Estimation de paramètres pour des modèles d'équations différentielles via la théorie du contrôle

Quentin Clairon [1] & Nicolas Brunel [2]

[1] *Laboratoire de Mathématiques et Modélisation d'Evry. Université d'Evry Val d'Essonne, 23 Bd. de France, 91037 EVRY CEDEX. (quentin.clairon@univ-evry.fr)*

[2] *Laboratoire de Mathématiques et Modélisation d'Evry - UMR CNRS 8071. Université d'Evry Val d'Essonne, 23 Bd. de France, 91037 EVRY CEDEX. / École nationale supérieure d'informatique pour l'industrie et l'entreprise (ENSIIE), 1 sq Résistance, 91025 EVRY CEDEX. (nicolas.brunel@ensiie.fr)*

Résumé. Les équations différentielles ordinaires sont couramment utilisées pour modéliser des processus biologiques (modèles SIR, réseaux de régulation génétiques). Elles se présentent sous la forme générale $\dot{X} = f(t, X, \theta)$ où $X(t)$ est la fonction de régression representánt l'évolution des variables biologiques d'interêt. Néanmoins ces dernières contiennent un paramètre θ potentiellement de grande dimension qui doit être estimé à partir d'observations bruitées et partielles du système étudié.

La définition implicite du modèle à l'échelle de la dynamique et non des observations ainsi que sa grande sensibilité à la valeur des paramètres mettent souvent en échec les méthodes statistiques classiques d'estimation (MLE , NLS etc...) et justifie la recherche de nouvelles méthodes d'estimation. Par ailleurs les méthodes classiques offrent rarement des critères adaptés pour diagnostiquer une éventuelle mauvaise spécification du modèle supposé.

Ici nous proposons une relaxation du modèle à l'échelle de la dynamique en introduisant le nouveau modèle $\dot{X} = f(t, X, \theta) + u(t)$. Nous définirons grâce à cette relaxation un estimateur minimisant à la fois l'écart aux données et au modèle initial. Pour ce faire nous utiliserons un résultat venu de la théorie du contrôle, le principe du maximum de Pontryagin, afin de définir la fonction optimale u et de marginaliser le cout dans l'espace de dimension infinie des fonctions u possibles.

Nous verrons sur des exemples que cette méthode nous permet d'avoir des estimations plus précises que les estimateurs classiques et que le terme de relaxation permet d'explorer la validité du modèle proposé.

Mots-clés. Systèmes dynamiques, Equations différentielles, Problème inverse, Estimation de paramètres

Abstract. Ordinary differential equations (ODE's) are widespread tools to model biological process (SIR model, genetic regulation network). There are under the form $\dot{X} = f(t, X, \theta)$ where $X(t)$ represents the evolution of the biological variables of interest. They rely on parameter θ which are of critical importance in terms of dynamic and need to be estimated directly from noisy and partial observations of the system.

The implicit definition of the regression function at the dynamic scale and not the observation scale as well as its great sensibility to parameter value often leads the classic statistic estimation methods (MLE, NLS) to fail and legitimate the search of alternative to classic approaches. Moreover the classic methods rarely propose misspecification criteria along the parametric estimation.

We present here an estimation method based on a model relaxation at the dynamic scale by introducing the new model $\dot{X} = f(t, X, \theta) + u(t)$. Thanks to this relaxation we define an estimator minimizing both the discrepancy with the data and with the initial model. For doing so we use a result coming from control theory, Pontryagin maximum principle, in order to define the optimal function u and profile the cost in the space of possible functions u.

We will see on examples our method produces more precise estimation than classical estimators and the relaxation terms allows us to explore the validity of the proposed model.

Keywords. Dynamical Systems, Differential Equations, Inverse problem, Parameter estimation.

Introduction

On considère un processus dynamique défini grâce à une équation différentielle ordinaire (EDO) comme solution du problème de Cauchy:

$$\begin{aligned} \dot{x} &= f(t, x, \theta) \\ x(0) &= x_0 \end{aligned} \tag{1}$$

où $x \in \mathbb{R}^d$ et θ appartient à un sous-ensemble de Θ of $\mathbb{R}^p$ et f est un champ de vecteur défini sur $[0, T] \times \mathbb{R}^d \times \Theta$ à valeur dans $\mathbb{R}^d$. On appelle X_{θ,x_0} la solution de (1) pour un paramètre θ et une condition initiale $X_{\theta,x_0}(0) = x_0$ donnés, elle représente l'évolution des variables biologiques d'interêt. Ce formalisme est couramment utilisé pour modéliser des processus dynamiques en physique, chimie, biologie etc (voir [5], [13],[4],[6] par example).

On s'intérèsse ici au problème d'estimation de paramètre, i.e ayant à disposition n observations $(Y_1, \ldots, Y_n)$ bruitées et partielles du système:

$$Y_i = CX^*_{\theta^*, x_0^*}(t_i) + \epsilon_i$$

de la vraie trajectoire $X^*_{\theta^*, x_0^*}$ on veut estimer la valeur du paramètre θ^* (et la condition initiale x_0^* si elle n'est pas connue, dans tout les cas on notera θ même si l'on considère l'ensemble de paramètre (θ, x_0)). Ici $0 = t_1 < t_2 \ldots < t_n = T$ sont les temps d'observations répartis sur $[0, T]$, et C la matrice d'observation de taille $d' \times d$.

