

Estimation Methods for Population Pharmacokinetic Models using Stochastic Sampling Approach

Kwang-Hee Kim^a · Jeong-Hwa Yoon^a · Eun-Kyung Lee^{a,1}

^aDepartment of Statistics, Ewha Womans University

(Received March 5, 2015; Revised March 30, 2015; Accepted March 30, 2015)

Abstract

This study is about estimation methods for the population pharmacokinetic and pharmacodynamic model. This is a nonlinear mixed effect model, and it is difficult to find estimates of parameters because of non-linearity. In this study, we examined theoretical background of various estimation methods provided by NONMEM, which is the most widely used software in the pharmacometrics area. We focused on estimation methods using a stochastic sampling approach - IMP, IMPMAP, SAEM and BAYES. The SAEM method showed the best performance among methods, and IMPMAP and BAYES methods showed slightly less performance than SAEM. The major obstacle to a stochastic sampling approach is the running time to find solution. We propose new approach to find more precise initial values using an ITS method¹ to shorten the running time.

Keywords: Nonlinear mixed effect model, stochastic sampling, maximum likelihood estimation method, EM algorithm.

1. 서론

약동학 모형은 약물 투여 후 시간에 따른 혈중농도의 변화를 이용하여 약물이 체내에서 흡수, 분포, 대사, 그리고 배설되는 과정을 약물의 특성을 나타내는 약동학 모수들을 이용하여 모형화한 것이다. 이들 약동학 모수들은 개인마다 다르게 나타나며 개인 간 차이는 인구학적 특성이나 병리학적 특성 등으로 설명될 수 있다. 집단 약동학 모형은 개인 내 변동과 개인 간 변동을 고려한 비선형 혼합효과모형의 형태로 구현된다 (Lee, 2010). $-2\log\text{likelihood}$ 를 목적함수로 정의하고 이를 최소화하는 최대우도추정치를 구하는 것이 목표이나 비선형 혼합효과모형의 경우 목적함수가 모수들의 비선형 형태와 적분형태를 포함하고 있어 일반적인 모형추정방법으로는 추정이 쉽지 않다.

NONMEM (Beal 등, 1992)은 현재 집단 약동/약력학 모형 추정을 위하여 가장 널리 쓰이고 있는 소프트웨어로 다양한 모형추정방법을 제공하고 있다. NONMEM에서 제공하고 있는 모형 추정방법은 크게 세 가지로 나누어진다. 첫 번째는 선형근사방법을 이용하는 것이다. 테일러시리즈 전개를 이용하여 목

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0003840).

¹Corresponding author: Department of Statistics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul, 120-750, Korea. E-mail: lee.eunk@ewha.ac.kr

적합수를 구하고자하는 모수에 대하여 선형근사식으로 바꾸어준 후 추정을 하는 방법이다. 전개를 하는 방법과 제약조건에 따라 FO 방법과 FOCE 방법, 그리고 Laplacian 방법 등이 있다 (Lee, 2010). 이들은 고전적인 방법으로 현재까지도 가장 많이 쓰이고 있는 방법이다. 그러나 선형근사를 이용하므로 근사식과 정확한 우도함수의 차이에서 오는 오차로 인하여 추정 모수에 대한 정확도가 떨어지게 된다.

두 번째 방법은 EM 알고리즘 (Dempster 등, 1977)을 이용하여 추정하는 방법이다. 이는 선형근사방법과는 달리 정확한 우도함수를 이용하게 되므로 모수를 정확하게 추정할 수 있다. 그러나 조건부 기대값을 구하는 기대화 단계(E-step)에서 모수에 대한 비선형 함수의 적분을 계산해야 하지만 이 적분 또한 모형이 비선형의 혼합효과모형이므로 정확한 값을 구하기가 쉽지 않다. 이를 해결하기 위하여 확률적 표본추출(stochastic sampling)을 이용한 방법들이 개발되었다. IMP와 IMPMAP 방법에서는 중요도 표본(importance sampling) 방법으로 추출한 표본을 이용하여 적분값을 계산한다. SAEM(Stochastic Approximation EM; Kuhn과 Lavielle, 2005) 방법은 기대화 단계를 시뮬레이션 단계와 확률근사 단계로 나누어 수행함으로써 적분을 계산하고 이들을 최대화 단계와 함께 반복적으로 수행하여 모수를 추정하게 된다. 세 번째 방법은 베이지안 추정을 이용하는 방법으로 모수에 대한 사전분포를 가정하고 사후분포로부터 표본을 추출하여 추정하는 방법이다.

현재까지도 선형 근사를 이용한 고전적인 방법이 많이 쓰이고 있기는 하나, 오차로 인한 정확도의 문제가 자주 제기되고 있고 또한 초기치에 따라 수렴여부가 크게 달라지게 되어 복잡한 모형일수록 추정에 어려움이 커진다. EM 알고리즘이나 베이지안을 이용한 방법은 2010년 이후에 NONMEM에 제공된 새로운 방법이다 (Bauer, 2010). 초기에는 확률적 표본추출로 인하여 계산시간이 오래 걸린다는 점과 사전분포 가정에 대한 거부감 등으로 많이 이용되지 않았으나 컴퓨터의 성능이 향상되고 또한 초기치에 크게 영향을 받아 수렴하지 않거나 혹은 오버플로우(overflow) 등의 문제로 프로그램이 정지하는 경우가 많은 선형근사방법에 비해 초기치의 영향을 덜 받고 좀 더 안정적으로 추정을 해 주어 사용자의 입장에서 모수의 추정이 쉬워짐에 따라 점점 사용빈도가 늘어나고 있다. 본 연구에서는 NONMEM에서 제공하고 있는 추정방법들의 특성을 이론적으로 살펴보고 이들을 기존의 선형근사 방법과 모의시험을 통하여 모형추정에서의 성능을 비교, 분석하여 집단 약동/약력학 모형 추정을 위한 최적의 방안을 제시하고자 한다.

