

ANNABELLE DIOT

ALAIN MALLET

RENÉ BRUNO

**Adaptation des quadratures de Gauss-Hermite  
pour améliorer l'estimation dans les modèles  
non linéaires à effets mixtes**

*Journal de la société française de statistique*, tome 143, n° 1-2 (2002),  
p. 195-204

[http://www.numdam.org/item?id=JSFS\\_2002\\_\\_143\\_1-2\\_195\\_0](http://www.numdam.org/item?id=JSFS_2002__143_1-2_195_0)

© Société française de statistique, 2002, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# ADAPTATION DES QUADRATURES DE GAUSS-HERMITE POUR AMÉLIORER L'ESTIMATION DANS LES MODÈLES NON LINÉAIRES À EFFETS MIXTES

Annabelle DIOT <sup>1, 2</sup>, Alain MALLET <sup>1</sup>, René BRUNO <sup>3</sup>

## RÉSUMÉ

Plusieurs méthodes ont été proposées permettant d'estimer les différents paramètres des modèles non linéaires à effets mixtes. Fréquemment utilisées en modélisation pharmacocinétique et pharmacodynamique de population, ces méthodes sont pour la plupart basées sur une approximation de la fonction vraisemblance. C'est la précision de cette approximation qui gouverne la qualité d'estimation des paramètres fixes et de la distribution des effets aléatoires. Certaines propriétés concernant les effets aléatoires, telle que la multimodalité, risquent de passer inaperçues pour cause d'une approximation trop grossière. Afin d'obtenir un degré d'approximation satisfaisant, une approche méthodologique basée sur l'utilisation des quadratures de Gauss-Hermite dans le calcul de la vraisemblance est proposée. Ce travail montre comment cette méthode s'adapte au problème d'approximation de vraisemblance dans les modèles non linéaires à effets mixtes même en présence d'hétérogénéités dans une population.

## ABSTRACT

Several methods have been proposed to estimate the different parameters entering the non linear mixed effects models. Frequently used in population pharmacokinetics/pharmacodynamics, nearly all of these methods rely upon approximations of the likelihood function. The accuracy of this approximation is likely to govern the accuracy of the estimation of the fixed parameters and of the distribution of the random effects; subtle features such as multimodality may be lost due to a rough approximation. In order to compute likelihood up to a satisfactory degree of approximation, a methodological approach is proposed, based on the use of the Gauss-Hermite quadrature. The aim of this work is to show how this quadrature is adjustable to the problem of the likelihood approximation in non linear mixed effects models with for example heterogeneities in a population.

---

1. Inserm U436, CHU Pitié Salpêtrière, 91 bd de l'Hôpital, 75013 Paris, France

2. IRIS, 6 place des pléiades, 92415 Courbevoie, France

3. Genentech, Inc., South San Francisco; E-mail : adi@biomath.jussieu.fr

## 1. Introduction

Afin de décrire au mieux le processus pharmacocinétique d'un médicament (processus décrivant son administration, sa présence dans l'organisme, puis son élimination) il est nécessaire de développer des modèles rendant compte d'une part des mécanismes physiologiques et physiopathologiques mis en jeu et d'autre part des variabilités biologiques. Les modèles doivent contenir une dimension déterministe ainsi qu'une dimension probabiliste conférant à certains des paramètres le statut de variables aléatoires. Ce travail a donc pour cadre technique les modèles mixtes, et plus précisément, de par leur application à la pharmacocinétique, les modèles non linéaires à effets mixtes. Le but de l'approche méthodologique proposée est de parvenir à une estimation précise des paramètres fixes et de la distribution des effets aléatoires dans ces modèles, distribution que nous avons choisie sous forme de mélange de lois normales afin de pouvoir décrire certaines hétérogénéités au sein de la population de patients.

Cette estimation repose sur la maximisation de la vraisemblance. Cependant, dans ces modèles non linéaires, la fonction vraisemblance s'exprime par une intégrale que l'on ne peut calculer exactement. Afin d'en obtenir une approximation d'un niveau satisfaisant, l'utilisation des quadratures numériques de Gauss-Hermite est proposée.