En principe cela peut être réalisé en utilisant les estimateurs classiques comme les moindres carrés non-linéaire défini par minimisation du critère:

$$R(\theta) = \sum_{i=1}^{n} \|Y_i - CX_{\theta, x_0}(t_i)\|_2^2$$

Il existe de nombreuses méthodes pour estimer une trajectoire $\hat{Y}(t)$ directement à partir des données en utilisant des méthodes non-paramétriques ([19]). On a ainsi un estimateur de la vraie courbe $t \mapsto CX^*_{\theta^*, x_0^*}(t)$ qui permet de définir un critère sous la forme d'une intégrale et similaire aux moindres carrés non-linéaire:

$$R(\theta) = \int_0^T \left\| \hat{Y}(t) - CX_{\theta, x_0}(t) \right\|_2^2 dt$$

Mais dans le cadre spécifique où la fonction de régression est définie de manière implicite par un modèle d'EDO les méthodes d'estimations classiques comme les moindres carrés où le maximum de vraisemblance conduisent à résoudre des problèmes d'optimisation non-linéaires en très grande dimension. Ces difficultées décritent par Ramsay viennent de l'intégration numérique répétées de L'EDO requise par l'algorithme d'optimisation et plus important de la présence de nombreux minima locaux. Bien que des approches métaheuristiques peuvent être utilisées pour contourner ce dernier problème (voir [14] et [18] pour la présentation et la comparaison de certaines de ces méthodes), elles mettent en lumière le caractère mal posé du problème statistique inverse ([3]) et légitime la recherche de méthodes alternatives aux approches classiques. Ainsi de nouvelles méthodes ont émérgé comme les approches bayésiennes hierarchiques ([10],[16]), ainsi que celles s'appuyant sur des estimateurs non-paramétriques de la courbe afin de regulariser le problème inverse (l'approche par "generalized smoothing" [17],[15] et les estimateurs "two step" [2],[1], [12], [7]). Recemment des méthodes basées sur le calcul variationnel ont été dévelopé [11].

Relaxation de l'EDO et théorie du contrôle

Reste qu'un problème majeur de ces méthodes est que la plupart supposent le modèle comme correct mais fournissent rarement un critère pour déceler une eventuelle mauvaise specification du modèle. C'est un point critique lorsque la dynamique du sytème n'est pas très bien connue ou comprise. La méthode de "generalized smoothing" peut donner des indices sur la présence d'une mauvaise spécification mais elle souffre d'un biais venant de l'utilisation de l'approximation dans une base de dimension finie de la perturbation et de l'utilisation de la dérivée de l'estimateur non-paramétrique. Il faut aussi noter les travaux de Hooker proposant des critères sur la présence ou non de mauvaise spécification ([8]) ainsi que sur leur nature ([9]), mais elles nécéssitent une estimation paramétrique préalable et souffre des même biais que ceux décrit pour l'approche par "Generalized Smoothing".

Dans notre cas nous allons relaxer la contrainte imposée sur la dynamique de la fonction de régression en considerant qu'elle est décrite par l'EDO relaxée (ou perturbée):

$$\dot{x}(t) = f(t, x, \theta) + u(t) \tag{2}$$

on définit $X_{\theta,x_0,u}$ la solution de (2) pour un paramètre θ, une condition initiale x_0 et une fonction u.

De manière similaire à ([8]) la fonction u est un terme de forcing introduite à l'échelle de la dynamique pour representer la distance entre la trajectoire observée et le modèle. Nous voulons minimiser à la fois la distance entre la trajectoire et les données et la norme $\|u\|_{L^2}$ afin de minimiser l'écart avec le modèle initial (1). Ainsi on veut minimiser la fonction de coût:

$$C_\lambda(u, \theta) = \int_0^T \left\| CX_{\theta,x_0,u}(t) - \widehat{Y} \right\|_2^2 dt + \lambda \int_0^T \|u(t)\|_2^2 \, dt \tag{3}$$

