

Processus de branchement multitypes conditionnés à l'extinction tardive et illustration en analyse des risques épidémiologiques

Sophie Péniçon

*IRTG SMCP Berlin / Université de Potsdam
INRA Jouy-en-Josas*

Neuvième Colloque "Jeunes Probabilistes et Statisticiens"
Le Mont-Dore
3-7 mai 2010



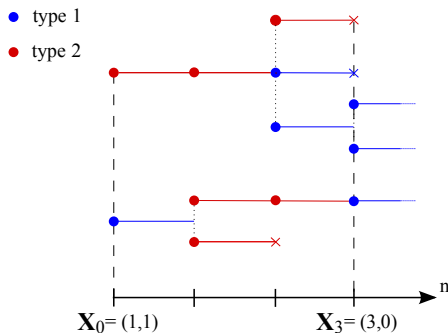
Plan

- 1 Processus de branchement multitype
 - Introduction
 - Quelques propriétés
- 2 Modélisation en épidémiologie: exemple de l'ESB
- 3 Quantification de l'infection
- 4 Analyse des risques liés à une extinction très tardive
 - Trajectoires conditionnées à une extinction très tardive
 - Estimation et prédiction
- 5 Un mot sur le processus de diffusion de Feller

Processus de branchement multitype

Processus de Bienaymé-Galton-Watson (BGW)

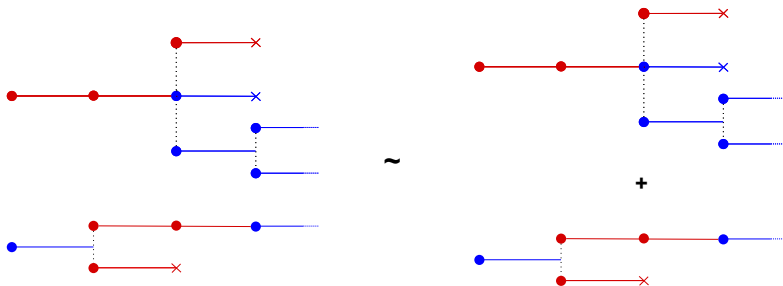
- d types de particules
- processus de Markov à valeurs dans \mathbb{N}^d



- distribution de la descendance $(p_1(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^2}$, $(p_2(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^2}$
 $p_i(k_1, k_2)$ = probabilité qu'une particule de type i donne naissance à k_1 particules de type 1 et k_2 particules de type 2

Propriété de branchement

= propriété additive



Extinction vs. explosion

- matrice moyenne \mathbf{M}
 $m_{ij} :=$ nombre moyen de descendants de type j pour une particule de type i
- Théorème de Perron-Frobenius \Rightarrow si $\exists n$ tel que $\mathbf{M}^n > \mathbf{0}$ alors \mathbf{M} a une valeur propre maximale réelle ρ qui est > 0 et simple

Le processus est alors dit **irréductible**. Sous cette hypothèse:

Théorème

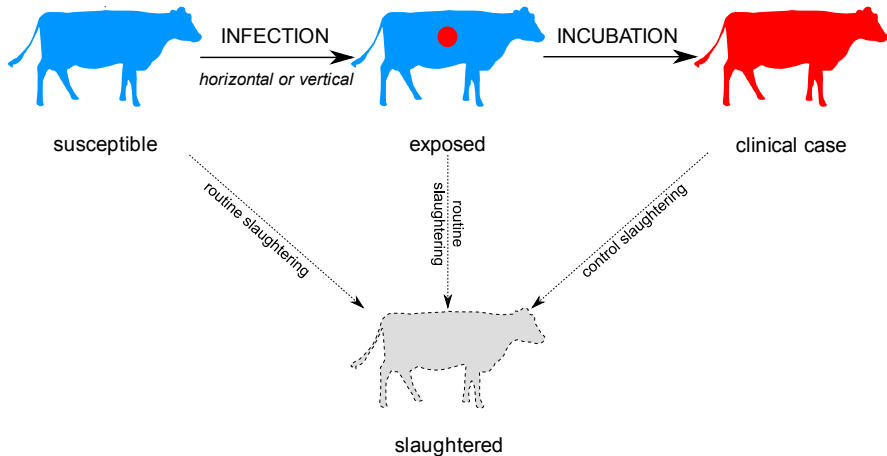
$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \mathbf{X}_n = \mathbf{0} \right) + \mathbb{P} \left(\lim_{n \rightarrow \infty} \mathbf{X}_n = \infty \right) = 1$$

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \mathbf{X}_n = \mathbf{0} \right) = 1 \iff \rho \leq 1$$

