

Estimation du paramètre d'infection dans les différentes phases d'une épidémie modélisée par un processus de branchement

Sophie Péni^sson
(Université Paris-Est Créteil, France)

en collaboration avec Christine Jacob (INRA)

Rencontre de l'ANR MANEGE
École Polytechnique, Paris

14 novembre 2012

Introduction

Contexte

- maladie infectieuse modélisée par un processus de branchement multitype
- tous les paramètres du modèle sont supposés connus, hormis le paramètre "quantifiant" l'infection

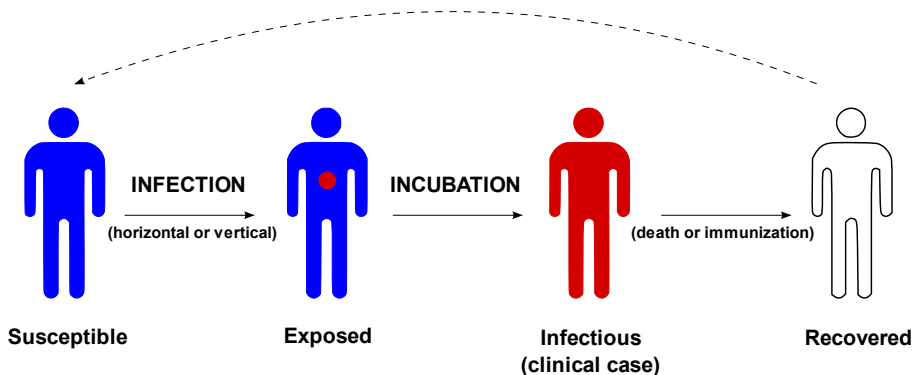
Objectif = estimation du paramètre d'infection

- peu d'estimateurs applicables à ce contexte dans la littérature sur les processus de branchement
- nous cherchons à construire des estimateurs adaptés aux différentes phases de l'épidémie (sur- ou sous-criticalité) et pour différentes propriétés asymptotiques (lorsque le temps ou le nombre initial d'individus tend vers l'infini)

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Le modèle épidémique

Maladie infectieuse générale de type SEIR



Puisque les états de santé S and E sont souvent indifférenciables, nous construisons un modèle stochastique basé uniquement sur l'incidence des cas cliniques (état I).

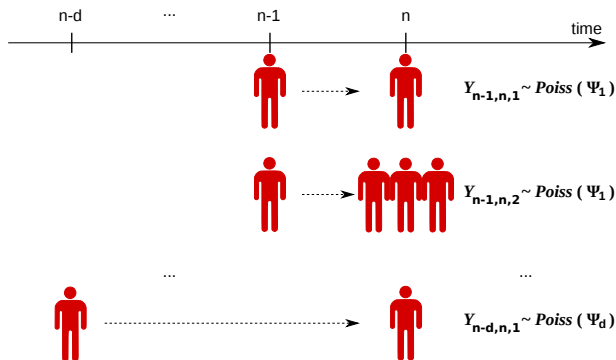
Le modèle épidémique

X_n = incidence de cas au temps n

$$X_n = \sum_{k=1}^d \sum_{i=1}^{X_{n-k}} Y_{n-k,n,i}$$

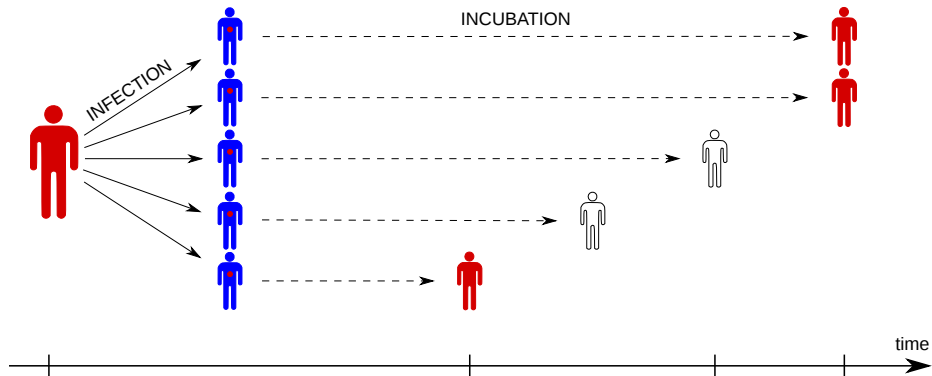
$$\mathcal{L}(Y_{n-k,n,i}) = \text{Poisson}(\Psi_k)$$

$\{Y_{n-k,n,i}\}_i$ i.i.d. et $\{Y_{n-k,n,i}\}$ indépendants de $\{Y_{n-k',n,i}\}$ pour $k \neq k'$



Justification intuitive du modèle

Production "indirecte" de nouveaux cas



Justification intuitive du modèle

Le modèle $(X_n)_n$ peut être obtenu comme limite d'un processus de branchement multitype âge- et taille de population-dépendant prenant en compte **tous les états de santé** et modélisant la variabilité de nombreux facteurs individuels tels que la **reproduction**, la **survie**, la **transmission** de la maladie et la période d'**incubation**.

