

研究論文

Laplace-Metropolis 알고리즘에 의한 다항로짓모형의 변수선택에 관한 연구

김혜중·이애경

동국대학교 통계학과

Laplace-Metropolis Algorithm for Variable Selection in Multinomial Logit Model

Hea-Jung Kim and Ae-Kyung Lee

Department of Statistics, Dongguk University

Abstract

This paper is concerned with suggesting a Bayesian method for variable selection in multinomial logit model. It is based upon an optimal rule suggested by use of Bayes rule which minimizes a risk induced by selecting the multinomial logit model. The rule is to find a subset of variables that maximizes the marginal likelihood of the model. We also propose a Laplace-Metropolis algorithm intended to suggest a simple method for estimating the marginal likelihood of the model. Based upon two examples, artificial data and empirical data examples, the Bayesian method is illustrated and its efficiency is examined.

1. 서론

범주형자료에 대한 통계적 분석은 공학, 생의학, 사회과학 등의 응용분야에서 최근 널리 사용되고 있으며, 반응범주의 유형에 따라 이항반응자료와 다항반응자료로 구분된다. 이들 중 다항반응자료의 회귀분석에 사용되는 다항반응로짓모형(multinomial response logit model)은 반응변수에 대응하는 공변량의 구조에 따라 다항로짓모형(multinomial logit model)과 일반화로짓모형(generalized logit model)으로 구분된다 (Chung, 1995 참조).

반응변수 $Y_i (i=1, \dots, n)$ 가 $J+1$ 개 범주 중 j 번째 범주의 반응을 보일 때 다항로짓 모형은 아래와 같이 정의한다.

$p(Y_i = j | \beta) =$

$$\frac{\exp(\mathbf{v}_{ij}'\boldsymbol{\delta} + \mathbf{w}_i'\boldsymbol{\gamma}_j)}{\sum_{j=1}^{J+1} \exp(\mathbf{v}_{ij}'\boldsymbol{\delta} + \mathbf{w}_i'\boldsymbol{\gamma}_j)},$$

$$j = 1, 2, \dots, J+1. \quad (1.1)$$

여기서 v_{ij} 는 i 번째 개체가 j 번째 범주를 택할 경우의 특성을 나타내는 $p \times 1$ 인 공변량벡터이고, w_i 는 범주선택에 무관하게 개체 i 의 고유한 특성을 나타내는 $q \times 1$ 인 공변량벡터이다. 그리고, $\beta = (\delta, \gamma_1, \dots, \gamma_J)'$ 는 $(p+Jq) \times 1$ 인 회귀계수벡터이며 모형의 표준화벡터인 $\gamma_{J+1} = 0$ 이다(McFadden, 1974 참조). 이 모형은 반응변수 Y_i 에 대응되는 공변량벡터에 개체 i 의 고유특성 (w_i)과 선택범주 j 를 반영하는 개체 i 의 특성 (v_{ij})을 함께 설명하는 이산형선택모형 (discrete choice model)으로 특히 품질경영 및 마케팅 분야에서 소비자의 의사결정에 대한 성향분석이나 판별분석을 목적으로 자주 사용되며 McFadden 모형 또는 조건부로짓모형(conditional logit model)이라고도 한다 (McFadden, 1974; Chib, Greenberg 및 Chen, 1998 참조).

이러한 특성을 지닌 다항로짓모형의 추정은 McFadden(1974)와 Chib, Greenberg 및 Chen(1998)에 의해 연구되었다. 그러나, 간단한 모형으로 반응변수와 공변량들 간에 관계를 설명하고, 정확한 예측을 위한 최적모형 (best fitted model)의 설정에 필요한 변수선택기준에 대한 연구는 아직 미진한 상태이다. 그 이유는 모형의 우도함수가 복잡하여, 일반화 회귀모형의 변수선택기준으로 널리 사용되는 로그 우도비, 스코어 통계량 및 Wald 통계량의 근사분포에 대한 연구가 없음에 기인한다.

본 논문에서는 이러한 문제점에 착안하여

베이저안 접근법으로 다항로짓모형의 변수선택기준 및 최적의 변수선택법을 제안하고자 한다. 이를 위하여, 본 논문은 Lew와 Raftery(1997)가 제안한 Laplace-Metropolis 알고리즘으로 다항로짓모형의 주변우도함수를 근사추정하는 절차를 제시하고, 추정된 주변우도함수를 변수선택의 기준으로 삼아 베이즈 규칙(Bayes rule)에 근거한 최적의 변수선택방법을 제안하고자 한다. 또한 모의실험과 경험적 자료분석을 통해 제안된 변수선택법의 유용성을 평가할 것이다.

2. 베이저안 변수선택법

식 (1.1)에서 정의된 다항로짓모형의 내포모형(nested model)을 모수벡터 $\beta = (\delta, \gamma_1, \dots, \gamma_J)'$, $(\delta_1, \delta_2, \dots, \delta_p)$ 를 포함하는 집합 B 에 대하여, $(p_{M_k} + Jq) \times 1$, $(p_{M_k} < p)$ 인 모수벡터 $\beta^{M_k} = (\delta^{M_k}, \gamma_1^{M_k}, \dots, \gamma_J^{M_k})'$, $(\beta^{M_k} \subset B)$ 를 가진 모형을 M_k 라 하자. 이에 대응하는 공변량벡터들을 각각 $v_{ij}^{M_k}$ 와 $w_j^{M_k}$ 로 나타내면, 내포모형 M_k 의 우도함수는

$$l(\beta^{M_k} | \mathbf{y}, M_k) = \prod_{i=1}^n \prod_{j=1}^{J+1} \left\{ \frac{\exp(v_{ij}^{M_k} \delta^{M_k} + w_i^{M_k} \gamma_j^{M_k})}{\sum_{j=1}^{J+1} \exp(v_{ij}^{M_k} \delta^{M_k} + w_i^{M_k} \gamma_j^{M_k})} \right\}^{d_{ij}},$$

$$\text{단, } \sum_{j=1}^{J+1} d_{ij} = 1, \quad d_{ij} = \begin{cases} 1, & \text{if } Y_i = j \\ 0, & \text{if } Y_i \neq j \end{cases} \quad (2.1)$$

이다. 여기서 $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ 는 $n \times 1$ 인 반응변수 Y 의 관측값을 나타낸다. 모형 M_k 에 주어진 모수벡터의 사전확률

밀도함수 $p(\beta^{M_k})$ 가 적절한(proper) 경우, β^{M_k} 의 사후확률밀도함수는 다음과 같다.

$$p(\beta^{M_k} | y, M_k) \propto l(\beta^{M_k} | y, M_k) p(\beta^{M_k})$$

$$= \prod_{i=1}^m \prod_{j=1}^{I_i} \left\{ \frac{\exp(v_{ij}^{M_k} \delta^{M_k} + w_i^{M_k} \gamma_j^{M_k})}{\sum_{i=1}^m \exp(v_{ij}^{M_k} \delta^{M_k} + w_i^{M_k} \gamma_j^{M_k})} \right\}^{d_{ij}} \cdot p(\beta^{M_k}). \quad (2.2)$$

그리고, 모형 M_k 의 주변우도함수(marginal likelihood)는

$$p(y | M_k) = \int_{\Omega} l(\beta^{M_k} | y, M_k) p(\beta^{M_k}) d\beta^{M_k},$$

$$\Omega \subset R^{p_{M_k} + J_k}, \quad (2.3)$$

이다.

