Bayesian data-driven model discovery under uncertainty

M. Ceoloni, F. Fatone, F. Fedeli

Supervised by:
Prof. A. Guglielmi, Prof. A. Manzoni

November 20, 2020

Introduction

Model identification and parameters estimation starting from raw data is of extreme importance in many fields of application (recently even more...)

Introduction

Model identification and parameters estimation starting from raw data is of extreme importance in many fields of application (recently even more...)

IL FAMOSO RO

Coronavirus, ecco regione per regione l'indice di trasmissibilità della malattia

L' Ro indica il numero di infezioni prodotte da una persona nell'arco del suo periodo infettivo. Un dato associato al tempo che intercorre nel passaggio della malattia fra un infetto primario e quelli secondari

Introduction

Model identification and parameters estimation starting from raw data is of extreme importance in many fields of application (recently even more...)

IL FAMOSO RO

Coronavirus, ecco regione per regione l'indice di trasmissibilità della malattia

L' Ro indica il numero di infezioni prodotte da una persona nell'arco del suo periodo infettivo. Un dato associato al tempo che intercorre nel passaggio della malattia fra un infetto primario e quelli secondari

In some of these applications is critical to define a clear uncertainty quantification framework.

Our work aims at providing Bayesian tools for:

- Model identification and selection for dynamical systems
- Model parameters estimation (uncertainty quantification)

Numerical Framework

For the first part of parameters estimation we relied on a **numerical framework** implemented exploiting **TensorFlow** automatic differentiation capabilities.

Considering the dynamical system $\dot{\boldsymbol{x}} = f(\boldsymbol{x}, t; \boldsymbol{\theta})$, we used **gradient descent** in order to approximate

$$\hat{\boldsymbol{\theta}} = argmin_{\boldsymbol{\theta}} \sum_{i=1}^{n} ||\boldsymbol{x}(t_i + \Delta t_i) - h_{\boldsymbol{\theta}}(\boldsymbol{x}(t_i))||^2$$

with $h_{\boldsymbol{\theta}}(\boldsymbol{x}(t_i))$ being the output of a numerical ODE solver (RK4).

Numerical Framework

For the first part of parameters estimation we relied on a **numerical framework** implemented exploiting **TensorFlow** automatic differentiation capabilities.

Considering the dynamical system $\dot{\boldsymbol{x}} = f(\boldsymbol{x}, t; \boldsymbol{\theta})$, we used **gradient descent** in order to approximate

$$\hat{\boldsymbol{\theta}} = argmin_{\boldsymbol{\theta}} \sum_{i=1}^{n} ||\boldsymbol{x}(t_i + \Delta t_i) - h_{\boldsymbol{\theta}}(\boldsymbol{x}(t_i))||^2$$

with $h_{\boldsymbol{\theta}}(\boldsymbol{x}(t_i))$ being the output of a numerical ODE solver (RK4).

In the initial deterministic pre-processing phase we use L^1 regularization to come up with realistic estimates to feed the Bayesian framework with:

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^{n} ||x_i - \hat{x}_i||^2 + \beta ||\boldsymbol{\theta}||$$

Bayesian Framework

The **Bayesian framework** presented in Yang, Bhouri, Perdikaris (2020) considers a **Lasso-like model** in order to perform effective model selection

Dynamical system $\quad \dot{\mathbf{x}} = f(\mathbf{x}, t; \boldsymbol{\theta}), \quad \mathbf{x} \in \mathbb{R}^D,\ t \in \mathbb{R},\ f : \mathbb{R}^D \to \mathbb{R}^D, \boldsymbol{\theta} \in \Theta$

- $p(\gamma, \lambda, \boldsymbol{\theta} | \mathbf{x}(t+\Delta t), \mathbf{X}(t)) \propto p(\mathbf{x}(t+\Delta t) | \mathbf{X}(t), \boldsymbol{\theta}, \gamma) p(\boldsymbol{\theta}|\lambda) p(\gamma) p(\lambda)$

  $\boldsymbol{\theta}$ parameters of the dynamical system
  γ discrepancy observed data - model's predictions
  λ usual shrinkage parameters for $\boldsymbol{\theta}$

- Sampling from the posterior is performed using **Hamiltonian Monte Carlo** in order to prevent possible issues when the number of parameters gets high

- Moment matrix for the HMC sampling has been selected to provide an optimal and independent sampling

The model

We first implemented the Lotka-Volterra (Prey-Predator model):

$$\begin{cases} \dot{x}_1 = \theta_1 x_1 + \theta_2 x_1 x_2 \\ \dot{x}_2 = \theta_3 x_1 x_2 + \theta_4 x_2 \end{cases}$$

using the suggested Lasso-like shrinkage model for the bayesian framework:

