Semi-semi-Markov processes: a new class of processes for formalizing and generalizing state-dependent individual-based models

Christine Jacob¹ and Anne-France Viet²

¹ Applied Mathematics and Informatics Unit, INRA, Jouy-en-Josas, France

² Unit of Animal Health Management, Veterinary School, INRA, Nantes, France



_

N

Origin: epidemiological problem

Study the propagation of a disease (BVD : Bovine Viral Diarrhoea) in a dairy herd 3 types of individual transitions :

- Branching transitions (births, deaths)
- Group changes according to physiological status and age :
 - 4 groups : calves, heifers before breeding, heifers after breeding, dairy cows
- Health status changes which depend on the population infection
- ⇒ individual-based stochastic model : the individual transitions are population-dependent and semi-Markovian
- ⇒ propagation of a random process (disease) on a random graph (vertices : individuals)

Individual-based models

- used in population dynamics when individuals are marked by personal characteristics or when the next state-change is driven by complex rule decisions which depend on the current state of the population
- calculate empirical distributions at the scale of the population from simulated individual trajectories ("bottom-up approach")
 - litterature on individual-based models has considerably increased thanks to the increase of the computers capacity and the popularization of informatics
 - no mathematical formalism : validation of this approach? how is the process at the level of the population?



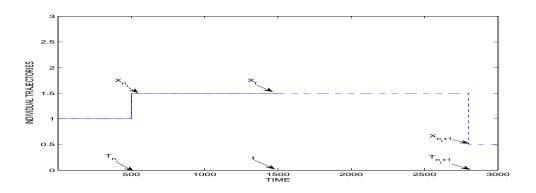
Build a rigorous mathematical formalism of these models at the population level (top-down)

⇒ good readability of the different model components independently of the programming language

> validate and supplement the empirical distributions by analytical results

Homogeneous Semi-Markov Process (SMP) for one individual ω

Feller W. (1964), Çinlar (1975), Kulkarni, V. (1995), Becker G. et al. (1999), Iosifescu M. (1999)



Countable state space, jumps at random times $\{X_t\}_{t\in\mathbb{R}^+}$ is an homogeneous semi-Markov process if

$$X_t = X_{n_t} 1_{\{n_t = \sup\{n: T_n \le t\}\}}, \ T_n: n$$
th jump time, $X_n:$ state at $T_n, \ \{X_n, T_n\}$ MRP

Law of $\{X_t\}_t \Longleftrightarrow$ law of $\{X_n, T_n\}_n \Longleftrightarrow$ transition laws of $\{X_n, T_n\}_n$

Assumption : $\{X_n, T_n\}$ is a MRP

$$P(X_{n+1}=j,\Delta T_{n+1}\leq \tau|X_n,\ldots,X_0,T_n,\ldots,T_0) = P(X_{n+1}=j,\Delta T_{n+1}\leq \tau|X_n)$$

$$\stackrel{homog.}{=} P(X_1=j,\Delta T_1\leq \tau|X_0)$$

$$\Delta T_{n+1} \stackrel{defin.}{=} T_{n+1} - T_n \text{ (waiting time between 2 jumps)}$$

Kernels (transition laws)

$$Q_{i,j}(\tau) \stackrel{defin.}{=} P(\Delta T_1 \le \tau, X_1 = j | X_0 = i)$$

$$= P(\Delta T_1 \le \tau | X_1 = j, X_0 = i) P(X_1 = j | X_0 = i)$$

$$= F_{i,j}(\tau) P(i,j)$$

 $F_{i,j}(\tau)$: cdf of the sojourn time in i before jumping in j; P(i,j): transition probability of $\{X_n\}$

