

On an Optimal Bayesian Variable Selection Method for Generalized Logit Model

Hea-Jung Kim¹⁾, Ae Kyoung Lee²⁾

Abstract

This paper is concerned with suggesting a Bayesian method for variable selection in generalized logit model. It is based on Laplace-Metropolis algorithm intended to propose a simple method for estimating the marginal likelihood of the model. The algorithm then leads to a criterion for the selection of variables. The criterion is to find a subset of variables that maximizes the marginal likelihood of the model and it is seen to be a Bayes rule in a sense that it minimizes the risk of the variable selection under 0-1 loss function. Based upon two examples, the suggested method is illustrated and compared with existing frequentist methods.

Keywords : Laplace-Metropolis Algorithm, Generalized Logit Model, Variable Selection

1. 서론

범주형자료에 대한 통계적 분석법은 생의학, 사회과학등의 응용분야에서 최근 널리 사용되고 있으며, 반응범주에 따라서 이항반응모형과 다항반응모형으로 통계적 모형을 설정할 수 있다. 다항반응모형으로 널리 사용되는 다항반응로짓모형(multinomial response logit model)은 반응변수 Y 에 대응하는 공변량 X 의 구조에 따라 일반화로짓모형(generalized logit model)과 다항로짓모형(multinomial logit model)으로 구분된다(Choong-Geun Chung, 1995 참조). 일반화로짓모형은 이항로짓모형을 일반화시킨 모형으로, $(J+1)$ 개의 범주를 갖는 명목형 반응변수 Y 는 기준범주반응값(baseline-category outcome value) $Y=0$ 에 대해 다음과 같이 J 개의 다항로짓함수로 정의된다.

$$\begin{aligned} g_j(\mathbf{x}) &= \log \left[\frac{P(Y=j|\mathbf{x})}{P(Y=0|\mathbf{x})} \right] \\ &= \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \cdots + \beta_{jp}x_p \\ &= \boldsymbol{\beta}'_j \mathbf{x}, \quad j = 1, 2, \dots, J. \end{aligned} \tag{1.1}$$

1) Professor, Department of Statistics, Dongguk University, Seoul 100-715, Korea
E-mail : kim3hj@mail.dgu.ac.kr

2) Graduate Student, Department of Statistics, Dongguk University, Seoul 100-715, Korea

여기서 $\beta'_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})$ 와 $\mathbf{x} = (1, x_1, x_2, \dots, x_p)'$ 는 각각 $1 \times (p+1)$ 인 모수 벡터와 $(p+1) \times 1$ 인 공변량 벡터(covariate vector)를 나타낸다. 따라서 각 반응범주에 대한 조건부확률은

$$\begin{aligned} \pi_j(\mathbf{x}) &= P(Y = j | \mathbf{x}) \\ &= \frac{e^{\beta'_j \mathbf{x}}}{\sum_{j=0}^J (e^{\beta'_j \mathbf{x}})} \end{aligned} \quad (1.2)$$

로 정의되며, 여기서 모수 $\beta_0 = 0$ 이다. 그러므로, n 개의 반응변수 Y_1, Y_2, \dots, Y_n 과 이에 대응하는 공변량 벡터 $\mathbf{x}_i, i = 1, \dots, n$, 가 주어졌을 때 일반화로지트모형의 조건부우도함수(conditional likelihood function)는

$$\begin{aligned} l(\boldsymbol{\beta} | \mathbf{y}) &= \prod_{i=1}^n [\pi_0(\mathbf{x}_i)^{y_{i0}} \pi_1(\mathbf{x}_i)^{y_{i1}} \dots \pi_J(\mathbf{x}_i)^{y_{iJ}}], \\ \text{단, } \sum_{j=0}^J \pi_j &= 1, \quad y_{ji} = \begin{cases} 1, & \text{if } Y_i = j \\ 0, & \text{if } Y_i \neq j \end{cases} \end{aligned} \quad (1.3)$$

이다.

본 논문에서는 일반화로지트모형의 설정에서 효과적으로 공변량의 변수선택하는 방법을 베이지 접근법으로 제안하고자한다.

일반화로지트모형에서의 변수선택이란 모형에서 정의된 J 개의 모든 이항로지트모형을 동시에 고려하여 반응변수에 대응되는 공변량과의 관계를 가장 잘 설명하는 변수군을 선택하는 것으로 현재 일반화로지트모형에서의 변수선택법에 대한 연구는 미진한 상태이다. 개별화된 분석법(Wijesinha et al, 1983)을 이용하여 Begg와 Gray(1984)가 일반화로지트모형을 개별화된 이항로지트모형으로 분리하여, 분리된 각각의 이항로지트모형의 근사적인 추정에 의해 일반화로지트모형을 추정하였고, 이를 이용하여 David와 Hosmer(1989)는 최적모형 선택을 위한 판정기준으로 로그 우도비와 Wald χ^2 통계량을 사용하여 단계별 변수선택법을 제안하였다. 그러나, 이를 기준으로 변수선택을 하는 경우 다음과 같은 문제점들이 있음을 지적할 수 있다. 첫째, 결과가 일치되지 않아 최적모형의 도출에 문제점을 지니고 있다(David와 Hosmer, 1989). 둘째로 변수선택 과정에서 이들 변수선택 기준에 대한 유의성 검정이 매 단계마다 유의수준 α 하에서 이루어지므로 결국 여러 단계 변수선택을 거쳐 얻어진 최적모형에 대한 검정에서의 유의수준 결정에 어려움이 발생된다(Berry, Hayes, Latour와 Schlotzhaur, 1995 참조). 셋째로 변수선택 기준의 분포는 근사분포 이론에 의해 얻어진 점근적인 χ^2 -분포를 따르므로 소표본인 경우 각 단계에서 선택된 변수의 유의성 검정의 검정력에 문제가 있다. 한편, 최적변수선택기준으로는 p 개의 공변량에 대해 2^p 개의 모든 가능한 일반화로지트모형에 대해 AIC, BIC와 같은 모형선택기준에 의해 최적 변량군을 선택하는 방법이 있으나, 복잡한 우도함수의 근사 추정에 의해서만 AIC와 BIC를 추정 가능한 문제를 지닌다.