Par exemple, en notant N_n (resp. E_n , I_n) le nombre total d'individus (resp. l'incidence des individus infectés et des cas cliniques) à l'instant n ,

$$E_n = \sum_{i=1}^{N_{n-1}} \underbrace{\delta_{n-1,n,i}(E)}_{\substack{\text{infection au temps } n \\ \text{de l'individu } i}} + \sum_{i=1}^{N_{n-1}} \sum_{j=1}^{Z_{n,i}} \underbrace{\delta_{n-1,n,i,j}(E)}_{\substack{\text{infection au temps } n \\ \text{du nouveau-né } j \text{ de } i}}$$

$$I_n = \sum_{k=1}^{a_{\max}-1} \sum_{i=1}^{E_{n-k}} \underbrace{\delta_{n-k,n,i}^E(I)}_{\substack{\text{incubation de } i \\ \text{du temps } n-k \text{ à } n}}$$

où a_{\max} est l'âge maximal de survie, $Z_{n,i}$ correspond aux nouveaux-nés de i au temps n , et où les δ_{\dots} sont des variables de Bernoulli valant 1 si l'action correspondante est réalisée.

Justification intuitive du modèle

N_n (resp. E_n , I_n) = nombre total d'individus (resp. incidence des individus infectés et des cas cliniques) à l'instant n ,

$$E_n = \sum_{i=1}^{N_{n-1}} \underbrace{\delta_{n-1,n,i}(E)}_{\text{infection au temps } n \text{ de l'individu } i} + \sum_{i=1}^{N_{n-1}} \sum_{j=1}^{Z_{n,i}} \underbrace{\delta_{n-1,n,i,j}(E)}_{\text{infection au temps } n \text{ du nouveau-né } j \text{ de } i}$$
$$I_n = \sum_{k=1}^{a_m-1} \sum_{i=1}^{E_{n-k}} \underbrace{\delta_{n-k,n,i}^E(I)}_{\text{incubation de } i \text{ du temps } n-k \text{ à } n}$$

En supposant principalement que

- la maladie est rare à l'instant initial,
- l'infection horizontale est de type Reed-Frost : probabilité pour un individu S d'être infecté à l'instant n est $1 - \left(1 - \frac{\theta}{N_{n-1}}\right)^{I_n}$,

alors on obtient $(X_n)_n$ comme limite de $(I_n)_n$ lorsque $N_0 \rightarrow \infty$.

Le modèle épidémique vu comme un processus de branchement multitype

$$\mathcal{L}(X_n | (X_{n-1}, \dots, X_{n-d})) = \text{Poisson}\left(\sum_{k=1}^d X_{n-k} \Psi_k\right)$$

$(X_n)_n$ processus d -Markovien 1-dimensionnel

Définissons $\mathbf{X}_n = (X_n, X_{n-1}, \dots, X_{n-d+1})$ et $\Psi = (\Psi_1, \dots, \Psi_d)$.

$$\mathcal{L}(\mathbf{X}_n | \mathbf{X}_{n-1}) = (\text{Poisson}(\mathbf{X}_{n-1} \cdot \Psi), X_{n-1}, \dots, X_{n-d+1})$$

$(\mathbf{X}_n)_n$ processus 1-Markovien d -dimensionnel

Proposition

$(\mathbf{X}_n)_n$ est un processus de branchement multitype, où les d types correspondent à la "mémoire" de $(X_n)_n$.

Le modèle épidémique vu comme un processus de branchement multitype

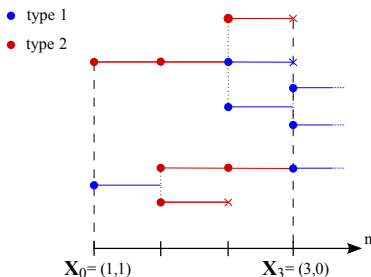
$$\mathbf{X}_n = (X_n, X_{n-1}, \dots, X_{n-d+1})$$

$$\mathcal{L}(\mathbf{X}_n | \mathbf{X}_{n-1}) = (\mathcal{Poisson}(\mathbf{X}_{n-1} \cdot \Psi), X_{n-1}, \dots, X_{n-d+1})$$

$(\mathbf{X}_n)_n$ est un processus de branchement multitype, où les d types correspondent à la "mémoire" de $(X_n)_n$.

