

Modélisation stochastique d'une épidémie SIR

Manon Costa

4 octobre 2011

Table des matières

1	Le modèle SIR	2
1.1	Un bref historique	2
1.2	Construction de Sellke	2
1.3	Approximation du comportement au début de l'épidémie	3
2	Extension sur un graphe	4
2.1	Le graphe des configurations	5
2.2	Propagation de l'épidémie sur un tel graphe	5
2.3	Limite en grande population	8
3	Perspectives	10

Résumé

Nous introduirons ici quelques modèles stochastiques de propagation d'épidémies. Les modèles épidémiologiques habituellement étudiés sont des modèles déterministes. Toutefois s'ils peuvent être plus difficiles à étudier et moins bien refléter la réalité, les modèles probabilistes sont une façon naturelle de modéliser l'évolution d'une épidémie : chaque individu a une certaine probabilité d'être infecté par la maladie. Une part importante de l'étude de ces problèmes stochastiques va être de déterminer si, quand la taille de la population augmente, ils convergent vers un problème déterministe.

Nous nous intéresserons plus particulièrement à des épidémies se transmettant d'individu en individu, et conférant en cas de rétablissement une immunité aux contacts suivants. Les maladies infantiles, sexuellement transmissibles ou même la grippe correspondent à ce genre de pathologie. On appelle ce type de modèle, une épidémie SIR (Susceptibles, Infectés, Rétablis). La population est séparée en trois catégories : les Susceptibles qui sont les individus sains qui sont sensibles à la maladie. Les Infectés qui sont atteints par l'épidémie et peuvent donc la transmettre pendant une certaine période de contagion. A la fin de cette période, ils deviennent Rétablis et sont immunisés contre une nouvelle infection.

1 Le modèle SIR

1.1 Un bref historique

Les premiers modèles épidémiologiques sont apparus au début du XXème siècle [1]. C'est en 1927 que Kermack et McKendrick proposent le premier modèle complet pour modéliser une épidémie. L'idée principale vient du fait qu'en temps discret, le nombre de nouvelles infections est proportionnel au produit du nombre d'infectés et de susceptibles. Ce modèle déterministe en population fermée peut alors s'écrire sous la forme suivante : si $x(t), y(t), z(t)$ représentent le nombre d'individus susceptibles, infectés et rétablis au temps t , alors :

$$\begin{aligned}x'(t) &= -\lambda x(t)y(t) \\y'(t) &= \lambda x(t)y(t) - \gamma y(t) \\z'(t) &= \gamma y(t)\end{aligned}$$

Les paramètres λ et γ représentent des constantes régissant respectivement l'infection et le rétablissement des individus.

Le premier modèle stochastique a été présenté par Reed et Frost en 1928, sur une échelle de temps discret (typiquement, une génération), il décrit les probabilités d'évolution des différentes classes de la population. A chaque génération, l'état de la population ne dépend que de la génération précédente, et les probabilités d'évolution sont binomiales. En gardant les mêmes notations, on a donc :

$$\begin{aligned}P(Y_{j+1} = y_{j+1} | X_0 = x_0, Y_0 = y_0, \dots, X_j = x_j, Y_j = y_j) \\&= P(Y_{j+1} = y_{j+1} | X_j = x_j, Y_j = y_j) \\&= \binom{x_j}{y_{j+1}} (1 - q^{y_j})^{y_{j+1}} (q^{y_j})^{x_j - y_{j+1}}\end{aligned}$$

C'est en 1949 que Bartlett propose ce qui sera le modèle épidémiologique standard. Au temps initial, la population se compose de m infectés et n susceptibles. Chaque infecté a une période de contagion qui suit une loi I (moyenne i variance σ^2) indépendante de celle des autres. Et pendant cette période il rencontre un autre individu donné suivant les instants de saut d'un processus de Poisson d'intensité λ/n . Tous ces processus sont indépendants entre eux et des périodes de contagion. Ainsi, pendant temps dt , la probabilité qu'un infecté rencontre un individu initialement susceptible vaut λ . On appellera $E_{n,m}(\lambda, I)$ une telle épidémie.

Un des paramètres importants dans ce modèle est le nombre d'infections attendues par un individu (*Basic reproduction number*) $R_0 = \lambda i$. On verra qu'en fonction de sa valeur, l'épidémie se propage à toute la population ou reste confinée à un plus petit nombre d'individus.

Le système ainsi construit n'est pas markovien dans le cas général, il faut pour cela choisir la loi I telle qu'elle est la propriété d'absence de mémoire (par exemple une exponentielle).

Dans la suite nous donnerons quelques propriétés de ce modèle avant de voir comment le généraliser à d'autres situations.

1.2 Construction de Sellke

Une construction équivalente, proposée par Sellke en 1983, se base sur l'idée d'affecter à chaque individu une capacité de résistance à l'épidémie et de définir, à partir du nombre d'infectés, une pression de l'épidémie et un critère d'infection sur la résistance de l'individu.

On va indexer la population de la façon suivante : les infectés initiaux seront notés $-(m-1), -(m-2), \dots -1, 0$ et les susceptibles de 1 à n . On associe à chaque individu un temps de contagion $I_{-(m-1)}, \dots, I_n$ i.i.d. de loi I . Et à chaque susceptible un seuil individuel Q_1, \dots, Q_n i.i.d. suivant une loi exponentielle de paramètre 1.

A chaque temps t , si $Y(t)$ est le nombre d'infectés par la maladie, on définit ainsi la pression de l'épidémie :

$$A(t) = \frac{\lambda}{n} \int_0^t Y(u) du$$

Un susceptible j est infecté au moment où la pression dépasse le seuil Q_j et devient alors contagieux pendant le temps I_j .