이러한 문제점을 고려하여 본 논문에서는 베이지 접근법으로 변수선택방법을 제시하고자 한다. 이 방법은 Lewis와 Raftery(1997)가 제안한 Laplace-Metropolis 방법으로 일반화로지트모형의 주변 우도함수(marginal likelihood)를 계산하는 방법을 제안하고, 이를 이용하여 서로 다른 공변량군을 포함한 모형을 비교하는 변수선택의 기준으로 한다. 또한 제안된 변수선택기준이 실제적으로 유용

한지를 살펴보기 위해 모의실험과 경험적인 자료분석을 통하여 제안된 모형선택의 유용성을 비교 검토하고자 한다.

2. 주변우도함수의 Laplace 근사

함수의 적분에 사용되는 Laplace 근사법은 p -차원 벡터 \mathbf{u} 의 함수 $h(\mathbf{u})$ 의 Taylor 급수전개에 의해 다음과 같이 적분을 근사하는 방법이다(de Bruijn, 1970).

$$f(\mathbf{u}) = \int \exp \{h(\mathbf{u})\} d\mathbf{u} \approx (2\pi)^{p/2} |A|^{-1/2} \exp \{h(\mathbf{u}^*)\}, \tag{2.1}$$

여기서 \mathbf{u}^* 는 단봉모형(unimodal)인 $h(\mathbf{u})$ 를 최대화시키는 \mathbf{u} 의 값이며, A 는 \mathbf{u}^* 에서 계산된 h 의 음의 역헤시안(minus inverse hessian)을 나타내며, 이 근사법을 이용하면 추정된 추정량의 추정오차는 근사적으로 $O(n^{-1})$ 이 된다(Tierney와 Kadane, 1986 참조).

Laplace 근사법을 이용하여 일반화로지트모형 (1.1)의 변수선택에 필요한 주변우도함수의 근사에 적용시키면 다음과 같다. 식 (1.1)에서 정의된 일반화로지트모형의 내포모형(nested model)을 $J \times (p^{M_k} + 1)$, $p^{M_k} < p$, 인 모수 $\beta^{M_k} = (\beta_1^{M_k}, \beta_2^{M_k}, \dots, \beta_j^{M_k})'$ 를 가진 모형이라 하고, 이를 M_k 으로 나타내면, 모수벡터의 사전확률밀도함수 $p(\beta^{M_k})$ 가 적절한(proper) 경우, β^{M_k} 의 사후확률밀도함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} p(\beta^{M_k} | \mathbf{y}, M_k) &\propto l(\beta^{M_k} | \mathbf{y}, M_k) p(\beta^{M_k}) \\ &= \prod_{i=1}^n \prod_{j=0}^I \left\{ \frac{e^{g_i^{M_k}(\mathbf{x})}}{\sum_{j=0}^I e^{g_j^{M_k}(\mathbf{x})}} \right\}^{y_{ij}} p(\beta^{M_k}). \end{aligned} \tag{2.2}$$

그리고, 모형 M_k 의 주변우도함수는

$$p(\mathbf{y} | M_k) = \int_{\Omega} l(\beta^{M_k} | \mathbf{y}, M_k) p(\beta^{M_k}) d\beta^{M_k}, \quad \Omega = R^{J(p^{M_k} + 1)}, \tag{2.3}$$

이다. 식 (2.3)을 앞에서 정의한 Laplace 근사법에 적용시키면

$$\begin{aligned} p(\mathbf{y} | M_k) &= \int_{\Omega} e^{h(\beta^{M_k})} d\beta^{M_k} \\ &\approx (2\pi)^{J(p^{M_k} + 1)/2} |H^{*M_k}|^{-1/2} l(\beta^{*M_k} | \mathbf{y}, M_k) p(\beta^{*M_k}) \end{aligned}$$

이 된다. 여기서 β^{*m} 은 $h(\beta^{M_k}) = \log \{l(\beta^{M_k} | \mathbf{y}, M_k)p(\beta^{M_k})\}$ 를 최대화시키는 β^{M_k} 의 값이며, 행렬 H^{*m} 은 $\beta^{M_k} = \beta^{*m}$ 에서 계산된 $h(\beta^{M_k})$ 의 음의 역헤시안이다.

따라서 $p(\mathbf{y} | M_k)$ 의 Laplace 근사 추정식에 대수를 취하면 다음과 같다.

$$\begin{aligned} P_{LM_k} &= \log p(\mathbf{y} | M_k) \\ &\approx \frac{J(p^{M_k} + 1)}{2} \log \{2\pi\} + \frac{1}{2} \log \{|H^{*m}| \} + \log \{l(\beta^{*m} | \mathbf{y}, M_k)\} + \log \{p(\beta^{*m})\}. \end{aligned} \quad (2.4)$$

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

(결과 1) p 개의 공변량을 가진 일반화로지트모형 (1.1)의 모든 내포모형은 $\xi = 2^p - 1$ 개 이고 이들을 각각 모형 M_1, M_2, \dots, M_ξ 로 나타내면 이들 중 최적모형 M_l 은 다음을 만족시킨다.