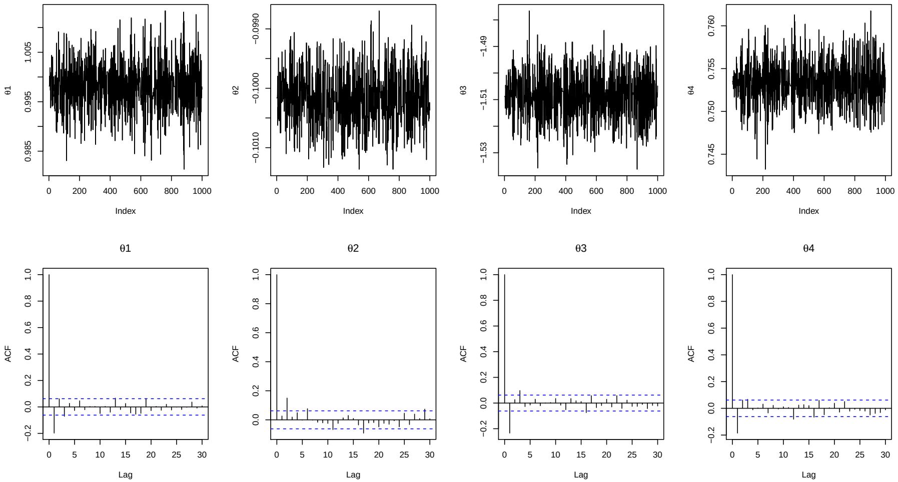
$$p(\boldsymbol{x}(t+\Delta t)|\boldsymbol{x}(t), \boldsymbol{\theta}, \gamma) = \prod_{i=1}^{N} \mathcal{N}(\boldsymbol{x}(t_i + \Delta t_i); h_{\boldsymbol{\theta}}, \gamma^{-1})$$

$$\mathrm{p}(\boldsymbol{\theta}|\lambda) = Laplace(\boldsymbol{\theta}; \lambda^{-1})$$

$$\mathrm{p}(\log\ \lambda) = Gamma(log\lambda; n_{par} + 1, 1)$$

$$\mathrm{p}(\log\ \gamma) = Gamma(log\gamma; 1, 1)$$

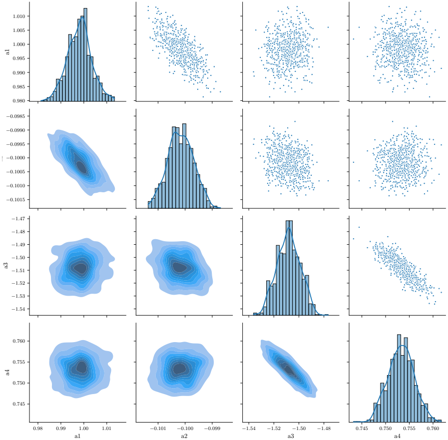
Traceplots of parameters



Effective sizes are: $\theta_1 = 1491.978$, $\theta_2 = 700.4196$, $\theta_3 = 1388.171$, $\theta_4 = 1151.698$

Pairplots of parameters

Pair plot shows a really good sampling from the parameters' space



- Good "cloudy" pattern
- Better results than in the reference paper
- Important correlation between variables ($R^2_{1,2} = 0.5$, $R^2_{3,4} = 0.735$)

Estimation of parameters

In order to estimate the parameters we used 4D kernel smoothing estimation and found the Maximum A Posteriori estimate:

$$\boldsymbol{\theta}_{\boldsymbol{MAP}} = arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}, \lambda, \gamma | \mathcal{D})$$

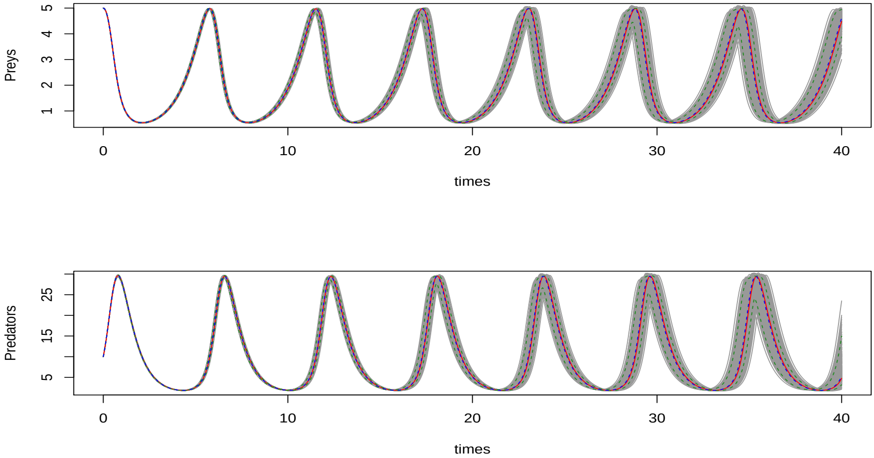
We found the following:

