TCP Transl Clin Pharmacol
2015;23(1):01-07
http://dx.doi.org/10.12793/tcp.2015.23.1.1

TUTORIAL

R-based reproduction of the estimation process hidden behind NONMEM® Part 1: first-order approximation method

Min-Gul Kim[1], Dong-Seok Yim[2] and Kyun-Seop Bae[3,4]*

[1]Clinical Pharmacology Unit, Biomedical Research Institute, Chonbuk National University Hospital, Jeonju 561-712, Republic of Korea, [2]PIPET (Pharmacometrics Institute for Practical Education and Training), The Catholic University of Korea, Seoul 137-701, Republic of Korea, [3]Department of Clinical Pharmacology and Therapeutics, Asan Medical Center, Seoul 138-736, Republic of Korea, [4]University of Ulsan College of Medicine, Seoul 138-736, Republic of Korea

*Correspondence: K. S. Bae; Tel: +82-2-3010-4611, Fax: +82-2-3010-4623, E-mail: ksbae@amc.seoul.kr

Received 28 May 2015
Revised 12 Jun 2015
Accepted 15 Jun 2015

Keywords
NONMEM,
modelling,
First order approximation,
R

pISSN: 2289-0882
eISSN: 2383-5427

NONMEM® is the most-widely used nonlinear mixed effects modelling tool introduced into population PK/PD analysis. Even though thousands of pharmaceutical scientists utilize NONMEM® routinely for their data analysis, the various estimation methods implemented in NONMEM® remain a mystery for most users due to the complex statistical and mathematical derivations underlying the algorithm used in NONMEM®. In this tutorial, we demonstrated how to directly obtain the objective function value and post hoc η for the first order approximation method by the use of R. We hope that this tutorial helps pharmacometricians understand the underlying estimation process of nonlinear mixed effects modelling.

Introduction

The nonlinear mixed effects method that was introduced to the biomedical research field approximately 30 years ago is now an essential platform in population pharmacokinetic and pharmacodynamic (PK/PD) data analysis.[1-3]

NONMEM® is the most-widely used nonlinear mixed effects modelling tool introduced for population PK/PD analysis since its prototype was developed by Lewis B. Sheiner and Stuart L. Beal of the University of California, San Francisco.[4] It is programmed in FORTRAN code and has evolved to its current version, 7.3. NONMEM® is regarded as the standard pharmacometrics tool. Even though thousands of pharmaceutical scientists utilize NONMEM® routinely for their data analysis, the various estimation methods implemented in NONMEM® remain a mystery for most users due to the complex statistical and mathematical derivations underlying the algorithm used in NONMEM®. Wang systematically derived the objective functions for the major estimation methods used in NONMEM® and clearly demonstrated their relationships through the derivation.[5] Wang derived the first-order (FO) method from an approximation of the Laplacian method, whereas we derived the FO method directly from the maximum likelihood estimation method.

R is a computer language and suite of libraries for statistical and mathematical computation. It has a strong but relatively small base system compared with other commercial software like SAS® and SPSS®. However, R has robust functions for scientific computation and many add-in packages for particular problems and situations. Many of R's standard functions are coded in the R language itself, making it easy to follow their algorithmic flow.

The objective of this tutorial is to illustrate how to directly obtain the objective function value and the post hoc η using R. This tutorial gives a detailed derivation of the FO approximation method in NONMEM®. To be consistent with the terminology, the notations are defined in Table 1.

Basic principles of the first-order approximation method used in NONMEM®

The following equations are used for PK/PD modelling with NONMEM®:

$$F = f(\theta, \eta, x)$$
$$Y = f(F, \varepsilon)$$
$$\eta \sim MVN(0, \Omega)$$
$$\varepsilon \sim MVN(0, \Sigma)$$