- $\rho < 1$: processus **sous-critique**
- $\rho = 1$: processus **critique**
- $\rho > 1$: processus **sur-critique**

Modélisation en épidémiologie: exemple de l'ESB

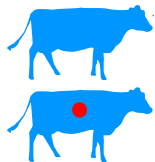
Maladie SEIR



Problématique

- quantification de l'infection persistant après la suppression des farines animales
- année d'extinction de l'épidémie
- évolution de l'épidémie en cas d'extinction très tardive

Remarque: les états de santé S et E sont indifférentiables



apparently healthy
animals



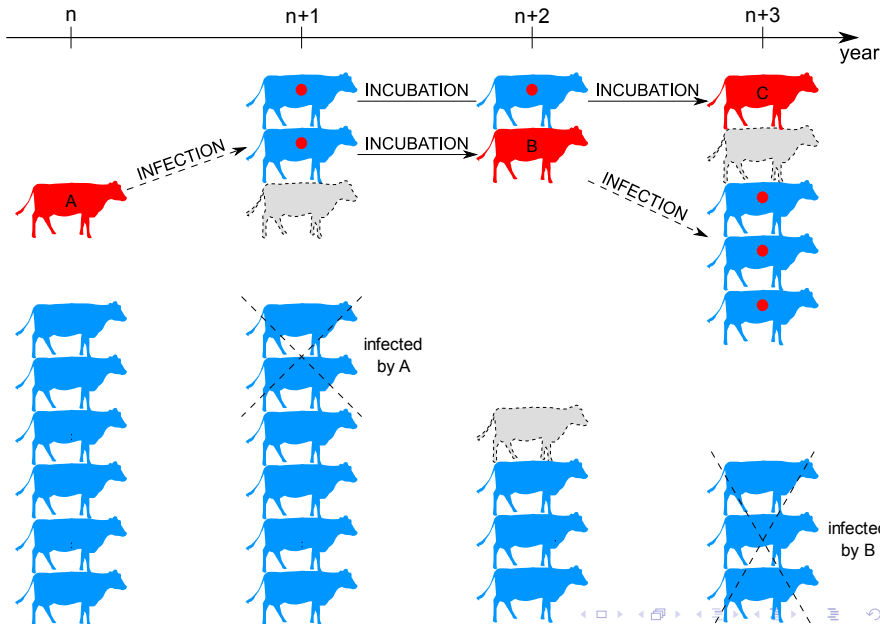
clinical cases

But: construire un modèle basé sur l'incidence des **cas cliniques**

Non rigoureux:

- modèle déterministe
- modélisation stochastique directe des cas cliniques

Plus rigoureux: modélisation stochastique de tous les états de santé



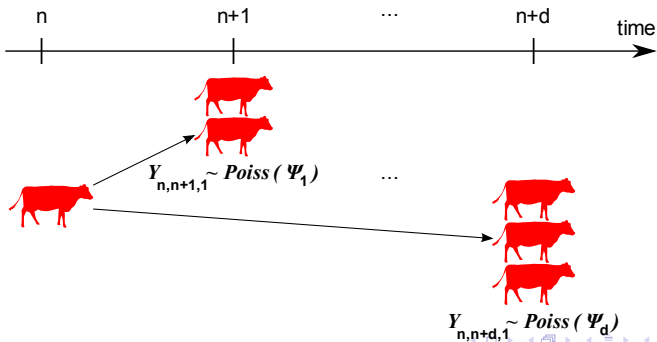
Modèle limite (C.Jacob)

Obtenu lorsque la population totale initiale temps vers ∞

Hypothèses clés: maladie SEIR rare, infection de type Reed-Frost

X_n := incidence de cas cliniques pour l'année n

$$X_n = \sum_{k=1}^d \sum_{i=1}^{X_{n-k}} Y_{n-k,n,i}, \quad \{Y_{n-k,n,i}\}_i \text{ i.i.d. } Y_{n-k,n,i} \stackrel{\mathcal{D}}{\sim} \text{Poisson}(\Psi_k(\theta_0))$$



$$\Psi_k(\theta_0) := \theta_0 P_{inc.}(k) \frac{\sum_{i=k+1}^{d+1} S_i}{\sum_{j=1}^{d+1} S_j} + p_{mat.} P_{inc.}(k) \frac{S_{k+1}}{\sum_{j=1}^{d+1} S_j}.$$

- $d = 9, 10$ ans étant l'âge maximal observé pour l'abattage du bétail
- θ_0 = paramètre inconnu d'infection horizontale persistante après la suppression de farine animale en 1989 (θ_0 = nb moyen par infectieux et par an d'animaux infectés *via* ingestion de prions excrétés par d'autres animaux vivants);
- $p_{mat.}$ = paramètre d'infection maternelle (proba pour un nouveau-né de mère infectieuse d'être infecté à la naissance);
- S_k = proba pour un animal apparemment sain de survivre au moins k années;
- $P_{inc.}(k)$ = proba que la période d'incubation intrinsèque soit de k années.