Ce processus est équivalent à la propagation d'un SIR. Cette construction facilite l'obtention de résultats exacts quand à la taille finale de l'épidémie.

Expression de la taille finale de l'épidémie

Exprimons la taille finale de l'épidémie, Z , en fonction de ces nouveaux paramètres. On réordonne les individus susceptibles en fonction de leur seuil croissant, on note alors $Q_{(1)}, \dots, Q_{(n)}$ ces seuils. L'épidémie s'arrête lorsque la pression exercée par les individus infectés n'est pas suffisante pour atteindre le seuil suivant. Lorsqu'un individu est infecté j , il participe à la pression pendant un temps I_j . Après la $i^{\text{ème}}$ infection, la pression reste inférieure à $\frac{\lambda}{n} \sum_{j=-(m-1)}^i I_j$. Donc si $Q_{(i+1)}$ est supérieur à cette valeur, l'épidémie s'arrête. Ainsi :

$$Z = \min\{i \geq 0, Q_{(i+1)} > \frac{\lambda}{n} \sum_{j=-(m-1)}^i I_j\}$$

Dans la suite on notera A la pression finale :

$$A = \frac{\lambda}{n} \sum_{j=-(m-1)}^Z I_j$$

Loi de probabilité de Z

Grâce à cette construction, on va pouvoir donner la loi de probabilité de la taille finale de l'épidémie.

Théorème 1.1. *Dans un tel modèle d'épidémie, les probabilités P_k^n que k individus initialement susceptibles aient été infectés par l'épidémie sont données par la formule suivante :*

$$\forall 0 \leq l \leq n, \sum_{k=0}^l \binom{n-k}{l-k} \frac{P_k^n}{\Phi(\lambda(n-l)/n)^{k+m}} = \binom{n}{l}$$

Pour calculer ces valeurs, on utilise le lemme suivant :

Lemme 1 (Identité de Wald pour les épidémies). *Soit une épidémie $E_{n,m}(\lambda, I)$, de pression finale A . Alors, pour tout $\theta \geq 0$, on a :*

$$E\left(\frac{e^{-\theta A}}{\Phi(\lambda\theta/n)^{Z+m}}\right) = 1$$

où Φ est la transformée de Laplace de I .

1.3 Approximation du comportement au début de l'épidémie

Quand la proportion de susceptible est importante, les infectés vont alors avec une forte probabilité, entrer en contact avec des individus qui sont toujours susceptibles, la propagation s'effectue donc comme dans un processus de branchement. Un argument de couplage va nous permettre de justifier cette idée.

Rappels sur les processus de branchement

On considère qu'au temps 0, il y a m ancêtres et que chaque individu a une durée de vie indépendante et de loi I . Pendant sa vie, un individu donne naissance à chaque saut d'un processus de Poisson d'intensité λ indépendant de I de des autres individus. On note $E_m(\lambda, I)$ une telle propagation. On note $Y(t)$ la taille de la population au temps t et D le nombre d'enfant par individu. A la $\nu^{\text{ème}}$ génération, en moyenne il y a $mE(D)^\nu$ individus. Intuitivement, on peut donc penser que suivant la valeur de $E(D)$, le comportement de la population va être différent.

Il est facile de voir que q la probabilité d'extinction du processus est solution de $\theta = E(\theta^D) = F(\theta)$, ou F est la fonction génératrice de D .

On a en fait plus précis :

Théorème 1.2. *La probabilité q d'extinction du processus de branchement est le plus petit point fixe de F . Si la variable aléatoire Z possède une espérance m , et que $m \leq 1$, alors 1 est le seul point fixe de F sur l'intervalle $[0; 1]$, et la probabilité d'extinction est 1.*

Si non, F admet exactement un point fixe sur $[0; 1]$, qui est la probabilité d'extinction du processus.

Dans le cas qui nous intéresse, le nombre de descendant à bien une moyenne et celle ci vaut λi le basic reproduction number.

Construction du couplage

On construit d'abord le processus de branchement $E_m(\lambda, I)$ puis on se donne l'espace $(\Omega, \mathcal{F}, \mathbf{P})$ contenant les vies de chaque individu $(\mathcal{H}_{-(m-1)}, \mathcal{H}_{-(m-2)}, \dots)$: chaque \mathcal{H}_i contient à la fois la durée de vie de chaque individu et les instants de naissance de ses descendants. Et $(U_i)_{i \geq 1}$ des variables i.i.d. uniformes sur $[0, 1]$ indépendantes des \mathcal{H}_i

On construit alors $E_{n,m}(\lambda, I)$ de la façon suivante :

- Les ancêtres initiaux sont les premiers infectés,
- A n fixé, on numérote de 1 à n les susceptibles. Un contact dans l'épidémie se produit lors d'une naissance. Le i ème individu touché est celui d'indice $C_i = [nU_i] + 1$. Si cet individu est encore susceptible, il devient infecté. Sinon l'individu du processus de branchement est considéré (ainsi que ses descendants) comme fantôme et ne prendra plus place dans le couplage.
- La mort d'un individu non fantôme correspond à son rétablissement.

Dans cette construction, les deux processus sont égaux jusqu'à l'apparition du premier fantôme à un instant T_n . Or tout choix fini de label C_i sera distinct alors lorsque n tend vers l'infini, on a , pour tout t_0 fixé $P(T_n > t_0) \rightarrow 1$.