(정의 1: de Bruijn, 1970). Laplace 근사법은 p -차원 벡터 u 의 함수 $h(u)$ 의 Taylor 급수전개에 의해 다음과 같이 적분을 근사하는 방법이다

$$f(u) = \int \exp\{h(u)\} du$$

$$\approx (2\pi)^{p/2} |A|^{-1/2} \exp\{h(u^*)\}, \quad (2.4)$$

여기서 u^* 는 단봉모형(unimodal)인 $h(u)$ 를 최대화시키는 u 의 값이며, A 는 u^* 에서 계산된 h 의 음의 역헤시안(minus inverse hessian)을 나타낸다.

앞서 정의한 Laplace 근사법을 이용하여 다항로짓모형 (2.1)의 주변우도함수를 근사시키면

$$p(y | M_k) = \int_{\Omega} e^{h(\beta^{M_k})} d\beta^{M_k}, \quad \Omega \subset R^{p_{M_k} + J_k},$$

$$\approx (2\pi)^{(p_{M_k} + J_k)/2} |H^{*M_k}|^{-1/2} l(\beta^{M_k} | y, M_k) \times p(\beta^{M_k})$$

이 되고, 여기서 $h(\beta^{M_k}) = \log\{l(\beta^{M_k} | y, M_k) p(\beta^{M_k})\}$ 이다. 또한 β^{*M_k} 는 $h(\beta^{M_k})$

를 최대화시키는 β^{M_k} 의 값이고, 행렬 H^{*M_k} 은 $\beta^{M_k} = \beta^{*M_k}$ 에서 계산된 $h(\beta^{M_k})$ 의 음의 역헤시안을 나타낸다. 그리고, 이때 발생하는 근사오차는 $O(n^{-1})$ 이 된다(Tierney와 Kadane, 1986 참조). $p(y | M_k)$ 의 Laplace 근사식에 대수를 취하면 다음과 같다.

$$P_{LM_k} = \log p(y | M_k) \quad (2.5)$$

$$\approx \frac{(p_{M_k} + J_k)}{2} \log\{2\pi\} + \frac{1}{2} \log\{|H^{*M_k}|\}$$

$$+ \log\{l(\beta^{*M_k} | y, M_k)\} + \log\{p(\beta^{*M_k})\}.$$

이와 같이, 식 (2.3)인 주변우도함수의 계산에 Laplace 근사법을 이용하면 다항로짓모형의 변수선택에 유용한 다음의 두 가지 결과를 얻을 수 있다.

(결과 1). p_{M_k} 개의 공변량을 가진 다항로

모형 (1.1)의 모든 내포모형은 $\xi = 2^{p_{M_k}} - 1$ 개이고 이들을 각각 모형 M_1, M_2, \dots, M_{ξ} 로 나타내면, 모형의 사전확률이 같다는 가정하에서 최적모형 M_l 은 다음을 만족시킨다.

$$P_{LM_l} = \max_{k=1, \dots, \xi} \{P_{LM_k}\}, \quad l \in \{1, 2, \dots, \xi\}.$$

(증명). 각 모형의 사전확률이 $\pi_k = \Pr(M_k)$, $k = 1, 2, \dots, \xi$, 일 때 모형을 다항로짓모형으로 적합시킬 경우의 위험(risk)은 다음

과 같다.

$$R(M_l | y) = \frac{\sum_{k=1}^{\xi} L(M_k, M_l) \pi_k p(y | M_k)}{\sum_{j=1}^{\xi} \pi_j p(y | M_j)},$$

$l = 1, 2, \dots, \xi$.

여기서, 손실함수 $L(M_k, M_l)$ 을 일반적으로 사용하는 0-1 손실함수인 $L(M_k, M_l) =$

$1 - \delta(M_k, M_l)$, 즉 $M_k = M_l$ 이면 $\delta(M_k, M_l) = 1$ 이고, $M_k \neq M_l$ 인 경우 $\delta(M_k, M_l) = 0$ 인 손실함수를 사용하면, 최소의 위험을 가진 베이즈규칙(Bayes rule)은 모형들 중에서 $\pi_k p(\mathbf{y} | M_k)$ 값을 최대로 하는 모형 M_l 을 최적모형으로 택하는 것이다. 그러므로, 모형의 사전확률이 모두 같다는 가정 하에서 $\log p(\mathbf{y} | M_k) \approx P_{LM_k}$ 을 적용하면 위 결과가 성립된다.

(결과 2). 각 모형의 사전확률이 $\pi_k = \Pr(M_k)$, $k = 1, 2, \dots, \xi$, 일 때 모형 M_k 가 다항로짓모형으로 적합될 사후확률은 다음과 같다.

$$\begin{aligned} \Pr(M_k | \mathbf{y}) &= \frac{\pi_k p(\mathbf{y} | M_k)}{\sum_{j=1}^{\xi} \pi_j p(\mathbf{y} | M_j)} \\ &\approx \frac{\pi_k (2\pi)^{p_{M_k}} |H^{*M_k}|^{1/2} l(\beta^{*M_k} | \mathbf{y}, M_k)}{\sum_{j=1}^{\xi} \{\pi_j (2\pi)^{p_{M_j}} |H^{*M_j}|^{1/2} l(\beta^{*M_j} | \mathbf{y}, M_j)\}} \\ &\quad \cdot \frac{p(\beta^{*M_k})}{p(\beta^{*M_k})}, \quad k = 1, 2, \dots, \xi. \end{aligned}$$

그러므로, 주어진 자료로부터 β^{*M_k} 와 H^{*M_k} , $k = 1, \dots, \xi$, 값을 추정하면, (결과 1)에 의해 최적다항로짓모형을 찾을 수 있을 뿐 아니라 그 모형이 적합될 사후확률도 (결과 2)로부터 얻을 수 있다.

