The Korean Journal of Applied Statistics (2014) DOI: http://dx.doi.org/10.5351/KJAS.2014.27.6.909
27(6), 909–922

Bayesian Inference with Inequality Constraints

Man-Suk Oh[a,1]

[a]Department of Statistics, Ewha Womans University

(Received October 21, 2014; Revised November 6, 2014; Accepted November 10, 2014)

Abstract

This paper reviews Bayesian inference with inequality constraints. It focuses on ⅰ) comparison of models with various inequality/equality constraints on parameters, ⅱ) multiple tests on equalities of parameters when parameters are under inequality constraints, ⅲ) multiple test on equalities of score parameters in models for contingency tables with ordinal categorical variables.

Keywords: Order restricted, multiple test, Markov chain Monte Carlo, Savage-Dickey density ratio, Bayes factor.

1. 서론

모수에 어떤 형태의 부등 제한 조건($>, <, =$)을 부여하는 것이 타당한 경우가 있다. 예를 들면, 진통제의 농도를 4 단계로 증가시키면서 네 그룹의 쥐에 주사할 경우 그룹별 위 점막의 평균 손상정도 $\mu_1, \ldots, \mu_4$ 가 농도의 증가에 따라 증가 또는 동일, 즉, $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$의 부등 조건을 가정하는 경우이다. 이렇듯 모수에 부등 제한조건이 존재하는 경우 이를 고려한 추론은 그렇지 않은 추론에 비하여 부등조건에 부합하는 합리적인 추정치를 제공하며 검정력을 향상시키는 장점이 있다 (Shang 등,2008).

부등 제한 조건을 반영한 고전적 추론에 관해서는 Marcus와 Peritz (1976), Williams (1977), Marcus (1982), Robertson 등 (1988), Hayter (1990), Liu (2001), Liu 등 (2002), Nashimoto와 Wright (2005) 등 예전부터 많은 연구가 이루어져 왔다. 그러나 고전적 추론에서는 일반적으로 검정 통계량의 분포가 복잡하여 수리적으로 구하기 어렵다는 문제점이 있다. 또한 고전적 가설검정의 경우 다중 검정이 어렵기 때문에 예를 들면 귀무가설이 $H_0 : \mu_1 = \cdots = \mu_4$일 경우 대립가설은 $H_1 : \mu_1 \leq \cdots \leq \mu_4, \mu_1 < \mu_4$가 되는데 만약 귀무가설을 기각했을 경우 어떤 모수들이 서로 다른지 판별하기가 어렵다.

부등 제한 조건을 반영한 베이지안 추론은 고전적 추론에 비하여 여러 장점들이 있다. 첫째, 사전분포의 영역을 주어진 부등 영역으로 국한시킴으로써 제한 조건에 부합하는 사후추정치를 쉽게 얻을 수 있다. 둘째, 모형 선택, 변수 선택을 포함한 가설 검정에서 여러 개의 가설을 동시에 검정하는 다중 동시 검정이 가능하다. 셋째, 복잡한 수리적 계산이 필요치 않으며 소표본에도 적용이 가능하다. 넷째, 모형에 존재하는 모수의 수 뿐만 아니라 부등 제한 조건으로 인하여 야기되는 모형의 복잡성을 모형선택에 반영할 수 있다 (Klugkist 등,2005; Oh, 2013).

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2013R1A1A2005481).
[1]Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 120-750, Korea. E-mail: msoh@ewha.ac.kr

Markov chain Monte Carlo(MCMC) 기법의 발달로 간단하지 않은 사후분포에서도 사후표본의 추출이 가능하게 되었다. 사후 분포가 부등제한 조건을 만족하는 영역으로 국한된 경우 MCMC 기법을 이용하여 사후표본을 추출하고 이로부터 추정치를 구하면 조건을 만족하는 사후 추정치를 구할 수 있으므로 부등 제한 조건을 반영한 베이지안 추론에서 추정은 비교적 쉽다. 반면, 가설검정은 추정 만큼 단순하지 않기 때문에 대부분의 최근 연구는 모형선택 또는 변수선택을 위한 가설검정에 집중되어 있다. 따라서 이 논문에서는 가설검정에 초점을 맞추어 특히 다음의 세 가지 주제에 대하여 기존의 연구를 살펴보기로 하겠다.

첫 번째 주제는 여러 가지 부등 제한 조건 가설들을 비교하는 것이다. 앞서 진통제 예에서는 $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$의 부등제한 조건을 가정하는 것이 타당하지만 어떤 경우에는 μ_i의 순서에 대한 여러 가설이 존재하는 경우가 있다. 예를 들면, 고통을 감내하는 시간에 대한 실험에서 대상을 네 그룹으로 나누어 그룹1에는 부정적이며 고통과 관련된 사진을, 그룹2에는 부정적이지만 고통과는 관련 없는 사진을, 그룹3에는 부정적과 고통 둘다 관련 없는 사진을, 그룹4에는 긍정적인 사진을 보여주었다 (Klugkist 등,2005). 심리학자들은 네 가지 다른 이론을 비교하고자 한다. 첫째 이론은 그룹1의 고통 감내 시간이 그룹2 보다 짧고, 그룹3의 고통감내 시간이 그룹4 보다 짧다. 그러나 그룹 2와 3 간에는 어떤 가정도 하지 않는다. 두 번째 이론은 그룹1부터 그룹 4의 고통 감내 시간은 단조 증가한다. 세 번째 이론은 그룹 2의 부정적 사진은 고통으로부터 관심을 분산시키는 효과가 있어서 오히려 아무 관련 없는 사진을 본 그룹3 보다 고통 감내 시간이 길지만 그러나 그룹4 보다는 짧다. 네 번째 이론은 사진은 아무 영향을 미치지 못한다. 위의 네 이론을 μ_i에 대한 식으로 표현하면,

$$
\begin{aligned}
&M_1 : \mu_1 < \mu_2,\ \mu_3 < \mu_4, \\
&M_2 : \mu_1 < \mu_2 < \mu_3 < \mu_4, \\
&M_3 : \mu_1 < \mu_3 < \mu_2 < \mu_4, \\
&M_4 : \mu_1 = \mu_3 = \mu_2 = \mu_4
\end{aligned} \tag{1.1}
$$

이다. 위 식에서 콤마(,)는 아무 제한 조건을 설정하지 않는다는 것을 나타낸다. 이 때 추론의 목적은 어떤 가설을 선택할 것인가 하는 것이다. 이 논문의 2장에서는 이 주제를 다루기로 한다.

두 번째 주제는 모수에 동등성(=)을 포함한 부등 제한 조건이 성립한다고 할 경우, 이 부등 제한 조건 하에서 동등성의 검정이다. 예를 들면, 앞의 진통제 예에서 그룹들의 평균 부작용의 정도가 $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$ 를 만족한다고 할 때,

$$
\begin{aligned}
&K_0 : \mu_1 < \mu_2 < \mu_3 < \mu_4, \\
&K_1 : \mu_1 = \mu_2 < \mu_3 < \mu_4, \\
&K_2 : \mu_1 < \mu_2 < \mu_3 = \mu_4, \\
&K_3 : \mu_1 = \mu_3 = \mu_2 = \mu_4
\end{aligned} \tag{1.2}
$$

의 네 가설 중에서 어느 것이 가장 적절한가 하는 것이다. K_0는 레벨의 증가에 따라 부작용의 정도가 유의하게 단조 증가한다, K_1은 네 그룹의 부작용의 정도 차이가 없다, K_2는 진통제의 농도를 첫 번째 레벨에서 두 번째 레벨로 높이더라도 부작용의 정도에 미치는 영향이 없지만 그 다음 두 레벨의 농도 증가는 유의한 영향을 미친다, K_3는 처음 세 레벨의 증가는 부작용의 정도를 증가시키지만 마지막 네 번째 증가는 이미 세 번째에서 부작용의 정도가 한계에 도달했으므로 더 이상의 유의미한 증가가 일어나지 않는다는 것을 의미한다. 3장에서는 이 주제를 다루기로 한다.