Pour chaque θ on introduit le coût marginalisé sur l'espace des perturbations:

$$S_{n,\lambda}(\theta) = \min_{u \in L^2} C_\lambda(u, \theta)$$

et on définit notre estimateur paramétrique comme l'estimateur de vraisemblance marginalisée:

$$\widehat{\theta_n} = \arg\min_{\theta \in \Theta} S_{n,\lambda}(\theta)$$

Le problème semble ici de minimiser (3) dans un espace de dimension infini mais grâce aux résultats issues de la théorie du contrôle, pour un θ donné, le principe du maximum de Pontryagin nous donne une expression connue pour le perturbation optimale $\overline{u}$ minimisant le coût (3) sous la forme:

$$\overline{u} = \frac{1}{2\lambda} \overline{p_\theta}(t)$$

en introduisant $\overline{X_\theta}(t)$ et $\overline{p_\theta}(t)$ solution de l'EDO avec conditions aux bords:

$$\begin{cases} \dot{\overline{X_\theta}}(t) = f(t, \overline{X_\theta}(t), \theta) + \frac{1}{2\lambda} \overline{p_\theta}(t) \\ \dot{\overline{p_\theta}}(t) = -\frac{\partial f}{\partial x}(t, \overline{X_\theta}(t), \theta)^T \overline{p_\theta}(t)(t) + 2C^T \left(C\overline{X_\theta}(t) - \widehat{Y}(t) \right) \\ \left(\overline{X_\theta}(0), \overline{p_\theta}(T) \right) = (x_0, 0) \end{cases}$$

On peut ainsi evaluer la valeur du coût minimal:

$$S_{n,\lambda}(\theta) = \int_0^T \left\| C\overline{X_\theta}(t) - \widehat{Y}(t) \right\|_2^2 dt + \frac{1}{4\lambda} \int_0^T \left\| \overline{p_\theta}(t) \right\|_2^2 dt$$

Par ailleurs, une fois l'estimateur $\widehat{\theta_n}$ obtenu, la valeur de $\widehat{u}$ correspondante:

$$\overline{u} = \frac{1}{2\lambda} \overline{p_{\widehat{\theta_n}}}(t)$$

nous donne sous la forme d'une fonction la distance entre le modèle initial et le modèle relaxé (2). Elle constitue une aide au diagnostic pour la présence d'une éventuelle mauvaise spécification du modèle. Cette fonction est obtenue simultanement avec l'estimation paramétrique et sans approximation dans une base finie. Il est à noter que cette distance est aussi calculée pour les variables non observées, étant donnée que la relaxation est à l'échelle de la dynamique et non des observations.

Nous verrons sur différents modèles linéaires et non-linéaires quelles sont les performances de notre estimateur et nous les comparerons avec les NLS ainsi que l'approche "generalized smoothing". Nous comparerons également les performances de ces estimateurs sur un example de modèle mal-spécifié.

Bibliographie

References

[1] N. J-B. Brunel. Parameter estimation of ode's via nonparametric estimators. *Electronic Journal of Statistics*, 2:1242–1267, 2008.

[2] N. J-B. Brunel and Q. Clairon F. D'Alche-Buc. Parameter estimation of ordinary differential equations with orthogonality conditions. *JASA*, 2014. To Appear.

[3] Hein W Engl, Christoph Flamm, Philipp Kügler, James Lu, Stefan Müller, and Peter Schuster. Inverse problems in systems biology. *Inverse Problems*, 25(12), 2009.

[4] C.P. Fall, E.S. Marland, J.M. Wagner, and J.J. Tyson, editors. *Computational Cell Biology*. Interdisciplinary applied mathematics. Springer, 2002.

[5] R. E. Fuguitt and J.E. Hawkins. Rate of Thermal Isomerization of a-Pinene in the Liquid Phase. *J.A.C.S*, 319(39), 1947.

[6] Albert Goldbeter. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, 1997.

[7] S. Gugushvili and C.A.J. Klaassen. Root-n-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. *Bernoulli*, to appear, 2011.

[8] G. Hooker. Forcing function diagnostics for nonlinear dynamics. *Biometrics*, 65:928–936, 2009.

[9] G. Hooker and S. Ellner. Goodness of fit in nonlinear dynamics: Mis-specified rates or mis-specified states? Technical report, Cornell University, 2013. arXiv:1312.0294.

[10] Y. Huang and H. Wu. A bayesian approach for estimating antiviral efficacy in hiv dynamic models. *Journal of Applied Statistics*, 33:155–174, 2006.

[11] D. Kaschek and J. Timmer. A variational approach to parameter estimation in ordinary differential equations. *BMC Systems Biology*, 6:99, 2012.

[12] H Liang and H. Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 103(484):1570–1583, December 2008.

[13] H.P. Mirsky, A.C. Liu, D.K. Welsh, S.A. Kay, and F.J. Doyle III. A model of the cell-autonomous mammalian circadian clock. *PNAS*, 106(27):11107–11112, July 2009.

[14] C.G. Moles, P. Mendes, and J.R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13:2467–2474, 2003.

[15] Xin Qi and Hongyu Zhao. Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *The Annals of Statistics*, 1:435–481, 2010.

[16] A.E. Raftery and L. Bao. Estimating and projecting trends in hiv/aids generalized epidemics using incremental mixture importance sampling. *Biometrics*, 66:1162–1173, 2010.

[17] J.O. Ramsay, G. Hooker, J. Cao, and D. Campbell. Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society (B)*, 69:741–796, 2007.

[18] M. Rodriguez-Fernandez, J.A. Egea, and J. R Banga. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BioMed Central*, 2006.

[19] D. Ruppert, M.P. Wand, and R.J. Carroll. *Semiparametric regression.* Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2003.