

Web-based Supplementary Materials for Maximum Likelihood Estimation in Dynamical Models of HIV by Guedj et al.

J. Guedj,^{1,2} R. Thiébaud^{1,2,*} and D. Commenges^{1,2}

¹INSERM, U875 (Biostatistique), Bordeaux, F-33076, FRANCE

²Université Victor Segalen Bordeaux 2, Bordeaux, F-33076, FRANCE

1. Web Appendix A: Computation of $\int_{\mathbf{u} \in \mathbb{R}^q} \mathcal{L}_{\mathcal{F}_i|\mathbf{u}^{(i)}}(\mathbf{u})\phi(\mathbf{u})d\mathbf{u}$

We want to compute a function such as: $I = \int_{\mathbf{u} \in \mathbb{R}^q} f_i(\mathbf{u})\phi(\mathbf{u})d\mathbf{u}$. Some authors (Liu and Pierce, 1994; Pinheiro and Bates, 2000) underlined the impact of a transformation of the integrand leading to

$$I = \frac{\exp\left(-\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{2}\right)}{\det(K)} \int_{\mathbf{u} \in \mathbb{R}^q} f_i(K^{-1}\mathbf{u} + \hat{\mathbf{u}}) \exp[-(K^{-1}\mathbf{u})'\hat{\mathbf{u}}] \exp\left\{\frac{\mathbf{u}'[I - (KK')^{-1}]\mathbf{u}}{2}\right\} \phi(\mathbf{u})d\mathbf{u}$$

where

$$\hat{\mathbf{u}} = \underset{\mathbf{u} \in \mathbb{R}^q}{\operatorname{argmax}} f_i(\mathbf{u})\phi(\mathbf{u}), \quad K'K = -\frac{\partial^2 \log(f_i(\mathbf{u})\phi(\mathbf{u}))}{\partial \mathbf{u} \partial \mathbf{u}'} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}$$

where K is an upper triangular matrix with positive diagonal elements. Using the notations of the manuscript, we rewrite $\hat{\mathbf{u}}$ as:

$$\left\{ \begin{array}{l} \hat{\mathbf{u}} = \underset{\mathbf{u} \in \mathbb{R}^q}{\operatorname{argmax}} \left\{ \sum_{m \leq M, j \leq n_{im}} \frac{1}{\sigma_m^2} \left[(Y_{ijm} - g_m(X(t_{ijm}), \tilde{\boldsymbol{\xi}}^{(i)}(\mathbf{u})) \right]^2 + \mathbf{u}'\mathbf{u} \right\} \\ \tilde{\xi}_l^{(i)}(\mathbf{u}) = \phi_l + \mathbf{z}_l^{(i)'}\boldsymbol{\beta}_l + \sum_{l' \leq q} \sum_{l'' \geq l'} u_{l'} w_{ll''}^{(i)} a_{l''l'} \end{array} \right.$$

Some problems can occur when looking for $\hat{\mathbf{u}}$. The maximization over the multidimensional and infinite region \mathbb{R}^q is potentially time consuming and

* email: rt@isped.u-bordeaux2.fr

difficult to reach when the starting values of maximization algorithm are far from the modal value. So, it is necessary to determine a threshold beyond which a maximum is not considered as the global maximum. By considering that the distribution of $\frac{f_i(\hat{\mathbf{u}})}{n_i}$ is close to a chi-square distribution with one degree of freedom near the maximum likelihood, we consider in the whole procedure that if $\frac{f_i(\hat{\mathbf{u}})}{n_i} > 3$, the maximization procedure should be restarted with different starting values to find the global maximum. We propose a systematic procedure, which turns out to be efficient to find good starting points \mathbf{u}_0 . We propose to maximize $f_i(\mathbf{u})\phi(\mathbf{u}) = \mathcal{L}_{\mathcal{F}_i|\mathbf{u}(i)}(\mathbf{u})\phi(\mathbf{u})$ along each random effect separately, which enable to reduce a multidimensional optimization to an optimization of dimension 1, that is to find $\hat{u}_i^{(r)}$ for $r = 1, \dots, q$:

$$\begin{cases} \hat{u}_r^{(i)} = \underset{u \in \mathbb{R}}{\operatorname{argmax}} \left\{ \sum_{m \leq M, j \leq n_{im}} \frac{1}{\sigma_m^2} \left[(Y_{ijm} - g_m(X(t_{ijm}, \tilde{\boldsymbol{\xi}}_r^{(i)}(u))) \right]^2 + u^2 \right\} \\ \tilde{\boldsymbol{\xi}}_{l,r}^{(i)}(u) = \phi_l + \mathbf{z}_l^{(i)'} \boldsymbol{\beta}_l + u \sum_{l'' \geq r} w_{ll''}^{(i)} a_{l''r} \end{cases} \quad (1)$$