Alors, pour tout t_0 fixé, si $Y_n(t)$ est le nombre d'infecté dans le modèle $E_{n,m}(\lambda, I)$ et $Y(t)$ le nombre d'individus dans le processus de branchement, on a la convergence suivante :

$$Y_n(t_0) \longrightarrow Y(t_0) p.s.$$

Monotonie

Un autre couplage lié à la construction de Sellke va nous permettre de comparer deux épidémies $E_{n,m}(\lambda, I)$ et $E_{n,m}(\lambda', I)$ où $\lambda < \lambda'$.

Pour les deux épidémies, on utilise les même périodes de contagion $I_{-(m-1)}, \dots, I_n$ et les même seuils épidémiques Q_1, \dots, Q_n . Seule la pression diffère et on a toujours $A(t) \leq A'(t)$. En particulier on a donc pour tout $j \geq 1$, $T'_j \leq T_j$.

2 Extension sur un graphe

Le modèle SIR, s'étend de manière naturelle sur un graphe. Dans le modèle initial, les infectés entrent en contact avec des individus de tout type choisit aléatoirement à un taux moyen β et se rétablissent à un taux moyen γ . Le graphe permet de considérer une hétérogénéité des contacts dans la population, et précise les relations entre les individus d'une population. Newman a montré [7] que le comportement de l'épidémie sur le graphe, peut être étudié par une percolation sur ce même graphe. On considère alors que la transmission de l'épidémie sur un graphe sépare les arêtes en deux types d'arêtes : les ouvertes par lesquelles l'épidémie a été transmise, et les autres. Dans ce cas, des résultats de percolation permettent d'avoir des renseignements sur la taille des composantes connexes et donc des excursions de l'épidémies (*outbreak*) a la fin de l'épidémie. De tels résultats donnent des informations sur l'importance de l'épidémie, mais pas sur son évolution au cours du temps. C'est cette dernière approche que nous allons maintenant étudier en nous restreignant à un graphe particulier.

2.1 Le graphe des configurations

Le graphe des configurations (*Configuration Model Graph*) est un graphe aléatoire pour lequel sont donnés, les degrés des noeuds qui sont distribués de façon indépendantes selon la loi $P(d = k) = p_k$, pour $k \geq 1$. Donc, autour de chaque noeud x il y a d_x demi arêtes. Puis on apparie aléatoirement les demi-arêtes, ce qui donne un graphe de relations/arêtes entre noeuds/individus. Une telle construction est possible dès que la somme des degrés est paire. Cette construction ne donne pas un graphe simple : il peut y avoir des boucles ou des arêtes parallèles. Ainsi, chaque graphe possible a la même probabilité d'être construit : c'est pourquoi on parle de graphe des configurations.

Dans la suite, les noeuds du graphe représenteront les individus de la population et chaque arête, un contact entre deux individus.

2.2 Propagation de l'épidémie sur un tel graphe

Ce graphe nous donne, pour chaque individu, les informations suivantes : nombre d'arêtes, de voisins et l'état des voisins. Pour décrire l'état du système, il est nécessaire de connaître l'état de chaque individu, ainsi que ses connexions. Pour l'ensemble S_t des susceptibles, la mesure (somme de mesures de Dirac) :

$$\mu_t^S = \sum_{x \in S_t} \delta_{d_x} \quad (2.1)$$

décrit la répartition des degrés des susceptibles.

Pour ce qui est des infectés I_t , et des rétablis R_t on va décrire l'état des arêtes reliant un infecté (resp. rétabli) à un susceptible dans S_t . Ces arêtes jouent un rôle particulier car elles peuvent encore être actives dans l'épidémie. Si x est un infecté à l'instant t , son degré en direction des susceptibles à cet instant sera noté : $d_x(S_t)$. Ce qui nous amène à définir :

$$\mu_t^{IS} = \sum_{x \in I_t} \delta_{d_x(S_t)} \quad \mu_t^{RS} = \sum_{x \in R_t} \delta_{d_x(S_t)}$$

La connaissance de ces trois mesures nous donne celle du système entier à un instant donné.

En effet, à partir de ces mesures, il est possible de retrouver les nombres d'individus de chaque classe, mais aussi d'avoir des informations sur le nombre d'arêtes sortant des susceptibles, ainsi que de connexions possibles IS et RS. On notera \mathcal{X}^p la fonction $x \rightarrow x^p$, ce qui permet par exemple d'écrire : $\langle \mu_t^{RS}, \mathbf{1} \rangle = R_t$ et $N_t^{RS} = \langle \mu_t^{RS}, \mathcal{X} \rangle$ le nombre de demi arêtes de R_t vers S_t . De même, les quantités I_t (resp. S_t) et N_t^{IS} (resp. N_t^S) s'obtiennent facilement à partir de la mesure μ_t^{IS} (resp. μ_t^S).

Au temps $t = 0$, la population est constituée de m individus infectés et n susceptibles. Le degré de chacun de ces individus est alors tiré indépendamment des autres suivant une loi fixée à l'avance, ce qui permet de construire le graphe des configurations associé. Dans l'évolution de l'épidémie, on ne connaît pas tout le graphe dès le début. On a seulement la partie de l'information contenue dans les relations entre infectés et rétablis et les relations possibles avec les susceptibles. En effet, ce n'est qu'au cours de l'évolution que l'on découvre qui infecte qui et comment ces nouveaux infectés interagissent avec la population. Le point clé de ce modèle est qu'il est possible de construire simultanément le graphe et l'épidémie. Ceci permet de faciliter la mise en équation du problème ainsi que la simulation. Pour modéliser la différence entre les relations que l'on connaît et les relations possibles qu'un individu peut former, on dessinera des arêtes lorsque l'on sait qu'une connexion a eu lieu, et des demi-arêtes lorsque l'on sait qu'une connexion peut avoir lieu, mais pas encore avec quel autre individu. Ces dernières seront considérées comme des demi-arêtes libres.