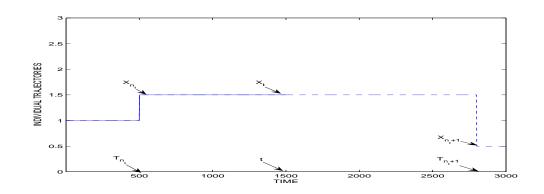
Kernels ← Transition rates

$$\lambda_{i,j}(\tau) \stackrel{def.}{=} \lim_{\Delta \tau \to 0} \frac{P(\Delta T_{n+1} \in (\tau, \tau + \Delta \tau), X_{n+1} = j | X_n = i, \Delta T_{n+1} > \tau)}{\Delta \tau}$$

$$\lambda_{i,j}(\tau) = \frac{\dot{Q}_{i,j}(\tau)}{1 - \sum_{j} Q_{i,j}(\tau)} \iff Q_{i,j}(\tau) = \int_{0}^{\tau} \lambda_{i,j}(u) exp(-\int_{0}^{u} \sum_{j} \lambda_{i,j}(s) ds) du$$

Particular case: Markov process:

$$\lambda_{i,j}(\tau) = \lambda_{i,j} = \lambda_i P(i,j)$$
, for any τ , $F_{i,j}(\tau) = 1 - \exp(-\lambda_i \tau)$
 $\mathbf{P}(t) = \exp(\Lambda t)$, $\mathbf{P}[i,j](t) \stackrel{def.}{=} P(X(t) = j|X(0) = i)$



${\bf x}$ Event-driven simulation algorithm based on $P(i,j)F_{i,j}(\tau)$

Current jump time and jump state : $(t_n,i)\Longrightarrow$ determine the next jump time and jump state (t_{n+1},j)

Simulate j according to $\{P(i,j')\}_{j'}$

Simulate $\tau \stackrel{def.}{=} t_{n+1} - t_n$ according to $F_{i,j}(.)$

Probability law of the process (renewal equations)

$$P(X_t = j | X_0 = i) \stackrel{notat.}{=} P_{i,j}(t) = \left[1 - \sum_{j'} Q_{i,j'}(t)\right] 1_{\{j=i\}} + \sum_{k} \int_{0}^{t} dQ_{i,k}(s) P_{k,j}(t-s)$$

Matrix process : $\mathbf{P}[i,j](t) \stackrel{def.}{=} P_{i,j}(t)$

$$\mathbf{P} = (\mathbf{I} - \mathbf{Q}^{\Sigma}) + \mathbf{Q} * \mathbf{P} \Longrightarrow \mathbf{P} = \sum_{n=0}^{\infty} \mathbf{Q}^{n*} * (I - \mathbf{Q}^{\Sigma}), \quad \mathbf{Q} * \mathbf{P}(t) = \sum_{k} \int_{0}^{t} Q_{i,k}(s) P_{k,j}(t - s)$$

Approximate solutions

9

Empirical distribution (simulations)

$$\mathbf{P} \simeq \sum_{n=0}^{n_t} \mathbf{Q}^{*n} * (I - \mathbf{Q}^{\Sigma})$$

Recursive solution of P by discretization of time

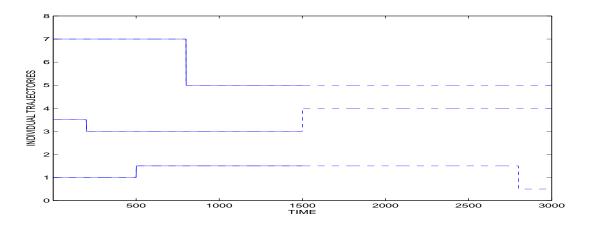
Upper and lower bounds of Li and Luo (2005)

Stationary law (all the states are recurrent) or quasi-stationary law (conditioned on staying in the nonabsorbing state)

Homogeneous Semi-Semi-Markov Processes for a closed population $\Omega = \{\omega_l\}_{l < N}$

Set of MRP : $\{\{(X_m^{(l)}(\omega_l),T_m^{(l)}(\omega_l))\}_m\}_{l=1,\dots,N}$, not synchronized and population-dependent

Goal: determine the distribution of the population process $\mathcal{X}_t = \{X_t^{(1)}, ..., X_t^{(N)}\}$

