

메타분석에서 그룹화 임의효과 모형의 베이지안 해석

정윤식¹⁾ 정호진²⁾

요약

본 논문은 의학분야에서 주로 사용되는 메타분석 중 그룹화 임의효과 모형(grouped random effects model)을 프라빗 연결함수(probit link function)를 이용하여 베이즈적 관점에서 연구하였다. 이때 프라빗 함수를 사용하기 위해 잠재변수를 정의하였고, 사전분포를 달리한 세가지 모형을 고려하였다. 주어진 세가지 모형들에서 적합한 모형선택을 위하여 베이즈 인자(Bayes factor, BF)와 유사베이즈 인자(pseudo-Bayes factor, PsBF)를 이용하였다. 깃스샘플러와 메트로폴리스 알고리즘을 이용하여 베이지안 계산상의 어려움을 해결하였다. 예로써, 새로운 간질약에 대한 효과를 조사하기 위하여 앞에서 제시된 방법으로 해석하였다.

주요용어: 그룹화 임의효과 모형, 깃스샘플러, 메트로폴리스 알고리즘, 베이즈 인자, 프라빗 연결함수.

1. 서론

의학분야에서 널리 이용되는 메타분석은 각각 독립인 연구로부터 나온 결과들의 정보를 종합하는 양적인 방법이다. Glass(1976)는 실험군과 대조군의 연구효과가 동질적일때는 모수효과(fixed effects) 모형의 타당성을 보였다. 한편 Hedges(1983)는 연구효과가 동질적이지 아닐때는 임의효과 모형(random effects model)이 타당함을 보였다. 일반적으로 실험군과 대조군의 차이처럼 연구효과가 이질적(heterogeneous)인 경우, Carlin(1992), Morris and Normand(1992)등은 임의효과 모형이나 계층적 모형(hierarchical model)을 사용하였다. 임의효과 모형은 연구들 각각의 효과와 연구들간의 효과의 변화를 모형화하는데 적합하다. 그러므로 여기서 우리는 다음과 같은 그룹화 임의효과 모형에 관심을 갖고자 한다.

우선 각각의 연구들이 독립적으로 수행된다고 가정한다. 또한, 각 연구에서 실험들의 성공 확률은 같다고 한다. 이 때, n_i 는 i 번째 연구의 실험 시행 횟수, y_i 는 i 번째 연구의 실험 성공 횟수, p_i 는 i 번째 연구의 성공률을 나타낸다. 일반적으로 p_i 의 추정치는 $\hat{p}_i = \frac{y_i}{n_i}$ 로 사용한다. 이들 연구들이 K 개의 서로 다른 그룹으로 나누어지며, 총 N 개의 연구들이 있다고 가정하자. 모든 연구들은 어느 한 그룹에 속하여야 한다. 이때, 이들 모형은 다음과 같이 표현할 수 있다.

$$\pi(y_i | \theta_i, \mu_k, \sigma_k^2) \stackrel{ind.}{\sim} \text{Binomial}(n_i, p_i), i = 1, \dots, N$$

1) (609-735) 부산시 금정구 장전동 산 30, 부산대학교 통계학과, 부교수

E-mail: yschung@hyowon.pusan.ac.kr

2) (156-010) 서울시 동작구 신대방동 460-18, 기상연구소, 예보연구실, 연구원

와

$$\pi(\theta_i | \mu_k, \sigma_k^2) \sim G(\mu_k, \sigma_k^2), \text{ for some } k = 1, \dots, K \quad (1.1)$$

여기서 $\theta_i = g(p_i)$, g 는 임의의 함수이고, G 는 θ_i 의 가능한 분포함수를 나타내며, μ_k, σ_k^2 는 그 것의 평균과 분산을 나타낸다. 예로써, 이때 임의 효과를 나타내기 위하여 주어진 평균을 $\mu_k = \mu + \delta_k$ 로 표현하면, δ_k 는 그룹 k 에 대한 임의효과라 생각할 수 있다.

메타분석이 자주 서로 유사하지 않은 연구들로부터 나온 결과를 부적절하게 종합하였고, 이런 경우에 통계적 추론은 매우 의심스럽게 되었다. 이러한 문제는 그룹화된 독립적인 연구들에 대해 베이지안 계층적 모형의 구조를 통해 해결될 수 있다.

이때, 통계적 추론에서 일반적으로 고전적 방법인 최대우도 추정치(MLE)를 이용하나, 이는 점근적 성질에 있어서 다수의 자료가 요구된다. 이에 비해 베이지안 분석은 사후분포를 근거로 추론이 이루어지므로 자료의 수와는 상관이 없다. 특히 메타분석에서는 일반적으로 관측치의 수가 적음으로, 베이지안 분석이 보다 훨씬 적합하다. 또한 계층적 베이지안 구조는 연구들의 그룹간 이질성뿐 아니라 각각의 연구들사이의 이질성도 묘사하고 설명하는 자연스러운 도구를 제공한다.

본 연구에서는 예로써, 새로운 간질약인 프로가바이드 (progabide)의 효과를 연구한 15편의 연구로 이루어진 실제 자료를 통해 프라빗 연결함수(probit link function)를 이용하여 제안된 베이지안 방법으로 여러 모형들 가운데 가장 좋은 모형을 선택하기 위하여 베이즈 인자와 유사베이즈 인자를 이용하며, 이때 복잡한 계산은 마코프체인 몬테칼로 방법으로 해결하였다. 이때 최상의 모형에서 임의효과를 분석하여 새로운 약의 효용성을 논한다. Larose와 Dey(1997)는 같은 모형에 로짓 연결함수를 이용하였으므로 본 논문의 결과와 비교하여 연결함수에 민감한 것을 보였다.

본 논문의 구성은 다음과 같다. 2장에서 계층적 그룹화 임의효과 모형의 베이지안 구조를 설명하고, 3장에서 베이지안 계산의 어려움은 90년대 들어와 Gelfand와 Smith(1990)가 제안하고 많은 베이지안 통계학자들에 의하여 연구되어진 깁스 샘플러로 해결하였으며, 이때 관심있는 사후분포는 구조식들이 매우 복잡하여 난수 생성의 문제점은 메트로폴리스 알고리즘을 이용하여 설명한다. 4장은 베이지안 모형선택을 위한 몇가지 기준(Bayes factor, pseudo-Bayes factor)을 설명하고, 그것들의 근사적 식들을 소개한다. 5장에서는 실제 자료에 앞에서 제시된 방법들을 이용하여 모형선택 결과를 설명하고, 선택된 모형에서 새로운 약의 효능을 논한다. 마지막으로 6장에서 본 논문의 결론과 앞으로의 과제를 언급한다.

