

영과잉 순서형 프로빗 모형의 베이지안 추론: 한국형 베일리-III 검사 자료에의 적용*

석 혜 은 방 희 정 김 수 영†

이화여자대학교 심리학과

불법약물 복용, 도박, 흡연, 음주 등 문제행동의 수준을 변별하는 문항이 일반인을 포함하여 측정되면 영(zero) 범주 반응값이 과도하게 나타난다. 이러한 자료를 영과잉(zero-inflated) 자료라고 하는데 일반적인 모형을 사용하면 결과를 왜곡하거나 본연의 의미를 해석하는데 한계가 있을 수 있다. 관찰된 값이 영비율이 높은 순위범주형(ordered categorical) 변수이고, 영 값들이 두 개의 구별된 집단과 관련이 있는 경우 Harris와 Zhao(2007)의 영과잉 순서형 프로빗(zero-inflated ordered probit; ZIOP) 모형을 적용할 수 있다. 본 연구에서는 ZIOP 모형의 베이지안 분석기법을 제시하고 이것을 한국형 베일리-III 영유아 검사의 적응행동척도 자료에 적용한다. ZIOP 모형의 첫 단계에서는 영 범주에 응답한 영유아들을 특정행동을 습득하기 이전의 절대적 비적응행동 영유아집단(genuine non-participants)과 적응행동이 나타났거나 현재는 적응행동을 보여주지 않지만 상황에 따라 적응행동을 표현할 가능성이 있는 잠재적 적응행동 영유아집단(potential participants)으로 구분하는 이분형 프로빗 모형을 적용한다. 두 번째 단계에서는 잠재적 적응행동 영유아 집단과 1 이상의 범주에 속한 실제적 적응행동 영유아 집단을 합해서 전체집단으로 보고 순서형 프로빗 모형을 적용한다. 본 연구에서는 영비율이 높은 적응행동척도 자료에 OP 모형과 ZIOP 모형을 적용한 후, 고전적 추정방법과 베이지안 추정방법으로 각각 모형을 비교하고, 또한 각 설명변수에 대한 주변효과를 살펴봄으로써 영비율이 높은 소표본 자료에 대한 두 가지 추정법의 차이를 탐색한다. MCMC 방법 중에 깃스샘플링 기법을 적용하였으며 R 프로그래밍 언어를 이용하여 추정 알고리즘을 프로그래밍하고 자료를 분석하였다. 이러한 결과에 기초하여 본 연구의 의의와 실제적 적용가능성, 그리고 제한점과 후속연구에 대하여 논의하였다.

주요어 : 영과잉 순서형 프로빗 모형, 베이지안 추론, MCMC, 깃스샘플링, 베일리 검사

* 본 연구는 2012년 보건복지부 사회서비스 R&D ‘한국형 영유아 발달검사 서비스 개발’연구 과제지원금에 의해 수행된 논문임(과제번호: HI12C2024).

본 연구는 제1저자의 2015년 이화여자대학교 대학원 장학금 지원에 의한 논문임.

† 교신저자: 김수영, 이화여자대학교 심리학과, (03760) 서울시 서대문구 이화여대길 52

Tel: 02-3277-3792, E-mail: suyong.kim@ewha.ac.kr

순위범주형(ordered-categorical)으로 측정된 반응값이 종속변수의 역할을 하는 경우 자료의 연속성 및 정규성 등을 가정한 선형회귀모형을 사용하는 것은 적절치 않다. 이와 같은 자료의 특성을 고려한 몇 가지 전통적인 분석모형들(예, 순서형 프로빗 모형, 순서형 로지스틱 모형 등)이 제안되고 사용되어 왔다. 하지만 영(zero) 값이 매우 높은 빈도로 발생할 때는 이러한 모형도 높은 영의 비율을 설명하는데 한계가 있다. 이 경우 영 값이 두 개의 구분되는 집단으로부터 발생하였다고 가정하고, 이분형 프로빗(binary probit) 모형과 순서형 프로빗(ordered probit) 모형을 혼합하여 종속변수에 영향을 주는 설명변수들을 탐색하는 영과잉 순서형 프로빗(zero-inflated ordered probit, ZIOP; Harris & Zhao, 2007) 모형이 상대적으로 최근에 제안되었다. 본 연구에서는 영비율이 높은 적응행동 자료에 베이지안 방법을 이용하여 ZIOP 모형을 추정하고, 심리학이나 교육학 등 사회과학 분야에서 모형의 활용가능성을 논의하고자 한다.

사회과학 분야에서 응답자의 반응을 측정하기 위하여 가장 빈번하게 이용되는 방법은 아마도 리커트 척도일 것이다. 많은 연구자가 리커트식 반응을 선호하는 이유는 상대적으로 구성하기 쉽고 내적 일관성을 높일 수 있으며 다양한 정의적 특성을 측정하는데 적합하기 때문이다(Nunnally & Bernstein, 1994). 리커트 척도는 4점, 5점, 7점 등 다양한 단계를 선택할 수 있는데, 적절한 단계를 선택하는 것은 응답 반응의 정확성은 물론 자료의 신뢰도와 도구의 측정학적 양호도를 검증하는 데에도 커다란 영향을 준다. 리커트 척도로 측정된 자료를 이용하여 통계분석을 할 때, 측정의 단계수가 매우 적은 경우(예, 5개 미만)에는

순위범주형 변수로서의 특성을 반영한 적절한 자료 분석 방법을 적용하는 것이 필요하다. 일반적인 순위범주형 변수의 취급에도 주의가 필요하지만, 특히 어느 한 범주가 높은 빈도로 나타날 때는 모형의 선택과 추정결과의 해석에 더욱 주의를 기울여야 한다. 일반적으로 도박, 음주, 청소년 비행 등의 문제행동 수준을 변별하는 문항반응 값은 많은 영을 포함하고 있다. 예를 들어, 불법약물복용과 관련한 설문에서 “당신은 ○○약물을 얼마나 자주 복용하십니까?”라는 질문에 대한 반응값으로 영이 과도하게 나타날 수 있는데, 이러한 자료를 영과잉 자료(zero-inflated data)라고 한다. 과도한 영 값이 존재하는 빈도자료(count data) 또는 순위범주형 자료에 일반적인 통계모형을 사용하면, 분석 결과가 왜곡되거나 본연의 의미를 해석하는데 어려움이 발생할 수 있다. 빈도자료는 순위범주형 자료와 형태의 측면에서 유사하고, 분석모형의 발전 과정도 공유하는 부분이 상당히 많기 때문에 함께 살펴볼 필요가 있다. 일반적으로 도박·음주 등의 문제행동 횟수를 나타내는 빈도자료를 분석하는 경우에는 주로 포아송(Poisson) 회귀모형을 사용하는데, 영이 과잉되면 이러한 모형들이 제대로 작동하지 않는다. 이는 높은 영의 비율로 인해 분산이 평균보다 커지고, 평균과 분산이 같다는 포아송 분포의 기본 가정이 적합하지 않게 되는 이른바 과대산포(over-dispersion)의 문제가 발생하기 때문이다. 빈도자료뿐만 아니라 순위범주형 자료에서도 영과잉 특성이 나타나면, 자료의 평균이 영에 가까워지므로 정적 왜도(positive skewness)가 커지는 문제가 있다. 이에 따라 정규성을 가정하는 일반적인 통계모형은 자료의 형태를 제대로 반영하지 못하고 이에 기반한 통계적 추론이 적절치 못

하게 된다.

영과잉 회귀모형의 발전과정과 모수추정

앞서 잠시 언급했듯이, 빈도자료를 분석하는 회귀모형은 포아송 분포의 특성을 사용하는 경우가 가장 일반적이다. 예를 들어, 제품의 생산공정 각 단계에서 단위시간당 발생하는 불량품의 개수가 갖는 확률분포로 포아송 분포를 사용하는 것은 오랜 기간 당연하게 생각되어 왔다. 하지만 점차 생산공정의 정확도가 높아지고 기술이 발달되어 공정이 안정적으로 운영되면서, 제품의 단위시간당 불량품 수가 매우 낮아 영의 비율이 과도하게 관찰되어 더 이상 포아송 분포를 적용하는 것이 적절치 않게 되었다. Lambert(1992)는 수학적 인 모형으로만 인식되었던 영과잉 포아송(zero-inflated poisson, ZIP) 분포(Singh, 1963)를 이용하여 설명변수의 효과를 추정할 수 있는 회귀모형을 영과잉 자료에 적용하였다. 이와 같은 영과잉 포아송 회귀모형의 기본 발상은 이분형 로지스틱 회귀모형과 포아송 회귀모형을 결합한 혼합모형(mixture model)을 사용하여 사건이 발생할 확률과 발생빈도를 구분하여 모형화하는 것이다. ZIP 모형으로부터 영과잉 음이항(zero-inflated negative binomial, ZINB; Wang, 2003) 모형, 영과잉 허들(zero-inflated hurdle; Gurm & Trivedi, 1996) 모형 등이 진화하였다. 최근에는 이러한 모형들을 이용한 연구가 정신의학 및 의료서비스 분야 뿐만 아니라 음주자료, 성격장애, 중독 등의 심리학 분야에도 적용되고 있다(Atkins, Baldwin, Zheng, Gallop, & Neighbors, 2013; Musio, Sauleau, & Buemi, 2010; Neelon, O'Malley, & Normand, 2010; Wright, Pincus, & Lenzenweger, 2012).