L'évolution de l'épidémie se propage alors comme suit. Chaque infecté dispose d'une horloge signifiant la fin de la durée de contagion. Toutes ces horloges sont indépendantes et suivent une loi exponentielle de paramètre β . De même chaque demi arête IS est munie d'une horloge de loi exponentielle de paramètre r qui marque la rencontre avec une autre demi arête.

Commençons par introduire quelques notations qui permettront de numéroter les individus en fonction de leur degré. Si μ est une mesure finie de \mathbf{N} dans lui-même, (qui représentera ensuite des degrés d'individus), on

note $F_\mu(m) = \mu(\{0, \dots, m\})$ la fonction de répartition de μ . $F_\mu(m)$ est alors le nombre d'individus de degrés inférieur ou égal à m . Puis on définit son inverse à droite $F_\mu^{-1}(i) = \inf\{m, F_\mu(m) \geq i\}$ qui donne donc le degré du $i^{\text{ème}}$ individus quand on les classe par degrés croissant. On notera maintenant pour tout $1 \leq i \leq S_t$

$$\gamma_i(\mu_t^S) = F_{\mu_t^S}^{-1}(i)$$

le degré du $i^{\text{ème}}$ individu susceptible au temps t .

De même on définira $\gamma_i(\mu_t^{IS})$ et $\gamma_i(\mu_t^{RS})$ le nombre de demi arêtes au temps t du $i^{\text{ème}}$ infecté (resp. rétabli) vers l'ensemble des susceptibles.

Ces notations permettent en particulier de retrouver les distributions à chaque instant des degrés des susceptibles et du nombre de demi-arêtes IS et RS. En effet, si i est tiré uniformément parmi les entiers de 1 à I_t alors $\gamma_i(\mu_t^{IS})$ donne le nombre de demi arêtes IS d'un individu tiré au hasard parmi les infectés.

Le système est donc régi par des événements qui se produisent aux instants de sauts de processus de Poisson indépendants. A un temps t , le prochain événement aura lieu après un temps distribué suivant une loi exponentielle de paramètre $\beta I_t + r N_t^{IS}$. A ce temps T on a deux possibilités :

1) L'un des individus se rétablit avec probabilité $\beta I_t / (\beta I_t + r N_t^{IS})$

On choisit uniformément parmi I_{T-} quel individu i va passer du statut I à R puis on met à jour les mesures :

$$\begin{aligned} \mu_T^{IS} &= \mu_{T-}^{IS} - \delta_{\gamma_i(\mu_{T-}^{IS})} \\ \mu_T^{RS} &= \mu_{T-}^{RS} + \delta_{\gamma_i(\mu_{T-}^{IS})} \end{aligned} \quad (2.2)$$

2) Un infecté entre en contact avec un individu susceptible

Une des demi arêtes IS infecte un individu susceptible. Le choix de cette demi arête interviendra un peu plus loin : on s'intéresse d'abord à l'individu qui va être infecté.

Le nouvel infecté est choisi de la façon suivante : parmi toutes les demi arêtes dont l'une des extrémités est un susceptible (c'est à dire parmi N_{T-}^S), on choisit celle par laquelle passera l'infection uniformément. Ainsi la probabilité d'atteindre un individu de degré k vaut $k \mu_{T-}^S(\{k\}) / N_{T-}^S$.

Ce nouvel infecté, établit lui aussi des connexions avec les infectés et les rétablis dont l'on notera le nombre respectivement j et ℓ . Ces liaisons sont choisies de manière uniforme parmi les demi-arêtes restantes. Ainsi, étant donné k le degré du nouvel infecté, la probabilité de former j arêtes II et ℓ arêtes IR suit une loi hypergéométrique :

$$p(j, \ell | k-1) = \frac{\binom{N_{T-}^{IS}-1}{j} \binom{N_{T-}^{RS}}{\ell} \binom{N_{T-}^S - N_{T-}^{IS} - N_{T-}^{RS}}{k-1-j-\ell}}{\binom{N_{T-}^S-1}{k-1}} \quad (2.3)$$

Pour mettre à jour les mesures, il faut savoir exactement entre quels individus se sont produit les liaisons : les deux vecteurs $u = (u_1, \dots, u_{I_{T-}})$ et $v = (v_1, \dots, v_{R_{T-}})$ indiqueront combien de liaisons chaque individu infecté (resp. rétabli) forme avec le nouvel infecté : u_i (resp. v_i) est le nombre de liaison du $i^{\text{ème}}$ infecté (resp. rétabli) avec le nouvel infecté.

Ces vecteurs sont soumis à des contraintes liées au nombre de demi arêtes disponibles pour former les liaisons.

