

Plans d'expériences D-optimaux pour modèles non linéaires

S.Hattou ^a

^a Sanofi, centre de recherche de Montpellier
371 rue du Professeur Joseph Blayac
34184 Montpellier cedex 04
stephane.hattou@sanofi.com

Mots clefs : Matrice jacobienne, Plan D-optimal, Modèle non linéaire, Algorithme de Fedorov.

Cette étude traite de la construction de plans d'expériences D-optimaux dans le cas de modèles non linéaires. Sous R, il existe des packages spécifiques pour certains types de modèles non linéaires, comme le package « DoseFindings » pour les modèles dose-réponse. Le package « AlgDesign » ne traite quant à lui que les modèles polynômiaux linéaires. Pour un modèle non linéaire quelconque (logistique, Michaelis-Menten, composé d'un système d'équations différentielles, ...), **il n'existe pas à ce jour de fonction disponible dans R permettant de réaliser ce type de calcul**. Dans cette étude, des fonctions indépendantes ont été programmées afin :

1. D'une part générer, pour un modèle non linéaire donné, la matrice jacobienne résultante des dérivées partielles par rapport aux coefficients du modèle.
2. D'autre part calculer, à partir d'une matrice de variance-covariance quelconque et d'un ensemble donné de points candidats, un plan D-optimal en utilisant l'algorithme de Fedorov.

Ce dernier point a fait l'objet d'une attention particulière lors de la programmation, afin de pouvoir gérer les problèmes numériques induits par l'inversion de la matrice hessienne.

Un modèle $Y = \eta(X, \theta^*) + \varepsilon$ est dit non linéaire par rapport aux paramètres lorsque au moins une dérivée première de la fonction du modèle par rapport aux paramètres dépend d'au moins un de ces paramètres. Supposons que l'on ait une première estimation θ ($a_0^0, a_1^0 \dots$) des paramètres. Le développement limité de Y au voisinage de cette estimation est alors donné par $Y_i - \eta(x_i) = \frac{\partial \eta(x_i)}{\partial \theta_0} \Delta \theta_0 + \dots + \frac{\partial \eta(x_i)}{\partial \theta_n} \Delta \theta_n + \varepsilon_i$ qui correspond au problème de régression linéaire $Y - \eta(X) = J(\theta) \Delta \theta + \varepsilon$, $J(\theta)$ étant la matrice jacobienne telle que $J(\theta) = \frac{\partial \eta(x_i, \theta)}{\partial \theta_j}$.

L'application des formules de la régression linéaire conduit alors à $\Delta \theta = [J(\theta)^T J(\theta)]^{-1} [J(\theta)^T \{Y - \eta(X)\}]$. Connaissant le vecteur des corrections $\Delta \theta$, il est possible de calculer de nouvelles estimations $a_0^1, a_1^1 \dots$ des paramètres. Le processus est répété jusqu'à convergence des estimations.

Un estimateur est dit non biaisé si l'espérance mathématique sur les observations possibles de θ conduisent à θ^* . Cet estimateur non biaisé est dit efficace si sa covariance atteint la borne inférieure de l'inégalité de Cramér-Rao : $V_\theta = E_{Y/\theta} [(\theta_{(Y)} - \theta^*) \times ((\theta_{(Y)} - \theta^*)^T)] \geq M_F^{-1}(\theta^*)$. Le premier terme de cette inégalité présente la matrice de covariance des paramètres. Le deuxième terme présente la matrice d'information de Fisher, paramètre fondamental lors de la mise en œuvre d'une stratégie expérimentale optimale. La matrice de covariance des paramètres traduit la façon dont les estimées sont dispersées autour des vrais paramètres. On souhaite que cette dispersion soit aussi faible que possible. Si l'on admet la validité de l'approximation du premier ordre pour le modèle au voisinage de la solution, la matrice de covariance des

paramètres peut être calculée par : $M_F^{-1}(\theta^*) = [J(\theta)^T(V_\theta)^{-1} J(\theta)]^{-1}$, V_θ étant la matrice de covariance des mesures. Le critère D-optimal pour la planification d'expériences en vue de l'estimation paramétrique est une fonctionnelle de la matrice d'information de Fisher $M_F(\theta^*)$. Minimiser ce critère revient à maximiser $\det[M_F(\theta^*)]$. L'optimisation de ce critère conduit à minimiser le volume des ellipsoïdes de confiance asymptotiques de θ^* . Dans le cas d'un protocole D-Optimal, $\text{cov}(\theta^*)$ est donc proportionnel au volume contenu à l'intérieur d'une ellipsoïde de confiance spécifique de θ^* dans l'espace des paramètres. Une stratégie appropriée pour calculer les x_i à utiliser lors de la régression est alors de minimiser $\{\text{cov}(\theta^*)\}$ sur des valeurs de x_i appartenant à un intervalle donné. Si les erreurs de mesure ε_{ij} sont gaussiennes, et que les erreurs sont pondérées par l'inverse de la covariance soit V_θ^{-1} , alors $\text{cov}(\theta^*) \approx M_F^{-1}(\theta^*)$.

La procédure de calcul mise en place sous R consiste à :

1. Calculer numériquement la matrice jacobienne $J(\theta)$, pour un modèle non linéaire donné, à l'aide de la fonction *jacobian* du package *numDeriv*. Le modèle non linéaire est défini préalablement sous la forme d'une fonction R.
2. Estimer le nombre et les valeurs optimales des points x_i , à partir d'un ensemble de points candidats donné, qui maximise le déterminant $[J(\theta)^T(V_\theta)^{-1} J(\theta)]$. L'algorithme d'échange de Fedorov a été programmé pour réaliser ces calculs. L'utilisation de la fonction *ginv()*, du package *MASS*, a été préférée à la fonction *solve()*, moins robuste dans le cas d'inversion de matrices singulières. Cet algorithme peut aussi être utilisé dans le cas de modèles linéaires ; il suffit de remplacer la matrice hessienne $[J(\theta)^T J(\theta)]$ par la matrice de variance covariance ($X^T X$) du modèle linéaire à estimer, en entrée de la fonction R construite.

Les fonctions ainsi créés ont été utilisées pour calculer les temps optimaux auxquels il était nécessaire de réaliser des mesures pour calculer avec la meilleure précision les coefficients d'un modèle :

1. Logistique 3 paramètres. Le nombre d'expériences réalisées a été diminué par 2, par rapport à une approche traditionnelle.
2. Constitué d'un système d'équations différentielles, modélisant une suite de réactions chimiques. Les paramètres qui ont été déterminés dans cette étude représentent les constantes cinétiques de chacune des réactions.

Références

- [1] Ogungbenro, K., Graham, G., Gueorguieva, I., Aarons, L., (2005). The use of a modified exchange algorithm to optimize sampling times for population pharmacokinetic experiments. *Computer Methods and Programs in Biomedicine* 80, 115-125
- [2] Atkinson, A.C., Bogacka, B., (2002). Compound and other optimum design for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems* 61, 17-33
- [3] De Aguiar, P.F., Bourguignon, B., Khots, M.S., Massart, D.L., Phan-Tan-Luu, R., (1995). D-optimal designs. *Chemometrics and Intelligent Laboratory Systems* 30, 199-210
- [4] Nguyen, Nam-Ky., Miller, A.J. (1992). A review of some exchange algorithms for constructing discrete D-optimal designs. *Computational Statistics & Data Analysis*, 489-498