Inference on epidemic models with time-varying parameters: methodology and preliminary applications

Joseph Dureau and Konstantinos Kalogeropoulos

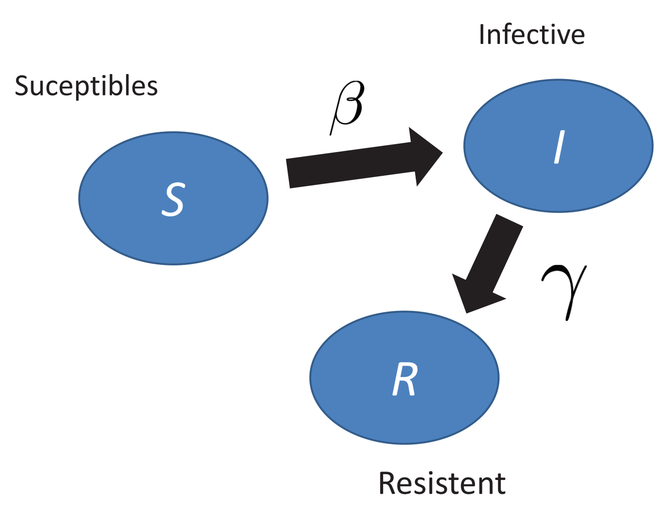
Department of Statistics, London School of Economics, UK

Email: j.dureau@lse.ac.uk

1 Why have time-varying parameters?

Inference on epidemic models is an active topic of research. Motivations are multiple: exploring mechanisms, testing theories, monitoring control interventions, surveilling upcoming epidemics.

A very typical compartmental model would be:



Whith the following definitions:

- S_t : proportion of the population that is susceptible, that can be infected
- I_t : proportion of the population that is infected and infective
- R_t : proportion of the population that is resistent, that is not infective any more and cannot be infected again
- β : transmission rate
- γ : recovery rate

Usually, inference is made for constant values of β and γ.

However, there are many reasons for β to be time-varying:

- Climate forcing is likely to have an impact on immunity and virus transmission
- Contact patterns evolve according to holidays, school/work periods, seasonal migrations,...
- Individual awareness to an epidemic can spontaneously decrease, or at the contrary increase under the influence of preventive measures
- etc...

2 How to model these time variations?

Fully parametric models for time-varying transmission rates have been explored:

- sinusoidal
- low-dimensional polynoms, splines,...

\+ tractable inference with classic MCMC algorithms
− limiting and arbitrary model choice

"Semi-parametric" models, on the other hand, have been used:

- random walk diffusion (Cazelles and Chau 1997, Mathematical Biosciences)

\+ very flexible model
− inference implied gaussian approximations (Extended Kalman Filter)

Our proposition

- use a diffusion process for β_t's trajectory, typically a geometric Brownian motion to preserve positivity
- apply novel MCMC algorithms to solve the inference problem, with low-informative priors on the diffusion coefficients

Classic SIR model:

$$\begin{cases} dS_t &= -\beta S_t I_t dt \\ dI_t &= (\beta S_t I_t - \gamma I_t)dt \\ dR_t &= \gamma I_t dt \end{cases}$$

Becomes $\Rightarrow\Rightarrow\Rightarrow$

Time-varying β model

$$\begin{cases} dS_t &= -\beta_t S_t I_t dt \\ dI_t &= (\beta_t S_t I_t - \gamma I_t)dt \\ dR_t &= \gamma I_t dt \\ d\log\beta_t &= \sigma_\beta dB_t \end{cases}$$

Going further...

- Try other diffusion processes (Ornstein-Uhlenbeck processes, integrated random walks, ...)
- Chose model from expert knowledge and/or indicators as the Bayes factor and the DIC.

3 A challenging inference problem

Objective

We want, under the following notations,

- X_t : dynamic vector of compartments populations
- θ : static parameters
- β_t : dynamic parameters
- $g(.|y)$: observation process model
- n : number of observations $(y_1, .., y_n)$
- N : number of particles

to explore the **posterior density** $p((\mathbf{X}_t, \beta_t, t \in [0,T]), \theta | y_{1:n})$.

Difficulties

- it is a high-dimensional density
- the posterior density and the Kolmogorov forward equation are intractable

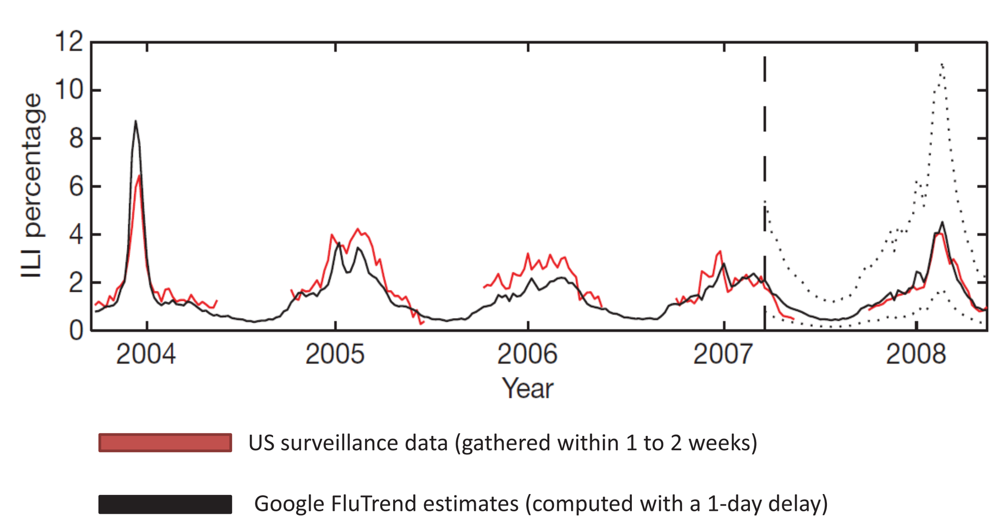
Estimating time-varying parameters with a Particle MCMC algorithm
(Andrieu et al. 2010, JRSS.B)