2. 집단 약동학 모형의 정의

집단 약동학 모형을 일반적으로 정의하면 다음과 같다.

$$\begin{aligned} y_{ij} &= g(\text{dose}_i, \phi_i, t_{ij}) + h(\text{dose}_i, \phi_i, t_{ij})\varepsilon_{ij}, \\ \phi_i &= \mu(\theta, x_i) + \eta_i, \\ i &= 1, \dots, N, \quad j = 1, \dots, n_i, \\ \eta_i &\sim N_p(0, \Omega), \\ \varepsilon_{ij} &\sim N(0, \sigma^2), \end{aligned} \quad (2.1)$$

여기서 y_{ij} 는 i 번째 피험자의 j 번째 시간 t_{ij} 에 관측된 혈중농도를, $g(\text{dose}_i, \phi_i, t_{ij})$ 는 약동학 모형을 표현한 함수를 나타낸다. dose_i 는 주입된 약물의 양을 나타내며, ϕ_i 는 i 번째 피험자의 약동학 모수를 나타내며 이는 집단 약동학 모수인 θ 와 피험자의 특성을 나타내는 공변수 x_i 에 대한 함수인 $\mu(\theta, x_i)$ 와 개인 간 변동을 나타내는 η_i 로 표현된다. Ω 는 η_i 의 분산으로 개인 간 변동을 설명해준다. 위의 모형에서 $h(\text{dose}_i, \phi_i, t_{ij}) = 1$ 인 경우는 선형적 오차를 갖는 모형이 되며 $h(\text{dose}_i, \phi_i, t_{ij}) = g(\text{dose}_i, \phi_i, t_{ij})$ 인 경우에는 비례적 오차를 갖는 모형이 된다.

이때의 우도함수는 다음과 같이 정의될 수 있다.

$$\begin{aligned} L(\theta, \Omega, \sigma^2; y) &= \prod_{i=1}^N \int p(y_i, \phi_i; \theta, \Omega, \sigma^2) d\phi_i \\ &= \prod_{i=1}^N \int p(y_i; \phi_i, \sigma^2) p(\phi_i; \theta, \Omega) d\phi_i. \end{aligned} \quad (2.2)$$

식 (2.2)의 우도함수를 최대로 하는 θ , Ω , 그리고 σ^2 를 구하는 것이 목표이다. 그러나 우도함수가 구하고자 하는 모수들의 비선형형태일 뿐만 아니라 ϕ_i 에 대한 적분을 풀어야 하는 상황이므로 최대우도추정치를 직접 구하는 것은 쉽지 않다.

이를 해결하기 위하여 여러가지 방법들이 제안되었다. 이들 방법은 크게 선형근사를 이용한 방법, EM 알고리즘을 이용한 방법, 그리고 베이지안 방법의 세 가지로 분류될 수 있다. 본 연구에서는 각각의 방법들을 이론적으로 자세히 살펴보고자 한다.

2.1. 선형근사를 이용한 방법

선형근사 방법에서는 우도함수에 나타나는 적분을 풀기 위하여 η_i 에 대한 선형근사를 이용하는 방법으로 NONMEM에서는 근사방법과 정도에 따라 FO(first order), FOCE(first order conditional estimation), 그리고 Laplacian(LAP)의 세 가지 방법을 제공하고 있다 (Wang, 2007).

먼저 y_i 와 ϕ_i 의 결합분포에 대한 로그우도함수를 다음과 같이 나타내자.

$$\begin{aligned} \log p(y_i, \phi_i; \theta, \Omega, \sigma^2) &= \text{const} - \frac{1}{2} \log |V_i(\phi_i, \sigma^2)| - \frac{1}{2} (y_i - g_i(\phi_i))^T V_i(\phi_i, \sigma^2)^{-1} (y_i - g_i(\phi_i)) \\ &\quad - \frac{1}{2} \log |\Omega| - \frac{1}{2} (\phi_i - \mu(\theta, x_i))^T \Omega^{-1} (\phi_i - \mu(\theta, x_i)) \\ &= l_{c_i}(\theta, \Omega, \sigma^2; y_i, \phi_i), \end{aligned} \quad (2.3)$$

여기서 $V_i(\phi_i, \sigma^2) = \text{diag}(h(\text{dose}_i, \phi_i, t_{ij})^2 \sigma^2)$, $g_i(\phi_i) = [g(\text{dose}_i, \phi_i, t_{i1}), \dots, g(\text{dose}_i, \phi_i, t_{in_i})]$ 이다. 그러므로 목적함수는

$$\begin{aligned} \text{OBJ} &= -2l(\theta, \Omega, \sigma^2; y) \\ &\approx -2 \sum_{i=1}^N l_{c_i}(\theta, \Omega, \sigma^2; y_i, \eta_{i0}) + \log |\Omega| + \eta_{i0}^T \Omega^{-1} \eta_{i0} + \log |\Omega^{-1} - \nabla^2 l(\eta_{i0})| \\ &\quad - (\Omega^{-1} \eta_{i0} - \nabla l(\eta_{i0}))^T (\Omega^{-1} - \nabla^2 l(\eta_{i0}))^{-1} (\Omega^{-1} \eta_{i0} - \nabla l(\eta_{i0})) \end{aligned} \quad (2.4)$$

로 나타낼 수 있다. 이는 우도함수를 $\eta_i = \eta_{i0}$ 에 대하여 테일러 급수전개를 시킨 것으로

$$\begin{aligned} l_{c_i}(\theta, \Omega, \sigma^2; y_i, \eta_{i0}) &= \log(p(y_i, \phi_i; \theta, \Omega, \sigma^2))_{\eta_i = \eta_{i0}}, \\ \nabla l_i(\eta_{i0}) &= \left[\frac{\partial}{\partial \eta_i} l_{c_i}(\theta, \Omega, \sigma^2; y_i, \phi_i) \right]_{\eta_i = \eta_{i0}} \\ \nabla^2 l_i(\eta_{i0}) &= \left[\frac{\partial^2}{\partial \eta_i \partial \eta_i^T} l(\theta, \Omega, \sigma^2; y_i, \phi_i) \right]_{\eta_i = \eta_{i0}} \end{aligned} \quad (2.5)$$

이다.

FO 방법은 $\eta_i = 0$ 에 대하여 선형근사를 시킨 우도함수를 이용하는 방법으로 이차 미분을 피하기 위하여 $\nabla^2 l(0)$ 대신 $E_{Y_i} [\nabla l(\eta_i) \nabla l(\eta_i)^T]_{\eta_i=0}$ 을 이용한다.

$$\begin{aligned} \text{OBJ}_{FO} = & \sum_{i=1}^N -2l_c(\theta, \Omega, \sigma^2; y_i, \eta_{i0} = 0) + \log |\Omega| + \log \left| \Omega^{-1} - E \left[\nabla l(\eta_i) \nabla l(\eta_i)^T \right]_{\eta_i=0} \right| \\ & - \nabla l(0)^T \left(\Omega^{-1} - E_{Y_i} \left[\nabla l(\eta_i) \nabla l(\eta_i)^T \right]_{\eta_i=0} \right)^{-1} \nabla l(0). \end{aligned} \quad (2.6)$$