Définissons alors

$$\mathcal{U} = \bigcup_{m=1}^{\infty} \mathbf{N}^m \quad (2.4)$$

Si n est un entier et μ une mesure finie de \mathbf{N} dans \mathbf{N} , l'ensemble des vecteurs, indiquant une configuration possible des n connexions parmi des éléments dont les degrés sont donnés par μ est alors décrit par :

$$\mathcal{U}(n, \mu) = \{u = (u_1, \dots, u_{\langle \mu, \mathbf{1} \rangle}) \in \mathcal{U} / \forall 1 \leq i \leq \langle \mu, \mathbf{1} \rangle, u_i \leq F_\mu^{-1}(i) \text{ et } \sum_{i=1}^{\langle \mu, \mathbf{1} \rangle} u_i = n\} \quad (2.5)$$

Avec ces notations les vecteurs u et v précédents qui indiquent la répartition des connexion IS (resp. RS) qui deviennent II (resp. RI) lors de la nouvelle infection appartiennent respectivement à $\mathcal{U}(j+1, \mu_{T-}^{IS})$ et $\mathcal{U}(\ell, \mu_{T-}^{RS})$. En effet, il se forme j liaisons entre le nouvel infecté et les infectés et 1 liaison qui est responsable de l'infection. Toutes les demi arêtes en eu dans ses liaisons sont tirées uniformément dans N_{T-}^{IS} . Donc $u \in \mathcal{U}(j+1, \mu_{T-}^{IS})$. Et de même pour les ℓ liaisons vers les rétablis sont tirées parmi N_{T-}^{RS} . Les probabilités de tirer u et v sont alors données par les formules suivantes :

$$\begin{aligned}\rho(u|j+1, \mu_{T-}^{IS}) &= \frac{\prod_{i=1}^{I_{T-}} \binom{\gamma_i(\mu_{T-}^{IS})}{u_i}}{\binom{N_{T-}^{IS}}{j+1}} \mathbf{1}_{u \in \mathcal{U}(j+1, \mu_{T-}^{IS})} \\ \rho(v|\ell, \mu_{T-}^{RS}) &= \frac{\prod_{i=1}^{R_{T-}} \binom{\gamma_i(\mu_{T-}^{RS})}{v_i}}{\binom{N_{T-}^{RS}}{\ell}} \mathbf{1}_{v \in \mathcal{U}(\ell, \mu_{T-}^{RS})}\end{aligned}\tag{2.6}$$

La première expression par exemple traduit le fait que pour chaque infecté i , il faut tirer indépendamment le nombre d'arêtes u_i parmi celles disponibles de telle façon que u soit dans $\mathcal{U}(j+1, \mu_{T-}^{IS})$.

Une fois choisis les paramètres k, j, l, u et v , l'évolution s'écrit donc au temps T :

$$\begin{aligned}\mu_T^S &= \mu_{T-}^S - \delta_k \\ \mu_T^{IS} &= \mu_{T-}^{IS} + \delta_{k-1-j-\ell} + \sum_{i=1}^{I_{T-}} \delta_{\gamma_i(\mu_{T-}^{IS})-u_i} - \delta_{\gamma_i(\mu_{T-}^{IS})} \\ \mu_T^{RS} &= \mu_{T-}^{RS} + \sum_{i=1}^{R_{T-}} \delta_{\gamma_i(\mu_{T-}^{RS})-v_i} - \delta_{\gamma_i(\mu_{T-}^{RS})}\end{aligned}\tag{2.7}$$

Ainsi construite, une épidémie sur le graphe des configurations, vérifie un système couplé d'équations différentielles stochastiques basées sur les mesures de Poisson. On considère deux mesures ponctuelles de Poisson pour décrire les deux phénomènes de la propagation : infection et rétablissement.

- Pour décrire l'infection d'un individu, on utilise $dQ_1(s, k, \theta_1, j, \ell, \theta_2, u, \theta_3, v, \theta_4)$ définie sur $\mathbf{R}^+ \times E_1$ avec $E_1 = \mathbf{N} \times \mathbf{R}^+ \times \mathbf{N} \times \mathbf{N} \times \mathbf{R}^+ \times \mathcal{U} \times \mathbf{R}^+ \times \mathcal{U} \times \mathbf{R}^+$ avec comme intensité, les mesures de comptage et de Lebesgue associées aux différents espaces ;
- et pour ce qui est du rétablissement, on se donne $dQ_2(s, i)$ définie sur $\mathbf{R}^+ \times \mathbf{N}$ de mesure d'intensité le produit de la mesure de Lebesgue et de la mesure de comptage.

On notera :

$$\mu_t^S(k) \lambda_t(k) = r N_t^{IS} \frac{k \mu_t^S(k)}{N_t^S}\tag{2.8}$$

Le système s'écrit alors :

$$\mu_t^S = \mu_0^S - \int_0^t \int_{E_1} \delta_k \mathbf{1}_{\theta_1 \leq \lambda_{s-}(k) \mu_{s-}^S(k)} \mathbf{1}_{\theta_2 \leq p_{s-}(j, \ell | k-1)} \mathbf{1}_{\theta_3 \leq \rho(u|j+1, \mu_{s-}^{IS})} \mathbf{1}_{\theta_4 \leq \rho(v|\ell, \mu_{s-}^{RS})} dQ_1\tag{2.9}$$

$$\begin{aligned}\mu_t^{IS} &= \mu_0^{IS} + \int_0^t \int_{E_1} \left[\delta_{k-(j+\ell+1)} + \sum_{i=1}^{I_{s-}} \delta_{\gamma_i(\mu_{s-}^{IS})-u_i} - \delta_{\gamma_i(\mu_{s-}^{IS})} \right] \\ &\quad \times \mathbf{1}_{\theta_1 \leq \lambda_{s-}(k) \mu_{s-}^S(k)} \mathbf{1}_{\theta_2 \leq p_{s-}(j, \ell | k-1)} \mathbf{1}_{\theta_3 \leq \rho(u|j+1, \mu_{s-}^{IS})} \mathbf{1}_{\theta_4 \leq \rho(v|\ell, \mu_{s-}^{RS})} dQ_1 \\ &\quad - \int_0^t \int_{\mathbf{N}} \beta \delta_{\gamma_i(\mu_{s-}^{IS})} \mathbf{1}_{i \leq I_{s-}} dQ_2\end{aligned}\tag{2.10}$$

