Parameter estimation of ODE's with regression splines: applications to biological networks

Nicolas Brunel

IBISC - FRE CNRS 2873, Université d'Evry, France

March 28 2007 - PESB'07, Manchester

Joint work with C. Klaassen (University of Amsterdam – EURANDOM, The Netherlands) and F. d'Alché-Buc (IBISC)

Biological networks and dynamical processes

Some facts

- Biological networks have a dynamical behavior,
- Macro molecules interact with each other within the cell,
- Useful data may be concentration profiles.

Objective

- Identify mathematical models describing the evolution of the concentrations (in a mechanistic way), usually Ordinary Differential Equations (ODE's).
- Concerned biological networks are metabolic or gene regulatory networks, signaling pathways.

Biological networks and dynamical processes

Some facts

- Biological networks have a dynamical behavior,
- Macro molecules interact with each other within the cell,
- Useful data may be concentration profiles.

Objective

- Identify mathematical models describing the evolution of the concentrations (in a mechanistic way), usually Ordinary Differential Equations (ODE's).
- Concerned biological networks are metabolic or gene regulatory networks, signaling pathways.

Assumptions on biochemical kinetics

- Classical mathematical description is deterministic and comes from biochemical kinetics.
- Equations are nonlinear and derived from law of mass action and Michaelis Menten kinetics: (phenomenological) parameters to determine.
- Structure of interactions is known, as the functional form of the equations.

Outline

Plan

Statistical setting

Structure of the model

$$\dot{x}(t) = f(x(t), \theta), t \in [0, T], x(t) \in \mathbb{R}^d$$

- Functional form of the vector field f is known and nonlinear
- Initial Value Problem: unknown initial conditions $x(0) = x_0$

The solution is denoted $\phi(t, \Theta)$, with $\Theta = (\theta, x_0)$

Objective: Estimation of θ from noisy concentration profiles

$$y_i = \phi(t_i, \Theta^*) + \epsilon_i, i = 1, \ldots, n$$

with i.i.d. noise ϵ_i (density $g(\cdot)$, with finite variance), x_0 kind of a nuisance parameter. All the system is observed.

Statistical setting

Structure of the model

$$\dot{x}(t) = f(x(t), \theta), t \in [0, T], x(t) \in \mathbb{R}^d$$

- Functional form of the vector field f is known and nonlinear
- Initial Value Problem: unknown initial conditions $x(0) = x_0$

The solution is denoted $\phi(t, \Theta)$, with $\Theta = (\theta, x_0)$

Objective: Estimation of θ from noisy concentration profiles

$$y_i = \phi(t_i, \Theta^*) + \epsilon_i, i = 1, \ldots, n$$

with i.i.d. noise ϵ_i (density $g(\cdot)$, with finite variance), x_0 kind of a nuisance parameter. All the system is observed.

Direct estimation

Standard approach: MLE for the estimation of $\Theta^* = (x_0^*, \theta^*)$

Classical nonlinear regression problem.

$$\hat{\Theta} \in \arg\max_{\Theta} \sum_{i=1}^{n} \log g(y_i - \phi(t_i, \Theta))$$

MLE is statistically efficient but

- ϕ is unknown (integration needed)
- parameter space is big (unknown initial conditions)

$\Rightarrow$ difficult optimization in a big space and heavy computations.

New characterization of the solution

Uniqueness of solution ϕ for one $\Theta = (x_0, \theta)$

If f smooth, $\phi(\cdot, \Theta^*) = \arg\min R_*^q(x) = \|\dot{x} - f(x, \theta^*)\|_q^q$ $(q > 0)$, when x is smooth and s.t. $x(0) = x_0^*$ ($\|\cdot\|_q = L^q$ norm)

Conversely, a fundamental property for the inference of the system from observations should be **identifiability**