영과잉 빈도자료에 대한 분석모형의 발전과 더불어, Harris와 Zhao(2007)는 영과잉 순위범주형 자료의 분석을 위해 ZIOP 모형을 제안하였다. 이들은 영 비율이 76%인 호주의 담배 소비자자료에 이 모형을 적용하여 담배 소비양상에 영향을 주는 주요한 요인을 탐색하였다. ZIOP 모형의 접근방법은 과도한 비율의 영이 개념적으로 두 개의 구분되는 집단(genuine non-participants vs. potential participants)에서 발생하였다고 가정하고 2단계로 분석모형을 적용하는 것이다. 예를 들어, 최근 한달 간의 흡연정도를 순위범주형 문항(예, 0, 1, 2, 3의 리커트 척도)을 이용해서 측정할 때, ZIOP 모형은 영 반응값을 보인 사람들에 대해 (1)흡연경험이 전혀 없거나 흡연에 대한 자발적 동기와 흥미도 없고 장차 흡연에 참여할 의사가 전혀 없는 절대적 비참여자(genuine non-participants) 집단과 (2)자료 수집기간 동안 일시적으로 건강상 이유 또는 경제적인 이유 때문에 흡연을 하지 않았지만 미래에 흡연을 할 가능성이 있는 잠재적 참여자(potential participants) 집단으로 구분한다. 위와 같은 아이디어를 가지고, 첫 단계에서는 행동(흡연)의 발생확률을 설명하는 이분형(0 vs. 1)¹⁾ 프로빗 모형을 적용하고 두 번째 단계에서는 행동의 정도를 설명하는 순서형(0, 1, 2, 3)²⁾ 프로빗 모형을 적용한다. ZIOP 모형에 대한 논리는 이후에 더 자세히 설명한다.

1) 여기서 0은 관찰값 0 중에서 절대적 비참여자 집단을 의미하고, 1은 관찰값 0 중에서 잠재적 참여자 집단으로 분류된 0과 나머지 1, 2, 3 반응값을 보인 참여자들을 의미한다.

2) 여기서 0, 1, 2, 3은 ZIOP 모형의 첫 단계에서 1로 분류된 참여자집단만을 의미한다. 즉, 여기서의 0은 오직 잠재적 참여자집단만을 가리킨다.

이러한 영과잉 자료의 특성을 반영한 모형의 추정에는 주로 고전적 추론(classical inference) 방법으로 수행되어 왔다. 고전적 추론은 모수가 특정한 값으로 고정되어 있다고 믿는 빈도론자(frequentist)의 관점에서 우도함수를 이용하여 모수를 추정한다. 이 방법은 대표본 근사 이론을 사용하기 때문에 영과잉 자료와 같이 분포가 매우 치우친 경우 안정적인 추정치를 얻으려면 표본크기가 매우 커야한다는 단점이 있다. 한편 고전적 추론방법과는 달리 모수가 고정되어 있지 않고 어떤 분포를 따르는 확률변수라고 가정하는 베이저안 추론(Bayesian inference) 방법도 있다. 베이저안 추론은 관찰된 자료의 우도함수뿐만 아니라 적절한 사전 분포(prior distribution)에 대한 정보를 함께 이용하여 모수의 사후분포(posterior distribution)를 추정하고 이로부터 추정치를 계산한다. 기존의 고전적인 추론 방법과 비교해서 베이저안 추론 방법의 장점은 다음과 같다. 첫째는 연구자가 추정 전에 알고 있는 유용한 정보를 사전분포를 통해 추정의 과정에 통합할 수 있고, 둘째는 표본의 크기가 작은 경우에 상대적으로 더 신뢰성 있는 분석을 할 수 있으며, 셋째는 결과의 해석이 고전적인 방법보다 쉽다(예, 신뢰구간의 해석). 특히 모수에 대한 사전정보가 충분한 경우 더욱 정확한 추정과 추론이 가능하고, 다루는 문제가 복잡한 경우 고전적인 방법보다 유용한 것으로 알려져 있다. 이에 대한 경험적인 연구로 Ghosh, Mukhopadhyay와 Lu(2006)는 ZIP 모형을 이용해서 표본의 크기가 작을 때 베이저안 방법이 고전적인 방법에 비해 더 효율적이라는 결과를 제시하였고, Gurmur와 Dagne(2012)는 수리적으로 매우 복잡한 ZIOP 모형에 베이저안 접근법이 더욱 실제적이며 계산적인 측면에서도

유연성이 있다고 주장하였다.

ZIOP 모형을 적용한 국내·외 연구

앞서 언급한 바와 같이 영과잉 모형이 발견되어 오면서 응용분야가 다양해지고 있음에도 불구하고, 국내 심리학 및 교육학 등의 사회과학 분야에서 영과잉 모형을 적용한 연구는 매우 미흡한 실정이다. 영과잉 모형을 심리학이나 교육학 분야에 적용한 국내 연구로는 홍성두, 여승수, 김남순, 박상희, 이선재(2009)가 지적장애 연구에 ZIP 모형과 ZINB 모형의 활용 방안을 탐색한 연구가 있고, 김준엽, 신혜숙(2009)이 청소년 비행의 발생여부와 빈도에 영향을 미치는 학교 및 개인수준의 요인을 탐색하는 방법으로 다층 ZIP 모형을 제시하고 활용가능성을 살펴본 연구가 있는 정도이다. 하지만 이와 같은 연구들은 영과잉 빈도 자료에 적용할 수 있는 모형에 대한 것이고 심리학이나 교육학 연구에서 빈번하게 사용되는 영과잉 순위범주형 자료에는 적절하지 않다.

심리학 학술 데이터베이스인 PsycINFO에서 “zero inflated”로 검색하면 약 300여 편의 응용논문이 검색되지만, 약 98%가 ZIP 모형 또는 ZINB 모형을 이용한 논문들이며 ZIOP 모형을 적용한 논문은 5편(2%) 정도에 불과하다. 이는 ZIOP 모형이 아직은 일반 연구자들에게는 생소하고 쉽게 접근하기 어려운 모형일 뿐만 아니라, 쉽게 접근할 수 있는 통계프로그램이 개발되지 않았기 때문이다. 따라서 본 연구에서는 ZIOP 모형과 베이저안 방법의 기본적인 논리를 제시하고, 한국형 베일리-III 영유아검사(Bayley Scales of Infant and Toddler Development-Third Edition, Bayley-III; Bayley, 2006)의 적응행동척도 자료³⁾에 ZIOP 모형을

적용한다. 적용된 모형은 베이지안 방법을 통하여 모수를 추정하고 주변효과를 살펴보며, 이를 통해 심리학이나 교육학 분야에서 ZIOP 모형의 활용가능성을 탐색한다.

ZIOP 모형과 베이지안 추론

적응행동척도의 자료 분석을 위해서 알아야 할 개념인 ZIOP 모형과 베이지안 추론에 대하여 설명한다. ZIOP 모형은 이분형 프로빗 모형과 순서형 프로빗 모형이 결합된 형태이므로 제대로 사용하기 위해서는 이 두 가지 모형에 대한 이해가 필요하다. 또한 본 연구에서는 최대우도 추정과 같은 고전적 추론이 아닌 베이지안 추정 방법을 이용하였기 때문에 방법의 특징과 장점에 대하여 살펴본다.

영과잉 순서형 프로빗 모형

빈도자료의 과도한 영 값을 설명하기 위하여 포아송 모형을 ZIP 모형으로 확장하듯이, 순위범주형 자료의 과도한 영 값을 설명하기 위해 Harris와 Zhao(2007)는 순서형 프로빗 모형을 ZIOP 모형으로 확장한다. 예를 들어, 마약복용이라든지 음주 후 문제행동의 정도 등이 순위범주형(예, 0, 1, 2, 3)으로 수집된 경우, Harris와 Zhao(2007)는 1단계 이분형 프로빗 모형과 2단계 순서형 프로빗 모형을 이용하여

ZIOP 모형을 위계적으로 구성한다. 첫 번째 단계에서는 문제행동에 참여하는가 또는 그렇지 않은가에 따라 전체를 두 집단으로 나누는 이분형 프로빗 모형을 사용하고, 두 번째 단계에서는 문제행동에 참여하는 집단의 행동 정도(0, 1, 2, 3)를 순서형 프로빗 모형을 이용하여 분석한다. 이때 주의해야 할 것은 문제행동에 참여하는 집단도 설문 당시에 문제행동을 하지 않고 있을 수 있다는 사실이다. 즉, Harris와 Zhao(2007)는 리커트 척도의 자료 수집에서 영으로 보고된 값들의 일부는 문제행동을 전혀 하지 않는 절대적 비참여자의 영 값으로, 그리고 나머지 영 값은 문제행동을 할 수도 있었지만 자료수집 기간 동안 하지 않은 잠재적 참여자의 영 값으로 가정하였다.

본 연구에서는 ZIOP 모형을 구성하는 이분형 프로빗 모형과 순서형 프로빗 모형을 간략하게 설명한다. 기본적으로 반응변수 y 가 이분형인 프로빗 모형에서는 사건이 일어날 확률(p)을 모형화 한다. 설명변수 \mathbf{x} 와 p 의 관계가 S자 모양으로 나타나는 경향이 있기 때문에 프로빗 모형은 \mathbf{x} 와 p 의 비선형관계를 식 (1)과 같이 표준정규분포의 누적확률분포함수(cumulative distribution function; cdf) $\Phi(\cdot)$ 를 이용해 표기한다.

$$\Pr(y = 1|\mathbf{x}) = p = \Phi(\mathbf{x}'\boldsymbol{\beta}). \quad (1)$$

위에서 $\boldsymbol{\beta}$ 는 이분형 프로빗 모형의 모수(계수) 벡터이다. 사실 이분형 프로빗 모형을 자세히 설명하고자 하면, 이분형(0, 1) 변수 y 의 기저에 정규분포를 따르는 잠재반응변수(latent response variable) y^* 가 존재한다는 개념을 이해할 필요가 있다. 정규분포를 따르는 잠재반응변수 $y^*(y^* \sim N[\mathbf{x}'\boldsymbol{\beta}, 1])$ 는 프로빗 모형을

3) 베일리-III 영유아검사의 적응행동 척도는 주양육자가 각 문항을 읽고 영유아가 보이는 행동의 빈도에 따라 '아직못함(0), 거의안함(1), 가끔 함(2), 항상 함(3)' 중에서 하나를 응답하도록 되어 있는 순위범주형 자료이며, 난이도가 높은 문항들이 반응값은 영범주의 비율이 높게 나타난다.