2. 그룹화 임의효과모형의 베이지안 구조

이 절부터 표기상의 간편함을 위하여 다음과 같은 표기를 사용한다. 즉, $[X, Y], [X|Y]$ 와 $[X]$ 들은 결합, 조건부 및 주변분포함수들을 각각 나타낸다.

K 개의 서로 다른 그룹으로 나누어진 N 개의 연구들이 있다고 가정하자. 반응변수가 이진 자료이고 연구들이 독립이라 가정하자. 우도함수로는 이항 분포가 적절하다. 그러므로 식(1.1)에서 주어진 그룹화 임의효과 모형의 구조식에서 θ_i 의 분포함수를 정규분포로 가정

하면 다음과 같이 표현된다.

$$\pi(y_i | \theta_i, \mu, \delta_k, \sigma_k^2) \stackrel{ind}{\sim} \text{Binomial}(n_i, p_i), \quad i = 1, \dots, N$$

와

$$\pi(\theta_i | \mu, \delta_k, \sigma_k^2) \sim \text{Normal}(\mu + \delta_k, \sigma_k^2), \quad k = 1, \dots, K \quad (2.1)$$

여기서 임의의 함수 g 는 $\theta_i = g(p_i)$ 이고, y_i 는 i 번째 연구의 실험 성공 횟수, n_i 는 i 번째 연구의 실험 시행 횟수, p_i 는 i 번째 연구의 성공률, 즉, $\hat{p}_i = \frac{y_i}{n_i}$, $0 < p_i < 1$ 이고, $k = 1, \dots, K$ 에 대해 δ_k 는 그룹 k 의 임의효과이다. 앞으로 N_k 는 그룹 k 에 있는 연구수이고, N 는 총 연구수이다. 즉, $N = \sum_{i=1}^K N_i$ 이고 수식 표현의 편이상 $N_0 = 0$ 이라 하자.

일반적으로 p_i 의 연결함수들로는 완비 로그-로그(complementary log-log), 로짓(logit), 프라빗(probit)함수 등이 있다. Larose와 Dey (1997)는 p_i 에 대해 로짓함수를 고려하였고 즉, $\theta_i = \frac{p_i}{1-p_i}$ 이다. 본 논문에서는 p_i 에 대한 프라빗 함수를 연구하고자 한다. 즉, $\theta_i = \Phi^{-1}(p_i)$, $-\infty < \theta_i, \mu, \delta_k < \infty$, 여기서 $\Phi(\cdot)$ 는 누적표준 정규분포이다. 예를들어 $K = 2$ 이면 단일눈가림(single blind)와 이중눈가림(double blind) 연구와 같이 서로 다른 두 그룹을 고려한 형태이다.

첫단계 사전분포로 $\theta_i = \Phi^{-1}(p_i)$ 가 실수 전체에서 정의되므로 정규분포를 사용한다. 이제 θ_i 의 평균과 분산의 구조를 살펴보자. 전체 평균 μ 에 그룹임의효과 δ_k , $k = 1, \dots, K$ 를 더한 가법평균구조를 이용한다. K 그룹 각각의 분산이 동일하다는 정보가 없기때문에, θ_i 의 분산은 그룹 k , $k = 1, \dots, K$ 에 의존한다고 가정한다.

베이지안 구조 아래서 μ, δ_k 와 σ_k^2 , $k = 1, \dots, K$ 의 특정한 사전분포를 결정해야 한다. μ, δ_k 와 σ_k^2 이 모두 독립이라 가정하면 이들의 결합 사전분포 함수는

$$\pi(\mu, \delta, \sigma^2) = \pi(\mu) \times \pi(\delta) \times \pi(\sigma^2)$$

이고, 여기서 $\delta = (\delta_1, \dots, \delta_K)$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)$ 이다. 여기서, 사전분포함수들은 Larose와 Dey(1997)와 같은 사전분포함수들을 이용한다. 이때 그들이 제안한 μ, δ_k 와 σ_k^2 의 세가지 다른 형태의 사전분포 함수들은 다음과 같다. 세가지 다른 형태를 갖는 사전분포 함수들에서 μ 와 $\delta_1, \dots, \delta_K$ 의 사전분포들은 다음과 같이 모두 같다고 가정하자.

$$\pi(\mu) \sim N(a, b), \quad \pi(\delta_k) \sim N(c_k, d_k), \quad k = 1, \dots, K. \quad (2.2)$$

여기서 $N(a, b)$ 는 평균 a , 분산 b 인 정규분포를 나타낸다. a, b, c_k 와 d_k 는 상수들이라 한다. 즉, μ 와 δ 는 위와 같은 동일한 사전분포를 갖는 반면에 σ_k^2 는 다른 사전분포를 갖는 세 모형을 다음과 같이 고려한다.

$$(M_1): \pi(\sigma_k^2) \sim \text{IG}(e_k, f_k), \quad (2.3)$$

$$(M_2): \pi(\sigma_k^2) \propto \frac{s_k}{(s_k + \sigma_k^2)^2}, \quad (2.4)$$

와

$$(M_3): \pi(\sigma_k^2) \propto \left(1 + \frac{\sigma_k^2}{s_k^2}\right)^{-1}. \quad (2.5)$$

여기서 s_k^2 , $k = 1, \dots, K$ 는 그룹 k 의 표본분산의 조화평균이고, IG 는 역 감마 분포함수를 나타낸다. e_k 와 f_k 는 알려진 상수들이다. 식(2.4)에서 σ_k^2 의 사전분포는 DuMouchel (1994)에 의해 제안된 로그-로지스틱(log-logistic)함수이다. 또한, 식(2.5)에서 사용된 σ_k^2 의 로그코쉬(log Cauchy) 사전분포는 매우 넓게 퍼져있다.

일반적인 프라빗 모형하에서 θ_i 와 p_i 의 관계식은 $\Phi^{-1}(p_i) = \theta_i$ 이다. Albert와 Chib (1993)의 방법을 이용하여, 잠재변수 $W_i = (W_{i1}, \dots, W_{ini})$ 를 다음과 같이 정의할 수 있다. 여기서 W_{ij} 들은 서로 독립이며, $N(\theta_i, 1)$, $j = 1, \dots, n_i$ 을 따른다. 만약 $W_{ij} > 0$ 이면 $Y_{ij} = 1$ 이고, $W_{ij} \leq 0$ 이면 $Y_{ij} = 0$ 이다. W_{ij} 는 알려져 있지 않지만 Y_{ij} 는 주어져 있으므로, W_{ij} 분포는 절단된 정규분포를 따른다고 할 수 있다. 이제 Y, W 가 주어진 θ, μ, δ_k 와 σ_k^2 의 결합사후밀도함수는

$$\begin{aligned} [\theta, \mu, \delta_k, \sigma_k^2 | Y, W] &\propto [W|\theta][\theta|\mu, \delta_k, \sigma_k^2][\mu, \delta_k, \sigma_k^2][Y|W, \theta] \\ &\propto [W|\theta][\theta|\mu, \delta_k, \sigma_k^2][\mu, \delta_k, \sigma_k^2][Y|W] \\ &\propto [W|\theta][\theta|\mu, \delta_k, \sigma_k^2][\mu, \delta_k, \sigma_k^2] \\ &\propto \prod_{i=1}^N \prod_{j=1}^{n_i} e^{-\frac{(w_{ij} - \theta_i)^2}{2}} \{\theta|\mu, \delta_k, \sigma_k^2\}[\mu, \delta_k, \sigma_k^2] \end{aligned} \quad (2.6)$$

이고, Y 는 W 에만 의존한다.

