

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

Quentin Clairon

Laboratoire Analyse et Probabilités, Université d'Evry (France)

Onzième colloque jeunes probabilistes et statisticiens 2014

Introduction des modèles étudiés et du problème d'estimation

Modèles d'équations différentielles

On s'intéresse au cas où la fonction de régression de dimension d est définie comme solution d'une d'équation différentielle ordinaire (EDO):

$$\begin{aligned}\frac{dX}{dt} &= f(t, X, \theta) \\ X(0) &= x_0\end{aligned}\tag{1}$$

- ▶ Modèle implicite: Donne la relation entre le taux de variation de la fonction de régression $\frac{dX}{dt}$ et $X(t)$ elle même.
- ▶ f, x_0, θ : respectivement champ de vecteur, condition initiale et ensemble de paramètres définissant le modèle.
- ▶ X_{θ, x_0} : Fonction de régression (appelée aussi variables d'état) solution de (1) pour une valeur de θ et une condition initiale x_0 donnée.
- ▶ Existence et unicité de X_{θ, x_0} mais souvent sans expression analytique possible.

EDO: Exemple d'utilisation (1)

Couramment utilisé en physique, chimie ([5]) et plus récemment en biologie (modèle SIR, biologie cellulaire [4],[12], réseaux de régulation génétique [2] etc...).

- ▶ Exemple 1: Réaction chimique (Isomerisation α -Pinene [5]) :

$$\dot{X}_1 = -(\theta_1 + \theta_2)X_1$$

$$\dot{X}_2 = \theta_1 X_1$$

$$\dot{X}_3 = \theta_2 X_1 - (\theta_3 + \theta_4)X_3 + \theta_5 X_5$$

$$\dot{X}_4 = \theta_3 X_3$$

$$\dot{X}_5 = \theta_4 X_3 - \theta_5 X_5$$

- ▶ Equation dérivée du principe de conservation de la masse.
- ▶ (X_1, \dots, X_5) : concentration des espèces chimiques d'intérêt.
- ▶ $\theta = (\theta_1, \dots, \theta_5)$: constante de réaction.

EDO: Exemple d'utilisation (2)

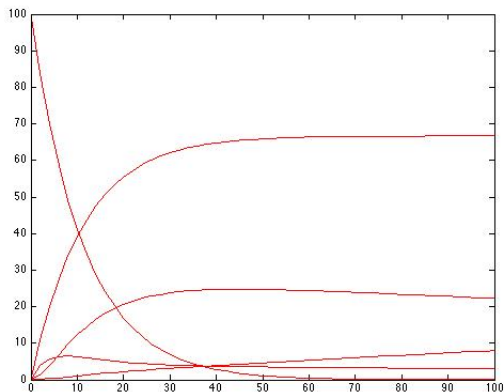


Figure: α -Pinene: Solution pour $x_0^* = (100, 0, 0, 0, 0)$ et $\theta^* = (5.93, 2.96, 2.05, 27.5, 4) \times 10^{-4}$

EDO: Exemple d'utilisation (3)

- ▶ Exemple 2: Modèle de FitzHugh-Nagumo ([4]) evolution en fonction du temps du voltage à travers la membrane d'un axone:

$$\begin{cases} \dot{V} &= c \left(V - \frac{V^3}{3} + R \right) \\ \dot{R} &= -\frac{1}{c} (V - a + bR) \end{cases} \quad (2)$$

- ▶ V : voltage
- ▶ R : variable de recouvrement (modélise les courants extérieurs)
- ▶ $\theta = (a, b, c)$: paramètres du modèle

EDO: Exemple d'utilisation (4)

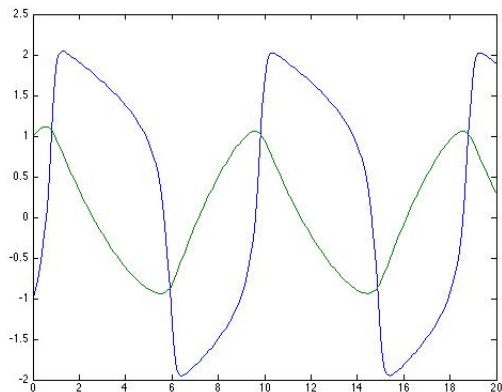


Figure: FitzHugh-Nagumo: Solution pour $(V_0; R_0) = (-1; 1)$ et $\theta^* = (0.2, 0.2, 3)$