Table 1. Definitions

Notation	Description
$F = f(\theta, \eta, \chi)$	The model predicted value (F), model parameter (θ), inter-individual random variability parameter (η), and covariates (χ) defining the structural model to explain PK/PD.
$\vec{F}_i = \vec{f}(\vec{\theta}, \vec{\eta}_{t,i}, \vec{0})$	Prediction vector (n_i x 1)
$G_i = \frac{\partial \vec{F}}{\partial \vec{\eta}}$	Partial derivatives of $\vec{F}_i$ with respect to etas (n_i x n_η)
$H_i = \frac{\partial \vec{Y}}{\partial \vec{\varepsilon}}$	Partial derivatives of $\vec{Y}_i$ with respect to epsilons (n_i x n_ε)
n_i	Number of observations for i-th subject
n_η	Number of etas
n_ε	Number of epsilons
$Y = f(F, \varepsilon)$	The observation (Y) given by F and the intra-individual variability (ε).
Y_i	The observation vector (Y) of i-th subject
Y_{ij}	The j-th observation (Y) of i-th subject
$\vec{Y}_i$	Observation vector (n_i x 1)
Σ	Sigma matrix (n_ε x n_ε)
Ω	Omega matrix (n_η x n_η)
$\varepsilon \sim MVN(0, \Sigma)$	Intra-individual (i.e., residual) random variability modeled using multivariate normal distribution with mean zero and covariance matrix Σ.
$\eta \sim MVN(0, \Omega)$	Inter-individual random variability modeled using multivariate normal (MVN) distribution with mean zero and covariance matrix Ω.

When F is a function of the predicted value without final residual variability, the observed value (Y) can be expressed as a function of F and residual variability. As shown above, $f(\theta, \eta, x)$ represents a structural model describing the relationship between the PK/PD observations and the model parameter θ, while η and ε represent the stochastic model components describing the randomness unexplained by the structural model. η represents inter-individual random variability with $E(\eta)=0$ and $V(\eta)=\omega^2$, whereas ε is the rest residual variability with $E(\varepsilon)=0$ and $V(\varepsilon)=\sigma^2$ (Therefore, θ, Ω and Σ are constants while η and ε are random variables. Before examining the details of mathematical equations, those not accustomed to mathematics and statistics may wish to view the R script first to more easily understand it.).

The typical PK/PD of the population is summarized by θ, and individual PK/PD parameters are expressed as a combination of θ and η. For example, systemic clearance is expressed as,

$$CL = \theta_{CL} \cdot \exp(\eta_{CL})$$

where θ_{CL} is the typical value and η_{CL} is the corresponding inter-subject variability following normal distribution with mean zero and covariance. In other words, the typical value of a PK/PD parameter is the value of the PK/PD parameter when $\eta = 0$. F is called the typical prediction when $\eta = 0$, while F is called the individual prediction when $\eta = \hat{\eta}$, the empirical Bayes estimate (EBE) of individual eta.[6]

Maximum likelihood estimation (MLE) is the approach used in NONMEM® for estimation of the fixed effect and random effect parameters.[7,8] MLE begins with writing a mathematical expression known as the likelihood function of the sample data. If Y is the measured observation, $E(Y)$ is the expectation of that observation by the model, and $V(Y)$ is the variance of the model, then the likelihood of the observation given the model is[9]:

$$L = \frac{1}{\sqrt{2\pi V(Y)}} e^{-\frac{1}{2V(Y)}(Y-E(Y))^2}$$

The probability (or likelihood) of a series of observations is the product of the probability (or likelihood) of individual observations. Thus, if there are n observations, the combined probability (likelihood) is the product of n individual probabilities (likelihoods), i.e., the likelihood of the i-the subject becomes:

$$L_i = \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi V(Y_{ij})}} e^{-\frac{1}{2V(Y_{ij})}(Y_{ij}-E(Y_{ij}))^2}$$

This can be transformed into the -2 log likelihood (-2LL):

$$-2\log(L_i) = n_i log(2\pi) + \sum_{j=1}^{n}\left(\log\left(V(Y_{ij})\right) + \frac{\left(Y_{ij} - E(Y_{ij})\right)^2}{V(Y_{ij})}\right)$$

NONMEM® minimizes the -2 log likelihood. Because the first part of the right side equation, $nlog(2\pi)$, is a constant (when the observation number n is fixed), minimization was directed to the second part of the right hand side of the above equation that is given by NONMEM® as the objective function value.

Therefore, the objective function for i-th subject (and if the $V(Y_{ij})$ is constant as $V(Y_i)$) can be denoted as:

$$OBJ_i = n_i \log V(Y_i) + \sum_{j=1}^{n_i}\frac{\left(Y_{ij} - E(Y_{ij})\right)^2}{V(Y_i)} ; (j = 1, \dots, n_i)$$

This function shows that the squared error, $(Y_{ij} - E(Y_{ij}))^2$, is weighted by the (inverse of) variance, $V(Y_i)$. However, Sheiner and Beal did not assume a normal distribution for this derivation. Therefore they called this method the 'extended least squares' objective function in NONMEM®[10-12] instead of the maximum likelihood estimation.