식(2.3) - (2.5)에서 정의된 세가지 형태의 모형들에 대한 결합사후 분포들은 식(2.6)을 이용하여 다음과 같이 표현할 수 있다. 첫번째로 모형 M_1 하에서 결합사후밀도함수, 즉 우도비함수 \times 사전분포는

$$\begin{aligned} &\prod_{i=1}^N \prod_{j=1}^{n_i} \exp\left\{-\frac{(w_{ij} - \theta_i)^2}{2}\right\} \prod_{k=1}^K \prod_{i=N_{k-1}+1}^{N_k} \left[\sigma_k^{-1} \exp\left\{-\frac{(\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}\right\}\right] \\ &\times \exp\left\{-\frac{(\mu - a)^2}{2b}\right\} \prod_{k=1}^K \exp\left\{-\frac{(\delta_k - c_k)^2}{2d_k}\right\} \frac{1}{(\sigma_k^2)^{e_k+1}} \exp\left\{-\frac{1}{f_k \sigma_k^2}\right\} \end{aligned} \quad (2.7)$$

에 비례하고, 모형 M_2 하에서, 결합사후밀도함수는

$$\begin{aligned} &\prod_{i=1}^N \prod_{j=1}^{n_i} \exp\left\{-\frac{(w_{ij} - \theta_i)^2}{2}\right\} \prod_{k=1}^K \prod_{i=N_{k-1}+1}^{N_k} \left[\sigma_k^{-1} \exp\left\{-\frac{(\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}\right\}\right] \\ &\times \exp\left\{-\frac{(\mu - a)^2}{2b}\right\} \prod_{k=1}^K \exp\left\{-\frac{(\delta_k - c_k)^2}{2d_k}\right\} \frac{s_k}{(s_k + \sigma_k^2)^2} \end{aligned} \quad (2.8)$$

에 비례한다. 또한 모형 M_3 하에서, 결합사후밀도함수는

$$\begin{aligned} & \prod_{i=1}^N \prod_{j=1}^{n_i} \exp\left\{-\frac{(w_{ij} - \theta_i)^2}{2}\right\} \prod_{k=1}^K \prod_{i=N_{k-1}+1}^{N_k} \left[\sigma_k^{-1} \exp\left\{-\frac{(\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}\right\}\right] \\ & \times \exp\left\{-\frac{(\mu - a)^2}{2b}\right\} \prod_{k=1}^K \exp\left\{-\frac{(\delta_k - c_k)^2}{2d_k}\right\} \left(1 + \frac{\sigma_k^2}{s_k^2}\right)^{-1} \end{aligned} \quad (2.9)$$

에 비례한다.

3. 계산상의 제안점

깁스샘플러는 완전조건 분포(Full Conditional Distribution)로부터 난수생성을 하여 반복수들을 늘려감으로써 각각의 난수들을 최신화(updated) 하고 그 최신화된 난수를 원하는 사후분포로부터 생성된 난수로 사용한다. 그러므로 식(2.7)로부터 깁스샘플러를 적용하기 위해서, 우선 모형 M_1 에 대한 완전조건부 분포들은 다음과 같다.

$$\begin{aligned} Y_{ij} = 1 \text{에 대하여} & \quad [W_{ij}|\theta] \propto N(\theta_i, 1), \quad W_{ij} > 0, \\ Y_{ij} = 0 \text{에 대하여} & \quad [W_{ij}|\theta] \propto N(\theta_i, 1), \quad W_{ij} \leq 0, \end{aligned} \quad (3.1)$$

$$[\sigma_k^2|\mu, \delta, \theta, y] = \text{IG} \left[\frac{N_k}{2} + e_k, \left(\frac{1}{f_k} + \frac{1}{2} \sum_{i=N_{k-1}+1}^{N_k} (\theta_i - (\mu + \delta_k))^2 \right)^{-1} \right], \quad (3.2)$$

$$[\delta_k|\mu, \sigma^2, \theta, y] = N \left[\frac{N_k d_k (\bar{\theta}_k - \mu) + c_k \sigma_k^2}{N_k d_k + \sigma_k^2}, \frac{d_k \sigma_k^2}{N_k d_k + \sigma_k^2} \right], \quad (3.3)$$

$$[\theta_i|\mu, \delta, \sigma^2, \theta_{(i)}, y] = N \left[\frac{\sigma_k^2 \sum_{j=1}^{n_i} w_{ij} + (\mu + \delta_k)}{n_i \sigma_k^2 + 1}, \frac{\sigma_k^2}{n_i \sigma_k^2 + 1} \right], \quad (3.4)$$

와

$$[\mu|\sigma^2, \delta, \theta, y] = N(M, V) \quad (3.5)$$

이다. 여기서,

$$y = (y_1, y_2, \dots, y_N)', \quad \theta = (\theta_1, \theta_2, \dots, \theta_N)', \quad \delta = (\delta_1, \dots, \delta_K)', \quad \sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)',$$

$$\theta_{(i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_N)', \quad \bar{\theta}_k = \frac{\sum_{i=N_{k-1}+1}^{N_k} \theta_i}{N_k}, \quad k = 1, \dots, K$$

$$M = \frac{\sum_{k=1}^K \sigma_{(k)}^2 b N_k (\bar{\theta}_k - \delta_k) + a \sigma_{1 \dots K}^2}{\sum_{k=1}^K \sigma_{(k)}^2 b N_k + \sigma_{1 \dots K}^2}, \quad V = \frac{b \sigma_{1 \dots K}^2}{\sum_{k=1}^K \sigma_{(k)}^2 b N_k + \sigma_{1 \dots K}^2},$$

$$\sigma_{1 \dots K}^2 = \prod_{i=1}^K \sigma_i^2, \quad \sigma_{(k)}^2 = \sigma_1^2 \sigma_2^2 \dots \sigma_{k-1}^2 \sigma_{k+1}^2 \dots \sigma_K^2, \quad k = 1, \dots, K.$$