Rappel : processus de branchement multitype général

- d types d'individus
- chaque individu produit indépendamment des autres un nombre aléatoire de descendants de tous types, selon une loi de probabilité ne dépendant que de son propre type
- $\mathbf{X}_n = (X_{n,1}, \dots, X_{n,d})$ composition de la population à l'instant n



Le modèle épidémique vu comme un processus de branchement multitype

Soit f_i la fonction génératrice de la loi de reproduction d'un individu de type i : pour tout $\mathbf{r} \in [0, 1]^d$,

$$f_i(\mathbf{r}) = \mathbb{E} \left(\mathbf{r}^{\mathbf{X}_1} \mid \mathbf{X}_0 = \mathbf{e}_i \right) = \mathbb{E} \left(r_1^{X_{1,1}} \dots r_d^{X_{1,d}} \mid \mathbf{X}_0 = \mathbf{e}_i \right),$$

et soit \mathbf{M} la matrice moyenne, où m_{ij} est le nombre moyen d'individus de type j produit par un individu de type i . Alors

$$\begin{cases} f_i(\mathbf{r}) := e^{-(1-r_1)\Psi_i} r_{i+1}, & i = 1 \dots d-1 \\ f_d(\mathbf{r}) := e^{-(1-r_1)\Psi_d} \end{cases}$$

$$\mathbf{M} = \begin{pmatrix} \Psi_1 & 1 & 0 & \dots & 0 \\ \Psi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \Psi_{d-1} & 0 & \dots & \dots & 1 \\ \Psi_d & 0 & \dots & \dots & 0 \end{pmatrix}$$

Le modèle épidémique vu comme un processus de branchement multitype

La racine de Perron ρ_0 de \mathbf{M} déterminant la classe de criticalité du processus de branchement et le taux de croissance asymptotique de $\mathbb{E}(\mathbf{X}_n)$ n'est pas calculable explicitement.

En revanche, puisque $\sum_{k=1}^d \Psi_k \rho_0^{-k} = 1$ on déduit que ρ_0 et $R_0 = \sum_{k=1}^d \Psi_k$ satisfont la même propriété de seuil, et donc que

Proposition

$R_0 < 1 \iff$	<i>le processus $\{\mathbf{X}_n\}_n$ est sous-critique</i>
$R_0 = 1 \iff$	<i>critique</i>
$R_0 > 1 \iff$	<i>surcritique</i>

R_0 correspond au nombre moyen total de cas secondaires générés en tout temps par un infectieux, et est donc une généralisation du taux de reproduction de base.

Le paramètre d'infection

Dans ce qui suit, on suppose que les Ψ_k , $k = 1 \dots d$, dépendent de façon affine d'un unique paramètre inconnu θ_0 :

$$\Psi_k(\theta_0) = a_k \theta_0 + b_k,$$

où $a_k > 0$ et $b_k \geq 0$.

Ceci correspond par exemple au cas où $(\mathbf{X}_n)_n$ est obtenu comme limite d'un processus prenant en compte tous les états de santé. On a alors

$$\Psi_k(\theta_0) = P_{inc}(k) \sum_{a=k+1}^{a_m} P_{age}(a) \theta_0 + p_{mat} P_{inc}(k) P_{age}(k+1),$$

où θ_0 est le nombre moyen par infectieux et par unité de temps de nouveaux individus infectés, que l'on nommera **paramètre d'infection**.

Remarque

Le paramètre d'infection est une fonction explicite de la racine de Perron :

$$\theta_0 = \frac{1 - \sum_{k=1}^d b_k \rho_0^{-k}}{\sum_{k=1}^d a_k \rho_0^{-k}}.$$

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Quelques estimateurs de ρ_0 dans la littérature

- **Cas monotype** $d = 1$ [Harris, 1948]

Estimateur maximum de vraisemblance de m_0 ($\rho_0 = a\theta_0 + b$)

$$\hat{m}_n^{MLE} = \frac{X_1 + \dots + X_n}{X_0 + \dots + X_{n-1}}$$

- ▶ Consistance lorsque $n \rightarrow \infty$ dans le cas surcritique sur l'ensemble de non-extinction.

