이지안 추정치에 변화를 가져오기는 할 것이다.

4.1 문항모수의 추정

문항반응이론의 기본 가정인 조건부 독립성과 일차원성을 만족하는지를 가설검정하기 위하여 Stout (1987) 의 Dimtest 를 실행한 결과, 이들 가정이 기각되지 않았다. 따라서 BILOG 를 이용하여 3-PL 모형에서, 문항모수는 MMAP (각 모수에 대해 사전 pdf 가 주어진 상태에서 주변 사후분포를 최대로 하여 추정) 방법으로 추정하였다. 능력모수는 EAP (expected a posterior) 방법으로 추정하였다. 능력의 사전 pdf 는 자료로부터 추정 (empirical Bayes) 하였는데, 대략 평균이 0 이 되고 표준편차가 1 이 되도록 변환하였다. 문항모수의 사전 pdf 는 default 로서 난이도, 변별도, 추측도에 대하여 각각 정규분포, 대수 정규분포, 베타분포로 주어졌다 (부록에 BILOG 실행를 위한 프로그램이 제시되어 있으므로 참조 바람). BILOG 내의 구체적인 계산 과정, 자료의 변환 및 사전분포의 선정에 대해서는 Mislevy and Bock (1990) 을 참조하기 바란다.

<표 4.1>은 각 문항의 모수추정치이다. 여기서 threshold, slope, asymptote 는 각각 난이도, 변별도, 추측도를 말한다. Intercept 는 난이도 곱하기 변별도에 음수를 취한 값으로 능력 0 에서의 logit $(=a_j(\theta-b_j))$ 값을 뜻하며, dispersn 은 변별도의 역수이다. Chisq 는 각 문항에서의 3-모수 로지스틱 모형에 대한 적합도 검정을 위한 우도비 카이제곱 통계량 값이다 (각 모수의 추정치의 표준오차와 Chisq 에 대한 p-value 는 아래 표에서는 생략함). 따라서 매우 큰 카이제곱 값은 그

문항에 대한 3-모수 모형의 부적합성을 나타낸다. <표 4.1>에서는 5%의 유의수준에서 3개의 (6번, 13번, 17번) 문항이 부적합으로 판정되었다. 4지 선다형 문항이므로 추측모수가 1/4 보다 큰 문항은 바람직하지 못하다고 볼 수 있는데, 5개 (1번, 10번, 21번, 24번, 28번)의 문항에서 나타났다. 변별도가 낮은 문항 또한 좋지 못하다고 할 수 있는데 대략 5개 (1번, 10번, 13번, 16번, 24번)의 문항에서 나타났다.

