Approximation of epidemic models by diffusion processes and their statistical inferencedes

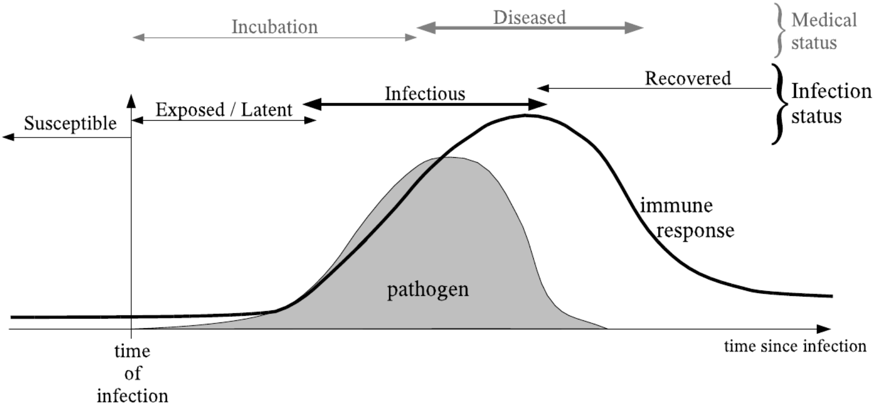
Catherine Larédo [1,2]

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

December 6, 2015

Based on Guy[1,2], Larédo, Vergu[1] (JMB 2014)

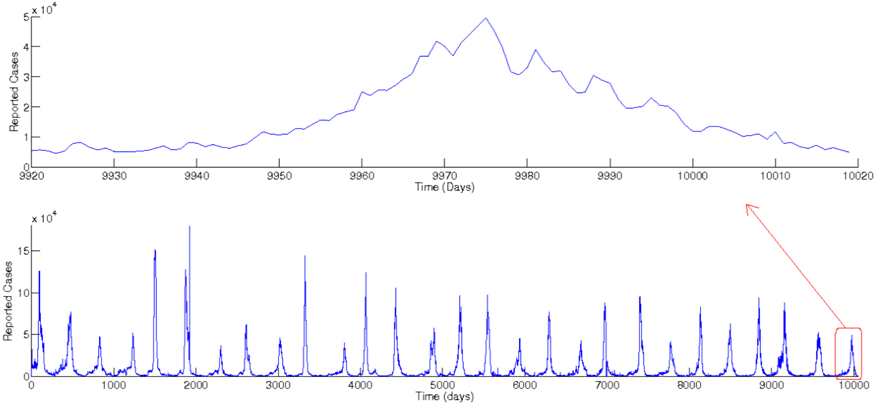
Context of infectious diseases (individual infection)



Modeling scheme for individual infection

- Compartmental models for describing the infection status : (S) Susceptible; (E) Exposed/Latent ; (I) Infectious/Infected ; (R) Removed.
- Difficulty in detecting the infection status ⇒ systematically noisy data.

Various dynamics at the ppopulation scale: (epidemics with one outbreak or recurrent outbreaks)



Influenza like illness cases in France ("Sentinelles" surveillance network)

Main important issues (1)

Determine the key parameters of the epidemic dynamics

- Basic reproduction number R_0 (average nb of secondary cases by one primary case in an entirely susceptible population)
- Average infectious time period d
- Latency period, etc...

Based on the available data

- Exact times of infection beginning and ending are not observed.
- Data are collected at fixed times (daily, weekly .. data)
- Temporally aggregated data.
- Sampling and reporting errors
- Some disease stages cannot be observed.

Main important issues (2)

Provide a common framework for developping estimation methods as accurate as possible given the data available

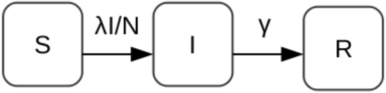
SEVERAL POSSIBLE WAYS TO OVERCOME THIS PROBLEM OF MISSING OR INCOMPLETE DATA

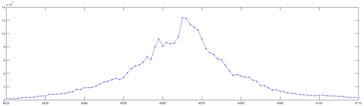
1. Develop algorithms to simulate the unobserved missing data.
2. Existence of lots of computer intensive methods in this domain.
3. Difficult to use for large populations.
4. Results are often unstable.

ANOTHER CHOICE HERE

1. Consider separately the model and the available data.
2. Study the properties of the observations derived from the model
3. Investigate inference based on these properties
4. Develop algorithms fast to implement in relation with the previous step

A simple mechanistic model for a single outbreak : *SIR*





S, I, R numbers of Susceptibles, Infected, Removed.
λ: transmission rate , γ: recovery rate.

