Learning parameters in ODEs

Two-step estimator

Florence d'Alché-Buc
Joint work with Nicolas Brunel and Paola Bouchet

Joint work with Minh Quach and Nicolas Brunel
IBISC FRE 3190 CNRS, Université d'Évry-Val d'Essonne, France

Coming back to classical parameter estimation in ODE

- General model is

$$\dot{x}(t) = f(t, x(t), \theta)$$

  with initial condition $x(0) = x_0$. f smooth enough for existence and uniqueness of the solution $\phi(\cdot, (x_0, \theta))$.

- Assumption: there exists a true parameter (x_0^*, θ^*) such that the observations are

$$y_i = \phi(t_i, (x_0^*, \theta^*)) + \epsilon_i$$

  where ϵ_i is a white noise (i.e. $x(t_i), y_i \in \mathbb{R}^d$),i.e. **ALL** the concentration profiles are observed.

- A basic nonlinear regression problem we have to estimate the big parameter $\theta^* = (x_0^*, \theta^*)$ from $(t_i, y_i)_{i=1,\ldots,n}$.

Coming back to classical parameter estimation in ODE

- General model is

$$\dot{x}(t) = f(t, x(t), \theta)$$

  with initial condition $x(0) = x_0$. f smooth enough for existence and uniqueness of the solution $\phi(\cdot, (x_0, \theta))$.
- Assumption: there exists a true parameter (x_0^*, θ^*) such that the observations are

$$y_i = \phi\left(t_i, (x_0^*, \theta^*)\right) + \epsilon_i$$

  where ϵ_i is a white noise (i.e. $x(t_i), y_i \in \mathbb{R}^d$),i.e. **ALL** the concentration profiles are observed.

- A basic nonlinear regression problem we have to estimate the big parameter $\boldsymbol{\theta}^* = (x_0^*, \theta^*)$ from $(t_i, y_i)_{i=1,\ldots,n}$.

Two-step estimators

1. Functional estimation from $(t_i, y_i)_{i=1,\dots,n}$:
   1. estimate $\hat{\phi}_n$ with nonparametric estimators (Splines, Support Vector Regression, or your prefered nonparametric estimato, ...)
   2. estimate the derivative $\dot{\phi}$ with $\hat{\dot{\phi}}_n$ (typically $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$)
2. Minimize the discrepancy between the two estimators of the derivatives measured

$$R_n^2(\theta) = \left\| \hat{\dot{\phi}}_n - f(t, \hat{\phi}_n, \theta) \right\|_{L^2}^2 = \int_0^1 \left\| \hat{\dot{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|_2^2 dt$$

The two step estimator is

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} R_n^2(\theta)$$

Initially proposed by Varah 1982 (with LS splines), commonly used since then.

Two-step estimators

1. Functional estimation from $(t_i, y_i)_{i=1,\dots,n}$:
   1. estimate $\hat{\phi}_n$ with nonparametric estimators (Splines, Support Vector Regression, or your prefered nonparametric estimato, ...)
   2. estimate the derivative $\dot{\phi}$ with $\dot{\hat{\phi}}_n$ (typically $\dot{\hat{\phi}}_n = \dot{\hat{\phi}}_n$)
2. Minimize the discrepancy between the two estimators of the derivatives measured

$$R_n^2(\theta) = \left\| \dot{\hat{\phi}}_n - f(t, \hat{\phi}_n, \theta) \right\|_{L^2}^2 = \int_0^1 \left\| \dot{\hat{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|_2^2 dt$$

The two step estimator is

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} R_n^2(\theta)$$

Initially proposed by Varah 1982 (with LS splines), commonly used since then.

Two-step estimators

1. Functional estimation from $(t_i, y_i)_{i=1,\ldots,n}$:
   1. estimate $\hat{\phi}_n$ with nonparametric estimators (Splines, Support Vector Regression, or your prefered nonparametric estimato, . . .)
   2. estimate the derivative $\dot{\phi}$ with $\dot{\hat{\phi}}_n$ (typically $\dot{\hat{\phi}}_n = \dot{\hat{\phi}}_n$)
2. Minimize the discrepancy between the two estimators of the derivatives measured

$$R_n^2(\theta) = \left\|\dot{\hat{\phi}}_n - f(t, \hat{\phi}_n, \theta)\right\|_{L^2}^2 = \int_0^1 \left\|\dot{\hat{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta)\right\|_2^2 dt$$

The two step estimator is

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} R_n^2(\theta)$$

Initially proposed by Varah 1982 (with LS splines), commonly used since then.

Advantages

- No numerical integration, $R_n^2(\theta)$ is easier to compute and minimize
- Componentwise optimization (decoupled equations): for $j = 1, \ldots d$,

$$\hat{\theta}_n^{[j]} = \arg\min_{\theta^{[j]}} \left\| \hat{\dot{\phi}}_n^j - f_j(t, \hat{\phi}_n, \theta^{[j]}) \right\|_{L^2}^2$$

  optimization takes place in a smaller space.