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

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

$$\begin{aligned} \Pr(M_k | \mathbf{y}) &= \frac{p_k p(\mathbf{y} | M_k)}{\sum_{j=1}^{\xi} p_j p(\mathbf{y} | M_j)} = [1 + \sum_{j \neq k} \frac{p_j}{p_k} B_{jk}]^{-1} \\ &\approx \frac{p_k (2\pi)^{Jp^{M_k}} |H^{*m}|^{1/2} l(\beta^{*m} | \mathbf{y}, M_k) p(\beta^{*m})}{\sum_{j=1}^{\xi} p_j (2\pi)^{Jp^{M_j}} |H^{*m}|^{1/2} l(\beta^{*m} | \mathbf{y}, M_j) p(\beta^{*m})}, \quad k = 1, 2, \dots, \xi. \end{aligned}$$

여기서 B_{jk} 는 모형 M_k 에 대한 M_j 모형의 적합도를 나타내는 베이즈요인(bayes factor)이다 (Berger와 Pericchi, 1996 참조).

3. Metropolis-Hastings 알고리즘

Laplace 근사 식 (2.4)를 계산하기 위해 필요한 β^{*m} 와 H^{*m} 의 값은 β^{M_k} 의 사후확률밀도함수로부터 얻을 수 있다. 그러나, 식 (2.2)와 같이 복잡한 사후확률밀도함수로부터 β^{*m} 와 H^{*m} 를 분석적인 방법으로 계산해 내는 것은 사실상 어렵다. 최근에 MCMC (Markov Chain Monte Carlo) 방법을 이용하여 β^{*m} 와 H^{*m} 를 추정하는 여러방법이 개발되었으며, 그 중 Chib(1995)의 방법이 대표적이나 이 방법은 β^{M_k} 의 조건부사후확률함수가 폐쇄형(closed form)인 경우만 사용 가능한 문제점을 지니고 있다. 또한, MCMC 방법은 일반적으로 반복적인 모의실험에 근거하므로

사후표본에 이상치들이 포함될 가능성이 있다. 이런 경우 사후표본의 평균으로 β^{*m} 와 H^{*m} 를 추정하여 (2.4)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 여러가지 형태가 제안되어 있다(Chib와 Greenberg, 1996).

이 절에서는 β^{*m} 와 H^{*m} 의 추정에 Chib와 Greenberg(1996)에 의해 제안된 재단된독립과정(tailored independence chain)하에서 도출된 생성밀도함수인 다변량정규분포 $f_N(\beta^{M_k} | \mu^*, \tau^2 \Sigma)$ 를 사용하여 Metropolis-Hastings 알고리즘을 구성하고자 한다. 여기서 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, Gelman와 Gilks, 1994; Muller, 1993 참조). 선택된 생성밀도함수를 이용하여 초기값 $\beta^{M_k^{(0)}}$ 이 주어진 재단된독립과정하에서 (t) 상태(state)에서 $(t+1)$ 상태로 전이되는 $\beta_t^{M_k}$ 은

$$\beta_t^{M_k} = \mu^* + z, \quad z \sim N_{p^*}(0, \tau^2 \Sigma), \quad p^* = J \times (p^{M_k} + 1). \quad (3.1)$$

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.

- (절차 1) 생성밀도함수 $f_N(\beta^{M_k^{(t)}} | \mu^*, \tau^2 \Sigma)$ 로부터 $\beta_t^{M_k}$ 를 발생시킨다.
- (절차 2) 균등분포 $U(0, 1)$ 으로부터 u 를 발생시킨다.

(절차 3) 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 선택 된다.

(i) $u \leq \alpha(\beta^{M_k^{(t)}} , \beta_t^{M_k})$ 인 경우 :

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

(ii) $u > \alpha(\beta^{M_k^{(t)}} , \beta_t^{M_k})$ 인 경우 :

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

이와 같이하여 생성밀도함수인 $f_N(\beta^{M_k^{(t)}} , \beta^{M_k^{(t+1)}})$ 가 다음과 같은 가역조건(Tierney, 1994 참조)인

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

을 만족할 때까지 (절차 1) ~ (절차 3)를 반복시행하여 N개의 사후표본을 얻는다.

3.2 알고리즘의 수렴성 진단

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

$$\{\beta^{M_k^{(1)}} , \beta^{M_k^{(2)}} , \beta^{M_k^{(3)}} , \dots , \beta^{M_k^{(t)}} , k = 1, 2, \dots, \xi\}. \tag{3.2}$$

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)를 결정해야 한다. t^* 의 결정은 사후표본의 수렴여부를 진단하기 위해 제안된 다음과 같은 진단방법들을 이용할 수 있다.

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

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

여기서, 식 (2.2)에 의해

$$-2 \ln(\text{사후확률밀도함수}) = -2 \ln \left[\prod_{i=1}^n \prod_{j=0}^I \left\{ \frac{e^{g_j^{M_k}(x)}}{\sum_{j=0}^I e^{g_j^{M_k}(x)}} \right\}^{y_i} p(\boldsymbol{\beta}^{M_k}) \right] \Big|_{\boldsymbol{\beta}^{M_k} = \boldsymbol{\beta}^{M_k(t^*)}}$$

단, $t^* \in \{t \mid t \geq t^*\}$ 이다.

위와 같은 진단법을 사용함으로써 Metropolis-Hastings 알고리즘에 의해 생성된 $\boldsymbol{\beta}^{M_k}$ 의 사후표본에 대한 수렴여부를 확인할 수 있고, 수렴하는 시점에 의해 적당한 t^* 의 값을 결정할 수 있다. 따라서 수렴을 만족하는 사후표본을 얻기 위해서는 제안된 Metropolis-Hastings 알고리즘으로 사후표본을 $T = (t^* + N)$ 번 반복추출한 다음, 수렴시점 t^* 까지 추출된 사후표본은 제거하고 남은 $N = (T - t^*)$ 개를 사용하여 $\boldsymbol{\beta}^{M_k}$ 와 H^{M_k} 를 추정한다.