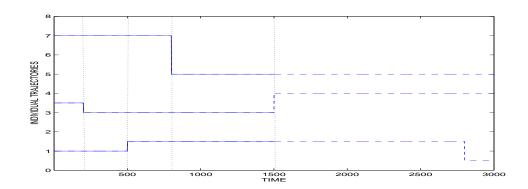


Particular case : the MRP are i.i.d. communication networks, asymptotic distributions in the heavy-tailed case $1-F(t)=t^{-\alpha}L(t)$ (Mikosh and Resnick, 2005, Mitov and Yanev, 2006)

Steps

- $-\operatorname{define}\ \{\mathcal{X}_t(\Omega)\}_{t\in\mathbb{R}^+}\operatorname{from}\ \{(\mathcal{X}_n(\Omega),\mathcal{T}_n(\Omega))\}n\in\mathbb{N}, \operatorname{defined}\operatorname{itself}\operatorname{from}\operatorname{the}\operatorname{MRP}\ \{\{(X_m^{(l)}(\omega_l),T_m^{(l)}(\omega_l))\}_{m\in\mathbb{N}}\}$
- deduce the kernel (transition law) of the population process from the individual kernels
- deduce transition rates, probability law, simulation algorithm,...

Definition of $\{\mathcal{X}_t(\Omega)\}$ from $\{(\mathcal{X}_n(\Omega), \mathcal{T}_n(\Omega))\}$, defined itself from the individual MRP



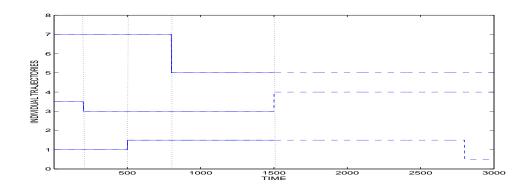
$$\mathcal{X}_{t}(\Omega) \stackrel{def.}{=} \mathcal{X}_{n_{t}}(\Omega) \stackrel{def.}{=} \{X_{m_{l,t}}^{(l)}(\omega_{l})\}_{l}$$

$$m_{l,t}(\omega_{l}) \stackrel{def.}{=} \sup\{m : T_{m}^{(l)}(\omega_{l}) \leq t\}$$

$$n_{t}(\Omega) \stackrel{def.}{=} \sum_{l} m_{l,t}(\omega_{l})$$

$$\mathcal{T}_{n_{t}} \stackrel{def.}{=} \sup_{l} \sup_{m} \{T_{m}^{(l)}(\omega_{l}) \leq t\}$$

Law of $\{\mathcal{X}_t(\Omega)\}_t \iff \text{law of } \{\mathcal{X}_n, \mathcal{T}_n\}_n \iff \{P(\mathcal{X}_{n+1} = J, \Delta \mathcal{T}_{n+1} \leq \tau | \mathcal{F}_n(I))\}_{n,I,J}$ (kernels)



$$\mathcal{F}_n(I) = \{\mathcal{X}_n = I, \mathcal{X}_{n-1} = I_{n-1}, ..., \mathcal{X}_0 = I_0, \mathcal{T}_n = t_n, \mathcal{T}_{n-1} = t_{n-1}, ..., \mathcal{T}_0 = t_0\} \text{ : past until } (t_n, I)$$

Kernel

$$P(\mathcal{X}_{n+1} = J, \Delta \mathcal{T}_{n+1} \leq \tau | \mathcal{F}_n(I)) = P(\Delta \mathcal{T}_{n+1} \leq \tau | \mathcal{X}_{n+1} = J, \mathcal{F}_n(I)) P(\mathcal{X}_{n+1} = J | \mathcal{F}_n(I))$$

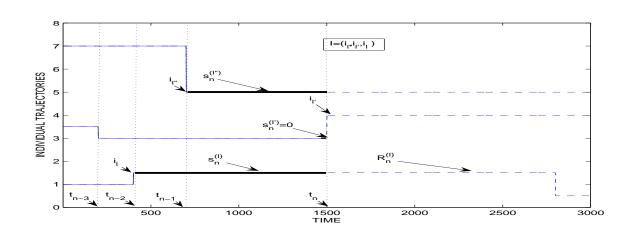
$$\stackrel{notation}{=} F_{\mathcal{F}_n(I),J}(\tau) P(\mathcal{F}_n(I),J)$$

