

PLANIFICATION D'EXPÉRIENCE OPTIMALE ET ÉQUATIONS DIFFÉRENTIELLES : RÉVISION DES BASES ET MÉTHODOLOGIE

Julien Cornebise⁽¹⁾ & Bruno Boulanger⁽²⁾

(1) *LSTA - Université Paris VI, 175 rue du Chevaleret, Boîte 158, 75013 Paris, France, cornebise@ccr.jussieu.fr*

(2) *European Early Phase Statistics, Lilly Services S.A., Rue Granbonpré, 11, B-1348 Mont-Saint-Guibert, Belgique, boulanger-bruno_e@lilly.com*

Mots clés: Biopharmacie, Planification d'expérience, Équations différentielles, Approche individuelle, Approche population.

Abstract: This presentation is aiming at researchers, practitioners, and PhD students, discovering the field of nonlinear optimal design of experiments, or willing to review the essential basics. Focusing on methodology, it mentions more particularly the way usual theory can be applied to models based on ordinary or delayed differential equations systems without known closed-form solution. After having described most current models in the fields of pharmacokinetics, pharmacodynamic, bio-processes, and having summarized key ideas of optimal design of experiments, we will see that there is no need for anything more complicated than adding a numerical approximation layer on top of classical tools: numerical differentiation, numerical integration of differential equations systems, ...

Résumé : Cette présentation est destinée aux chercheurs, praticiens, et doctorants, découvrant le domaine des plans d'expériences optimaux non-linéaires, ou souhaitant en revoir les bases essentielles. Orientée sur la méthodologie, elle mentionne plus particulièrement la façon dont la théorie usuelle peut être appliquée aux modèles basés sur des systèmes d'équations différentielles ordinaires ou retardées sans solution connue sous forme analytique. Après avoir détaillé les modèles les plus souvent rencontrés dans les domaines d'applications tels que la pharmacocinétique, pharmacodynamique, bio-processus, et résumé les idées-clés de la planification d'expérience optimale, nous verrons qu'il s'agit simplement de superposer aux outils classiques une couche d'approximation numérique : différentiation numérique, intégration numérique du système d'équations différentielles, ...

1. Modélisation

Nous nous intéressons en premier lieu aux modèles employés dans les domaines de la biopharmacie, pharmacocinétique, pharmacodynamique, et qui sont très semblables à ceux présents dans de nombreuses autres disciplines telles la chimio-cinétique ou l'étude des bioprocessus. De tels modèles comportent typiquement :

- une partie déterministe, dans notre cas un système d'équations différentielles ordinaires ou retardées provenant d'un modèle compartimental, dont un ou plusieurs compartiments sont observés, pas nécessairement la totalité,
- une partie aléatoire représentant la fluctuation des mesures, variabilité intra-individu,
- et enfin une éventuelle partie aléatoire liée à la population, variabilité inter-individus.

Nous avons donc la forme générale suivante :

$$\begin{cases} \frac{\partial x}{\partial t} &= f(x, \theta, t), \\ \eta &= h(x, \theta, t), \\ x(t_0) &= x_0 \end{cases} \quad (1)$$

où les fonctions f et h sont connues et quelconques (en supposant les conditions de régularité adéquates), $t \in T$ est la variable indépendante (le temps, dans la plupart des cas, avec $T = \mathbb{R}^+$), x est le vecteur des e variables d'états, $\theta \in \mathbb{R}^p$ est le vecteur des paramètres (constant par rapport au temps) du modèle, et η est le vecteur de sortie d -dimensionnel mesuré, fonction des variables d'états. Un cas courant est celui où η est un projecteur canonique, et où seuls certains compartiments du modèle sont observés, à des temps donnés.

