

한성필, 조용순, 김형섭, 배균섭

***nmw: R* 패키지와 *Edison* 사이언스 앱을 이용한 비선형 혼합효과 모델링과 시뮬레이션**



Contents

| | |
|--|------------|
| List of Tables | v |
| List of Figures | vii |
| 책머리에 | ix |
| 1 서론 | 1 |
| 2 이론 및 계산방법 | 3 |
| 2.1 Maximum Likelihood Estimation | 3 |
| 2.2 First-order conditional estimation (FOCE) method . | 4 |
| 2.3 설치 | 4 |
| 2.4 자료 | 5 |
| 3 결과 | 9 |
| 4 결론 | 13 |
| 5 감사의 글 | 15 |
| Appendix | 17 |
| 세션정보 | 17 |
| 참고문헌 | 19 |



List of Tables

| | |
|-------------------------------------|----|
| 3.1 nmw R 패키지에서 정의된 함수 설명 | 10 |
|-------------------------------------|----|



List of Figures

| | | |
|-----|--|----|
| 2.1 | R에 내장된 Theoph 자료. 이 자료를 사용하여 분석을 시행하였습니다. | 6 |
| 3.1 | nmw R 패키지를 활용한 비선형 혼합효과 모델 사용 흐름도 | 10 |
| 3.2 | PRED 함수를 정의하여 numDeriv 패키지의 deriv() 함수를 사용하여 정의하고 InitStep() 함수의 인자로 입력함 | 12 |
| 4.1 | 에디슨 앱의 입력화면 | 13 |



책머리에

본 연구의 목적은 비선형 혼합효과 모델링 과정을 이해하고자 개발된 R 패키지의 사용법을 소개하고 이를 활용한 Edison 사이언스앱의 활용법을 제시하는 것입니다. `nmw R Package`의 기본적인 원리를 설명하고, R 패키지의 함수 사용법을 보이며, 에디슨 앱에서 구현된 활용 예시를 논하고자 합니다. 이를 바탕으로 비선형 혼합효과 모델링의 원리를 익힐 수 있고, 더 나아가 약물의 용법/용량, 모델 파라미터의 변화에 따른 혈중농도 등의 약동학 변화에 대한 다양한 시뮬레이션으로 응용될 수 있습니다.

한성필¹, 조용순², 김형섭², 배균섭²

¹ 가톨릭대학교 약리학교실 ² 서울아산병원 임상약리학과, 울산대학교 병원, 서울특별시 송파구 올림픽로 43길 88

E-mail: 한성필 shan@catholic.ac.kr¹, 배균섭 ksbae@acp.kr²

¹<mailto:shan@catholic.ac.kr>

²<mailto:ksbae@acp.kr>



1

서론

비선형 혼합효과 모델링(Nonlinear mixed effect modeling)을 위해 가장 흔히 쓰이는 도구는 NONMEM®이라고 하는 소프트웨어입니다. 현재 모델기반 약물개발(Model-based drug development)를 위한 계량약리학 분야에서 가장 표준적인 도구로 생각되고 있으나 그 동작하는 원리를 이해하기는 쉽지 않습니다. 공개소프트웨어인 R을 통해서 NONMEM의 계산과 알고리즘을 구현하는 시도가 있었고, 성공적으로 결과값을 재현할 수 있음을 보였습니다. (Kim et al., 2015; Bae and Yim, 2016) 이 연구를 통해서 NONMEM의 동작원리를 단계별로 실행하며 학습할 수 있게 되었으며, 더 나아가 저자들은 공개 소프트웨어인 nmw R 패키지를 개발 (Bae, 2018), 배포하여 이러한 원리를 누구나 비용없이 접근할 수 있게 하였습니다.



2

이론 및 계산방법

2.1 Maximum Likelihood Estimation

우도함수를 최대화하면서 모수를 추정하는 방법입니다. 관측된 표본에 기초하여 관측 불가능한 파라미터(모수)를 추정하는 방법론 중 하나로, 표본들로부터 알려지지 않은 모집단 확률분포의 형태를 추정해가는 방법론입니다. (An estimation method that finds a combination of parameters which maximizes the likelihood of an observation dataset)

MLE는 주어진 Dataset(관측결과, D)의 발생 가능성(Likelihood, 우도)이 가장 높은 모수를 추정하고, 이를 바탕으로 최적의 f를 도출합니다. 이를 수식으로 표현하면 다음과 같습니다.