X_n processus **d -Markovien 1-dimensionnel**, de loi de transition

$$X_n | (X_{n-1}, \dots, X_{n-d}) \stackrel{\mathcal{D}}{\sim} \text{Poisson} \left(\sum_{k=1}^d X_{n-k} \psi_k(\theta_0) \right),$$

ou $\mathbf{X}_n := (X_n, X_{n-1}, \dots, X_{n-d+1})$ processus **1-Markovien d -dimensionnel**,

$$\mathbf{X}_n | \mathbf{X}_{n-1} \stackrel{\mathcal{D}}{\sim} (\text{Poisson}(\mathbf{X}_{n-1} \cdot \boldsymbol{\Psi}(\theta_0)), X_{n-1}, \dots, X_{n-d+1}).$$

Proposition

\mathbf{X}_n est un processus de branchement à d types.

Matrice moyenne

$$\mathbf{M}(\theta_0) = \begin{pmatrix} \psi_1(\theta_0) & 1 & 0 & \dots & 0 \\ \psi_2(\theta_0) & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \psi_{d-1}(\theta_0) & 0 & \dots & \dots & 1 \\ \psi_d(\theta_0) & 0 & \dots & \dots & 0 \end{pmatrix}$$

On a $[\mathbf{M}(\theta_0)]^d > \mathbf{0}$. Soit ρ la plus valeur propre maximale de $\mathbf{M}(\theta_0)$. Alors

$$\rho \leq 1 \iff \sum_{k=1}^d \Psi_k(\theta_0) \leq 1.$$

On peut écrire

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

où a_k et b_k sont connus. Alors

Proposition

$$\begin{aligned} \text{extinction p.s. de l'épidémie} &\iff \sum_{k=1}^d \Psi_k(\theta_0) \leq 1 \\ &\iff \theta_0 \leq \frac{1 - \sum_{k=1}^d b_k}{\sum_{k=1}^d a_k} \simeq 23. \end{aligned}$$

Estimation du paramètre d'infection θ_0

Conditional Least Squares Estimator

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

Proposition (S.P.)

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

$$\lim_{\mathbf{x}_0 \rightarrow \infty} \sqrt{\frac{\sum_{k=1}^n \sum_{i=1}^d a_i X_{k-i}}{\sigma^2(\hat{\theta}_{\mathbf{x}_0)}}} (\hat{\theta}_{\mathbf{x}_0} - \theta_0) \stackrel{D}{=} \mathcal{N}(0, 1).$$

où $m_{ij}^{(k)}(\theta)$ denote l'entrée (i, j) de la matrice $[\mathbf{M}(\theta)]^k$, et

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

Résultats numériques

$$|\mathbf{X}_0| = |\mathbf{X}_{1997}^{obs}| = X_{1997}^{obs} + \dots + X_{1989}^{obs} = 167977$$

$$\hat{\theta}_{\mathbf{X}_0} = 2.4486$$

$$\mathbb{P}\left(\theta_0 \in [2.4000 ; 2.4971]\right) \simeq 0.95$$

Conclusion

Existence d'une source d'infection horizontale persistant après 1989, *mineure* mais *non-nulle*. Nb moyen d'animaux infectés par cette voie est seulement de qqes unités par infectieux et par an (infection par farines animales de l'ordre de 1000).

Permet l'estimation de l'année d'extinction de l'épidémie, du nombre de cas à venir etc.

Conditionnement à une extinction très tardive

But: Étude des trajectoires à extinction très tardive:

$$\mathbb{P}(\mathbf{X}_n = \cdot | \mathbf{X}_{n+k} \neq \mathbf{0}), \quad k \text{ très grand.}$$

Approximation par

$$\mathbb{P}(\mathbf{X}_n^* = \cdot) := \lim_{k \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = \cdot | \mathbf{X}_{n+k} \neq \mathbf{0}).$$

Proposition (A.Joffe & S.Dallaporta (2008))

Supposons $\rho \leq 1$. Alors

$$\mathbb{P}(\mathbf{X}_{n+1}^* = \mathbf{j} | \mathbf{X}_n^* = \mathbf{i}) = \frac{\mathbf{1} \cdot \mathbf{j} \cdot \boldsymbol{\xi}(\theta_0)}{\rho \mathbf{i} \cdot \boldsymbol{\xi}(\theta_0)} \mathbb{P}(\mathbf{X}_{n+1} = \mathbf{j} | \mathbf{X}_n = \mathbf{i}),$$

où $\boldsymbol{\xi}(\theta_0)$ est le vecteur à droite normalisé de $\mathbf{M}(\theta_0)$ pour ρ .