세 번째 주제는 부등제한 조건하에서 분할표 분석이다. 부등 제한 조건을 반영한 추론이 가장 활발히 이루어지고 있는 분야가 범주형 자료에 대한 분할표의 분석이다. 그런데 분할표에 대한 모형은 일반적인 분산분석 모형 또는 회귀 모형과 다르며 identifiability를 만족시키기 위한 제한 조건들이 추가로 존재하기 때문에 좀더 복잡하다. 본 논문에서는 부등 제한 조건을 반영한 베이지안 분할표 분석을 4장에서 따로 다루고자 한다.

마지막으로 5장에서는 추후 연구주제들을 제시한다.

2. 부등 제한 조건들의 비교

자료 y_{ij} 가 다음과 같은 ANOVA 모형을 따른다고 하자.

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, n,\ i = 1, \ldots, k,$$

여기에서 y_{ij}는 i번째 처리의 j번째 대상으로부터 얻은 관측치, μ_i는 i번째 처리 그룹의 평균, ε_{ij}는 $N(0, \sigma^2)$따르는 오차항이다.

우리의 관심은 식 (1.1) 에서와 같이 μ_i에 관한 여러 개의 부등 (<,>,=) 조건들을 비교하는 것이다. 설명을 간단히 하기 위해서 k=4일 때 식 (1.1)의 가설들을 검정하는 문제에 대하여 기술하기로 하겠다.

부등제한 조건들에 대한 고전적 추론 방법은 Barlow 등 (1972), Robertsan 등 (1988), Silvapulle와 Sen (2004), Dykstra 등 (2002), Shyamal과 Dunson(2005) 등의 많은 연구가 있다. 그러나 고전적 검정법은 1장에서 언급한 문제점들 외에도 비내포(non-nested)된 모형들의 비교가 어렵다는 문제점이 있다 (Marden 등, 2002; Klugkist 등, 2005).

흔히 쓰이는 모형선택 기준인 AIC(Akaike, 1987)이나 BIC(Schwartz, 1978)는 모형의 적합도와 복잡성을 둘 다 고려하는 장점이 있다. 그런데 AIC와 BIC에서는 모형의 복잡성을 모형에 포함된 모수의 수로 측정하기 때문에 부등조건에 대한 가설검정에는 적절치 않다. 예를 들면, 식 (1.1)에서 모형 M_1, M_2, M_3에서 모수의 수는 아무런 제약 조건이 없는 완전 모형

$$M_0 : \mu_1, \mu_2, \mu_3, \mu_4$$

에서의 모수의 수와 같다. Anraku (1999)와 Kuiper 등 (2011, 2012)는 이 문제를 해결하기 위하여 부등조건을 모형의 복잡성에 반영하는 모형선택 기준을 제시하였으나 이 방법은 비교하는 모형의 수가 많을 때 복잡성을 추정하기 어렵다.

부등 조건들을 비교하는 문제에서 베이지안 방법은 다른 방법들에 대한 좋은 대안이다. 왜냐하면 베이지안 방법은 모형의 사후확률에 기반하는데, 사후확률은 모형에 존재하는 모수의 수 뿐만 아니라 모수공간이 제한됨으로써 발생되는 모형의 복잡성을 자동적으로 반영하기 때문이다 (Moreno, 2005; Kugkist 등, 2005; Hoijtink 등, 2008).

Klugkist 등 (2005)는 포괄적 사전분포를 이용하여 부등조건들을 비교하는 베이지안 기법을 제안하였다. 즉, μ_i에 대한 아무 제약이 없는 완전모형에서 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$에 대한 사전 분포를 π_0라 할 때, 부등제한조건이 있는 모형의 사전 분포로 π_0를 주어진 조건으로 제한한 분포를 사용할 것을 제안하였다. 예를 들면, 식 (1.1)의 M_1에서 $\boldsymbol{\mu}$의 사전분포는 $\pi_1(\boldsymbol{\mu}) = \pi_0(\boldsymbol{\mu})\ I(\mu_1 < \mu_2, \mu_3 < \mu_4)$를 선택하는 것이다. 이 사전분포를 사용하면 M_1과 M_0의 베이즈 상수는, 두 모형의 사전확률이 같다고 할 때,

$$\mathrm{BF}_{10} = \frac{P_0(\mu_1 < \mu_2, \mu_3 < \mu_4 \mid y)}{P_0(\mu_1 < \mu_2, \mu_3 < \mu_4)} \tag{2.1}$$

로 주어진다. 이 때 $P_0(\cdot \mid y)$와 $P_0(\cdot)$은 각각 완전모형 M_0하에서 사후확률과 사전확률이다. 베이즈 상수의 추정은 M_0에서 $\boldsymbol{\mu}$의 사전표본과 사후표본을 추출하여 부등조건을 만족시키는 표본의 비율을 사용하여 추정한다.

Klugkist 등 (2005)의 이 방법은 식 (1.1)의 M_1, M_2, M_3와 같이 동등성(=)를 포함하지 않은 모형에서는 가능하지만 M_4와 같이 동등성을 가진 모형에는 적용할 수 없다. M_0에서 가정한 연속 사전분포에서 μ_i들이 같을 확률은 0이기 때문이다.

Laudy와 Hoijtink (2007)과 Klugkist와 Hoijtink (2007)은 이 문제의 해결책으로 M_4를 유사동등성 가설

$$M_4^* : \mu_1 \approx \mu_2 \approx \mu_3 \approx \mu_4$$

로 대체하여 검정하는 방법을 제안하였다. 여기에서 $\mu_1 \approx \mu_2$는 아주 작은 상수 ϵ에 대하여 $\mid \mu_1 - \mu_2 \mid < \epsilon$를 의미한다. Mulder 등 (2010), Klugkist 등 (2007), Van Wesel 등 (2011)은 이를 더욱 확장하여

$$M_5^* : \mu_1 < \mu_2 < \mu_3 \approx \mu_4$$

와 같이 부등조건과 유사 동등성이 결합된 가설도 검정할 수 있는 방법을 제시하였다. 여기에서는 간단히 $M_6 : \mu_1 = \mu_2$와 M_6에 대응하는 유사동등성 가설 $M_6^* : \mu_1 \approx \mu_2$의 검정에 대하여 정리하기로 하겠다.