Quelques estimateurs de ρ_0 dans la littérature

- **Cas monotype** $d = 1$ [Harris, 1948]

$$\hat{m}_n^{MLE} = \frac{X_1 + \dots + X_n}{X_0 + \dots + X_{n-1}}$$

- **Cas multitype** $d \geq 1$ [Asmussen & Keiding, 1978]

Estimateur maximum de vraisemblance basé sur l'observation de toutes les combinaisons parents-enfants

$$\hat{\rho}_n^{MLE} = \rho(\hat{\mathbf{M}})$$

$$\hat{m}_{ij} = \frac{Z_1^i(j) + \dots + Z_n^i(j)}{Z_0^i(j) + \dots + Z_{n-1}^i(j)}$$

où $Z_k^i(j)$ est le nombre d'individus à la génération k de type j dont les parents sont de type i

- ▶ Consistance et normalité asymptotique lorsque $n \rightarrow \infty$ dans le cas surcritique sur l'ensemble de non-extinction

Quelques estimateurs de ρ_0 dans la littérature

- **Cas monotype** $d = 1$ [Harris, 1948]

$$\hat{m}_n^{MLE} = \frac{X_1 + \dots + X_n}{X_0 + \dots + X_{n-1}}$$

- **Cas multitype** $d \geq 1$ [Asmussen & Keiding, 1978]



$$\hat{\rho}_n^{MLE} = \rho(\hat{\mathbf{M}}), \quad \hat{m}_{ij} = \frac{Z_1^i(j) + \dots + Z_n^i(j)}{Z_0^i(j) + \dots + Z_{n-1}^i(j)}$$



$$\hat{\rho}_n = \frac{|\mathbf{X}_1| + \dots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \dots + |\mathbf{X}_{n-1}|}$$

où $|\mathbf{X}_k| = X_{k,1} + \dots + X_{k,d}$.

Consistance et normalité asymptotique lorsque $n \rightarrow \infty$ dans le cas surcritique sur l'ensemble de non-extinction.

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Phase de croissance, $n \rightarrow \infty$

On déduit de $\hat{\rho}_n = \frac{|\mathbf{X}_1| + \dots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \dots + |\mathbf{X}_{n-1}|}$ et de ses propriétés l'estimateur suivant

$$\hat{\theta}_n := \frac{1 - \sum_{k=1}^d b_k (\hat{\rho}_n)^{-k}}{\sum_{k=1}^d a_k (\hat{\rho}_n)^{-k}}$$

qui satisfait :

Théorème

Supposons que $(\mathbf{X}_n)_n$ est surcritique. Alors, sur l'ensemble de non-extinction, $\hat{\theta}_n$ est fortement consistant

$$\lim_{n \rightarrow \infty} \hat{\theta}_n \stackrel{a.s.}{=} \theta_0$$

et asymptotiquement normalement distribué.

Le comportement asymptotique de $\hat{\theta}_n - \theta_0$ dépend qualitativement des tailles relatives de ρ_0 et de λ^2 , où λ est une valeur propre de \mathbf{M}_0 , choisie en fonction de la décomposition canonique de Jordan.

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$

On notera \mathbf{a} et \mathbf{b} les vecteurs d -dimensionnels de coordonnées a_k et b_k , par $\mathbf{x} \cdot \mathbf{y}$ le produit scalaire usuel et par $|\cdot|$ la norme L^1 .

Considérons l'estimateur des moindres carrés conditionnel (CLSE) basé sur le processus $X_k / \sqrt{\mathbf{a} \cdot \mathbf{X}_{k-1}}$:

$$\hat{\theta}_{|\mathbf{X}_0|} = \arg \min_{\theta \in \Theta} \sum_{k=1}^n \frac{[X_k - \mathbb{E}_{\theta}(X_k | \mathbf{X}_{k-1})]^2}{\mathbf{a} \cdot \mathbf{X}_{k-1}} = \frac{\sum_{k=1}^n (X_k - \mathbf{b} \cdot \mathbf{X}_{k-1})}{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}.$$

La forme explicite provient du fait que $\mathbb{E}_{\theta}(X_k | \mathbf{X}_{k-1}) = \boldsymbol{\psi}(\theta) \cdot \mathbf{X}_{k-1}$.