Dans une première partie nous présentons les adaptations et optimisations à effectuer au niveau de ces quadratures afin de pouvoir les utiliser en s'assurant une bonne qualité d'approximation de la fonction vraisemblance. Nous présentons dans une seconde partie une application dans le domaine de la pharmacocinétique mettant l'accent sur l'un des intérêts majeurs de cette approche : la prise en compte d'hétérogénéités. Enfin, nous concluons en proposant une stratégie d'analyse pour utiliser au mieux cette approche, tant par la qualité des résultats que par le temps d'implémentation.

## 2. Méthodes

### 2.1. Vraisemblance et Quadrature

En analyse de population, où les sujets sont indépendants conditionnellement aux paramètres et à la distribution des effets aléatoires, la fonction vraisemblance est égale au produit des vraisemblances des individus, celles-ci étant de la forme :

$$L_i = \int_{R^n} g_i(y_i, \theta, Z_i, \eta) \cdot h(\eta) d\eta$$

avec  $y_i$  les observations du sujet  $i$ ,  $Z_i$  les covariables du sujet  $i$ ,  $\theta$  les paramètres fixes,  $\eta \in R^n$  les paramètres aléatoires,  $g_i$  la vraisemblance individuelle du sujet  $i$  pour une valeur fixée des paramètres  $\theta$  et  $\eta$ , et  $h$  la distribution des effets aléatoires. Nous avons choisi  $h$  sous forme d'un mélange de  $p$  lois normales de moyenne  $\mu_j$  et de variance-covariance  $Q'_j Q_j$  :

$$h(\eta) = \sum_{j=1}^p \left\{ \alpha_j \cdot (2\pi)^{-n/2} \cdot |Q'_j Q_j|^{-1/2} \cdot \exp \left[ -\frac{1}{2} (\eta - \mu_j)' \cdot (Q'_j Q_j)^{-1} \cdot (\eta - \mu_j) \right] \right\}$$

La fonction vraisemblance liant les observations aux valeurs des paramètres fixes et aux distributions des effets aléatoires est généralement impossible à calculer exactement. Afin d'obtenir une approximation de cette fonction, nous proposons l'utilisation des quadratures numériques de Gauss-Hermite.

La formule brute de la quadrature de Gauss-Hermite d'ordre  $r$  pour une fonction univariée  $f$  est la suivante :

$$\int_R f(x) \cdot \exp(-x^2) \cdot dx \approx \sum_{q=1}^r \left[ f(\xi^{(q)}) \cdot A^{(q)} \right] \quad (1)$$

où  $\xi^{(q)}$  sont les nœuds de la quadrature, et  $A^{(q)}$ , les poids de la quadrature. Ces nœuds et ces poids sont issus des polynômes orthogonaux d'Hermite de degré  $r$ . Ces polynômes sont de la forme

$$P_r(x) = (-1)^r \cdot \exp(x^2) \cdot \frac{d^r (\exp(-x^2))}{dx^r}(x)$$

Les nœuds sont les zéros de ces polynômes et les poids sont obtenus à l'aide du calcul des polynômes dérivés. Ces  $\xi^{(q)}$  et  $A^{(q)}$  sont donc fixés et tabulés dans des tables que l'on peut trouver dans l'ouvrage de Stroud et Secrest (1966).

Plusieurs changements de variables sont nécessaires afin d'adapter les quadratures de Gauss-Hermite à la fonction vraisemblance. Par ailleurs, quelques ajustements seront également mis en œuvre pour s'assurer de la bonne précision de cette estimation.

## 2.2. Adaptation et Qualité d'approximation

Si  $f$  est multivariée avec  $x \in R^n$ , on montre que l'approximation de l'intégrale de  $f(x) \cdot \exp(-x'x)$  par la quadrature s'écrit sous la forme d'une somme de  $r^n$  éléments du type de ceux de la quadrature de base (1).

L'élévation de l'ordre de la quadrature permet une amélioration de la qualité de l'approximation, ceci au prix d'un temps de calcul numérique accru.