By using a first-order Taylor series approximation of f about true eta value (η_t, unknown to the investigator) of η and 0 of ε respectively, $E(Y)$ and $V(Y)$ were derived as follows:
For an individual with η_t,

$$Y = f(\theta,\eta,\varepsilon) \cong f(\theta,\eta_t,0) + \left.\frac{\partial f}{\partial \eta}\right|_{\eta=\eta_t,\varepsilon=0}(\eta-\eta_t) + \left.\frac{\partial f}{\partial \varepsilon}\right|_{\eta=\eta_t,\varepsilon=0}(\varepsilon-0)$$

$$\frac{\partial f(\theta,\eta,0)}{\partial \eta} = g(\theta,\eta,0) = \frac{\partial F}{\partial \eta}$$

$$\frac{\partial f(\theta,\eta,\varepsilon)}{\partial \varepsilon} = h(\theta,\eta,\varepsilon) = \frac{\partial F}{\partial \varepsilon}$$

$$E(Y) \cong f(\theta,\eta_t,0) - g(\theta,\eta_t,0)\eta_t$$

$$V(Y) \cong [g(\theta,\eta_t,0)]^2\omega^2 + [h(\theta,\eta_t,0)]^2\sigma^2 + 2\cdot g(\theta,\eta_t,0)\cdot h(\theta,\eta_t,0)\cdot COV(\eta,\varepsilon)$$

Because the covariance between η and ε is assumed to be zero, the last term of approximated $V(Y)$, $2\cdot g(\theta, \eta_t, 0)\cdot h(\theta, \eta_t, 0)\cdot COV(\eta, \varepsilon)$, is eliminated.

$$V(Y) \cong [g(\theta,\eta_t,0)]^2\omega^2 + [h(\theta,\eta_t,0)]^2\sigma^2$$

In NONMEM®, the FO method assumes η_t is zero; therefore, $E(Y)$ becomes

$$E(Y) \cong f(\hat{\theta},0,0) - g(\hat{\theta},0,0)\cdot 0 = f(\hat{\theta},0,0)$$

Meanwhile, there are different ways to obtain $V(Y)$ for each INTERACTION estimation option in NONMEM®. This option considers the interaction of η and ε and uses $\hat{\eta}$ (estimate of true eta, η_t) instead of 0 for the calculation of the variance of Y. This is a little contradictory in that the FO method assumes EBE is zero. Therefore, the authors discourage using the INTERACTION option in the FO method, while with other estimation methods like FOCE and LAPLACIAN, the use of INTERACTION is generally recommended. This is because the INTERACTION option did not improve the output with the additive residual error model, while it gave more accurate estimates for other residual error models (proportional error, combined additive and proportional error, power function error, etc.). Because this article only deals with the FO method, $V(Y)$ was assumed as follows:

$$V(Y) \cong [g(\hat{\theta},0,0)]^2\hat{\omega}^2 + [h(\hat{\theta},0,0)]^2\hat{\sigma}^2$$

If the POSTHOC option is used, NONMEM® will estimate η based on the final estimates of θ, Ω and Σ for each individual after the end of the estimation step. This estimate of each subject's $\eta(\hat{\eta})$ has several names – post hoc η, realized η, EBE of η, or maximum a posteriori (MAP) estimate of η. NONMEM® constructs an objective function in order to estimate post hoc η in addition to the objective function for θ, Ω and Σ. The objective function for post hoc η in NONMEM® is as follows[6]:

$$OBJ^{eta} = \log(\sigma^2) + \sum_{j=1}^{n_i}\frac{\left(y_{ij} - f(\theta,\eta,0)\right)^2}{\sigma^2} + \frac{(\eta-0)^2}{\omega^2}$$

$$\hat{\eta} = \arg\min(OBJ^{eta})$$

Vector notations

To reproduce the estimation algorithm in the above section using R, equations were converted into vector notation because R is optimized to calculate using vectors.
For an individual,

$$\vec{Y}_t = \vec{f}(\vec{\theta},\vec{\eta},\vec{\varepsilon}) \cong \vec{f}(\vec{\theta},\vec{\eta}_t,\vec{0}) + G_i(\vec{\theta},\vec{\eta}_t,\vec{0})(\vec{\eta}-\vec{\eta}_t) + H_i(\vec{\theta},\vec{\eta}_t,\vec{0})(\vec{\varepsilon}-\vec{0})$$

$$E(\vec{Y}_i) \cong \vec{f}(\vec{\theta},\vec{\eta}_t,\vec{0}) - G_i(\vec{\theta},\vec{\eta}_t,\vec{0})\vec{\eta}_t \qquad \vec{\eta}_t = \vec{0} \text{ in FO method}$$

$$V(\vec{Y}_i) \cong G_i(\vec{\theta},\vec{\eta}_t,\vec{0})\Omega G_i(\vec{\theta},\vec{\eta}_t,\vec{0})^T + diag\left(H_i(\vec{\theta},\vec{\eta}_t,\vec{0})\Sigma H_i(\vec{\theta},\vec{\eta}_t,\vec{0})^T\right) \quad \text{if } COV(\vec{\eta},\vec{\varepsilon}) = 0$$