On observe, pour $i = 1, \dots, n$, $y_i = \eta(t_i, \theta^*) + \varepsilon_i$ où t_i sont les instants d'échantillonnage, θ^* est la vraie valeur des paramètres (inconnue), et ε_i sont les erreurs de mesure. La fonction $\eta(t, \theta)$ est, dans notre cas, non-linéaire en les paramètres, rendant impossible l'utilisation des outils les plus classiques. Diverses hypothèses peuvent être posées sur les erreurs d'observations ε_i , les plus classiques étant choisies parmi les suivantes : erreurs identiquement distribuées, centrées, décorrélées, indépendantes, de variance constante par rapport au temps (cas dit homoscédastique) ou dépendante du temps (cas hétéroscédastique). Ce dernier cas rajoute un niveau de complexité au modèle, en posant par exemple $\mathbb{V}[y_i] = \sigma^2 \times \{g(\eta(t_i, \theta), \beta)\}^2$ où β est le vecteur des paramètres du modèle d'erreur et g une fonction connue. Des choix courants sont $g(\mu_i, \beta) = \beta\mu_i$, $g(\mu_i, \beta) = \exp(\mu_i\beta)$, ou $g(\mu_i, z_i, \beta) = \sigma_1 + \sigma_2\mu_i$ avec $\beta = (\sigma_1, \sigma_2)$. Ces cas particuliers propres au contexte univarié ($d = 1$) se généralisent aisément au cas multivarié, avec toutefois une explosion du nombre de paramètres dans les matrices de variance-covariance.

La prise en compte de l'aspect population s'obtient avec une couche supplémentaire de modélisation. Dans le cas de N individus distincts d'où proviennent les observations, pour $j = 1, \dots, N$, chaque coordonnée du vecteur θ_j des paramètres de l'individu j sera vu comme fonction d'un paramètre θ , dit effet fixe, commun à l'ensemble de la population, et d'un paramètre dit effet aléatoire, b_{lj} , $l = 1, \dots, p$. Souvent, les b_{lj} indépendants et identiquement distribués selon une loi normale centrée de matrice de variance-covariance diagonale d'éléments $\omega_1, \dots, \omega_p$. Les deux approches les plus employées sont $\theta_{lj} = \theta + b_{lj}$ (modèle additif), et $\theta_{lj} = \theta \exp(b_{lj})$ (modèle multiplicatif). Cette modélisation de l'approche population, commune à beaucoup de travaux, peut en particulier être retrouvée sous une forme détaillée dans Gueorguieva et al. (2005), Retout

et al. (2001,2002,2003a,2003b), ou Davidian et Giltinan (1997), qui s'intéresse tout particulièrement à l'estimation dans ce cadre.

2. Planification d'expérience

La question de planifier les expériences, c'est à dire de choisir les instants t_i d'échantillonnage, afin d'obtenir les "meilleures" estimations possibles des paramètres ou la "meilleure" discrimination entre plusieurs modèles n'est pas récente. L'un des articles fondateurs est celui de Box et Lucas (1959). La théorie a depuis été largement étudiée, particulièrement dans Silvey (1980) qui parcourt en grand détail et en moins d'une centaine de pages toutes les fondations, et dans Atkinson et Donev (1992), d'un abord plus facile car plus intuitif mais moins profond. Cette littérature ne mentionne pas explicitement les cas où la réponse est solution d'une équation différentielle, mais nous verrons en section 3 qu'elle reste valide. L'idée essentielle repose sur le lien entre la Matrice d'Information de Fisher (abrégée FIM), ou plus exactement la FIM espérée et la qualité des estimateurs.

Dans l'approche individuelle (en délaissant donc l'aspect population), la FIM d'une expérience observant y_1, \dots, y_n est définie par $FIM(\theta) = \mathbb{E} \left[-\frac{\partial^2 \log p(y_1, \dots, y_n, \theta)}{\partial \theta^2} \right]$ où $p(y_1, \dots, y_n, \theta)$ est la vraisemblance des observations pour le paramètre θ . Sous hypothèse d'indépendance des observations, $p(y_1, \dots, y_n, \theta) = \prod_{i=1}^n p(y_i, \theta)$, la FIM se simplifie en $FIM(\theta) = \sum_{i=1}^n \mathbb{E} \left[-\frac{\partial^2 \log p(y_i, \theta)}{\partial \theta^2} \right]$. Dans le cas hétéroscédastique normal avec variance unitaire des erreurs, on obtient la simplification supplémentaire $FIM(\theta) = \sum_{i=1}^n \mathbb{E} \left[\left(\frac{\partial \eta(t_i, \theta)}{\partial \theta} \right)^T \left(\frac{\partial \eta(t_i, \theta)}{\partial \theta} \right) \right]$.