<표 4.1> MMAP 추정법에 의하여 구한 문항모수의 추정치

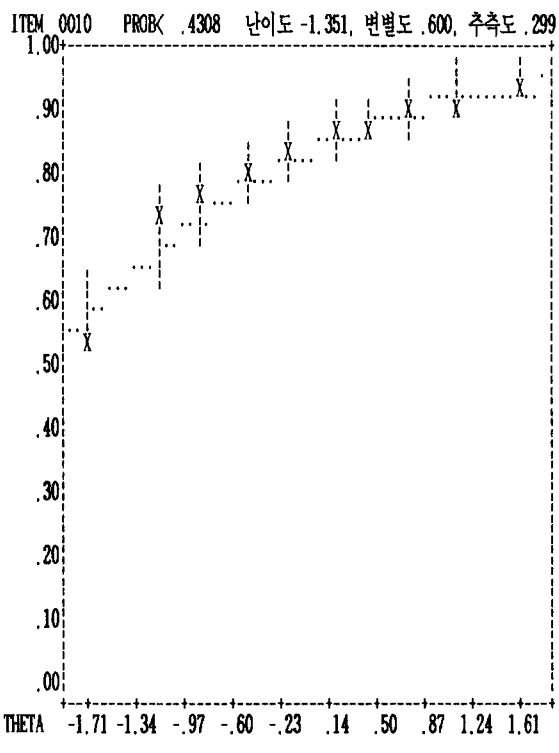
ITEM	INTERCEPT	SLOPE	THRESHOLD	DISPERSN	ASYMPTOTE	CHISQ	DF
0001	-1.092	.756	1.445	1.323	.350	7.5	9.0
0002	.669	1.105	-.606	.905	.160	6.7	7.0
0003	.839	1.169	-.718	.855	.162	11.4	7.0
0004	-.370	1.001	.369	.999	.241	7.2	9.0
0005	-.743	1.043	.713	.959	.171	9.7	9.0
0006	.617	.935	-.660	1.069	.114	35.8	8.0
0007	1.096	.995	-1.102	1.005	.200	8.9	6.0
0008	-.750	.970	.774	1.031	.160	5.5	9.0
0009	-1.313	1.216	1.080	.822	.234	6.6	9.0
0010	.811	.600	-1.351	1.667	.299	8.0	8.0
0011	.381	.901	-.423	1.110	.186	4.8	8.0
0012	-.430	1.178	.365	.849	.211	4.6	8.0
0013	-.160	.660	.242	1.516	.193	18.0	9.0
0014	.438	1.435	-.305	.697	.155	2.4	7.0
0015	-.890	1.948	.457	.513	.220	10.4	7.0
0016	.495	.758	-.653	1.319	.211	8.5	8.0
0017	.459	1.097	-.418	.911	.121	21.2	8.0
0018	.080	.965	-.083	1.037	.203	7.3	8.0
0019	.052	1.524	-.034	.656	.245	6.5	7.0
0020	-1.645	1.032	1.594	.969	.177	14.2	9.0
0021	-.558	1.957	.285	.511	.289	13.2	7.0
0022	.692	1.191	-.581	.840	.150	4.4	7.0
0023	-.092	1.079	.085	.927	.218	9.0	8.0
0024	-1.051	.759	1.384	1.318	.314	3.5	9.0
0025	-.723	1.046	.691	.956	.172	13.7	9.0
0026	.090	.847	-.106	1.180	.283	10.2	8.0

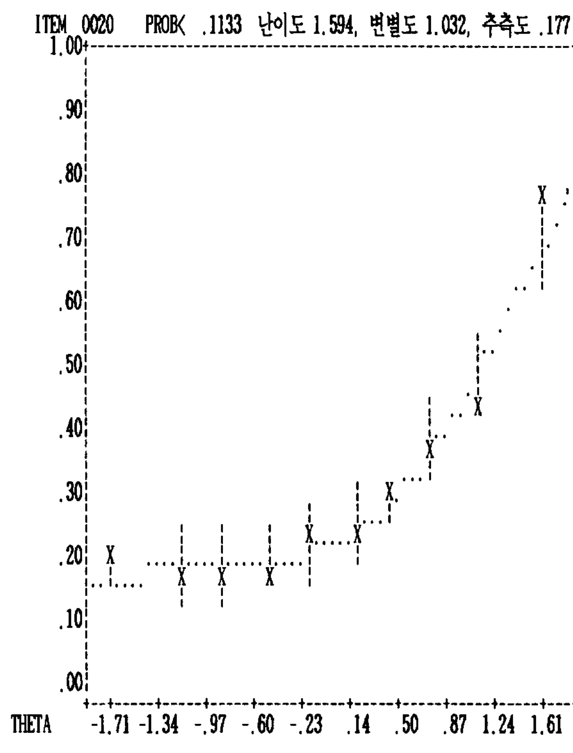
4.2 문항반응 곡선

문항모수의 추정치를 바탕으로 한 몇개 문항의 반응곡선이 <그림 4.*> 에 있다. 이 그림을 봄으로서 능력평가에 좋은 문항과 안 좋은 문항을 가려낼 수 있다. <그림 4.1> 은 문항 10번에 대한 반응곡선으로서 매우 쉬운 문항 임을 보여주고 있다. 또한 <그림 4.2>는 문항 20번의 반응곡선으로서 매우 어려운 문항임을 알수 있다. 이 문항은 난이도 1.594 로서 26개 문항중 가장 어려운 문항으로 나타났다. 여기서 10번 문항은 능력이 낮은 피험자에게만 제대로 반응하고, 20번 문항은 능력이 높은 피험자에게만 제대로 반응하게 된다. 즉, 능력이 낮은 피험자를 평가하는데 20번 문항은 별 의미가 없으며, 반대로 10번 문항은 능력이 많은 피험자를 평가하는데 별 기여를 못한다.