Because $\vec{Y}_t$ was approximated by a first-order Taylor series, only the first-order terms for both $\vec{\eta}$ and $\vec{\varepsilon}$ were shown. Due to the assumption of multivariate normality of $\vec{\eta}$ and $\vec{\varepsilon}$, the above $\vec{Y}_t$ also follows multivariate normal distribution with mean $E(\vec{Y}_t)$ and the variance-covariance matrix $V(\vec{Y}_t)$. Thus, $-2LL_i$ were derived by using estimated values for unknown true values as follows:

$$-2LL_i = n_i\log(2\pi) + \log|V(\vec{Y}_i)| + \left(\vec{Y}_i - E(\vec{Y}_i)\right)^T V(\vec{Y}_i)^{-1}\left(\vec{Y}_i - E(\vec{Y}_i)\right)$$
$$\cong n_i\log(2\pi) + \log|V(\vec{Y}_i)| + (\vec{Y}_i - \vec{f}(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0}) + G_i(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0})\hat{\vec{\eta}}_{t,i})^T V(\vec{Y}_i)^{-1}(\vec{Y}_i - \vec{f}(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0}) + G_i(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0})\hat{\vec{\eta}}_{t,i})$$

Thus, the objective function of NONMEM® was denoted as

$$OBJ_i = \log|V(\vec{Y}_i)| + \left(\vec{Y}_i - E(\vec{Y}_i)\right)^T V(\vec{Y}_i)^{-1}\left(\vec{Y}_i - E(\vec{Y}_i)\right)$$
$$\cong \log|V(\vec{Y}_i)| + (\vec{Y}_i - \vec{f}(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0}) + G_i(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0})\hat{\vec{\eta}}_{t,i})^T V(\vec{Y}_i)^{-1}(\vec{Y}_i - \vec{f}(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0}) + G_i(\vec{\theta},\hat{\vec{\eta}}_{t,i},\vec{0})\hat{\vec{\eta}}_{t,i})$$

Overall objective function value is just the sum of each individual objective function value. Thus,

$$\mathrm{OBJ} = \sum_i \mathrm{OBJ}_i = \sum_i \log|V(\vec{Y}_i)| + \sum_i \left(\vec{Y}_i - E(\vec{Y}_i)\right)^T V(\vec{Y}_i)^{-1}\left(\vec{Y}_i - E(\vec{Y}_i)\right)$$
$$\cong \sum_i \log|V(\vec{Y}_i)| + \sum_i (\vec{Y}_i - \vec{f}(\vec{\theta}, \hat{\vec{\eta}}_{t,i}, \vec{0}) + G_i(\vec{\theta}, \hat{\vec{\eta}}_{t,i}, \vec{0})\hat{\vec{\eta}}_{t,i})^T V(\vec{Y}_i)^{-1}(\vec{Y}_i - \vec{f}(\vec{\theta}, \hat{\vec{\eta}}_{t,i}, \vec{0}) + G_i(\vec{\theta}, \hat{\vec{\eta}}_{t,i}, \vec{0})\hat{\vec{\eta}}_{t,i})$$

Because the FO method assumes $\vec{\eta}_t$ is zero, this equation could be simplified as follows:

$$OBJ \cong \sum_i \log|V(\vec{Y}_i)| + \sum_i \left(\vec{Y}_i - \vec{f}(\vec{\theta}, \hat{\vec{\eta}}_{t,i}, \vec{0})\right)^T V(\vec{Y}_i)^{-1}\left(\vec{Y}_i - \vec{f}(\vec{\theta}, \hat{\vec{\eta}}_{t,i}, \vec{0})\right)$$

where

$$V(\vec{Y}_i) \cong G_i(\vec{\theta}, \vec{0}, \vec{0})\Omega G_i(\vec{\theta}, \vec{0}, \vec{0})^T + diag\left(H_i(\vec{\theta}, \vec{0}, \vec{0})\Sigma H_i(\vec{\theta}, \vec{0}, \vec{0})^T\right)$$