- Intuitive interpretation: Riemann discretization of the integral turns the optimization problem in a classical nonlinear regression problem

$$R_n^2(\theta) \approx \sum_{t_k} \left(\hat{\dot{\phi}}_n^j(t_k) - f_j(t_k, \hat{\phi}_n(t_k), \theta^{[j]}) \right)^2 \Delta t_k$$

Advantages

- No numerical integration, $R_n^2(\theta)$ is easier to compute and minimize
- Componentwise optimization (decoupled equations): for $j = 1, \ldots d$,

$$\hat{\theta}_n^{[j]} = \arg\min_{\theta^{[j]}} \left\| \hat{\dot{\phi}}_n^j - f_j(t, \hat{\phi}_n, \theta^{[j]}) \right\|_{L^2}^2$$

  optimization takes place in a smaller space.

- Intuitive interpretation: Riemann discretization of the integral turns the optimization problem in a classical nonlinear regression problem

$$R_n^2(\theta) \approx \sum_{t_k} \left(\hat{\dot{\phi}}_n^j(t_k) - f_j(t_k, \hat{\phi}_n(t_k), \theta^{[j]}) \right)^2 \Delta t_k$$

Advantages

- No numerical integration, $R_n^2(\theta)$ is easier to compute and minimize
- Componentwise optimization (decoupled equations): for $j = 1, \dots d$,

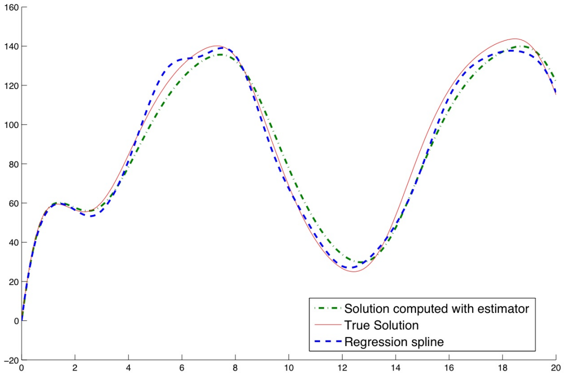
$$\hat{\theta}_n^{[j]} = \arg\min_{\theta^{[j]}} \left\| \hat{\dot{\phi}}_n^j - f_j(t, \hat{\phi}_n, \theta^{[j]}) \right\|_{L^2}^2$$

  optimization takes place in a smaller space.

- Intuitive interpretation: Riemann discretization of the integral turns the optimization problem in a classical nonlinear regression problem

$$R_n^2(\theta) \approx \sum_{t_k} \left(\hat{\dot{\phi}}_n^j(t_k) - f_j(t_k, \hat{\phi}_n(t_k), \theta^{[j]}) \right)^2 \Delta t_k$$

The reconstructed curves (repressilator, all parameters learnt)



Estimated parameters (repressilator)

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)

.

Integration of shape constraints (master work with Paola Bouchet)

- Goal: in case of very small dataset, use prior knowledge to constrain the solution
- Some biologists know if the system is oscillating or not if it comes back to equilibrium ...
- Good nonparametric estimators $\hat{\phi}_n$ of ϕ^* can be obtained with prior information, such as
  - positivity, monotony, convexity i.e. shape-constrained inference,
  - known initial (boundary) values,
  - “semiparametric” estimation

Meaningful decomposition

- Ameliorate the estimation of ϕ^* by writing the decomposition

$$\phi^*(t) = S(t) + N(t)$$

S : Main shape, trend

N : Transient behavior, perturbation w.r.t a reference situation.

"Refined" examples of possible shapes for S:

- Periodic solution (limit cycle) of nonlinear ODE

$$S(t) = \sum_{k=0}^{\infty} b_k \cos(2\pi k\omega t + \phi_k)$$

- Likely ("normal") parameters values $\theta_1, \ldots, \theta_\ell$

$$S(t) = \phi(t, (x_0, \theta_1)) \text{ or } \sum_{k=1}^{\ell} b_k \phi(t, (x_0, \theta_k))$$

Meaningful decomposition

- Ameliorate the estimation of ϕ^* by writing the decomposition

$$\phi^*(t) = S(t) + N(t)$$

S : Main shape, trend

N : Transient behavior, perturbation w.r.t a reference situation.