$\delta_1 = \mu_2 - \mu_1$으로 정의하면, $\mu_1 \approx \mu_2$는 $-\epsilon < \mu_1 - \mu_2 < \epsilon$으로 표현되기 때문에 식 (2.1)을 적용하면

$$\mathrm{BF}_{6^*0} = \frac{P_0(\mid \delta_1 \mid < \epsilon \mid y)}{P_0(\mid \delta_1 \mid < \epsilon)}$$

으로 주어진다. 그런데 위 식의 분자 분모는 ϵ이 작으면 주어진 좁은 영역에 속하는 표본이 적어 정확한 추정이 어렵다. 표본의 수가 크지 않으면 만족하는 표본이 없을 수도 있다. 이 문제를 해결하기 위해 Klugkist와 Hoijtink (2007)은 단계적 방법을 제안하였는데 우선 어느 정도 큰 ϵ_1을 사용하여 $M_6^{*(1)} :\mid \delta_1 \mid < \epsilon_1$과 M_0를 비교하는 $BF_{6^*0}^{(10)}$를 계산한 다음, $\epsilon_2 < \epsilon_1$에 대하여 $M_6^{*(2)} :\mid \delta_1 \mid < \epsilon_2$를 정의하고 $M_6^{*(1)}$와 $M_6^{*(2)}$를 비교하는 베이즈 상수 $BF_{6^*}^{(21)}$를 추정한다. 이 때 표본은 $M_6^{*(1)}$에서 추출하기 때문에 이미 $\mid \delta_1 \mid < \epsilon_1$을 만족한 상태에서 $\mid \delta_1 \mid < \epsilon_2$를 만족하는 표본은 상당수 있을 것이다. 이와 같이 ϵ을 줄여가면서 반복한 다음 베이즈 상수들을 곱해주면 BF_{6^*0}를 얻을 수 있다. Wetzels 등 (2010)는 이 방법에서 BF_{6^*0}가 M_6와 M_0를 비교하는 BF_{60}로 수렴함을 증명하였다. 그러나 유사 동등성의 경우 ϵ을 점점 줄여가면서 반복하는 과정은 간단하지 않으며 계산시간이 오래 걸릴 수도 있다.

Oh (2014b)는 엄밀한 부등(>, <)과 동등성(=) 조건의 혼합을 포함하여 임의의 부등 제한조건에 적용할 수 있는 간단한 베이즈 상수를 제안하였다. 이 방법에서는 유사 동등성을 사용하지 않고 동등성 가정을 검정한다. 예를 들면, 위의 M_5^*대신 $M_5 : \mu_1 < \mu_2 < \mu_3 = \mu_4$를 검정하는 것이다. M_5에서 $\delta_3 = \mu_4 - \mu_3$에는 0에서의 점확률을 그리고 동등성 가정이 없는 나머지 모수들에 대해서는 연속분포를 사용하는 합성 사전분포를 가정하는데, 이 때 연속분포로 π_0를 제한하는 표괄적 사전분포를 사용한다. M_5와 M_0에 대한 베이즈 상수를 구하기 위해 2단계 방법을 사용하는데 먼저 $M_1 : \mu_1 < \mu_2 < \mu_3 < \mu_4$와 M_0의 베이즈 상수를 구하면 $BF_{10} = P(\mu_1 < \mu_2 < \mu_3 < \mu_4 \mid y)/P(\mu_1 < \mu_2 < \mu_3 < \mu_4)$이므로 이는 쉽게 추정할 수 있다. 다음, M_5와 M_1를 비교하는 베이즈 상수를 구한다면 $\mathrm{BF}_{50} = \mathrm{BF}_{51} \times \mathrm{BF}_{10}$로 구할 수 있다.

Oh (2014b)는 Savage-Dickey density ratio를 이용하여 $\mathrm{BF}_{51} = \pi_1(\delta_3 = 0 \mid y)/\pi_1(\delta_3 = 0)$가 됨을 보였다. 여기에서 $\pi_1(\delta_3 = 0 \mid y)$와 $\pi_1(\delta_3 = 0)$는 각각 M_1하에서 δ_3의 주변 사후밀도함수와 주변 사전

밀도함수의 0에서의 값이다. 주변 사후 밀도함수는 δ_3의 조건부 사후분포가 주어지고 모형으로부터 추출된 모수의 사후표본이 있다면 Rao-Blackwellization을 이용하여 쉽게 추정할 수 있다. δ_3가 스칼라가 아닌 벡터인 경우에는 δ_3의 조건부 사후밀도함수는 주어지지 않을 수 있다. 하지만 많은 경우에 δ_3의 각 원소의 조건부 사후밀도함수는 주어지는데, Oh (2014b)는 각 원소의 조건부 사후밀도함수를 이용하여 δ_3 벡터의 주변 사후밀도 함수를 추정하는 기법을 제안하였다. 이에 대한 보다 제세한 사항은 3장에서 기술하기로 한다.

3. 부등제한조건하에서의 동등성 검정

$\mu_1, \ldots, \mu_k$에 대하여 예를 들어 $\mu_1 \leq \cdots \leq \mu_k$ 부등제한조건이 존재함을 안다고 가정하자. 그런데 이 부등 제한 조건에서 $\leq$는 '작다'($<$) 또는 '같다'($=$)를 의미한다. 앞서 진통제 예를 보면, 진통제의 농도가 높아지면 부작용이 증가하거나 같거나 둘 중 하나인데 식 (1.2)에서와 같이 $\mu_1 \leq \cdots \leq \mu_k$ 조건 하에서 μ_i들 간에 $<$와 $=$의 조합으로 된 가설들을 비교하는 것이 추론의 목적이 될 수 있다.

식(1.2) 의 예에서는 4개의 가설을 비교하는 예를 들었지만 실은 어느 레벨에서 차이가 없는지 정보가 없다면 동등성($=$)과 단조증가($<$)의 모든 가능한 조합을 고려해야 할 것이다. 이 경우 k개의 그룹이 존재한다면 최대 2^{k-1}개의 가설에 대한 비교검정이 요구될 수 있다. 이 중 $K_0 : \mu_1 < \cdots < \mu_k$를 제외한 나머지 가설들은 최소 하나의 동등성을 포함하고 있다.

부등제한조건 하에서 동등성에 대한 베이지안 가설 검정에서는 동등성을 포함하는 모형에서 degenerate 되는 모수에 대한 점확률과 나머지 모수들에 대한 영역이 제한된 연속분포를 합성한 합성 사전분포를 사용해야 한다. 따라서 2^{k-1}개의 모형을 비교한다면 사후분포가 2^{k-1}개의 서로 다른 "제한된" 분포들의 합성으로 주어지는 데 이 때 베이즈 상수 또는 사후확률의 계산이 쉽지 않다.

이 문제점을 피하기 위하여 Dunson과 Neelon (2003)은 고전적 추론에서 사용하는 isotonic regression 변환 (Bacchetti, 1989; Morton-Jones 등, 2000)을 이용한 하이브리드 방법을 제안하였다. 일단 모수 $\boldsymbol{\mu}$에 대한 사전분포로 제한조건을 반영하지 않은 연속 분포를 선택하고, 이로부터 유도된 사후분포로부터 MCMC 기법을 사용하여 사후표본를 얻은 다음 이들을 isotonic regression 변환하였다. 변환으로 새로이 얻은 표본은 주어진 부등제한조건을 만족하므로 이를 이용하여 조건을 만족하는 $\boldsymbol{\mu}$의 추정치를 얻을 수 있다. 또한 변환 전의 표본과 변환 후의 표본의 관계를 이용하여 $H_0 : \mu_1 = \cdots = \mu_k$와 $H_1 : \mu_1 \leq \cdots \leq \mu_k$, $\mu_1 < \mu_k$를 검정할 수 있다. 이 방법은 사용이 용이하고 어떤 형태의 부등제한 조건에도 사용할 수 있는 장점이 있다. 그러나 이 방법은 어떤 통계적 모형에 해당하는지 확실치 않고 (Taylor 등, 2007) 또한 H_0를 기각할 경우 어떤 평균들 사이에 엄격한 부등관계($<$)가 존재하는지 보여주지 못한다.