$$\begin{aligned}
\mu_t^{RS} &= \mu_0^{RS} + \int_0^t \int_{E_1} \left[\sum_{i=1}^{R_{s-}} \delta_{\gamma_i(\mu_{s-}^{RS}) - v_i} - \delta_{\gamma_i(\mu_{s-}^{RS})} \right] \\
&\quad \times \mathbf{1}_{\theta_1 \leq \lambda_{s-}(k) \mu_{s-}^S(k)} \mathbf{1}_{\theta_2 \leq p_{s-}(j, \ell | k-1)} \mathbf{1}_{\theta_3 \leq \rho(u | j+1, \mu_{s-}^{IS})} \mathbf{1}_{\theta_4 \leq \rho(v | \ell, \mu_{s-}^{RS})} dQ_1 \\
&\quad + \int_0^t \int_{\mathbf{N}} \beta \delta_{\gamma_i(\mu_{s-}^{IS})} \mathbf{1}_{i \leq I_{s-}} dQ_2
\end{aligned} \tag{2.11}$$

Remarque 1 : Les auteurs de [4] ont préféré garder des processus de Poissons avec des mesures d'intensité simples et complexifier leurs équations par la présence d'indicatrices. On verra plus explicitement lors du calcul des parties martingales et à variation finie de ces mesures en quoi ces équations reflètent bien la construction expliquée précédemment. On peut déjà remarquer que les variables k, j, l, u, v représentent les même quantités que celles définies précédemment, et les variables en θ sont celles qui donnent l'intensité de la mesure de Poisson.

Remarque 2 : La construction définit précédemment, permet de fabriquer une solution de ce système. Pour montrer que celle-ci est unique sur tout intervalle de temps borné, il suffit alors de remarquer que les mesures de Poisson sont définies sur un espace de masse finie. Alors, les processus de Poisson ne sauteront qu'un nombre fini de fois sur tout espace de temps compact, et donc cette construction définit une unique solution au système.

Remarque 3 : A partir de ces équations, il est possible de calculer la décomposition en martingale et partie à variations finies de chacune des mesures. Le but est d'écrire

$$\langle \mu_t^{IS}, f \rangle = \langle \mu_0^{IS}, f \rangle + A_t^{IS,f} + M_t^{IS,f} \tag{2.12}$$

Ces résultats proviennent directement des propriétés du calcul stochastique semi martingales à variations finies [8, 5].

La partie à variation finie correspond au compensateur de la mesure $\langle \mu_t^{IS}, f \rangle$. Et on obtient directement le crochet de la partie martingale, comme le carré des sauts.

2.3 Limite en grande population

Ce qui nous intéresse est le comportement limite en grande population. On indexera toutes les mesures par n qui croit proportionnellement à la taille de la population et on va travailler sur les versions renormalisées de ces mesures :

$$\mu_t^{(n),S} = \frac{1}{n} \mu_t^{n,S}, \quad \mu_t^{(n),IS} = \frac{1}{n} \mu_t^{n,IS}, \quad \mu_t^{(n),RS} = \frac{1}{n} \mu_t^{n,RS}$$

Le but de ce paragraphe est de montrer que les suites de mesures préalablement définies convergent vers un triplet de mesures $\bar{\mu} = (\bar{\mu}^S, \bar{\mu}^{IR}, \bar{\mu}^{RS})$ solution du système d'équations différentielles ordinaires suivant sous certaines hypothèses.

Ces équations ont été proposées par Volz [9] et mettent en jeu les proportions d'arêtes IS, RS et SS qui s'écrivent $\bar{p}_s^I = \frac{\bar{N}_s^{IS}}{\bar{N}_s^S}$, \bar{p}_s^R et \bar{p}_s^S . On notera aussi θ_t la probabilité qu'un individu de degré un reste susceptible jusqu'au temps t :

$$\theta_t = \exp\left(-r \int_0^t \bar{p}_s^I ds\right)$$

Ainsi, si l'individu est de degré k la probabilité qu'il soit encore susceptible au temps t est θ_t^k . Ceci permet de comprendre facilement l'équation pour $\bar{\mu}_t^S$:

$$\langle \bar{\mu}_t^S, f \rangle = \sum_{k \geq 0} \mu_0^S(k) \theta_t^k f(k) \tag{2.13}$$

Pour comprendre les deux autres équations, il faut voir que le terme de rémission $\int_0^t \beta \langle \bar{\mu}_s^{IS}, f \rangle ds$ est le même que pour les équations stochastiques. Pour ce qui est de l'infection, le taux d'infection donné précédemment par λ s'écrit maintenant $rk\bar{p}^I$. Et que comme les arêtes du graphes sont simples presque surement, le nombre de connexions entre le nouvel infecté et l'ensembles des infectés (et des rétablis) est distribué suivant une loi multinomiale : c'est la deuxième intégrale. Chaque infecté (ou rétabli) est ensuite choisi suivant la distribution biaisée en degré ; d'où la troisième intégrale.