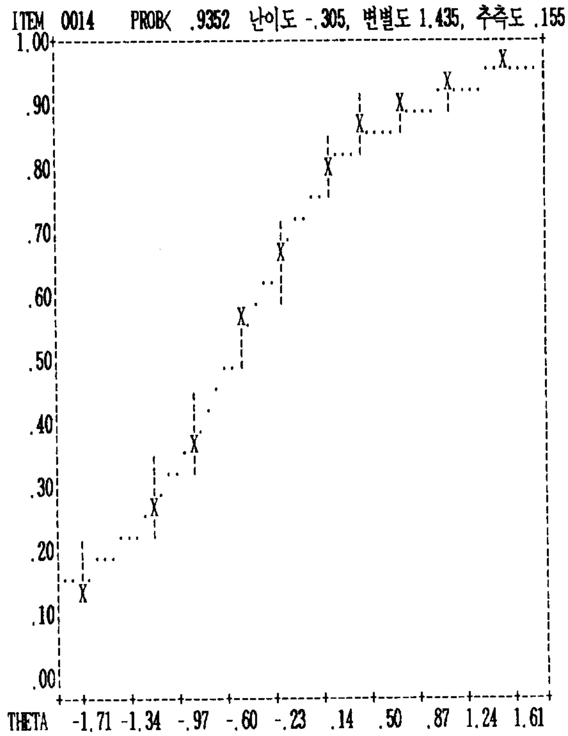
한편, 문항 14번은 능력 평가를 위해 매우 바람직한 반응곡선을 보이고 있는데, <그림 4.3>이 그것이다. 이는 낮은 추측모수, 적절한 난이도와 1.435 의 적절한 변별도를 가지면서 반응곡선이 직선에 가까운데, 이는 그 문항을 맞출 가능성이 피험자의 능력에 선형적으로 정비례하게 나타남으로서 피험자의 능력을 그대로 평가하는데 목적이 있는 학력고사에 알맞는 문항이다. (한편 변별도가 매우 높은 문항은 예를들어 합격 불합격을 가리는 선발고사 등에 알맞다.) 즉, 이 문항은 능력이 낮은 피험자와 보통 및 높은 피험자에 대하여 모두 반응함으로서, 어떤 능력의 피험자를 평가하는데도 도움이 되는 문항이다.

반면, <그림 4.4> 는 문항 24번의 반응곡선으로서 바람직하지 못한 문항의 한 예를 보여주고 있다. 실제로 추측도는 매우 높은데 변별도는 낮고 난이도는 높은 문항이다. 따라서 이 문항은 능력이 많은 피험자도 정답하기 힘든 (약 80% 맞출 확률), 매우 어려운 문항임에도 불구하고 추측에 의해서 맞출 가능성이 높은 문항이다. 즉, 어느 정도 능력이 있는 피험자라도 자신의 능력으로 맞추기보다는 추측에 의해서 더 맞출 가능성이 높다는 이야기가 되어, 정확한 능력 평가에 바람직하지 못하다.