FOCE 방법은 $(\partial/\partial\eta_i)l_{c_i}(\theta, \Omega, \sigma^2|y_i, \phi_i) = 0$ 의 해인 $\eta_i = \hat{\eta}_i$ 에 대하여 선형근사를 시킨 우도함수를 사용하는 방법으로 FO에서와 마찬가지로 이차 미분을 피하기 위하여 $\nabla^2 l(\hat{\eta}_i)$ 대신 $E_{Y_i} [\nabla l(\eta_i) \nabla l(\eta_i)^T]_{\eta_i=\hat{\eta}_i}$ 을 이용한다. 이때 $\nabla l(\hat{\eta}_i) = 0$ 이 되므로 목적함수는 다음과 같이 정의할 수 있다.

$$\begin{aligned} \text{OBJ}_{FOCE} = & \sum_{i=1}^N -2l_c(\theta, \Omega, \sigma^2; y_i, \eta_{i0} = \hat{\eta}_i) + \log |\Omega| + \hat{\eta}_i^T \Omega^{-1} \hat{\eta}_i \\ & + \log \left| \Omega^{-1} - E_{Y_i} \left[\nabla l(\eta_i) \nabla l(\eta_i)^T \right]_{\eta_i=\hat{\eta}_i} \right|. \end{aligned} \quad (2.7)$$

Laplacian 방법은 FOCE와 마찬가지로 $\eta_i = \hat{\eta}_i$ 에 대하여 선형근사를 시킨 우도함수를 사용하는 방법으로 $\nabla^2 l_i(\hat{\eta}_i)$ 에서의 이차미분을 그대로 사용한다.

$$\begin{aligned} \text{OBJ}_{LAP} = & \sum_{i=1}^N -2l_{c_i}(\theta, \Omega, \sigma^2; y_i, \eta_{i0} = \hat{\eta}_i) + \log |\Omega| + \hat{\eta}_i^T \Omega^{-1} \hat{\eta}_i \\ & + \log \left| \Omega^{-1} - \nabla^2 l_i(\hat{\eta}_i) \right|. \end{aligned} \quad (2.8)$$

FO, FOCE, 그리고 Laplacian 방법에서는 위에서 정의한 각각의 목적함수를 이용하여 이를 최소화하는 모수를 추정하게 된다.

3. EM 알고리즘과 확률적 표본추출방법을 이용한 추정방법

EM 알고리즘 (Dempster 등, 1977)은 구하고자 하는 모수의 최소충분통계량에 대한 조건부 기대값을 계산하는 E-step과 이를 최대화하는 M-step을 반복적으로 수행하여 최대우도추정치를 구하는 방법으로 식 (2.1)의 모형에 적용하면 다음과 같이 정리할 수 있다.

E-step

$$S_{m,k+1} = E \left[\tilde{S}_m(y, \phi) | y, \theta_k, \Omega_k, \sigma_k^2 \right], \quad m = 1, \dots, M, \quad (3.1)$$

여기서 $\tilde{S}_m(y, \phi)$ 는 최소충분통계량을 나타낸다.

식 (2.1)의 집단 약동학 모형에서는

$$\begin{aligned} \tilde{S}_1(y, \phi) &= \sum_{i=1}^N \phi_i, \\ \tilde{S}_2(y, \phi) &= \sum_{i=1}^N \phi_i \phi_i^T, \\ \tilde{S}_3(y, \phi) &= \sum_{i=1}^N \sum_{j=1}^{n_i} \left(\frac{y_{ij} - g(\text{dose}_i, \phi_i, t_{ij})}{h(\text{dose}_i, \phi_i, t_{ij})} \right)^2 \end{aligned} \quad (3.2)$$

를 이용한다.

M-step

$$(\theta_{k+1}, \Omega_{k+1}, \sigma_{k+1}^2) = \arg \max L(\theta, \Omega, \sigma^2; S_{1,k+1}, \dots, S_{M,k+1}). \quad (3.3)$$

EM 알고리즘은 위의 E-step과 M-step을 반복적으로 수행하여 θ , Ω , 그리고 σ^2 의 최대우도추정치를 구하는 방법이다.

먼저

$$\begin{aligned} \bar{\phi}_i &= E[\phi_i | y, \theta_k, \Omega_k, \sigma_k^2] = \int \phi_i p(\phi_i | y, \theta_k, \Omega_k, \sigma_k^2) d\phi_i, \\ \bar{B}_i &= E[(\phi_i - \bar{\phi}_i)(\phi_i - \bar{\phi}_i)^T | y, \theta_k, \Omega_k, \sigma_k^2] = \int (\phi_i - \bar{\phi}_i)(\phi_i - \bar{\phi}_i)^T p(\phi_i | y, \theta_k, \Omega_k, \sigma_k^2) d\phi_i \end{aligned}$$

라고 정의하자. 여기서 $\bar{\phi}_i$ 를 구하기 위해서는 ϕ_i 의 사후분포가 필요하나 집단 약동학모형에서는 ϕ_i 의 사후분포가 알려진 분포의 형태가 아닌 복잡한 형태가 되며 이로 인하여 E-step의 수행이 쉽지 않다. 1990년대 이후 사후분포의 기대치를 구하는 E-step을 해결하기 위하여 다양한 방안들이 제안되어 왔으며 (Celeux와 Diebolt, 1992; Delyon 등, 1999) 집단 약동학 모형 추정에서는 2000년대 이후 본격적으로 이들을 이용한 추정방법을 이용하기 시작하였다 (Kuhn과 Lavielle, 2005; Mentré, 2008). IMP와 IMPMAP 방법에서는 주표본기법(importance sampling method)를 이용, SAEM 방법에서는 확률적 근사(stochastic approximation) 방법을 이용하고 있다. 이들 방법을 좀 더 자세히 살펴보기로 한다. 본 논문에서는 집단 약동학 모형의 중요모수인 θ 와 Ω 를 증점적으로 고려한다.

3.1. IMP와 IMPMAP

IMP(IMPOrtance sampling)과 IMPMAP(IMPOrtance sampling with Mode A Posteriori) 방법은 NONMEM에서 제공되고 있는 추정방법으로 몬테칼로 적분(Monte Carlo integration)을 이용하여 E-step의 적분을 계산한다. 식 (3.1)의 기대값 계산에는 ϕ_i 의 사후분포로부터의 표본 추출이 필요하다. 그러나 식 (2.1)과 같은 모형에서 사후분포는 복잡한 형태를 하고 있으며 이로부터 직접 표본을 추출하는 것은 쉽지 않다. 이를 위하여 주표본 기법(importance sampling method; Robert와 Casella, 2013)을 이용하고 있다. NONMEM에서는 초기 표본생성함수로 $N(\hat{\phi}_i, (\Omega^{-1} + \hat{S}_i^{-1})^{-1})$ 를 이용한다. 여기서 $\hat{\phi}_i$ 는 사후최빈값으로 $p(y_i, \phi_i; \theta, \Omega, \sigma^2)$ 의 최빈값을 나타내며, \hat{S}_i^{-1} 는 $p(y_i | \hat{\phi}_i, \sigma^2)$ 의 정보행렬(information matrix)로 $-\nabla l(\eta_{i0})$ 를 나타낸다. 이는 식 (2.7)로부터 유추된 것으로 ϕ_i 에 대한 초기 정보로 유용하게 쓰인다.