Initialize θ
Set $W_1^j = \frac{1}{N}$
for $IndIt = 1$ to $NbIterations$ **do**
  Sample θ^* from $Q(\theta, .)$
  $L(\theta^*) = 1$
  for $i = 1$ to $n-1$ **do**
    for $j = 1$ to N **do**
      Sample $(X_{i+1}^j, \beta_{i+1}^{\theta^*,j})$ from $p(.,.|X_i^j, \theta^*, \beta_i^{\theta^*,j})$
      Noting $Y_{i+1}^j = h((X_t^j, t \in [0, t_{i+1}]))$,
      set $\alpha^j = g(Y_{i+1}^j | y_t)$ and $W_{i+1}^j \propto \alpha^j$
    end for
    $L(\theta^*) = L(\theta^*) * (\sum_{j=1}^N W_i^j \alpha^j)$
    Resample $(X_{i+1}^j, \beta_{i+1}^{\theta^*,j})$ according to (W_{i+1}^j), set $W_{i+1}^j = \frac{1}{m}$
  end for
  Accept θ^* with probability $1 \wedge \frac{L(\theta^*)Q(\theta^*,\theta)}{L(\theta)Q(\theta,\theta^*)}$
  Sample j^{rand} from $1, .., NbParticules$
  Keep θ and $\beta_{1:n}^{\theta,j^{rand}}$
end for

4 Preliminary application: surveilling Influenza outbreaks from Google's FluTrend data

Google FluTrend Data: Estimates of Influenza-Like Illnesses cases

(Ginsberg et al. 2008, Nature)



A simple model for Influenza:

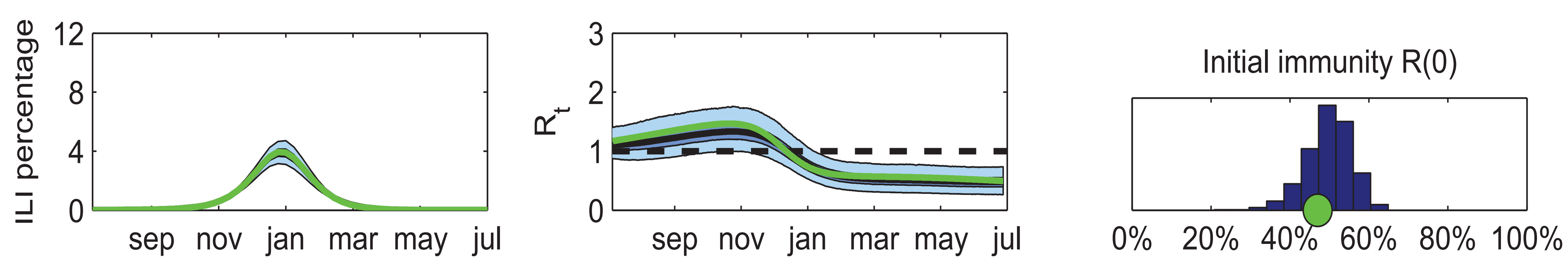
$$\begin{cases} dS_t &= -\beta_t S_t I_t dt \\ dE_t &= (\beta_t S_t I_t dt - kE_t)dt \\ dI_t &= (kE_t - \gamma I_t)dt \\ dR_t &= \gamma I_t dt \\ d\log\beta_t &= \sigma_\beta dB_t \\ g(.|y) &= \mathcal{N}(y, \sigma_{obs} y) \end{cases}$$

Note: E is the group of individuals who were infected but are not infectious yet. k^{-1} is the referred to as the latency period. Informative priors were taken for k and γ, based on bibliography.

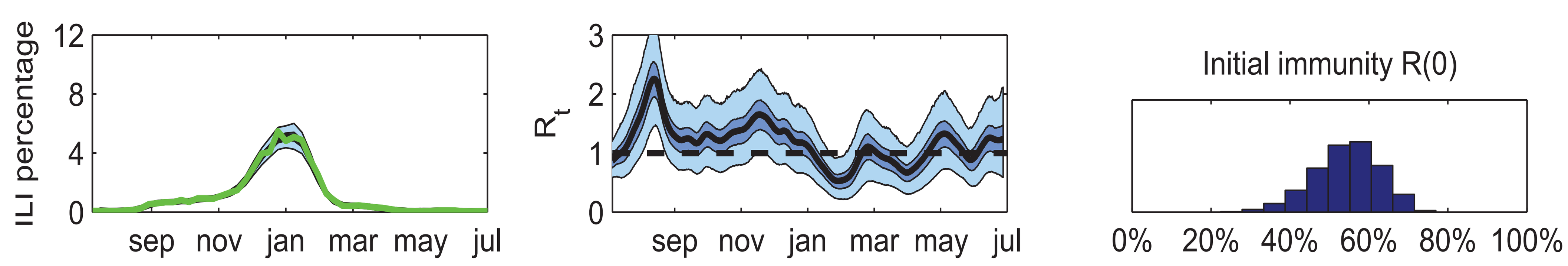
Questions:

- How transmittable is the upcoming strain of influenza?
- Does the effective reproduction rate $R_t = \frac{\beta_t S_t}{\gamma TotPop}$ vary along time?
- What is the population immunity to the upcoming strain of influenza?

a) Validating the algorithm on simulated data



b) 2008-2009 epidemic in France, a "classic" seasonal epidemic



c) 2009-2010 epidemic in France, the H1N1 pandemic