

Introduction aux Processus Stochastiques

Projet 1: Chaînes de Markov en temps discret

Analyse de la propagation d'un virus au sein d'un réseau

Profs.: Amir ABOUBACAR et Pierre GEURTS

Année académique 2019-2020

Ce travail est à réaliser par groupe de 2 étudiants maximum pour les informaticiens et 3 étudiants maximum pour les mathématiciens. Il doit être remis via la plateforme de soumission de Montefiore pour le dimanche 19/04/2019 à 23h59.

L'archive soumise devra contenir les fichiers suivants :

- Votre rapport avec les réponses aux questions au format pdf.
- Le code permettant de répondre aux questions.

Le travail peut être réalisé au choix en utilisant MATLAB, Mathematica, R, Python, Java, ou C.

Contexte général et objectifs

Dans ce projet, et en lien avec l'actualité, on se propose de simuler la propagation d'un virus au sein d'une population de N individus en utilisant un modèle simplifié basé sur les chaînes de Markov en temps discret. Ce modèle a été proposé par exemple dans cet article [1]. La première partie consiste à étudier le modèle exact dans le cas où le nombre d'individus est réduit. La deuxième partie consiste à traiter le cas où N est grand via des simulations numériques.

Modèle de propagation du virus

On suppose une population de N individus pouvant être dans trois catégories par rapport à un virus¹ :

- Susceptibles (S) : des individus sains n'ayant jamais été infectés par le virus
- Infectieux (I) : des individus infectés par le virus et donc malades
- Immunisés (R) : des individus ayant été infectés précédemment mais guéris et maintenant immunisés contre une réinfection.

Pour simplifier le modèle, on ignorera les individus décédés et on ne tiendra pas compte non plus du fait qu'un individu est mis en quarantaine dès qu'il montre les symptômes de la

1. https://fr.wikipedia.org/wiki/Modèles_compartimentaux_en_épidémiologie

maladie. Ces deux simplifications n'auront cependant pas un grand impact sur la propagation de la maladie. En effet, du point de vue de la propagation du virus, les personnes décédées et les personnes en quarantaine sont équivalentes à des personnes immunisées qui ne peuvent plus transmettre la maladie.

Contrairement à des modèles plus macroscopiques basés sur des équations différentielles², on souhaite dans ce projet étudier la propagation de la maladie au niveau des individus eux-mêmes en prenant en compte les interactions entre eux. Pour modéliser ces interactions, on supposera donné un graphe (non dirigé) reliant ces individus, sous la forme d'une matrice d'adjacence $W \in \{0, 1\}^{N \times N}$. Un élément $W_{i,j}$ vaudra 1 si les individus i et j sont susceptibles de se rencontrer physiquement (et donc de se transmettre le virus), 0 sinon.

Etant donné ces hypothèses, le modèle de propagation en temps discret proposé est le suivant :

- Un individu infectieux au temps t a une probabilité β d'infecter au temps $t + 1$ chacun des individus susceptibles auquel il est connecté dans le graphe (les individus infectieux ou immunisés ne peuvent plus être ré-infectés)
- Un individu infectieux au temps t a une probabilité μ d'être guéri, et donc immunisé, au temps $t + 1$. La guérison d'un patient au temps $t + 1$ n'aura pas d'impact sur sa faculté à infecter d'autres individus au temps t selon la première règle.

La probabilité β modélise le fait qu'un individu ne rencontre pas nécessairement tous ses contacts à chaque pas de temps et qu'interagir avec une autre individu ne va pas nécessairement mener à une infection.

Tel que décrit le processus est un processus de Markov en temps discret. Les états de la chaîne correspondante sont représentés par les états, parmi 3, de l'ensemble des N individus de la population. Le nombre d'états de la chaîne est donc 3^N .

1 Etude du modèle exact

Dans un premier temps, on se propose d'étudier le modèle exact dans le cas où N est petit. On vous demande pour cela d'écrire un programme informatique permettant à partir d'un nombre N d'individus et d'une matrice d'adjacence W arbitraire de calculer la matrice de transition de la chaîne de Markov correspondant au processus décrit ci-dessus. On considérera dans cette section deux matrices d'adjacence W^{lin} telle que $W_{i,j}^{lin} = 1$ si et seulement si $|i - j| = 1$ et W^{full} telle que $W_{i,j}^{full} = 1$ si et seulement si $i \neq j$.

Dans votre rapport, répondez aux questions suivantes :

1. Justifiez que le modèle proposé est bien un processus de Markov et caractérisez de la manière la plus précise possible la chaîne de Markov définie par le processus de propagation du virus. Est-elle (a)périodique, irréductible, régulière, absorbante ? Comment ces propriétés dépendent-elles du graphe de connexion W ?
2. Combien de temps en moyenne faudra-t-il à un individu pour guérir une fois infecté ?
3. En vous basant sur la matrice de transition, calculez et tracez sur un graphe l'évolution en fonction du temps t de la proportion moyenne d'individus dans chacune des trois catégories (S , I , et R) pour les deux matrices W^{lin} et W^{full} . Fixez dans les deux cas N à 6, β à 0.5, μ à 0.2 et supposez qu'initialement un seul individu est infecté (avec

2. Par exemple, ici : <http://gabgoh.github.io/COVID/index.html>.

la même probabilité que ce soit chacun des N individus). Comparez et commentez les courbes que vous obtenez.