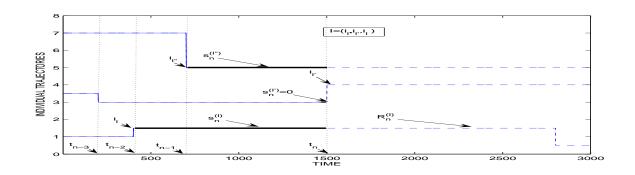
$$\stackrel{notation}{=} Q_{\mathcal{F}_n(I),J}(\tau)$$

 \Longrightarrow calculate $F_{\mathcal{F}_n(I),J}(.)$, $P(\mathcal{F}_n(I),J)$ from the individual transitions defined from (t_n,I)

Assumptions given the past until (t_n, I)

- 1. A1: the $\{(remaining\ waiting\ time\ R_n^{(l)},\ next\ jump\ state\ X_{m_n+1}^{(l)})\}_l$ are mutually independent
- 2. A2: for each l, the law of $R_n^{(l)}$ (remaining time in i_l before jumping in j_l) depends only on I and on $s_n^{(l)}$ (time already spent in i_l) and on I
- 3. A3: the probability for l to jump from i_l to j_l depends only on i_l , j_l , and I: $P^{(l)}(i_l|I,j_l)$
- 4. A4: the next population jump is defined by the individual jump which occurs the first $Q_{\mathcal{F}_n(I).J_l}(\tau) = P(\min_l \{R_n^{(l')}\} = R_n^{(l)}, R_n^{(l)} \le \tau, X_{m_n+1}^{(l)} = j_l | \mathcal{F}_n(I))$





Proposition. Let $I \rightarrow J_l : i_l \rightarrow j_l$. Then

$$\begin{split} dF_{\mathcal{F}_{n}(I),J_{l}}(\tau) &\overset{def.}{=} \frac{dQ_{\mathcal{F}_{n}(I),J_{l}}(\tau)}{P(\mathcal{F}_{n}(I),J_{l})} = \frac{\int_{0}^{\tau} \Pi_{l'\neq l}(1-\sum_{j_{l'}\in\mathcal{X}_{l'}(I)}Q_{i_{l'}|I,j_{l'}}^{(l')|s_{n}^{(l')}}(\tau))dQ_{i_{l}|I,j_{l}}^{(l)|s_{n}^{(l)}}(\tau)}{\int_{0}^{\infty} \Pi_{l'\neq l}(1-\sum_{j_{l'}\in\mathcal{X}_{l'}(I)}Q_{i_{l'}|I,j_{l'}}^{(l')|s_{n}^{(l')}}(\tau))dQ_{i_{l}|I,j_{l}}^{(l)|s_{n}^{(l)}}(\tau)} \quad \textbf{(cdf of } \Delta\mathcal{T}_{n+1}\textbf{)} \\ P(\mathcal{F}_{n}(I),J_{l}) &\overset{def.}{=} \int_{0}^{\infty} dQ_{\mathcal{F}_{n}(I),J_{l}}(\tau) = \int_{0}^{\infty} \Pi_{l'\neq l}(1-\sum_{j_{l'}\in\mathcal{X}_{l'}(I)}Q_{i_{l'}|I,j_{l'}}^{(l')|s_{n}^{(l')}}(\tau))dQ_{i_{l}|I,j_{l}}^{(l)|s_{n}^{(l)}}(\tau), \\ Q_{i_{l}|I,j_{l}}^{(l)|s_{n}^{(l)}}(\tau) &= \frac{F_{i_{l}|I,j_{l}}^{(l)}(s_{n}^{(l)}+\tau)-F_{i_{l}|I,j_{l}}^{(l)}(s_{n}^{(l)})}{1-F_{i_{l}|I,j_{l}}^{(l)}(s_{n}^{(l)})}P^{(l)}(i_{l}|I,j_{l}): \text{ individual prior kernel} \end{split}$$