IMPMAP 방법에서는 이후의 반복적인 알고리즘에서도 사후최빈값을 이용한 초기표본생성함수로 부터 표본을 추출한다. 그러나 IMP 방법에서는 초기 이후에는 $N(\bar{\phi}_i^{(k-1)}, \bar{B}_i^{(k-1)})$ 로 부터 표본을 추출한다. IMP 방법의 알고리즘을 자세히 살펴보면 다음과 같다.

E-step of k^{th} iteration

1. $k = 1$ 인 경우

- $N(\hat{\phi}_i, (\Omega^{-1} + \hat{S}_i^{-1})^{-1})$ 로 부터 $\phi_{i,r}^{(k)}$ 의 표본을 추출한다 ($r = 1, \dots, R$).

여기서 $\hat{\phi}_i$ 는 사후최빈값으로 $p(y_i, \phi_i; \theta, \Omega, \sigma^2)$ 의 최빈값을, \hat{S}_i^{-1} 는 $p(y_i | \hat{\phi}_i, \sigma^2)$ 의 정보행렬(information matrix)를 나타낸다.

$$- u_{i,r}^{(k)} = \frac{p(\phi_{i,r}^{(k)} | y_i)}{p(\phi_{i,r}^{(k)}; \hat{\phi}_i, (\Omega^{-1} + \hat{S}_i^{-1})^{-1})} \text{을 계산한다.}$$

2. $k > 1$ 인 경우

$$- N(\bar{\phi}_i^{(k-1)}, \bar{B}_i^{(k-1)}) \text{로 부터 } \phi_{i,r}^{(k)} \text{의 표본을 추출한다 } (r = 1, \dots, R).$$

$$- u_{i,r}^{(k)} = \frac{p(\phi_{i,r}^{(k)} | y_i)}{p(\phi_{i,r}^{(k)}; \bar{\phi}_i, \bar{B}_i^k)} \text{을 계산한다.}$$

$$3. z_{i,r}^{(k)} = \frac{u_{i,r}^{(k)}}{\sum_{r=1}^R u_{i,r}^{(k)}}.$$

$$4. \bar{\phi}_i^{(k)} = \sum_{r=1}^R \phi_{i,r}^{(k)} z_{i,r}^{(k)}.$$

$$5. \bar{B}_i^{(k)} = \sum_{r=1}^R (\phi_{i,r}^{(k)} - \bar{\phi}_i^{(k)}) (\phi_{i,r}^{(k)} - \bar{\phi}_i^{(k)})^T z_{i,r}^{(k)}.$$

M-step of k^{th} iteration

식 (3.3)을 최소화하기 위하여 다음의 헤시안 행렬(Hessian matrix)를 이용한다.

$$1. H_{\theta^{(k)}} = \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \theta} \right) \Omega^{-1} \left(\frac{\partial \mu_i}{\partial \theta} \right)^T$$

$$2. g_{\theta^{(k)}} = \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \theta} \right) \Omega^{-1} (\bar{\phi}_i^{(k)} - \hat{\mu}_i^{(k)})$$

$$3. \theta^{(k+1)} = \theta^{(k)} + H_{\theta^{(k)}}^{-1} \cdot g_{\theta^{(k)}}$$

$$4. \Omega^{(k+1)} = \frac{1}{N} \sum_{i=1}^N (\bar{\phi}_i - \hat{\mu}_i^{(k+1)}) (\bar{\phi}_i - \hat{\mu}_i^{(k+1)})^T + \frac{1}{N} \sum_{i=1}^N \bar{B}_i^{(k)}$$

여기서 $\hat{\mu}_i^{(k+1)} = \mu(\theta^{(k+1)}, x_i)$ 를 나타낸다.

IMP MAP 방법에서는 E-step의 과정 1을 모든 k 에 대하여 수행하는 것으로 EM 알고리즘의 모든 과정에서 사후최빈값을 이용하여 표본을 추출한다. 이후의 나머지 과정인 E-step의 과정 3, 4, 5, 그리고 M-step은 IMP 방법과 동일하다.

3.2. ITS

ITS(Iterative Two Stage)는 EM 알고리즘에서의 E-step에서 기대치를 구하는 대신 사후최빈값을 이용하는 방법으로 적분을 수행하지 않고 한 점의 값으로 대체한다. 확률표본을 이용하지 않고 하나의 값만을 이용하므로 정확한 추정치를 구하기 힘들나 상대적으로 추정에 필요한 계산의 속도는 매우 빠르다.

Algorithm of k^{th} iteration

$$1. \bar{\phi}_i^{(k)} = \hat{\phi}_i.$$

2. $\bar{B}_i^{(k)} = \hat{B}_i = \left(\Omega^{(k)-1} + \hat{S}_i^{-1} \right)^{-1}$.
3. $\theta^{(k+1)} = \theta^{(k)} + H_{\theta^{(k)}}^{-1} \cdot g_{\theta^{(k)}}$.
4. $\Omega^{(k+1)} = \frac{1}{N} \sum_{i=1}^N \left(\bar{\phi}_i^{(k)} - \hat{\mu}_i^{(k+1)} \right) \left(\bar{\phi}_i^{(k)} - \hat{\mu}_i^{(k+1)} \right)^T + \frac{1}{m} \sum_{i=1}^N \bar{B}_i^{(k)}$.

위의 과정을 수렴할 때까지 반복한다.

3.3. SAEM

SAEM(Stochastic Approximation EM)방법은 E-step의 적분을 모의시험단계(simulation step)과 확률적 근사 단계(stochastic approximation step)로 나누어 수행함으로써 해결하는 방법으로 M-step은 IMP에서와 같은 방법으로 수행한다. 일반적인 SAEM 알고리즘의 모의시험단계와 확률적 근사단계는 다음과 같다 (Kuhn과 Lavielle, 2005).

Simulation step of k^{th} iteration

1. $p(\phi_i | y, \theta_k, \Omega_k, \sigma_k^2)$ 로부터 $\phi_i^{(k)}$ 의 표본을 추출한다 ($i = 1, \dots, N$).
그러나 대부분의 경우 $p(\phi_i | y, \theta_k, \Omega_k, \sigma_k^2)$ 로부터 직접 표본을 추출하는 것은 불가능하므로 Metropolis-Hastings 알고리즘 (Robert와 Casella, 2013)을 이용하여 표본을 추출한다.