Intérêt des modèles d'EDO

- ▶ Interprétabilité du modèle, description mécanique du processus étudié (traduction naturelle en terme mathématique des lois de conservations par ex).
- ▶ Peut reproduire une grande variété de mouvements complexes.
- ▶ Bon cadre d'approximation de processus stochastique ([13],[8]).
- ▶ Problème directe de prédiction bien traitée par des méthodes numériques (approximation de la solution [10]) et qualitatives (étude de stabilité, de présence de cycle limite).

Importance des paramètres (1)

Modèles souvent surparamétrisés.

Avoir une connaissance précise de la valeur des paramètres est important:

- ▶ Pour eux-même: Dus à leur interprétabilité leurs valeurs nous renseignent sur les éléments clés de la dynamique du système étudié.
- ▶ Pour le problème direct: Leurs valeurs sont critiques pour correctement prédire l'évolution du système étudié (théorie de la bifurcation).

Importance des paramètres (2)

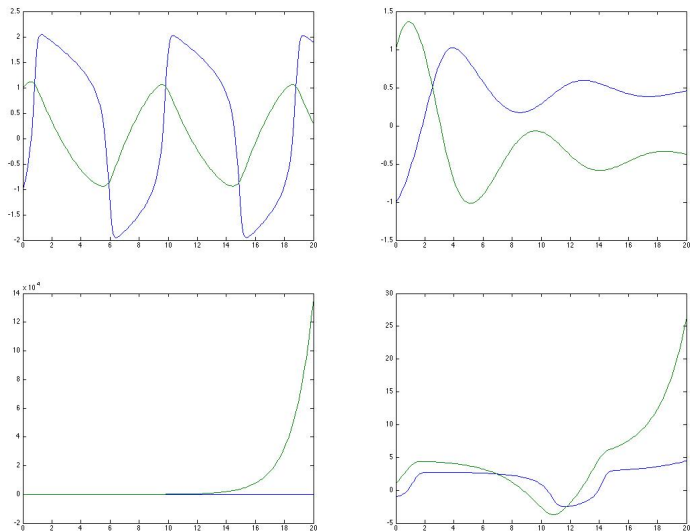


Figure: FitzHugh-Nagumo: Solution pour $(V_0; R_0) = (-1; 1)$ et pour différentes valeurs de θ

Problème étudié: Estimation de paramètre

A partir d'observations partielles et bruitées (Y_1, \dots, Y_n) du système:

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

où

- ▶ $0 = t_1 < t_2 \cdots < t_n = T$ temps d'observations aléatoirement répartis sur un intervalle $[0, T]$
- ▶ C la matrice d'observation de taille $d' \times d$ avec $d' \leq d$
- ▶ ϵ_i l'erreur de mesure.

On cherche à estimer θ^* sachant que X_{θ^*, x_0^*} solution de:

$$\begin{aligned} \frac{dX}{dt} &= f(t, X, \theta) \\ X(0) &= x_0 \end{aligned}$$

pour $\theta = \theta^*$ et $x_0 = x_0^*$.

Méthode d'estimation existante

Méthode

- ▶ Maximum de vraisemblance: équivalent à l'estimateur des moindres carrés non-linéaire si l'erreur de mesure est supposé gaussienne ([11], [15], [14], [3]):

$$\begin{aligned} (\hat{\theta}_{NLS}, \hat{x}_{0NLS}) &= \arg \min_{\theta, x_0} \|Y - X_{\theta, x_0}(\cdot)\|_2^2 \\ &:= \arg \min_{\theta, x_0} \sum_i (Y_i - X_{\theta, x_0}(t_i))^2 \end{aligned} \quad (3)$$

- ▶ Inférence Bayésienne ([9], [6], [1], [7]): On définit généralement l'estimateur de Bayes par:

$$(\hat{\theta}_B, \hat{x}_{0B}) = \mathbb{E}[\pi(\theta, x_0 \mid Y_1, \dots, Y_n)] \quad (4)$$

- ▶ $\pi(\theta, x_0 \mid Y_1, \dots, Y_n)$: loi à posteriori
- ▶ 4 approximé avec utilisation d'algorithmes MCMC ("Markov chain Monte Carlo") pour simuler des couples $(\theta^{(i)}, x_0^{(i)})$ asymptotiquement selon la loi $\pi(\theta, x_0 \mid Y_1, \dots, Y_n)$.