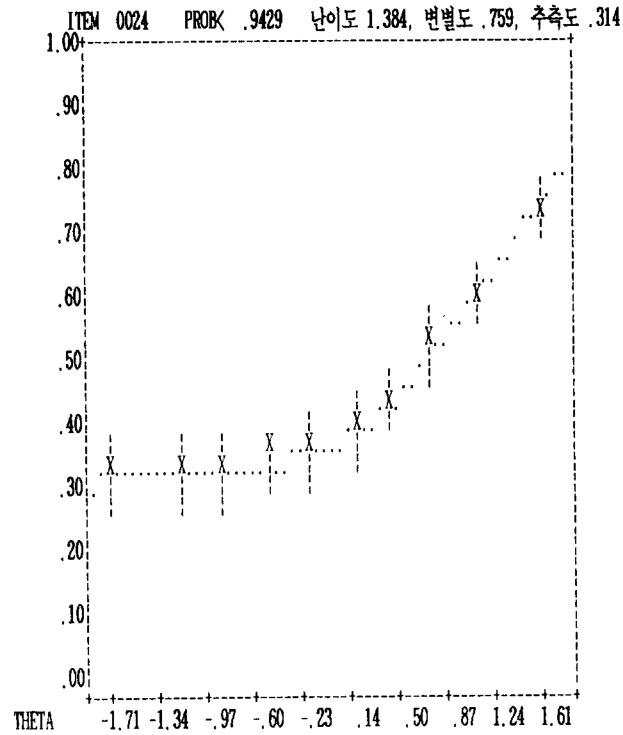




<그림 4.1> 문항 10번의 반응곡선으로서 매우 쉬운 문항임. <그림 4.2>문항 20번의 반응곡선으로서 매우 어려운 문항임.



<그림 4.3> 문항 14 번의 반응곡선으로서 바람직한 문항임.



<그림 4.4> 문항 24 번의 반응곡선으로서 안 좋은 문항임.

4.3 피험자 능력의 추정

일단 문항모수를 위와같이 추정한 뒤, 능력모수는 EAP 방법으로 추정하였다. 여기서 능력의 사전 pdf 는 자료로부터 추정하였는데 (empirical Bayes), 추정된 능력은 평균이 50, 표준편차가 20 이 되도록 변환하여 출력하도록 하였다. 실제로는 4404 명 각각에 대하여 능력추정치가 출력되었으나, 설명을 위해 <표 4.2>에는 그 일부만 실었다.

26개 모두를 정답한 피험자의 점수는 94.63 이 되었고, 단지 3개만을 맞춘 피험자의 점수는 6.92 (가장 낮은 점수) 이다. 같은 갯수 만큼 정답한 피험자라도 능력 추정치는 다름을 볼 수 있다. 이와같이 차이가 나는 이유는 같은 갯수를 맞췄다고 해도 서로 다른 문항을 맞췄을 것이고, 또 각 문항들은 다른 모수를 가지기 때문이다. 예를들어 쉬운 문항을 맞춘 경우에는 낮은 가중치를 주고 어려운 문항을 맞춘 경우에는 높은 가중치를 주는 것과 같은 이치로 인한 차이이다.

능력의 추정치에 대한 표준오차는 다음과 같이 정의되는 검사의 정보함수 (information function) 로 부터 구해진다.

$$I(\theta) = -E\left\{ \frac{\partial^2}{\partial \theta^2} \log L(\underline{u} \mid \theta) \right\}$$

(4.1)

이것은 Fisher 의 정보함수로서, 실제 값은 BILOG 내에서 Gauss-Newton 최적화의 마지막 단계에서 구해진다. 추정된 능력의 표준오차 (S.E.) 는, 추정치가 일치통계량이라는 가정 하에서, 그 능력에서의 정보함수의 역수의 제곱근으로 주어진다.

<표 4.2> 각 피험자의 능력의 추정치와 표준오차 (극히 일부만 수록함)

WEIGHT	SUBTEST	TRIED	RIGHT	PERCENT	ABILITY	S.E.
1.00	MATH93	26	17	.6538	55.1721	5.2066
1.00	MATH93	26	22	.8462	73.0335	5.8943
1.00	MATH93	26	22	.8462	66.4753	8.2908
1.00	MATH93	26	16	.6154	51.6821	7.7328
1.00	MATH93	26	26	1.0000	94.6313	11.7729
1.00	MATH93	26	13	.5000	41.6396	5.9801
1.00	MATH93	26	20	.7692	64.4008	8.3720
1.00	MATH93	26	7	.2692	20.1373	9.5378
1.00	MATH93	26	21	.8077	60.4900	7.1202
1.00	MATH93	26	20	.7692	57.7616	4.7949
1.00	MATH93	26	3	.1154	8.2553	9.4004
1.00	MATH93	26	23	.8846	76.1159	7.5307
1.00	MATH93	26	21	.8077	68.4106	7.7836
1.00	MATH93	26	11	.4231	37.1159	7.1726