$\mathbf{u}_0^{(i)} = (u_{10}^{(i)}, \dots, u_{q0}^{(i)})$ turns out to be a good starting point for maximization with: $u_{r0}^{(i)} = \frac{\hat{u}_r^{(i)} * f(\hat{u}_r^{(i)}) * \phi(\hat{u}_r^{(i)})}{\sum_{l=1, \dots, q} [f(\hat{u}_l^{(i)}) * \phi(\hat{u}_l^{(i)})]}$.

Maximization was then achieved by Marquardt algorithm (Marquardt, 1963) and $K'K$ was computed by numerical derivative.

Note that if the values used for the set $(\phi_l, \boldsymbol{\beta}_l, A, \boldsymbol{\sigma})$ in equation (1) are equal to the final estimated values (see Table 2), then $\hat{\mathbf{u}}$ is also the posterior mode $\hat{\mathbf{u}}^{(i)}$ as introduced in section 3.3 for computing the individual predictions.

2. Web Appendix B: Systems of sensitivity equations

In this appendix, we show an example for the semi-analytical computation of the $\frac{\partial g_m(X(t, \boldsymbol{\xi}^{(i)}))}{\partial \boldsymbol{\xi}^{(i)}}$, $m \leq M$ by using the systems of sensitivity equations.

The system of ODE used in section 4.2 is:

$$\begin{cases} \frac{dQ}{dt} = \lambda + \rho T - \alpha Q - \mu_Q Q \\ \frac{dT}{dt} = \alpha Q - (1 - \eta)\gamma T V_I - \rho T - \mu_T T \\ \frac{dT^*}{dt} = (1 - \eta)\gamma T V_I - \mu_{T^*} T^* \\ \frac{dV_I}{dt} = \omega \mu_{T^*} \pi T^* - \mu_v V_I \\ \frac{dV_{NI}}{dt} = (1 - \omega) \mu_{T^*} \pi T^* - \mu_v V_{NI} \end{cases}$$

We take as an example (for sake of clarity, we omit the index i):

$$\begin{cases} \xi_1 = \eta \text{ (the effect of treatment)} \\ g_1(X(t, \boldsymbol{\xi})) = \log_{10}(V_I(t, \boldsymbol{\xi}) + V_{NI}(t, \boldsymbol{\xi})), \text{ (the total quantity of free virions)} \end{cases}$$

We have:

$$\frac{\partial g_1(X(t, \boldsymbol{\xi}))}{\partial \xi_1} = \frac{1}{\log(10)} \frac{V_{I, \xi_1}(t, \boldsymbol{\xi}) + V_{NI, \xi_1}(t, \boldsymbol{\xi})}{V_I(t, \boldsymbol{\xi}) + V_{NI}(t, \boldsymbol{\xi})},$$

where $V_{I, \xi_1}(t, \boldsymbol{\xi})$ and $V_{NI, \xi_1}(t, \boldsymbol{\xi})$ are obtained by simultaneously solving the original system and the system obtained by differentiating the original system with respect to $\xi_1 = \eta$. That it, we have to solve the following ODE system:

$$\begin{cases} \frac{dQ}{dt} = \lambda + \rho T - \alpha Q - \mu_Q Q \\ \frac{dT}{dt} = \alpha Q - (1 - \eta)\gamma T V_I - \rho T - \mu_T T \\ \frac{dT^*}{dt} = (1 - \eta)\gamma T V_I - \mu_{T^*} T^* \\ \frac{dV_I}{dt} = \omega \mu_{T^*} \pi T^* - \mu_v V_I \\ \frac{dV_{NI}}{dt} = (1 - \omega) \mu_{T^*} \pi T^* - \mu_v V_{NI} \\ \frac{dQ_{\xi_1}}{dt} = \rho T_{\xi_1} - \alpha Q_{\xi_1} - \mu_Q Q_{\xi_1} \\ \frac{dT_{\xi_1}}{dt} = \alpha Q_{\xi_1} + \gamma T V_I - (1 - \eta)\gamma T_{\xi_1} V_I - (1 - \eta)\gamma T V_{I, \xi_1} - \rho T_{\xi_1} - \mu_T T_{\xi_1} \\ \frac{dT_{\xi_1}^*}{dt} = -\gamma T V_I + (1 - \eta)\gamma T_{\xi_1} V_I + (1 - \eta)\gamma T V_{I, \xi_1} - \mu_{T^*} T_{\xi_1}^* \\ \frac{dV_{I, \xi_1}}{dt} = \omega \mu_{T^*} \pi T_{\xi_1}^* - \mu_v V_{I, \xi_1} \\ \frac{dV_{NI, \xi_1}}{dt} = (1 - \omega) \mu_{T^*} \pi T_{\xi_1}^* - \mu_v V_{NI, \xi_1} \end{cases}$$