정규분포의 경우 n_i 개의 관찰값을 갖는 대상 i에 대해서 $-2 \log \text{likelihood}$ ($-2LL$)은 다음과 같이 표현할 수 있습니다.

$$-2LL_i = n_i \cdot \log(2\pi) + \sum_{j=1}^{n_i} \log(\sigma_{ij}^2) + \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_{ij})^2}{\sigma_{ij}^2}$$

위 함수를 최소화 (우도를 최대화) 하는 모수의 조합은 무엇일까요? 그 결과가 maximum likelihood estimator/estimate (MLE) 이고 그 목적함수는 다음과 같습니다.

$$O_i = \sum_{j=1}^{n_i} \log(\sigma_{ij}^2) + \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_{ij})^2}{\sigma_{ij}^2}$$

MLE의 장점은 매우 직관적이라는 점입니다. 그리고 별다른 가정을 하지 않고 주어진 관측결과를 바탕으로 쉽게 계산할 수 있습니다. (물론, 관측결과가 이항분포를 따를 것이라는 가정이 들어갔지만, 다른 추정법에 비해서는 가정이 약한 편입니다.)

MLE는 빈도론을 기반으로 합니다. 관측결과에 의존해서 판단하기 때문에 관측결과의 크기(관측횟수)와 질(관측결과의 정확성)에 따라서 모수 추정의 성패가 크게 좌우됩니다.

2.2 First-order conditional estimation (FOCE) method

FOCE는 FO 가정법보다 더 복잡합니다. EBE를 각 iteration 마다 추정하기 때문입니다.

The FOCE is more complex than the first-order (FO) approximation method because it estimates the empirical Bayes estimate (EBE) for each iteration. By contrast, it is a further approximation of the Laplacian (LAPL) method, which uses second-order expansion terms. FOCE without INTERACTION can only be used for an additive error model, while FOCE with INTERACTION (FOCEI) can be used for any error model. The formula for FOCE without INTERACTION can be derived directly from the extension of the FO method, while the FOCE with INTERACTION method is a slight simplification of the LAPL method. Detailed formulas and R scripts are presented here for the reproduction of objective function values by NONMEM.

2.3 설치

먼저 R 패키지의 설치 및 로딩 과정은 다음의 명령어를 R 콘솔에 입력함으로써 가능합니다. NONMEM Workshop 2017에서 사용된 nmw 패키지입니다. (Kim et al., 2015; Bae and Yim, 2016; Bae, 2018)

```
install.packages('nmw')
library(nmw)
```

NONMEM을 통해 얻은 THEO-FO.OUT와 비교할 때 값이 같습니다.

Scaling factor는 소스 코드를 보고 알아낸 것입니다.

2.4 자료

```
DataAll <- Theoph
colnames(DataAll) <- c("ID", "BWT", "DOSE", "TIME", "DV")
DataAll[, "ID"] <- as.numeric(as.character(DataAll[, "ID"]))
```

이후 ?nmw 명령어를 통해 내장되어 있는 예제를 볼 수 있고 이를 실행해 볼 수 있습니다. 사용되는 자료는 R에 내장되어 있는 Theoph이며 이 자료는 12 명의 환자에서 테오필린 320 mg을 경구 투여한 후 24시간 동안 혈장에서 측정한 농도를 포함하고 있습니다. 그림을 그려보면 다음과 같습니다. (Figure 2.1)

```
lattice::xyplot(DV ~ TIME | as.factor(ID), data=DataAll, type="b")
```

```
nTheta = 3
nEta = 3
nEps = 2

THETAinit = c(2, 50, 0.1)
OMinit = matrix(c(0.2, 0.1, 0.1, 0.1, 0.2, 0.1, 0.1, 0.1, 0.2),
                 nrow=nEta, ncol=nEta)
SGinit = diag(c(0.1, 0.1))

LB = rep(0, nTheta) # Lower bound
UB = rep(1000000, nTheta) # Upper bound
```