3.3 알고리즘의 로버스트성

제안된 알고리즘에 얻은 $\boldsymbol{\beta}^{M_k}$ 의 사후표본 (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 이상치에 로버스트한 사후표본의 L_1 중심 (L_1 center)을 사용하면 해결할 수 있다.

따라서, 이상치에 로버스트한 $\boldsymbol{\beta}^{*M_k}$ 은

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

를 최소화하는 $\boldsymbol{\beta}^{M_k^{(r)}}$ 값으로 구할 수 있다. 여기서 t^* 는 사후표본의 수렴시점을 나타내고, $|\cdot|$ 는 절대거리(absolute metric)를 의미한다(Small, 1990 참조). 또한, 모형 M_k 의 주변우도함수의 근사추정에 필요한 H^{*M_k} 의 값은 $\boldsymbol{\beta}^{M_k}$ 의 사후확률밀도함수에서 수치적인 방법에 의해 구한다. 이 방법은 표본의 크기에 따라 다르며, 대표본인 경우에는 최우추정량 $\boldsymbol{\beta}^{*M_k}$ 의 점근적인 정규성에 의해 사후표본의 표본분산행렬을 이용하여 H^{*M_k} 를 추정하고 소표본일 경우는 Rousseeuw와 van Zomeren(1990)에 의해 제안된 가중된 분산행렬을 이용하여 H^{M_k} 의 로버스트 추정값을 구한다. 따라서, 위와 같은 추정법에 의해 추정된 $\boldsymbol{\beta}^{M_k}$ 와 H^{M_k} 의 값을 $\boldsymbol{\beta}^{*M_k}$ 와 H^{*M_k} 라 하면 식 (2.4)에 정의된 Laplace 근사식의 추정값 \hat{P}_{LM_k} 은

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

(3.3)

가 된다. 그러므로, \hat{P}_{LM_k} 을 2장에서 논의된 (결과 1)의 P_{LM_k} 에 대입을 시키면 최적의 모형 M_l 을 선택할 수 있고 M_l 에 포함된 변수들이 곧 일반화로지트모형의 최적군이 된다.

4. 모의실험 및 경험적 자료분석

이 장에서는 본 논문에서 제안된 Metropolis-Hastings 알고리즘에 의해 주변우도함수를 근사 추정하여 얻은 변수선택 기준의 유효성을 모의실험에 의해 얻은 자료와 경험적 자료를 통해 평가해보았다.

4.1 모의실험

모의실험에 사용된 일반화로지트모형은 다음과 같다.

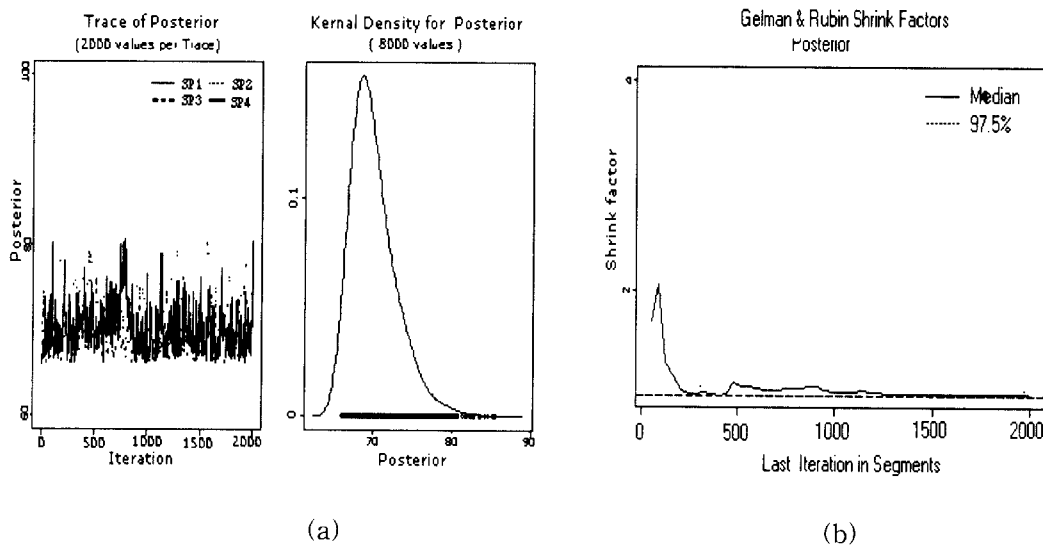
$$\begin{aligned} g_j(\mathbf{x}_i) &= \ln \left[\frac{P(Y_i = j | \mathbf{x}_i)}{P(Y_i = 0 | \mathbf{x}_i)} \right] \\ &= \beta_{j0} + \beta_{j1}x_{1i} + \beta_{j2}x_{2i} + \beta_{j3}x_{3i} \\ &= \boldsymbol{\beta}'_j \mathbf{x}_i, \quad j = 1, 2, \quad i = 1, 2, \dots, 50. \end{aligned} \quad (4.1)$$

여기서 반응변수 Y_i 는 반응값 0, 1, 2를 가지며, 기준범주의 반응값 0에 대한 다항로지트함수를 식 (4.1)과 같이 설정하였다. 위 모형에서 회귀계수는 각각 $\boldsymbol{\beta}'_1 = (0, 2, 2)$, $\boldsymbol{\beta}'_2 = (0, 4, 4)$ 으로 놓고, 공변량 $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i})'$, $i = 1, \dots, 50$,의 관측값은 $N(0, I_3)$, $I_3 = \text{diag}(1, 1, 1)$,로 부터 발생시켰으며, 각 공변량 벡터 \mathbf{x}_i 관측에 대해 식 (4.1)로부터 얻은 반응변수 Y_i 의 분포가 주어지면 이것을 이용하여 관측값 Y_i , $i = 1, \dots, 50$,을 발생시켜 모의실험에 사용된 크기 50인 공변량과 반응변수에 대한 자료를 얻었다.