또한, 모형 M_2 와 모형 M_3 에 대해 완전조건 분포함수들은 식(3.2)를 제외하고 모형 M_1 의 완전조건 분포함수들과 같다. 즉, 두 모형 M_2 와 M_3 에서, $[\delta_k | \mu, \sigma^2, \theta, y]$, $[\theta_i | \mu, \delta, \sigma^2, \theta_{(i)}, y]$, $[\mu | \sigma^2, \delta, \theta, y]$ 는 각각 식(3.3), (3.4)와 (3.5)에 있다. 모형 M_2 에 대해서는 식(3.2) 대신,

$$[\sigma_k^2 | \mu, \delta, \theta, y] \propto (\sigma_k^2)^{-\frac{N_k}{2}} \frac{s_k}{(s_k + \sigma_k^2)^2} \exp\left(-\frac{\sum_{i=N_{k-1}+1}^{N_k} (\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}\right) \quad (3.6)$$

이며, 모형 M_3 에 대해서는 식(3.2) 대신,

$$[\sigma_k^2 | \mu, \delta, \theta, y] \propto (\sigma_k^2)^{-\frac{N_k}{2}} \left(1 + \frac{\sigma_k^2}{s_k^2}\right)^{-1} \exp\left(-\frac{\sum_{i=N_{k-1}+1}^{N_k} (\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}\right) \quad (3.7)$$

이다. σ_k^2 을 생성할 때, 모형 M_2 와 모형 M_3 에 대해서, 완전조건부분포의 형태가 난수생성이 쉽지 않으므로 메트로폴리스 알고리즘을 이용해야만 한다. 메트로폴리스 알고리즘을 이용하기 위해, 적절한 후보-생성(candidate-generating) 밀도함수 (Chib and Greenberg, 1995)가 필요하다. 예를들어, 만약 $\pi(t)$ 는 $\pi(t) \propto \psi(t)h(t)$ 로 쓰여지고, 여기서 $h(t)$ 는 쉽게 난수 생성이 되는 밀도함수이고, $\psi(t)$ 는 균등하게 유계를 갖는다. 그래서 이런 후보들을 생성하기 위해서 $q(x, y) = h(y)$ 로 둔다. 이 경우에 전이확률은 ψ 함수 계산만을 요구한다. 이때 선택 확률은

$$\alpha(x, y) = \min\left\{\frac{\psi(y)}{\psi(x)}, 1\right\} \quad (3.8)$$

이다. 그러므로 모형 M_2 에 대해서, 식(3.6)을 통하여

$$h(\sigma_k^2) = \text{IG}\left(\frac{N_k}{2} - 1, \frac{2}{\sum_{i=N_{k-1}+1}^{N_k} (\theta_i - (\mu + \delta_k))^2}\right), \psi(\sigma_k^2) = \frac{s_k}{(s_k + \sigma_k^2)^2} \quad (3.9)$$

로 둘 수 있고, 이와 유사하게, 모형 M_3 에 대해서 식(3.7)을 통하여

$$h(\sigma_k^2) = \text{IG}\left(\frac{N_k}{2} - 1, \frac{2}{\sum_{i=N_{k-1}+1}^{N_k} (\theta_i - (\mu + \delta_k))^2}\right), \psi(\sigma_k^2) = \left(1 + \frac{\sigma_k^2}{s_k^2}\right)^{-1} \quad (3.10)$$

라 둘 수 있다. 그러므로 식(3.8)를 이용하면 모형 M_2 , M_3 에서 각각의 선택확률들을 식(3.9)와 (3.10)에서 주어진 ψ 의 식을 갖고 구할 수 있다.

이제 정규분포, $N(\theta_i, 1)$, $i = 1, \dots, N$,로부터 식(3.1)에서 W_{ij} 을 생성하는 방법을 소개한다. 먼저 $j = 1, \dots, y_i$ 에 대해서, $0 < W_{ij} < \infty$ 인 $N(\theta_i, 1)$ 로부터 W_{ij} 를 생성하기 위해서 변수 변환을 한다. $X_{ij} = W_{ij} - \theta_i$ 라 놓으면 $X_{ij} \sim N(0, 1)$, $-\theta_i < X_{ij} < \infty$ 가 된다. 일대일(one-for-one) 방법에 의해서 $X_{ij} = \Phi^{-1}[\Phi(-\theta_i) + u(\Phi(\infty) - \Phi(-\theta_i))] = \Phi^{-1}[\Phi(-\theta_i) + u(1 - \Phi(-\theta_i))]$ 가 된다. 여기서 u 는 구간 $(0, 1)$ 에서의 일양분포로부터 생성된 난수이다. 따라서 $W_{ij} = X_{ij} + \theta_i$ 이다. 이와 유사하게 $j = y_i + 1, \dots, n_i$ 에 대해서, $-\infty < W_{ij} \leq 0$ 인 $N(\theta_i, 1)$ 로부터 W_{ij} 를 생성하기 위해서 $X_{ij} = W_{ij} - \theta_i$ 라 두자. 즉, $X_{ij} \sim N(0, 1)$, $-\infty < X_{ij} \leq -\theta_i$ 이다. 앞에서와 같이 일대일(one-for-one) 방법을 이용하면, $X_{ij} = \Phi^{-1}[u(\Phi(-\theta_i) - \Phi(-\infty))] = \Phi^{-1}[u(\Phi(-\theta_i))]$ 가 된다. 그러므로 $W_{ij} = X_{ij} + \theta_i$ 이다.

김스샘플러는 다음 두단계 과정으로 이루어진다.

1단계 : 초기치 $\theta_i^{(0)}, \mu^{(0)}, \delta_k^{(0)}$ 와 $\sigma_k^{2(0)}, i = 1, \dots, N, k = 1, \dots, K$ 를 갖고 식(3.1)의 절단 정규분포로부터 $W^{(1)}$ 을 생성한다.

2단계 : 일반적인 깃스 반복은 다음과 같다.

$$\begin{aligned} \mu^{(1)} &\leftarrow [\mu | \sigma_k^{2(0)}, \delta_k^{(0)}, \theta_i^{(0)}, W^{(1)}; k = 1, \dots, K; i = 1, \dots, N] \\ k &= 1, \dots, K \text{에 대하여} \\ \sigma_k^{2(1)} &\leftarrow [\sigma_k^2 | \mu^{(0)}, \delta_k^{(0)}, \theta_i^{(0)}, W^{(1)}, \sigma_l^{2(1)}, l < k \text{ 와 } \sigma_l^{2(0)}, l > k] \\ \delta_k^{(1)} &\leftarrow [\delta_k | \sigma_k^{2(1)}, \mu^{(1)}, \theta_i^{(0)}, W^{(1)}, \delta_l^{(1)}, l < k, \delta_l^{(0)}, l > k; i = 1, \dots, N] \\ i &= 1, \dots, N \text{에 대하여} \\ \theta_i^{(1)} &\leftarrow [\theta_i | \sigma_k^{2(1)}, \delta_k^{(1)}, \mu^{(1)}, \theta_l^{(1)}, l < i, \text{ 와 } \theta_l^{(0)}, l > i, W^{(1)}; k = 1, \dots, K] \end{aligned}$$