Remarques :

- dans le cas linéaire $\mathbf{b} = \mathbf{0}$, le CLSE $\hat{\theta}_{|\mathbf{X}_0|}$ correspond également au MLE basé sur les observations X_0, \dots, X_n
- la variance conditionnelle du terme d'erreur est bornée

Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$

Théorème

Supposons que $\forall i = 1 \dots d$, $\lim_{|\mathbf{x}_0| \rightarrow \infty} X_{-i+1} |\mathbf{X}_0|^{-1} = \alpha_i \in [0, 1]$. Alors, pour toute classe de criticalité, $\hat{\theta}_{|\mathbf{x}_0|}$ est fortement consistant et asymptotiquement normalement distribué :

$$\lim_{|\mathbf{x}_0| \rightarrow \infty} \hat{\theta}_{|\mathbf{x}_0|} \stackrel{a.s.}{=} \theta_0,$$

$$\lim_{|\mathbf{x}_0| \rightarrow \infty} \sqrt{\frac{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}{\sigma^2(\hat{\theta}_{|\mathbf{x}_0|})}} \left(\hat{\theta}_{|\mathbf{x}_0|} - \theta_0 \right) \stackrel{D}{=} \mathcal{N}(0, 1).$$

Avec

$$\sigma^2(\theta) := \theta + \frac{\sum_{k=1}^n \sum_{i,j=1 \dots d} \alpha_j b_i m_{ji}^{(k-1)}(\theta)}{\sum_{k=1}^n \sum_{i,j=1 \dots d} \alpha_j a_i m_{ji}^{(k-1)}(\theta)},$$

et où $m_{ij}^{(k)}(\theta)$ est la (i, j) -ème entrée de la k -ème puissance de la matrice $\mathbf{M}(\theta)$.

Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$

- **Consistance forte.** Propriété de branchement du processus $(\mathbf{X}_n)_n$:

$$X_{k,i} = \sum_{j=1}^{X_{0,1}} X_{k,i,j}^{(1)} + \dots + \sum_{j=1}^{X_{0,d}} X_{k,i,j}^{(d)},$$

où $X_{k,i,j}^{(l)}$ est la i -ème coordonnée d'un BP multitype à l'instant k initialisé par une particule de type l . LFGN :

$$\lim_{|\mathbf{X}_0| \rightarrow \infty} \frac{\sum_{j=1}^{X_{0,l}} X_{k,i,j}^{(l)}}{X_{0,l}} \stackrel{\text{a.s.}}{=} m_{li}^{(k)}(\theta_0).$$

- **Normalité asymptotique.**

$$\sqrt{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}} \left(\hat{\theta}_{|\mathbf{X}_0|} - \theta_0 \right) = \frac{\sum_{i=1}^d \sum_{j=1}^{\sum_{k=1}^n X_{k-i}} \dot{Y}_{k-i,k,j}}{\sqrt{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}}.$$

où les $\{\overset{\circ}{\zeta}_{k-i,k,j}\}_j$ sont i.i.d. et les $\{\overset{\circ}{\zeta}_{k-i,k,j}\}_{i,j}$ indépendants. TCL pour une somme d'un nombre aléatoire de v.a. indépendantes

$$\lim_{|\mathbf{X}_0| \rightarrow \infty} \frac{\sum_{j=1}^{\sum_{k=1}^n X_{k-i}} \overset{\circ}{\zeta}_{k-i,k,j}}{\sqrt{\sum_{k=1}^n X_{k-i}}} \stackrel{\mathcal{D}}{=} \mathcal{N}(0, a_i \theta_0 + b_i).$$

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Phase de décroissance, $n \rightarrow \infty$

Conditionnement à la non-extinction

Soit $(\mathbf{Y}_n)_n$ avec pour probabilités de transition

$$\mathbb{P}(\mathbf{Y}_n = \mathbf{j} | \mathbf{Y}_{n-1} = \mathbf{i}) := \mathbb{P}(\mathbf{X}_n = \mathbf{j} | \mathbf{X}_{n-1} = \mathbf{i}, \mathbf{X}_n \neq \mathbf{0}).$$

On considère le CLSE

$$\hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \sum_{k=1}^n \frac{[Y_k - \mathbb{E}_{\theta}(Y_k | \mathbf{Y}_{k-1})]^2}{\mathbf{a} \cdot \mathbf{Y}_{k-1}}$$

Remarque : la variance conditionnelle du terme d'erreur est bornée

Phase de décroissance, $n \rightarrow \infty$

Conditionnement à la non-extinction

Théorème

Supposons que $(\mathbf{X}_n)_n$ est sous-critique. Alors $\hat{\theta}_n^*$ est fortement consistant

$$\lim_{n \rightarrow \infty} \hat{\theta}_n^* \stackrel{\text{a.s.}}{=} \theta_0,$$

et asymptotiquement normalement distribué

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^n f'(\hat{\theta}_n^Y, \mathbf{Y}_k)^2}{\sqrt{\sum_{k=0}^n f'(\hat{\theta}_n^Y, \mathbf{Y}_k)^2 f(\hat{\theta}_n^Y, \mathbf{Y}_k) (\mathbf{a} \cdot \mathbf{Y}_k)^{-1/2}}} (\hat{\theta}_n^Y - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