"Refined" examples of possible shapes for S:

- Periodic solution (limit cycle) of nonlinear ODE

$$S(t) = \sum_{k=0}^{\infty} b_k \cos(2\pi k\omega t + \phi_k)$$

- Likely (''normal") parameters values $\theta_1, \ldots, \theta_\ell$

$$S(t) = \phi\left(t, (x_0, \theta_1)\right) \text{ or } \sum_{k=1}^{\ell} b_k \phi(t, (x_0, \theta_k))$$

Constraints and semiparametric SVR

- Classical SVR (with RKHS $\mathcal{H}$ and associated kernel $k(\cdot,\cdot)$):

$$\hat{\phi}_n = \arg\min_{f\in\mathcal{H}} \sum_{i=1}^{n} L_\epsilon(y_i - f(t_i)) + C\,\|f\|_{\mathcal{H}}^2 \implies \hat{\phi}_n(t) = b + \sum_{i\in SV} c_i k(t_i, t)$$

with $L_\epsilon(x) = \max(|x| - \epsilon, 0)$, C=trade-off constant, SV = set of Support Vectors (k is typically a Gaussian kernel).

- Semiparametric SVR: $S \in \text{span}\,\{\psi_1, \ldots, \psi_\ell\}$:

$$\hat{\phi}_n = \arg\min_{N\in\mathcal{H}} \sum_{i=1}^{n} L_\epsilon\left(y_i - (S(t_i) + N(t_i))\right) + C\,\|N\|_{\mathcal{H}}^2$$

$$\implies \hat{\phi}_n(t) = \underbrace{\sum_{k=1}^{\ell} b_k \psi_k(t)}_{\hat{S}(t)} + \underbrace{\sum_{i\in SV} c_i k(t_i, t)}_{\hat{N}(t)}$$

Coefficients b_k, c_i are computed as solution of constrained convex (quadratic) problem. If $\psi_k \in \mathcal{H}$, then $\hat{S} \perp \hat{N}$ in $\mathcal{H}$.

Constraints and semiparametric SVR

- Classical SVR (with RKHS $\mathcal{H}$ and associated kernel $k(\cdot,\cdot)$):

$$\hat{\phi}_n = \arg\min_{f\in\mathcal{H}} \sum_{i=1}^{n} L_\epsilon(y_i - f(t_i)) + C\,\|f\|^2_{\mathcal{H}} \implies \hat{\phi}_n(t) = b + \sum_{i\in SV} c_i k(t_i, t)$$

with $L_\epsilon(x) = \max(|x| - \epsilon, 0)$, C=trade-off constant, SV = set of Support Vectors (k is typically a Gaussian kernel).

- Semiparametric SVR: $S \in \text{span}\,\{\psi_1, \ldots, \psi_\ell\}$:

$$\hat{\phi}_n = \arg\min_{N\in\mathcal{H}} \sum_{i=1}^{n} L_\epsilon\left(y_i - (S(t_i) + N(t_i))\right) + C\,\|N\|^2_{\mathcal{H}}$$

$$\implies \hat{\phi}_n(t) = \underbrace{\sum_{k=1}^{\ell} b_k \psi_k(t)}_{\hat{S}(t)} + \underbrace{\sum_{i\in SV} c_i k(t_i, t)}_{\hat{N}(t)}$$

Coefficients b_k, c_i are computed as solution of constrained convex (quadratic) problem. If $\psi_k \in \mathcal{H}$, then $\hat{S} \perp \hat{N}$ in $\mathcal{H}$.

Comparisons of constraints in Repressilator

Use of $\hat{\phi}_n^{period}$ with $\ell = 2, \omega = 0.55$; $\hat{\phi}_n^{ode}$, with $\theta_1 = \theta^* + 0.2$, $\theta_2 = \theta^* - 0.1$; $\hat{\phi}_n^{cons}$ with $\hat{\phi}_n^{cons}(0) = \phi^*(0)$ and $\hat{\phi}_n^{cons}(T) = \phi^*(T)$.

	True Parameter	$\hat{\phi}_n$	$\hat{\phi}_n^{period}$	$\hat{\phi}_n^{ode}$	$\hat{\phi}_n^{cons}$
v_1	150	150.0	113.2	149.1	149.17
v_2	80	79.99	87.6	78.2	78.4
v_3	100	101.98	81.2	101.8	100.6
$k_{1,2}$	50	50.5	66.6	50.4	50.5
$k_{2,3}$	40	40.4	53.2	40.1	40.5
$k_{3,1}$	50	49.65	39.0	48.9	50.2
k_1	1	0.98	1.24	1.23	0.99
k_2	2	1.96	1.9	1.91	1.95
k_3	3	2.85	3.21	3.18	2.8

Table: Mean of the two-step estimator computed with different estimators of the true solution of the data when $T = 40$ observations, computed with 100 Monte Carlo runs.

Conclusion and perspectives

Some comments

- First approach: deal with hidden variables, non special attention to initial condition value nor solution to ODEs (partial integration)
- Second approach: no hidden variables, nonparametric estimation of the solution of the ODE, possibility to integrate shape constraints

Current and future work

- Combine the advantage of both
  1. State-space model: By searching for parameters θ and initial condition value, that lead to an approximate solution of the ODES
  2. State-space model: making the nonparametric solution appear in the equations