$$\begin{aligned} \langle \bar{\mu}_t^{IS}, f \rangle &= \langle \bar{\mu}_0^{IS}, f \rangle - \int_0^t \beta \langle \bar{\mu}_s^{IS}, f \rangle ds + \int_0^t \sum_{k \geq 0} rk\bar{p}_s^I \sum_{j+\ell+m=k-1} \binom{n}{j, \ell, m} (\bar{p}_s^I)^j (\bar{p}_s^R)^\ell (\bar{p}_s^S)^m f(m) \bar{\mu}_s^S(k) ds \\ &\quad + \int_0^t \sum_{\geq 0} rk\bar{p}_s^I (1 + (k-1)\bar{p}_s^I) \sum_{k' \geq 0} (f(k'-1) - f(k')) \frac{k' \bar{\mu}_s^{IS}(k')}{N_s^{IS}} \bar{\mu}_s^S(k) ds \end{aligned} \quad (2.14)$$

$$\langle \bar{\mu}_t^{RS}, f \rangle = \langle \bar{\mu}_0^{RS}, f \rangle + \int_0^t \beta \langle \bar{\mu}_s^{IS}, f \rangle ds + \int_0^t \sum_{k \geq 0} rk\bar{p}_s^I (k-1)\bar{p}_s^R \sum_{k' \geq 0} (f(k'-1) - f(k')) \frac{k' \bar{\mu}_s^{RS}(k')}{N_s^{RS}} \bar{\mu}_s^S(k) ds \quad (2.15)$$

Pour obtenir l'unicité de la solution de ce système (si elle existe) on se place sur un ensemble de mesures plus restreint, où l'on a un contrôle sur certains moments : pour tout $\varepsilon \geq 0$ et tout A ,

$$\mathcal{M}_{\varepsilon, A} = \{ \mu \in \mathcal{M}_F(\mathbf{N}), \langle \mu, \mathbf{1} + \mathcal{X}^5 \rangle \leq A \text{ et } \langle \mu, \mathcal{X} \rangle \geq \varepsilon \} \quad (2.16)$$

et on définira également :

$$\mathcal{M}_{0+, A} = \{ \mu \in \mathcal{M}_F(\mathbf{N}), \langle \mu, \mathbf{1} + \mathcal{X}^5 \rangle \leq A \text{ et } \langle \mu, \mathcal{X} \rangle > 0 \} \quad (2.17)$$

Théorème 2.1. *Pour tout $A > 0$, le système (2.13)-(2.14)-(2.15) possède au plus une solution dans l'espace $\mathcal{C}(\mathbf{R}^+, \mathcal{M}_{0, A} \times \mathcal{M}_{0, A} \times \mathcal{M}_{0, A})$.*

On montre d'abord que pour tout T fixé et $\varepsilon > 0$, le système (2.13)-(2.14)-(2.15) a au plus une solution dans $\mathcal{C}([0, T], \mathcal{M}_{0, A} \times \mathcal{M}_{\varepsilon, A} \times \mathcal{M}_{0, A})$. Pour montrer maintenant l'unicité recherchée, il suffit de vérifier que le terme N_t^{IS} ne s'annule jamais.

Théorème 2.2. *On suppose que les suites $(\mu_0^{(n), S})_{n \geq 0}$ et $(\mu_0^{(n), IS})_{n \geq 0}$ et $(\mu_0^{(n), RS})_{n \geq 0}$ convergent vers des mesures $\bar{\mu}_0^S$, $\bar{\mu}_0^{IS}$ et $\bar{\mu}_0^{RS}$ dans $\mathcal{M}_F(\mathbf{N})$ muni de la convergence faible. Et qu'il existe $0 < \bar{\varepsilon} < A$ tels que*

$$\forall n, (\mu_0^{(n), S}, \mu_0^{(n), IS}, \mu_0^{(n), RS}) \in \mathcal{M}_{0, A}^3, \text{ et que } \langle \bar{\mu}_0^{IS}, \mathcal{X} \rangle > \bar{\varepsilon} \quad (2.18)$$

Alors quand n tend vers l'infini, la suite $(\mu^{(n)})_{n \geq 0}$ converge en distribution dans $\mathcal{D}(\mathbf{R}^+, \mathcal{M}_{0, A} \times \mathcal{M}_{0+, A} \times \mathcal{M}_{0, A})$ vers $\bar{\mu}$ pour la topologie de la convergence étroite.

Parmi ces hypothèses, il est important pour que la convergence reste vraie que le nombre d'arêtes IS ne s'annule jamais. C'est pourquoi on définit pour tout $\varepsilon' > 0$:

$$t_{\varepsilon'} = \inf \{ t, \langle \bar{\mu}_t^{IS}, \mathcal{X} \rangle < \varepsilon' \}$$

Démonstration. La preuve va en fait se dérouler de la façon suivante, pour tout $\varepsilon' \leq \bar{\varepsilon}$, on va montrer que l'on a convergence dans $\mathcal{D}([0, t_{\varepsilon'}], \mathcal{M}_{0, A} \times \mathcal{M}_{0+, A} \times \mathcal{M}_{0, A})$.

Le moment $\langle \bar{\mu}_t^{IS}, \mathcal{X} \rangle$ ne s'annulant jamais, on aura alors la convergence sur tous les intervalles $[0, T]$, vers une limite continue, ce qui équivaut à la convergence dans $\mathcal{D}(\mathbf{R}^+, E)$ [2].

Soit $\varepsilon < \varepsilon'$. On va travailler sur la suite de processus arrêtés en les temps d'arrêts :

$$\tau_\varepsilon^n = \inf \{ t \geq 0, \langle \mu_t^{(n), IS}, \mathcal{X} \rangle < \varepsilon \} \quad (2.19)$$

On montrera d'abord que cette suite est tendue, puis que ses limites satisfont, sur un certain intervalle de temps, les équations différentielles voulues.