Avec

$$f(\theta, \mathbf{x}) = \mathbb{E}_\theta \left(\frac{Y_n}{\sqrt{\mathbf{a} \cdot \mathbf{x}}} \mid \mathbf{Y}_{n-1} = \mathbf{x} \right) = \frac{\Psi(\theta) \cdot \mathbf{x}}{\sqrt{\mathbf{a} \cdot \mathbf{x}} (1 - \mathbf{1}_{\{x_1 + \dots + x_{d-1} = 0\}} e^{-\Psi_d(\theta)x_d})}$$

et où $f'(\theta, \mathbf{x})$ est la dérivée par rapport à θ .

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Phase de décroissance, $n \rightarrow \infty$

Conditionnement à la non-extinction dans le futur lointain

Soit $(\mathbf{Z}_n)_n$ avec pour probabilités de transition

$$\mathbb{P}(\mathbf{Z}_n = \mathbf{j} | \mathbf{Z}_{n-1} = \mathbf{i}) = \mathbb{P}(\mathbf{X}_n = \mathbf{j} | \mathbf{X}_{n-1} = \mathbf{i}, \lim_{k \rightarrow \infty} \mathbf{X}_k \neq \mathbf{0}).$$

Proposition

Le processus $(\mathbf{Z}_n)_n$ (Q -processus associé à $(\mathbf{X}_n)_n$) a pour loi de transition conditionnelle

$$\mathcal{L}(Z_n | \mathbf{Z}_{n-1}) = \text{Poisson}(\mathbf{Z}_{n-1} \cdot \boldsymbol{\Psi}(\theta_0)) * \mathcal{B}(p(\theta_0, \mathbf{Z}_{n-1}))$$

où

$$p(\theta, \mathbf{x}) = \frac{\xi_1 \mathbf{x} \cdot \boldsymbol{\Psi}}{\xi_1 \mathbf{x} \cdot \boldsymbol{\Psi} + \sum_{k=2}^d x_k \xi_k},$$

ξ étant le vecteur propre à droite normalisé de \mathbf{M}_0 pour la valeur propre ρ_0 .

Phase de décroissance, $n \rightarrow \infty$

Conditionnement à la non-extinction dans le futur lointain

On considère le CLSE

$$\hat{\theta}_n^{**} = \arg \min_{\theta \in \Theta} \sum_{k=1}^n \frac{[Z_k - \mathbb{E}_\theta(Z_k | \mathbf{Z}_{k-1})]^2}{\mathbf{a} \cdot \mathbf{Z}_{k-1}}$$

Remarques :

- la variance conditionnelle du terme d'erreur est bornée
- en considérant cet estimateur on fait une hypothèse invérifiable sur l'évolution future de l'épidémie, mais grâce à cette estimation on peut dans le cadre de cette hypothèse faire des prédictions sur les trajectoires les plus dangereuses

Phase de décroissance, $n \rightarrow \infty$

Conditionnement à la non-extinction dans le futur lointain

Théorème

Supposons que $(\mathbf{X}_n)_n$ est sous-critique. Alors $\hat{\theta}_n^{**}$ est fortement consistant et asymptotiquement normalement distribué :

$$\lim_{n \rightarrow \infty} \hat{\theta}_n^{**} \stackrel{a.s.}{=} \theta_0,$$

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^n g'(\hat{\theta}_n^{**}, \mathbf{Z}_k)^2}{\sqrt{\sum_{k=0}^n g'(\hat{\theta}_n^{**}, \mathbf{Z}_k)^2 h(\hat{\theta}_n^{**}, \mathbf{Z}_k)}} \left(\hat{\theta}_n^{**} - \theta_0 \right) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

Avec

$$g(\theta, \mathbf{x}) = \mathbb{E}_\theta \left(\frac{Z_n}{\sqrt{\mathbf{a} \cdot \mathbf{x}}} \mid \mathbf{Z}_{n-1} = \mathbf{x} \right),$$

$$h(\theta, \mathbf{x}) = \mathbb{E}_\theta \left(\left(\frac{Z_n}{\sqrt{\mathbf{a} \cdot \mathbf{x}}} - g(\theta, \mathbf{x}) \right)^2 \mid \mathbf{Z}_{n-1} = \mathbf{x} \right).$$