위 코드는 같이 θ (nTheta), η (nEta), \times (nEps) 개수를 먼저 지정하고 초기값 (THETAinit, OMinit, SGinit)을 설정하는 과정이며 LB, UB는 θ 값 추정의 상한값과 하한값을 정해주는 것으로 적합 (fit)을 성공적으로 수행하기 위해 적절한 범위를 지정해야 합니다. 초기값, 하한값, 상한값은 InitStep() 함수의 인자로 입력되어 initialization step에 사용됩니

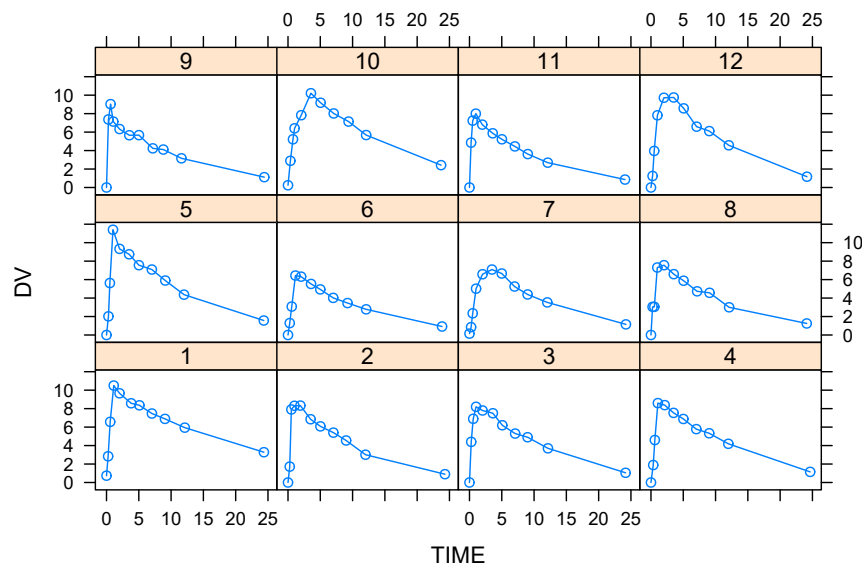


FIGURE 2.1: R에 내장된 Theoph 자료. 이 자료를 사용하여 분석을 시행 하였습니다.

다. (Figure 2) 이때 PRED 함수 (Figure 3)도 InitStep()의 입력 인자로 사용되는데, 이는 약동학 모델을 미분방정식 형태로 구현한 함수입니다. {Kang et al. (2012)}

InitStep()의 추정방식으로 세가지가 지원되며 ZERO (First Order Approximation Method), COND (First Order Conditional Estimation with Interaction Method), LAPL(Laplacian Approximation with Interaction Method)에 따라 PRED 함수의 추정 방법이 달라지게 됩니다. [1, 2] 이후 EstStep(), CovStep(), PostHocEta(), TabStep() 함수를 실행하며 계산을 수행해 나가고 각 함수에 대해 Table 1에서 설명하였습니다.

```
##### First Order Approximation Method
InitStep(DataAll, THETAinit=THETAinit, OMin=OMin, SGinit=SGinit,
         LB=LB, UB=UB, Pred=PRED, METHOD="ZERO")
(EstRes = EstStep())           # 4 sec
(CovRes = CovStep())           # 2 sec
```



```
PostHocEta() # Using e$FinalPara from EstStep()  
TabStep()
```



3

결과

InitStep(), EstStep() 후 계산된 값은 다음과 같습니다. OFV (Objective Function Value), 최종 모수 (final parameter)를 계산한 것이 출력됩니다.

```
$`Initial OFV`  
[1] 141.3076  
  
$Time  
Time difference of 3.356192 secs  
  
$Optim  
$Optim$par  
[1] 0.560417594 -0.167835388 0.148962362 0.995143048 0.056166719 0.151227211 -1.03246852  
[10] -0.956899772 -0.205559310  
  
$Optim$value  
[1] 57.32106  
  
$Optim$counts  
function gradient  
74 74  
  
$Optim$convergence  
[1] 0  
  
$Optim$message  
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"  
  
$`Final Estimates`
```

```
[1] 3.16946754 38.25213460 0.10501808 1.19823325 0.13747849 0.03134899 0.37015671 0.04
[10] 0.01207782 0.05427434
```