5. 결론 및 논의와 연구과제

이상에서 우리는 문항반응함수에 바탕하여 문항 및 능력모수의 추정방법들과 그 오차에 대해서 공부하였다. 추정방법들을 대략 살펴보면, 먼저 결합우도함수을 최대화 함으로서 능력모수와 문항모수를 동시에 추정하는 결합우도 추정법 (Joint MLE, JMLE) 이 있다. 이 방법은 LOGIST 라는 컴퓨터 프로그램의 보급과 함께 널리 사용되고 있으나 몇가지의 이론상의 문제점이 있다. 한편 JMLE 의 한 문제점을 해결한 것으로, 능력모수에 대한 충분통계량이 주어진 조건부 우도함수를 최대화 함으로서 문항모수들을 추정하는 조건부 최우추정법 (Conditional MLE) 이 있는데, 이는 오직 Rasch 모형 (즉, 1PL) 에만 유효하다는 한계를 가지고 있다. 한편 능력모수 θ 를 하나의 확률변수로 보고 θ 의 사전분포 함수를 이용해 우도함수를 문항모수들만의 함수로 변형시킨 다음 이를 최대화 함으로서 문항모수들을 먼저 추정하고, 이 문항모수들의 추정치를 이용하여 다시 능력 모수를 추정하는 주변 최우추정법 (Marginal MLE) 이 있다. 이 방법은 Bock and Aitkin (1981) 에 의한 EM 알고리즘과 가우시안 구적 수치적분법의 적용으로 이론상으로 그리고 계산상으로 적절한 방법으로 평가 받고 있으며 BILOG 라는 컴퓨터 프로그램의 보급 함께 널리 사용되고 있다. 또한 문항모수들도 확률변수로 간주하면서 그들의 사전분포함수를 이용하여 모수를 추정하는 베이지안 방법과 이들의 혼합적 방법들을 살펴 보았다.

본 논문에서는 특히 BILOG 에서 적용되고 있는 MMLE 와 베이지안 추정법에 주목하였고, 이 방법을 실제로 93년도 대입학력고사 수학 II 문항에 적용하였다. 여기서 제시된 표준오차의 공식들은 모두 대표본의 가정하에서 얻어지는 것들이다. 따라서 문항수가 적은 경우나 피험자수가 적은 경우에는 주의하여야 할 것이다. 이 논문에서는 주어진 자료가 우리가 세운 로지스틱 문항반응모형에 얼마나 잘 적합한가 (model-data fit) 를 다루는 자세한 연구는 소개되지 않았다. 또한 문항 및 능력에 따른 정보함수에 관한 구체적 논의와 그 유용성에 대하여 다루지 않았는데, 이는 성태제 (1991) 또는 이종성 (1990) 에서 찾아볼 수 있다.

문항반응이론을 이용하여 교육측정 및 평가에 도움이 될 수 있는 연구 분야로는 검사의 비교 및 동등화, 편파문항의 검색, 문제은행의 개발, 능력별 수준검사 등이다. 검사의 동등화 기법과 그것의 대학수학능력 시험으로의 적용에 관하여는 황소림 (1993) 을 참조하고, 편파된 문항의 검색에 관하여는 박정수, 노석준 (1993) 을 보기 바란다.

문항반응이론은 아직 일차원성과 국소적 독립성의 가정이 만족되지 않을 경우의 문항 및 능력의 추정과 그의 응용에 실제로 사용할 수 있을 정도의 충분한 연구가 되어있지 않다고 한다. 이러한 방향에서의 앞으로의 문항반응이론의 기초적인 연구가 기대된다 하겠다. 또한 선택형 문제가 아닌 서술형 또는 주관식 문제와 같이 다단계 점수나 부분점수 (partial credit) 를 주는 검사의 분석 등에 관한 이론의 개발 및 실용화가 절실히 요구된다 하겠다. 이 분야에 대해서는 지은림 (1993) 과 Baker (1992) 를 참조하기 바란다.

통계학적 연구의 입장에서 볼 때, 모수 및 능력 추정치의 성격에 관한 규명이 매우 미약한 실정이다. 예를들어 JMLE 에서 피험자 및 문항수가 동시에 무한히 증가할 때, 일치추정량이 되는지 또는 근사적 정규성에 관한 이론 등은 잘 정립되어 있지 않다. MMLE 의 통계적 성질 또한 확실히 확립되지는 않았다. 베이지안 추정법에 있어서도 그 추정치의 사후분산의 계산이나 방법의 robustness 등에서 많은 연구가 요구하고 있다.