Phase de décroissance, $n \rightarrow \infty$

Propriétés asymptotiques de $\hat{\theta}_n^*$ et $\hat{\theta}_n^{**}$.

$$f(\theta, \mathbf{x}) = \mathbb{E}_\theta\left(\frac{Y_n}{\sqrt{\mathbf{a} \cdot \mathbf{x}}} \mid \mathbf{Y}_{n-1} = \mathbf{x}\right), \quad g(\theta, \mathbf{x}) = \mathbb{E}_\theta\left(\frac{Z_n}{\sqrt{\mathbf{a} \cdot \mathbf{x}}} \mid \mathbf{Z}_{n-1} = \mathbf{x}\right)$$

- Puisque f'' et g'' ne sont pas linéaires la méthode usuelle est inapplicable (développement de S'_n au voisinage de θ_0 et LFGN pour martingales appliquée à S'_n)
- On utilise dans les deux cas la récurrence positive des chaînes de Markov irréductibles $(\mathbf{Y}_n)_n$ et $(\mathbf{Z}_n)_n$. Pour le Q -processus $(\mathbf{Z}_n)_n$, la mesure stationnaire π est la mesure de Yaglom ν biaisée en taille :

$$\pi(\mathbf{i}) = \frac{\mathbf{i} \cdot \boldsymbol{\xi} \nu(\mathbf{i})}{\sum_{\mathbf{k} \in \mathbb{N}^d} \mathbf{k} \cdot \boldsymbol{\xi} \nu(\mathbf{k})}, \quad \mathbf{i} \in \mathbb{N}^d,$$

où ν est définie par la propriété $\forall \mathbf{i}, \mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = \mathbf{i} \mid \mathbf{X}_0 = \mathbf{j}, \mathbf{X}_n \neq \mathbf{0}) = \nu(\mathbf{i}).$$

- Généralisation des critères nécessaires et suffisants pour la consistance et la normalité dans des modèles de régression non-linéaire [Jacob, 2010]

- 1 Le modèle épidémique et le paramètre d'infection
- 2 Quelques estimateurs dans la littérature
- 3 Estimateurs du paramètre d'infection
 - Phase de croissance, $n \rightarrow \infty$
 - Toutes phases de l'épidémie, $|\mathbf{X}_0| \rightarrow \infty$
 - Phase de décroissance, $n \rightarrow \infty$
 - Conditionnement à la non-extinction
 - Conditionnement à la non-extinction dans le futur lointain
- 4 Illustration avec l'épidémie d'ESB

Illustration avec l'épidémie d'ESB

Incidence de cas entre 1989 et 2011

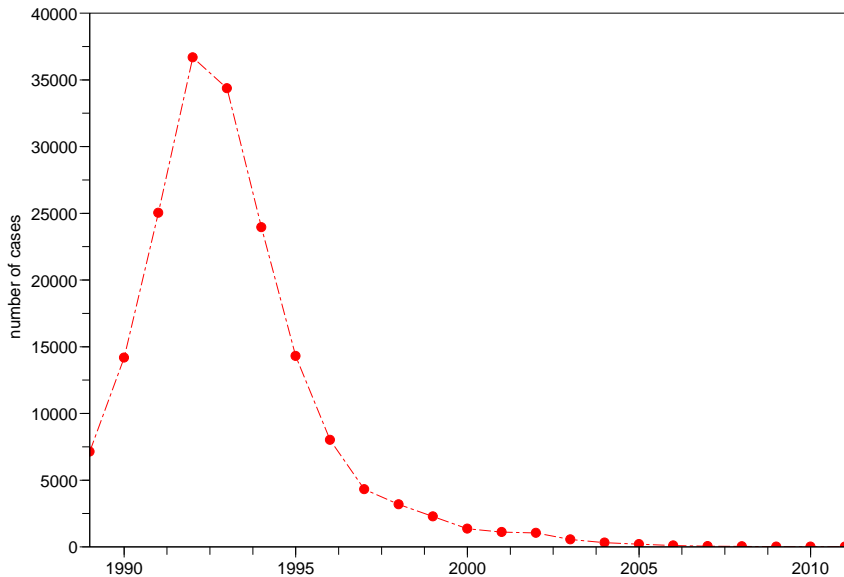


Illustration avec l'épidémie d'ESB

- Avant les mesures sanitaires en 1988 : approximativement 1000 nouveaux animaux infectés par infectieux et par an
- Après les mesures sanitaire : on pose $d = a_{max} - 1 = 9$. Pour chaque $k = 1 \dots d$,

$$\Psi_k(\theta_0) = P_{inc}(k) \sum_{a=k+1}^{a_{max}} P_{age}(a) (\theta_0 + \mathbf{1}_{\{a=k+1\}} p_{mat}).$$