3. Laplace Metropolis 알고리즘

변수선택 기준인 대수주변우도함수의 Laplace 근사식 (2.5)를 계산하는데는 β^{*M_k} 와 H^{*M_k} 의 추정이 필요하다. 이들은 β^{M_k} 의 사

후확률밀도함수로부터 추정된다. 그러나, 식 (2.2)와 같이 복잡한 사후확률밀도함수로부터 β^{*M_k} 와 H^{*M_k} 를 분석적인 방법으로 추정하는 것은 사실상 어렵다. 최근에 들어 McFadden (1989), Albert와 Chib(1993), McCulloch, Polson 및 Rossi(1994) 그리고 Stern(1997)등에 의해 전통적인 방법 (classical method)과 모의 표본을 이용한 베이즈 방법(simulation based Bayes method)으로 다항로짓모형을 유효하게 추정하는 방법들이 다양하게 개발되었으나, 본 논문은 Gelfand와 Smith(1990), Chib와 Greenberg(1996)등이 여러 모형의 추정에 사용한 MCMC(Markov Chain Monte Carlo) 방법을 다항로짓모형의 추정에 맞게 변형시켜 다항로짓모형 및 주변우도함수의 Laplace 근사식을 추정하는 방법을 개발하고자 한다. 참고로, MCMC방법에 의한 모형의 주변우도함수추정에는 Chib(1995)이 제안한 Gibbs 표본 이용법이 대표적이거나, 다항로짓모형과 같이 단일 모수벡터 β^{M_k} 를 가진 모형의 추정에는 적당하지 않다. 또한, Gibbs 표본법은 일반적으로 반복적인 모의 실험에 근거하므로 사후표본에 이상치들이 포함될 가능성이 있다. 이런 경우 사후표본의 평균으로 β^{*M_k} 와 H^{*M_k} 를 추정하여 (2.5)를 계산하면, (2.3)식에 정의된 모형 M_k 의 주변우도함수의 근사 추정이 이상치에 로버스트(robust)하지 못한 결과를 초래하게 된다.

본 장에서는 이러한 문제점들을 고려하여 Metropolis-Hastings 알고리즘으로 (2.3)식을 유효하게 추정하는 방법을 제안하고자 한다.

3.1 Metropolis-Hastings 알고리즘

Metropolis-Hastings 알고리즘을 이용하여

(2.3)식에 정의된 모형 M_k 의 주변우도함수를 근사 추정하기 위해서는 마야코프 과정에서 정의된 생성밀도함수(generating density)의 선택이 필요하다. 일반적으로 생성밀도함수는 위치모수와 척도모수를 가진 분포족(family of distribution)에 의해 정의되며, 이러한 분포족으로는 여러 가지 형태가 제안되어 있다(Chib와 Greenberg, 1995).

이 절에서는 β^{*m} 와 H^{*m} 의 추정에 Chib와 Greenberg(1996)가 제안한 재단된 독립과정(tailored independence chain)하에서 도출된 생성밀도함수를 사용하여 Metropolis-Hastings 알고리즘을 구성하고자 한다. 생성밀도함수로는 다변량정규분포 $f_N(\beta^{M_k} | \mu^*, \tau^2 \Sigma)$ 을 사용하며, μ^* 는 $\log p(\beta^{M_k} | \mathbf{y}, M_k)$ 를 최대화시키는 최빈수이고, Σ 는 이 값에서 계산된 $\log p(\beta^{M_k} | \mathbf{y}, M_k)$ 의 음의 역헤시안인 $-\left(\frac{\partial^2 \log p(\beta | \mathbf{y}, M_k)}{\partial \beta \partial \beta'}\right)^{-1}$ 이다.

그리고, τ 는 표본채택률(acceptance rate)를 조절하기 위한 조율요인(tuning factor)이다(Roberts, Gelman, Gilks, 1994; Muller, 1993 참조). 선택된 생성밀도함수를 이용하면 초기값 $\beta^{M_k^{(0)}}$ 이 주어진 재단된 독립과정하에서 (t) 상태(state)에서 $(t+1)$ 상태로 전이되는 $\beta_t^{M_k}$ 은

$$\beta_t^{M_k} = \mu^* + z, \quad z \sim N_{p_{M_k}}(0, \tau^2 \Sigma), \quad (3.1)$$

의 과정으로 설정된다. 따라서, β^{M_k} 의 사후표본은 식 (3.1)에 의해 생성된 $\beta_t^{M_k}$, $t = 0, 1, \dots, N$, 에 대해 이동확률(probability of move) $\alpha(\beta^{M_k^{(t)}}, \beta_t^{M_k})$ 를 고려하여

다음과 같이 Metropolis-Hastings 알고리즘을 구성할 수 있다.

(단계1). 생성밀도함수인

$f_N(\beta^{M_k} | \mu^*, \tau^2 \Sigma)$ 로 부터 $\beta_t^{M_k}$ 를 발생시킨다.

(단계2). 균등분포 $U(0, 1)$ 으로부터 u 를 발생시킨다.

(단계3). 이동확률

$$\alpha(\beta^{M_k^{(t)}}, \beta_t^{M_k}) = \min \{1, \frac{p(\beta_t^{M_k} | \mathbf{y}, M_k) f_N(\beta^{M_k^{(t)}} | \mu^*, \tau^2 \Sigma)}{p(\beta^{M_k^{(t)}} | \mathbf{y}, M_k) f_N(\beta_t^{M_k} | \mu^*, \tau^2 \Sigma)} \}$$

에 의해 (t) 상태에서 $(t+1)$ 상태로 전이하는 $\beta^{M_k^{(t+1)}}$ 는 다음 조건에 의해 선택된다.

(i) $u \leq \alpha(\beta^{M_k^{(t)}}, \beta_t^{M_k})$ 이면 이동확률 $\alpha(\beta^{M_k^{(t)}}, \beta_t^{M_k})$ 을 갖고 $\beta^{M_k^{(t+1)}} = \beta_t^{M_k}$ 를 채택하고,

(ii) $u > \alpha(\beta^{M_k^{(t)}}, \beta_t^{M_k})$ 이면 이동확률 $1 - \alpha(\beta^{M_k^{(t)}}, \beta_t^{M_k})$ 을 갖고 $\beta^{M_k^{(t+1)}} = \beta^{M_k^{(t)}}$ 를 채택한다.