Shang 등 (2008)은 연속된 두 평균의 차이를 $\delta_i = \mu_{i+1} - \mu_i$로 정의하고 $\boldsymbol{\mu}$에 대한 추론을 $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{k-1})$에 대한 추론으로 바꾸어 수행하였다. 각 δ_i의 사전분포로 0에서의 점확률과 0보다 큰 구간으로 제한된 연속분포의 합성분포를 가정한다. 또한 δ_i들은 서로 독립이라 가정한다. MCMC 기법을 사용하여 δ_i들의 사후표본을 추출하고 이를 이용하여 각 i에 대하여

$$H_{0i} : \mu_i = \mu_{i+1} \text{ vs. } H_{1i} : \mu_i < \mu_{i+1}$$

를 검정하는 베이지안 방법을 제안었다. H_{0i}중 하나 이상 기각되면 $H_0 : \mu_1 = \cdots = \mu_k$를 기각할 수 있기 때문에 H_0와 $H_1 : \mu_1 \leq \cdots \leq \mu_k$, $\mu_1 < \mu_k$의 검정도 가능하다.

그러나, 각 i에 대하여 H_{0i}와 H_{1i} 를 독립적으로 검정하는 짝검정(pairwise test)은 특히 $\delta_{i-1} = \mu_i - \mu_{i-1}$과 $\delta_i = \mu_{i+1} - \mu_i$가 서로 연관되어 있기 때문에 잘못된 결론을 유도할 수 있다는 것을 Oh와 Shin

(2011), Oh (2013)에서 예시로 보여주고 있다. 따라서 $\mu_1, \ldots, \mu_k$에 대한 동등성 검정에서 짝검정이 아닌 동시 다중 검정이 요구된다.

Oh와 Shin (2011)은 부등 제한 조건하에서 '$<$'와 '$=$'의 모든 가능한 조합을 동시에 검정하는 베이지안 방법을 제안하였다. 식 (1.2)의 모형을 보면, $K_0 : \delta_1 > 0, \ldots, \delta_{k-1} > 0$는 동등성이 하나도 존재하지 않는 모형이다. K_0에는 degenerate되는 모수가 하나도 없기 때문에 모수에 부등 영역으로 제한된 연속 사전분포 π_0를 가정할 수 있고 따라서 사후분포도 부등 조건으로 제한된 연속분포이다. K_0에서 δ_i의 사후표본은 깁스 표본기법을 사용하면 쉽게 추출할 수 있다. 동등성이 포함된 모형, 예를 들면 $K_1 : \delta_1 = 0, \delta_2 > 0, \delta_3 > 0$에서는 사전분포로 $\pi_1(\delta_2, \delta_3) = \pi_0(\delta_2, \delta_3 \mid \delta_1 = 0)$를 가정한다. 이같은 포괄적 사전분포 하에서 K_1과 K_0의 베이즈 상수를 구하면 Savage-Dickey density ratio(Dickey와 Lieutz, 1970; Dickey, 1971, 1976; Verdinelli와 Wasserman, 1995)에 의하여

$$\mathrm{BF}_{10} = \frac{\pi_0(\delta_1 = 0 \mid y)}{\pi_0(\delta_1 = 0)}$$

이다. 여기에서 $\pi_0(\delta_1 = 0 \mid y)$와 $\pi_0(\delta_1 = 0)$ 는 각각 K_0하에서 δ_1의 주변 사후밀도함수와 주변사전밀도 함수의 0에서의 값이다.

이를 확장하면 $K_3 : \delta_1 = \delta_2 = \delta_3 = 0$와 K_0의 베이즈 상수는

$$\mathrm{BF}_{30} = \frac{\pi_0(\delta_1 = \delta_2 = \delta_3 = 0 \mid y)}{\pi_0(\delta_1 = \delta_2 = \delta_3 = 0)}$$

이다. 이를 일반화시키면, 하나 이상의 δ_i를 0으로 가정하는 축소모형과 K_0의 베이즈 상수는 축소모형에서 0으로 가정하는 δ_i들의 주변 사후 밀도함수와 주변 사전 밀도함수의 비율로 주어진다. 보통 주변 사전 밀도함수는 알려진 함수이므로 주변 사후 밀도함수만 추정하면 베이즈 상수를 추정할 수 있다.

여기에서 주목할 점은 어떤 축소모형에 대해서도 주변 사후 밀도함수가 항상 K_0하에서의 주변 사후 밀도함수라는 점이다. 2^{k-1}개의 모든 가능한 동등성 가설의 동시 검정을 위해서 $\boldsymbol{\delta}$의 모든 가능한 부분집합에 대해서 K_0하에서 주변 사후 밀도함수 값만 추정하면 되므로, 다중 가설검정 문제가 $\boldsymbol{\delta}$의 부분집합의 주변 사후 밀도함수 추정의 문제로 바뀌게 된다.

K_0에서 $\boldsymbol{\delta}$에 대하여 $\delta_i>0$영역으로 제한된 정규사전분포를 가정하면 $\boldsymbol{\delta}$들의 공통 사후 분포는 제한된 다변량 정규분포로 주어지고 각 원소 δ_i의 조건부 사후분포는 $\delta_i > 0$로 제한된 일변량 정규분포로 주어지므로 깁스표본기법을 이용하여 $\boldsymbol{\delta}$의 사후표본은 쉽게 추출할 수 있다. 그러나 $\boldsymbol{\delta}$의 분포가 제한되어 있기 때문에 $\boldsymbol{\delta}$의 부분집합의 주변 사후 밀도함수는 수리적으로 구할 수 없다. Oh (1999)는 이 문제를 깁스 표본기법에서 표본의 균형공식을 이용하여 해결하는 병법을 제안하였다. 이 방법을 적용하면,

$$\begin{aligned}
&\pi_0(\delta_1 = \delta_2 = \delta_3 = 0 \mid y) \\
&= E[\pi_0(\delta_1 = 0 \mid \delta_2, \delta_3, \varphi, y) \times \pi_0(\delta_2 = 0 \mid \delta_1 = 0, \delta_3, \varphi, y) \times \pi_0(\delta_3 = 0 \mid \delta_1 = \delta_2 = 0, \varphi, y)]
\end{aligned}$$

이다. δ_i의 조건부 사후분포는 제한된 일변량 정규분포이기 때문에 δ_i의 조건부 사후밀도함수는 식이 주어진다. 기대치는 항상 K_0하에서 $\boldsymbol{\delta}$와 그 외의 모수 φ의 사후분포에 대한 것이므로 K_0로부터 추출된 사후 표본만 있으면 $\boldsymbol{\delta}$의 임의의 부분집합에 대한 주변 사후 밀도함수를 모두 동시에 추정할 수 있다.

Oh와 Shin (2011)의 방법은 기존의 역점프 MCMC나 Stochastic Search Variable Selection(SSVS) 와 달리 모든 가능한 모형을 적합시키지 않고 단지 K_0로부터 추출된 한 세트의 표본만 있으면 가설들을 동시에 검정할 수 있다는 장점이 있다. 그러나 각 원소에 대한 완전 조건부 밀도함수를 매 MCMC 표본마다 반복적으로 계산해야 하기 때문에 k가 클 경우 시간이 오래 걸릴 수 있다.