Uniqueness of the parameter θ for given ϕ

ϕ^* is the true solution and let $R_*^q(\theta) = \|\dot{\phi}^* - f(\phi^*, \theta)\|_q^q$:

$$\theta^* = \arg\min_\theta \|\dot{\phi}^* - f(\phi^*, \theta)\|_q^q$$

New characterization of the solution

Uniqueness of solution ϕ for one $\Theta = (x_0, \theta)$

If f smooth, $\phi(\cdot, \Theta^*) = \arg\min R_*^q(x) = \|\dot{x} - f(x, \theta^*)\|_q^q$ $(q > 0)$, when x is smooth and s.t. $x(0) = x_0^*$ ($\|\cdot\|_q = L^q$ norm)

Conversely, a fundamental property for the inference of the system from observations should be **identifiability**

Uniqueness of the parameter θ for given ϕ

ϕ^* is the true solution and let $R_*^q(\theta) = \|\dot{\phi}^* - f(\phi^*, \theta)\|_q^q$:

$$\theta^* = \arg\min_\theta \|\dot{\phi}^* - f(\phi^*, \theta)\|_q^q$$

New characterization of the solution

Uniqueness of solution ϕ for one $\Theta = (x_0, \theta)$

If f smooth, $\phi(\cdot, \Theta^*) = \arg\min R^q_*(x) = \|\dot{x} - f(x, \theta^*)\|^q_q$ $(q > 0)$, when x is smooth and s.t. $x(0) = x^*_0$ ($\|\cdot\|_q$=L^q norm)

Conversely, a fundamental property for the inference of the system from observations should be **identifiability**

Uniqueness of the parameter θ for given ϕ

ϕ^* is the true solution and let $R^q_*(\theta) = \|\dot{\phi}^* - f(\phi^*, \theta)\|^q_q$:

$$\theta^* = \arg\min_\theta \|\dot{\phi}^* - f(\phi^*, \theta)\|^q_q$$

General Principle of two-step estimator

Nonparametric estimation

- Estimation of ϕ^* with an L^q consistent estimator $\hat{\phi}_n$
- Estimation of the derivative ϕ^*: $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$

Plug-in

Replace ϕ^* by $\hat{\phi}_n$ in R_*^q: $R_n^q(\theta) = \|\dot{\hat{\phi}}_n - f(\hat{\phi}_n, \theta)\|_q^q$

Definition of the **two-step estimator** (M-estimator)

$$\hat{\theta}_n \in \arg\min_{\theta} R_n^q(\theta)$$

If f is uniformly Lipschitz and $\forall \epsilon > 0$, $\sup_{\|\theta-\theta^*\|>\epsilon} R_*^q(\theta) > 0$ (identifiability) then $\hat{\theta}_n \xrightarrow{P} \theta^*$.

General Principle of two-step estimator

Nonparametric estimation

- Estimation of ϕ^* with an L^q consistent estimator $\hat{\phi}_n$
- Estimation of the derivative ϕ^*: $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$

Plug-in

Replace ϕ^* by $\hat{\phi}_n$ in R^q_*: $R^q_n(\theta) = \|\dot{\hat{\phi}}_n - f(\hat{\phi}_n, \theta)\|^q_q$

Definition of the **two-step estimator** (M-estimator)

$$\hat{\theta}_n \in \arg\min_\theta R^q_n(\theta)$$

If f is uniformly Lipschitz and $\forall \epsilon > 0$, $\sup_{\|\theta-\theta^*\|>\epsilon} R^q_*(\theta) > 0$ (identifiability) then $\hat{\theta}_n \xrightarrow{P} \theta^*$.

General Principle of two-step estimator

Nonparametric estimation

- Estimation of ϕ^* with an L^q consistent estimator $\hat{\phi}_n$
- Estimation of the derivative ϕ^*: $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$

Plug-in

Replace ϕ^* by $\hat{\phi}_n$ in R^q_*: $R^q_n(\theta) = \|\dot{\hat{\phi}}_n - f(\hat{\phi}_n, \theta)\|^q_q$

Definition of the **two-step estimator** (M-estimator)

$$\hat{\theta}_n \in \arg\min_{\theta} R^q_n(\theta)$$

If f is uniformly Lipschitz and $\forall \epsilon > 0$, $\sup_{\|\theta-\theta^*\|>\epsilon} R^q_*(\theta) > 0$ (identifiability) then $\hat{\theta}_n \xrightarrow{P} \theta^*$.