We can derive equations in the NONMEM® user's guide from the above equation.

$$OBJ_i = log|V(\vec{Y}_i)| + \left(\vec{Y}_i - E(\vec{Y}_i)\right)^T V(\vec{Y}_i)^{-1}\left(\vec{Y}_i - E(\vec{Y}_i)\right)$$
$$= log|C_i| + (\vec{Y}_i - \vec{F}_i)^T C_i^{-1}(\vec{Y}_i - \vec{F}_i)$$
$$= log|C_i| + (\vec{Y}_i - \vec{F}_i)^T C_i^{-1/2} C_i^{-1/2}(\vec{Y}_i - \vec{F}_i)$$
$$= log|C_i| + (\vec{Y}_i - \vec{F}_i)^T \left[C_i^{-1/2}\right]^T C_i^{-1/2}(\vec{Y}_i - \vec{F}_i) \quad (\because C \text{ is symmetric})$$
$$= log|C_i| + \left[C_i^{-1/2}(\vec{Y}_i - \vec{F}_i)\right]^T C_i^{-1/2}(\vec{Y}_i - \vec{F}_i) \quad (\because (AB)^T = B^T A^T)$$
$$= log|C_i| + R_i^T R_i$$

where

$$\vec{R}_i = C_i^{-1/2}(\vec{Y}_i - \vec{F}_i)$$

$\vec{R}_i$ is called the 'Weighted Residual (WRES)' in NONMEM®, and this equation is the same as the equation in the NONMEM® Users Guide I, p34-41. Because WRES is defined in this way, the sign of WRES could be different from 'Residual (RES)'. $C_i^{-1/2}$ could be calculated using spectral decomposition, as was shown in R code 2.

To estimate $\hat{\vec{\eta}}_{t,i}$ of true $\vec{\eta}_{t,i}$ for i-th subject, final estimates of $\hat{\theta}$, Ω and Σ obtained after the end of estimation step were used.

$$OBJ(\vec{\eta}_{t,i}) = \log|V(\vec{Y}_i)| + \left(\vec{Y}_i - \vec{f}(\vec{\theta}, \vec{\eta}_{t,i}, \vec{0})\right)^T V(\vec{Y}_i)^{-1}\left(\vec{Y}_i - \vec{f}(\vec{\theta}, \vec{\eta}_{t,i}, \vec{0})\right) + \vec{\eta}_{t,i}{}^T \Omega^{-1} \vec{\eta}_{t,i}$$
$$\hat{\vec{\eta}}_{t,i} = \mathrm{argmin}\left(OBJ(\vec{\eta}_{t,i})\right)$$

Implementation

Computing environment

NONMEM® 7.3 (ICON Development Solutions, Ellicott City, MD) running on a GNU GFortran version 4.6.3 FORTRAN compiler under MS-Windows 7 (64 bit) was used for the computation of objective function value. For the R software, R 3.1.2 for MS-Windows was used.

Dataset and population pharmacokinetic model

The THEO dataset in the NONMEM® distribution media was used for exemplary computing of the objective function value. This dataset is readily available, and its use will make it easy to compare the results from other available methods of the future. The THEO dataset has 132 observations from 12 subjects (11 observations per subject following an oral administration of 320 mg theophylline). A one-compartment PK model with first-order absorption, as shown in the model equations below, was used for the calculation of the objective function value.

$$F = \frac{Dose}{V} \cdot \frac{k_a}{k_a - k}(e^{-k \cdot t} - e^{-k_a \cdot t})$$
$$ka = \theta_1 \cdot \exp(\eta_1)$$
$$V = \theta_2 \cdot \exp(\eta_2)$$
$$k = \theta_3 \cdot \exp(\eta_3)$$
$$Y = F + F \cdot \varepsilon_1 + \varepsilon_2$$
$$\vec{\eta} = \begin{bmatrix}\eta_1 \\ \eta_2 \\ \eta_3\end{bmatrix} \sim MVN(\vec{0}, \Omega) \quad \Omega = \begin{bmatrix}\omega_{11}{}^2 & \omega_{12}{}^2 & \omega_{13}{}^2 \\ \omega_{21}{}^2 & \omega_{22}{}^2 & \omega_{23}{}^2 \\ \omega_{31}{}^2 & \omega_{32}{}^2 & \omega_{33}{}^2\end{bmatrix} = \Omega^T$$
$$\vec{\varepsilon} = \begin{bmatrix}\varepsilon_1 \\ \varepsilon_2\end{bmatrix} \sim MVN(\vec{0}, \Sigma) \quad \Sigma = \begin{bmatrix}\sigma_{11}{}^2 & 0 \\ 0 & \sigma_{22}{}^2\end{bmatrix}$$

where *ka*, *V*, *k* are absorption rate constant, volume of distribution, and elimination rate constant, respectively.