Stochastic approximation of k^{th} iteration

1. $S_{m,k+1} = S_{m,k} + \gamma_k \left(\tilde{S}_m(y, \phi^{(k)}) - S_{m,k} \right)$,

여기서 γ_k 는 알고리즘의 수렴속도와 관련이 있는 $[0, 1]$ 사이의 값으로 K 번의 반복 이후에서는 $\gamma_k = (k - K)^{-1}$ 을 이용한다. K 는 대체로 50~100사이로 결정한다. M-step은 IMP에서와 같은 방식으로 수행한다.

NONMEM에서는 SAEM의 모의시험단계를 다음과 같이 3단계로 나누어 시행한다 (Bauer, 2010).

1. $N(\hat{\mu}_i, \hat{\Omega})$ 로부터 $\phi_i^{(k1)}$ 의 표본을 추출한다.
2. $N(\phi_i^{(k1)}, Z)$ 로부터 $\phi_i^{(k2)}$ 의 표본을 추출한다.
여기서 $Z = \kappa \hat{\Omega}$ 를 나타내며 κ 는 보정계수(scaling factor)이다.
3. $N(\phi_{il}^{(k2)}, z_{il})$ 로부터 $\phi_{il}^{(k)}$ 의 표본을 추출한다.
여기서 $\phi_{il}^{(k2)}$ 는 $\phi_i^{(k2)}$ 의 l 번째 원소를 나타내며, z_{il} 은 Z 의 l 번째 대각원소를 나타낸다 ($l = 1, \dots, p$).

또한 확률적 근사 단계는 다음과 같이 수행한다.

- $S_{m,k+1} = \frac{k-1}{k} S_{m,k} + \frac{1}{k} \tilde{S}_m(y, \phi^{(k)})$

4. 베이지안 방법

식 (2.1)의 모형을 베이지안 방법으로 추정하기 위하여 추정모수의 사전분포(prior distribution)을 다음과 같이 가정한다.

$$\theta \sim N(\mu_\theta, V_\theta) \quad (4.1)$$

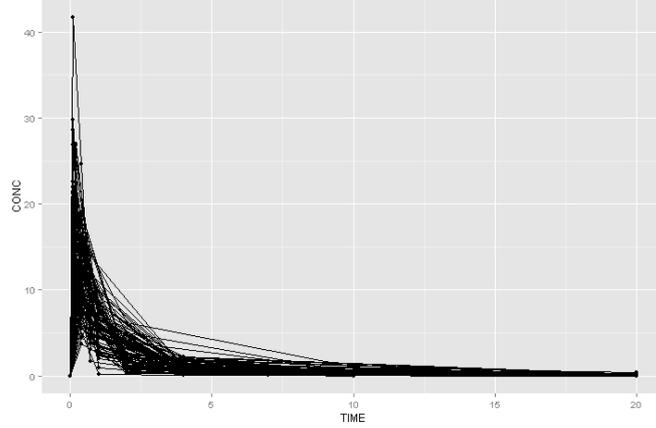


Figure 5.1. Time vs. drug concentration profile plot of 1000 patients with IV bolus

$$\begin{aligned}\Omega^{-1} &\sim \text{Wishart}(df_{\Omega}, (df_{\Omega}\Omega_0)^{-1}) \\ \sigma^{-2} &\sim \text{Gamma}\left(\frac{df_{\sigma}}{2}, \frac{df_{\sigma}}{2\sigma_0^2}\right),\end{aligned}$$

이 때의 각 모수들의 조건부 확률분포는 다음과 같다.

$$\begin{aligned}\theta|\cdot &\sim N(V(N\Omega^{-1}\bar{\phi} + V_{\theta}^{-1}\mu_{\theta}), V) \\ \Omega^{-1}|\cdot &\sim \text{Wishart}\left(N + df_{\Omega}, \left(\sum_{i=1}^N (\phi_i - \mu(\theta, x_i))^T (\phi_i - \mu(\theta, x_i))\right)^{-1}\right) \\ \sigma^{-2}|\cdot &\sim \text{Gamma}\left(\frac{\sum_{i=1}^N n_i + df_{\sigma}}{2}, \frac{df_{\sigma}}{2\sigma_0^2} + \sum_i \sum_j \left(\frac{y_{ij} - g(\text{dose}_i, \phi_i, t_{ij})}{h(\text{dose}_i, \phi_i, t_{ij})}\right)^2\right),\end{aligned}\tag{4.2}$$

여기서 $V = (N\Omega^{-1} + V_{\theta}^{-1})^{-1}$, $\bar{\phi} = (1/N)\sum \phi_i$ 를 나타낸다. NONMEM에서는 베이지안 방법으로서의 추정을 위해서 깃스 샘플링(Gibbs sampling; Gilks 등, 1996) 방법을 이용하며 식 (4.2)의 조건부 분포를 이용하여 반복적으로 표본을 추출하여 모수를 추정하게 된다.

5. 예제

2장에서 살펴본 여러가지 집단 약동학 모형의 추정방법들의 성능을 비교해보기 위하여 NONMEM 7.2에서 제공하고 있는 example1의 자료를 이용하였다. 이는 1000명의 피험자에게 정맥주사로 약물을 1회 투여한 후 시간에 따라 혈중농도를 6번씩 측정한 자료로 측정 시간은 0.1, 0.2, 0.4, 0.7, 1, 2, 4, 7, 10, 20, 40, 70시간 중 각 피험자마다 랜덤으로 6번을 선택하여 측정하였다. Figure 5.1은 측정 혈중농도의 프로파일 그림으로 처음 20시간까지의 변화를 좀 더 정확히 보기 위하여 20시간까지 만을 그린 그림이다.

추정 모형으로는 이구획 단회정주모형(two compartment IV bolus model)을 이용하였으며 모형의 식

Table 5.1. The result of two compartment IV bolus model with various estimation methods

| method | OBJ | θ_1 | θ_2 | θ_3 | θ_4 | run time(sec) |
|-------------|------------------|------------|------------|------------|------------|---------------|
| FO | -896.149 | 1.45 | 1.77 | 0.021 | 2.43 | 61.20 |
| FOCE | -1142.433 | 1.69 | 1.61 | 0.820 | 2.39 | 8.76 |
| Laplacian | -1123.627 | 1.63 | 1.52 | 0.849 | 2.42 | 143.55 |
| ITS | -1143.036 | 1.68 | 1.59 | 0.813 | 2.37 | 8.97 |
| IMP | -1144.358 | 1.64 | 1.56 | 0.750 | 2.35 | 227.54 |
| IMPMPAP | -1142.868 | 1.64 | 1.56 | 0.747 | 2.35 | 226.05 |
| SAEM | -1155.519 | 1.64 | 1.55 | 0.752 | 2.35 | 144.10 |
| BAYES | -1154.222 | 1.63 | 1.55 | 0.742 | 2.34 | 386.47 |