Consequence: $Q_{\mathcal{F}_n(I),J_l}(.)$ depends only on the current state I, the next state J_l and $\{s_n^{(l)}\}_l$

7

Time-driven simulation algorithm $\iff \lambda_{\mathcal{F}_n(I),J}(.)$

Event-driven simulation algorithm $\iff Q_{\mathcal{F}_n(I),J}(.) = F_{\mathcal{F}_n(I),J}(.)P(\mathcal{F}_n(I),J)$

Determine the next jump (t_{n+1}, J) from (t_n, I) and $\{s_n^{(l)}\}$, $I = (i_1, ..., i_l, ..., i_N)$: for each individual l,

- 1. choose j_l according to $\{P^{(l)}(i_l|I,j_l)\}_{j_l}$
- 2. simulate a remaining waiting time $r_n^{(l)}$ in i_l before jumping into j_l

Then $r_n^{(l)} = \min_{l'} \{r_n^{(l')}\}$ defines the next jump time and the next state $J_l = (i_1, ..., j_l, ..., i_N)$

Transition rates ← kernels

$$\lambda_{\mathcal{F}_n(I),J}(\tau) \stackrel{def.}{=} \lim_{\Delta\tau \to 0} \frac{P(\Delta \mathcal{T}_{n+1} \in (\tau, \tau + \Delta\tau), \mathcal{X}_{n+1} = J | \mathcal{F}_n(I), \Delta \mathcal{T}_{n+1} > \tau)}{\Delta\tau}$$

Proposition

$$\lambda_{\mathcal{F}_n(I),J}(\tau) = \frac{\dot{Q}_{\mathcal{F}_n(I),J}(\tau)}{1 - \sum_{J} Q_{\mathcal{F}_n(I),J}(\tau)}, \tau \in \mathbb{R}^+$$

$$Q_{\mathcal{F}_n(I),J}(\tau) = \int_0^{\tau} \lambda_{\mathcal{F}_n(I),J}(u) exp(-\int_0^t \sum_{J} \lambda_{\mathcal{F}_n(I),J}(s) ds) du$$

Corollary. Assume (Exp): $F_{i_l|I,j_l}^{(l)}(\tau) = 1 - \exp(-\lambda_{i_l|I} \tau)$

Then the SSMP is a MP, and for all I not absorbing

$$dF_{I,J_l}(\tau) = dF_I(\tau) = (\sum_{l'} \lambda_{i_{l'}|I}) \exp(-\sum_{l'} \lambda_{i_{l'}|I} \tau) d\tau$$

$$P(I,J_l) = P^{(l)}(i_l|I,j_l) \frac{\lambda_{i_l|I}}{\sum_{l'} \lambda_{i_{l'}|I}}.$$

$$\lambda_{I,J_l}(\tau) = \lambda_{i_l|I} P^{(l)}(i_l|I,j_l) = \lambda_{i_l|I,j_l}$$

$$\lambda_{I}(\tau) = \sum_{J} \lambda_{I,J}(\tau) = \sum_{l} \lambda_{i_l|I}$$

$$\mathbf{P}(t) = \exp(\Lambda t)$$

Consequence. Under (Exp), if I is not an absorbing state, then

$$m_I = ig[\sum_{l'} \lambda_{i_{l'}|I}ig]^{-1}$$
 (mean time in I)