Difficultés rencontrées

1. Intégration répétée par méthode numériques de l'EDO augmentant le temps de calcul.
 - 1.1 Moindres carrées: Nécessaire pour résoudre (3)
 - 1.2 Inférence Bayésienne: Nécessaire pour les algorithmes MCMC
2. Besoin d'estimer la condition initiale x_0 en plus de θ , augmentation de la dimension du problème.
3. Grande sensibilité, dépendance non-linéaire de X_{θ, x_0} à θ .
 - 3.1 Moindres carrées: (3) présente une surface irrégulière, aux variations brusques et aux nombreux minima locaux.
 - 3.2 Inférence Bayésienne: Mauvaise estimation de $\pi(\theta, x_0 \mid Y_1, \dots, Y_n)$, convergence lente des algorithmes MCMC, $\pi(\theta, x_0 \mid Y_1, \dots, Y_n)$ multi-modale
4. Problèmes d'identifiabilités du modèle.
 - 4.1 Moindres carrées: Nombreux minima locaux pour (3)
 - 4.2 Inférence Bayésienne: Qualité de la loi à priori capitale pour assurer la convergence vers $\pi(\theta, x_0 \mid Y_1, \dots, Y_n)$

Difficultés rencontrées: Synthèse

- ▶ Problème d'identifiabilité généralement aggravé par le caractère partiellement observée du système.
- ▶ Les problèmes sont dues à la différence d'échelle entre le modèle et les estimateurs.
 - ▶ Un modèle d'EDO est défini à l'échelle de la dynamique.
 - ▶ Les estimateurs classiques sont construits au niveau des observations.

On propose un nouvelles approche d'estimation incorporant l'information sur la dynamique du système apporté par l'EDO afin de régulariser le problème inverse.

Nouvelle méthode d'estimation

Introduction de l'EDO perturbée

Relaxation sur la dynamique imposée par l'EDO (1), on introduit l'EDO perturbée:

$$\begin{aligned}\frac{dX}{dt} &= f(t, X, \theta) + u \\ X(0) &= x_0\end{aligned}\tag{5}$$

et $X_{\theta, x_0, u}$: solution de (5) pour une valeur de θ , une condition initiale x_0 et une perturbation u donnée.

Introduction de la fonction de coût

On introduit le coût general:

$$C_\lambda(u, \theta, x_0) = \int_0^T \left\| CX_{\theta, x_0, u}(t) - \hat{Y} \right\|_2^2 dt + \lambda \int_0^T \|u(t)\|_2^2 dt \quad (6)$$

Ici:

- ▶ $\int_0^T \left\| CX_{\theta, x_0, u}(t) - \hat{Y} \right\|_2^2 dt$: distance entre la trajectoire observée et les données
- ▶ $\int_0^T \|u(t)\|_2^2 dt$: distance entre le modèle initial et le modèle perturbée

\hat{Y} est un estimateur non-paramétrique obtenu directement à partir des données.

Définition de l'estimateur

On définit S_λ le coût profilé sur l'ensemble des perturbations possibles:

$$S_\lambda(\theta, x_0) := \min_u C_\lambda(u, \theta, x_0)$$

ainsi que notre estimateur:

$$\left(\widehat{\theta}_{OC}, \widehat{x}_{0OC} \right) := \arg \min_{\theta, x_0} S_\lambda(\theta, x_0)$$

Introduction d'une pénalisation de distance au modèle à l'échelle de la dynamique.