	θ_1	θ_2	θ_3	θ_4
True	1	-0.1	-1.5	0.75
MAP	0.9979098	-0.1003935	-1.506524	0.7530146
Rel. error	2.0902e-3	3.935e-3	4.349e-3	4.019e-3

The relative error is really small, allowing us to provide accurate estimates on the underlying differential model.

Estimation of the model

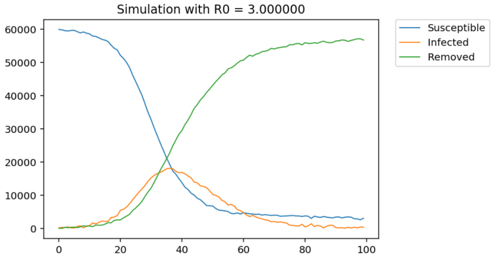
Simulations using HMC sampled parameters gave promising results



SIR model

Then, we directed our attention to the **SIR model**:

$$\begin{cases} \dot{S} = -\dfrac{\beta S I}{N} \\ \dot{I} = \dfrac{\beta S I}{N} - \gamma I \\ \dot{R} = \gamma I \end{cases}$$

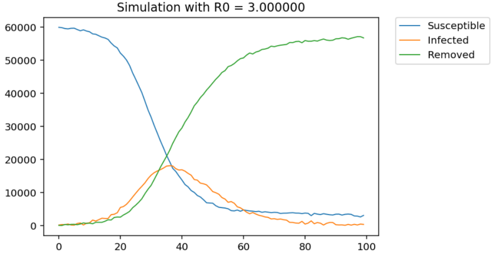


We first simulated the model with $N = 60000$, $\beta = 0.3$ and $\gamma = 0.1$, with initial conditions $S_0 = 59900$, $I_0 = 100$, $R_0 = 0$, adding Fractional Brownian Motion noise with $hurst = 0.2$.

SIR model

Then, we directed our attention to the **SIR model**:

$$\begin{cases} \dot{S} = -\frac{\theta_1 SI}{N} + \theta_5 R \\ \dot{I} = \frac{\theta_2 SI}{N} - \theta_3 I \\ \dot{R} = \theta_4 I - \theta_5 R \end{cases}$$



In this synthetic scenario, we tried to test the robustness in the model identification framework, adding a fake parameter θ_5 and distinguishing between θ_1, θ_2 and θ_3, θ_4.

Preliminary Results

After sampling with HMC, we look at the posteriors and highest density intervals of $\theta_1 - \theta_2$, $\theta_3 - \theta_4$ and θ_5.

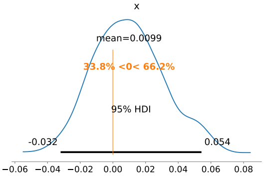


Figure: $\theta_1 - \theta_2$

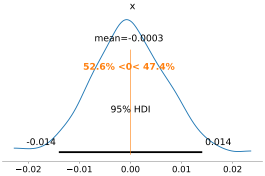


Figure: $\theta_3 - \theta_4$

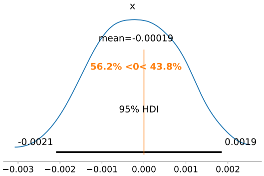


Figure: θ_5

We easily see they do contain 0 in a very central position, thus confirming good model identification properties.

We could then proceed by estimating only the relevant parameters without confounding factors.

Further developments

- Investigating the **robustness** of the model assumptions (e.g. changing priors, hyperparameters).
- Testing the model's performance in a **black-box** identification setting, where the number of parameters to estimate quickly grows.
- Exploring more complex differential models for epidemics, hopefully gathering **nCov-19** epidemiological data and testing the model's performance on **real-world data**.
- Extending the algorithm to deal with states which **cannot be observed** directly.

Thank You

Y. Yang, M.A. Bhouri, P. Perdikaris (2020). Bayesian differential programming for robust systems identification under uncertainty. ArXiv pre-print, submitted to Proceedings of the Royal Society A.

Hoffman MD, Gelman A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15, 1593–1623.

R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud. (2018). Neural ordinary differential equations. 32nd International Conference on Neural Information Processing Systems (NIPS'18).