\mathbf{X}_n^* est le **Q-processus** associé à \mathbf{X}_n .

Quasi-stationnarité et stationnarité

Théorème (A.Joffe & F.Spitzer (1967))

Supposons $\rho < 1$, alors

$$\nu_{\theta_0}(\mathbf{j}) := \lim_n \mathbb{P}(\mathbf{X}_n = \mathbf{j} \mid \mathbf{X}_0 = \mathbf{i}, \mathbf{X}_n \neq \mathbf{0})$$

existe, ne dépend pas de \mathbf{i} , et est une mesure de probabilité.

On appelle la mesure de probabilité ν_{θ_0} la limite de Yaglom.

Proposition (A.Joffe & S.Dallaporta (2008))

Supposons $\rho < 1$ et $\mathbb{E}(\mathbf{X}_1 \ln \mathbf{X}_1) < \infty$. Alors \mathbf{X}_n^* est récurrent positif de probabilité stationnaire

$$\pi_{\theta_0}(\mathbf{j}) := \frac{\mathbf{j} \cdot \xi(\theta_0) \nu_{\theta_0}(\mathbf{j})}{\sum_{\mathbf{k} \in \mathbb{N}^d} \mathbf{k} \cdot \xi(\theta_0) \nu_{\theta_0}(\mathbf{k})}.$$

Application au processus épidémiologique

Rappel: le processus \mathbf{X}_n a pour loi de transition

$$X_n | \mathbf{X}_{n-1} \stackrel{\mathcal{D}}{\sim} \text{Poisson}(\mathbf{X}_{n-1} \cdot \Psi(\theta_0)).$$

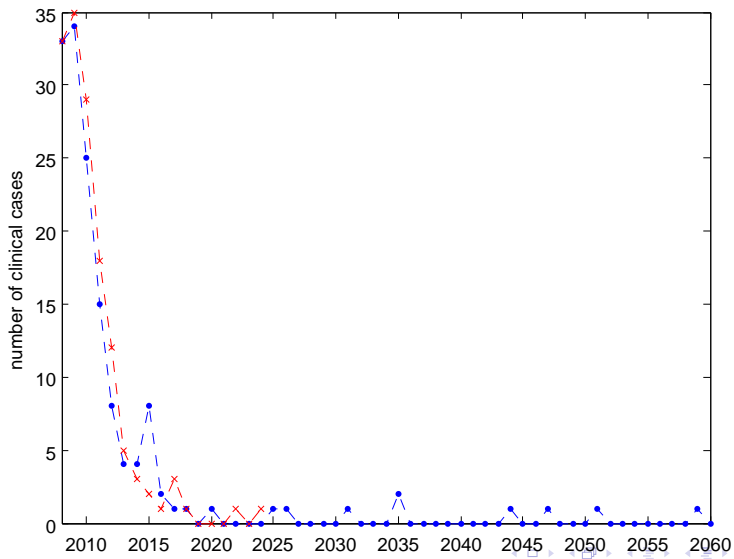
Le Q-processus associé \mathbf{X}_n^* vérifie alors:

Proposition (S.P.)

$$X_n^* | \mathbf{X}_{n-1}^* \stackrel{\mathcal{D}}{\sim} \text{Poisson}(\mathbf{X}_{n-1}^* \cdot \Psi(\theta_0))$$

$$* \mathcal{B} \left(\frac{\xi_1(\theta_0) \mathbf{X}_{n-1}^* \cdot \Psi(\theta_0)}{\xi_1(\theta_0) \mathbf{X}_{n-1}^* \cdot \Psi(\theta_0) + \sum_{i=2}^d X_{n-i+1}^* \xi_i(\theta_0)} \right).$$

Simulation d'une trajectoire du processus X_n (en rouge) et du processus X_n^* (en bleu), pour le même paramètre d'infection $\theta_0 = \hat{\theta}_{\mathbf{x}_0} = 2.4486$.



Estimation et prédiction

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

Proposition (S.P.)