4. 순서적 제한조건을 지닌 분할표의 추론

분할표에서 부등 제한조건이 대두되는 가장 흔한 경우는 행변수 또는 열변수가 순서적 범주형 변수인 경우이다. 변수가 순서적 범주를 가질 때 이 변수의 범주에 대응하는 스코어 모수에 주어진 순서를 반영하는 것이 타당할 것이다. 순서적 부등 제한 조건을 반영한 분할표 분석은 순서를 반영하지 않는 추론에 비하여 앞서 언급한 여러 장점들로 인하여 일찍이 연구자들의 관심을 끌어 왔다. 고전적 추론 방법으로는 Agresti 등 (1987), Ritov와 Gillula (1993), Galindo-Garre와 Vermunt (2004, 2005)가 있으며 Agresti와 Coull (2002)는 순서적 제한 조건을 가지는 분할표의 분석을 정리하였다. 반면 베이지안 추론에 대한 연구는 Iliopoulos 등 (2007), Hoijtink 등 (2008), Tarantola 등 (2008), Iliopoulos 등 (2009), Oh (2014a) 등으로 최근에 들어서야 관심을 끌고 있다. 베이지안 기법의 개발이 상대적으로 부족하고 늦은 이유는 분할표에 대한 로그선형 모형의 경우 MCMC 기법의 적용이 쉽지 않은 등 계산상의 문제 때문으로 추측된다.

순서적 제한조건을 지닌 분할표에 대한 베이지안 기법을 설명하기 위하여 행변수 X와 열변수 Y가 둘 다 순서적 범주형 변수인 경우를 고려해보자. 이 경우 흔히 사용되는 Goodman (1979)의 RC모형은 분할표에서 (i,j)범주의 돗수가 Poisson(θ_{ij})를 따른다고 할 때 다음과 같다.

$$\lambda_{ij} = \log(\theta_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \phi\mu_i\nu_j, \quad i = 1,\ldots,r,\ j = 1,\ldots,c.$$

여기에서 $\mu_1,\ldots,\mu_r$은 X의 r개 범주에 대한 스코어 모수, $\nu_1,\ldots,\nu_j$는 Y의 c개 범주에 대한 스코어모수이며 ϕ는 전반적인 X와 Y의 관련성의 측도이다.

RC 모형에서 모수들은 identifiability를 위해 $\lambda_1^X = \lambda_1^Y = \mu_1 = \nu_1 = 0$, $\mu_r = \nu_c = 1$를 만족해야 한다. 또한 순서적 제한조건에 의해, X와 Y의 순서적 범주들이 단조감소가 아니라고(non-decreasiong)할 때 $\mu_1 \le \cdots \le \mu_r$, $\nu_1 \le \cdots \le \nu_c$가 성립한다. 이 둘을 조합하면

$$\begin{aligned} &\lambda_1^X = \lambda_1^Y = 0, \\ &\quad 0 = \mu_1 \le \cdots \le \mu_r = 1, \quad \mu_1 < \mu_r, \\ &\quad 0 = \nu_1 \le \cdots \le \nu_c = 1, \quad \nu_1 < \nu_c \end{aligned} \tag{4.1}$$

의 제한조건을 가진다.

RC 모형에서 스코어 모수의 조건부 사후분포는 표본 생성이 용이한 분포로 주어지지 않는데 이에 위의 제한조건까지 부가되어 모수의 사후표본 추출이 쉽지 않다. Iliopoulos 등 (2007)은 RC 모형에서 깁스 표본 기법내에서 메트로폴리스-헤스팅스 기법을 사용하여 식 (4.1)를 만족하는 모수의 사후표본을 추출하고 이를 이용하여 추정치를 구하는 베이지안 방법을 제안하였다.

분할표 분석에서 모수의 추정치를 구하는 것 외에 또 다른 주된 관심사는 스코어 모수의 동등성에 대한 검정이다. 만약 어떤 두 연속된 스코어 모수가 일치한다고 하면, 예를들어 $\mu_1 = \mu_2$이면, 두 범주는 동일한 범주로 통합될 수 있으므로 두 범주를 합쳐 모형을 단순화 시킬 수 있고 해석도 용이해지기 때문이다. 식 (4.1)의 제한조건 하에서 인접한 범주들의 동등성에 대한 가설은 아래와 같이 총 $(2^{r-1}-1)(2^{c-1}-1)$개가 존재한다. 우리의 관심은 이 중 어느 모형(가설)들이 자료와 잘 적합되는가 하는 것이다.

$$\begin{aligned} &L_0 : 0 = \mu_1 < \ldots < \mu_r = 1, \quad 0 = \nu_1 < \ldots < \nu_c = 1 \\ &L_1 : 0 = \mu_1 = \mu_2 < \mu_3 < \ldots < \mu_r = 1, \quad 0 = \nu_1 < \ldots < \nu_c = 1 \end{aligned}$$

$$L_2 : 0 = \mu_1 < ... < \mu_r = 1, \quad 0 = \nu_1 = \nu_2 < \nu_3 < ... < \nu_c = 1$$
$$L_3 : 0 = \mu_1 = \mu_2 = \mu_3 < ... < \mu_r = 1, \quad 0 = \nu_1 < ... < \nu_c = 1 \tag{4.2}$$
$$\vdots$$

Iliopoulos 등 (2009)는 순서적 부등 제한 조건을 가지는 분할표에서 동등성에 대한 모든 가능한 모형을 동시에 검정하는 베이지안 방법을 처음으로 제시하였다. 이 방법은 역점프 MCMC 기법을 사용하여 가능한 모형들을 방문하고 방문한 모형에서는 Iliopoulos 등 (2007) 방법으로 표본을 추출하며, 각 모형의 사후확률은 방문횟수의 상대적 비율로 추정한다. 그러나 잘 알려진 바와 같이, 역점프 MCMC는 후보 모형들을 이동방문하게 되는데 모형 방문 후 그 모형의 모수들을 추정하기 때문에 결국 주어진 반복수 내에서 모든 가능한 모형들을 적합시키는 셈이 되어 후보 모형의 수가 많을 경우 효율이 떨어질 수 있다. 또한 확률을 상대적 방문횟수로 추정하는 것도 MCMC 반복수가 매우 크지 않으면 정확한 추정치를 얻기 어렵다.

Oh (2014a)는 부등제한을 가진 RC 모형에 대한 베이지안 추론에서 발생하는 계산상의 어려움을 적절한 잠재변수를 도입하여 해결하는 방법을 제안하였다. 이는 이항자료에 대한 프로빗 모형에서 잠재변수를 도입하여 깁스표본기법의 적용을 가능하게 한 Albert와 Chib (1991) 방법의 아이디어와 비슷하다.

RC 모형에 잠재변수 $Z_{ij} \sim N(\lambda_{ij}, 1)I(\Phi^{-1}F(y_{ij} - 1 \mid \lambda_{ij}) < Z_{ij} - \lambda_{ij} < \Phi^{-1}F(y_{ij} \mid \lambda_{ij}))$를 도입하면 $P(\Phi^{-1}F(y_{ij} - 1 \mid \lambda_{ij}) < Z_{ij} - \lambda_{ij} < \Phi^{-1}F(y_{ij} \mid \lambda_{ij}) = F(y_{ij} \mid \lambda_{ij}) - F(y_{ij} - 1 \mid \lambda_{ij}) = P(Y_{ij} = y_{ij} \mid \lambda_{ij})$가 성립한다. 여기에서 Φ는 표준정규분포의 누적분포함수이고 $F(\cdot \mid \lambda_{ij})$는 평균 $e^{\lambda_{ij}}$인 포아송 분포의 누적분포함수이다. 따라서 이산변수 Y_{ij}를 제한된 정규분포를 따르는 잠재변수 Z_{ij}로 대체할 수 있다.