여기서 화살표는 밀도함수로부터 난수 생성을 의미한다. 위 두 단계는 $\theta_i^{(0)}, \mu^{(0)}, \delta_k^{(0)}, \sigma_k^{2(0)}$ 와 $W^{(0)}$ 를 $\theta_i^{(1)}, \mu^{(1)}, \delta_k^{(1)}, \sigma_k^{2(1)}$ 와 $W^{(1)}$ 로 최신화하는 것이다. 베이지안 계산을 위해, 4개의 병렬체인을 사용한다. 깃스샘플러는 20,000번 반복하고, 메트로폴리스 알고리즘은 15,000번 반복한다. 처음 15,000번은 버리고, 나머지 5,000번 중 10번째 값만 사용한다. 난수 생성을 위해 Fortran IMSL을 이용하였다. 깃스 결과로부터 관심있는 추정치를 얻을 수 있다. $\{\theta_i^{(g)}, \delta_k^{(g)}, \sigma_k^{2(g)}, \mu^{(g)}\}_{g=1}^B$ 는 깃스 결과이고, 여기서 B 는 깃스샘플러를 B 번 반복했음을 의미한다. 예를 들어, 모형 M_1 에서, p_i 의 사후 평균은 $\hat{p}_i = \frac{1}{B} \sum_{g=1}^B \Phi(\theta_i^{(g)})$, $i = 1, \dots, N$ 로 추정된다.

4. 베이지안 모형선택

베이지안(Bayes factor, BF)와 유사베이지안(pseudo-Bayes factor, PsBF)를 사용하여 앞에서 주어진 서로 다른 세 모형 M_1, M_2 와 M_3 중 하나를 선택하는 방법을 설명한다.

4.1. 베이지안 인자

모형선택을 위하여 아래와 같은 H_0 와 H_1 중 하나를 선택하기로 가정하자.

$$H_0 : \text{model } M_1 \text{ is chosen vs. } H_1 : \text{model } M_2 \text{ is chosen.}$$

일반적인 베이지안 모형선택 절차는 다음과 같다. w_i 를 모형 M_i , $i = 1, 2$ 의 사전확률이라 두고, $\pi(x; M_i)$ 을 모형 M_i 의 예측(predictive)분포라 하자. 즉,

$$\{m\}_i = \pi(x; M_i) = \int \pi(x|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i.$$

만약 x 가 관측된 자료이면, $w_i\pi(x; M_i)$ 값이 더 큰 모형을 선택한다. 이때, 일반적으로 특별한 정보가 없으면 $w_i = \frac{1}{2}$ 로 두고, M_2 에 대한 M_1 의 베이지안인자를 계산한다.

$$BF_{12} = \frac{\pi(x; M_1)}{\pi(x; M_2)} = \frac{\{m\}_1}{\{m\}_2}. \tag{4.1}$$

Jeffrey (1961), Kass와 Raftery (1995)는 베이즈인자값에 대한 구간별 해석을 제안하였고, 일반적으로 $BF_{12} > 1$ 이면 모형 M_2 보다 모형 M_1 이 더 적합한 모형으로 선택된다. 더 일반적으로 설명하기 위해 $\pi(x|\theta)$ 는 X 의 분포 함수이고, $\pi(\theta)$ 은 θ 의 사전분포 함수이라 하자. 여기서 예측분포 함수 $[m] = \int \pi(x|\theta)\pi(\theta)d\theta$ 을 주표본 기법(importance sampling method)으로 근사하고자 한다. 이때, 주표본 함수(importance sampling function)로써 사후분포 함수 $\pi(\theta|x)$ 을 고려한다. 그리고 마코프체인 몬테칼로 방법(MCMC method)을 이용하여 예측분포 함수 $[m]$ 의 근사식을 구할 수 있다. 먼저, 메트로폴리스 알고리즘과 깁스샘플러로 사후분포 함수 $\pi(\theta|x)$ 로부터 난수를 생성한다. 이때 생성된 난수들을 깁스 결과(Gibbs output)라 하고 $\{\theta^{(g)}\}_{g=1}^B$ 로 표기한다. 여기서 B 는 깁스샘플러를 B 번 반복했음을 의미한다. 다음으로 몬테칼로 방법을 이용하면, 예측분포 함수의 근사식은 다음과 같이 구할 수 있다.

$$[\hat{m}] = \frac{\sum_{g=1}^B w_g \pi(x|\theta^{(g)})}{\sum_{g=1}^B w_g} \quad (4.2)$$

여기서, $w_g = \frac{\pi(\theta^{(g)})}{\pi(\theta^{(g)}|x)}$ 이다. 또한 $\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{[m]}$ 이므로, $w_g = \frac{[m]}{\pi(x|\theta^{(g)})}$ 로 표현 할 수 있다. 그러므로, 식(4.2)인 X 의 예측분포 함수의 근사식은 다음과 같이 다시 표현된다.

$$[\hat{m}] = \left[\frac{1}{B} \sum_{g=1}^B \frac{1}{\pi(x|\theta^{(g)})} \right]^{-1} \quad (4.3)$$

Kass와 Raftery (1995)는 식(4.3)를 언급하였고, Chung (1997)은 베이저안 접근법을 사용하여 이진 회귀모형의 연결함수를 선택하기위해 식(4.3)를 사용하였다. 예로써, 2장에서 논의된 모형 M_1 와 모형 M_2 의 베이즈인자의 근사식은 식(4.3)을 이용하면 다음과 같이 얻어진다.

$$BF_{12} = \frac{[m]_1}{[m]_2} = \frac{\left[\frac{1}{B} \sum_{g=1}^B \left\{ \prod_{i=1}^N \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w_{ij1}^{(g)} - \theta_{i1}^{(g)})^2}{2}} \right\}^{-1} \right]^{-1}}{\left[\frac{1}{B} \sum_{g=1}^B \left\{ \prod_{i=1}^N \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w_{ij2}^{(g)} - \theta_{i2}^{(g)})^2}{2}} \right\}^{-1} \right]^{-1}} \quad (4.4)$$

여기서 $w_{ij1}^{(g)}$ 와 $\theta_{i1}^{(g)}$ 는 모형 M_1 의 g 번째 깁스 결과이고, $w_{ij2}^{(g)}$ 와 $\theta_{i2}^{(g)}$ 는 모형 M_2 의 g 번째 깁스 결과이다.