Une autre expression souvent rencontrée est celle du cas linéaire en les paramètres, où $\eta(t_i, \theta) = \gamma(t_i)^T \theta$, où γ est un vecteur de fonctions du temps et où les y_i sont i.i.d. de loi normale et de variance unitaire. La FIM devient alors $FIM(\theta) = FIM = \sum_{i=1}^n \gamma(t_i)^T \gamma(t_i)$. Ces diverses expressions doivent être gardées à l'esprit afin de mettre en parallèle les différents articles sur le sujet.

Dans l'approche population, le principe reste inchangé, seule l'expression de la FIM diffère, puisque la vraisemblance doit alors prendre en compte les valeurs des paramètres fixes et aléatoires. Dans ces cas, il est très difficile d'obtenir une expression explicite pour cette dernière. Deux possibilités : utiliser une approche type Monte-Carlo, en simulant différentes populations et en effectuant une moyenne sur la FIM prise en ces diverses populations, ce qui donne une approximation numérique d'une forme exacte de la FIM, ou utiliser une approximation de Taylor pour obtenir un calcul numérique exact d'une forme approchée de la FIM. Cette dernière solution est la moins coûteuse en temps de calcul, et est au coeur des travaux de Retout et al. (2001,2002,2003a,2003b).

Le lien entre la FIM et la qualité de l'expérience peut-être vu de différentes façons : la principale est au travers de la borne de Cramer-Rao, qui minore la variance des estimateurs sans biais des paramètres par l'inverse de la FIM. Le lien peut aussi être intuité en observant la forme des matrices de variance covariance des estimateurs donnés par exemple dans Seber et Wild (1989) ou Davidian et Giltinan (1997). Dans le domaine des bio-

processus, une autre approche consiste à maximiser l’“identifiabilité pratique” du modèle, ce qui permet de retrouver la FIM.

Il est important de remarquer que, hors du cas particulier d’une réponse fonction linéaire en les paramètres, la FIM dépend non seulement des instants d’échantillonnage, mais aussi de la valeur vraie du paramètre θ , qui est précisément inconnue et que l’on cherche à estimer ! C’est pourquoi les plans d’expériences sont appelés localement optimaux : ils sont conçus en pré-supposant une valeur a priori du paramètre, basée sur la connaissance de l’expert du domaine, et ne sont donc optimaux que dans un voisinage de cette valeur. Des simulations permettent d’avoir une vision de la taille de ce voisinage. Une approche bayésienne mettant une loi a priori sur les paramètres (au lieu d’une simple valeur, loi mettant toute la masse en un point) permet plus de souplesse, mais ne sera pas abordée ici. Nous renvoyons à l’article de Chaloner et Verdinelli (1995). Un protocole d’expérimentation séquentiel, où les expériences sont planifiées en fonction des résultats précédents, sur un mode itératif, peut être employé mais est potentiellement onéreux.