Table 5.2. The result of two compartment IV bolus model with 100 simulated data

| Method | OBJ | RMSE(%) | | | | | | | | | run time (초) |
|-------------|-----------------|-------------|-------------|-------------|-------------|---------------|---------------|---------------|---------------|---------------|-----------------|
| | | θ_1 | θ_2 | θ_3 | θ_4 | ω_{11} | ω_{22} | ω_{33} | ω_{44} | σ^2 | |
| FO | -251.84 | 5.25 | 13.89 | 24.08 | 21.68 | 58.72 | 31.66 | 38.29 | 192.39 | 2560.07 | 11.37 |
| FOCE | -1053.71 | 3.83 | 4.46 | 12.28 | 2.91 | 14.48 | 20.55 | 27.27 | 21.08 | 862.81 | 10.22 |
| LAP | -728.03 | 13.61 | 18.58 | 92.17 | 12.76 | 47.02 | 61.53 | 98.71 | 68.02 | 5217.55 | 50.14 |
| ITS | -1045.05 | 3.47 | 3.54 | 11.01 | 2.50 | 15.61 | 25.36 | 35.15 | 26.67 | 878.45 | 9.41 |
| IMP | -832.46 | 9.73 | 12.42 | 32.96 | 9.73 | 45.06 | 28.23 | 43.90 | 122.69 | 8485.09 | 38.36 |
| IMPMPAP | -1064.13 | 2.64 | 3.15 | 9.49 | 2.44 | 15.54 | 22.13 | 29.13 | 26.49 | 872.45 | 43.30 |
| SAEM | -1076.39 | 2.63 | 3.09 | 9.61 | 2.45 | 15.11 | 19.11 | 26.07 | 22.82 | 878.89 | 114.02 |
| BAYES | -1076.30 | 2.68 | 3.17 | 9.30 | 2.43 | 15.98 | 20.96 | 24.63 | 21.64 | 920.12 | 165.04 |
| ITS-IMP | -1068.38 | 2.65 | 3.18 | 9.49 | 2.43 | 15.66 | 22.22 | 30.78 | 25.87 | 896.17 | 18.34 |
| ITS-IMPMPAP | -1066.33 | 2.65 | 3.20 | 9.42 | 2.41 | 15.70 | 22.25 | 30.26 | 26.48 | 882.60 | 19.55 |
| ITS-SAEM | -1067.60 | 2.70 | 3.25 | 10.18 | 2.53 | 16.12 | 25.00 | 42.18 | 33.21 | 976.38 | 73.61 |
| ITS-BAYES | -1076.29 | 2.68 | 3.09 | 9.30 | 2.43 | 15.89 | 20.93 | 24.24 | 22.11 | 920.47 | 174.16 |

은 다음과 같다.

$$g(\text{dose}, \phi, t) = \frac{\text{dose} \cdot (\alpha - k_{21})}{V_1 \cdot (\alpha - \beta)} e^{-\alpha t} + \frac{\text{dose} \cdot (k_{21} - \beta)}{V_1 \cdot (\alpha - \beta)} e^{-\beta t}, \tag{5.1}$$

$$\alpha + \beta = k_e + k_{12} + k_{21},$$

$$\alpha \cdot \beta = k_e \cdot k_{21}.$$

Table 5.1은 Figure 5.1의 자료를 이구획 단회정주모형(two compartment IV bolus model)에 적합시킨 결과이다. 이는 NONMEM에서 자동으로 계산해주는 초기치를 이용하여 추정을 한 결과를 정리한 것이다. 여기서 $\theta_1 = k_e V_1$ 은 청소율(clearance), $\theta_2 = V_1$ 는 중심구획(central compartment)의 용적(volume of distribution), $\theta_3 = V_1 k_{12}/k_{21}$ 는 주변구획(peripheral compartment)의 용적(volume of distribution), 그리고 $\theta_4 = k_{21} V_1$ 는 구획 간 청소율(inter-compartmental clearance)를 나타낸다. OBJ는 $-2\log\text{likelihood}$ 값으로 여러 추정방법 중 SAEM 방법의 OBJ가 최소로 모형을 가장 잘 추정하고 있음을 알 수 있다. 또한 BAYES 방법도 SAEM 방법과 비슷한 성능을 보이고 있으며 IMP와 IMPMPAP 방법도 크게 뒤떨어지지 않는 성능을 보이고 있다. FOCE와 ITS는 확률표본을 이용한 방법들에 비하여 조금 떨어지는 성능을 보이고 있으나 시간을 고려하면 크게 나쁘지 않음을 알 수 있다. 확률표본을 이용한 추정방법들 중 SAEM의 경우 ITS나 FOCE에 비하여 18배 정도의 시간이, BAYES 방법의 경우에는 40배 이상의 시간이 소요되고 있다.

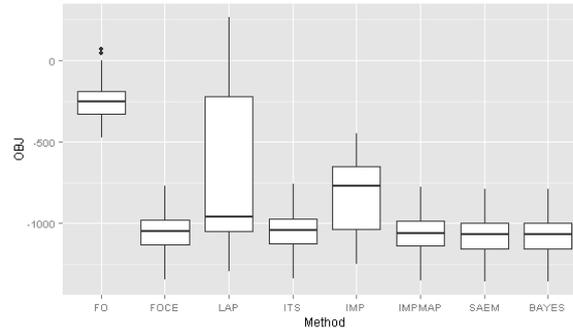


Figure 5.2. Comparison of the objective function values with various estimation methods

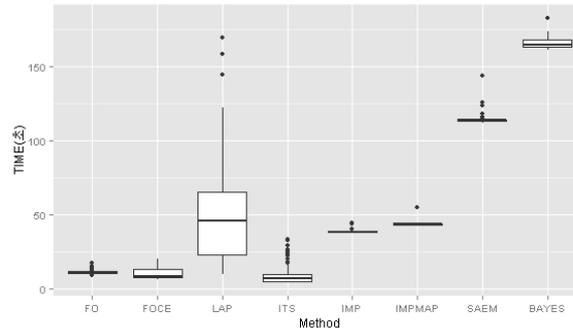


Figure 5.3. Comparison of the elapsed time to estimate parameters with various estimation methods