Pour assurer une bonne précision dès les ordres faibles, nous proposons d'utiliser la quadrature de Gauss-Hermite adaptée selon des procédures voisines de celles proposées en univarié par Liu et Pierce (1994) et en multivarié par Pinheiro et Bates (1995). Cette adaptation permet d'évaluer la vraisemblance individuelle en des valeurs tenant compte du support de la fonction. Elle repose sur deux principes : la définition des abscisses de la quadrature issues du centrage des nœuds et leur ajustement basé sur une multiplication des valeurs des abscisses par une matrice facteur d'échelle.

## ADAPTATION DES QUADRATURES DE GAUSS-HERMITE

Le centrage consiste à définir les abscisses de la quadrature comme les nœuds centrés à la valeur des effets aléatoires qui maximise l'intégrande  $f(x) \cdot \exp(-x'x)$ , ces valeurs étant obtenues en annulant le gradient de  $f(x) \cdot \exp(-x'x)$ . Quant à la matrice facteur d'échelle, elle consiste à redimensionner les abscisses de la quadrature afin de mieux couvrir l'ensemble des valeurs des effets aléatoires pour lesquelles la valeur de l'intégrande est importante. La matrice facteur d'échelle  $K$  choisie vérifie  $K'K = H$  avec  $H$  inverse du hessien de  $-\frac{1}{2} \log\{f(x) \cdot \exp(-x'x)\}$ . Cependant, il réside un problème dans le choix de cette matrice facteur d'échelle en multivarié. En effet, plusieurs décompositions de la hessienne sont possibles :  $K'K = H$  possède une infinité de solutions de la forme  $OD^{1/2}P^{-1}$  avec  $O$  matrice orthogonale quelconque,  $D$  et  $P$  les matrices respectivement diagonales et orthogonales obtenues par diagonalisation de  $H$ . Dans cet article, deux décompositions ont été choisies : décomposition par Cholesky, de par son utilisation plus courante dans ce type de décomposition (Pinheiro et Bates (1995); Procédure NLMIXED de SAS) et décomposition par diagonalisation de  $H$  dont la forme est plus proche de celle de l'ensemble des solutions. À partir de la matrice facteur d'échelle obtenue par diagonalisation (qui se révélera produire de meilleurs résultats), nous avons mis en place une procédure d'optimisation.

Cette optimisation a pour point de départ la matrice  $K_0 = D^{1/2}P^{-1}$ . Le domaine de recherche est l'ensemble des matrices  $K$  de la forme  $MK_0$ ,  $M$  diagonale, et le critère d'optimisation est la différence entre la quadrature d'ordre  $r$  et la quadrature d'ordre  $r+2$ , approximation de l'erreur d'estimation reposant selon Moslard (1972) sur la formule :  $|\text{intégrale} - \text{quadrature}_r| \approx |\text{quadrature}_r - \text{quadrature}_{r+2}|$ . Cette recherche est effectuée à chaque approximation de l'intégrale.

Plusieurs niveaux pour cette optimisation sont alors possibles : si l'on choisit d'appliquer un même facteur sur chaque effet aléatoire, on peut prendre  $M$  scalaire et procéder ainsi à une optimisation de dimension 1, cela repose sur un ajustement identique dans chaque direction de l'espace des effets aléatoires. En revanche, si l'on veut un facteur différent sur chaque effet aléatoire, on prend alors  $M$  diagonale et on procède à une optimisation de dimension  $n$ . On a alors un ajustement plus ou moins important dans chaque dimension.

En résumé, ces choix de matrice facteur d'échelle vont donner lieu à quatre méthodes. On aura donc la première reposant sur la décomposition par Cholesky, la deuxième sur la décomposition par diagonalisation sans optimisation, la troisième ajoutant à cette dernière décomposition une optimisation en dimension 1 et la dernière avec une optimisation en dimension  $n$ .

Cette méthodologie est illustrée dans le paragraphe suivant.

Le programme d'estimation des paramètres des modèles basé sur les quadratures de Gauss-Hermite est réalisé en Fortran. En ce qui concerne la procédure de calcul de gradient et de la hessienne, cela se fait par la méthode des différences finies selon l'algorithme de Ridders (Press *et al.* (1992)). Les procédures de minimisation et de maximisation reposent sur une méthode de simplex.