Calcul du coût profilé

- ▶ Problème: Définir $S_\lambda(\theta, x_0)$ revient à résoudre un problème d'optimisation en dimension infinie.
- ▶ Solution: Utiliser un résultat venu du contrôle optimal, le Principe du minimum de Pontryagin, donnant une caractérisation nécessaire du contrôle optimale \bar{u} (i.e permettant d'atteindre $\min_u C_\lambda(u, \theta, x_0)$) pour un couple (θ, x_0) donné.

Minimum de Pontryagin

Selon le principe du minimum de Pontryagin pour (θ, x_0) donné le contrôle optimale est sous la forme:

$$\bar{u}(t) = \frac{1}{2\lambda} p_{\theta, x_0}(t)$$

en introduisant $p_{\theta, x_0}(t)$ nommé vecteur adjoint tel que le couple $(X_{\theta, x_0, \bar{u}}, p_{\theta, x_0})$ soit solution de:

$$\begin{aligned}\dot{X}_{\theta, x_0, \bar{u}} &= f(t, X_{\theta, x_0, \bar{u}}, \theta) + \frac{1}{2\lambda} p_{\theta, x_0} \\ \dot{p}_{\theta, x_0} &= -\frac{\partial f}{\partial x}(t, X_{\theta, x_0, \bar{u}}, \theta)^T p_{\theta, x_0} + 2C^T (CX_{\theta, x_0, \bar{u}} - \hat{Y}) \\ (X_{\theta, x_0, \bar{u}}(0), p_{\theta, x_0}(T)) &= (x_0, 0)\end{aligned}$$

Ainsi:

$$S_\lambda(\theta, x_0) = \int_0^T \left\| CX_{\theta, x_0, \bar{u}}(t) - \hat{Y}(t) \right\|_2^2 dt + \frac{1}{4\lambda} \int_0^T \|p_{\theta, x_0}(t)\|_2^2 dt$$

Intérêt supplémentaire

- ▶ Présence d'une misspecification de modèle i.e le vrai processus suit une EDO différente de (1)
- ▶ Le contrôle optimale \bar{u} obtenu pour l'estimation $(\widehat{\theta}_{OC}, \widehat{x}_{0OC})$ mesure la distance entre le modèle supposé et la trajectoire observée. Il peut ainsi donner des indications sur une éventuelle mauvaise spécification.

Résultats et exemple

Résultat cas linéaire

Ici

$$f(t, X, \theta) = A(t, \theta)X + r(t, \theta)$$

- ▶ C1: Le modèle est identifiable à (θ^*, x_0^*) i.e:

$$\forall (\theta, x_0) \in \Theta \times \mathcal{X}; CX_{\theta, x_0} = CX_{\theta^*, x_0^*} \implies \begin{cases} \theta = \theta^* \\ x_0 = x_0^* \end{cases}$$

- ▶ C2: $(t, \theta) \rightarrow A(t, \theta)$ et $(t, \theta) \rightarrow r(t, \theta)$ C^2 sur $[0, T] \times \Theta$
- ▶ C3: $\frac{\partial^2 S_\lambda}{\partial \theta^T \partial \theta}(\theta^*)$ est inversible
- ▶ C4: (Y_i, t_i) i.i.d avec $Var(Y_i | t_i) = \sigma I_{d'}$ ($\sigma < +\infty$) et les t_i distribués uniformément sur $[0, T]$
- ▶ C5: \hat{Y} estimateur obtenu par spline de regression dans une base de taille K avec: $\sqrt{n}K^{-s} \rightarrow 0$ et $\frac{K^2}{n} \rightarrow 0$

Alors: Sous ces conditions $\hat{\theta} - \theta^*$ est asymptotiquement normal et:

$$\hat{\theta} - \theta^* = O_P(n^{-1/2})$$

Exemple: Comparaison avec les NLS

Protocole: Pour chaque modèle, avec un θ^* et x_0^* fixé on effectue 100 simulations d'estimations. Pour chacune d'elle on génère 50 observations uniformément réparties sur $[0, T]$ avec ajout d'un bruit gaussien d'écart type $\sigma = 15\%$ de la valeur moyenne des variables d'états.