이해하기 위한 핵심 개념이며, $y^* = \mathbf{x}'\beta + \epsilon$ 의 관계를 가정한다. 하지만 이에 대한 것은 아래에서 반복적으로 설명이 될 것이므로 중복 설명을 피하기 위하여 여기서는 생략한다.

다음으로, 순서형 프로빗 모형은 확률변수 y 가 관찰가능하며 0, 1, 2, 3 등의 순서가 있는 경우에 이분형 프로빗 모형과 마찬가지로 관찰된 순위형 값들의 기저에 연속적인 정규분포(y^* 분포)가 있다고 가정한다. 이러한 가정에 반응변수 y 와 설명변수 \mathbf{x} 의 관계를 설정할 때, 먼저 정규분포를 따르는 잠재반응변수 y^* 와 설명변수 \mathbf{x} 를 연결하고 이후에 경계(threshold)값 모수 τ 를 통해 관찰된 반응변수 y 와 잠재반응변수 y^* 를 연결하는 모형을 설정한다. 순서형 프로빗 모형은 식 (2)와 같으며 여기서 추정해야할 모수는 계수벡터 β 와 τ 이다.

$$y^* = \mathbf{x}'\beta + \epsilon, y = \begin{cases} 0, & y^* \leq \tau_1 \\ 1, & \tau_1 < y^* \leq \tau_2 \\ 2, & \tau_2 < y^* \leq \tau_3 \\ 3, & \tau_3 < y^* \end{cases} \quad (2)$$

바로 위에서 설명한 이분형 프로빗 모형과 순서형 프로빗 모형은 사실 범주의 개수가 다르다는 것만 제외하면 근본적으로 같은 모형이라고 볼 수 있다.

이제 Harris와 Zhao(2007)의 ZIOP 모형을 다음과 같이 두 단계로 설명한다⁴⁾. ZIOP 모형은 첫 번째 단계에서 비참여자 집단($S=0$)과 참여자 집단($S=1$)을 식별하는 이분변수 S 를 정의한다. 여기서 이분변수 S 는 잠재반응변수

4) 두 단계에서 적용되는 2가지 모형에 대한 표기법은 식 (1), (2)와는 구별하여, Harris와 Zhao(2007)에서 제안된 모형을 재해석하여 정리한 것이다.

S^* 와 연결($S^* > 0$ 이면 $S=1$)되는데 참여자 집단에 포함될 경향성을 나타내는 잠재 반응변수 S^* 는 식 (3)과 같이 주어진다.

$$S^* = \mathbf{x}'\beta + \epsilon. \quad (3)$$

여기서 \mathbf{x} 는 두 집단 중 참여자 집단을 선택하는데 영향을 주는 설명변수이고, β 는 설명변수의 계수, 그리고 ϵ 은 표준정규분포를 따르는 오차이다. 따라서 참여자 집단에 속할 확률은 식 (4)와 같이 표현될 수 있다.

$$\Pr(S=1|\mathbf{x}) = \Pr(S^* > 0|\mathbf{x}) = p = \Phi(\mathbf{x}'\beta). \quad (4)$$

첫 번째 단계에서 참여자 집단($S=1$)이 결정되면, 참여자 집단에서 관찰될 수 있는 반응수준 \tilde{y} 는 0, 1, 2, 3 등으로 표현할 수 있다. 이 때 $\tilde{y}=0$ 은 처음 수집된 관찰값 $y=0$ 중에서 문제행동을 보이지는 않았지만 언제라도 보일 수 있는 잠재적 참여자집단으로 구분된 사람들의 반응값이다. 두 번째 단계에서는 \tilde{y} 의 이면에 연속적인 정규분포가 있다고 가정하는 두 번째 잠재변수 \tilde{y}^* 를 통해서 식 (5)와 같이 순서형 프로빗 모형을 정의한다.

$$\tilde{y}^* = \mathbf{z}'\gamma + \delta. \quad (5)$$

여기서 \mathbf{z} 는 ZIOP 모형의 두 번째 단계에서 행동의 정도에 영향을 주는 설명변수이고, γ 는 설명변수의 계수, 그리고 δ 는 표준정규분포를 따르는 오차이다. 식 (5)와 같이 잠재변수 \tilde{y}^* 와 설명변수 \mathbf{z} 를 먼저 연결한 후에 경계값 모수 τ 를 통해 \tilde{y} 와 \tilde{y}^* 를 식 (6)과 같이 연결한다.

다. 추정하고자 하는 연속형 모수를 θ 라고 하면 베이즈 정리(Bayes' theorem)를 이용한 모수 θ 의 사후분포는 식 (9)와 같이 계산된다.

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int_{\theta} f(y|\theta)\pi(\theta) d\theta} \propto f(y|\theta)\pi(\theta). \quad (9)$$

위에서 y 는 관찰값, $\pi(\theta)$ 는 θ 의 사전분포, $f(y|\theta)$ 는 우도함수 또는 결합밀도함수, $\int_{\theta} f(y|\theta)\pi(\theta) d\theta$ 는 y 의 주변분포, 그리고 $\pi(\theta|y)$ 는 θ 의 사후분포이다. 식 (9)로부터 베이즈 추론은 자료 및 모수에 대한 사전정보가 모수를 추정하는 과정에 반영된다는 것을 알 수 있다.

일반적으로 사후분포는 복잡한 형태의 다변량 분포이며 수리적으로 유도하기 어려운 경우가 많다. Metropolis, Rosenbluth, Rosenbluth, Teller와 Teller(1953)에 의해 처음으로 제안된 마코프 연쇄 몬테카를로(Markov chain Monte Carlo; MCMC) 방법은 변수가 많은 다차원 적분을 계산하기 위해서 개발되었다. 몬테카를로 적분은 주어진 분포에서 표본을 추출한 후에 표본들의 평균을 사용하여 기댓값의 근사치로 사용하며 엄청난 계산량이 요구된다. 컴퓨터의 발달로 인해 계산이 빨라지고 Geman과 Geman(1984)이 깃스샘플링(Gibbs sampling)을 제안하면서 MCMC 방법이 일반화되고 있다. 깃스샘플링 기법은 결합사후분포를 계산하기 위한 표본을 얻기 위해서 각 모수의 조건부 사후분포(conditional posterior distribution)로부터 무선 표본을 반복적으로 생성하는 표집방법이다. 예를 들어, 관심이 있는 모수가 $\theta_1, \theta_2, \theta_3$ 라면 각 모수의 조건부사후분포는 $\pi(\theta_1|y, \theta_2, \theta_3), \pi(\theta_2|y, \theta_1, \theta_3), \pi(\theta_3|y, \theta_1, \theta_2)$ 로

표현할 수 있다. 이러한 분포와 반복적인 과정(iterative process)을 통해서 결합분포를 한 단계씩 갱신(update)시켜 나가는 방법(모수 추정치 세트가 단계마다 계속 갱신됨)이 깃스샘플링기법이다. 이 무선 표본들의 극한분포는 적절한 조건하에서 결합사후밀도함수가 되며 정상분포(stationary distribution)의 성격을 갖게 된다.

깃스샘플링 알고리즘을 이용하여 본 연구에서 추정해야하는 3개의 모수벡터 β, γ, τ 에 대한 사후모수의 추정과정은 다음과 같다. 첫 번째 단계에서는 각각의 모수에 대한 초기값 $\theta^{(0)} = [\beta^{(0)}, \gamma^{(0)}, \tau^{(0)}]$ 을 정한다. 다음 단계에서는 각 모수의 초기값을 새로운 값으로 갱신하면서 차례대로 3개의 모수벡터 β, γ, τ 를 조건부사후분포로부터 추출한다⁵⁾. 이 단계를 충분히 많이(일반적으로 10,000번 이상) 반복하면, 마지막 m 번째 단계에서의 모수 추정치 $\theta^{(m)} = [\beta^{(m)}, \gamma^{(m)}, \tau^{(m)}]$ 은 근사적으로 결합사후분포 $\pi(\theta|y)$ 를 따르게 된다. 깃스샘플링 과정에서 초기에 수집된 표본은 추정치의 계산을 위해 사용하지 않고 버리는 것이 일반적인데 이 부분을 버린 구간(burn-in phase)이라고 한다. 깃스샘플링 과정에서 앞부분의 표본은 모수의 초기값에 크게 의존하는 속성이 나타나는데, 충분한 시간이 지나 안정된 후에는 초기값에 관계없이 조건부사후분포에 기초한 표본을 얻을 수 있다. 설명한 바와 같이 깃스샘플링기법은 결합분포를 이용한 다차원 문제를 조건부분포를 이용한 일차원 문제로 단순화시키는 장점이 있다. 결합분포는 복잡하다

5) ZIOP 모형의 각 모수 β (첫 번째 단계에서의 설명변수의 계수), γ (두 번째 단계에서의 설명변수의 계수), π (프로빗 모형의 경계값 모수)의 완전 조건부 사후분포는 부록1에 제시하였다.

라도 조건분포의 유도가 가능한 ZIOP 모형 등에 더욱 적절한 추정방법일 수 있다.