### 3. Application

#### 3.1. Molécule et Modèle Pharmacocinétique

Ce travail est basé sur le modèle pharmacocinétique et les valeurs a priori des paramètres de population  $(\theta, \alpha_j, \mu_j, Q_j^2)$  et des caractéristiques individuelles des patients d'une étude clinique de phase II. Cette étude concerne un médicament cardiovasculaire injecté par bolus IV (dose  $D_{iv}$ ) suivi de plusieurs doses sous cutanées  $D_k$  administrées à des dates  $t_k$ . Elle porte sur 448 patients atteints d'un infarctus du myocarde. Plusieurs concentrations du médicament sont testées à des dates diverses ( $m_i$  dates pour le sujet  $i$ ), ce qui donne 788 observations non équilibrées. De plus, nous aurons deux effets aléatoires :  $n = 2$

Le modèle utilisé est un modèle dit ouvert à un compartiment de distribution, dont les paramètres sont la clairance ( $CL$ ), le volume ( $V$ ) et la constante d'absorption ( $Ka$ ). On note  $Ke = CL/V$ . On exprime alors la concentration du médicament dans l'organisme à la date  $t$  par :

$$c(t) = \frac{D_{iv}}{V} \exp(-Ke \cdot t) + \sum_{t_k < t} \left\{ \frac{D_k}{V} \frac{Ka}{(Ka - Ke)} [\exp(-Ke(t - t_k)) - \exp(-Ka(t - t_k))] \right\}$$

Un modèle complémentaire exprime la dépendance de la clairance vis à vis de deux covariables ( $Z_1$  le poids du patient et  $Z_2$  la clairance de la créatinine) et introduit deux effets aléatoires. Ce modèle est le suivant :

$$\begin{aligned} CL &= (\theta_1 + \theta_4 \cdot Z_1 + \theta_5 \cdot Z_2) \cdot \exp(\eta^{(1)}) \\ V &= \theta_2 \cdot \exp(\eta^{(2)}) \\ Ka &= \theta_3 \end{aligned}$$

Les observations sont constituées des mesures de concentrations :  $y_{ij} = c(t_{ij}) + \varepsilon_{ij}$  où  $\varepsilon_{ij}$  reflète l'erreur de mesure. Pour l'ensemble des observations réalisées chez le patient  $i$ , on a  $y_i = c(t_i) + \varepsilon_i$  avec  $y_i \in R^{m_i}$ . La densité de probabilité de  $y_i$  conditionnellement au paramètre aléatoire individuel  $\eta_i$ , notée  $g_i$  (cf. §2.1) résulte de celle de  $\varepsilon_i$ . Ainsi, avec  $\varepsilon_i \sim N(0, \sum_i' \sum_i)$ , et  $\sum_i' \sum_i = \sigma^2 I_{m_i}$ ,  $\sigma^2$  étant dénommée la variance résiduelle, on a :

$$g_i(y_i, \theta, Z_i, \eta_i) = (2\pi)^{-m_i/2} \cdot \left| \sum_i' \sum_i \right|^{-1/2} \cdot \exp \left[ -\frac{1}{2} (y_i - c(t_i))' \cdot \left( \sum_i' \sum_i \right)^{-1} \cdot (y_i - c(t_i)) \right]$$

### 3.2. Simulations

Afin de tester l'aptitude de cette approche méthodologique dans l'estimation des paramètres des modèles non linéaires à effets mixtes et la détection de mélanges au niveau de la distribution des effets aléatoires représentant dans notre application une détection de sous-populations hétérogènes, nous avons procédé à une série de simulations.