Computation of objective function value using user-written code in R

The NONMEM® control stream used to obtain the objective function value, final estimates of θ, Ω and Σ, individual objective function values, and G and H matrices are shown below. Trailing 1s in G11, G21, G31, H11, and H21 are NONMEM's convention.

```
; NONMEM control stream to obtain G and H matrix
  and individual objective function values

$PROBLEM THEOPHYLLINE ORAL
$INPUT ID AMT TIME DV WT
$DATA THEO.CSV
$PRED
 DOSE = 320
 KA   = THETA(1) * EXP(ETA(1))
  V   = THETA(2) * EXP(ETA(2))
  K   = THETA(3) * EXP(ETA(3))
  F   = DOSE / V * KA / (KA - K) * (EXP(-K*TIME) -
        EXP(-KA*TIME))
  Y   = F + F * EPS(1) + EPS(2)
```

```
$THETA
  (0, 2)
  (0, 50)
  (0, 0.1)

$OMEGA BLOCK(3)
  0.2
  0.1  0.2
  0.1  0.1  0.2

$SIGMA
  0.1
  0.1

$EST MAX=9999 METHOD=ZERO
$TABLE ID OBJI FIRSTONLY NOAPPEND ONE
  HEADER FORMAT=s1F14.7 FILE=OBJI.tab
$TABLE ID TIME DV PRED G11 G21 G31 H11 H21
 NOAPPEND ONEHEADER FILE=GH.tab
```

Prior to computing the objective function value in R, two matrix-related functions, one for converting a vector of the lower triangular part to a full matrix (R code 1) and the other for calculating the square-root-inverse of a matrix using spectral decomposition (R code 2), were developed.

```
# R code 1
ltv2mat = function(vec) {
 LENGTH = length(vec)
 DIM = as.integer(round((sqrt(8*LENGTH+1)-1)/2,0))
 if (DIM*(DIM+1)/2 != LENGTH) return(NULL)
 mat = matrix(nrow=DIM, ncol=DIM)
 mat[upper.tri(mat, diag=TRUE)] = vec
 mat[lower.tri(mat)] = t(mat)[lower.tri(mat)]
return(mat)
}
```

```
# R code 2
mat.sqrt.inv = function(mat) {
 ei = eigen(mat)
 d = abs(ei$values)
 d2 = 1 / sqrt(d)
 ans = ei$vectors %*% diag(d2) %*% t(ei$vectors)
return(ans)
}
```

The PRED function needs to be coded in R to calculate the F, G, and H vectors or matrices for the objective function value (R code 3). F is the typical prediction in the FO method. G is the partial derivative of F with respect to η evaluated at $\hat{\eta}$, which is necessary to calculate the total objective function value for θ, Ω and Σ. H is the partial derivative of Y with respect to ε.

$$G1 = \frac{\partial F}{\partial \eta_1} = \frac{\partial F}{\partial k_a}\frac{\partial k_a}{\partial \eta_1} = \left[\frac{-Dose \cdot k}{V(K_a - k)}(e^{-k\cdot t} - e^{-k_a\cdot t}) + \frac{Dose \cdot k_a}{V(K_a - k)}(t \cdot e^{-k_a\cdot t})\right] \cdot k_a$$

$$G2 = \frac{\partial F}{\partial \eta_2} = \frac{\partial F}{\partial V}\frac{\partial V}{\partial \eta_2} = \left[-\frac{Dose \cdot k_a}{V^2(K_a - k)}(e^{-k\cdot t} - e^{-k_a\cdot t})\right] \cdot V$$

$$G3 = \frac{\partial F}{\partial \eta_3} = \frac{\partial F}{\partial k}\frac{\partial k}{\partial \eta_3} = \left[\frac{Dose \cdot k_a}{V(K_a - k)^2}(e^{-k\cdot t} - e^{-k_a\cdot t}) + \frac{Dose \cdot k_a}{V(K_a - k)}(-t \cdot e^{-k\cdot t})\right] \cdot k$$