이때 생성밀도함수인 $f_N(\beta^{M_k^{(t)}}, \beta^{M_k^{(t+1)}})$ 는 다음과 같은 가역조건(Tierney, 1994 참조)을 만족한다.

$$p(\beta^{M_k^{(t)}} | \mathbf{y}, M_k) f_N(\beta^{M_k^{(t)}}, \beta^{M_k^{(t+1)}}) \alpha(\beta^{M_k}, \beta^{M_k^{(t+1)}}) \\ = p(\beta^{M_k^{(t+1)}} | \mathbf{y}, M_k) f_N(\beta^{M_k^{(t+1)}}, \beta^{M_k^{(t)}}) \alpha(\beta^{M_k^{(t+1)}}, \beta^{M_k})$$

3.2 알고리즘의 수렴성 진단

앞 절에서 제안된 Metropolis-Hastings 알고리즘을 반복 시행하여 얻은 모형 M_k 에 대한 β^{M_k} 의 사후표본을

$$\{\beta^{M_k(1)}, \beta^{M_k(2)}, \beta^{M_k(3)}, \dots, \beta^{M_k(t)}\},$$

$$k = 1, 2, \dots, \xi. \quad (3.2)$$

이라고 하자. 그러면, 에르고딕 정리(ergodic theorem)에 의해 $\beta^{M_k(t)}$ 의 결합사후분포는 β^{M_k} 의 결합사후분포로 접근하게 된다. 따라서 충분히 큰 $t = t^*$ 에 대해 $\beta^{M_k(t^*)}$ 의 값은 β^{M_k} 의 사후분포로부터 생성된 값으로 간주된다(Tierney, 1994 참조). 이렇게 수렴성을 만족하는 사후표본을 얻기 위해서는 t^* (burn-in time)의 결정이 필요하다. t^* 의 결정은 사후표본의 수렴여부를 진단하기 위해 제안된 다음과 같은 진단법들을 사용할 수 있다.

방법1) 정상분포(stationary distribution)로부터 멀리 떨어져 있는 초기값을 다양하게 변화시킨 다중체인을 고려하여 $-2 \log$ (사후확률밀도함수)의 trace plot과 커널함수로 β^{M_k} 의 사후표본에 대한 수렴성을 확인하는 방법(Cowles와 Carlin, 1996 참조).

방법2) Gelman와 Rubin(1992)에 의해 제안된 $-2 \log$ (사후확률밀도함수) 값에 대한 shrink factor 그래프와 shrink factor 값으로 수렴성을 확인하는 방법.

여기서 식(2.2)에 의해

$$\log(\text{사후확률밀도함수}) =$$

$$\log \left[\prod_{i=1}^n \prod_{j=1}^{t+1} \left\{ \frac{\exp(\mathbf{v}_{ij}' \boldsymbol{\delta}^{M_k} + \mathbf{w}_i' \boldsymbol{\gamma}_j)}{\sum_{j=1}^{t+1} \exp(\mathbf{v}_{ij}' \boldsymbol{\delta}^{M_k} + \mathbf{w}_i' \boldsymbol{\gamma}_j)} \right\}^{d_{ij}} \cdot p(\boldsymbol{\beta}^{M_k}) \right] \Big|_{\boldsymbol{\beta}^{M_k} = \boldsymbol{\beta}^{M_k(t^*)}$$

단, $t' \in \{t \mid t \geq t^*\}$ 이다.

위와 같은 진단법을 사용함으로써 Metropolis-Hastings 알고리즘에 의해 생성된

β^{M_k} 의 사후표본에 대한 수렴여부를 확인할 수 있고, 수렴하는 시점에 의해 t^* 의 값이 결정된다.

3.3 알고리즘의 로버스트성

제안된 알고리즘에서 얻은 β^{M_k} 의 사후표본인 (3.2)에는 이상치들이 포함될 가능성이 있다(Lewis와 Raftery, 1997 참조). 이러한 사후표본으로 모형 M_k 의 주변우도함수를 근사 추정하면 추정결과가 이상치에 로버스트하지 못한 경우를 초래한다. 이러한 문제점은 Small(1990)이 제안한 사후표본의 L_1 중심(L_1 center)을 사용하여 해결할 수 있다. 즉, 이상치에 로버스트한 $\beta^{*M_k} = (\beta_1^{*M_k}, \dots, \beta_{p_{M_k}}^{*M_k})'$ 는

$$d(\boldsymbol{\beta}^{M_k(r)}) = \sum_{i=t^*+1}^N \sum_{j=1}^s |\beta_j^{M_k(i)} - \beta_j^{M_k(r)}|,$$

$$r = t^* + 1, \dots, N, \quad s^* = p_{M_k},$$

를 최소화하는 $\beta^{M_k(r)}$ 값으로 구할 수 있다.

여기서 t^* 는 사후표본의 수렴시점을 나타내고, $|\cdot|$ 는 절대거리(absolute metric)를 의미한다. 또한, 모형 M_k 의 주변우도함수의 근사추정에 필요한 행렬 H^{*M_k} 의 값은 β^{M_k} 의 사후확률밀도함수를 이용하여 수치적인 방법으로 구한다. 일반적으로 대표본인 경우에는 최우추정량 β^{M_k} 의 점근적인 정규성(asymptotic normality)에 의해 사후표본의 표본공분산행렬을 이용하여 H^{*M_k} 를 추정할 수 있으나, 사후표본에 이상치가 존재하는 경우에는 로버스트성을 만족하지 못하므로 본 논문에서는 Rousseeuw와 van Zomeren(1990)이 제안한 최소부피타원체추정량에 근거한 가중

<표 1> 식 (4.1)에 정의된 참 모형이 M_6 일 때, 제안된 변수선택기준으로 선택된 최적모형에 대한 모의실험 결과

| | 모 형 | | | | | | |
|--------------------|------------------------|----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | $M_1(x_1)$ | $M_2(x_2)$ | $M_3(x_3)$ | $M_4(x_1x_2)$ | $M_5(x_1x_3)$ | $M_6(x_2x_3)$ | $M_7(x_1x_2x_3)$ |
| 모형선택횟수 및 비율 | 0(0.0) | 0(0.0) | 0(0.0) | 0(0.0) | 0(0.0) | 194(0.97) | 6(0.03) |
| 모형의 사후확률 (표준오차) | 9.10E-14 (7.46E-13) | 1.21E-9 (9.10E-9) | 0.025723 (0.156607) | 9.12E-11 (6.83E-10) | 0.000088 (0.000687) | 0.850126 (0.197815) | 0.124063 (0.144348) |

된 분산행렬을 이용하여 H^{M_k} 의 로버스트 추정값을 구한다.