Estimation du paramètre d'infection θ_0 (infection horizontale) :

On a $|\mathbf{X}_0| = 167977$, $n = 14$ et un seuil de criticalité pour θ_0 déduit de $R_0 = 1$: $\theta_{crit} = 23$.

$$\hat{\theta}_{|\mathbf{x}_0|} = 2.4328 \quad I_{95\%} = [2.3847; 2.4810]$$

$$\hat{\theta}_n = 9.8086$$

$$\hat{\theta}_n^* = 2.4328 \quad I_{95\%} = [2.4004; 2.4653]$$

$$\hat{\theta}_n^{**} = 2.4310 \quad I_{95\%} = [2.3828; 2.4792]$$

Étude Bayésienne antérieure : $\hat{\theta}^{MAP} = 2.43$

Illustration avec l'épidémie d'ESB

Simulation de 5 trajectoires de $(\mathbf{X}_n)_n$ pour le paramètre d'infection

$$\hat{\theta}_{|\mathbf{x}_0|} = 2.4328$$

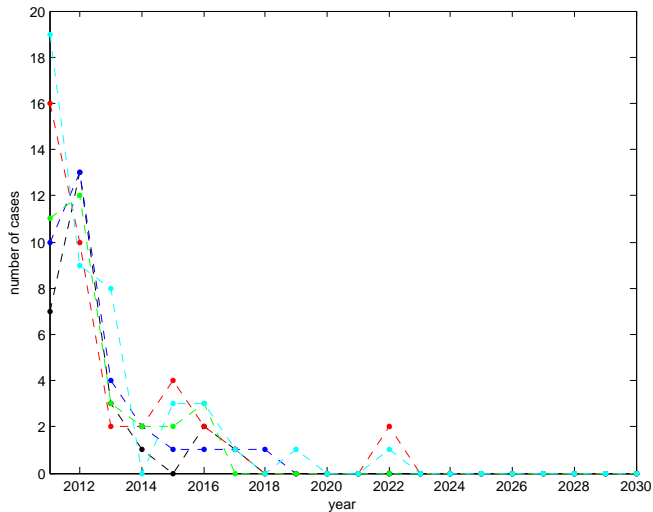
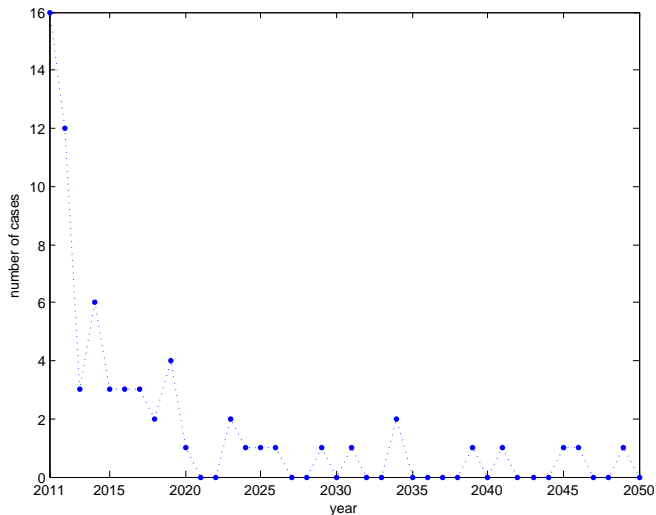






Illustration avec l'épidémie d'ESB

Simulation d'une trajectoire de $(\mathbf{Z}_n)_n$ pour le paramètre d'infection

$$\hat{\theta}_n^{**} = 2.4310$$



Références

-  ASMUSSEN, S. AND KEIDING, N. (1978). Martingale central limit theorems and asymptotic estimations theory for multitype branching processes. *Adv. Appl. Prob.* **10**, 109–129.
-  HARRIS, T. E. (1948). Branching processes. *Ann. Math. Statistics* **19**, 474–494.
-  JACOB, C. (2009). Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.* **38**, 566–597.
-  PÉNISSON, S. and JACOB, C. (2012). Stochastic methodology for the study of an epidemic decay phase, based on a branching model. *Int. J. Stoch. Anal. Applications of Stochastic Processes in Biology and Medicine*.