EstStep() 후 CovStep() 함수를 실행하면, 표준 오차, 분산 행렬, correlation 행렬, 역분산 행렬, 고유값, R, S 행렬 등을 출력하게 됩니다. PostHocEta()는 각 대상자 별로 ETA에 대한 EBE 값을 보여주며, TabStep()은 각 농도 포인트 별로 예측값과 관측값 및 잔차 등을 표 형식으로 보여 주게 됩니다. (Table 3.1)

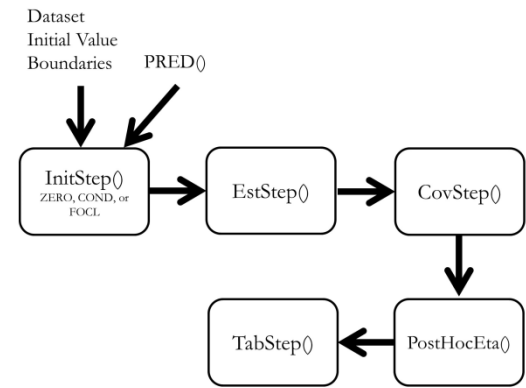


FIGURE 3.1: nmw R 패키지를 활용한 비선형 혼합효과 모델 사용 흐름도

TABLE 3.1: nmw R 패키지에서 정의된 함수 설명

| 함수명 | 설명 | 기능 |
|------------|---------------------|---|
| InitStep() | Initialization Step | 초기값 지정 및 추정 방법 등을 설정 합니다. |
| EstStep() | Estimation Step | 초기화 후 objective function value, 최종 모수 등을 계산합니다. |
| CovStep() | Covariance Step | EstStep() 후 얻어진 최종모수를 통해 표준 오차, 분산 행렬, correlation 행렬, 역분산 행렬, 고유값, R, S 행렬 등을 계산합니다. |

| | | |
|--------------|-----------------------------------|--|
| PostHocEta() | Empirical bayesian estimate (EBE) | 각 대상자 별로 ETA에 대한 EBE 값을 보여줍니다. |
| TabStep() | Table Step | 각 농도 포인트 별로 예측값과 관측값 및 잔차 등을 표 형식으로 보여줍니다. |

```

FGD = deriv(~DOSE/(TH2*exp(ETA2))*TH1*exp(ETA1)/(TH1*exp(ETA1)
            (exp(-TH3*exp(ETA3)*TIME)-exp(-TH1*exp(ETA1)*TIME)
            c("ETA1", "ETA2", "ETA3"),
            function.arg=c("TH1", "TH2", "TH3", "ETA1", "ETA2"),
            func=TRUE,
            hessian=(e$METHOD == "LAPL"))
H = deriv(~F + F*EPS1 + EPS2, c("EPS1", "EPS2"), function.arg=

PRED = function(THETA, ETA, DATAi)
{
  FGDres = FGD(THETA[1], THETA[2], THETA[3], ETA[1], ETA[2], ETA[3])
  Gres = attr(FGDres, "gradient")
  Hres = attr(H(FGDres, 0, 0), "gradient")

  if (e$METHOD == "LAPL") {
    Dres = attr(FGDres, "hessian")
    Res = cbind(FGDres, Gres, Hres, Dres[,1,1], Dres[,2,1], Dres[,3,1])
    colnames(Res) = c("F", "G1", "G2", "G3", "H1", "H2", "D11", "D21", "D31")
  } else {
    Res = cbind(FGDres, Gres, Hres)
    colnames(Res) = c("F", "G1", "G2", "G3", "H1", "H2")
  }
  return(Res)
}

```

FIGURE 3.2: PRED 함수를 정의하여 numDeriv 패키지의 deriv() 함수를 사용하여 정의하고 InitStep() 함수의 인자로 입력함

4

결론

본 연구에서는 NONMEM의 동작원리를 단계별로 실행하는 nmw R 패키지의 기본적인 원리와 함수 사용법이 소개되었습니다. Nmw 패키지가 사용되어 개발된 에디슨 앱을 사용할 수 있으며 다양한 값과 초기값을 바꾸면서 모델링이 어떻게 변화하는지를 관찰할 수 있습니다. (Figure 4.1) 또한 함께 개발된 에디슨 앱을 통하여 더 많은 학생들이 이러한 비선형 혼합효과 모델에 대해서 이해할 수 있을 것으로 기대됩니다. 이를 바탕으로 구현된 모델링 원리는 약물의 용법/용량, 모델 파라미터의 변화에 따른 혈중농도 등의 약동학 변화에 대한 다양한 시뮬레이션 등으로 응용될 수 있을 것입니다.