$$H1 = \frac{\partial Y}{\partial \varepsilon_1} = F$$

$$H2 = \frac{\partial Y}{\partial \varepsilon_2} = 1$$

Where G1, G2, and G3 are partial derivatives of F with respect to η_1, η_2, and η_3, respectively, and H1 and H2 are the partial derivatives of Y with respect to ε_1 and ε_2, respectively.

```
# R code 3
PRED =  function(THETA, ETA, DATA){
 DOSE  = 320
 TIME  = DATA$TIME

 KA = THETA[1] * exp(ETA[1])
  V  = THETA[2] * exp(ETA[2])
  K  = THETA[3] * exp(ETA[3])
  F  = DOSE/V*KA/(KA-K) * (exp(-K*TIME) - exp(-K
       A*TIME))

 G1 = ((DOSE/V/(KA-K)-DOSE/V*KA/(KA-
      K)^2)*(exp(-K*TIME)-exp(-KA*TIME))+DOSE/
      V*KA/(KA-K)*(exp(-KA*TIME)*TIME)) * KA

 G2 = -(DOSE/V^2*KA/(KA-K)*(exp(-K*TIME)-exp(-
      KA*TIME))) * V

 G3 = (DOSE/V*KA/(KA-K)^2*(exp(-K*TIME)-
       exp(-KA*TIME))-DOSE/V*KA/(KA-K)*(exp(-
       K*TIME)*TIME)) * K

 H1 = F
 H2 = 1
 return(cbind(F, G1, G2, G3, H1, H2))
}
```

In order to mimic NONMEM® in R, final estimates of θ, Ω and Σ from NONMEM® results were used to calculate the objective function value. For obtaining the final estimates, we used the .xml file from the NONMEM® output files. The objective function value is the sum of individual objective function values (R

code 4).

```
# R code 4
DATASET = read.csv("THEO.csv", header=TRUE)
names(DATASET) = c("ID", "AMT", "TIME", "DV",
"WT")
ID = unique(DATASET$ID)
OBJ = 0; OBJI.TAB = data.frame(); FGH.TAB = data.
frame(); FIT.TAB = data.frame()
NETA = 3; NEPS = 2

# Final estimate of theta, omega, sigma obtained from
prior NONMEM results
THETA = c(3.1693436270752584, 38.251728321164123,
0.10501216077638235)
OMEGA = c(1.198144159941804, 0.13748126980
932032, 0.0313481792427366263, 0.3700142043422
2171, 0.0433917711628219599, 0.25055031320292215)
SIGMA = c(0.0120775735988250220, 0.0000000000
000000, 0.0542660303512642050)

# Calculate the objective function value
for(i in ID) {
 DATA = DATASET[DATASET$ID==ID[i],]
  ETA  = rep(0, NETA)  # ETA is a vector of zeros for
         FO method.
  FGH  = PRED(THETA, ETA, DATA)
    Y  = DATA$DV
    F  = FGH[,1]
    G  = FGH[,2:4]
    H  = FGH[,5:6]

 OM = ltv2mat(OMEGA)
  SG = ltv2mat(SIGMA)

  RES  = Y - F
 COV  = G %*% OM %*% t(G) + diag(diag(H %*% SG
         %*% t(H)))
 WRES = mat.sqrt.inv(COV) %*% RES
  OBJI  = log(det(COV)) + t(WRES) %*% WRES

 OBJ = OBJ + OBJI
 OBJI.TAB = rbind(OBJI.TAB, cbind(ID=i, OBJI))
 FGH.TAB = rbind(FGH.TAB, cbind(ID=i, FGH))
 FIT.TAB = rbind(FIT.TAB, cbind(DATA, F, RES,
           WRES))
}
```