이와 같은 모의실험을 200번 반복하여 얻은 자료를 이용하여 식 (4.1)에서 정의되는 모든 내포 모형 M_k , $k = 1, \dots, 7$,의 주변우도함수를 제안된 알고리즘으로 근사 추정하고, 이것을 변수선택 기준에 적용하여 선택한 최적모형이 (4.1)에 정의된 참 모형과 일치하는지를 평가하고자 한다. 여기서 모든 내포모형에 포함된 모수벡터 $\boldsymbol{\beta}^{M_k}$ 의 사전확률분포는 $p(\boldsymbol{\beta}^{M_k}) \sim N(0, 10I_{p^{M_k}})$, $p^{M_k} < 3$, $p^{M_k} \times p^{M_k}$ 인 단위행렬 $I_{p^{M_k}}$ 를 사용하였으며, 이것으로 식 (2.2)와 (2.4)를 유도하였다.

또한, 제안된 알고리즘으로부터 얻은 사후표본의 수렴성을 진단하기 위해 3.2절에서 논의된 Gelman과 Rubin(1992)의 수렴성 진단법인 shrink factor값을 사용하였다. 이 값의 계산에는 Best, Cowles 그리고 Vines(1996)에 의해 개발된 S-PLUS함수인 CODA 함수를 이용하였으며, (그림 1)

은 이 함수로부터 얻은 것이다. 먼저, 식 (4.1)로부터 정의되는 모든 내포모형 M_k 각각에 대한 β^{M_k} 의 4가지 초기값을 ($\hat{\beta}_{LSE}^{M_k}, \hat{\beta}_{MLE}^{M_k}, \hat{\beta}_{MLE}^{M_k} \pm$ 표준편차)으로된 알고리즘을 2000번 반복하여 β^{M_6} 의 사후표본을 발생한 결과 모든 내포모형 M_k 에 대해 β^{M_k} 의 사후표본체인이 수렴하는 것으로 나타났으며, 그 중에서 참 모형 M_6 에 대한 수렴 진단결과만을 (그림 1)로 나타내었다. (그림 1)은 주어진 4개의 다중체인으로부터 추정된 β^{M_6} 의 사후밀도함수에 $-2 \ln$ 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 그래프로써, shrink factor 값이 1.01이 되어 사후표본이 실사후표본에 수렴하고 있음을 보여주고 있다(shrink factor 값이 1일 경우 완전수렴을 나타냄). 따라서, Metropolis-Hastings 알고리즘을 이용하여 추출된 사후표본의 수렴 시점은 $t^* = 2000$ 으로 정하고, $(t^* + N)$ 개 추출된 사후표본에서 t^* 개를 제거하고 남은 $N = 10000$ 개를 주변우도함수 추정에 사용할 사후표본으로 설정하였다.



(그림 1) (a) $-2 \ln$ (사후확률밀도함수)에 대한 trace plot과 커널함수 ;
 (b) $-2 \ln$ (사후확률밀도함수)에 대한 shrink factor.

이와 같은 절차로 얻어진 사후표본을 이용하여 (2.3)식에 정의된 주변우도함수를 근사 추정하였으며, 이를 사용하여 식 (4.1)에서 정의되는 모든 내포모형들 중에서 최적모형을 선택한 결과는 (표 1)과 같다. (표 1)에 의하면, 200번의 모의실험 중 참 모형인 M_6 을 최적모형으로 선택된 비율이 타 모형에 비해 월등히 큰 값을 나타내고 있다.

그러므로, 모의실험결과 제안된 변수선택기준은 일반화로지트모형의 변수선택에서 유효하게 사용될 수 있음을 확인할 수 있다. (표 1)에서 괄호안의 숫자는 200번 반복 실험 중 최적모형으로 선택된 비율을 나타낸다.

(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(0.00)	16(0.08)	10(0.05)	1(0.005)	0(0.00)	164(0.82)	9(0.045)

4.2 경험적인 자료분석

이 절에서는 본 논문에서 제안된 변수선택기준을 *Applied Logistic Regression*(David와 Hosmer, 1989)에 수록된 자료에 적용하였다. 이 자료는 유방 x-선 조영법(mammography)에 관한 여성들의 인지도, 태도 그리고 행동을 나타내는 Mammography Experience 자료이다.

자료의 구성은 다음과 같다.

공변량 벡터 x :

x_1 : 유방암 증상이 없다면 유방 x-선을 조영해 볼 필요성이 없다고 생각하는가.

(강한긍정 또는 긍정 (0); 부정 또는 강한부정(1))

x_2 : 유방 x-선 조영법의 효능정도(5-20).

x_3 : 모계쪽의 유방암 병력 유무. : (유(1); 무(0))

x_4 : 스스로 유방암 검사를 받도록 누군가에 의해 권유를 받았던 경험의 유무.

(유(1); 무(0))

x_5 : 새로운 유방암을 발견하는데 유방 x-선 조영법의 기여도.