추정방법들을 좀 더 심도있게 비교하기 위하여 위의 자료와 이구획 단회 정주모형을 이용하여 모의실험을 수행하였다. SAEM에서의 추정치를 모의시험을 위한 모수로 사용하여 100개의 모의자료를 생성한 후 다양한 추정방법을 이용하여 이구획 단회정주모형에 적합시켜 목적함수의 값과 추정치를 계산하고 각 추정에서 소요된 시간을 측정하였다. 이에 대한 결과는 Table 5.2에 정리되어 있다. 각 추정방법별 OBJ의 평균을 비교해 보면 FO 방법이 가장 좋지 않은 성능을, SAEM 방법이 가장 좋은 성능을 보이고 있다 (Figure 5.2). BAYES 방법도 SAEM과 비슷한 성능을 보이고 있으며 IMPMAP도 크게 뒤떨어지지 않는 성능을 보이고 있다. 그러나 IMP 방법은 OBJ값 뿐 아니라 θ_3 와 Ω , 그리고 σ^2 에서 상당히 큰 RMSE값을 나타내고 있어 안정적으로 추정하지 못하고 있음을 알 수 있다. IMPMAP, SAEM, BAYES 방법은 대부분의 추정치에서 비슷한 RMSE값을 보이고 있으며 ω_{44} 와 σ^2 에서는 FOCE가 가장 좋은 성능을 보이고 있다.

Figure 5.3은 추정방법별 추정시간을 상자그림을 나타낸 것으로 FO, FOCE, 그리고 ITS가 가장 단 시간에 추정치를 계산함을 알 수 있다. LAP는 자료에 따라 매우 다른 추정시간을 보이고 있으며 이를 통하여 LAP 방법의 수렴이 자료에 따라 크게 다른 양상을 보이고 있다는 것을 알 수 있다. IMP와 IMPMAP는 비슷한 시간이 소요되며 SAEM은 2분에 가까운 시간이 소요되고 있다. 또한 이들 방법은 몇 개의 데이터를 제외하고는 매우 일정한 시간이 소요되고 있음을 알 수 있다. BAYES 방법은 거의 3분에 가까운 시간이 소요되고 있다.

Figure 5.4는 추정방법별 θ 추정치의 편의(bias)를 상자그림으로 나타낸 것으로 모든 θ 에서 IMPMAP,

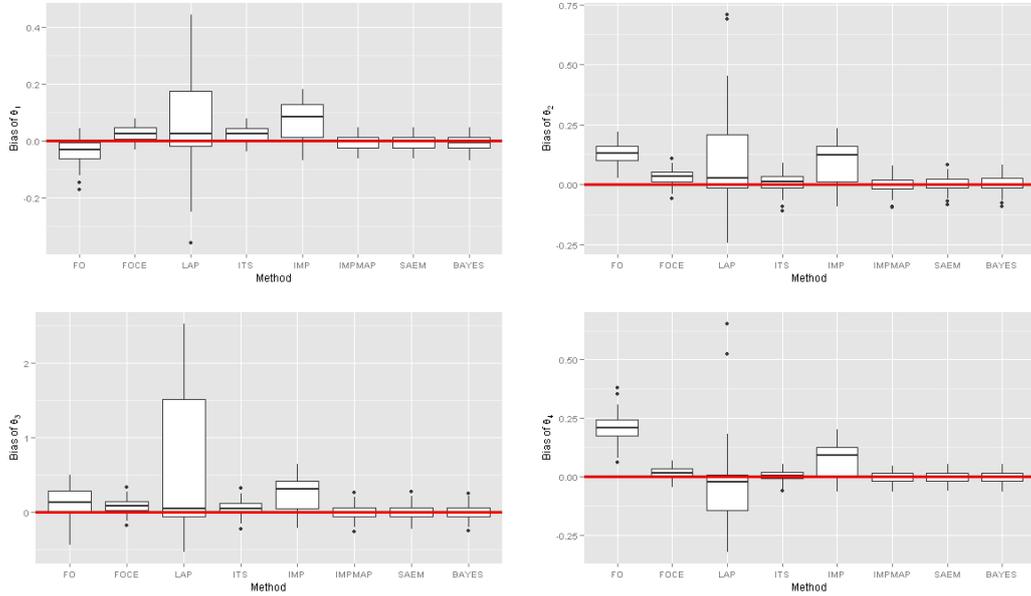


Figure 5.4. Comparison of the biases of θ with various estimation methods

SAEM, 그리고 BAYES 방법은 비슷한 성능을 보이고 있으며 FOCE와 ITS가 그 다음으로 좋은 성능을 보이고 있다. FOCE의 경우 θ 를 과대추정하는 경향이 있다. FO와 LAP 방법, 그리고 IMP 방법의 성능은 좋지 않음도 확인할 수 있다. Figure 5.5은 Ω 에 대한 추정치의 편의(bias)를 나타내는 그림으로 FOCE, IMPMAP, SAEM, 그리고 BAYES 방법들은 모의자료에 대한 추정값들의 변동이 크지 않으므로 매우 안정적으로 추정을 하고 있음을 알 수 있다.

모의시험을 통하여 살펴본 결과 SAEM 방법과 BAYES 방법이 가장 안정적으로 추정을 하고 있으며 이들 중 SAEM이 좀 더 빠른 시간에 추정을 함을 알 수 있다. 또한 추정시간을 고려하면 FOCE도 나쁘지 않은 성능을 보이고 있으나 좀 더 정확한 추정을 위해서는 SAEM을 사용하는 것이 바람직하다.

ITS 방법은 가장 빠른 시간에 추정치를 찾아주고 있으며 FO나 IMP보다는 훨씬 좋은 성능을 보이고 있다. 이를 초기치 설정에 이용하여 확률적 표본을 이용하는 방법의 성능이 개선되는지를 알아보기 위하여 IMP, IMPMAP, SAEM, 그리고 BAYES 방법에 적용해 보았다. 이들의 결과는 Table 5.2의 아랫단 부분에 나타나있다. IMP와 IMPMAP의 경우 추정에 소요되는 시간은 반으로 줄어들고 성능도 개선된 것을 볼 수 있다. 특히 IMP 방법의 경우 ITS 초기치를 이용하는 경우에는 안정적으로 추정을 하고 있음을 알 수 있다. SAEM의 경우도 시간이 많이 단축되는 것을 볼 수 있으나 성능은 SAEM 만을 이용한 경우보다 좋지 않은 것을 알 수 있다. BAYES의 경우 ITS를 사용한 경우와 그렇지 않은 경우에 시간과 성능면에서 크게 차이를 보이지 않고 있음을 알 수 있다. 이를 통하여 IMP, IMPMAP, 그리고 SAEM을 이용할 경우 ITS의 초기치 사용을 고려해 볼 수 있으며 특히 IMP를 이용할 경우에는 좀 더 안정적인 추정치를 얻기 위해서는 ITS 초기치를 사용하는 것이 좋을 것을 파악할 수 있다.