지금까지는 문항반응이론에서의 모수추정방법에 대해 고찰하였으나, 대부분의 실제의 경우에는 고전검사이론과 문항반응이론에서의 방법들의 혼합에 의한 문항 및 검사의 분석이 바람직할 것이다. 고전검사이론에 바탕한 검사분석은 문항반응이론에 비하여 훨씬 이해하기 쉬어서, 문항반응분석의 결과를 더 잘 이해할 수 있도록 도와준다 (고전검사이론과 문항반응이론의 비교는 이종성 (1986) 에 잘 나와있다). 따라서 고전검사이론은 문항반응이론을 적용하기 이전의 보조적인 방법으로서 사용될 수 있고, 결국 검사의 특성을 좀 더 잘 파악하는데 도움이 될 것이다.

6. 참고문헌

[1] 박 정수 (1992). "문항반응이론에서의 일차원성에 관한 통계적 가설 검정", 「한국교육」, 제 19 호, 73-87.

[2] 박 정수, 노 석준 (1993), "문항 및 검사의 편파성 검정을 위한 통계적 방법", 「교육평가연구」, 제 6 권 2 호, 95-122.

[3] 성 태제 (1991), 「문항반응이론 입문」, 양서원. 서울.

[4] 지 은림 (1993), "서답형 문항을 위한 부분점수모형", 「교육평가연구」, 제 6 권 2 호, 241-258.

[5] 이 종성 (1986), "고전검사이론과 문항반응이론", 「교육평가연구」, 제 1 권 1 호, 183-194.

[6] 이 종성 역 (1990), 「문항반응이론과 응용」 (Lord (1980) 의 번역), 대광문화사. 서울.

[7] 황 소림 (1993), "대학수학능력시험 제6차, 제7차 실험평가의 문항특성과 피험자 능력점수의 동등화", 「교육평가연구」, 제 6 권 2 호, 287-314.

[8] Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimates, *Journal of the Royal Statistical Society,* Series B, Vol. 32, 283-301.

[9] Baker, F.B. (1992). *Item Response Theory: Parameter Estimation Techniques,* Marcel Dekker, N.Y.

[10] Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item

parameters: An application of an EM algorithm, *Psychometrika,* Vol. 46, 443-459.
[11] Bock, R. D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items, *Psychometrika,* Vol. 35, 179-197.
[12] Hambletone, R. K. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications,* Kluwer-Nijhoff, Boston.
[13] Harwell, M. R., Baker, F. B. and Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: a diadic, *Journal of Educational Statistics,* Vol. 13, 243-271.
[14] Hattie, J. (1985). Methodology Review: Assessing unidimentionality of tests and items, *Applied Psychological Measurement,* Vol. 9, 139-164.
[15] Hulin, C. L., Drasgow, F. and Parsons, C. K. (1983). *Item Response Theory,* Dow Jones-Irwin, Homewood, Illinois.
[16] Lord, F. M. (1980). *Applications of item response theory to practical testing problems,* Erlbaum Publishing Co., Hillsdale, NJ.
[17] Mislevy, R. J. (1986). Bayes modal estimation in item response models, *Psychometrika,* Vol. 51, 177-195.
[18] Mislevy, R. J. and Bock, R. D. (1990). *BILOG 3: Item analysis and test scoring with binary logistic models,* Second edition, Scientific Software, Inc., Mooresville, IN.
[19] Stout, W. F. (1987). A nonparametric approach for assesing latent trait unidimension- ality, *Psychometrika,* Vol. 52, 589-617.
[20] Wingersky, M. S., Barton, M. A. and Lord, F. M. (1982). *LOGIST user's guide,* Educational Tesing Service. Princeton, NJ.

<부록> 1993 대입학력고사 수학 II 에 대한 BILOG 프로그람

```
>comment
 Bilog program for math93.dat with 3-PL

>global    dfname='a:math93b.dat', nparm=3, save;
>save      parm='math93.par';
>length    nitems=26;
>input     ntotal=26, sample=4404, nalt=4, nidch=1;
           (a1,26a1)
>test      tname=math93;
>calib     plot=1.0, case=1, free, tprior, float;
>score     method=2, idist=3, rsc=3, loc=50, scale=20, info=2;
```