General Principle of two-step estimator

Nonparametric estimation

- Estimation of ϕ^* with an L^q consistent estimator $\hat{\phi}_n$
- Estimation of the derivative $\dot{\phi}^*$: $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$

Plug-in

Replace ϕ^* by $\hat{\phi}_n$ in R_*^q: $R_n^q(\theta) = \|\dot{\hat{\phi}}_n - f(\hat{\phi}_n, \theta)\|_q^q$

Definition of the **two-step estimator** (M-estimator)

$$\hat{\theta}_n \in \arg\min_{\theta} R_n^q(\theta)$$

If f is uniformly Lipschitz and $\forall \epsilon > 0$, $\sup_{\|\theta-\theta^*\|>\epsilon} R_*^q(\theta) > 0$ (identifiability) then $\hat{\theta}_n \xrightarrow{P} \theta^*$.

Comments

- General and adaptable to $\dot{x}(t) = f(t, u(t), x(t); \theta)$ where u is known continuous time input variable,
- Flexibility in the choice of $\hat{\phi}_n$ (Nadaraya-Watson, Wavelets, Splines,...),
- Choice of L^q distance (for instance $q = 1, 2$),
- Discretization of R_n^q for minimization with Rieman sum:

$$R_n^q(\theta) \simeq \sum_i (\dot{\hat{\phi}}_n(s_i) - f(\hat{\phi}_n(s_i), \theta))^q (s_{i+1} - s_i)$$

it corresponds to a second nonlinear regression: explanation of the derivative with the current states.

Comments

- General and adaptable to $\dot{x}(t) = f(t, u(t), x(t); \theta)$ where u is known continuous time input variable,
- Flexibility in the choice of $\hat{\phi}_n$ (Nadaraya-Watson, Wavelets, Splines,. . .),
- Choice of L^q distance (for instance $q = 1, 2$),
- Discretization of R_n^q for minimization with Rieman sum:

$$R_n^q(\theta) \simeq \sum_i (\dot{\hat{\phi}}_n(s_i) - f(\hat{\phi}_n(s_i), \theta))^q (s_{i+1} - s_i)$$

it corresponds to a second nonlinear regression: explanation of the derivative with the current states.

Comments

- General and adaptable to $\dot{x}(t) = f(t, u(t), x(t); \theta)$ where u is known continuous time input variable,
- Flexibility in the choice of $\hat{\phi}_n$ (Nadaraya-Watson, Wavelets, Splines,...),
- Choice of L^q distance (for instance $q = 1, 2$),
- Discretization of R_n^q for minimization with Rieman sum:

$$R_n^q(\theta) \simeq \sum_i (\dot{\hat{\phi}}_n(s_i) - f(\hat{\phi}_n(s_i), \theta))^q (s_{i+1} - s_i)$$

it corresponds to a second nonlinear regression: explanation of the derivative with the current states.

Comments

- General and adaptable to $\dot{x}(t) = f(t, u(t), x(t); \theta)$ where u is known continuous time input variable,
- Flexibility in the choice of $\hat{\phi}_n$ (Nadaraya-Watson, Wavelets, Splines,...),
- Choice of L^q distance (for instance $q = 1, 2$),
- Discretization of R_n^q for minimization with Rieman sum:

$$R_n^q(\theta) \simeq \sum_i (\dot{\hat{\phi}}_n(s_i) - f(\hat{\phi}_n(s_i), \theta))^q (s_{i+1} - s_i)$$

it corresponds to a second nonlinear regression: explanation of the derivative with the current states.