For the POSTHOC process, the OBETA function is necessary to calculate the objective function for post hoc η (R code 5). In order to minimize the objective function values for post hoc η, the 'L-BFGS-B' method in the 'optim' function was used with zero initial values. Additionally, the standard error of post hoc η was calculated (R code 6).[6] After the POSTHOC process, G and H matrices were recalculated using the final estimates of θ, Ω, Σ, and $\hat{\eta}$ for each individual (R code 7).

```
# R code 5
OBETA = function(ETA){
 FGH = PRED(THETA, ETA, DATA)
 Y  = DATA$DV
 F  = FGH[,1]
 H  = FGH[,(NETA+2):(NETA+NEPS+1)]
 COV = diag(diag(H%*%SG%*%t(H)))
 isum = log(det(COV))+t(Y-F)%*%solve(COV)%*%(Y-
F)+t(ETA)%*%solve(OM)%*%ETA
 return(isum)
}
```

```
# R code 6
ETA = cbind(ID, matrix(nrow=length(ID), ncol=2*3))
colnames(ETA) = c("ID", "ETA1", "ETA2", "ETA3",
"seETA1", "seETA2", "seETA3")
for(i in ID){
DATA = DATASET[DATASET$ID==ID[i],]
#optim: general-purpose optimization based on some
algorithms pre-defined.
#L-BFGS-B method: based on quasi-Newton method,
which allows box constraints.
 RES = optim(rep(0,NETA), OBETA, method="L-
BFGS-B", hessian=T)
 ETA[i, 2:(NETA+1)] = RES$par
 ETA[i, (NETA+2):(2*NETA+1)] = sqrt(diag(2*abs(sol
ve(RES$hessian))))
}
```

```
# R code 7
EBEGH.TAB = data.frame()
for(i in ID){
 DATA = DATASET[DATASET$ID==ID[i],]
 FGH  = PRED(THETA, ETA[i, 2:(NETA+1)], DATA)
  EBEGH.TAB = rbind(EBEGH.
TAB, cbind(ID=i, FGH[,2:(NETA+1)],
FGH[,(NETA+2):(NETA+NEPS+1)]))
}
```

There was no difference between the estimates in the NONMEM® output file and the results from our R codes, e.g., F, G,

and H vectors or matrices, individual objective function values, or objective function.

Summary

The nonlinear mixed effects model aims to describe the data observed in a population of subjects. NONMEM® has become the gold standard, both in the pharmaceutical industry and academia. But it is difficult to understand the estimation methods implemented in NONMEM®. The classical estimation method, FO approximation, is relatively comprehensible compared with other estimation methods. This method was reproduced using the R language, which is more user-friendly and easy to learn than other programming languages such as FORTRAN or C++.

In this tutorial, we demonstrate the calculation steps of the F, G and H vectors or matrices using final estimates of θ, Ω, Σ and the objective function value. In fact, NONMEM® produces the updated θ, Ω and Σ from the ZXMIN1 subroutine to minimize the value of the objective function. ZXMIN1 uses the derivative-free Davidon-Fletcher-Powell algorithm, which is a type of quasi-Newton algorithm. This optimization process for estimation of θ, Ω and Σ is more complex than η and considers bound constraints, scaling, reparameterization, etc. Therefore, we chose not to use this process.

We hope that this tutorial helps pharmacometricians understand the underlying estimation process of nonlinear mixed effects modelling.

Conflict of interest

Nothing to declare

References

1. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. J Pharmacokinet Biopharm 1980;8:553-571.
2. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. J Pharmacokinet Biopharm 1981;9:635-651.
3. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data. J Pharmacokinet Biopharm 1983;11:303-319.
4. Beal S, Sheiner L, Boeckmann A, Bauer R. NONMEM user's guides. Ellicott city, MD, USA, 1989-2009.
5. Wang Y. Derivation of various NONMEM estimation methods. J Pharmacokinet Pharmacodyn 2007;34:575-593.
6. Kang D, Bae KS, Houk BE, Savic RM, Karlsson MO. Standard Error of Empirical Bayes Estimate in NONMEM® VI. Korean J Physiol Pharmacol 2012;16:97-106. doi: 10.4196/kjpp.2012.16.2.97.
7. Beal S, Sheiner L. The NONMEM System. Am Stat 1980;34:118-119.
8. Plan EL, Maloney A, Mentre F, Karlsson MO, Bertrand J. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. AAPS J 2012;14:420-432. doi: 10.1208/s12248-012-9349-2.
9. Fisher D, Shafer S. Fisher/Shafer NONMEM Workshop Pharmacokinetic and Pharmacodynamic Analysis with NONMEM. Het Pand, Ghent, Belgium, 2007.
10. Peck CC, Beal SL, Sheiner LB, Nichols AI. Extended least squares nonlinear regression: a possible solution to the "choice of weights" problem in analysis of individual pharmacokinetic data. J Pharmacokinet Biopharm 1984;12:545-558.
11. Sheiner LB, Beal SL. Pharmacokinetic parameter estimates from several least squares procedures: superiority of extended least squares. J Pharmacokinet Biopharm 1985;13:185-201.
12. Peck CC, Sheiner LB, Nichols AI. The problem of choosing weights in nonlinear regression analysis of pharmacokinetic data. Drug Metab Rev 1984;15:133-148.