Les conditions ont été les suivantes :

- Le nombre de patients, leurs covariables et leurs dates d'observations sont identiques à celles du jeu de données originel.
- Les valeurs théoriques des paramètres fixes sont les valeurs estimées sur le jeu de données originel en utilisant la méthode très couramment utilisée en pharmacocinétique développée par Beal et Sheiner (1992) : NONMEM. Celle-ci repose sur une linéarisation du modèle cinétique par rapport aux paramètres autour de leurs valeurs moyennes.
- La distribution de  $\eta^{(2)}$  est celle estimée sur le jeu de données originel.
- La distribution de  $\eta^{(1)}$  est simulée sous forme d'un mélange de deux lois normales ( $p = 2$ ) afin de simuler deux sous-populations : l'une avec de basses clairances et la seconde avec de fortes clairances. Les constituantes du mélange choisi préservent les moyenne et variance de  $\eta^{(1)}$  estimées sur le jeu de données originel. Ce mélange est présenté dans la figure 1.
- Les effets aléatoires  $\eta^{(1)}$  et  $\eta^{(2)}$  sont non corrélés.

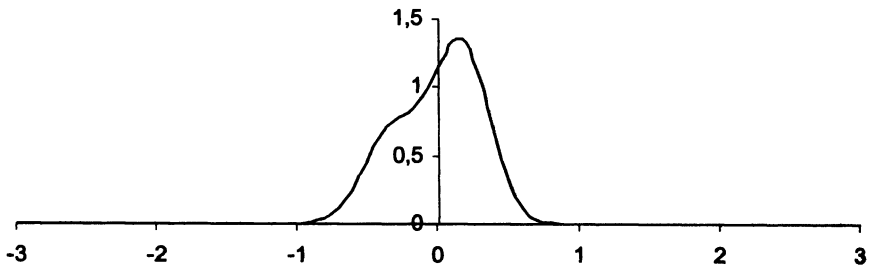


FIG. 1. - Distribution de  $\eta^{(1)}$  :  $\alpha_1 = \frac{1}{3}$ ,  $\alpha_2 = \frac{2}{3}$ ,  $\mu_1^{(1)} = -0.32$ ,  $\mu_2^{(1)} = 0.16$ ,  $(Q_1^{(1)})^2 = (Q_2^{(1)})^2 = 0.04$

Le travail a été effectué en parallèle avec des quadratures d'ordre 3 et 9 pour chaque méthode (ces méthodes étant présentées dans la table 1) afin de mieux apprécier l'impact de chaque ajustement sur la qualité d'approximation.

Plusieurs répétitions ont été réalisées pour chaque méthode. Les résultats de ces différentes méthodes seront comparés entre eux mais également avec les résultats obtenus par NONMEM.

## ADAPTATION DES QUADRATURES DE GAUSS-HERMITE

TABLE 1. – Ajustements utilisés pour chaque méthode de Gauss-Hermite (Codé en O=Oui/N=Non).

	QUADRATURE			
	Centrage	Facteur d'échelle Cholesky	Facteur d'échelle Diagonalisation	Optimisation
Ghbrute	N	N	N	N
Ghcent	O	N	N	N
Ghcho	O	O	N	N
Ghdiag	O	N	O	N
Ghopt	O	N	O	O

### 3.3. Résultats

En ce qui concerne la phase d'estimation, les effets aléatoires ont été supposés non corrélés et ont été estimés sous la forme d'un mélange de deux composantes.

Nous avons fixé le nombre de répétitions en considérant la stabilité des résultats obtenus. 30 répétitions ont été jugées nécessaires. La table 2 en annexe présente donc les résultats d'estimations des paramètres de population avec les moyennes, écarts types et carrés des erreurs quadratiques moyennes sur les 30 répétitions pour chaque méthode.

Les résultats de la table 2 (Annexe) montrent que toutes les méthodes donnent des estimations plus précises que NONMEM. Par ailleurs, à l'ordre 3, le centrage seul se révèle insuffisant voire plus mauvais que Ghbrute. Cela est probablement dû à la sensibilité de l'approximation sans facteur d'échelle : si la méthode est très sensible au point calculé sans facteur d'échelle, le fait de simplement centrer les nœuds peut à la suite de problèmes numériques, les éloigner des valeurs annulant le gradient de  $f(x) \cdot \exp(-x'x)$ . En revanche, la combinaison du centrage et du facteur d'échelle améliore la précision de tous les paramètres, alors que l'augmentation de l'ordre, elle, améliore essentiellement l'estimation de la distribution des effets aléatoires, la variance résiduelle et la détection d'hétérogénéités.