Computational Advantages

- no integration needed
- Componentwise estimation

$$\forall i = 1, \dots, d, \dot{x}_i = f_i(x_i, x_{[i]}, \theta_{[i]})$$

f_i=i^{th} component of f, $x_{[i]}$= list of components of x involved in f_i, $\theta_{[i]}$= list of components of parameter θ involved in f_i. Minimization of d criteria

$$R^q_{n,i}(\theta_{[i]}) = \|\dot{\hat{\phi}}_{i,n} - f_i(\hat{\phi}_{i,n}, \hat{\phi}_{[i],n}, \theta_{[i]})\|^q_q$$

and $\hat{\theta}_{n,[i]} = \arg\min R^q_{n,[i]}(\theta_{[i]})$

$\Rightarrow$ Easy optimization step.

Computational Advantages

- no integration needed
- Componentwise estimation

$$\forall i = 1, \dots, d, \dot{x}_i = f_i(x_i, x_{[i]}, \theta_{[i]})$$

f_i=i^{th} component of f, $x_{[i]}$= list of components of x involved in f_i, $\theta_{[i]}$= list of components of parameter θ involved in f_i.

Minimization of d criteria

$$R^q_{n,i}(\theta_{[i]}) = \|\dot{\hat{\phi}}_{i,n} - f_i(\hat{\phi}_{i,n}, \hat{\phi}_{[i],n}, \theta_{[i]})\|^q_q$$

and $\hat{\theta}_{n,[i]} = \arg\min R^q_{n,[i]}(\theta_{[i]})$

$\Rightarrow$ Easy optimization step.

Computational Advantages

- no integration needed
- Componentwise estimation

$$\forall i = 1, \ldots, d, \dot{x}_i = f_i(x_i, x_{[i]}, \theta_{[i]})$$

f_i=i^{th} component of f, $x_{[i]}$= list of components of x involved in f_i, $\theta_{[i]}$= list of components of parameter θ involved in f_i.
Minimization of d criteria

$$R^q_{n,i}(\theta_{[i]}) = \|\dot{\hat{\phi}}_{i,n} - f_i(\hat{\phi}_{i,n}, \hat{\phi}_{[i],n}, \theta_{[i]})\|^q_q$$

and $\hat{\theta}_{n,[i]} = \arg\min R^q_{n,[i]}(\theta_{[i]})$

$\Rightarrow$ Easy optimization step.

Plan

Splines

Vector space and B-splines

- Splines on $[0, T]$ are **piecewise polynomial functions** of a given maximal degree k, for a given **break points** sequence $\boldsymbol{\xi}$ formed of $0 = \xi_0 < \xi_1 < \cdots < \xi_l < \xi_{l+1} = T$). Regularity conditions are controlled with ν_i continuity conditions at internal knots $\xi_i, i = 1, 2, \ldots, l$.
- It is a vector space spanned by B-splines $B_1, \ldots, B_K$ ($K = kl - \sum_{i=2}^{l} \nu_i$ is the dimension).
  B-splines $\{B_1, \ldots, B_K\}$ form a basis of $\mathbb{S}_{\boldsymbol{\xi},k,\nu}$, and have compact support and are nearly orthogonal.

Notations: $\mathbf{B}(\cdot) = (B_1(\cdot), \ldots, B_K(\cdot))^\top$

Nonparametric regression

Nonparametric regression: Least Squares Estimation

From time series $(y_i)_{i=1,\ldots,n}$, with $y_i = (y_i^j)_{j=1,\ldots,d}$, the regression function is estimated as $\hat{\phi}_{j,n}(t) = \hat{c}_{j,n}^\top \mathbf{B}(t)$ with

$$\forall j = 1, \ldots, d, \hat{c}_{j,n} = \arg\min_{c_n} \sum_{i=1}^{n} (y_i^j - c_n^\top \mathbf{B}(t_i))^2$$

Asymptotics

Essential point: Bias-Variance trade-off

Solution $\hat{\phi}_n$ is approximated by a vector space of dimension K_n.

- smooth data and reduce variance,
- inexact reconstruction of ϕ^*.