Nous allons chercher à minimiser la borne de Cramer-Rao, et donc à maximiser la FIM. Le sens auquel nous entendons expérience “optimale” peut désormais être défini comme le maximum d’une fonction de la FIM associée à cette expérience : différentes fonctions, ou critères, sont couramment employées. Le plus répandu est le critère de D-optimalité, où l’on maximise le déterminant de la FIM, ce qui revient à minimiser le volume des régions de confiance, comme expliqué par exemple dans Atkinson et Donev (1992). La théorie des plans d’expériences, permettant d’aboutir à des résultats essentiels comme l’équivalence entre certains critères, une majoration du nombre d’instants d’échantillonnage, repose sur la vue d’une expérience comme une mesure de probabilité discrète sur l’espace des instants d’échantillonnage, dont les poids sont des fractions acceptant le nombre d’expériences n pour dénominateur. En utilisant ensuite les plans continus (appelés plans approchés par Silvey (1980)), pour lesquels la mesure de probabilité devient quelconque, de nombreux résultats sont démontrés à l’aide d’outils d’analyse comme la dérivée au sens de Fréchet, ou d’outils d’algèbre comme le théorème de Caratheodory sur l’enveloppe convexe d’un ensemble de points, qui permet de revenir à des mesures optimales à support discret mais avec des poids positifs réels. Dans une dernière étape, le retour à des mesures du premier type ne permet d’obtenir que des plans qui approchent l’optimalité, sans forcément l’atteindre. Ces détours donnent une justification théorique au résumé plus simple de la planification : maximiser une fonction des instants d’échantillonnage, ici le déterminant de la FIM associée. Ceci est effectué à l’aide d’algorithmes répandus de recherche d’optima de fonctions multivariées.

3. Approche numérique et Optimisation

Lors que la fonction η est solution d’un système d’équations différentielles, il est courant qu’elle ne soit pas connue sous forme analytique. Cela ne change rien à la théorie exposée plus haut, mais requiert l’ajout d’une couche supplémentaire dans la méthodologie : intégration et dérivation numériques. Une référence incontournable pour les algorithmes

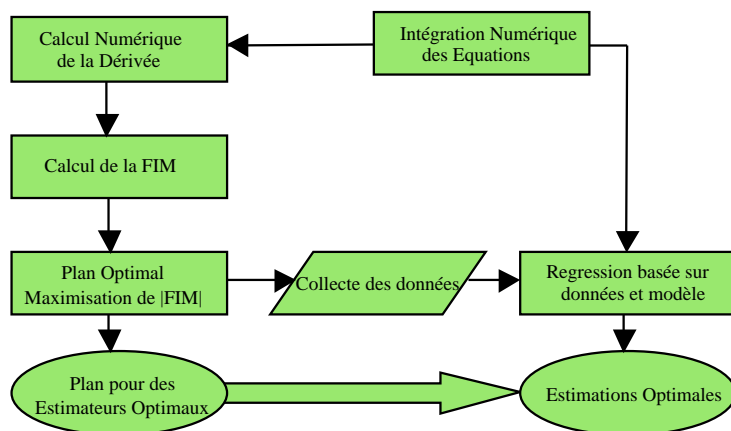


Figure 1: Mise en commun des étapes : résumé de la méthodologie

numériques est Press et al. (1992). Il est en effet possible d’approcher la réponse pour une valeur fixée des paramètres, sur un intervalle de temps choisi par l’utilisateur, à l’aide de méthodes d’intégration telles que Runge-Kutta et ses variantes. De telles routines de résolution numérique de systèmes d’équations différentielles sont souvent disponibles dans les bibliothèques scientifiques. Le solveur d’équations de Matlab a par exemple été particulièrement détaillé dans Samphine et al. (1997).

La différentiation numérique s’obtient quant à elle en approximant la dérivée par le taux de variation $\frac{\partial \eta}{\partial \theta}(t, \theta) \approx \frac{\eta(t, \theta+h) - \eta(t, \theta)}{h}$ où h est un pas choisi arbitrairement petit. Plusieurs méthodes existent (e.g. différence avant, arrière, centrale), mais il est important de garder à l’esprit qu’un soin certain doit être apporté aux considérations numériques. Un pas h choisi trop grand entraînera une mauvaise approximation, mais un pas trop petit posera des problèmes de précision. Un autre exemple de dangers de cette méthode se pose pour les grandes valeurs de θ : un h trop petit sera “absorbé”, suite à la précision finie de la représentation à virgule flottante des réels. C’est pourquoi certains programmes utilisent $\frac{\partial \eta}{\partial \theta}(t, \theta) \approx \frac{\eta(t, (1+h_1)\theta+h_2) - \eta(t, (1-h_1)\theta-h_2)}{2(h_1\theta+h_2)}$ où par exemple $h_1 = 10^{-4}$ et $h_2 = 10^{-6}$.