4.2. 유사 베이즈 인자

유사 베이즈인자 PsBF (Geisser와 Eddy, 1979)는 다음과 같이 정의된다.

$$PsBF(M_1, M_2) = \frac{\prod_{r=1}^N \pi(x_r; M_1|x_{(r)})}{\prod_{r=1}^N \pi(x_r; M_2|x_{(r)})} \quad (4.5)$$

여기서 $\pi(x_r; M_i|x_{(r)})$ 는 $i = 1, 2$ 에 대한 모형 M_i 하에서의 교차확인 예측밀도함수(cross-validation predictive density)이며, $x_{(r)}$ 는 r 번째 자료가 제외된 것을 의미한다. 교차확인 예측밀도함수는 다음과 같다.

$$\pi(x_r|x_{(r)}) = \int \pi(x_r|\theta, x_{(r)})\pi(\theta|x_{(r)})d\theta . \quad (4.6)$$

이것은 조건부 예측좌표(conditional predictive ordinate, CPO)로 정의가 되고, Gelfand, Dey 와 Chang (1992)에 의하여 제안되었다. 다른 모형들에 비해 CPO값이 더 큰 모형을 선호한다. 사후분포 $\pi(\theta|x)$ 로부터 난수생성이 가능하고, 주어진 θ 에 대해 관측치가 조건적으로 독립일때, 즉, $\pi(x_r|\theta, x_{(r)}) = \pi(x_r|\theta)$ 이므로, 식(4.6)에 의해 CPO의 추정치가 계산된다. 또한 MCMC방법의 하나로 Dey, Chen 과 Chang (1997)은 식(4.5)의 추정치로 다음과 같이 제안하였다.

$$\hat{\pi}(x_r|x_{(r)}) = B\left[\sum_{g=1}^B \pi(x_r|\theta^{(g)})^{-1}\right]^{-1} \quad (4.7)$$

x_r 의 조건부 밀도함수의 조화평균은 사후 난수생성 값으로 계산된다. 식(4.5)와 식(4.7)을 이용하면, 2장에서 정의된 모형 M_1 와 모형 M_2 의 유사베이즈인자는 다음과 같이 얻어진다.

$$PsBF_{12} = PsBF(M_1, M_2) = \prod_{i=1}^N \frac{\left[\sum_{g=1}^B \left\{ \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w_{ij1}^{(g)} - \theta_{i1}^{(g)})^2}{2}} \right\}^{-1} \right]^{-1}}{\left[\sum_{g=1}^B \left\{ \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w_{ij2}^{(g)} - \theta_{i2}^{(g)})^2}{2}} \right\}^{-1} \right]^{-1}}$$

여기서 $w_{ij1}^{(g)}$ 와 $\theta_{i1}^{(g)}$ 는 모형 M_1 의 g 번째 깃스 결과이고, $w_{ij2}^{(g)}$ 와 $\theta_{i2}^{(g)}$ 는 모형 M_2 의 g 번째 깃스 결과이다.

5. 사례분석

5.1. 실제자료

Fisher와 Van Belle (1993)는 새로운 간질약(progabide)의 효과를 조사한 $N = 15$ 개의 독립인 임상실험을 수집하였다. 반응변수는 무처리(baseline)와 비교했을때, 환자의 발작 횟수가 50퍼센트 이상 감소했는가 하는것이다. 표5.1에서 Studies항의 열거된 번호는 본 연구에서 사용되어진 연구논문의 순서이고, original항은 원 자료의 연구 논문 순서이다. 여기서, n_i 은 i 번째 연구의 총 환자수, y_i 는 i 번째 연구의 호전된 환자 수, \hat{p}_i 는 i 번째 연구의 호전된 환자의 비, $\hat{\theta}_i = \Phi^{-1}(\hat{p}_i)$ 은 i 번째 연구에서 관측된 환자의 호전비에 대한 프라빗 함수 값이다. 예를들면, studies 1은 원 자료에서도 첫번째 자료로써 30명의 환자중 17명의 환자는 발작횟수가 50퍼센트 이상 감소하였음을 나타낸다. 그러므로 이 자료를 2장에서 정의된 프로빗 모형으로 구성하기 위해서 $j = 1, \dots, 17$ 에 대하여 $y_{1j} = 1$ 이고, $j = 18, \dots, 30$ 에 대

하여 $y_{ij} = 0$ 로 나타낼 수 있다. 같은 방법으로 모든 i 에 대하여 y_{ij} 값을 정할 수 있다. 처음 6개는 개방연구(open study)로, 나머지 9개는 폐쇄연구(closed study)로 실험이 이루어졌다. 여기서 개방연구는 단일눈가림(single blind) 임상실험인 반면, 폐쇄 연구는 이중눈가림(double blind) 임상실험을 의미한다.

표 5.1: 프로그가바이드(Progabide) 자료

	<i>studies</i>	<i>original</i>	n_i	y_i	\hat{p}_i	$\hat{\theta}_i$
개방 연구	1	1	30	17	0.5666	0.1677
	2	7	16	8	0.5000	0.0000
	3	9	69	41	0.5942	0.2383
	4	10	23	13	0.5652	0.1641
	5	11	42	32	0.7619	0.7124
	6	15	151	90	0.5960	0.2430
폐쇄 연구	7	2	20	5	0.2500	-0.6744
	8	3	20	9	0.4500	-0.1256
	9	4	17	3	0.1765	-0.9287
	10	5	15	7	0.4667	-0.0835
	11	6	18	8	0.4444	-0.1398
	12	8	17	9	0.5294	0.0737
	13	12	19	1	0.0526	-1.6201
	14	13	51	12	0.2353	-0.7215
	15	14	59	17	0.2881	-0.5589

5.2. 모형선택 결과

2장과 같이 세가지 모형 M_1 , M_2 와 M_3 을 고려하자. 주어진 세 모형 모두 식(2.2)는 $a = 0, b = 10, c_k = 0$ 와 $d_k = 10$ 으로 동일한 값을 갖는다. 모형 M_1 에 대해서, 식(2.3)은 $e_k = 0.001$ 와 $f_k = 1000$ 를 갖고, 모형 M_2 와 M_3 에 대해서, 식(2.4)와 식(2.5)의 s_k^2 는 그룹 k 의 표본분산의 조화평균으로 정한다.