따라서, 위와 같은 추정법에 의해 추정된 β^{M_k} 와 H^{M_k} 의 값을 β^{*M_k} 와 H^{*M_k} 로 나타내면 식 (2.5)에 정의된 Laplace 근사식의 추정값 \hat{P}_{LM_k} 은

$$\begin{aligned} \hat{P}_{LM_k} &= \log p(\mathbf{y} | M_k) & (3.3) \\ &\approx \frac{p_{M_k} + Jq}{2} \log \{2\pi\} + \frac{1}{2} \log \{ |H^{*M_k}| \} \\ &\quad + \log \{ l(\beta^{*M_k} | \mathbf{y}, M_k) \} + \log \{ p(\beta^{*M_k}) \}. \end{aligned}$$

가 된다.

그러므로, \hat{P}_{LM_k} 을 2장에서 유도한 (결과 1)의 P_{LM_k} 에 대입을 시키면 최적의 모형 M_l 을 선택할 수 있고, 이때 M_l 에 포함된 변수들은 다항로짓모형의 최적변수군이 된다. 같은 방법에 의해 (결과 2)로부터 각 모형이 자료에 적합될 사후확률을 추정할 수 있다.

4. 모의실험 및 경험적 자료분석

이 장에서는 주변우도함수를 Laplace근사와 Metropolis 알고리즘으로 추정하여 구한 변수선택기준이 실제로 자료분석에 유용한지를 살펴보기 위해 모의실험과 경험적인 예제를 통하여 그 유용성을 검토하고자 한다.

4.1 모의실험

모의실험에 사용된 다항로짓모형은 다음과 같다.

$$p(Y_i = j | \beta) = \frac{\exp(\mathbf{v}_{ij}'\delta + w_i\gamma_j)}{\sum_{j=1}^4 \exp(\mathbf{v}_{ij}'\delta + w_i\gamma_j)}, \quad j = 1, 2, 3, 4, \quad i = 1, 2, \dots, 50. \quad (4.1)$$

여기서 반응변수 Y_i 의 영역은 네 가지 범주의 반응값인 {1, 2, 3, 4}이며, 회귀계수 $\beta = (\delta_1, \delta_2, \delta_3, \gamma_1, \gamma_2, \gamma_3)'$ = (0, 2, 3, 5, 4, 2)'로 놓았다. 위 모형에서 증화된 공변량 $\mathbf{v}_{ij} = (v_{ij1}, v_{ij2}, v_{ij3})'$, $i = 1, 2, \dots, 50$; $j = 1, 2, 3, 4$,의 관측값은 $N(0, I_{j3})$, $I_{j3} = \text{diag}(1, 1, 1)$ 로부터 발생시켰으며, $w_i = 1$ 로 정의하였다.

이와 같이 얻은 자료를 이용하여 식 (4.1)에서 정의되는 모든 내포모형 M_k , $k = 1, 2, \dots, 7$,의 주변우도함수를 추정하고, 이것을

변수선택기준에 적용하여 (결과 1)을 이용하여 최적모형을 찾는 방법을 200번 반복한 결과는 <표 1>과 같다. <표 1>에 의하면, 200번의 모의 실험 중 참 모형인 M_6 을 최적모형으로 선택된 비율과 사후확률이 타 모형에 비해 월등히 큰 값을 나타내고 있어, 제안된 변수선택법이 유효한 것임을 보여 주고 있다.

모의실험에서 모든 내포모형에 포함된 모수벡터 β^{M_k} 의 사전확률분포는 $p(\beta^{M_k}) \sim N(0, 100I_{p_{M_k}})$, 으로 가정하여 모형의 선택에 사전확률분포의 영향을 줄이고, 자료에 의해 모형이 선택되도록 하였다. 여기서 $I_{p_{M_k}}$

< 그림 1 >은 임의로 주어진 모수의 초기값에 의해 얻은 4개의 다중체인으로부터 추정된 β^{M_6} 의 사후밀도함수에 $-2 \log$ 를 취해 얻어진 결과로 그림 (a)에는 각 체인의 trace 변화를 그래프로 나타낸 trace plot과 모든 체인으로부터 추정된 사후밀도함수의 커널을 보여준다. 그림 (b)는 추정된 shrink factor 그래프로서, shrink factor 값이 1.03이 되어 사후표본이 실사후분포에 수렴하고 있음을 보여주고 있다(shrink factor 값이 1일 경우 완전수렴을 나타냄).

위와 같은 진단결과 수렴 시점은 $t^* = 2000$ 이었으며, t^* 개를 제거하고 남은 $N = 10000$ 개의 사후표본들을 주변우도함수

<표 2> 경험적 자료의 210개 관측값 중에서 첫 번째, 두 번째 자료

| y_i | TIME | INVT | INVC | GC | H1 | P1 | IND1 (w_1) | IND2 (w_2) | IND3 (w_3) |
|-------|------|------|------|----|----|----|-------------------|-------------------|-------------------|
| 4 | 69 | 59 | 100 | 70 | 35 | 1 | 1 | 0 | 0 |
| 4 | 34 | 31 | 372 | 71 | 0 | 0 | 0 | 1 | 0 |
| 4 | 35 | 25 | 417 | 70 | 0 | 0 | 0 | 0 | 1 |
| 4 | 0 | 10 | 180 | 30 | 0 | 0 | 0 | 0 | 0 |
| 4 | 64 | 58 | 68 | 68 | 30 | 2 | 1 | 0 | 0 |
| 4 | 44 | 31 | 354 | 84 | 0 | 0 | 0 | 1 | 0 |
| 4 | 53 | 25 | 399 | 85 | 0 | 0 | 0 | 0 | 1 |
| 4 | 0 | 11 | 255 | 50 | 0 | 0 | 0 | 0 | 0 |

는 $p_{M_k} \times p_{M_k}$ 인 단위행렬이다. 또한, 제안된 알고리즘으로부터 얻은 사후표본의 수렴성 진단은 3.2절에서 논의된 Gelman 과 Rubin(1992)의 수렴성 진단법으로 실시하였다. 진단에 필요한 계산은 Best, Cowles와 Vines(1996)에 의해 개발된 S-PLUS함수인 CODA 함수를 사용하였으며, <그림 1>은 그 함수로부터 계산된 결과이다.