4. Toujours en vous basant sur la matrice de transition, calculez dans les deux mêmes cas qu'au point précédent le temps moyen nécessaire à la disparition totale du virus. Expliquez les différences obtenues.

2 Etude sur base de simulations

Dans le cas où N est grand, calculer explicitement la matrice de transition est impossible, le nombre d'états devenant trop élevé. Dans l'article [1], les auteurs dérivent une approximation du modèle en faisant l'hypothèse que les états des individus sont indépendants les uns des autres à chaque pas de temps t , c'est-à-dire :

$$P(X_1(t), \dots, X_N(t)) = P(X_1(t)) \dots P(X_N(t)),$$

où $X_i(t)$ est l'état de l'individu i au temps t . Sur base de ces hypothèses, il est possible de calculer facilement l'évolution des $P(X_i(t))$ de chaque individu au cours du temps. Cette hypothèse d'indépendance n'est cependant pas vraie en général et cette approche peut mener à des conclusions erronées.

Dans ce projet, on propose plutôt d'utiliser une approche empirique. On peut en effet estimer les mêmes courbes et statistiques qu'au point précédent en se basant sur une simulation de la chaîne de Markov. On vous demande pour cela d'écrire un programme permettant de générer une réalisation aléatoire de la chaîne, représentée par l'évolution au cours du temps de l'état des N individus, étant donné une matrice W et des valeurs de β et μ fixées a priori. Sur base de cette réalisation, vous devrez être capable de mesurer à chaque pas de temps les proportions d'individus dans les trois classes et de calculer le temps nécessaire à la disparition du virus. Pour répondre aux questions ci-dessous, une matrice d'adjacence W^{big} vous est fournie représentant un graphe défini sur 2000 individus comportant 3993 arêtes (non dirigées). Cette matrice a été générée aléatoirement selon un modèle "scale-free".

Sur base de votre programme, répondez aux questions suivantes :

1. Expliquez pourquoi l'hypothèse d'indépendance posée par les auteurs de l'article [1] n'est en général pas vérifiée.
2. Dans les mêmes conditions que les sous-questions 2 et 3 de la section précédente, générez 10 réalisations de la chaîne de Markov et reportez sur un graphe l'évolution de la moyenne des proportions des trois catégories sur ces 10 réalisations. Calculez également la moyenne du temps de disparition des individus infectieux sur ces 10 réalisations. Vérifiez que ces résultats confirment les résultats de la section précédente.
3. Calculez les mêmes courbes pour le graphe W^{big} fourni en utilisant des valeurs $\beta = 0.5$ et $\mu = 0.2$ en supposant qu'initialement 0.5% des individus (au hasard) sont infectés. Faites des moyennes sur un nombre suffisant de réalisations pour que vos estimations soient stables.
4. Supposons que la courbe précédente représente l'évolution de l'épidémie dans le cas où aucune mesure n'est prise et que l'objectif principal est de réduire le nombre total de personnes infectées simultanément (afin d'éviter l'engorgement des hôpitaux), tout en cherchant également à réduire le nombre total de personnes infectées (pour réduire le nombre de décès) et dans une moindre mesure la durée de l'épidémie. En utilisant et, éventuellement en modifiant le modèle, montrez l'impact des quatre mesures suivantes :

- (a) Réduire la probabilité de transmission de la maladie en recommandant aux gens de se tenir à une certaine distance les uns des autres, de se laver les mains, etc.
- (b) Réduire les interactions entre les personnes par des mesures de confinement (fermeture des écoles, des transports en commun, etc.).
- (c) Vacciner un certain pourcentage de la population (en supposant qu'un vaccin est disponible).
- (d) Traiter les patients avec un médicament qui permettrait d'accélérer leur guérison.

Pour chaque mesure, expliquez comment la modéliser et étudiez son effet sur les critères mentionnés ci-dessus. Comparez également les mesures entre elles. Pour cette comparaison, vous pouvez par exemple fixer un nombre total de lits d'hôpital disponibles et implémenter chacune des mesures pour que le nombre total de patient infectés ayant besoin d'une hospitalisation soit inférieur au nombre de lits (en supposant qu'un pourcentage fixe d'individus infectés a besoin d'être hospitalisé). Les mesures peuvent ensuite être comparées sur base des critères secondaires.

3 Resources

L'énoncé du projet et le graphe fourni pour la deuxième partie est disponible sur ce site : <http://www.montefiore.ulg.ac.be/~geurts/math1222.html>. Pour illustrer le fonctionnement du modèle, une simulation de propagation du virus sur une grille rectangulaire est également téléchargeable sur le site (n'hésitez pas à générer votre propre visualisation). Pour toute question relative au projet, contactez Pierre Geurts (p.geurts@uliege.be). S'il y a lieu, une liste des questions les plus fréquemment posées sera également maintenue sur le site ci-dessus. Veuillez à consulter cette liste avant de contacter l'encadrant.

Références

- [1] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 10(4), January 2008.