Larose와 Dey(1997)는 같은 자료에 로짓모형을 이용하여 연구하였다. 그들의 결과는 서로 다른 세 모형에서 추정 값들의 차이가 거의 나타나지 않으므로 어느 모형을 선택하여도 무방하다. 그러나 본 연구인 프로핏 연결함수에서는 베이즈 인자 또는 유사 베이즈 인자를 이용한 세 모형의 차이가 뚜렷하므로 본 모형은 연결함수에 민감하다고 본 예를 통하여 알 수 있다. 이러한 관점에서 4장에서 제시한 모형 선택이 의미있다고 생각한다. 그러므로 4장에서 정의된 베이즈인자와 유사베이즈인자들의 근사식을 이용하여 주어진 세 모형 중 하나를 선택한다. 표 5.2는 BF 와 $PsBF$ 의 근사값을 나타낸다. 표 5.2에 의하면, BF_{21} 와 BF_{32} 모두 1보다 크고, 또한 $PsBF_{21}$ 와 $PsBF_{32}$ 들도 모두 1보다 크다. 그러므로 두 경우 모

표 5.1: 프라빗 연결 함수에서 BF 와 $PsBF$ 값

ij	BF_{ij}	$PsBF_{ij}$
21	$0.47717D + 02$	$0.68269D + 39$
32	$0.10532D + 17$	$0.54974D + 83$

표 5.2: 표 5.3: 모수들의 사후 성질들

모수	사후평균	95% 신뢰구간
σ_1^2	0.85912	(0.04261, 4.75419)
σ_2^2	0.68859	(0.10544, 2.47132)
μ	0.12729	(-3.51657, 4.02491)
δ_1	0.84118	(-3.02228, 4.53474)
δ_2	-0.07508	(-4.01330, 3.61432)
p_1	0.78706	(0.58018, 0.93906)
p_2	0.74203	(0.47614, 0.93479)
p_3	0.81781	(0.68606, 0.91753)
p_4	0.78833	(0.55783, 0.94465)
p_5	0.93634	(0.82428, 0.99238)
p_6	0.81924	(0.73287, 0.89350)
p_7	0.46056	(0.23297, 0.71667)
p_8	0.64683	(0.38968, 0.87766)
p_9	0.38301	(0.16782, 0.62925)
p_{10}	0.66139	(0.35829, 0.90696)
p_{11}	0.64704	(0.37496, 0.88539)
p_{12}	0.71761	(0.44606, 0.93499)
p_{13}	0.22043	(0.07233, 0.42513)
p_{14}	0.43367	(0.27951, 0.60110)
p_{15}	0.49351	(0.33487, 0.65095)
$\mu + \delta_1$	0.96847	(0.13309, 1.74886)
$\mu + \delta_2$	0.05220	(-0.53436, 0.63429)
WA_{open}	0.82501	(0.77034, 0.87503)
WA_{closed}	0.49936	(0.42647, 0.57637)
WA_{All}	0.68947	(0.64630, 0.73106)

두에서 $M_3 > M_2 > M_1$ 순으로 이루어 지므로 모형 M_3 가 세가지 모형들중 가장 잘 적합된 모형으로 선택된다. 세 가지 모형에서 각각 모수들의 베이저안 추정값들이 약간의 다른 값들로 나타난다. 그러나 여기서 모형 M_3 가 가장 잘 적합된 모형으로 선택되므로 M_3 에 대한 모수 추정값들을 표 5.3에 실고 그 값들로 해석하고자 한다.

위에서 선택된 모형 M_3 을 고려한다. 표 5.3 으로부터 σ_1^2 의 사후평균이 σ_2^2 의 그 것보다 크며, 또한 95% 신뢰구간도 σ_1^2 이 σ_2^2 보다 넓다. 그러므로 폐쇄연구가 개방연구보다 더 동질적인 그룹을 나타냄을 알수있다. 따라서, σ_1^2 로부터 개방연구에 대해 고정효과 모형 보다 임의효과 모형이 타당하다고 본다.

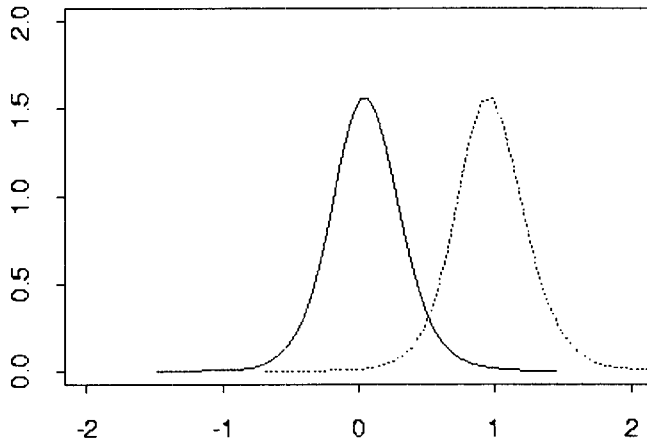


그림 5.1: $\mu + \delta_1$ 와 $\mu + \delta_2$ 의 사후분포. 개방연구: *dotted line*; 폐쇄연구: *solid line*.

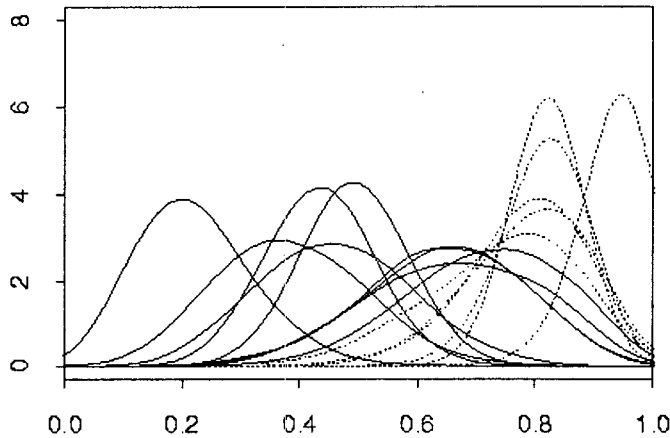


그림 5.2: p_i 의 사후분포. 개방 연구: *dotted lines*; 폐쇄 연구: *solid lines*.

$\mu + \delta_1$ 의 사후평균은 0보다 1에 가깝게 크나 $\mu + \delta_2$ 의 사후 평균은 0에 거의 가깝다. 또한 그림 5.1은 $\mu + \delta_1$ 와 $\mu + \delta_2$ 의 사후분포 도표이며, 이것으로도 $\mu + \delta_1$ 은 유의적으로 0보다 크다. 즉 개방연구에서 프로그바이드를 사용한 환자의 평균 호전률이 0.5보다 유의적으로 크다. 따라서 개방 연구는 프로그바이드의 효과가 있음을 나타낸다. 그림 5.2는 개방연구와 폐쇄연구의 p_i 에 대한 각각의 사후분포들을 나타낸다. 개방연구에 해당하는 p_i 의 사후 분포들은 모두 0.5보다 큰쪽으로 나타나며, 폐쇄연구중 5개 연구의 p_i 에 대한 사후 분포는 0.5보다 작게 나타난다.

마지막으로 Larose와 Dey (1997)에 의해 제안된 p_i 의 가중 평균을 다음과 같이 정의하고 설명한다.

$$WA_{all} = \sum_1^N n_i p_i / \sum_1^N n_i, \quad WA_{open} = \sum_1^{N_1} n_i p_i / \sum_1^{N_1} n_i \quad \text{and} \quad WA_{closed} = \sum_{N_1+1}^N n_i p_i / \sum_{N_1+1}^N n_i$$

은 각각 p_i 의 전체연구 가중평균, 개방연구 가중평균, 폐쇄연구 가중평균을 의미한다. 그림 5.3은 그러한 p_i 의 가중평균의 사후분포들을 그림으로 설명한다. p_i 의 값이 유의적으로 0.5보다 크므로, 개방연구와 전체연구에서 프로그바이드라는 신약은의 효과가 있다고 말할 수 있다.