6. 결과 및 논의

집단 약동/약력학 모형은 비선형 혼합효과모형으로 집단 특성을 나타내는 고정효과의 모수 추정 뿐 아

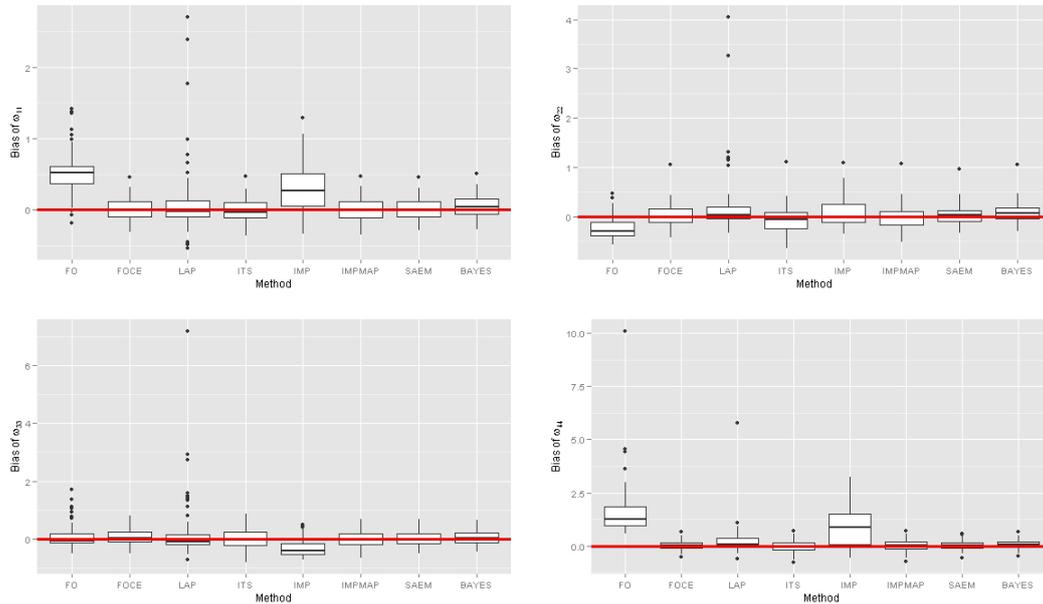


Figure 5.5. Comparison of the biases of Ω with various estimation methods

나라 집단 내의 개인 간 특성을 나타내는 랜덤 효과의 분산 추정이 중요한 역할을 하고 있다. 그러나 추정모수의 비선형성으로 인하여 추정에 어려움이 있다. 이를 위하여 고전적인 방법인 선형근사부터 최근 각광받고 있는 확률적 표본을 이용한 방법까지 다양한 방법들이 제안되어 있다. 본 논문에서는 집단 약동/약력학 모형 추정에 가장 널리 이용되고 있는 소프트웨어인 NONMEM에서 제공하고 있는 다양한 추정방법들을 이론적으로 살펴보고 이들의 성능을 비교, 분석해 보았다. 오랫동안 쓰여 온 선형근사를 이용한 FOCE 방법은 안정적으로 추정할 뿐만 아니라 추정소요시간도 다른 방법에 비하여 빠름을 알 수 있다. 그러나 선형근사로 인한 오차는 비선형성이 커짐에 따라 증가하게 된다. 확률적 표본을 이용한 방법으로는 SAEM이 가장 안정적으로 추정을 해주며 FOCE보다 훨씬 좋은 성능을 보이고 있다. IMPMAP 방법도 비슷한 성능을 보이고 있다. 그러나 이들 방법은 추정소요시간이 길다는 단점이 있으며 이를 보완하기 위한 방법으로는 좀 더 좋은 초기치를 이용하여 좀 더 빨리 수렴하도록 하는 방법이 있다. 정확한 추정치를 찾아주지는 못하지만 상대적으로 빠른 시간 내에 근사한 추정치를 찾아주는 ITS 방법을 초기치를 찾는 방법으로 이용한 후 확률적 표본을 이용한 방법들을 적용하면 좀 더 빠른 시간에 좋은 성능으로 추정치를 구할 수 있다.

References

- Bauer, R. J. (2010). *NONMEM 7 Technical Guide, Icon Development Solution*, Ellicott City, Maryland.
- Beal, S. L., Sheiner, L. B., Boeckmann, A. and Bauer, R. J. (1992). *NONMEM Users Guides, NONMEM Project Group*, University of California, San Francisco.
- Celeux, G. and Diebolt, J. (1992). A stochastic approximation type EM algorithm for the mixture problem, *Stochastics: An International Journal of Probability and Stochastic Processes*, **41**, 119–134.
- Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm, *Annals of Statistics*, 94–128.

- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman Hall/CRC, New York.
- Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models, *Computational Statistics and Data Analysis*, **49**, 1020–1038.
- Lee, E.-K. (2010). A statistical approach to the pharmacokinetic model, *The Korean Journal of Applied Statistics*, **23**, 511–520.
- Mentré, F. (2008). Stochastic EM algorithms in population pharmacokinetic-pharmacodynamic analyses, In *The American Conference on Pharmacometrics (ACoP)*, 9–12.
- Robert, C. and Casella, G. (2013). *Monte Carlo Statistical Methods*, Springer Science and Business Media.
- Wang, Y. (2007). Derivation of various NONMEM estimation methods, *Journal of Pharmacokinetics and Pharmacodynamics*, **34**, 575–593.

확률적 표본추출 방법을 이용한 집단 약동학 모형의 추정과 검증에 관한 고찰

김광희^a · 윤정화^a · 이은경^{a,1}

^a이화여자대학교 통계학과

(2015년 3월 5일 접수, 2015년 3월 30일 수정, 2015년 3월 30일 채택)

요약

본 논문에서는 집단 약동/약력학 모형 추정을 위한 다양한 추정방법들을 이론적으로 비교, 분석하였다. 특히 확률적 표본을 이용한 방법들인 IMP, IMPMAP, SAEM 방법과 베이지안 방법의 이론적 배경과 이들의 성능을 자세히 살펴보고, 기존의 선형근사를 이용한 FO, FOCE 등의 방법과 비교 분석하였다. 확률적 표본을 이용한 추정방법들이 추정에 많은 시간이 소요된다는 문제점을 개선하기 위하여 좀 더 좋은 초기치를 찾는 방안으로 상대적으로 짧은 시간에 정확한 추정치를 계산해주는 ITS 방법을 이용하였다.

주요용어: 비선형 혼합효과모형, 확률적 표본, 최대우도추정, EM 알고리즘.

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2010-0003840).

¹교신저자: (120-750) 서울특별시 서대문구 이화여대길 52, 이화여자대학교 통계학과.

E-mail: lee.eunk@ewha.ac.kr