Choix du couple (θ^*, x_0^*) :

- ▶ Pour α -Pinene: $x_0^* = (100, 0, 0, 0, 0)$,
 $\theta^* = (5.93, 2.96, 2.05, 27.5, 4) \times 10^{-4}$
- ▶ Pour FitzHugh Nagumo: $x_0^* = (-1, 1)$, $\theta^* = (0.2, 0.2, 3)$.

Exemple: Resultat

Précision de l'estimation mesurée par:

$$ARE = \frac{1}{Nb_{run}} \sum_{i=1}^{Nb_{run}} \frac{|\theta^* - \hat{\theta}|}{|\theta^*|}$$

et:

$$MSE = Bias(\hat{\theta}, \theta^*)^2 + Var(\hat{\theta})$$

α -Pinene		MSE ($\times 10^{-2}$)	ARE ($\times 10^{-2}$)
	$\widehat{\theta}_{NLS}$	0.26	8.47
	$\widehat{\theta}_{OC}$	0.11	7.01
FitzHugh-Nagumo		MSE ($\times 10^{-4}$)	ARE ($\times 10^{-3}$)
	$\widehat{\theta}_{NLS}$	0.14	4.80
	$\widehat{\theta}_{OC}$	0.11	4.08

Table: Resultats obtenus

Conclusion

Résultat prometteurs mais:

- ▶ Temps de calcul
- ▶ Sélection d'hyperparamètre
- ▶ Théorie (cas linéaire traité, cas non linéaire...)

Merci de votre attention



B. Calderhead, M Girolami, and N.D Lawrence.

Accelerating bayesian inference over nonlinear differential equations with gaussian processes.

In Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference, 2009.



M.B. Elowitz and S. Leibler.

A synthetic oscillatory network of transcriptional regulators.

Nature, 403(6767):335–338, 2000.



W.R. Esposito and C.A. Floudas.

Deterministic global optimization in nonlinear optimal control problems.

Journal of Global Optimization, 17:97–126, 2000.



R. FitzHugh.

Impulses and physiological states in theoretical models of nerve membrane.

Biophysical Journal, 6:445–466, 1961.



R. E. Fuguitt and J.E. Hawkins.

Rate of Thermal Isomerization of α -Pinene in the Liquid Phase.

J.A.C.S, 319(39), 1947.



A. Gelman, F. Bois, and J. Jiang.

Physiological pharmacokinetic analysis using population modeling and informative prior distributions.

Journal of the American Statistical Association, 91, 1996.



O. Ghasemi, M. Lindsey, T. Yang, N. Nguyen, Y. Huang, and Y. Jin.

Bayesian parameter estimation for nonlinear modelling of biological pathways.

BMC Systems Biology, 5, 2011.



D.T. Gillespie.

The chemical langevin equation.

Journal Of Chemical Physics, 113(1):297–306, 2000.



Y. Huang and H. Wu.

A bayesian approach for estimating antiviral efficacy in hiv dynamic models.

Journal of Applied Statistics, 33:155–174, 2006.



J. Hubbard and B. West.

Differential Equations: A dynamical Systems Approach.

Text in Applied Mathematics 5. Springer, 1997.



Z. Li, M.R. Osborne, and T. Prvan.

Parameter estimation of ordinary differential equations.

IMA Journal of Numerical Analysis, 25:264–285, 2005.



H. Miao, X. Xia, A. S. Perelson, and H. Wu.

On identifiability of nonlinear ode models and applications in viral dynamics.

SIAM Review, 53:3–39, 2011.



O. Radulescu, A. Muller, and A. Crudu.

Theoremes limites pour les processus de markov a sauts.

RSTI - TSI, 26(1):441–467, 2006.



M. Rodriguez-Fernandez, J.A. Egea, and J. R Banga.

Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems.

BioMed Central, 2006.



H. Wu, A. Kumar, H. Miao, J. Holden-Wiltse, T.R. Mosmann, A.M. Livingstone, G.T. Belz, A.S. Perelson, M.S. Zand, and D.J Topham.
Modeling of influenza-specific cd8+ t cells during the primary response indicates that the spleen is a major source of effectors.
The Journal Of Immunology, 187(9):4474–4482, 2010.