그런데 평균 e^{λ}를 갖는 포아송 분포의 누적분포함수 $F(k \mid \lambda)$, $k = 0, 1, 2, \ldots,$는

$$F(k \mid \lambda) \approx \Phi\left(-3\sqrt{k+1}\left[\left(\frac{e^{\lambda}}{k+1}\right)^{\frac{1}{3}} - 1 + \frac{1}{9(k+1)}\right]\right)$$

으로 근사시킬 수 있으므로 (Johnson과 Kotz, 1972, Ch 4), Z_{ij}의 제한조건을 아래와 같이 단순화시킬 수 있다.

$$\begin{aligned} &Z_{ij} \sim N(\lambda_{ij}, 1) \\ &\times I\left(-3\sqrt{y_{ij}}\left[\left(\frac{e^{\lambda_{ij}}}{y_{ij}}\right)^{\frac{1}{3}} - 1 + \frac{1}{9(y_{ij})}\right] < Z_{ij} - \lambda_{ij} < -3\sqrt{y_{ij}+1}\left[\left(\frac{e^{\lambda_{ij}}}{y_{ij}+1}\right)^{\frac{1}{3}} - 1 + \frac{1}{9(y_{ij}+1)}\right]\right). \end{aligned} \tag{4.3}$$

제한조건을 무시했을 때 Z_{ij}의 평균 λ_{ij}가 모수 $\lambda, \lambda_i^X, \lambda_j^Y, \mu_i, \nu_j$의 선형 결합이기 때문에 이들 모수에 식 (4.1)를 만족하는 제한된 정규 사전분포를 가정하면, 조건부 사후분포는 제한된 정규분포로 주어지므로 깁스표본기법를 적용하여 모수들의 사후표본을 추출하고 조건에 부합하는 사후추정치를 구할 수 있다. 또한, 주어진 조건을 만족하는 사후표본 추출이 가능하고 각 스코어 모수의 조건부 사후확률밀도함수를 계산할 수 있으므로, 3장에서와 같이 스코어 모수의 임의의 부분집합에 대한 주변 사후밀도함수를 추정하여 동등성에 대한 모든 가능한 모형을 동시에 검정할 수 있다.

Oh (2014a) 방법은 잠재변수의 도입과 포아송 분포의 누적분포함수에 대한 근사식을 이용하여 모수의 조건부 사후분포를 제한된 정규분포로 만듬으로써 로그선형 모형에서의 계산상의 어려움을 해결하였다는데 의미가 있다. 후보 모형을 모두 적합시키는 것이 아니라 하나의 모형 L_0만 적합시키면 되고, 깁스표본기법을 적용함으로써 MCMC표본 추출함수의 선택에 대하여 고민할 필요가 없으며, 조건부 사후확

Table 4.1. Age and the severity of dream disturbance

Age	severity of disturbance				total
	low			high	
	1	2	3	4	
5–7	7	4	3	7	21
8–9	10	15	11	13	49
10–11	23	9	11	7	50
12–13	28	9	12	10	59
14–15	32	5	4	3	44
total	100	42	41	40	223

Table 4.2. Posterior probabilities of the best 5 models in the drean disturbance data

Model	Posterior probability	
	S-D ratio	RJMCMC
$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	0.1987	0.1540
$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	0.1255	0.1620
$\mu_1 = \mu_2 < \mu_3 < \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	0.1041	0.0725
$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 < \nu_3 < \nu_4$	0.0897	0.0609
$\mu_1 = \mu_2 < \mu_3 < \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	0.0860	0.0877

률밀도함수가 주어지므로 베이즈 상수의 추정이 용이한 장점이 있다. 단, $\lambda, \lambda_i^X, \lambda_j^Y, \mu_i, \nu_j$의 조건부 사후분포에서 제한된 영역을 구할 때는 식 (4.3)의 제한영역을 λ_{ij}에 대하여 푼 다음 각 모수에 대한 제한영역을 구해야 하는데, 이 과정에서 λ_{ij}에 대한 일차원 방정식의 해를 수치적으로 구해야 하는 번거로움이 따른다.

실제 분석 예로 Maxwell (1961)의 악몽에 의한 수면방해(dream disturbance) 자료에 대한 모형 선택을 살펴보자.

Table 4.1에서와 같이 나이가 5개의 순서적 범주로, 수면방해의 정도가 4개의 순서적 범주로 구성되어 있다. 나이의 범주에 대한 스코어 모수를 $\mu_1, \ldots, \mu_5$라 하고 수면방해의 정도 범주에 대한 스코어 모수를 $\nu_1, \ldots, \nu_4$라 하자.

이 자료에 Iliopoulos 등 (2009)의 역점프 방법과 Oh (2014a)의 Savage-Dickey density ratio 방법을 적용하여 베이지안 모형 선택을 수행한 결과는 Table 4.2, Table 4.3 과 같다. 참고로, 동일 자료에 Agresti 등 (1987)는 순서적 제한을 가진 열효과(Column Effect) 모형을 적용하여 $\nu_1 < \nu_2 = \nu_3 < \nu_4$를 최적 모형으로 선택하였고 Ritov와 Gilula (1993)은 순서적 제한 조건을 지닌 대응분석모형(correspondence analysis model)을 적용하여 $\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$, $\nu_1 < \nu_2 = \nu_3 < \nu_4$ 모형을 최적모형으로 선택한 바 있다.

Table 4.2를 보면, Savage-Dickey density ratio 방법과 역점프 방법이 선택한 상위 5개의 모형은 일치한다. 단, Savage-Dickey density ratio 방법은 Ritov와 Gilula (1993)가 선택한 $\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$, $\nu_1 < \nu_2 = \nu_3 < \nu_4$ 모형을 최적으로 선택하였으나 역점프 방법은 위 모형에서 $\nu_3 < \nu_4$를

Table 4.3. Marginal posterior probabilities for equality of successive scores in the drean disturbance data

Row scores	Posterior probability		Column scores	Posterior probability	
	S-D ratio	RJMCMC		S-D ratio	RJMCMC
$\mu_1 < \mu_2$	0.224	0.285	$\nu_1 < \nu_2$	0.999	0.996
$\mu_2 < \mu_3$	0.982	0.940	$\nu_2 < \nu_3$	0.318	0.286
$\mu_3 < \mu_4$	0.402	0.391	$\nu_3 < \nu_4$	0.570	0.484
$\mu_4 < \mu_5$	0.976	0.964			

$\nu_3 = \nu_4$로 대체한 모형을 최적으로 선택하였다. Table 4.3에 주어진 이웃한 스코어 모수의 비동등성에 대한 주변 사후확률 추정치를 보면, 나이 범주의 스코어 모수에 대한 확률은 두 방법이 거의 비슷하다. 그러나 $\nu_3 < \nu_4$에 대한 주변 사후확률은 Savage-Dickey density ratio 방법에서는 0.5보다 크지만 역점프 방법에서는 0.5보다 작은데, 이는 두 방법에서 상위 2개 모형의 순서가 바뀌는 Table 4.2의 모형선택 결과와 부합한다.

5. 토의

이 장에서는 부등 제한조건하에서의 베이지안 추론에서 추후 연구주제가 될 수 있는 것들에 대하여 살펴보기로 하겠다. 기존의 연구에 대해 요구되는 개선책은 각 장에서 언급하였으므로 이 장에서는 전체적인 맥락에서 추후 연구가 필요하다고 생각되는 몇 가지 주제에 대하여 기술해 보기로 한다.

