Approximation of epidemic models by diffusion processes and their statistical inferencedes

Catherine Larédo [1,2]

[1] UR 341, MIA, INRA, Jouy-en-Josas
[2] UMR 7599, LPMA, Université Paris Diderot

December 2, 2015

TALK 4, CIMPA 2015
Assessment of estimators on simulated and real data set

Based on Guy, Laredo, Vergu, JMB, 2015 and JSFDS (2016)

Simulation study

Assessment of the inference method

- Simulation of epidemics with jump Markov processes
- *SIR* epidemic model simulated with Gillepsie's algorithm (1977)
- *SIRS* with time-dependent transmission rate and demography simulated with the τ-leap method (Cao,2005)

Accuracy of estimators

- According to the population size N
- Number of observations n
- Parmeter values ruling the epidemic

Simulation scheme

- Choose an epidemic scenario: mechanistic model, population size, parameters
- Perform 1000 simulations of the Pure jump Markov Process associated with this scenario.

Reference

- Compute the M.L.E of the Jump Process assuming that **all the jumps are observed**
- Compute the Fisher information I_{PJM} of the Pure Jump Model.
- Reference: this MLE together with the associated confidence interval.

Remark: $I_{PJM} = I_b$, Fisher information of continuous observation of the diffusion.

Computation of estimators for each simulation

- Choose a sampling interval Δ and keep only the observations of the simulation at times $i\Delta$ (with realistic values of $\Delta \geq 1$).
- Compute our estimators on these discrete data (Point estimators)
- Compute the theoretical confidence intervals (CI_{th}) based on our inference method

Joining all the 1000 simulations results

- Compute the empirical confidence intervals (CI_{emp}) based on the 1000 simulations;
- Compute the average point estimators.

Remark: Only non extinct trajectories are kept;
Criterion: Final epidemic size larger than $5\% S0$; $\Rightarrow$ Possible bias?

SIR model

Basic reproduction number: $R_0 = \frac{\lambda}{\gamma}$
Average infectious duration: $d = \frac{1}{\gamma}$

Schéma de simulation:

Parameter	Description	Values
R_0	basic reproduction number	1.5, 3
d	infectious period	3, 7 days
$T^{(1)}$	final time of observation	20, 40, 45, 100 days
N	population size	400, 1000, 10000
n	number of observations	5, 10, 20, 40, 45, 100

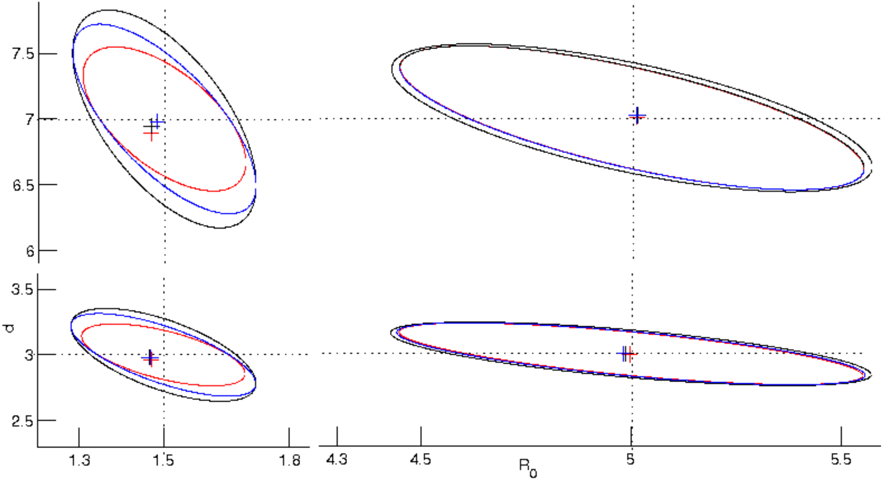
Table: Range of parameters for the *SIR* model defined in Section **??**. $^{(1)}$: T is chosen as the time point where the corresponding deterministic trajectory passes below the threshold of $1/100$.

