NONMEM estimation method derivation (Draft)

Yaning Wang

Definitions:

Y_{ij}: the response at jth time point for ith subject
ψ: mean parameter (θ) and residual variance-covariance matrix (Σ)
Ω: between-subject variance-covariance matrix
η_i: the random subject effect with mean 0 and variance-covariance matrix Ω
e_i: residual
C_i: variance of Yi
Mean-variance model:

$$y_i = f_i(h_i, q) + e_i$$

$$E(y_i) = f_i(h_i, q) \quad Var(y_i) = Var(e_i) = C_i(h_i, y)$$

$$y_i \mid h_i, y \sim N(f(h_i, q), \Sigma) \;\; (1)$$

$$h_i \mid \Omega \sim N(0, \Omega) \;\; (2)$$

Joint probability of y_i and η_i is

$$P(y_i, h_i \mid y, \Omega) = L_i(y, \Omega \mid y_i, h_i) = P(y_i \mid h_i, y) \cdot P(h_i \mid \Omega) = L_i(h_i, y \mid y_i) \cdot P(h_i \mid \Omega) = l_i \cdot h$$

(3)

where l_i is the likelihood of η_i and ψ for the data y_i and h is the density function of η_i.
Since η_i is not observable, we would like to obtain the marginal density function of y_i

$$P(y_i \mid y, \Omega) = L_i(y, \Omega \mid y_i) = \int P(y_i, h \mid y, \Omega) dh = \int l_i(h; y) \cdot h(h; \Omega) dh \;\; (4)$$

which is also the marginal likelihood of ψ and Ω for the data y_i.
The conditional density function of η_i is then

$$P(h_i \mid y, \Omega, y_i) = \frac{P(h_i, y_i \mid y, \Omega)}{P(y_i \mid y, \Omega)} = \frac{l_i(h_i; y) \cdot h(h_i; \Omega)}{L_i(y, \Omega \mid y_i)} \propto l_i(h_i; y) \cdot h(h_i; \Omega) \;\; (5)$$

Since ψ and Ω are obtained as the MLE of equation (4), equation (5) is called the empirical Bayes posterior distribution of η_i. The mode of equation (5) is the post-hoc estimate of η_i.
Since the closed form of equation (4) is difficult to obtain exactly, approximation is necessary to simplify the computation.

Laplacian approximation (First level of approximation)

Given a complex integral, $\int f(x)dx$, $f(x)$ is re-expressed as $e^{\log f(x)} = e^{g(x)}$ and g(x) can be approximated by a ***second-order*** Taylor expansion of g(x) about a point x0

$$\boldsymbol{g(x) \approx g(x_0) + (x - x_0)g'(x_0) + \frac{(x - x_0)^2}{2!} g''(x_0)}$$

The approximated integration is called a **first order** Laplacian approximation to the true integration.

If x_0 is $\hat{x}$, the mode of *f(x)*, the second term will be zero since $g'(x_0)=0$. Then

$$\int f(x)dx = \int e^{g(x)}dx \approx \int e^{g(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = f(x_0)\cdot\sqrt{\frac{2p}{-g''(x_0)}}$$ (see Q&A for derivation)

But if x_0 is not the mode, the result will be

$$\int f(x)dx = \int e^{g(x)}dx \approx \int e^{g(x_0)+(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = f(x_0)\cdot\int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx$$

$$= f(x_0)\cdot\sqrt{\frac{2p}{-g''(x_0)}}\cdot e^{\frac{-g'(x_0)^2}{2g''(x_0)}}$$ (6, see Q&A for derivation)

By definition,

$$-2\log l_i(h;y) = \Phi_i(h)$$

$$\Gamma_i(h) = \frac{d-2\log l_i(h;y)}{dh} = -2\frac{l_i'}{l_i}$$ (see Q&A for derivation) and $\frac{l_i'}{l_i} = \frac{\Gamma_i(h)}{-2}$

$$\Delta_i(h) = \frac{d\Gamma_i(h)}{dh} = -2\left(\frac{l_i'}{l_i}\right)'$$ and $\left(\frac{l_i'}{l_i}\right)' = \frac{\Delta_i(h)}{-2}$