Notations and assumptions:

- Closed population of size N ($\forall t,\ S(t) + I(t) + R(t) = N$).
- Homogeneous contacts in a well mixing population:
  $(S, I) \rightarrow (S-1, I+1)$ at rate $S\ \lambda\ \frac{I}{N}$
  $(S, I) \rightarrow (S, I-1)$ at rate $\gamma\ I$

Key parameters of this epidemic model

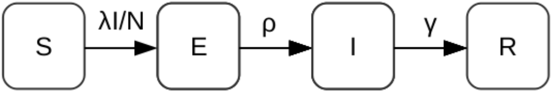
- Basic reproduction number $R_0 = \frac{\lambda}{\gamma}$.
- Average infectious period $d = \frac{1}{\gamma}$.

A minimal model for Ebola epidemics

- Explicit and detailed model in Legrand J., Grais R.F., Boelle P.Y. Valleron A.J. and Flahault A. (2007), Epidemiology & Infection
- Impossible to estimate parameters from available data.
- Due to identifiability problems.

A Minimal model for Ebola Transmission
Camacho et al. PLoS Curr, 2015;7.

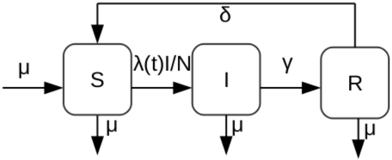
***SEIR* model with temporal transmission rate**

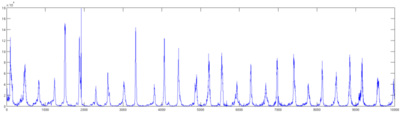


- $(S, E, I) \rightarrow (S-1, E+1, I)$ at rate $S\ \lambda(t)\frac{I}{N}$,
- $(S, E, I) \rightarrow (S, E-1, I+1)$ at rate $\rho\ E$
- $(S, E, I) \rightarrow (S, E, I-1)$ at rate $\gamma\ I$

A mechanistic model for recurrent epidemics

SIRS **model with seasonal forcing** (Keeling et Rohanni, 2011)





- δ: Immunity waning rate (per year)$^{-1}$,
- μ (Population renewal): birth rate and death rate (per decades)$^{-1}$
- $\lambda(t) = \lambda_0(1 + \lambda_1 cos(2\pi \frac{t}{T_{per}}))$,
- λ_0 Baseline transition rate, λ_1 :intensity of the seasonal effect,
- T_{per}: period of the seasonal trend.

Important: $\lambda_1 = 0 \Rightarrow$ Damping out oscillations $\Rightarrow$ need to have a temporal forcing

Appropriate model for recurrent epidemics in very large populations

Key parameters: $R_0 = \frac{\lambda_0}{\gamma+\mu}$, $d = \frac{1}{\gamma}$,Average waning period: $\frac{1}{\delta T_{per}}$

Recap on some mathematical approaches

Description

- p: number of health states (i.e. compartments).
- Depends on the model choice for describing the epidemic dynamics.
- *SIR* and *SIRS* models: $p = 3$.
- Adding a state "Exposed/latent" $\Rightarrow$ *SEIR* model: $p = 4$.
- Addition of states where individuals have similar behaviour with respect to the pathogen: age, vaccination, structured populations).

Some classical mathematical models

- Pure jump Markovprocess with state space $\mathbb{N}^p$: $Z(t)$.
- Deterministic models satifying an ODE on $\mathbb{R}^p$: $x(t)$.
- Gaussian Process with values in $\mathbb{R}^p$: $G(t)$.
- Diffusion process $X(t)$ satisfying a SDE on$\mathbb{R}^p$:

Links between these models?

Pure jump p- dimensional Markov process $Z(t)$

Simple and natural modelling of epidemics.

Notations

- Population size $N \Rightarrow Z(t) \in E = \{0, .., N\}^p$.
- Jumps of $Z(t)$: collection de functions $\alpha_\ell(\cdot) : E \to (0, +\infty)$, indexed by $\ell \in E^- = \{-N, .., N\}^p$
- For all $x \in E$, $0 < \sum_\ell \alpha_\ell(x) := \alpha(x) < \infty$.

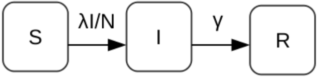
Pure jump Markov Process with state space de E: $Z(t)$

- $\Rightarrow$ Transition rate from $x \to x + \ell$: $\alpha_\ell(x)$,
- Q-matrix of $(Z(t))$: $Q = (q_{xy}, (x, y) \in E \times E)$
  if $y \neq x$, $q_{xy} = \alpha_{y-x}(x)$, and $q_{xx} = -\alpha(x)$.

$\star$ Each individual stays in state x with exponential holding time $\mathcal{E}(\alpha(x))$,
$\star$ Then, it jumps to another state according to a Markov chain with transition kernel $\mathbb{P}(x \to x + \ell) = \frac{\alpha_\ell(x)}{\alpha(x)}$.