Observations of all the jumps $\rightarrow$ MLE and theoretical confidence interval CI_{th}; for $\Delta = 1$, Contrast estimator and CI_{th}, $\Delta = T/10$

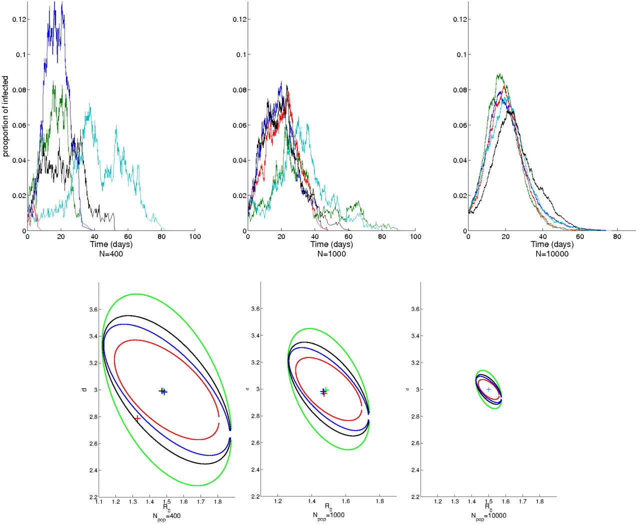
SIR theoretical confidence ellipsoids and estimators

Population size $N = 1000$; $R_0 = 1.5, 5$, $d = 3, 7$. Complete Obs.;$\Delta = 1$,$\Delta = T/10$

Average point estimator based on the 1000 simulations for (MLE et $\hat{\alpha}_{\epsilon,n}$)



Good results when varying the population size in the *SIR*



$R_0 = 1.5$; $d = 3$ (days); $T = 50$ (days); $\Delta = 1; 5; 10$ (days) and
$N = 4.10^2; 10^3; 10^4$

Results for small R_0 and large R_0

Results for a population size $N = 1000$, and various nb of obs. n

- Correlation between parameters: increases with d,decreases with R_0,
- Empirical CI: allways very tight $\Rightarrow$ not shown,
- Theoretical confidence intervals CI_{PJM} and CI_{th} for various samplings: very close
- No loss in estimation accuracy for $n = 40$ (1 obs/day) for large R_0.

Results when varying N

- Width of CI decreases with N; correlation not impacted.
- Given N, confidence ellipsoids are still very close, even for small n.
- $N = 400$ MLE biased while CE is OK.
- Very noisy sample paths.
- MLE optimal for "typical" realizations of the jump process.

SIRS model with seasonal forcing

Time-dependent transmission rate (Keeling and Rohani, 2011)

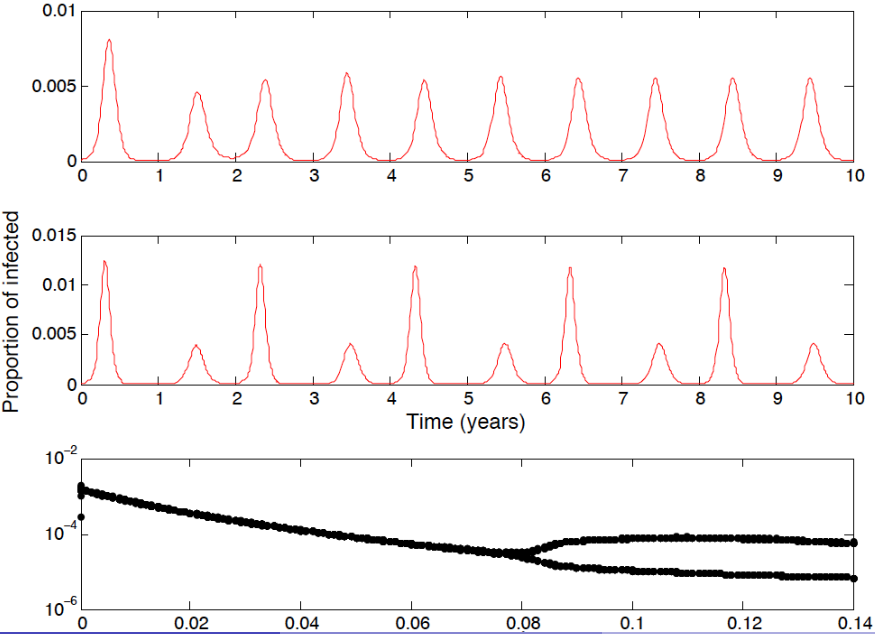
To avoid extinction, immigration flow η added $S \rightarrow I$: $\frac{\lambda(t)}{N} S(I + N\eta)$.