이 논문의 2장과 3장에서 언급한 ANOVA 모형에 대한 베이지안 추론은 회귀 계수에 대한 부등제한 조건이 있는 선형회귀 모형에 대한 추론으로 쉽게 확장시킬 수 있다. 정규분포를 따르지 않는 자료에 대한 일반화 선형모형에서는 적절한 잠재변수를 도입하면 선형모형에서와 같이 2장, 3장의 방법을 적용할 수 있다. 회귀계수에 부등 제한조건이 있는 일반화 선형 모형에 대한 기존의 베이지안 연구로는, 프로빗 모형과 로지스틱 모형에 대한 Oh (2013)의 연구가 있고 RC 모형 또는 RC 모형의 변환인 Row Effect 모형에 대해서는 4장에서 언급한 여러 연구가 있다. 그러나 다른 일반화 선형모형에 대한 연구는 아직 없는 실정이므로 이에 대한 연구가 추후 연구주제가 될 수 있겠다.

이 논문에서 언급한 대부분의 베이지안 추론에서는 계산상의 편리함을 위해서 포괄적 사전분포를 사용하였고 또 사전분포의 형태로 정규분포를 가정하였다. 정규 분포 이외의 사전분포를 사용하면 모수의 조건부 사후분포가 편리한 형태가 아니어서 깁스표본기법이 아닌 메트로폴리스-헤스팅스 기법으로 표본을 추출해야 하는데 메트로폴리스-헤스팅스 기법에서는 표본추출분포를 선택하는데 다소의 조율이 필요해서 비전문가가 사용하기에 어려움이 있을 수 있기 때문이다. 부등 제한 조건의 검정에서 사전분포가 미치는 영향에 대한 Klugkist와 Hoijtink (2007)의 연구에 의하면, 사전분포 형태의 영향은 미미하나 사전분포의 분산은 동등성 검정에 유의한 영향을 미칠 수 있다. 따라서 특별한 이유가 없는한 정규 사전분포의 사용은 무방하나 사전 분산을 변화시키면서 베이지안 검정 결과의 민감성을 체크할 필요가 있겠다.

정규 사전분포를 사용하는 경우 동등성에 대한 베이지안 검정은 사전분산이 커짐에 따라 동등성을 채택할 확률이 증가하는 Lindley's Paradox (1957)에서 자유롭지 못하다. 무정보 사전분포는 좋은 대안이 되지 못하는데, 왜냐하면 부적절성(improper) 때문에 베이즈 상수 계산에 적합치 않기 때문이다. 일반적인 베이지안 검정에서와 마찬가지로 최근 부등 제한조건하에서의 베이지안 검정에서도 객관적인 사전분포(objective prior)와 객관적인 베이즈 상수(objective Bayes factor)에 대한 연구가 활발히 진행되고 있다 (Mulder 등, 2010; Van Wesel 등, 2011; Hoijtink, 2013; Mulder, 2014a, 2014b). 이상의 연구에서 아직 다루지 않은 모형들에서 부등 제한조건을 검정할 때 적절한 객관적 사전분포의 제시, 가능한 객관적 사전분포들의 비교, 객관적 사전분포를 사용했을 때 사후표본 추출을 용이하게 하는 알고리즘의 개

발 등이 추후 연구주제가 될 수 있겠다.

ANOVA모형에서 k개의 그룹평균에 대한 동등성의 모든 가능한 조합의 수는 2^{k-1}이다. ANOVA 또는 회귀모형에서 고려하는 변수가 많을 경우 2^{k-1}가 매우 클 수 있는데, 이 때 기존의 방법들은 모두 계산상의 부담이 커져 실용성이 떨어진다. 계산상의 부담을 완화하여 2^{k-1}가 매우 큰 경우에도 적용 가능한 방법의 개발이 중요한 추후 연구 주제의 하나이다. 이 때 참고할 사항은, 총 2^{k-1} 개의 가능한 모형이 있지만 이 중 대부분의 모형은 적합도가 매우 낮고 단지 소수의 모형만이 자료를 잘 적합하는 모형일 가능성이 높다는 것이다. 따라서 초기에 적합도가 낮은 대부분의 모형을 걸러내고 적합도가 어느 정도 이상 되는 모형들만 본격적으로 비교하는 2단계 방법을 생각해볼 수 있겠다.

분할표 분석에서는, 여러 변수에 대한 다원 분할표(multi-way contingency table) 분석에 대한 연구가 추후 연구주제가 될 수 있겠다. 다원 분할표에서는 동등성의 가능한 조합의 수가 많을 뿐만 아니라 돗수가 0 또는 매우 작은 희박한 셀의 수가 많은 경우가 종종 발생하므로 이 두 문제를 고려한 베이지안 방법의 개발이 요구된다.

부등 제한조건을 반영한 베이지안 추론은 대부분 MCMC 기법을 사용하는데 비전문가가 알고리즘을 스스로 구현하여 사용하기에는 어려움이 따른다. 많은 연구자들이 개발한 프로그램을 요청시 제공하거나 웹사이트에 공개하고 있지만 프로그램에 익숙하지 않은 실무자들이 프로그램을 파악하여 제대로 사용하기는 쉽지 않다. 따라서 기존에 개발된 베이지안 방법을 현장의 실무자가 용이하게 사용할 수 있도록 사용자 중심의 패키지의 개발이 이론적 또는 방법적 연구 못지 않은 중요한 추후 연구주제라고 본다.

References

Akaike, H. (1987). Factor analysis and AIC, *Psychometrika*, **52**, 317–332.

Agresti, A., Chuang, C. and Kezouh, A. (1987). Order-restricted score parameters in association models for contingency tables, *Journal of the American Statistical Association*, **82**, 619–623.

Agresti, A. and Coull, B. A. (2002). The analysis of contingency tables under inequality constraints, *Journal of Statistical Planning and Inference*, **107**, 45–73.

Albert. J. and Chib, S. (1991). Bayesian Analysis of binary and polychotomous response Data, *Journal of the American Statistical Association*, **88**, 669–679

Anraku, K. (1999). An information criterion for parameters under a simple order restriction, *Biometrika*, **86**, 141–152

Bacchetti, P. (1989). Addictive isotonic models, *Journal of the American Statistical Association*, **84**, 289–294.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*, New York, NY: Wiley.

Dickey, J. (1971). The weighted likelihood ration, linear hypotheses on normal location parameters. *The Annals of Statistics*, **42**, 204–223.

Dickey, J. (1976). Approximate posterior distributions, *Journal of the American Statistical Association*, **71**, 680–689.

Dickey, J. and Lientz, B. P. (1970). The weighted likelihood ration, sharp hypotheses about chances, the order of a Markov Chain, *Annals of Mathematical Statistics*, **41**, 214–226.

Dunson, D. B. and Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models, *Biometrics*, **59**, 286–295.

Dykstra, R. L., Robertson, T. and Silvapulle, M. J. (2002). Statistical inference under inequality constraints, Special issue, *Journal of Statistical Planning and Inference*, **107**, 1–2.

Galindo-Garre, F. G.and Vermunt, J. K. (2004). The order restricted association model : Two estimation algorithms and issues in testing, *Psychometrika*, **68**, 614–654.

Galindo-Garre, F. G. and Vermunt, J. K. (2005). Testing log-linear models with inequality constraints: A comparison of asymptotic, bootstrap and posterior predictive p-values, *Statistical Netherlandica*, **59**, 82–94.

Goodman, L. A., (1979). Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association*, **74**, 537–552.

Hayter, A. J. (1990). A one-sided studentized range test for testing against a simple ordered alternative, *Journal of the American Statistical Association*, **85**, 778–785.