Un problème subsiste au niveau de la variance résiduelle qui n'est pas toujours très bien estimée, même après optimisation. Cependant, l'estimation de la distribution des effets aléatoires s'avère très satisfaisante dès le centrage et le facteur d'échelle. On retrouve notamment de bonnes estimations de variances équivalentes entre les deux composantes.

Si l'on se focalise sur l'effet du choix du facteur d'échelle, la première chose à souligner est le fait que la matrice facteur d'échelle obtenue par diagonalisation donne effectivement des résultats meilleurs que celle obtenue par Cholesky avec laquelle il est difficile d'estimer le paramètre d'hétérogénéité.



Par ailleurs, l'optimisation en dimension 1 d'ordre 3 est globalement équivalente à Ghdiag (diagonalisation) d'ordre 9 pour un temps de calcul voisin. Elle sera d'ailleurs comparativement plus efficace pour un plus grand nombre d'effets aléatoires (augmentation du temps de calcul moins importante selon des estimations de temps basées sur la construction du programme, des simulations avec un plus grand nombre d'effets aléatoires n'ayant pas encore été réalisées). L'optimisation en dimension 2 est la méthode qui donne les meilleures estimations avec les meilleures précisions parmi les méthodes d'ordre 3. Ses résultats sont globalement meilleurs que ceux de Ghdiag ordre 9 avec cependant une moins bonne estimation de la variance résiduelle.

#### 4. Conclusion

En adaptant les quadratures de Gauss-Hermite par centrage et en optimisant, un petit ordre suffit pour obtenir de bonnes estimations et de bonnes précisions. Cependant, il existe un réel problème au niveau du choix de la matrice facteur d'échelle car, selon ce choix, les résultats sont assez différents. Parmi les deux facteurs d'échelle proposés, nous avons retenu celui obtenu par diagonalisation pour ensuite procéder à une optimisation. L'intérêt numérique de cette approche réside dans le fait qu'elle permet un degré d'approximation ajustable selon l'ordre et le niveau de l'optimisation. Il reste ensuite à faire un compromis entre le temps d'implémentation et la qualité d'approximation. Une stratégie d'analyse est alors proposée : dans le cas où le modèle posséderait un ou deux effets aléatoires, nous conseillons un ordre élevé sans optimisation avec centrage et facteur d'échelle. Au-delà de deux effets aléatoires, il semble plus judicieux d'utiliser un ordre plus faible avec optimisation : le temps de calcul comparatif, évalué de façon prospective devrait évoluer en faveur des méthodes contenant des optimisations. Ces remarques nécessitent une confirmation par des simulations additionnelles. L'intérêt de cette approche dans son application en pharmacocinétique repose certes sur la qualité d'estimation des paramètres du modèle, mais essentiellement sur son aptitude à prendre en compte la présence de sous-populations grâce à l'utilisation des mélanges de lois normales en nombre plus ou moins important au niveau de la distribution des effets aléatoires. L'inclusion d'une procédure statistique de détermination du nombre de composantes détectées est en cours d'étude. Cette méthode voit donc son application dans l'adaptation individuelle des posologies selon les mécanismes physiologiques, physiopathologiques mais également les caractéristiques génomiques de chaque patient.

## 5. Références

- BEAL S.L., SHEINER L.B. (1992), *NONMEM user's guides*.
- LIU Q., PIERCE D.A. (1994), A note on Gauss-Hermite quadrature, *Biometrika*, 81, 624-629.
- MOSLARD R. (1972), *Estimation de l'erreur dans les quadratures numériques de Gauss*, Arcueil, Centre de Calcul Scientifique de l'Armement.
- PINHEIRO J.C., BATES D.M. (1995), Approximations to the loglikelihood function in the nonlinear mixed effects model, *Journal of computational and graphical statistics*, 4, 12-35.
- PRESS W.H. *et al.* (1992), *Numerical Recipes in Fortran 77 : The Art of Scientific Computing*, Cambridge University Press.
- STROUD A.H., SECREST D. (1966), *Gaussian Quadrature Formulas*, Englewood Cliffs NJ, Prentice-Hall.