Ce souci des erreurs numériques est aussi présent dans l’intégration, où il faut être très attentif à de petites variations des résultats : les erreurs sur les calculs des valeurs de $\eta(t_i, \theta)$ se cumulent avec celles de la dérivée, et enfin avec la prise du déterminant de la FIM, opération très sensible à de faibles variations. Imposer au solveur une tolérance d’erreur moins large que celle par défaut est indispensable.

Toujours dans l’optique de minimiser les erreurs numériques, il est préférable d’utiliser un algorithme d’optimisation qui ne nécessite pas d’avoir recours à la dérivée de la fonction objectif, au contraire des méthodes de gradient. Ainsi, le Simplex de Nelder-Mead, l’algorithme d’échange de Fedorov-Wynn, ou le recuit simulé sont des choix pertinents.

Enfin, le solveur permet de calculer, en une seule passe, les réponses pour une valeur

donnée des paramètres et à plusieurs instants d'échantillonnage. Ceci permet d'accélérer grandement l'exécution : le calcul des dérivées partielles $\frac{\partial \eta(t_i, \theta)}{\partial \theta_j}$ où $i = 1, \dots, n$ et $j = 1, \dots, p$ peut se faire en une seule fois pour tous les instants et pour j choisi, en effectuant la boucle externe sur les j , et en remplaçant la boucle interne sur les i par un seul appel au solveur d'équations différentielles.

Pour conclure, l'imbrication des différentes étapes présentées ici est résumé dans la Figure 1.

Bibliographie

- [1] Atkinson, A.C., et Donev, A.N (1992) *Optimum Experimental Designs*, Oxford Science Publications.
- [2] Box, G.E.P. et Lucas, H.L. (1959) Design of experiments in nonlinear situations. *Biometrika*, 46:77-90.
- [3] Davidian, M. et Giltinan, D.M. (1997), *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall.
- [4] Gueorguieva, I., Aarons, L., Ogungbenro, K., Jorga, K.M., Rodgers, T., and Rowland, M. (2005) Optimal design for multivariate response pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamic*, à paraître.
- [5] Le Bailly de Tillegem, C., Beck, B., Boulanger, B., et Govaerts, B. (2005) A fast exchange algorithm for designing focused libraries in lead optimization. *Journal of Chemical Information and Modeling*, 45:758-767.
- [6] Press, W.H., Teukolsky, S.A., Vetterling, W.T., et Flannery, B.P. (1992) *Numerical Recipes in C: the Art of Scientific Programming, Second Edition*, Cambridge University Press.
- [7] Retout, S., Duffull, S. et Mentré, F. (2001) Development and implementation of the population fisher information matrix for the evaluation of population pharmacokinetic designs. *Computer Methods and Programs in Biomedicine*, 65:141-151, 2001.
- [8] Retout, S., et Mentré, F. (2003a) Further developments of the fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, 13(2):209-227.
- [9] Retout, S., et Mentré, F. (2003b) Optimization of individual and population designs using Splus. *Journal of Pharmacokinetics and Pharmacodynamics*, 30(6):417-443.
- [10] Retout, S., Mentré, F., et Bruno, R. (2002) Fisher information matrix for nonlinear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine*, 21:2623-2639.
- [11] Seber, G.A.F et Wild, C.J. (1989) *Nonlinear Regression*, Wiley, New York.
- [12] Shampine, L.F. et Reichelt, M.W. (1997) The Matlab ODE suite. *SIAM Journal of Scientific Computing*, 18(1):1-22.
- [13] Silvey, S.D. (1980), *Optimal Design, An Introduction to the Theory for Parameter Estimation*, Chapman & Hall.