| Variable Name | Value | Description |
|---------------|-------------------------------------|-------------|
| Dataset: | Theoph | T |
| Method: | ZERO | T |
| nTheta: | 3 | T |
| nEta: | 3 | T |
| nEps: | 2 | T |
| THE TANH: | 2.50,0.1 | T |
| CMin: | 0.2,0.1,0.1,0.1,0.2,0.1,0.1,0.1,0.2 | T |
| SIGM: | 0.1,0.0,0.1 | T |

FIGURE 4.1: 에디슨 앱의 입력화면



5

감사의 글

2017년도 정부(미래창조과학부)의 재원으로 한국연구재단 첨단 사이언스·교육 허브 개발 사업의 지원을 받아 수행된 연구입니다. (NRF-2011-0020576)



세션정보

`sessionInfo()`

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 18363)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Korean_Korea.949 LC_CTYPE=Korean_Korea.949
## [3] LC_MONETARY=Korean_Korea.949 LC_NUMERIC=C
## [5] LC_TIME=Korean_Korea.949
##
## attached base packages:
## [1] stats    graphics grDevices utils    datasets methods  base
##
## other attached packages:
## [1] kableExtra_1.1.0    forcats_0.5.0      stringr_1.4.0
## [4] dplyr_0.8.5         purrr_0.3.4        readr_1.3.1
## [7] tidyr_1.0.3         tibble_3.0.1       ggplot2_3.3.0
## [10] tidyverse_1.3.0     lattice_0.20-41    nmw_0.1.4
## [13] numDeriv_2016.8-1.1 knitr_1.28
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.0 xfun_0.13          haven_2.2.0        colorspace_1.4-
1
## [5] vctrs_0.3.0      generics_0.0.2    viridisLite_0.3.0  htmltools_0.4.0
## [9] yaml_2.2.1       rlang_0.4.6       pillar_1.4.4       withr_2.2.0
## [13] glue_1.4.0       DBI_1.1.0         dbplyr_1.4.3       modelr_0.1.7
## [17] readxl_1.3.1     lifecycle_0.2.0   munsell_0.5.0      gtable_0.3.0
## [21] cellranger_1.1.0 rvest_0.3.5       evaluate_0.14      fansi_0.4.1
```

```
## [25] highr_0.8      broom_0.5.6    Rcpp_1.0.4.6   backports_1.1.6
## [29] scales_1.1.1   webshot_0.5.2  jsonlite_1.6.1 fs_1.4.1
## [33] png_0.1-7      hms_0.5.3      digest_0.6.25  stringi_1.4.6
## [37] bookdown_0.18  grid_4.0.0     cli_2.0.2      tools_4.0.0
## [41] magrittr_1.5   pacman_0.5.1   crayon_1.3.4   pkgconfig_2.0.3
## [45] ellipsis_0.3.0 xml2_1.3.2     reprex_0.3.0   lubridate_1.7.8
## [49] assertthat_0.2.1 rmarkdown_2.1  httr_1.4.1     rstudioapi_0.11
## [53] R6_2.4.1       nlme_3.1-147   compiler_4.0.0
```

참고문헌



Bibliography

- Bae, K.-S. (2018). *nmw: Understanding Nonlinear Mixed Effects Modeling for Population Pharmacokinetics*. R package version 0.1.4.
- Bae, K.-S. and Yim, D.-S. (2016). R-based reproduction of the estimation process hidden behind nonmem[®] part 2: First-order conditional estimation. *Translational and Clinical Pharmacology*, 24(4):161–168.
- Kang, D., Bae, K.-S., Houk, B. E., Savic, R. M., and Karlsson, M. O. (2012). Standard error of empirical bayes estimate in nonmem[®] vi. *The Korean Journal of Physiology & Pharmacology*, 16(2):97–106.
- Kim, M.-G., Yim, D.-S., and Bae, K.-S. (2015). R-based reproduction of the estimation process hidden behind nonmem[®] part 1: first-order approximation method. *Translational and Clinical Pharmacology*, 23(1):1–7.