추정에 사용하였다. 그리고, 사후표본의 표본채택률은 모수의 차원이 1인 경우는 45%, 차원이 6 이하인 경우는 25% 그리고 차원이 6 이상인 경우는 23% 정도로 표본이 채택되도록 τ 값을 각 차원에 따라 0.55, 0.35, 0.35으로 설정하였다(Roberts, Gelman, Gilks, 1994; Muller, 1993 참조).

<표 3> 주변우도함수 Laplace 근사 추정값 (\hat{P}_{LM_k})에 의한 변수선택 결과
(괄호 안의 값은 $\hat{P}_{M_k} = \Pr(M_k | y)$ 의 근사값)

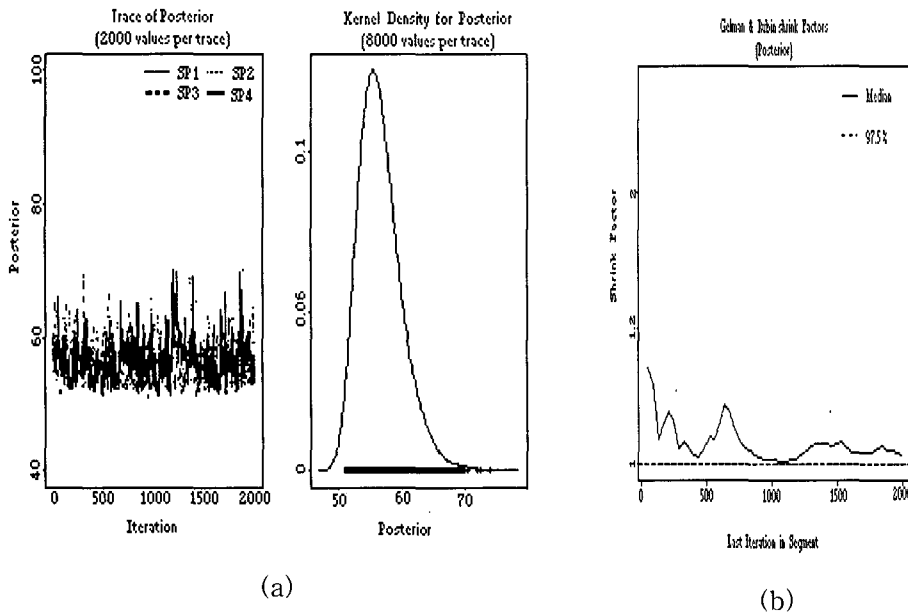
| 모형 | $\hat{P}_{LM_k} / \hat{P}_{M_k}$ | 모형 | $\hat{P}_{LM_k} / \hat{P}_{M_k}$ | 모형 | $\hat{P}_{LM_k} / \hat{P}_{M_k}$ |
|-----------------|----------------------------------|-----------------------------|----------------------------------|---|----------------------------------|
| $\nu_1 W$ | -225.6663 (0.0067133) | $\nu_1 \nu_2 \nu_3 W$ | -231.2577 (0.000025) | $\nu_1 \nu_2 \nu_3 \nu_5 W$ | -226.1065 (0.0043228) |
| $\nu_2 W$ | -335.9043 (8.937E-51) | $\nu_1 \nu_2 \nu_4 W$ | -235.0147 (5.8476E-7) | $\nu_1 \nu_2 \nu_3 \nu_6 W$ | -245.8885 (1.108E-11) |
| $\nu_3 W$ | -225.7628 (0.0060958) | $\nu_1 \nu_2 \nu_5 W$ | -239.5735 (6.1252E-9) | $\nu_1 \nu_2 \nu_4 \nu_5 W$ | -227.5552 (0.0010153) |
| $\nu_4 W$ | -225.7128 (0.0064083) | $\nu_1 \nu_2 \nu_6 W$ | -229.3815 (0.0001635) | $\nu_1 \nu_2 \nu_4 \nu_6 W$ | -234.8110 (7.1688E-7) |
| $\nu_5 W$ | -225.6708 (0.0066832) | $\nu_1 \nu_3 \nu_4 W$ | -233.2820 (3.3074E-6) | $\nu_1 \nu_2 \nu_5 \nu_6 W$ | -227.5552 (1.7326E-7) |
| $\nu_6 W$ | -225.7215 (0.0063528) | $\nu_1 \nu_3 \nu_5 W$ | -230.0825 (0.0000811) | $\nu_1 \nu_3 \nu_4 \nu_5 W$ | -238.1920 (2.4383E-8) |
| $\nu_1 \nu_2 W$ | -233.4548 (2.7825E-6) | $\nu_1 \nu_3 \nu_6 W$ | -222.2268 (0.2092637) | $\nu_1 \nu_3 \nu_4 \nu_6 W$ | -227.0434 (0.0016938) |
| $\nu_1 \nu_3 W$ | -225.2569 (0.0101097) | $\nu_1 \nu_4 \nu_5 W$ | -235.5223 (3.5199E-7) | $\nu_1 \nu_3 \nu_5 \nu_6 W$ | -227.2789 (0.0013384) |
| $\nu_1 \nu_4 W$ | -228.5860 (0.0003622) | $\nu_1 \nu_4 \nu_6 W$ | -221.1205 (0.632636) | $\nu_1 \nu_4 \nu_5 \nu_6 W$ | -225.9521 (0.0050445) |
| $\nu_1 \nu_5 W$ | -231.7204 (0.0000158) | $\nu_1 \nu_5 \nu_6 W$ | -228.6115 (0.0003531) | $\nu_2 \nu_3 \nu_4 \nu_5 W$ | -303.6990 (8.665E-37) |
| $\nu_1 \nu_6 W$ | -222.9541 (0.1011187) | $\nu_2 \nu_3 \nu_4 W$ | -305.5663 (1.339E-37) | $\nu_2 \nu_3 \nu_4 \nu_6 W$ | -306.7778 (3.987E-38) |
| $\nu_2 \nu_3 W$ | -294.8861 (5.823E-33) | $\nu_2 \nu_3 \nu_5 W$ | -299.4069 (6.336E-35) | $\nu_2 \nu_3 \nu_5 \nu_6 W$ | -297.3860 (4.78E-34) |
| $\nu_2 \nu_4 W$ | -302.0717 (4.411E-36) | $\nu_2 \nu_3 \nu_6 W$ | -296.9988 (7.041E-34) | $\nu_2 \nu_4 \nu_5 \nu_6 W$ | -298.2089 (2.099E-34) |
| $\nu_2 \nu_5 W$ | -303.7925 (7.892E-37) | $\nu_2 \nu_4 \nu_5 W$ | -302.7645 (2.206E-36) | $\nu_3 \nu_4 \nu_5 \nu_6 W$ | -298.9742 (9.766E-35) |
| $\nu_2 \nu_6 W$ | -303.2821 (1.315E-36) | $\nu_2 \nu_4 \nu_6 W$ | -296.7155 (9.347E-34) | $\nu_1 \nu_2 \nu_3 \nu_4 \nu_5 W$ | -233.5089 (2.636E-6) |
| $\nu_3 \nu_4 W$ | -297.6287 (3.75E-34) | $\nu_2 \nu_5 \nu_6 W$ | -304.7551 (3.014E-37) | $\nu_1 \nu_2 \nu_3 \nu_4 \nu_6 W$ | -232.0057 (0.0000119) |
| $\nu_3 \nu_5 W$ | -294.1466 (1.22E-32) | $\nu_3 \nu_4 \nu_5 W$ | -301.0677 (1.204E-35) | $\nu_1 \nu_2 \nu_3 \nu_5 \nu_6 W$ | -231.0377 (0.0000312) |
| $\nu_3 \nu_6 W$ | -296.6991 (9.501E-34) | $\nu_3 \nu_4 \nu_6 W$ | -299.7463 (4.512E-35) | $\nu_1 \nu_2 \nu_4 \nu_5 \nu_6 W$ | -233.1823 (3.6541E-6) |
| $\nu_4 \nu_5 W$ | -295.1209 (4.605E-33) | $\nu_3 \nu_5 \nu_6 W$ | -295.4668 (3.258E-33) | $\nu_1 \nu_3 \nu_4 \nu_5 \nu_6 W$ | -232.7538 (5.6089E-6) |
| $\nu_4 \nu_6 W$ | -290.4502 (4.916E-31) | $\nu_4 \nu_5 \nu_6 W$ | -291.8953 (1.159E-31) | $\nu_2 \nu_3 \nu_4 \nu_5 \nu_6 W$ | -307.4341 (2.068E-38) |
| $\nu_5 \nu_6 W$ | -299.4733 (5.929E-35) | $\nu_1 \nu_2 \nu_3 \nu_4 W$ | -229.5375 (0.0001399) | $\nu_1 \nu_2 \nu_3 \nu_4 \nu_5 \nu_6 W$ | -235.9185 (2.3684E-7) |