Marginal probability law of $\{X_t\}_t$: renewal equations

$$P(\mathcal{X}_{t} = J | I_{0}, \{s_{0}^{(l)}\}, t_{0}) = P(\Delta \mathcal{T}_{1} > t - t_{0} | I_{0}, \{s_{0}^{(l)}\}) 1_{\{J = I_{0}\}} + \sum_{I_{1} \neq I_{0}} \int_{t_{1} \in (t_{0}, t)} dP(\mathcal{X}_{1} = I_{1}, \Delta \mathcal{T}_{1} = t_{1} - t_{0} | I_{0}, \{s_{0}^{(l)}\}) P(\mathcal{X}_{t} = J | I_{1}, \{s_{1}^{(l)}\}, t_{1})$$

$$= P(\Delta \mathcal{T}_{1} > t - t_{0} | I_{0}, \{s_{0}^{(l)}\}) 1_{\{J = I_{0}\}} + \sum_{I_{1} \neq I_{0}} \int_{t_{1} \in (t_{0}, t)} dP((\mathcal{X}_{1}, S_{1}) = (I_{1}, \{s_{1}^{(l)}\}), \Delta \mathcal{T}_{1} = t_{1} - t_{0} | I_{0}, \{s_{0}^{(l)}\}) P(\mathcal{X}_{t} = J | I_{1}, \{s_{1}^{(l)}\}) P(\mathcal{X}_{t} = J | I_{1}, \{s_{1}^{(l)}\})$$

 \Longrightarrow Approximate solution : $\mathbf{P}(t) = [\sum_{n \geq 0}^{n_t} \mathbf{Q}^{\mathcal{Y}*n} * (\mathbf{I} - \mathbf{Q}^{\Sigma}))(t)$

Approximate solution : discretization of time

$$s_h^{(l)} = [s_{h-1}^{(l)} + t_h - t_{h-1}] 1_{\{t_h \notin \{T_m^{(l)}\}_m\}} \Longrightarrow \{s_h^{(l)}\}_l \stackrel{not.}{=} \Delta_{t_h - t_{h-1}}; \{s_0^{(l)}\}_l \stackrel{not.}{=} \Delta_0$$

$$\mathbf{P}_{\Delta_0}(t-t_0) = \mathbf{I} - \mathbf{Q}_{\Delta_0}^{\Sigma}(t-t_0) + \int_{t_1 \in (t_0,t)} d\mathbf{Q}_{\Delta_0}(t_1-t_0) \mathbf{P}_{\Delta_{t_1-t_0}}(t-t_1).$$

The discretization of the system using $t - t_0 = nh$, $t_1 - t_0 \in \{ih\}_{i \le n}$, leads to the solution

$$\begin{pmatrix} \mathbf{P}_{\Delta_0}(nh) \\ \mathbf{P}_{\Delta_h}((n-1)h) \\ \dots \\ \mathbf{P}_{\Delta_{(n-1)h}}(h) \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{\Delta_0}(0) & \mathbf{R}_{\Delta_0}(h) \dots \mathbf{R}_{\Delta_0}((n-1)h) \\ 0 & \mathbf{R}_{\Delta_h}(0) \dots \mathbf{R}_{\Delta_h}((n-2)h) \\ \vdots & \vdots & \vdots \\ 0 & 0 \dots \dots \mathbf{R}_{\Delta_{(n-1)h}}(0) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{B}_{\Delta_0,n} \\ \mathbf{B}_{\Delta_h,n} \\ \vdots \\ \mathbf{B}_{\Delta_{(n-1)h},n} \end{pmatrix}$$

 $\mathbf{R}_{\Delta}(ih) = \mathbf{I}\delta_{0,i} - a_i\dot{\mathbf{Q}}_{0,\Delta}(ih)(1-\delta_{0,i}),\ \delta_{0,i} = 1$ when i=0 (and is 0 otherwise), $i=0,\ldots,n-1$ $\mathbf{B}_{\Delta_{jh},n} = \mathbf{I} - \mathbf{Q}_{\Delta_{jh}}^{\Sigma}((n-j)h),\ j=0,\ldots,n-1$ $\{a_i\}_i$ depends on the numerical integration scheme,