Il est tout d'abord facile de voir que les suites $(\mu_t^{(n),S})_n$, $(\mu_t^{(n),IS})_n$ et $(\mu_t^{(n),RS})_n$ sont tendues pour tout $t > 0$ car on a une condition sur leurs moments. Pour avoir la tension dans $\mathcal{D}(\mathbf{R}^+, \mathcal{M}_{0,A}^3)$, commençons par quelques résultats sur la convergence des processus à valeurs mesures.

Théorème 2.3. *Soit E un espace métrique localement compact. Notons, (\mathcal{M}_F, v) l'ensemble des mesures finies et positives sur E , muni de la topologie vague.*

Soit P_n une suite de mesures de probabilités à valeurs dans $\mathcal{D}([0, T], \mathcal{M}_F)$ et $(f_k)_{k \geq 1}$ une suite de fonction dense dans $\mathcal{C}_0 = \{f \text{ continues et nulles à l'infini}\}$ avec $f_0 = 1$.

Si l'on a les conditions suivantes

- *Pour tout k , $(\langle P_n, f_k \rangle)_{n \geq 0}$ est tendue dans $\mathcal{D}([0, T], \mathbf{R})$,*
 - *Toute valeur d'adhérence de (P_n) est continue,*
 - *La suite des masses $(\langle P_n, \mathbf{1} \rangle)_{n \geq 0}$ est tendue dans $\mathcal{D}([0, T], \mathbf{R})$*
- alors (P_n) est tendue dans $\mathcal{D}([0, T], (\mathcal{M}_F, \acute{e}))$*

Ainsi, il nous suffit d'avoir la tension dans $\mathcal{D}(\mathbf{R}^+, \mathbf{R}^3)$ des suites $(\langle \mu^{(n),S}, f \rangle, \langle \mu^{(n),IS}, f \rangle, \langle \mu^{(n),RS}, f \rangle)_n$ avec $f \in \mathcal{B}_b(\mathbf{N}, \mathbf{R})$: ensemble des fonctions réelles bornées sur \mathbf{N} . (Ce qui nous donnera en particulier la tension des masses).

Comme on connaît les décompositions en semi-martingales de ces processus, on utilise le critère d'Aldous-Rebolledo qui implique que pour vérifier la tension d'une semi-martingale, il suffit d'avoir la tension à temps fixé et que la partie à variation finie et le crochet de la partie martingale vérifient le critère d'Aldous [5].

Ainsi, on a bien la tension recherchée, et par le théorème de Prokhorov, cette famille est relativement compacte, donc celle des processus arrêtés aussi. On choisit donc une valeur d'adhérence : comme ces processus font des sauts de taille tendant vers 0, on peut choisir cette limite continue. [2].

A partir de cela, on vérifie que la suite des mesures arrêtées vérifie presque l'équation voulue (c'est à dire à un terme près convergeant vers 0). Cette partie technique provient des propriétés du graphe et du fait qu'étant donné que le nombre de liaisons IS reste grand, on peut minorer le nombre de liaisons multiples que le nouvel infecté forme avec les rétablis. Ainsi, ce terme va tendre vers 0, ne laissant que le terme contenant des liaisons simples, qui correspond à l'équation déterministe.

Enfin, grâce aux hypothèses sur les moments et à des arguments de continuité, on montre que la suite d'origine converge.

□

3 Perspectives

Une épidémie est souvent reliée à un graphe particulier : par exemple, dans le cas du SARS ou d'autres maladies extrêmement contagieuses, il est opportun de considérer un graphe des relations aériennes pour mieux comprendre la propagation au niveau mondial [3]. D'autres travaux ont mis en évidence que les relations sexuelles par exemple pouvaient être modélisées par des graphes petit monde [6], ainsi pour l'étude des maladies sexuellement transmissibles, il serait très important de transposer ce type de résultats sur ces graphes particuliers.

Toutefois, il n'est pas possible de le faire de cette manière, car ici on a combiné l'aléa du graphe et celui de la maladie en un seul aléa. Cette méthode est très particulière au graphe des configuration. Faire évoluer une épidémie sur un graphe aléatoire plus complexe relève de la propagation de processus aléatoires en milieu aléatoire.

Références

- [1] H. Andersson and T. Britton. *Stochastic Epidemic Models and their Statistical Analysis*. Lecture Notes in Statistics, Springer Verlag, 2000.
- [2] P. Billingsley. *Convergence of Probability Measures, 2nd Edition*. Wiley-Interscience, 1999.

- [3] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases : the sars case study. *BMC Medicine*, 2007.
- [4] L. Decreusefond, J-S. Dhersin, P. Moyal, and V. C. Tran. Large graph limit for an sir process in random network with heterogeneous connectivity. *Annals of Applied Probability*, à paraître.
- [5] A. Joffe and M. Metivier. Weak convergence of sequences of semimartingales. *Advanced Applied Probability*, 1986.
- [6] K. Kaski and J.Saramäki. Modeling epidemics with dynamic small-world networks. *Journal of Theoretical Biology* 234/3, 413-421, 2005.
- [7] M.E.J Newman. Spread of epidemic disease on networks. *Physical Review E*66, 2002.
- [8] P. Protter. *Stochastic Integration and Differential Equations*. Springer Verlag, 1990.
- [9] E. Volz. Sir dynamics in structured populations with heterogeneous connectivity. *Journal of Mathematical Biology*, 2008.