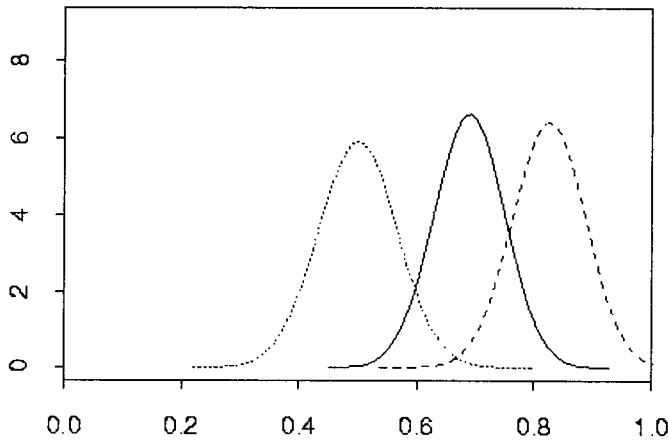


그림 5.3: p_i 의 가중 평균의 사후분포. 개방 연구: broken line; 폐쇄 연구: dotted line; 전체 연구: solid line.

6. 결론

본 논문에서는 메타분석에서 그룹화 임의효과 모형을 베이지안 관점에서 연구하였다. 그룹간 연구 뿐만 아니라 개별적 연구사이의 이질성을 설명할 수 있는 계층적 베이지안 구조가 제공되었다.

모든 자료는 이진자료이고, 연구들은 독립이라고 가정하기 때문에 프라빗 연결함수 $\theta_i = \Phi^{-1}(p_i)$ 에서 우도함수로써 이항분포가 정해졌다. 그룹의 임의효과를 구별하기 위하여 주로 θ_i 에 대해 정규분포를 사용한다. θ_i 의 정규분포의 모수로 평균과 분산에 대한 합리적인 사전분포를 각각 정규분포와 역감마 분포로 고려하였고, 계산상의 기술로써 Albert와 Chib (1993)의 방법을 통해 결합사후분포에 대해 잠재변수 W 를 정의하였다. 깁스샘플러와 메트로폴리스 알고리즘이 그룹화 임의효과 모형에 대한 베이지안 계산상의 어려움을 피하기 위해 이용된다. 주어진 모형의 여러 다른 사전분포중 가장 좋은 사전분포를 선택하기 위한 베이즈인자와 유사베이즈인자의 근사식을 제시한다. 예로서, 새로운 간질약인 프로가바이드의 효과를 조사한 자료를 사용한다.

앞으로의 과제로서, 본 논문에서 제안된 방법이 연결함수에 이론적으로 민감한지를 확인하고, 임의의 두 연결함수중 큰 값을 비교하는 베이즈인자의 측면에서 최선의 연결함수를 찾아내는 몇가지 기준이 제시되어야 할 것이다. 즉 어느 연결함수를 이용하는가 하는 문제이다. 또한, 초사전(hyper prior) 함수에 민감하다는 결론을 얻으므로, 연결함수와 초사전함수에 대한 민감성(robustness)의 연구도 앞으로의 과제라 생각한다.

참고문헌

- [1] Albert, A. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, 88, 669-679.
- [2] Carlin, B. (1992). Comment on hierarchical models for combining information and for meta-analyses, *Bayesian Statistics*, 4, Oxford, U.K., Oxford University Press.
- [3] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hasting algorithm, *American Statistician*, 49, 327-335.
- [4] Chung, Y. (1997). Sampling based approach to Bayesian analysis of binary regression model with incomplete data, *Journal of the Korean Statistical Society*, 26, 4, 493-503.
- [5] Dey, D., Chen, M-H. and Chang, H. (1997). Bayesian inference of nonlinear random effects model, *Biometrics*, 53, 1239-1252.
- [6] DuMouchel, W. (1994). Hierarchical Bayes linear models for meta-analysis, *Technical Report*, 27, National Institute of Statistical Sciences.
- [7] Fisher, L. and Van Belle, G. (1993). *Biostatistics: A Methodology for the Health Sciences*, New York, Wiley.
- [8] Geisser, S. and Eddy, W. (1976). A predictive approach to model selection, *Journal of the American Statistical Association*, 74, 153-160.
- [9] Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal

densities, *Journal of the American Statistical Association*, 85, 398-409.

- [10] Gelfand, A., Dey, D. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods, *Bayesian Statistics*, 4, 147-165.
- [11] Glass, G.V. (1976). Primary, secondary, and meta-analysis of research, *Educational Researcher*, 5, 3-8.
- [12] Hedges, L. (1983). A random effects model for effect sizes, *Psychological Bulletin*, 93, 388-395.
- [13] Jeffreys, H. (1961). *Theory of Probability* (3rd edition), London, Oxford University Press.
- [14] Kass, R.E. and Raftery, A.E. (1995). Bayes factors, *Journal of the American Statistical Association*, 90, 773-795.
- [15] Larose, D. and Dey, D. (1997). Grouped random effects models for Bayesian meta-analysis, *Statistics in Medicine*, 16, 1817-1829.
- [16] Morris, C. and Normand, S. (1992). Hierarchical models for combining information and for meta-analyses, *Bayesian Statistics*, 4, Oxford U.K., Oxford University Press.

[1999년 2월 접수, 1999년 9월 채택]

Bayesian Analysis of Grouped Random Effects Model in Meta Analysis

Younshik Chung¹⁾ Hojin Jeong²⁾

ABSTRACT

In this paper, we consider the grouped random effects model(GREM) in meta analysis which widely is used in the field of medicine. Since usually the number of studies is small in medical data, we investigate the Bayesian approach to GREM with probit link function under the three different plausible priors. When the probit link is applied to GREM, the structure of joint distribution is very complicated. Thus, the latent variables are defined. For selection of priors, Bayes factor and pseudo-Bayes factor are employed. In order to avoid the difficult Bayesian computation, the Markov chain Monte Carlo method(Gibbs sampler and Metropolis algorithm) is applied. Finally, our proposed methodology is explained and applied to real data which contains 15 independent studies of the efficacy of progabide.

Keywords: Bayes factor; Gibbs sampler; Grouped random effects model; Metropolis algorithm; Probit link function.

1) Associate Professor, Department of Statistics, Pusan National University.

E-mail: yschung@hyowon.pusan.ac.kr

2) Researcher, Forecast Research Lab., Meteorological Research Institute.