앞부분에서 베이지안 추론에 대해 설명한 바와 같이, ZIOP 모형의 모수를 베이지안 방법으로 추정하려면 모수벡터 β, γ, τ 의 사전분포($\pi(\beta), \pi(\gamma), \pi(\tau)$)를 정의해야 한다. 사전분포를 결정할 때 사후분포와 같은 형태로 정의하면 MCMC 기법을 용이하게 해주는 장점이 있다. 이러한 사전분포를 공액사전분포(conjugate prior distribution)라고 하는데 본 연구에서는 정규분포를 가정하였다. 여기서 각각의 모수 β, γ, τ 는 서로 독립이므로 결합사전분포(joint prior distribution)는 식 (10)과 같이 나타낼 수 있다.

$$\pi(\beta, \gamma, \tau) = \pi(\beta) \cdot \pi(\gamma) \cdot \pi(\tau). \quad (10)$$

결합사후분포는 우도함수와 결합사전분포를 곱한 것에 비례하는데, ZIOP 모형의 결합사후분포는 식 (11)과 같다.

$$\begin{aligned} \pi(\beta, \gamma, \tau | y) &\propto \left\{ \prod_{i=1}^n [\Phi(-x_i' \beta) + \Phi(x_i' \beta) \Phi(-z_i' \gamma)]^{I(y_i=0)} \right\} \\ &\cdot \left\{ \prod_{i=1}^n \prod_{j=1}^J [\Phi(x_i' \beta) (\Phi(\tau_j - z_i' \gamma)) - \Phi(\tau_{j-1} - z_i' \gamma)]^{I(y_i=j)} \right\} \\ &\cdot \pi(\beta) \cdot \pi(\gamma) \cdot \pi(\tau) \end{aligned} \quad (11)$$

모형의 적합성

Spiegelhalter, Best, Carlin과 Linde(2002)는 AIC(Akaike, 1973)와 BIC(Schwarz, 1978)에 상응하는 베이지안 적합도 기준으로 DIC(deviance information criterion)를 제시하였다. DIC는

MCMC를 통해서 추출된 사후표본을 이용하여 쉽게 계산될 수 있기 때문에 변수가 많아지고 모형이 복잡해지면 구하기 어려운 베이지안 상수(Bayes factor)보다 여러 모형에서 사용될 수 있다는 장점이 있다. DIC는 식(12)와 같다.

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D. \quad (12)$$

여기서 \bar{D} 는 모형의 적합도, p_D 는 유효한 모수의 수를 의미한다(Spiegelhalter, 2002). DIC는 AIC, BIC와 마찬가지로 비교하는 두 모형 중에 DIC 추정값이 더 작은 모형이 자료를 더욱 잘 설명한다고 해석한다.

적응행동 척도 자료에의 적용

연구자료

본 연구에서는 2013년 11월부터 1년간 진행된 한국형 베일리-III 영유아검사 표준화 2차 예비연구 중 적응행동영역 자료를⁶⁾를 사용하였다. 이 자료는 부모 또는 주양육자가 생후 16주에서 42개월에 해당하는 정상 영유아의 적응행동 정도 질문에 응답한 것이다. 총 900명 중에서 설명변수의 반응값이 결측된 14명을 제외한 886명의 반응자료가 분석에 사용되었다. 특히 ZIOP 모형을 적용하기 위해서 적응

6) 2006년에 개발된 베일리-III 영유아 검사는 인지, 언어, 운동, 사회-정서, 적응행동에 해당되는 5가지 영역을 종합적으로 측정한다. 한국형 베일리-III 영유아 발달검사 표준화 2차년도 연구를 수도권, 강원, 충청지역을 대상으로 표준화 연구를 진행하였으며, 적응행동 척도 자료는 김호정(2015)의 자료와 동일하다.

표 1. 적응행동 척도의 의사소통 문항들

문항번호	질문 내용
문20	앞서 일어난 일을 이야기할 때 과거 시제를 사용한다(예, 밥 먹었어).
문21	한 가지 주제를 가지고 3분 이상 이야기를 나눈다.
문22	적절하게 대화를 끝낸다(예, 계속해서 질문을 하지 않는다. 대화를 끝내려고 할 때 계속 말하려는 행동을 보이지 않는다).
문23	다른 사람이 이야기 할 때 부적절하게 끼어들지 않는다.
문24	전화번호를 말한다(예, 집 전화번호, 부모님 전화번호 등).

행동 척도의 하위영역인 의사소통 문항반응자료 중에서 영의 응답비율이 높은 5개 문항을 임의로 선택하였다(표 1). 인구통계학적 특성 변수로는 성별(0=여아, 1=남아), 모(母)교육수준, 가계수입, 월(月)령 변수를 사용하였다. 사용한 4개의 설명변수는 영유아 발달의 중요한 측면들을 잘 설명해 줄 수 있는 주요변수들이다(Bayley, 2006). 베이지안 추정과정은 R 3.1.3 (R core team, 2015)을 이용하였다. 김스알고리즘, 모형적합도 기준값 생성 등을 R 언어로 직접 프로그래밍하여 분석결과를 얻었다⁷⁾.

기술통계

각 문항에 대한 응답의 빈도와 분포는 각각 표 2와 그림 1에 제공된다. 문항 20 ~ 문항 23은 영 응답비율이 약 70%이고 문항 24는 영 비율이 92%로 나타나는 등 5개 문항 모두 뚜렷하게 영과잉 자료의 특성을 보이고 있다.

사후모수의 추정과 모형진단

본 연구에서는 1단계 이분형 프로빗 모형에서 절대적 비적응행동 집단과 잠재적 적응행동

동 집단의 구분에 영향을 주는 설명변수 x 와 2단계 순서형 프로빗 모형에서 적응행동 집단의 적응행동 정도에 영향을 주는 설명변수 z 로서 성별, 모교육수준, 가계수입, 월령변수를 동일하게 사용하였다. 연구목적에 따라 x 와 z 는 서로 얼마든지 다르게 선택할 수 있다.

김스샘플링기법을 이용하여 사후모수를 추정하기 위한 각 모수의 초기값은 다음과 같이 설정하였다. 먼저 절대적 비적응행동 집단과 잠재적 적응행동 집단을 식별하는 이항변수 S 는 반응변수의 관찰값이 영보다 큰 경우에 모두 1을 부여하고, 관찰값이 영인 경우에는 베르누이분포로 난수를 생성하여 임의로 절반은 1을 주고 나머지 절반은 0을 주었다. 다시 말해서, 적응행동이 이미 나타난 영유아집단은 S 의 초기값으로 1을 설정하고, 적응행동이 나타나지 않았던 영유아들 중 50%는 잠재적으로 적응행동이 나타날 가능성이 있는 영유아 집단(잠재적 적응행동 집단)으로 보고 나머지 50%는 적응행동이 절대로 나타나지 않는 집단(절대적 비적응행동 집단)으로 본 것이다. β 의 사전평균과 사전분산 초기값은 이분변수 S 를 반응변수로 두고 프로빗 모형을 적합시킨 결과에서 얻은 추정치로 설정하였다.

γ 의 사전평균과 사전분산 초기값은 적응행

7) 부록2에 R코드를 제시하였다.

표 2. 문항빈도분석결과

반응값	문20		문21		문22		문23		문24	
	N	%	N	%	N	%	N	%	N	%
0	628	70.9	638	72.0	618	69.8	627	70.8	815	92.0
1	21	2.4	48	5.4	31	3.5	44	5.0	26	2.9
2	56	6.3	108	12.2	130	14.7	177	20.0	13	1.5
3	181	20.4	92	10.4	107	12.1	38	4.3	32	3.6
총합	886	100.0	886	100.0	886	100.0	886	100.0	886	100.0

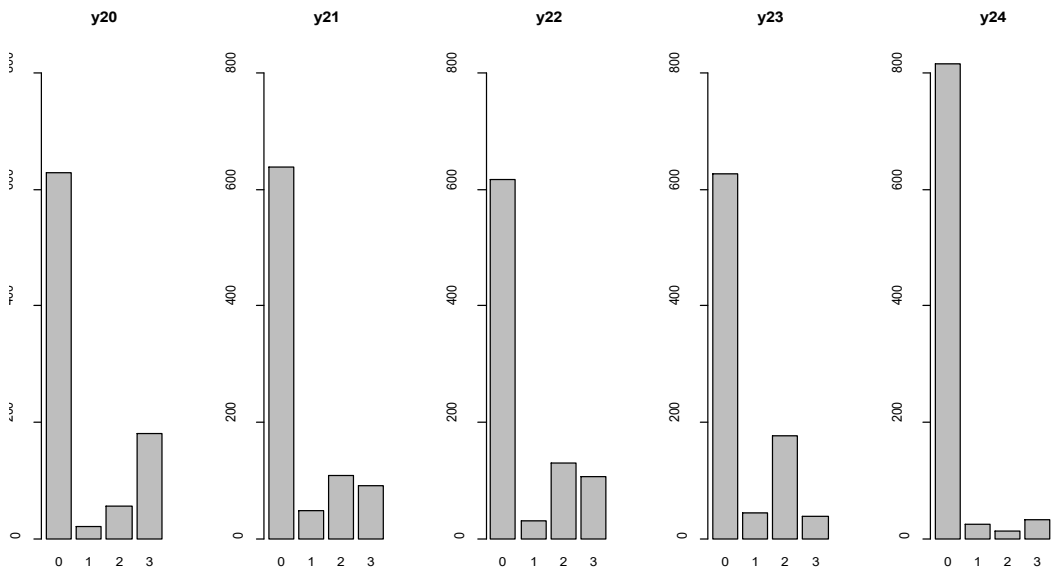


그림 1. 문항응답분포

동이 나타난 영유아 집단 자료와 잠재적 적응 행동 집단으로 분류된 영유아 집단(즉, $S = 1$) 자료에 순서형 프로빗 모형을 적용한 결과로 얻은 추정치로 설정하였다.⁸⁾

다음으로 깃샘플링알고리즘을 이용하여 총 110,000개의 난수 표본을 생성하였고, 안정

된 추정치를 취하기 위하여 초기 변인 구간의 표본 10,000개를 제거하였다. 또한 나머지 표본 중에서 자기상관이 최대한 나타나지 않도록 50의 배수가 되는 표본만을 선택하였고 결과적으로 약 2,000개의 사후표본을 베이지안 추론에 사용하였다. 사후표본이 안정적으로 추출되었는지 진단하는 방법으로 5개 문항에 대하여 모수벡터 β , γ , τ 의 주변사후분포, 자기상관함수그림(auto-correlation function plot)⁹⁾,

8) 이와 같이 자료로부터 얻은 추정치를 사전분포로 사용하는 방식을 경험적 베이지안 추정(empirical Bayesian estimation)이라고 한다.