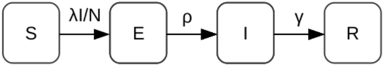
SIR, *SEIR* models in a closed population of size N

SIR epidemic model
$\forall t,\ S(t)+I(t)+R(t)=N \Rightarrow \Rightarrow Z(t)=(S(t),I(t)) \in E=\{0,\dots,N\}^2$
and $\ell = (-1,1),(0,-1)$



$(S,I)\to(S-1,I+1)$: $\alpha_{(-1,1)}(S,I)=\lambda S\frac{I}{N}$,
$(S,I)\to(S,I-1)$: $\alpha_{(0,-1)}(S,I)=\gamma I$

SEIR epidemic model:(Time-dependent process)
$Z(t)=(S(t),E(t),I(t)) \in E=\{0,\dots,N\}^3$.

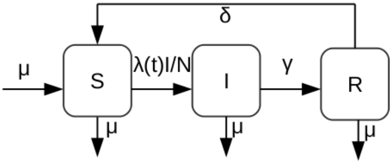


$\ell = (-1,1,0),(0,-1,1),(0,0,-1)$.

$(S,E,I)\to(S-1,E+1,I)$: $\alpha_{(-1,1,0)}(S,E,I)=\lambda(t)S\ \frac{I}{N}$,
$(S,E,I)\to(S,E-1,I+1)$: $\alpha_{(0,-1,1)}(S,E,I)=\rho I$,
$(S,E,I)\to(S,E,I-1)$: $\alpha_{(0,0,-1)}(S,E,I)=\rho I$

SIRS model with seasonal forcing in a closed population

$\Rightarrow Z(t) = (S(t), I(t)) \in E = \{0, \ldots, N\}^2$,
$\ell = (-1, 1), (0, -1), (1, 0), (-1, 0)$.



$(S, I) \rightarrow (S-1, I+1) : \alpha_{(-1,1)}(t; S, I) = \lambda(t) S \, \frac{I}{N}$,
$(S, I) \rightarrow (S, I-1) : \alpha_{(0,-1)}(S, I) = (\gamma + \mu) I$,
$(S, I) \rightarrow (S+1, I) : \alpha_{(1,0)}(S, I) = \mu N + \delta(N - S - I)$,
$(S, I) \rightarrow (S-1, I) : \alpha_{(-1,0)}(S, I) = \mu S$.
$\lambda(t) = \lambda_0(1 + \lambda_1 sin(2\pi \frac{t}{T_{per}})) \Rightarrow$ Time-inhomogeneous Markov process.

Remark: Simulations are easy with Gillespie's algorithm

Density dependent jump processes $Z(t)$ (1)

Extension of results Ethier et Kurz (2005) to time-dependent processes.

Notations

$\star$ if $y = (y_1, \ldots, y_p) \in \mathbb{R}^p, [y] = ([y_1], \ldots, [y_p])$, with $[y_i]$ integer part of y_i.
Transposition of a vector y or a matrix M: ${}^t y$, ${}^t M$.
$\star$ Gradient de $b(.) \in C(\mathbb{R}^p, \mathbb{R}^p)$: $\nabla b(y) = (\frac{\partial b_i}{\partial y_j}(y))_{ij}$.

Framework

$\star$ Constant population size $N \Rightarrow E = \{0, .., N\}^p$.
$\star$ Collection $\alpha_\ell(\cdot) : E \to (0, \infty)$ with $\ell \in E^- = \{-N, .., N\}^p$.
$\star$ Transition rates: $y \to y + \ell$: $\alpha_\ell(y) \Rightarrow q_{yz} = \alpha_{z-y}(y)$.

Normalization by the size N of $Z(t)$

- $Z_N(t) = \frac{Z(t)}{N} \Rightarrow$
- State space: $E_N = \{N^{-1}k, k \in E\}$;
- $Z_N(t)$ jump process on E_N with Q-matrix: if $x, y \in E_N, y \neq x$, $q^{(N)}_{xy} = \alpha_{N(y-x)}(x)$.

Density dependent jump processes $Z(t)$:(2)

Density dependent Process

Recall that the jumps ℓ of $Z(t)$ belong to $E^- = \{-N, .., N\}^p$.

Assumption (H):
(H1): $\forall (\ell, y) \in E^- \times [0,1],\ \frac{1}{N}\alpha_\ell([Ny]) \to \beta_\ell(y)$.
(H2): $\forall \ell,\ \ y \to \beta_\ell(y) \in C^2([0,1]^p)$.