Hoijtink, H. (2013). Objective Bayes factors for inequality constrained hypotheses, *International Statistical Review*, **81**, 207–229

Hoijtink, H., Klugkist, I. and Boelen, P. A. (2008). *Bayesian Evaluation of Informative Hypotheses*, Springer, New York.

Iliopoulos, G., Kateri, M. and Ntzoufras, I. (2007). Bayesian estimation of unrestricted and order-restricted association models for two-way contingency table, *Computational Statistics and Data Analysis*, **51**, 4643–4655.

Iliopoulos, G., Kateri, M. and Ntzoufras, I. (2009). Bayesian model comparison for the order-restricted RC association model, *Psychometrika*, **74**, 561–587.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics*, John & Wiley, New York.

Klugkist, I., Kata, B. and Hoijtink, H. (2005). Bayesian model selection using encompassing priors, *Statistica Neerlandica*, **59**, 57–59.

Klugkist, I., Laudy, O. and Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach, *Psychological Methods*, **10**, 477–493.

Klugkist, I. and Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models, *Computational Statistics and Data Analysis*, **51**, 6367–6379.

Kuiper, R. M., Hoijtink, H.J.A. and Silvapulle, M. J. (2011). An Akaike-type information criterion for model selection under inequality constraints, *Biometrika*, **98**, 495–501.

Kuiper, R. M., Hoijtink, H. J. A. and Silvapulle, M. J., (2012). Generalization of the order-restricted information criterion for multivariate normal linear models, *Journal of Statistical Planning and Inference*, **142**, 2454–2463.

Laudy, O. and Hoijtink, H. (2007). Bayesian methods for the analysis of inequality and equality constrained contingency tables, *Statistical Methods in Medical Research*, **16**, 123–138.

Lindley, D. V. (1957). A statistical paradox, *Biometrika*, **44**, 187–192.

Liu, L., Lee, C. C. and Peng, J. (2002). Max-min multiple comparison procedure for isotonic does-response curves, *Journal of Statistical Planning and Inference*, **107**, 133–141.

Liu, L. (2001). Simultaneous statistical inference for monotone dose-response means, Doctoral dissertation, *Memorial University of Newfoundland*, St, John's, Canada.

Marden, J. H. and Allen, L. R. (2002). Molecules, muscles, and machines: Universal performance characteristics of motors. *The National Academy of Sciences USA*, **99**, 4161–4166.

Marden, J. I. and Gao. Y. H.(2002). Rank-based procedures for structural hypotheses of the covariance matrix, *Sankhya*, **64**, 653–677

Marcus, R. and Peritz, E. (1976). Some simultaneous confidence bounds in normal models with restricted alternatives, *Journal of the Royal Statistical Society*, **38**, 157–165.

Marcus, R. (1982). Some results on simultaneous confidence intervals for monotone contrasts in one-way ANOVA model, *Communications in Statistics A*, **11**, 615–622.

Maxwell, I. E. (1961). Maxwell, E. A. Recent trends in factor analysis, *Journal of the Royal Statistical Society*, **124**, 49–59.

Moreno, E. (2005). Objective Bayesian methods for one-sided testing, *Journal of the American Statistical Association*, **14**, 181–198.

Monton-Jones, T., Diggle, P., Parker, L., Dickinson, H. O. and Binks, K. (2000). Addictive isotonic regression models in epidemiology, *Statistics in Medicine*, **19**, 849–859.

Mulder, J. (2014a). Prior adjusted default Bayes factors for testing (in) equality constrained hypotheses, *Computational Statistics and Data Analysis*, **71**, 448–463.

Mulder, J. (2014b). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification, *British Journal of Mathematical and Statistical Psychology*, **67**, 153–171.

Mulder, J., Hoijtink, H. and Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors, *Journal of Statistical Planning and Inference*, **140**, 887–906.

Nashimoto, K. and Wright, F. T. (2005). A note on multiple comparison procedures for detecting difference in simply ordered means, *Statistics and Probability Letters*, 73, 393–401.

Oh, M. S. (1999). Estimation of posterior density functions from a posterior sample, *Computational Statistics and Data Analysis*, **29**, 411–427.

Oh, M. S. and Shin, D. W. (2011). A unified Bayesian inference on treatment means with order constraints, *Computational Statistics and Data Analysis*, **55**, 924–934.

Oh, M. S. (2013). Bayesian multiple comparison of models for binary data with inequality constraints, *Statistics and Computing*, **23**, 481–490.

Oh, M. S. (2014a). Bayesian test one quality of score parameters in the order restricted RC association model, *Computational Statistics and Data Analysis*, **72**, 147–157.

Oh, M. S. (2014b). Bayesian comparison of model swith inequality and equality constraints. *Statistics and Probability Letters*, **84**, 176–182.

Ritov, Y. and Gilula, Z. (1993). The order restricted RC model for ordered contingency tables: estimation and testing for fit, *Annals of Statistics*, **19**, 2090–2101.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, John Wiley and Sons, New York.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.

Shang, J., Cavanaugh, J. E. and Wright, F. T. (2008). A Bayesian multiple comparison procedure for order-restricted mixed models, *International Statistical Review*, **76**, 268–284.

Shayamall, D. P. and Dunson, D. B. (2005). Estimation of order-restricted means from correlated data, *Biometrika*, **92**, 703–715.

Silvapulle, M. J. and Sen, P. K. (2004). *Constrained Statistical Inference*, Wiley, New York.

Tarantola, C., Consonni, G. and Dellaportas, P. (2008). Bayesian clustering of row effects models, *Journal of Statistical Planning and Inference*, **138**, 2223–2235.

Taylor, J. M. G., Wang, L. and Li, Z. (2007). Analysis on binary responses with ordered covariates and missing data, *Statistics in Medicine*, **26**, 3443–3458.

Van Wesel, F., Hoijtink, H. and Klugkist, I. (2011). Choosing priors for constrained analysis of variance: Methods based on training data, *Scandinavian Journal of Statistics*, **38**, 666–690.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio, *Journal of the American Statistical Association*, **90**, 613–618.

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F. and Wagenmakers, E. (2010). Bayesian parameter estimation in the expectancy valence model of the Iowa gambling task, *Journal of Mathematical Psychology*, **54**, 14–27.

Williams, D. A. (1977). Some inference procedures for monotonically ordered normal means, *Biometrika*, **64**, 9–14.

부등 제한 조건하에서의 베이지안 추론

오만숙[a,1]

[a]이화여자대학교 통계학과

(2014년 10월 21일 접수, 2014년 11월 6일 수정, 2014년 11월 10일 채택)

요 약

부등제한 조건 (>,<,=)과 관련된 베이지안 추론에서 다음의 세 가지 주제에 대하여 기존의 연구와 최근의 연구동향 그리고 추후 연구주제에 대하여 살펴보았다 : ⅰ) 모수에 대한 여러 부등제한 조건들의 비교, ⅱ) 모수에 부등제한 조건을 부여하는 것이 타당하다고 할 때 모수의 동등성에 관한 동시 다중 검정, ⅲ) 순서적 범주형 변수에 대한 분할표에서 스코어 모수에 순서적 부등제한 조건을 가정 할 때 스코어 모수의 동등성에 대한 다중 검정.

주요용어: 순서적 제한조건, 다중가설검정, 마코브 체인 몬테칼로, Savage-Dickey 밀도함수 비, 베이즈 상수.

이 연구는 한국 연구재단의 기본연구(2013R1A1A2005481)의 지원에 의한 연구임.

[1](120-750) 서울특별시 서대문구 이화여대길 52, 이화여자대학교 통계학과. E-mail: msoh@ewha.ac.kr