(전혀 없다(1); 약간 있다(2); 매우 많다(3))

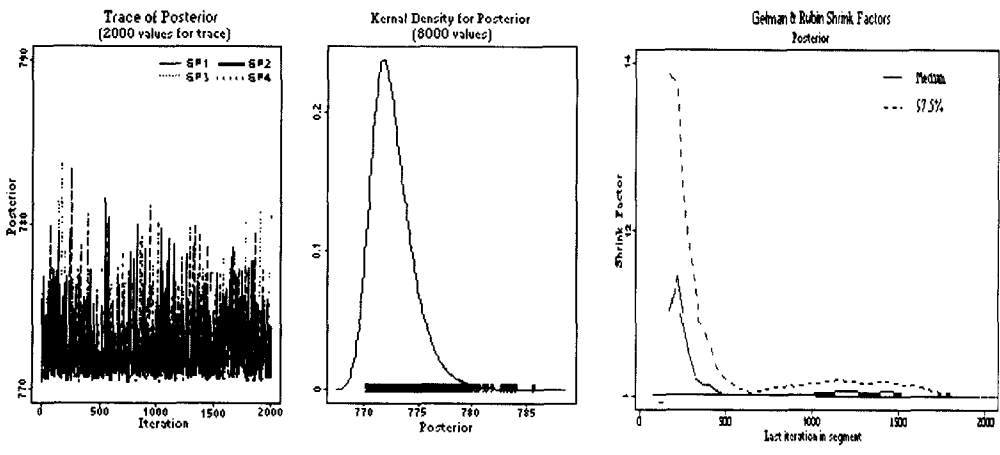
반응변수 Y : 유방 x-선 검사를 받아본 시기에 따라 다음과 같은 3개의 범주로 관측되었다.

(전혀없다(1); 지금부터 1년전에(2); 과거 몇 년전에(0))

반응변수 Y 는 3개의 범주를 갖는 명목형 반응변수로서 기준범주반응값 $Y = 0$ 에 대해 2개의 일변화로지함수를 다음과 같이 설정할 수 있다.

$$\begin{aligned}
 g_j(\mathbf{x}_i) &= \log \left[\frac{P(Y_i = j | \mathbf{x}_i)}{P(Y_i = 0 | \mathbf{x}_i)} \right] \\
 &= \beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \beta_{3j}x_{3i} + \beta_{4j}x_{4i} + \sum_{k=1}^2 \alpha_{kj}d_{ki} \quad (4.2) \\
 &= \boldsymbol{\beta}'_j \mathbf{x}_i, \quad j = 1, 2, \quad i = 1, 2, \dots, 412.
 \end{aligned}$$

여기서 $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}, \alpha_{1j}, \alpha_{2j})'$ 와 $\mathbf{x}_i = (1, x_{1i}, x_{2i}, x_{3i}, x_{4i}, \mathbf{d}_i)'$ 는 각각 1×7 인 모수벡터와 공변량 벡터를 나타내고, 가변수 벡터 $\mathbf{d}_i = (d_{i1}, d_{i2})$ 는 공변량 x_{5i} 의 관측값 중에서 $x_{5i} = 3$ (매우많다)를 기준공변량값(reference covariate value)으로 하여 나머지 관측값들을 가변수 d_{i1} 와 d_{i2} 로 나타내낸 것이다 ($x_{5i} = 1$ 인 경우 $d_{i1} = 1, d_{i2} = 0$; $x_{5i} = 2$ 인 경우 $d_{i1} = 0, d_{i2} = 1$; $x_{5i} = 3$ 인 경우 $d_{i1} = 0, d_{i2} = 0$). 그리고 식 (4.2)로부터 정의되는 32개의 상수항을 포함한 모든 내포모형 M_k 에 포함된 모수벡터 $\boldsymbol{\beta}^{M_k}$ 의 사전확률분포는 $p(\boldsymbol{\beta}^{M_k}) \sim N(0, 10I_{p^{M_k}})$, $p^{M_k} < 3$ 를 사용하였으며, 이 분포를 이용하여 식 (2.2)와 (2.4)를 유도하였다. 여기서 $I_{p^{M_k}}$ 는 $p^{M_k} \times p^{M_k}$ 인 단위행렬이다. 먼저, 제안된 Metropolis-Hastings 알고리즘에 의해 추출된 사후표본의 수렴성에 대한 진단은 앞의 모의실험과 같은 방법으로 검토하였다. 그 결과 제안된 알고리즘을 2000번 반복하면 식 (4.2)로부터 정의된 모든 내포모형 M_k 에 대한 $\boldsymbol{\beta}^{M_k}$ 의 사후표본은 수렴하는 것으로 나타났으며, 이러한 진단 결과는 (그림 2)에 의해 알 수 있다. 따라서, Metropolis-Hastings 알고리즘을 이용하여 추출된 사후표본의 수렴 시점은 $t^* = 2000$ 으로 정하고, $(t^* + N)$ 개 추출된 사후표본에서 t^* 개를 제거하고 남은 $N = 10000$ 개를 주변우도함수 추정에 사용할 사후표본으로 설정하였다.



(a) $-2 \ln$ (사후확률밀도함수)에 대한 trace plot과 커널함수 ;
 (b) $-2 \ln$ (사후확률밀도함수)에 대한 shrink factor.

이와 같은 절차로 얻어진 사후표본을 이용하여 얻은 (3.3)식의 주변우도함수 추정값(\hat{P}_{LM_i})와 모형의 사전확률 $\Pr(M_k)$ 이 동일하다는 조건하에서 구한 모형의 사후확률 $\Pr(M_k | \mathbf{y})$ 의 근사값을 기준으로 식 (4.2)에서 정의된 일반화로지트모형에 대한 변수선택 결과는 (표 2)와 같다.