4.2 경험적 자료분석

이 절에서는 본 논문에서 제안된 변수선택 기준을 Chib, Greenberg 및 Chen(1998)에 수록된 자료에 적용하였다.

이 자료는 Sydney, Melbourne, New South Wales 그리고 Australia를 통행할 때

사용되는 교통수단을 4가지의 범주(비행기(1), 기차(2), 버스(3), 자가용(4))로 분류하여 개체들(운전자)이 어느 범주(교통수단)를 선택하는가를 알아보기 위한 선택형자료이다. 반응변수 $Y_i, i = 1, 2, \dots, 210$ 의 관측값은 4가지의 교통수단을 의미하는 다항범주



<그림 1> (a) $-2 \log$ (사후확률밀도함수)에 대한 trace plot과 커널함수
 (b) $-2 \log$ (사후확률밀도함수)에 대한 shrink factor.

1, 2, 3, 4으로 관측되었으며, 이에 대응하는 공변량벡터 (v_{ij}', w_i') , $j = 1, 2, 3, 4$ 의 관측값은 다음과 같다.

- v_{ij1} : 고속터미널대기시간 (TIME)
- v_{ij2} : 운송시간 (INVT)
- v_{ij3} : 운송비용 (INVC)
- v_{ij4} : 일반화된 비용척도 (GC)
- v_{ij5} : 가사수입 · 비행기 (HI)
- v_{ij6} : 여행자 그룹규모 · 비행기(P1)

여기서 v_{ijp} , ($p=1, 2, \dots, 6$)는 개체 i 가 j 번째 범주를 택할 경우의 특성을 나타내는 p 번째 공변량값을 나타낸다. 위 공변량에서 고속터미널대기시간, 운송시간, 운송비용, 일

반화된 비용척도는 4가지 교통수단을 의미하는 다항범주 (1,2,3,4)의 특성을 나타내는 변량이고, (가사수입 · 비행기)와 (여행규모 · 비행기)는 개체특성을 나타내는 가사수입과 여행규모를 범주 (비행기=1)의 특성으로 변화시킨 공변량을 나타낸다. 그리고 개체 i 의 고유의 특성을 나타내는 공변량 w_i 의 값은 1이다. 그러므로, 공변량 벡터 $x_{ij}' = (v_{ij}', w_i)$ 는 다음과 같이 정의된다.

< 표 2 >는 앞서 정의된 공변량의 총화된 구조에 대한 이해를 돕기 위해 다항로짓모형의 경험적 자료분석에 사용된 실제 자료의 첫 번째와 두 번째 관측값을 나타낸다.

$$x_{ij}' = \begin{cases} (v_{i1}, 1, 0, 0) & \text{if } j=1 \\ (v_{i2}, 0, 1, 0) & \text{if } j=2 \\ (v_{i3}, 0, 0, 1) & \text{if } j=3 \\ (v_{i4}, 0, 0, 0) & \text{if } j=4 \end{cases}$$

따라서 반응변수 Y_i 에 대한 j 번째 범주에 대한 조건부반응확률은

$$p(Y_i = j | \beta) = \frac{\exp(v_{ij}'\delta + w_i\gamma_j)}{\sum_{j=1}^4 \exp(v_{ij}'\delta + w_i\gamma_j)} = \frac{\exp(x_{ij}'\beta)}{\sum_{j=1}^4 \exp(x_{ij}'\beta)},$$

$$j = 1, 2, 3, 4, i = 1, 2, \dots, 210. \quad (4.2)$$

으로 정의되며, $\beta = (\delta, \gamma_1, \gamma_2, \gamma_3)'$, $\delta = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6)'$ 는 9×1 인 회귀계수 벡터이고 표준화 계수 $\gamma_4 = 0$ 이다. 식 (4.2)로부터 정의되는 $(2^6 - 1)$ 개의 모든 내포모형 M_k 에 포함된 모수벡터 β^{M_k} 의 사전확률분포는 $p(\beta^{M_k}) \sim N(0, 100I_{p_{M_k}})$, 를 사용하였으며, 이 분포를 이용하여 식(2.3)와 (2.5)를 유도하였다.