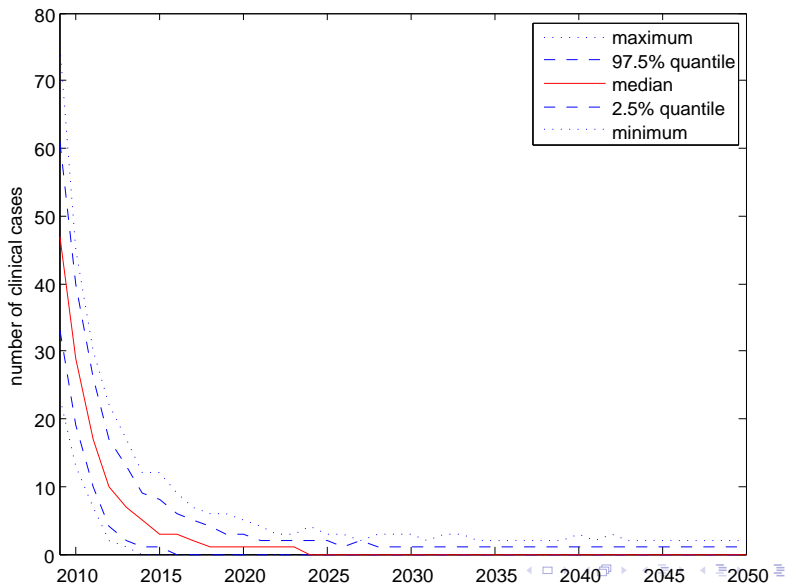
$$\lim_{n \rightarrow \infty} \hat{\theta}_n^* \stackrel{p.s.}{=} \theta_0,$$

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{\sum_{\mathbf{j} \in \mathbb{N}^d} f^*(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j})}{\sum_{\mathbf{j} \in \mathbb{N}^d} g^*(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j})} (\hat{\theta}_n^* - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

$$n = 11$$

$$\hat{\theta}_n^* = 2.4472$$

$$\mathbb{P}(\theta_0 \in [2.3988 ; 2.4956]) \approx 0.95$$

Simulation de 1000 trajectoires de X_n^* à partir de 2009

Un mot sur le processus de diffusion de Feller

Proposition (N.Champagnat & S.Roelly (2008) / S.P. (2010))

Soit \mathbb{P} la loi d'un processus de diffusion de Feller (resp. d'un BGW à temps continu) multitype, irréductible et (sous-)critique ($\rho \leq 0$). Alors la loi du Q -processus associé, défini par








$$\forall t \geq 0, \forall B \in \mathcal{F}_t, \mathbb{P}^*(B) := \lim_{\theta \rightarrow \infty} \mathbb{P}(B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0})$$

est une h -transformée de Doob de \mathbb{P} , satisfaisant: pour tout $t \geq 0$ et $\mathbf{x} \in \mathbb{R}_+^d$ (resp. $\mathbf{x} \in \mathbb{N}^d$), $\mathbf{x} \neq \mathbf{0}$,

$$\mathbb{P}_{\mathbf{x}}^* |_{\mathcal{F}_t} = \frac{\mathbf{X}_t \cdot \boldsymbol{\xi}}{\mathbf{x} \cdot \boldsymbol{\xi}} e^{-\rho t} \mathbb{P}_{\mathbf{x}} |_{\mathcal{F}_t}.$$

Proposition (S.P.)

De même que la diffusion de Feller multitype peut-être obtenue comme limite de BGW multitypes, le Q -processus associé peut-être obtenu comme limite de Q -processus associés aux BGW.

-  CHAMPAGNAT, N. and RØELLY, S. (2008). Limit theorems for conditioned multitype Dawson-Watanabe processes and Feller diffusions. *Electron. J. Probab.* **13**, 777–810.
-  DALLAPORTA, S. and JOFFE, A. (2008). The Q -process in a multitype branching process. *Int. J. Pure Appl. Math.* **42**, 235–240.
-  JACOB, C., MAILLARD-TEYSSIER, L., DENIS, J. B. and BIDOT, C. (2009). A branching process approach for the propagation of the BSE in Great-Britain. In *Branching processes and their Applications*, Lecture Notes in Statistics, Springer-Verlag, 227–242.
-  A. JOFFE, A. and SPITZER, F. (1967). On multitype branching processes with $\rho \leq 1$. *J. Math. Anal. Appl.* **19**, 409–430.
-  PÉNISSON, S. (2010). Continuous-time multitype branching processes conditioned on very late extinction. *ESAIM P&S*.
-  PÉNISSON, S. (2010). Estimation of the infection parameter in the different phases of an epidemic modeled by a branching process.
-  PÉNISSON, S. and JACOB, C. (2010). BSE epidemic in Great-Britain: prediction of the disease spread and study of the very late extinction case scenario, based on a stochastic branching model.