$\star$ μ: demography parameter; δ:immunity waning; $\gamma(= 1/d)$ recovery rate;
$\star$ $\lambda(t) = \lambda_0(1 + \lambda_1 sin(2\pi t/T_{per})) \Rightarrow$ New parameter $R := \frac{\lambda_0}{\gamma}$.
$\star$ $b(\theta; t, x) = \begin{pmatrix} -\lambda(t)s(i + \eta) + \delta(1 - s - i) + \mu(1 - s) \\ \lambda(t)s(i + \eta) - (\gamma + \mu)i \end{pmatrix}$.

ODE: dynamical system with bifurcation according to λ_1

$\star$ ODE: $\frac{dx}{dt} = b(\theta; t, x)$.
$\star$ Before bifurcation: annual oscillations with constant amplitude.
$\star$ After bifurcation: Biennial oscillations with unequal amplitudes.
$\star$ Bifurcation diagram w.r.t. λ_1.

Dynamics of the ODE



Parameter values chosen in the previous figure

- $N = 10^7$; $T_{per} = 365, \mu = 1/(50\,T_{per}), \eta = 10^{-6}$,
- $\lambda_0 = 0.5, \gamma = 1/3 \Rightarrow R_0 = 1.5, d = 3$; $\delta = 1/(2 \times 365)$,
- $(s_0, i_0) = (0.7, 10^{-4})$.
- Top panel: $\lambda_1 = 0.1$; middle panel: $\lambda_1 = 0.1$.
- Bottom panel: bifurcation diagram w.r.t. λ_1.

Choice of plausible values for modeling influenza seasonal outbreaks

⋆ Large population size: $N = 10^7$ to ensure

—> Sufficiently large "signal over noise" ratio.

—> Sufficient pool of S and I after each outbreak.

⋆ $\mu = 1/50$ years^{-1}; $T_{per} = 365$ (days), $\eta = 10^{-6}$

⋆ $R0 = 1.5$, $d = 3$, $\delta = 2 \Rightarrow$ bifurcation for $\lambda_1 = 0.07$;

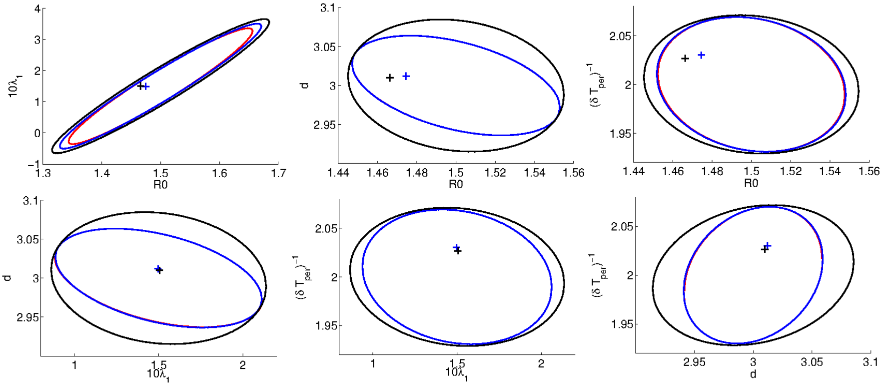
⋆ $\lambda_1 = 0.05$ and $\lambda_1 = 0.15$ (before and after bifurcation)

Numerically, these 2 scenarios have the characteristics of influenza seasonal outbreaks.

Simulation study: 1000 simulations of these 2 scenarios

- Known parameters: μ, T_{per}, η.
- Unknown parameters: R, d, λ_1, δ.
- Estimation of these parameters on each simulation.
- Results displayed with different projections of the 4-dimensional theoretical ellipsoid.

Estimation results and confidence ellipsoids for the *SIRS*



$R = 1.5; d = 3; \lambda_1 = 0.15, \delta = 2$ (days), $T = 20$ (years), $N = 10^7$.

Observations on $[0, T]$: Complete (MLE), $\Delta = 1$, $\Delta = 7$ (days).

Average point estimator and theoretical confidence ellipsoids.

Results for the *SIRS*

• Almost no correlation between parametres, except R_0 and λ_1

• Good accuracy of estimation for all parameters.

• Disposing one obs/day $\rightarrow$ accuracy identical to corresponding complet obs. of the epidemic process.

• One obs/per week $\rightarrow$ still reasonably accurate estimations