The initial conditions of this system are given by deriving the initial conditions of the original system with respect to ξ_1 ; here $t = 0$ is the initiation of therapy, so all the components are equal to zero $Q_{\xi_1}(0) = T_{\xi_1}(0) = T_{\xi_1}^*(0) = V_{I,\xi_1}(0) = V_{NI,\xi_1}(0) = 0$.

The expressions for the corresponding $\frac{\partial g_m(X(t, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial \tilde{\boldsymbol{\xi}}^{(i)}}$ are obtained by the relation:

$$\frac{\partial g_m(X(t, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial \tilde{\boldsymbol{\xi}}^{(i)}} = \frac{\partial g_m(X(t, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial \boldsymbol{\xi}^{(i)}} \frac{\partial \boldsymbol{\xi}^{(i)}}{\partial \tilde{\boldsymbol{\xi}}^{(i)}}$$

3. Web Appendix C: Implementation

Solver for ODE

The computation of $\mathcal{L}_{\mathcal{F}_i|\mathbf{u}^{(i)}}$ and $\mathbf{U}_{\mathcal{F}_i|\mathbf{u}^{(i)}}$ requires to solve numerically the original system of ODE (1) and its systems of sensitivity equations. Here we use a backward difference formula (BDF) method (Gear type method). We implemented the Livermore solver DLSODE, specially adapted for stiff systems (Radhakrishnan and Hindmarsh, 1993). Both relative and absolute precision were set at 10^{-11} .

Computation of the observed likelihood

The “exact” computation of $\mathcal{L}_{\mathcal{O}_i}$ can be split into: i) finding the maximum of the integrand $\mathcal{L}_{\mathcal{F}_i|\mathbf{u}^{(i)}}(\mathbf{u})\phi(\mathbf{u})$ and deducing the Hessian of the log of the integrand at the maximum (see Web Appendix A); ii) transforming the integrand and applying a program developed by Genz and Keister (1996); this routine enables to reach more precision than basic adaptive Gaussian quadrature based on Gauss-Hermite nodes. For $q=3$ (the case encountered in the application), the transformations used here make it possible to compute the observed log-likelihood $L_{\mathcal{O}_i}$ with an absolute precision of 10^{-3} in less than 100 evaluations of the integrand. Moreover it can be noted that the

time requested for adapting the integrals is completely negligible compared to the gain in the number of evaluations required.

Maximization

Concerning starting points, we suggest to use (when feasible) order of magnitude provided by empirical means and variances of individual fits. However, when few data per patients are available, the individual estimation procedure may fail. In this case, we suggest to estimate an approximated population model, in which all the patients have the same parameters, avoiding in a first approximation the multidimensional integrals. This enables to provide a reasonable order of magnitude.

The speed of convergence is then dependent from these values. For three random effects, it can be up to four hours on a Bi-Xeon 3.06 GHz 1024 MB RAM (program written in FORTRAN). We encourage to try different starting values to avoid local maximum. Nevertheless, the use of population model increases the practical identifiability and in our case we did not encounter problems of local extrema. The stopping value (convergence criterion) used here was $0.01 \cdot \dim(\theta)$ for real data study analysis and $0.05 \cdot \dim(\theta)$ for simulations (see Commenges et al. (2006) for the statistical meaning of the choice value)

REFERENCES

- Commenges, D., Jacqmin-Gadda, H., Proust, C. and Guedj, J. (2006). A Newton-like algorithm for likelihood maximization: the robust variance scoring algorithm. *arXiv:math.ST/0610402*.
- Genz, A. and Keister, B. D. (1996). Fully symmetric Interpolatory Rules for

- Multiple Integrals over Infinite Regions with Gaussian Weight. *Journal of Computational and Applied Mathematics* **71**, 299–311.
- Liu, Q. and Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika* **81**, 624–629.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11**, 431–441.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, London.
- Radhakrishnan, K. and Hindmarsh, A. C. (1993). Description and use of lsode, the livermore solver for ordinary differential equations. *LLNL Report UCR-ID-113855 Livermore, CA* **4**.