Semi-semi-Markovian branching process for individuals with a pregnancy period

 $\mathcal{X}_t(\Omega)$ takes values in $\mathcal{X} = \{\{(P_l)\}_{l \in \mathcal{L}}\}_{\mathcal{L}}$, where $P_l \in \mathcal{P} = \{pregnant, not \ pregnant, R\}$ $l \in \mathcal{L}$, l: (date of birth, number u of the individual among the individuals born at this date)

$$\mathcal{X}_t(\Omega) \stackrel{def.}{=} \mathcal{X}_{n_t}(\Omega)$$
 (1)

$$n_t(\Omega) \stackrel{def.}{=} \sum_{l \in \mathcal{L}_{n_t-1}(\Omega)} m_{p,l,t}$$
 (2)

$$m_{p,l,t} \stackrel{def.}{=} \sup\{m : T_m^{(p,l)} \le t\}, l \in \mathcal{L}_{n_t-1}(\Omega)$$
 (3)

$$\mathcal{X}_{n_t}(\Omega) \stackrel{def.}{=} \{X_{m_{p,l,t}}^{(B,l)}\}_{l \in \mathcal{L}_{n_t-1}(\Omega)} \tag{4}$$

$$X_{m_{p,l,t}}^{(B,l)} \stackrel{def.}{=} \{X_{m_{p,l,t}}^{(p,l)} \neq R, \{X_0^{(p,l')} \neq R\}_{l' \in \widetilde{Y}_{n_t,l}}\}$$
(5)

$$T_0^{(p,l')} \stackrel{def.}{=} T_{m_{n,l}}^{(p,l)}, \ l' \in \widetilde{Y}_{n_t,l}, \ l \in \mathcal{L}_{n_t-1}(\Omega)$$

$$\tag{6}$$

$$\mathcal{T}_{n_t}(\Omega) \stackrel{def.}{=} \sup_{l \in \mathcal{L}_{n_t-1}(\Omega)} \{ T_{m_{p,l,t}}^{(p,l)} \}. \tag{7}$$

$$\mathcal{L}_{n_t}(\Omega) \stackrel{def.}{=} \{l, \{labels\{\widetilde{Y}_{n_t,l}\}\}\}_{l \in \mathcal{L}_{n_t-1}(\Omega)}$$
(8)

Example : if $i_l^p \rightarrow j_l^B$: $pregnant \rightarrow not\ pregnant$, then

 $\mathbb{R} \quad F_{i_l^p|I,j_l^B}^{(B,l)}(.) = F_{pregnant,not\,pregnant}^{(B,l)}(.) \text{ (cdf of the pregnancy period),}$

 $P^{(B,l)}(i_l^p|I,j_l^B)$ is the probability for l to give birth to $\widetilde{Y}_{n,l}$ newborns at his next "jump" among the states $\{alive\ with\ \widetilde{Y}\ newborns\}_{\widetilde{Y}},R\}$

Spread of a disease in a branching population structured in groups

(4), (5), (6) replaced by

$$\mathcal{X}_{n_{t}}(\Omega) \stackrel{def.}{=} \{ (X_{m_{p,l,t}}^{(B,l)}, X_{m_{h,l,t}}^{(h,l)}, X_{m_{g,l,t}}^{(g,l)}) 1_{\{X_{m_{g,l,t}}^{(g,l)} \neq R\}} \}_{l \in \mathcal{L}_{n_{t}-1}(\Omega)}
X_{m_{p,l,t}}^{(B,l)} \stackrel{def.}{=} \{ X_{m_{p,l,t}}^{(p,l)}, \{ (X_{0}^{(p,l')}, X_{0}^{(h,l')}, X_{0}^{(g,l')}) 1_{\{X_{0}^{(g,l')} \neq R\}} \}_{l' \in \widetilde{Y}_{n_{t},l}} \} \}, \ l \in \mathcal{L}_{n_{t}-1}(\Omega)
T_{0}^{(c,l')} \stackrel{def.}{=} T_{m_{p,l,t}}^{(p,l)}, \ l' \in \widetilde{Y}_{n_{t},l}, \ l \in \mathcal{L}_{n_{t}-1}(\Omega), \ c \in \{p,h,g\}$$

Conclusion

- -Individual based models: empirical distributions based on individual simulated trajectories
- -Population process: kernel, simulation algorithm, probability law, approximated probability law, asymptotic behavior?

THANK YOU FOR YOUR ATTENTION!