The quality of approximation must increase as $n \longrightarrow \infty$.
The sequence of B-splines is denoted $(B_{k,n}(\cdot))_{k=1,\dots,K_n}$.

Adaptive number of knots

$K_n \longrightarrow \infty$ such that step size $|\xi_n|$ of the knots partition ξ_n tends to 0. The crux is that there exists

$$c_n \in \mathbb{R}^{K_n} \text{ with } \|\phi^* - c_n^\top \mathbf{B}_n\|_2 \longrightarrow 0 \text{ sufficiently "rapidly"}$$

Asymptotics

Essential point: Bias-Variance trade-off

Solution $\hat{\phi}_n$ is approximated by a vector space of dimension K_n.

- smooth data and reduce variance,
- inexact reconstruction of ϕ^*.

The quality of approximation must increase as $n \longrightarrow \infty$.
The sequence of B-splines is denoted $(B_{k,n}(\cdot))_{k=1,\dots,K_n}$.

Adaptive number of knots

$K_n \longrightarrow \infty$ such that step size $|\xi_n|$ of the knots partition ξ_n tends to 0. The crux is that there exists

$$c_n \in \mathbb{R}^{K_n} \text{ with } \|\phi^* - c_n^\top \mathbf{B}_n\|_2 \longrightarrow 0 \text{ sufficiently "rapidly"}$$

Properties

If $\hat{\phi}_n(t) = \sum_{k=1}^{K_n} c_k B_{k,n}(t)$ is a spline $K_n = O(n^{1/2m+3})$, f is C^m, $m \geq 1$, then:

- $(\hat{\theta}_n - \theta^*) = O(n^{-\frac{m+1}{2m+3}})$
- $n^{\frac{m+1}{2m+3}}(\hat{\theta}_n - \theta^*)$ is asympt. normal.

It is not $\sqrt{n}$ consistent estimator, but tends to it as f is smoother. Similar ideas have been used by Varah (1982), Madar *et al* (2003), Ramsay *et al* (2006).

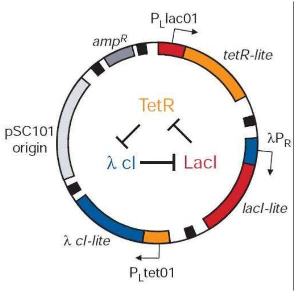
Plan

Repressilator

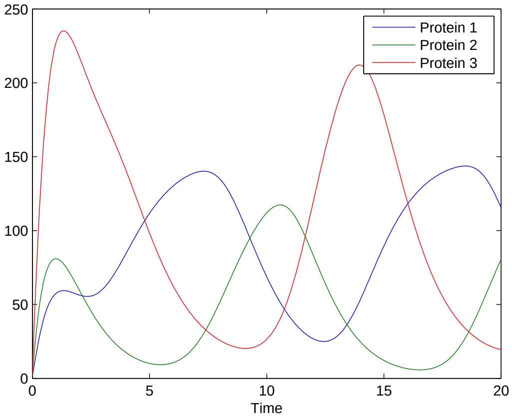
Gene regulatory network with 3 genes and 3 proteins (Elowitz et Leibler, 2000)

Oscillating regulatory networks of 3 mutually inhibited genes (via their induced proteins).

$$
\begin{aligned}
\dot{r}_1(t) &= v_1^{max} \frac{k_{12}^n}{k_{12}^n + p_2(t)^n} - k_1^r r_1(t) \\
\dot{r}_2(t) &= v_2^{max} \frac{k_{23}^n}{k_{23}^n + p_3(t)^n} - k_2^r r_2(t) \\
\dot{r}_3(t) &= v_3^{max} \frac{k_{31}^n}{k_{31}^n + p_1(t)^n} - k_3^r r_3(t) \\
\dot{p}_1(t) &= \gamma_1 r_1(t) - k_1^p p_1(t) \\
\dot{p}_2(t) &= \gamma_2 r_2(t) - k_2^p p_2(t) \\
\dot{p}_3(t) &= \gamma_3 r_3(t) - k_3^p p_3(t)
\end{aligned}
$$