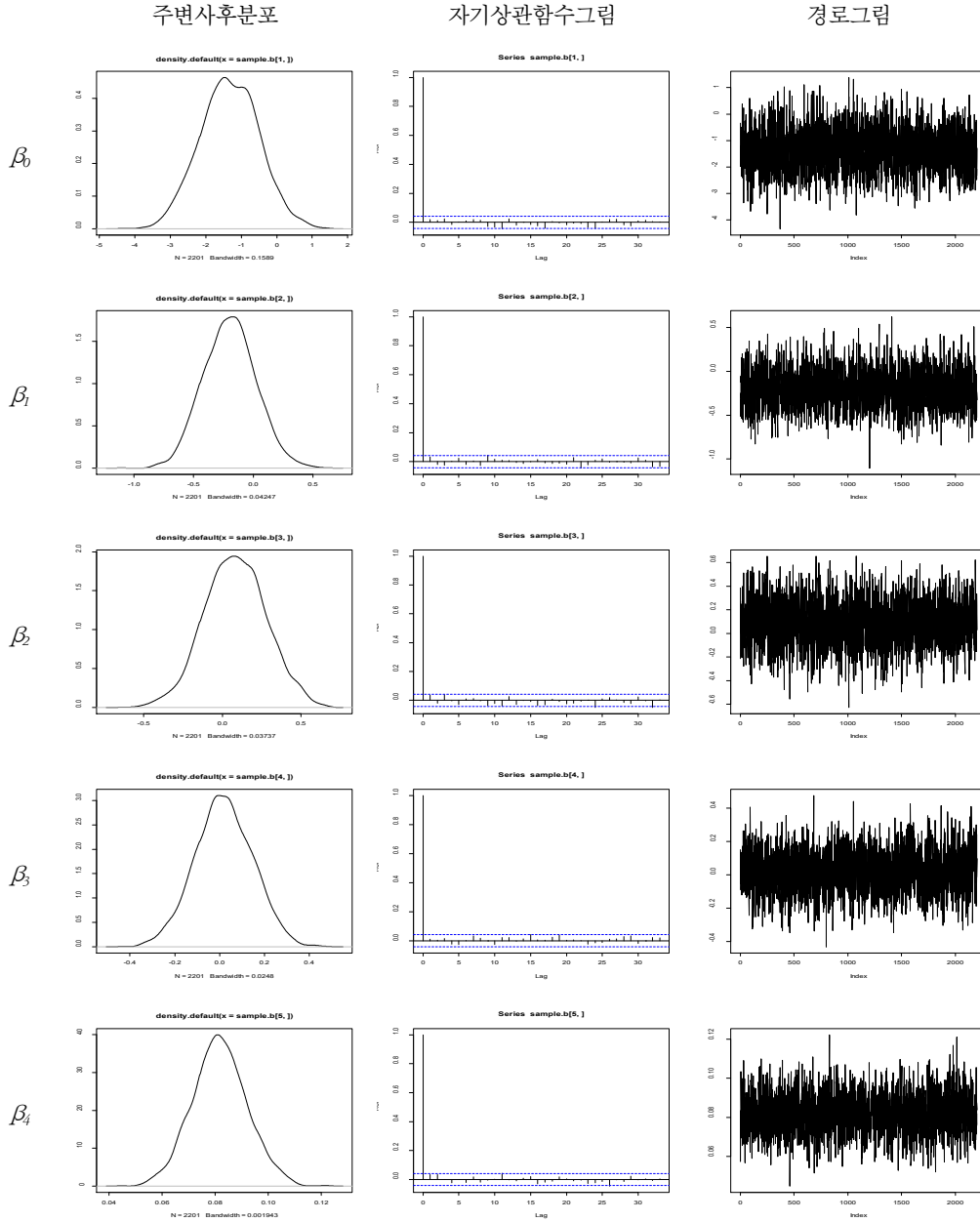


그림 2. β 의 주변사후분포, 자기상관함수그림, 경로그림

9) 자기상관함수는 현재의 시점과 이전의 시점의 값의 상관을 나타내는 함수이며, MCMC 방법으로 베이저안 추론을 하는 경우에는 마코프 연쇄 내의 상관관계를 확인하는데 주로 사용된다.

경로그림을 살펴보았다. 그림 2는 문항 20의 $\beta' = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]$ 의 주변사후분포, 자기상관함수 그림, 경로그림을 정리한 것이

다. 자기상관함수 그림은 급격하게 감소하는 형태를 보였고, 경로그림도 추정값의 범위를 크게 벗어나지 않았기 때문에 수렴이 잘 된 것으로 판단할 수 있다. 추출된 사후표본의 추정치들은 수렴이 잘되었으며, 나머지 문항들에 대한 β , γ , τ 의 사후표본도 대체적으로 수렴이 잘 되었음을 확인하였다.

베이지안 신뢰구간과 주변효과

표 3은 ZIOP 모형의 모수 β , γ , τ 의 사전 분포와 사후분포의 모수추정치, 표준오차, 베이지안 신뢰구간¹⁰⁾을 보여준다. 예를 들어, 문항 20을 이용한 분석에서 월령 계수(β_4)의 95% 베이지안 신뢰구간은 [.066, .099]이다. 95% 신뢰구간 내에 영이 포함되지 않았기 때문에 유의수준 5%에서 영가설($H_0: \beta_4^{\text{문항}20} = 0$)은 기각된다. 이러한 방법으로 모든 문항 내 모수들의 신뢰구간을 살펴보면 월령이 전체 문항에서 유의한 설명변수임을 확인할 수 있다. 따라서 영유아 시기의 월령은 영유아가 적응행동을 보여줄 가능성을 식별하는 변수이면서 동시에 적응행동이 얼마만큼의 강도로

나타나는지에 대한 반응정도와의 정적상관을 보이는 변수로 나타났다. 나머지 문항에서도 통계적으로 유의한 설명변수들을 비슷한 방식으로 해석할 수 있다.

표 4는 OP모형과 ZIOP모형의 결과를 고전적인 추정에 의한 모형적합도 기준인 AIC, BIC와 베이지안 추정에 의한 모형적합도 기준인 DIC로 구분하여 제시한 것이다. 영비율이 70%정도인 4개의 문항(문항 20 - 문항 23)에 대해서는 모든 기준에서 ZIOP모형이 OP모형보다 자료를 더욱 잘 설명하는 결과가 나타났다. 그러나 영비율이 90%인 24번 문항에서 AIC와 BIC는 반대의 결과가 나타났고, DIC에서는 ZIOP모형만 추정값이 산출되었다.

표 5는 순서형 반응범주별로 주변효과를 정리한 것이다. 먼저 비적응행동 영유아 집단(즉, $\Pr(y = 0)$)에 대한 성별의 주변효과를 살펴보면 남아가 여아보다 비적응행동을 보일 확률이 3.7% ~ 7.5% 높게 나타났고, 모교육수준, 가계수입, 월령이 높을수록 비적응행동을 보일 확률이 낮게 나타났다. 20번 문항에서 모교육수준이 높은 영유아가 모교육수준이 낮은 영유아보다 비적응행동을 보일 확률이 3.4% 낮게 나왔고 가계수입이 높은 영유아가 가계수입이 낮은 영유아보다 비적응행동을 보일 확률이 1.4% 낮게 나왔으며 월령이 높은 영유아가 월령이 낮은 영유아보다 비적응행동을 보일 확률이 3.1% 낮게 나왔다. 문항 21 ~ 문항 23에서도 비적응행동에 대한 모교육수준, 가계수입, 월령의 주변효과는 일관된 결과를 보였다. 비적응행동 집단의 세부집단인 절대적 비적응행동 집단과 잠재적 적응행동 집단에 대한 해석 역시 같은 방식으로 설명변수의 효과를 이해할 수 있다.

10) 베이지안 추론에서의 신뢰구간의 해석은 고전적 추론에서의 신뢰구간과는 차이가 있다. 고전적 추론에서 모수에 대한 95% 신뢰구간(confidence interval)의 해석은 표본을 반복하여 여러 번 추출하여 얻어진 다수의 신뢰구간들 중 약 95%가 모수를 포함한다고 해석한다. 이에 반해 신용구간(credibility interval)으로도 알려진 베이지안 신뢰구간은 하나의 구간이 바로 그 모수를 포함할 확률이 95%라고 해석한다. 국내의 베이지안 연구논문에서 여전히 많은 연구자들이 신뢰구간이라는 용어를 사용하고 있기 때문에, 본문에서는 국내 연구자들에게 친숙한 신뢰구간으로 용어를 사용하였다.

표 3. ZIOP 모형의 모수 β , γ 의 사전/사후 추정치, 표준오차, 95% 베이저안 신뢰구간

			사전분포의 모수추정		사후분포의 모수추정			
			추정치	표준오차	추정치	표준오차	95% 신뢰구간	
							2.5%	97.5%
문20	상수항	β_0	-0.823	1.267	-1.797	.815	-3.139	-.456
	성별	β_1	-.015	-.040	-.153	.210	-.499	.192
	모교육수준	β_2	.092	-.287	.189	.200	-.140	.519
	가계수입	β_3	.007	-.015	.018	.124	-.186	.223
	월령*	β_4	.049	-.003	.083	.010	.066	.099
	상수항	γ_0	-2.168	.978	-2.584	.429	-3.290	-1.878
	성별*	γ_1	-.225	-.048	-.194	.106	-.369	-.019
	모교육수준	γ_2	-.041	-.187	.083	.099	-.080	.247
	가계수입	γ_3	.068	-.014	.063	.063	-.041	.169
	월령*	γ_4	.090	-.005	.080	.006	.070	.090
문21	상수항	β_0	-1.035	1.271	-1.228	.794	-2.535	.078
	성별	β_1	-.073	.084	-.253	.216	-.610	.103
	모교육수준	β_2	.113	.082	.082	.201	-.247	.413
	가계수입	β_3	-.164	.029	.015	.124	-.188	.219
	월령*	β_4	.086	.001	.078	.010	.061	.095
	상수항	γ_0	-3.852	1.220	-3.001	.430	-3.710	-2.293
	성별	γ_1	-.202	.075	-.137	.112	-.031	.046
	모교육수준	γ_2	.256	.069	.137	.102	-.031	.306
	가계수입	γ_3	.078	.025	.079	.061	-.022	.181
	월령*	γ_4	.098	.001	.083	.006	.073	.094
문22	상수항	β_0	-1.917	1.315	-1.957	.854	-3.363	-.552
	성별	β_1	.123	.086	-.153	.217	-.510	.204
	모교육수준	β_2	.207	.084	.171	.204	-.165	.507
	가계수입	β_3	.019	.030	.037	.128	-.173	.248