And we know

$$h = \frac{1}{\sqrt{2p}\,|\Omega|^{\frac{1}{2}}}\cdot e^{-\frac{h'\Omega^{-1}h}{2}}$$ and

$$h' = \frac{1}{\sqrt{2p}\,|\Omega|^{\frac{1}{2}}}\cdot e^{-\frac{h'\Omega^{-1}h}{2}}\cdot\left(-\frac{2\Omega^{-1}h}{2}\right)$$

Therefore

$$\frac{h'}{h} = -\Omega^{-1}h$$ and $\left(\frac{h'}{h}\right)' = -\Omega^{-1}$

Since the integrand is

$$f(h) = l_i(h;y)\cdot h(h;\Omega),$$

$$g(h) = \log l_i(h;y) + \log h(h;\Omega)$$

$$g'(h) = \frac{l_i'}{l_i} + \frac{h'}{h} = -\frac{\Gamma_i(h)}{2} - \Omega^{-1}h$$

$$g''(h) = \left(\frac{l_i'}{l_i}\right)' + \left(\frac{h'}{h}\right)' = -\frac{\Delta_i(h)}{2} - \Omega^{-1}$$

$$-2\log L_i(y, \Omega \mid y_i) = -2\log\int l_i(h;y)\cdot h(h;\Omega)dh$$

$$\approx -2\log\left\{l_i(\hat{h};y)\cdot h(\hat{h};\Omega)\sqrt{\frac{2p}{|-g''(\hat{h})|}}\cdot e^{\frac{-g'(\hat{h})'g''(\hat{h})^{-1}g'(\hat{h})}{2}}\right\}$$ (apply equation 6)

$$= -2\log l_i(\hat{\eta}; y) - 2\log h(\hat{\eta};\Omega) - 2\log\sqrt{\frac{2\pi}{\left|-g''(\hat{\eta})\right|}} - 2\log e^{\frac{-g'(\hat{\eta})' g''(\hat{\eta})^{-1} g'(\hat{\eta})}{2}}$$

$$= \Phi_i(\hat{\eta}) - 2\log\frac{1}{\sqrt{2\pi}\left|\Omega\right|^{\frac{1}{2}}} - 2\log e^{-\frac{\hat{\eta}'\Omega^{-1}\hat{\eta}}{2}} - \log\frac{2\pi}{\left|-g''(\hat{\eta})\right|} + g'(\hat{\eta})' g''(\hat{\eta})^{-1} g'(\hat{\eta})$$

$$\propto \Phi_i(\hat{\eta}) + \log\left|\Omega\right| + \hat{\eta}'\Omega^{-1}\hat{\eta} + \log\left|-g''(\hat{\eta})\right| + g'(\hat{\eta})' g''(\hat{\eta})^{-1} g'(\hat{\eta})$$

$$= \Phi_i(\hat{\eta}) + \log\left|\Omega\right| + \hat{\eta}'\Omega^{-1}\hat{\eta} + \log\left|\Omega^{-1} + \frac{\hat{\Delta}_i}{2}\right| - \left(\frac{\hat{\Gamma}_i}{2} + \Omega^{-1}\hat{\eta}\right)'\left(\Omega^{-1} + \frac{\hat{\Delta}_i}{2}\right)^{-1}\left(\frac{\hat{\Gamma}_i}{2} + \Omega^{-1}\hat{\eta}\right) \quad (7)$$

If $\hat{\eta}$ is the mode of the equation (5), the last term in equation above is zero since $g'(\hat{\eta})$ is 0. Or

$$-2\log L_i(y, \Omega \mid y_i) = \Phi_i(\hat{\eta}) + \log\left|\Omega\right| + \hat{\eta}'\Omega^{-1}\hat{\eta} + \log\left|\Omega^{-1} + \frac{\hat{\Delta}_i}{2}\right|$$

FOCE (second level of approximation)

$\Delta_{i(k)}(\eta) \approx \frac{1}{2} E\left(\Gamma_{i(k)}(\eta)\Gamma_{i(k)}(\eta)'\right)$ (Should not this be an outer-product instead of an inter-product as in NONMEM manual in order to obtain a matrix?)