(표 2) Mammography Experience 자료에 대한 변수선택 결과
(c : 상수항, 괄호안의 값은 $\hat{P}_{M_k} = \Pr(M_k | \mathbf{y})$ 의 근사값)

모형	$\hat{P}_{LM_i} / \hat{P}_{M_k}$	AIC	모형	$\hat{P}_{LM_i} / \hat{P}_{M_k}$	AIC	모형	$\hat{P}_{LM_i} / \hat{P}_{M_k}$	AIC
c	-409.2071 (1.12E-13)		$cx_1x_2x_3$	-381.1401 (0.17376)	730.409	$cx_1x_3d_1d_2$	-386.2719 (0.00103)	743.261
cx_1	-386.1437 (0.00117)	756.677	$cx_1x_2x_4$	-381.2390 (0.15740)	732.578	$cx_1x_4d_1d_2$	-387.3258 (0.00036)	743.640
cx_2	-399.0663 (2.85E-9)	777.283	$cx_1x_3x_4$	-383.2843 (0.02036)	740.235	$cx_2x_3d_1d_2$	-397.5834 (1.26E-8)	762.721
cx_3	-407.1852 (8.49E-13)	800.340	$cx_2x_3x_4$	-396.1636 (5.19E-8)	763.040	$cx_2x_4d_1d_2$	-396.8646 (2.58E-8)	762.094
cx_4	-404.9472 (7.96E-12)	796.065	$cx_1x_2x_3x_4$	-380.1873 (0.45056)	725.793	$cx_3x_4d_1d_2$	-399.5409 (1.77E-9)	772.388
cx_1x_2	-381.3045 (0.14742)	737.192	cd_1d_2	-403.8891 (2.29E-11)	790.401	$cx_1x_2x_3d_1d_2$	-384.5770 (0.00559)	730.511
cx_1x_3	-384.0483 (0.00948)	748.005	$cx_1d_1d_2$	-387.8290 (0.00022)	751.831	$cx_1x_3x_4d_1d_2$	-383.8123 (0.01201)	732.021
cx_1x_4	-384.6213 (0.00535)	748.534	$cx_2d_1d_2$	-398.7119 (4.06E-9)	769.340	$cx_1x_2x_4d_1d_2$	-386.0781 (0.00125)	731.773
cx_2x_3	-397.8169 (9.94E-9)	770.234	$cx_3d_1d_2$	-402.3564 (1.06E-10)	782.021	$cx_2x_3x_4d_1d_2$	-396.7522 (2.88E-8)	755.991
cx_2x_4	-397.0686 (2.10E-8)	769.752	$cx_4d_1d_2$	-401.1050 (3.71E-10)	779.837	$cx_1x_2x_3x_4d_1d_2$	-383.8996 (0.01100)	725.415
cx_3x_4	-403.4594 (3.52E-11)	787.942	$cx_1x_2d_1d_2$	-385.1709 (0.00309)	737.162			

(표 2)는 (4.2)식에 정의된 일반화로지트모형의 모든 내포모형에 대해 얻은 \hat{P}_{LM_i} 와 $\Pr(M_k | \mathbf{y})$ 그리고 표본이론하에서 변수선택기준으로 널리 사용되는 AIC 값을 보여준다. 이에 의하면, 2장에서 논의된 (결과 1)에 따라서 제안된 변수선택기준에 의해 선택된 최적모형은 \hat{P}_{LM_i} 과 $\Pr(M_k | \mathbf{y})$ 이 가장 큰 모형인 $\text{logit}(\pi_j) = f(c, x_1, x_2, x_3, x_4)$, $j = 1, 2$, 모형으로 나타났다. 이와는 달리 AIC 기준에 의하면 $\text{logit}(\pi_j) = f(c, x_1, x_2, x_3, x_4, x_5)$, $j = 1, 2$ 가 최적모형인 것으로 판단되었다.

(표 3) 단계적 변수선택 결과 (c : 상수항, 괄호안의 값은 베イズ 추정값)

로짓 모형	공변량	자유도	단 계 1				단 계 2			
			회귀 모수	추정 오차	Wald χ^2 -값	p -값	회귀 모수	추정 오차	Wald χ^2 -값	p -값
1	c	1	-2.6775 (-3.0748)	1.4359 (1.2889)	3.48	0.0622	-1.7619 (-2.0900)	0.8483 (0.8807)	4.31	0.0378
	x_1	1	2.0890 (2.0322)	0.4575 (0.4503)	20.85	0.0000	2.2225 (2.3413)	0.4521 (0.4091)	24.16	0.0000
	x_2	1	-0.2530 (-0.2230)	0.0731 (0.0563)	11.96	0.0005	-0.2854 (-0.2457)	0.0716 (0.0661)	15.91	0.0001
	x_3	1	1.2983 (1.4640)	0.4333 (0.5201)	8.98	0.0027	1.3025 (1.3941)	0.4292 (0.3823)	9.21	0.0024
	x_4	1	1.2366 (0.2300)	0.5264 (0.4123)	5.52	0.0188	1.2132 (1.1171)	0.5213 (0.4478)	5.42	0.0199
	d_1	1	0.0899 (0.0012)	1.1607 (1.0968)	0.01	0.9383	-	-	-	-
	d_2	1	0.9643 (0.9845)	1.1262 (1.0629)	0.73	0.3919	-	-	-	-
2	c	1	-1.0147 (-0.9965)	1.0727 (0.6950)	0.89	0.3442	-1.7542 (-2.1421)	0.8092 (0.8369)	4.70	0.0302
	x_1	1	1.1192 (1.1313)	0.3573 (0.3511)	9.81	0.0017	1.1506 (1.2537)	0.3515 (0.3784)	10.71	0.0011
	x_2	1	-0.1662 (-0.1700)	0.0741 (0.0496)	5.03	0.0250	-0.1560 (-0.1165)	0.0711 (0.0717)	4.82	0.0281
	x_3	1	1.0171 (1.0820)	0.4536 (0.4868)	5.03	0.0249	1.0644 (1.2685)	0.4525 (0.4151)	5.53	0.0187
	x_4	1	1.0262 (1.0080)	0.5140 (0.4949)	3.99	0.0459	0.9581 (0.9217)	0.5072 (0.4287)	3.57	0.0589
	d_1	1	-0.8976 (-0.9174)	0.7142 (0.5871)	1.58	0.2088	-	-	-	-
	d_2	1	-0.6681 (-0.6651)	0.6876 (0.5895)	0.94	0.3312	-	-	-	-
ln (최대우도)			-348.7075				-352.8967			