먼저, 제안된 Metropolis-Hastings 알고리즘에 의해 추출된 사후표본의 수렴성에 대한 진단은 앞의 모의실험과 같은 방법으로 검토하였다. 그 결과 $t^*=2000$ 에서 모든 내포모형 M_k 에 대한 β^{M_k} 의 사후표본은 수렴하는 것으로 나타났다.

따라서, Metropolis-Hastings 알고리즘을 이용하여 추출된 사후표본의 수렴 시점은 $t^* = 2000$ 으로 정하고, $(t^* + N)$ 개 추출된 사후표본에서 t^* 개를 제거하고 남은 $N =$

10000개를 주변우도함수 추정에 사용할 사후표본으로 설정하였다.

사후표본의 표본채택률은 모수의 차원이 1인 경우는 45%, 차원이 6 이하인 경우는 25% 그리고 차원이 6 이상인 경우는 23% 정도로 표본이 채택되도록 τ 값을 각각 0.45, 0.35 그리고 0.20으로 조정하였다. 이와 같은 절차로 얻은 (3.3)식의 주변우도함수 추정값 (\hat{P}_{LM_k})와 모형의 사전확률 $\Pr(M_k)$ 이 동일하다는 조건하에서 구한 모형의 사후확률 $p(M_k | y)$ 의 근사값을 <표 3>에 나타내었다. <표 3>의 결과에 의하면, {터미널대기시간 (v_1), 일반화된 비용척도 (v_4), 여행자 규모 · 비행기 (v_6), IND1, IND2, IND3 }을 공변량으로 가진 모형이 자료에 가장 적합한 모형으로 나타났다.

5. 결론

본 논문은 베이지 접근법으로 다항로짓모형에 대한 변수선택방법을 제시하고, Lewis와 Raftery(1997)에 의해 제안된 Laplace Metropolis 방법을 이용하여 변수선택기준을 추정하는 방법을 제안하였다. 제안된 추정법은 다항로짓모형이 가진 주변우도함수의 Laplace 근사식을 제안된 독립과정(tailored independence chain)하에서 Metropolis-Hastings 알고리즘으로 추정하는 방법이다. 이 방법은 기존의 것과 달리 다음과 같은 두 가지 유용성을 갖고 있다. 첫째로, 모형의 사후확률밀도함수로부터 직접 주변우도함수를 근사 추정할 수 있으며, 둘째로 사후표본의 L_1 중심과 역헤시안 추정량을 이용하여 주

변우도함수를 추정하므로 이상치에 로버스트한 추정을 가능케 한다.

또한, 모의실험과 경험적인 자료분석을 통해 제안된 변수선택법의 유효성을 평가한 결과 두 자료분석에서 제안된 변수선택법이 모두 유효한 변수선택 결과를 보였다. 한편, 제안된 변수선택기준의 민감성을 평가하기 위해서는 β 에 대한 여러 형태의 사전확률분포 하에서 이루어진 변수선택결과들의 비교분석이 필요할 것이다. 이 부분에 대한 연구는 본 논문에서 다루지 않고 앞으로의 연구과제로 남겨 두었다.

참고문헌

- [1] Albert, J. H., and Chib, S.(1993), "Bayesian Analysis of Binary and Polychotomous Response Data", Journal of American Statistics Association, Vol. 88, pp. 669-679.
- [2] Best, N., Cowles, M. K. and Vines. K.(1997), "Coda Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output Version 0.40", Technical Report. University of Cambridge. MRC Biostatistics Unit.
- [3] Chib, S.(1995), "Marginal Likelihood from the Gibbs Output", Journal of American Statistics Association, Vol. 90, pp. 1313-1321.
- [4] Chib, S. and Greenberg, E.(1995) "Understanding the Metropolis-Hasting Algorithm", The American Statistician, Vol. 49, pp. 327-335.
- [5] Chib, S. and Greenberg, E.(1996), "Analysis of multivariate probit models", Biometrika
- [6] Chib, S., Greenberg, E and Chen, Y.(1998), "MCMC Methods for Fitting and Comparing Multinomial Response Models", Technical Report, University of Washinhton.
- [7] Choong-Geun Chung (1995), "A Guide to The Analysis of Categorical Dependent variables with SAS and SPSS"; <http://www.indiana.edu/~statmath/stat/all/cat/giant.html>.
- [8] Cowles, M. K. and Carlin, B. P.(1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," Journal of the American Statistical Association, Vol. 91, pp. 883-904.
- [9] de Bruijn, N. G.(1970), Asymptotic Methods in Analysis, Amsterdam: North Holland.
- [10] Gelman, A. E. and Rubin, D. B.(1992a), "Inference from iterative simulation using multiple sequence (with disscessed)", Statistical Science, Vol. 7, pp. 457-511.
- [11] Gelfand, A. E. and Smith, A. F. M.(1990), "Sampling-based approaches to calculating marginal densities", Journal of the American Statistical Association, Vol. 85, pp. 398-409.
- [12] Lewis, S. M. and Raftery, A. E.(1997), "Estimating Bayes Factors via Posterior Simulation With the Laplace-Metropolis Estimator", Journal of the American Statistical

- Association, Vol. 92, No. 438, pp. 648-655.
- [13] McFadden, D.(1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Economics*, ed. P. Zarembka, New York: Academic Press, pp. 105-142.
- [14] McFadden, D.(1989), "A method of simulated moments for estimation of discrete response methods without numerical integration", *Econometrica*, Vol. 57, pp. 995-1026.
- [15] McCulloch, R. E. and Rossi, P. E.(1994), "Exact likelihood analysis of the multinomial probit mode", *Journal of Econometrics*, Vol. 64, pp. 207-240.
- [16] Muller, P.(1993), "A Genetic Approach to Posterior Integration and Gibbs Sampling", manuscript.
- [17] Roberts, G. O., Gelman, A., and Gilks, W. R.(1994), "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," Technical Report, University of Cambridge.
- [18] Rousseeuw, P. J, and van Zomeren, B. C. (1990), "Unmasking multivariate outliers and leverage points(with discussion)", *Journal of the American Statistical Association*, Vol. 85, pp. 633-651.
- [19] Small, G. G. (1990), "A survey of multivariate medians", *Instructional Statistics Review*, Vol. 58, pp. 263-277.
- [20] Stern, S.(1997), "Simulation-based estimation", *Journal of Economic Literature*, Vol. 35, pp. 2006-2039.
- [21] Tierney, L.(1994), "Markov Chains for Exploring Posterior Distributions", *Annals of Statistics*, Vol. 22, pp. 1701-1762.
- [22] Tierney, L., and Kadane, J. B.(1986), "Accurate Approximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association*, Vol. 81, pp. 82-86.