표 3. ZIOP 모형의 모수 β , γ 의 사전/사후 추정치, 표준오차, 95% 베이지안 신뢰구간 (계속)

		사전분포의 모수추정		사후분포의 모수추정		95% 신뢰구간		
		추정치	표준오차	추정치	표준오차	95% 신뢰구간		
						2.5%	97.5%	
	월령*	β_4	.092	.001	.091	.010	.074	.108
	상수항	γ_0	-2.440	1.084	-2.442	.450	-3.183	-1.701
	성별	γ_1	-.048	.066	-.151	.010	-.325	.021
	모교육수준	γ_2	.022	.062	.117	.102	-.050	.286
	가계수입	γ_3	.124	.021	.044	.061	-.057	.145
	월령*	γ_4	.078	.001	.073	.001	.062	.084
문23	상수항	β_0	-1.083	1.287	-1.221	.908	-2.171	.272
	성별	β_1	-.127	.085	-.090	.217	-.446	.266
	모교육수준	β_2	-.155	.083	.028	.231	-.352	.408
	가계수입	β_3	.384	.295	.026	.127	-.183	.237
	월령*	β_4	.067	.001	.083	.010	.066	.100
	상수항	γ_0	-3.169	1.306	-2.709	.459	-3.465	-1.953
	성별	γ_1	.101	.078	-.155	.112	-.340	.029
	모교육수준	γ_2	.179	.075	.160	.113	-.026	.347
	가계수입	γ_3	.057	.026	.018	.625	-.084	.120
	월령*	γ_4	.085	.001	.078	.006	.068	.089
문24	상수항	β_0	.651	1.093	2.489	.651	1.417	3.561
	성별	β_1	-.232	.072	-.133	.175	-.421	.154
	모교육수준*	β_2	-.616	.070	-.890	.161	-1.156	-.623
	가계수입*	β_3	.520	.025	.399	.109	.218	.580
	월령	β_4	.015	.001	.003	.007	-.008	.015
	상수항	γ_0	-34.919	5.799	-35.291	.971	-36.889	-33.693
	성별	γ_1	.565	.288	.320	.223	-.046	.687
	모교육수준*	γ_2	6.516	.290	6.472	.237	6.081	6.862
	가계수입*	γ_3	-.488	.100	-.631	.149	-.877	-.385
	월령*	γ_4	.301	.001	.332	.015	.307	.357

주. *는 95% 베이지안 신뢰구간에서 유의한 설명변수임.

표 4. OP모형과 ZIOP모형의 적합도 비교

	고전적인 추정 기준				베이저안 추정 기준	
	AIC		BIC		DIC	
	OP	ZIOP	OP	ZIOP	OP	ZIOP
문20	767.56	679.82*	787.04	737.69*	834.26	692.22*
문21	927.83	920.60*	985.69	978.47*	979.67	500.28*
문22	967.71	783.52*	1025.57	841.38*	1054.03	579.73*
문23	904.85	876.35*	962.72	934.22*	921.33	513.09*
문24	467.89*	912.90	525.75*	970.77	N/A (추정불능)	1098.92

* 각 기준 내에서 OP모형과 ZIOP모형 중에 더 좋은 모형임.

표 5. ZIOP 모형의 주변효과

		Pr (y = 0)						Pr (y = 1)	Pr (y = 2)	Pr (y = 3)			
		절대적 비적응		잠재적 적응		합계							
		M	SD	M	SD	M	SD						
문20	성별	.055	.023	.020	.022	.075	.014	-.001	.001	-.003	.001	-.005	.004
	모교육수준	-.023	.009	-.010	.011	-.034	.007	.001	.001	.001	.001	.002	.001
	가계수입	-.005	.002	-.009	.007	-.014	.006	.001	.001	.001	.001	.001	.001
	월령	-.022	.009	-.008	.009	-.031	.006	.001	.001	.001	.001	.002	.001
문21	성별	.060	.024	.010	.016	.071	.013	-.003	.001	-.004	.003	-.002	.002
	모교육수준	-.038	.015	-.013	.015	-.052	.009	.001	.001	.002	.002	.001	.001
	가계수입	-.008	.003	-.010	.009	-.019	.007	.001	.001	.001	.001	.001	.001
	월령	-.022	.009	-.008	.009	-.031	.006	.001	.001	.001	.001	.001	.001
문22	성별	.035	.016	.018	.019	.053	.014	-.001	.001	-.004	.002	-.001	.001
	모교육수준	-.046	.021	-.008	.014	-.055	.013	.002	.001	.005	.003	.002	.001
	가계수입	-.009	.004	-.005	.005	-.014	.004	.001	.001	.001	.001	.001	.001
	월령	-.024	.011	-.006	.009	-.031	.007	.001	.001	.002	.001	.001	.001
문23	성별	.027	.011	.019	.018	.047	.013	-.001	.001	-.003	.002	-.001	.001
	모교육수준	-.047	.019	-.010	.014	-.057	.011	.002	.001	.005	.004	.001	.001
	가계수입	-.010	.004	-.000	.002	-.010	.002	.001	.001	.001	.001	.001	.001
	월령	-.023	.009	-.007	.009	-.031	.006	.001	.001	.002	.002	.001	.001
문24	성별	.042	.006	-.005	.013	.037	.014	-.001	.001	-.001	.001	-.002	.005
	모교육수준	.327	.051	-.129	.329	.198	.329	-.005	.013	-.003	.009	-.015	.040
	가계수입	-.139	.021	.016	.041	-.122	.045	-.002	.005	.001	.004	.006	.017
	월령	-.001	.001	-.005	.015	-.006	.015	.001	.001	.001	.003	.001	.001

결론 및 논의

본 연구에서는 문항반응 자료가 순위범주형이며 영과잉 특성을 나타낼 때, ZIOP 모형의 베이지안 추론 기법을 제시하고, 이것을 한국형 베일리-III 영유아 검사 적응행동척도 자료에 적용하여 활용가능성을 살펴보았다. 본 연구의 중요성과 의의는 다음과 같다.

첫째, 반응변수가 행동의 변별수준을 측정하는 순위범주형 자료인 경우에는 범주적 특성을 고려한 자료분석 방법을 도입하는 것이 결과의 왜곡을 줄일 수 있다는 측정학적 중요성을 강조한다. 측정도구 문항의 반응단계 수는 개발과정에서의 신뢰도와 타당도 분석은 물론 측정도구를 활용한 후속연구(예, 측정동일성, 매개효과, 조절효과 검증)에서 수행되는 자료분석과도 밀접한 관계가 있다. 리커트 응답반응에서의 적절한 단계 수는 구인, 문맥, 모집단에 따라 결정될 수 있는데, 일반적으로 5단계 이상이면서 11단계를 넘지 않을 것을 제안하고 있다(McCoach, Gable, & Madura, 2013). 왜냐하면 단계 수가 너무 적으면 변별도가 흐려지는 반면, 너무 많으면 측정오차를 초래할 수 있기 때문이다. 그런데 여기서 중요하게 짚고 넘어가야 하는 문제는 반응포맷이 기본적으로 범주형의 성질이 있다는 것이다. 척도개발자가 태도 또는 정의적 성격을 연속체(continuum)로 정의할지라도 연구자가 도구를 사용하는 순간 연속성 가정은 위배된다(Bollen, 1989). 이러한 맥락에서 단계의 수를 결정하여 측정하는 리커트 척도와 연속형 변수의 정규성을 가정한 통계모형의 수리적인 속성을 고려해볼 때, 반응포맷이 기본적으로 범주형이라면 범주형 자료분석 기법으로 다뤄져야 한다(Wirth & Edwards, 2007). 범주형 자

료에 대한 추정방법은 효율적이면서도 접근이 용이한 소프트웨어들이 다양하게 소개되고 있다(예, Mplus; Muthen & Muthen, 1998-2016).

둘째, 반응변수가 순위범주형으로 측정되고 영이 과도하게 포함되어 있는 베일리-III 영유아 검사 적응행동척도 자료에 ZIOP 모형을 직접 적용하고 활용가능성을 제시한 것이다. 영비율이 높은 순위범주형 자료에 OP모형과 ZIOP모형을 적용해 본 결과 ZIOP모형이 고전적인 방법과 베이지안 방법의 모형적합도 기준에서 OP모형보다 좋은 결과를 보였다. 또한 적응행동 검사자료에 활용 가능성을 확인해 본 결과, 적응행동에 영향을 주는 설명변수들이 선행연구결과와 대체적으로 일치하였으며(이복희, 2001; 유정희, 2008) 영 응답 반응을 의미상으로 두 개의 집단으로 나누어 주변효과를 살펴본 결과도 일반 연구자들에게 의미 있는 정보를 제공함을 확인하였다. 적응행동 관련 검사들은 주로 특수교육 또는 상담장면에서 다양한 진단과 평가를 위해 사용되고 있다. 적응행동의 결함은 정신지체와 발달 장애 영역에서 중요한 함의점이 있다. 본 연구에서 절대적 비적응행동 영유아집단에는 잠재적 정신지체와 발달장애가 나타날 가능성이 있는 영유아가 포함될 수 있고, 잠재적 적응행동 영유아 집단에는 검사 시점에는 적응행동 문항에 0 범주로 응답하였지만 시간이 흐르면 적응행동이 나타나서 비장애영유아로 분류될 수 있는 가능성을 가정하였다. 적응행동과 관련된 연구가 증가하였음에도 불구하고 여전히 구성개념에 대한 의견의 일치는 이루어지지 않고 있지만, 측정하고자 하는 구인은 주로 실제적 기술, 개념적 기술, 사회적 기술 등을 측정하기 때문에 기술을 수행하는 정도(예를 들면 0은 거의 안함, 1은 때때로, 2는 항상)로

반응값을 얻는다(이준석, 유재연, 신현기, 전병운, 고등영, 2007; 김호정, 2015, Harrison & Oakland, 2003). 따라서 이러한 모형으로 영유아 적응행동에 영향을 주는 다양한 예측변인들을 탐색하는 후속연구가 가능할 것이라고 생각한다.