(표 3)은 David와 Hosmer(1989)가 제안한 일반화로짓모형의 단계적 변수선택법에 의한 결과이다. 이들이 제안한 선택 기준인 로그 우도비와 Wald χ^2 통계량을 사용하여 변수선택한 결과는 다음과 같다. 먼저 로그 우도비검정 결과를 보면, 공변량 $\mathbf{d} = (d_1, d_2)$ 에 대한 로그 우도비검정

통계량값은 $8.3784 = -2[-352.8967 - (348.7075)]$ 이고, 이에 대한 유의확률값(p -값)이 0.07866 (자유도 4)가 되어 공변량 d_1 과 d_2 는 이 모형에서 유의한 것으로 나타났다. 그러나, Wald χ^2 통계량을 기준으로 한 변수선택에서는 공변량 d_1 과 d_2 에 관련된 회귀계수의 p -값이 각각 로짓 모형에 대해 0.9383(0.2088)과 0.3919(0.3312)로 유의하지 않은 것으로 나타났다. 이것은 식 (4.2)에 정의된 일반화로짓모형에 대한 변수선택에서 두 검정기준(로그 우도비와 Wald χ^2 통계량)이 서로 상반된 선택 결과를 가져와 이들 기준의 유효성에 문제가 있음을 보여준다. David와 Hosmer(1989)가 이 자료의 분석에서 제시한 최적의 일반화모형은 $\text{logit}(\pi_j) = f(x_1, x_2, x_3, x_4)$, $j=1,2$ 이다. 이 모형과 이 절에서 행한 변수선택 결과를 대비해 볼 때 본 논문에서 제안한 베이지스 기준과 Wald 기준이 David와 Hosmer의 결과와 일치된 모형을 선택하고 있다. 이는 일반화로짓모형의 변수선택에서 베이지스 기준이 표본이론하에서 제안된 AIC, 로그 우도비 및 Wald 기준 간에 불일치를 해결해 줄 수 있는 변수선택 기준이 됨을 알 수 있다.

5. 결론

본 논문은 베이지스 접근법으로 일반화로짓모형의 최적모형을 선택하기 위한 방법을 Lewis와 Raftery(1997)에 의해 제안된 Laplace-Metropolis 방법을 이용하여 제안하였다. 본 논문에서 제안된 Laplace-Metropolis 방법은 일반화로짓모형의 주변우도함수 Laplace 근사식을 재단된독립과정(tailored independence chain)하에서 제안된 Metropolis-Hastings 알고리즘으로 추정하는 방법이다. 이 추정법은 기존의 것과 달리 다음과 같은 두 가지 유용성을 갖추고 있다. 첫째, 조건부사후 밀도함수가 비폐쇄형(closed form)인 경우에도 주변우도함수를 근사 추정할 수 있다. 둘째로, 이상치에 로버스트한 사후표본의 L_1 중심과 역 헤시안 추정량을 이용하여 이상치에 로버스트한 주변우도함수를 추정할 수 있다.

본 논문에서는 이러한 유용성이 있는 Laplace-Metropolis 방법을 이용하여 일반화로짓모형의 주변우도함수를 근사 추정하는 방법을 제시하였고, 이를 이용하여 일반화로짓모형의 변수선택기준을 제시하였으며, 제시된 기준의 유효성을 모의실험과 경험적인 자료분석을 통해 평가하여 보았다. 그 결과 모의실험 및 경험적 자료분석에서 제안된 변수선택기준이 유효한 것으로 나타났다. 한편, 제시된 기준의 민감성을 평가하기 위해서는 β 에 대한 여러 형태의 사전확률분포하에서 이루어진 변수선택결과에 대한 분석이 필요할 것이다. 이 부분에 대한 연구는 본 논문에서 다루지 않고 앞으로의 연구과제로 남겨 두었다.

참고문헌

- [1] Begg, C. B., and Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regression. *Biometrika*, 71,1-10.
- [2] Berger, J. O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of American Statistics Association*, Vol 91, pp 109-122.
- [3] Jack J. Berry, Duane S. Hayes, Kristin R. Latour, David C. Schlotzhauer, (1995) *Logistic*

Regression Example, Using the SAS System.

- [4] Chib, S.(1995), "Marginal Likelihood from the Gibbs Output", *Journal of American Statistics Association*, Vol 90, pp 1313-1321.
- [5] Chib, S. and Greenberg, E.(1995a) "Understanding the Metropolis-Hasting Algorithm", *The American Statistician*, Vol 49 , pp 327-335 .
- [6] Choong-Geun Chung (1995), "A Guide to The Analysis of Categorical Dependent variables with SAS and SPSS" ; <http://www.indiana.edu/~statmath/stst/all/cat/stst/all/cat/catv2.html>.
- [7] 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, 883-904.
- [8] David W. Hosmer, J.(1989), *Applied Logistic Regression*, John Wiley & Sons, New York.
- [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*, 7, pp 457-511.
- [11] Lewis, S. M. and Raftery, A. E.(1997), "Estimating Bayes Factors via Posterior Simulation With the Laplace-Metropolis Estimator", Vol. 92,No. 438, pp 648-655.
- [12] Muller, P.(1993), "A Genetic Approach to Posterior Integration and Gibbs Sampling", manuscript.
- [13] 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.
- [14] 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.
- [15] Small, G. G. (1990), A survey of multivariate medians. *Instruduction Statistical Review*, 58, 263-277.
- [16] Tierney, L.(1994), "Markov Chains for Exploring Posterior Distributions", *Annals of Statistics*, Vol 22, pp 1701-1762.
- [17] 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.
- [18] Wijesinha, A., Begg, C. B., Funkenstein, H. H. & McNeil, B. J.(1983). Methodology for the differential diagnosis of a complex data-set: A case study using data from routine CT-scan examinations. *Medical Decision Masking* 3, 133-54.