Definition of the key quantities $b(.)$ and $\Sigma(.)$

$$b(y) = \sum_\ell \ell \beta_\ell(y), \quad \Sigma(y) = \sum_\ell \ \ell\, {}^t\ell\ \beta_\ell(y).$$

These quantities are wel defined since the number of jumps is finite.
Note that $b(y) = (b_k(y), 1 \le k \le p) \in \mathbb{R}^p$ and
$\Sigma(y) = (\Sigma_{kl}(y), 1 \le k, l \le p)$ is a p-dimensional matrix.

Approximations of $Z_N(.)$

$x(t) = x_0 + \int_0^t b(x(s))ds.$
$\Phi(t,u)$ solution de $\frac{\partial \Phi}{\partial t}(t,u) = \nabla b(x(t))\Phi(t,u)$; $\Phi(u,u) = I_p$.

Convergence Theorem

Assume (H1),(H2), and that $Z_N(0) \to x_0$ as $N \to \infty$.Then,
$\star$ $Z_N(.)$ converges $x(.)$ uniformly on $[0,T]$,
$\star$ $\sqrt{N}(Z_N(t) - x(t))$ converges in distribution to $G(t)$,
$\star$ $G(t)$ centered Gaussian process with
$Cov(G(t), G(r)) = \int_0^{t \wedge r} \Phi(t,u)\Sigma(x(u))\, {}^t\Phi(r,u)du.$

Proof: Ethier & Kurz (2005): $\alpha_l([Ny]) \equiv \beta_l(y)$; GLV (2014) for
(i) Jump rates $\alpha_l(.)$ satisfying (H)
(ii) Time-dependent jump rates $\alpha_l(t,x)$ with
$\frac{1}{N}\alpha_l(t,[Ny]) \to \beta_l(t,y) \Rightarrow b(t,y); \Sigma(t,y)$
(Proof based on general limit theorems (Jacod and Shiryaev)).

Diffusion approximation of $Z_N(t)$

Recap: $b(y) = \sum_\ell \ell \, \beta_\ell(y)$ and $\Sigma(y) := \sum_\ell \, \ell \, {}^t\ell \beta_\ell(y)$.
Let $f : \mathbb{R}^p \to \mathbb{R}$ a bounded measurable and $\mathcal{A}$ the generator of $Z(t)$

- $\mathcal{A}(f)(y) = \sum_\ell \alpha_\ell(y)(f(y+\ell) - f(y))$
- $\Rightarrow \mathcal{A}_N(f)(y) = \sum_\ell \alpha_\ell(Ny)(f(y + \frac{\ell}{N}) - f(y))$

Euristically : Expanding the generator $\mathcal{A}_N$ of $Z_N(.) = \frac{Z(t)}{N}$.
$\mathcal{A}_N(f)(y) = b(y)\nabla f(y) + \frac{1}{2N} \sum_{i,j=1}^p \Sigma_{ij}(y) \frac{\partial^2 f}{\partial y_i \partial y_j}(y) + o(1/N)$
$\Rightarrow \mathcal{A}_N(f)(y) = B_N(f)(y) + o(1/N)$.

Diffusion approximation of $Z_N(t)$: diffusion with generator B_N

$dX_N(t) = b(X_N(t))dt + \frac{1}{\sqrt{N}} \sigma(X_N(t))dB(t)$,
$B(t)$: p-dimensionnal Brownian motion et $\sigma(.)$ a square root of Σ ()
$\sigma(y) \, {}^t\sigma(y) = \Sigma(y)$.

Small pertubations of Dynamical systems

Freidlin & Wentzell (1978) ; Azencott (1982)

$$\epsilon = 1/\sqrt{N} \Rightarrow X_N(t) = X_\epsilon(t)$$

Link between the CLT for $(Z_N(.))$ and diffusion $(X_N(.)$

Expanding $X_\epsilon(t)$ with respect to ϵ,

- $X_\epsilon(t) = x(t) + \epsilon g(t) + \epsilon R_\epsilon(t)$,
- where $dg(t) = \nabla b(x(t))g(t)dt + \sigma(x(t))dB(t) \;\; ; \;\; g(0) = 0$,
- $sup_{t \leq T} \parallel \epsilon R_\epsilon(t) \parallel \rightarrow 0$ in probability as $\epsilon \rightarrow 0$.

Explicit solution of this stochastic differential equation

$\star$ $g(t) = \int_0^t \Phi(t,s)\sigma(x(s))dB(s)$, where

$\star$ $\Phi(t,u)$ s.t. $\frac{\partial \Phi}{\partial t}(t,u) = \nabla b(x(t))\Phi(t,u)$; $\Phi(u,u) = I_p$.

$\star$ $g(.)$: centered Gaussian process with same covariance matrix as $G(.)$.