Hessian matrix is approximated by a function of the gradient vector due to the difficulty for the direct computation of a Hessian matrix, which is a type of first order approximation. Therefore it is called first order conditional estimation.

FO (third level of approximation)

$\Delta_{i(k)}(\eta) \approx \frac{1}{2} E\left(\Gamma_{i(k)}(\eta)\Gamma_{i(k)}(\eta)'\right)$ **AND** $\hat{\eta} = \mathbf{0}$

$$-2\log L_i(y, \Omega \mid y_i) = \Phi_i(\mathbf{0}) + \log\left|\Omega\right| + \log\left|\Omega^{-1} + \frac{\hat{\Delta}_i}{2}\right| - \left(\frac{\hat{\Gamma}_i}{2}\right)'\left(\Omega^{-1} + \frac{\hat{\Delta}_i}{2}\right)^{-1}\left(\frac{\hat{\Gamma}_i}{2}\right)$$

This is equivalent to taking a first order Taylor expansion on the nonlinear structure model around $\hat{\eta} = \mathbf{0}$ (NONMEM manual I, where the name “first order” comes from) and then obtaining the maximum likelihood estimators. Therefore, the origins of “first order” in FO and FOCE are different. But if it is FOCE without interaction, the “first order” of FOCE can also be referred to as the first order Taylor expansion on the nonlinear structure model around $\hat{\eta}$.

Q&A

1. Q: Why $\int e^{g(x_0)+\frac{(x-x_0)^2}{2!}g'(x_0)}\, dx = f(x_0)\sqrt{\frac{2\pi}{-g''(x_0)}}$?

A: The left side of the equation comes from the second order Taylor expansion around the mode. On the right side of the equation, $f(x_0)$ is from $\exp(g(x_0))=\exp(\log f(x_0))=f(x_0)$, which is constant. The square root part is from the rest of the integral, which is the kernel of a normal density with mean $\mu=x_0$ and variance $\sigma^2=-1/g''(x_0)$. ($\int e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$)

2. Q: Why $\Gamma(h) = \frac{d[-2LL]}{dh} = -2\frac{l_i'}{l_i}$?

A: $\Gamma_i(h) = \frac{d-2\log l_i(h;y)}{dh} = -2\frac{l_i'}{l_i}$ follows the chain rule of derivative:

$$\frac{d\log(f(x))}{dx} = \frac{f'(x)}{f(x)}$$

3. Q: Why $f(x_0)\int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g'(x_0)}\, dx = f(x_0)\sqrt{\frac{2\pi}{-g''(x_0)}}\exp\left(\frac{-g'(x_0)^2}{2g''(x_0)}\right)$?

A: I will first describe the lengthy method, which, however, everyone with basic algebra knowledge and statistics will understand. Then I will describe the much shorter method which requires the knowledge of moment generating function for a random variable.

1. Lengthy one

$$(x-x_0)g'(x_0) + \frac{(x-x_0)^2}{2}g''(x_0) = xg'(x_0) - x_0g'(x_0) + \frac{x^2-2xx_0+x_0^2}{2}g''(x_0)$$

$$= \frac{x^2}{2}g''(x_0) - x[x_0g''(x_0) - g'(x_0)] + \frac{x_0^2}{2}g''(x_0) - x_0g'(x_0)$$

$$= \frac{g''(x_0)}{2}\left\{x^2 - 2\left[x_0 - \frac{g'(x_0)}{g''(x_0)}\right]x\right\} + \frac{x_0^2}{2}g''(x_0) - x_0g'(x_0)$$