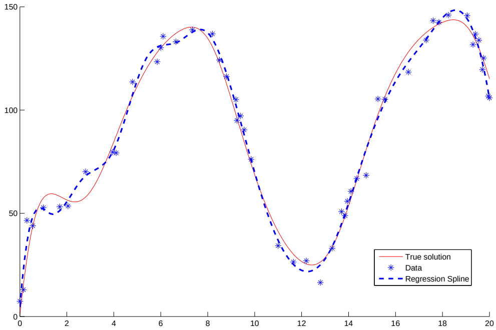


Evolution of protein concentrations



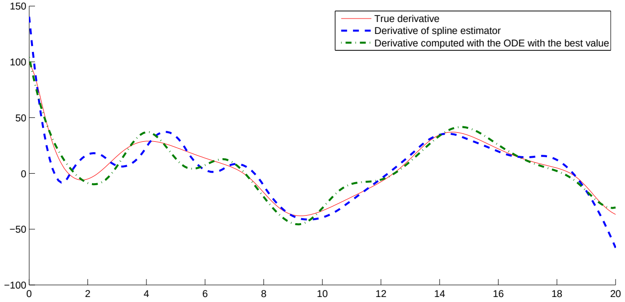
Nonparametric estimation of the solution by splines

Estimation of the solution $\phi(\cdot, \Theta^*)$ by C^2 cubic splines (15 knots) from 50 observations



Estimates of the derivative of $\hat{\phi}_n$

Discrepancy between $\dot{\hat{\phi}}_n$ (blue) and $f(\hat{\phi}_n, \theta)$ (green) is minimized in L^2.



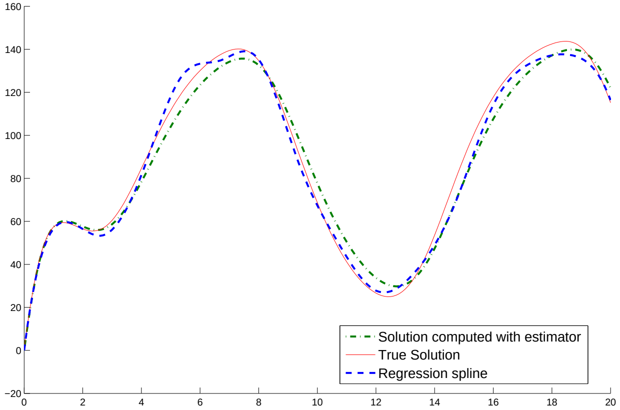
Estimated parameters

Component i	v_i^{max}	k_{ij}	k_i	γ_i	k_i^p
$i = 1$	134.3 (150)	50.5 (50)	0.9 (1)	0.96 (1)	0.97 (1)
$i = 2$	69 (80)	43 (40)	1 (1)	1.9 (2)	0.94 (1)
$i = 3$	125 (100)	47.4 (50)	1.1 (1)	2.9 (3)	0.97 (1)

and $n = 2.95(3)$.

The reconstructed curves

Comparison between ϕ^* (red), $\hat{\phi}_n$ (blue) and $\phi(\cdot, \hat{\Theta}_n)$ (green).



Plan

Conclusion

- New consistent estimator for ODE's using regression tools,
- Fast estimation (in particular for $q = 2$)

Perspectives

- Practical adaptive choice of knots for splines,
- Ameliorate $\hat{\theta}_n$ into an efficient $\sqrt{n}$-consistent estimator,
- Extend to other mathematical models of systems biology,
- Extend to other type of data (knock-out data).

Conclusion

- New consistent estimator for ODE's using regression tools,
- Fast estimation (in particular for $q = 2$)

Perspectives

- Practical adaptive choice of knots for splines,
- Ameliorate $\hat{\theta}_n$ into an efficient $\sqrt{n}$-consistent estimator,
- Extend to other mathematical models of systems biology,
- Extend to other type of data (knock-out data).