셋째, PsycINFO에서 검색된 5편의 ZIOP 모형 응용논문들이 고전적 추론방법으로만 접근한 반면, 본 연구에서는 고전적 추론방법과 베이저안 추론으로 접근하여 그 결과를 비교하고 소표본에서의 베이저안 추정의 활용가능성을 탐색한 것도 큰 의의가 있다고 할 수 있다. Ghosh 외(2006)는 1수준 ZIP모형을 이용하여 모의실험을 통해 사례수가 작고 영비율이 극단적으로 높은 경우에는 베이저안 추정이 고전적인 추정방법보다 95% 신뢰구간을 더 좁게 추정함으로써 베이저안 방법이 양호함을 주장하였다. 한편, Stegmueller(2013)는 다층모형(multilevel model)에서 적절한 집단 수를 제안하기 위해서 모의실험으로 고전적인 추정방법과 베이저안 추정방법을 비교하였다. 그 결과 집단의 수가 작고(5 - 30) 층위간 상호작용(cross-level interaction)이 포함된 다층모형의 경우 고전적인 추정방법에 의한 추정치와 신뢰구간이 심하게 왜곡될 수 있는 반면, 베이저안 접근은 훨씬 강건하고 안정적임을 보였다. 이와 같이 선행연구들을 살펴보면 베이저안 추정은 극단적인 값이 나타나는 자료구조 또는 복잡한 모형에서 고전적인 추정정보보다 더 잘 작동하는 것으로 해석할 수 있다. 본 연구에서도 고전적인 추정에 의한 모형적합도 결과를 살펴보면, 선택한 5개의 문항 중 영비율이 90%이상인 24번 문항은 영비율이 약 70%인 나머지 4개의 문항들과는 다르게 오히려 OP 모형이 더 좋다는 결과를 보였다. 그러나 베

이지안 추정에 의한 모형적합도 결과에서 OP 모형의 DIC는 추정이 불가능¹¹⁾하였지만 ZIOP 모형의 DIC는 추정되었다. 이러한 결과는 24번 문항의 실제 사례 수 886명 중 의미 있는 반응값(1, 2, 3으로 응답한 경우)의 사례 수는 단지 71명만 존재하는 상황에서 나타난 경우이다. 이는 Ghosh 외(2006)의 결과와 마찬가지로, 극단적인 영비율이 있는 소표본 자료에 ZIOP모형을 적용할 때 고전적인 추정방법으로는 왜곡된 결과를 나타낼 수 있는 가능성이 있음을 함의한다.

본 연구가 심리학이나 교육학 자료의 분석에 새로운 방법론적 틀을 제시한 것은 큰 의의라고 할 수도 있지만, 여느 연구와 마찬가지로 제한점이 존재한다. 이러한 제한점을 밝히고, 이 제한점과 더불어 후속 연구를 위한 제언을 하고자 한다. 첫째, ZIOP 모형은 개발된 지 10년이 안 되는 새로운 모형이고, 일반 연구자들이 쉽게 접근할 수 있는 통계 소프트웨어는 현재시점까지는 개발되지 않은 실정이다. 따라서 적응행동 자료에 대한 ZIOP 모형의 모수추정, OP모형과 ZIOP모형의 적합도 기준(AIC, BIC, DIC)추정은 통계분석 모듈 프로그래밍 방법에 대한 이론(Rizzo, 2007)에 기초하여 R에 내장된 통계함수와 R 언어로 알고리즘을 구성하여 분석을 실시하였다. 베이저안 방법은 고전적 방법과는 달리 MCMC 방법을 사용하여 가능한 많은 수의 안정적인 사후표본을 생성해야 하며 이를 위해 처음 생성된 표본은 사용하지 않고 버린다. 이러한 분석과정에서 사전분포의 선택, 초기값 설정, 알고리

11) 영비율이 70%인 4개 문항들의 절편(γ_0)의 사후 추정치는 '-4 ~ -1' 범위내에서 추정된 반면, 영비율이 90%인 24번 문항의 절편(γ_0)의 사후추정치는 -35로 매우 작게 추정되었다.

증 반복 수 등을 결정해야하며, 모형의 복잡도에 따라 성능이 좋은 컴퓨터와 오랜 추정시간이 요구된다. 하지만 Bugs, WinBugs, OpenBugs, 그리고 최근에는 R과 함께 구동하는 JAGS(just another Gibbs sampling) 등 다양한 베이지안 추정 소프트웨어가 제공되고 있으므로 일반연구자들도 시도해 볼 만한 가치가 있다고 생각된다.

둘째, 영과잉 문항의 영비율이 90% 이상인 경우, 70% 정도인 다른 문항들의 설명변수들에서 나타난 양상과 다른 결과를 보여주었다는 점이다. 이러한 결과가 전체 표본 수에 비해 관찰된 표본수가 매우 작기 때문인지 또는 측정하고자 하는 의사소통의 구인과 다르게 기능하는 문항이기 때문인지 추후 연구를 통해 밝혀야만 할 것이다. 전자의 문제라면 베이지안 추정의 사후분포를 형성하기 위해 전체 깁스샘플링 표본 중에서 어떤 표본들을 의미 있게 추출해야 하는지 컴퓨터 시뮬레이션 연구를 통해 확인할 수 있을 것이다. 후자의 문제라면 베일리-III 척도 자체에 대한 더욱더 심도 있는 연구가 필요할 것이다.

이러한 제한점에도 불구하고 본 연구에서 사용된 ZIOP 모형은 순위범주형 자료에서 영의 비율이 과도하게 높은 경우에 설명변수의 영향을 제대로 파악하기 위해 적용할 수 있는 의미 있는 모형이다. 이러한 영 과잉 특성 자료는 교육장면, 임상장면, 산업장면에서도 얼마든지 발생할 수 있다. 예를 들어, 교육장면에서는 정인지체아나 영재아 또는 낙제아를 진단하는 경우가 있을 수 있고, 임상장면에서는 인지능력(예. 치매선별검사), 음주·마약·도박·인터넷 등의 다양한 중독의 정도를 진단하고자 하는 경우가 있으며, 산업장면에서는 소비행동, 인사선발과 분류를 목적으로 진

단하고자 하는 경우가 있다. 연구자의 관심 있는 종속변수가 순위범주형이면서 과도하게 영 값이 포함되어 있고 또한 영 값이 의미상으로 두 가지 상황에서 생성되었다고 가정할 수 있다면, 선형회귀모형을 포함한 일반적인 회귀모형은 오차에 대한 가정이 위배되어 왜곡된 결과를 나타내며 적절한 모형이라 할 수 없다. 이러한 자료는 일반적인 회귀모형의 한계점을 극복하고 순서형 프로빗 모형을 확장한 ZIOP 모형을 적용하여 문항반응을 설명하는 요인들을 규명할 수 있을 것이다. 또한 문제행동이나 적응행동 등이 나타날 가능성에 대한 설명변수(x)와 행동의 정도에 대한 설명변수(z)를 다르게 사용할 수 있기 때문에 ZIOP 모형은 순서형 척도의 자료분석에 큰 장점이 있다.

후 주

1. 본 연구를 진행하는 과정에서 많은 도움을 주신 이화여대 심리학과 명예교수 김아영 교수님과 이화여대 통계학과 오만숙 교수님께 감사드립니다.
2. 본 연구가 발전될 수 있도록 아낌없는 조언을 해주신 한국측정평가심리학회 토론회, 질문자와 3분의 심사위원에게 감사드립니다.

참고문헌

- 김준엽, 신혜숙 (2009). 다층 영과잉 포아송 모형의 적용: 청소년 비행의 발생여부 및 발생빈도 분석, *교육평가연구*, 22(2), 471-493.

- 김호정 (2015). 베일리-III 검사를 사용한 미국과 한국 영유아 적응행동 비교 연구, 이화여자 대학교 박사학위논문.
- 유정희 (2008). 생활자립훈련중심 전환교육활동이 정신지체학생의 적응행동기술에 미치는 효과. 대구대학교 대학원 박사학위 논문.
- 이복희 (2001). 신체활동중심의 치료레크리에이션 활동이 정신지체아동의 적응행동 및 정서적 행동에 미치는 효과. 한국특수체육 학회지, 9(2), 73-87.
- 이준석, 유재연, 신현기, 전병운, & 고등영. (2007). 초등학생용 적응행동검사의 실제적 기술 영역 문항 타당성 검토. 특수교육 학연구, 41, 37-53.
- 홍성두, 여승수, 김남순, 박상희, 이선재 (2009). 지적장애 연구를 위한 영과잉 모형의 활용방안 탐색. 지적장애연구, 11(4), 179-197.
- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 255-265.
- Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., & Neighbors, C. (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors*, 27(1), 166-177.
- Bayley, N. (2006). *Bayley scales of infant and toddler development-third edition: Technical manual*. San Antonio, TX: Harcourt Assessment.
- Bollen, K. A. (1989). *Structural equation models with latent variables*. New York: Wiley.
- Ghosh, S. K., Mukhopadhyay, P., & Lu, J. C. J. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, 136(4), 1360-1375.
- Gurmu, S., & Dagne, G. A. (2012). Bayesian approach to zero-inflated bivariate ordered probit regression model, with an application to tobacco use. *Journal of Probability and Statistics*, 2012, 1-26.
- Gurmu, S., & Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics*, 14(4), 469-477.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721-741.
- Harris, M. N., & Zhao, X. (2007). A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics*, 141(2), 1073-1099.
- Harrison, P., & Oakland, T. (2003). Adaptive behavior assessment system (ABAS-II). *San Antonio, TX: The Psychological Corporation*.
- Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. New York, NY: Springer.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087-1092.
- Musio, M., Sauleau, E. A., & Buemi, A. (2010). Bayesian semi-parametric ZIP models with space-time interactions: an application to cancer registry data. *Mathematical Medicine and*

- Biology*, 27(2), 181-194.
- Muthén, L. K., & Muthén, B. O. (1998-2016). *Mplus User's Guide* 7.0. Los Angeles, CA; Muthén & Muthén.
- Neelon, B. H., O'Malley, A. J., & Normand, S.-L. T. (2010). A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling*, 10(4), 421-439.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society A*, 135, 370-384.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- Rizzo, M. L. (2007). *Statistical computing with R*. CRC Press.
- Singh, S. (1963). A note on inflated Poisson distribution. *Journal of the Indian Statistical Association*, 1, 140-144.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.
- Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, 57(3), 748-761.
- Wang, P. (2003). A bivariate zero-inflated negative binomial regression model for count data with excess zeros. *Economics Letters*, 78(3), 373-378.
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12(1), 58-79.
- Wright, A. G., Pincus, A. L., & Lenzenweger, M. F. (2012). An empirical examination of distributional assumptions underlying the relationship between personality disorder symptoms and personality traits. *Journal of Abnormal Psychology*, 121(3), 699-706.
- 1차원고접수 : 2016. 10. 11.
수정원고접수 : 2017. 02. 27.
최종게재결정 : 2017. 04. 05.