$$= \frac{g''(x_0)}{2}\left\{\left[x - \left(x_0 - \frac{g'(x_0)}{g''(x_0)}\right)\right]^2 - \left[x_0 - \frac{g'(x_0)}{g''(x_0)}\right]^2\right\} + \frac{x_0^2}{2}g''(x_0) - x_0g'(x_0)$$

$$= \frac{g''(x_0)}{2}\left[x - \left(x_0 - \frac{g'(x_0)}{g''(x_0)}\right)\right]^2 - \frac{g''(x_0)}{2}\left[x_0 - \frac{g'(x_0)}{g''(x_0)}\right]^2 + \frac{x_0^2}{2}g''(x_0) - x_0g'(x_0)$$

$$= \frac{g''(x_0)}{2}\left[x - \left(x_0 - \frac{g'(x_0)}{g''(x_0)}\right)\right]^2 - \frac{g''(x_0)}{2}\left[x_0 - \frac{g'(x_0)}{g''(x_0)}\right]^2 + \frac{x_0^2}{2}g''(x_0) - x_0g'(x_0)$$

$$= \frac{g''(x_0)}{2}\left[x - \left(x_0 - \frac{g'(x_0)}{g''(x_0)}\right)\right]^2 - \left[\frac{x_0^2}{2}g''(x_0) - x_0 g'(x_0) + \frac{g'(x_0)^2}{2g''(x_0)}\right] + \frac{x_0^2}{2}g''(x_0) - x_0 g'(x_0)$$

$$= \frac{g''(x_0)}{2}\left[x - \left(x_0 - \frac{g'(x_0)}{g''(x_0)}\right)\right]^2 - \frac{g'(x_0)^2}{2g''(x_0)}$$

Then

$$\int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = \int e^{\frac{g''(x)}{2}\left[x-\left(x_0-\frac{g'(x'_0)}{g''(x_0)}\right)\right]^2-\frac{g'(x_0)^2}{2g''(x_0)}}dx$$

$$= e^{-\frac{g'(x_0)^2}{2g''(x_0)}}\int e^{\frac{g''(x)}{2}\left[x-\left(x_0-\frac{g'(x'_0)}{g''(x_0)}\right)\right]^2}dx$$

$$= e^{\frac{-g'(x_0)^2}{2g''(x_0)}}\sqrt{\frac{2\pi}{-g''(x_0)}}$$ (again the integrand is the kernel of a normal density function)

2. Shorter one

We know the moment generating function (MGF), $M_x(t)$, of a normal random variable $x \sim N(\mu, \sigma^2)$ is

$$M_x(t) = E(e^{xt}) = \int \frac{1}{\sqrt{2\pi\sigma}} e^{xt} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

$$\int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = \int e^{yg'(x_0)+\frac{y^2}{2}g''(x_0)}dy$$ (What does this look like?)

This is just the kernel of MGF with $\mu = 0$, $t = g'(x_0)$ and $\sigma^2 = -\frac{1}{g''(x_0)}$.

$$\int e^{yg'(x_0)+\frac{y^2}{2}g''(x_0)}dy = \int e^{yt}e^{-\frac{y^2}{2\sigma^2}}dy = \sqrt{2\pi\sigma}\cdot\int \frac{1}{\sqrt{2\pi\sigma}}e^{yt}e^{-\frac{y^2}{2\sigma^2}}dy$$

$$= \sqrt{2\pi\sigma}\cdot e^{\frac{\sigma^2 t^2}{2}}$$

$$= \sqrt{\frac{2\pi}{-g''(x_0)}}\cdot e^{\frac{-g'(x_0)^2}{2g''(x_0)}}$$

BINGO! ☺

In fact, the first lengthy method is basically how the result for moment generating function is derived.

If you don't know too much about MGF, try the link (MGF) on my website.