Bayesian Analysis of KBSID-III Adaptive Behavior Data Using a Zero-Inflated Ordered Probit Model

Hye-Eun Seok

Hee-Jung Bang

Su-Young Kim

Ewha Womans University

Excessive zeros are frequently observed in response variables when behavioral characteristics in the development of children are assessed. For example, in the Korean Bayley Scales of Infant and Toddler Development-Third Edition (KBSID-III)-adaptive behavior test, zero scale was excessively recorded more than other scales, such as 1, 2, or 3. A regular ordered probit (OP) model can be used when more than two outcomes appear in ordinal dependent variables. However, it is not appropriate for an OP model to be used with zero-inflated ordinal data. An OP model also has a limitation when there are two semantically distinctive groups, genuine non-participant and potential participant groups. We applied a two-step zero-inflated ordered probit (ZIOP) model in a Bayesian framework to the KBSID-III-adaptive behavior data. In the first step, the adaptive group (potential adaptive group was included) was separated from the genuine non-adaptive group using a probit model. In the second step, an OP model was applied to the adaptive group. A Bayesian estimation procedure to the ZIOP model was carried out with a Gibbs sampling algorithm using the open-source software R. The utility of the ZIOP model with zero-inflated ordered categorical variables was verified by checking the marginal effect of predictors on the change in the probability of a certain category.

Key words : Zero-inflated, Ordered probit model, Bayesian inference, MCMC, Gibbs sampling, Bayley-III

부 록

부록1: 각 모수에 대한 완전 조건부 사후분포

ZIOP 모형의 주요모수인 β , γ , τ 의 완전 조건부 사후분포는 다음과 같다.

1) β 의 조건부 사후분포

$$\begin{aligned} \pi(\beta|else) &\propto \prod_{i=1}^n \phi(s_i^*|x_i'\beta, 1) \cdot \pi(\beta) \\ &\propto \exp\left[-\frac{1}{2}\left\{\sum_i (s_i^* - x_i'\beta)^2 + (\beta - \beta_0)'\Sigma_\beta^{-1}(\beta - \beta_0)\right\}\right] \\ &\propto \exp\left[-\frac{1}{2}\left\{(s^* - x'\beta)'\Sigma_\beta^{-1}(s^* - x'\beta) + (\beta - \beta_0)'\Sigma_\beta^{-1}(\beta - \beta_0)\right\}\right] \\ &\propto \exp\left[-\frac{1}{2}\left\{\beta'(x'x)\beta + \beta'\Sigma_\beta^{-1}\beta - 2\beta'x's^* - 2\beta'\Sigma_\beta^{-1}\beta_0\right\}\right] \\ &\sim N(\mu_\beta, \Sigma_\beta) \end{aligned}$$

유도과정을 생략하면 다음과 같이 간략하게 표현된다.

$$\pi(\beta|else) \propto \prod_{i=1}^n \phi(s_i^*|x_i'\beta, 1) \cdot \pi(\beta) \sim N(\mu_\beta, \Sigma_\beta)$$

이 때, $\phi(s_i^*|x_i'\beta, 1)$ 는 평균 $x_i'\beta$, 분산 1(즉, $N(x_i'\beta, 1)$)의 확률밀도함수를 의미하며, β 의 조건부 사후분포는 분산 $\Sigma_\beta = (x'x + \Sigma_\beta^{-1})^{-1}$, 평균 $\mu_\beta = \Sigma_\beta \cdot (x's^* + \Sigma_\beta^{-1}\beta_0)$ 인 정규분포를 따른다¹²⁾.

2) γ 의 조건부 사후분포

$$\pi(\gamma|s_i^* > 0, else) \propto \prod_{i=1, s_i^* > 0}^n \phi(y_i^*|z_i'\gamma, 1) \cdot \pi(\gamma) \sim N(\mu_\gamma, \Sigma_\gamma)$$

γ 의 조건부 사후분포는 분산 $\Sigma_\gamma = (z'z + \Sigma_\gamma^{-1})^{-1}$, 평균 $\mu_\gamma = \Sigma_\gamma \cdot (z'y^* + \Sigma_\gamma^{-1}\gamma_0)$ 인 정규분포를 따른다. 단, y_i^* 는 $s_i^* > 0$ 이다.

12) Σ 는 일반적으로 다변량 정규분포의 분산을 의미한다.

3) τ 의 조건부 사후분포

$$\begin{aligned} \pi(\tau_j | s_i^* > 0, \text{else}) &\propto \pi(\tau_j) \cdot \prod_{i=1, s_i^* > 0}^n [I(\tau_{j-1} \leq y_i^* < \tau_j)]^{I(y_i=j)} \cdot [I(\tau_j \leq y_i^* < \tau_{j+1})]^{I(y_i=j+1)} \\ &\sim N(\tau_{0j}, \sigma_j^2) \cdot I[\max(\tau_{j-1}, i : y_i = j, y_i^*) \leq \tau_j < (\tau_{j+1}, i : y_i = j+1, y_i^*)] \end{aligned}$$

단, y_i^* 는 $s_i^* > 0$ 이다.

부록2: R 코드

본 연구에 사용된 R 프로그램은 다음과 같다. 자료의 입력, 초기값 지정 등의 기본코드는 생략하고 깁스 샘플러 수행과정을 요약하여 제시한다.

```
##### 분석에 사용된 주요 패키지 설치 #####
install.packages("mvtnorm"); library(mvtnorm)
install.packages("msm"); library(msm)
install.packages("Zelig"); library(Zelig)
install.packages("MCMCpack"); library(MCMCpack)

##### 깁스 샘플러 : beta, gamma, tau의 사후 모수 추출 #####

N=length(y) ;
n.warm=10000; n.sim=110000 ; n.thin=50          # 번인구간, 표본개수, thinning구간

for(i in 1: (n.sim) {
  for(j in 1: N) {
    ##### s_star
    if(y[j]!=0){w <- as.numeric((1-pnorm(x[j,]%s.b)
      /((1-pnorm(x[j,]%s.beta))+(pnorm(x[j,]%s.beta)*pnorm(-x[j,]%s.gamma))))
      u<-runif(1,0,1)
      if(u<=w) {s_star<-rtnorm(1,x[j,]%s.b,1,lower=-Inf, upper=0 )}
      else {s_star<-rtnorm(1,x[j,]%s.b,1,lower=0, upper=Inf)}}
    for(k in 1:3){
      if(y[j]==k){s_star=rtnorm(1,x[j,]%s.b,sd=1,lower=0, upper=Inf)}}
    s.data[j]=s_star }

  var = solve(t(x)%s.x + solve(var.b))          ##### beta
  m = var%as.matrix(s.data) + var.b%as.matrix(s.data)
  s.beta = t(rmvnorm(1,m,var))

  s.data1= ifelse(s.data<0, 0, 1) ; data1 = cbind(data, s.data1) ; ifs1.data=data1[s.data1==1, ]
  gy = ifs1.data[,5]; gx = as.matrix(cbind(1,ifs1.data[,1:4]))

  for(j in 1: length(gy)){
    ##### y_star
    if(gy[j]!=0){y_star = rtnorm(1, gx[j,]%s.g,sd=1,lower=-Inf, upper=d0)}
    else if(gy[j]==1){y_star = rtnorm(1, gx[j,]%s.g,sd=1,lower=d0, upper=d1)}
    else if(gy[j]==2){y_star = rtnorm(1, gx[j,]%s.g,sd=1,lower=d1, upper=d2)}
    else {y_star = rtnorm(1, gx[j,]%s.g,sd=1,lower=d2, upper=d3)}
    s.y[j]=y_star
  }
  yd=cbind(gy, s.y)

  var = solve(t(gx)%s.gx + solve(var.g))          ##### gamma
  m = var%as.matrix(s.y) + (var.g%as.matrix(s.g0))
  s.gamma = t(rmvnorm(1,m,var))

  ##### tau
  for(kk in 1:3){tau[kk] = rtnorm(1,d[kk],10,lower=max(tau[kk-1], max(yd[gy==kk,][,2])),
  upper=min(tau[kk+1], min(yd[gy==kk+1,][,2])) )}
  s.tau = t(cbind(tau1,tau2,tau3))
  ps = table(s.data1)

  ##### 사후 표본(최종)
  if(i > (n.warm & as.integer((i-n.warm)/n.thin)==(i-n.warm)/n.thin) ) {
    sample.gamma = cbind(sample.gamma, s.gamma)
    sample.beta = cbind(sample.beta, s.beta)
    sample.tau = cbind(sample.tau, s.tau)
  }
}
```