

Sujets de mémoire 2023–2024

Prometteur : Nicolas Franco

nicolas.franco@unamur.be

1 Analyse de certaines données minimales pour la calibration d'un modèle épidémiologique avec infection asymptomatique

Contexte

Lors de l'apparition de l'épidémie de COVID-19 en Europe en mars-avril 2020, de nombreux modèles mathématiques ont dû tant bien que mal être calibrés. Les données disponibles à l'époque se résumaient à des données de décès et éventuellement d'hospitalisations. Or un virus comme le SARS-Cov-2 possède une transmission de type asymptomatique ou pre-symptomatique, ce qui fait qu'une grande partie des porteurs du virus sont indétectables via les effets directs de la maladie. Il est donc impossible de calibrer un modèle correctement sans obtenir des indéterminations ou des incertitudes extrêmement grandes, et cela peut donc mener à des projections pouvant être complètement erronées. Certaines données, si disponibles ou collectables, peuvent être ajoutées afin d'essayer de lever ces incertitudes, comme par exemple des données sérologiques ou des données de testing. Cette information serait intéressante à connaître dans le cadre d'une éventuelle future pandémie.

Objectifs

L'objectif principal de ce mémoire est de tester si la connaissance de certaines données permet de calibrer de façon suffisante un modèle de maladie infectieuse de type compartimental, avec des compartiments différents pour différentes évolutions de la maladie (asymptomatique, pre-symptomatique, symptomatique, hospitalisation, décès). En particulier, il sera demandé :

- De construire un modèle de ce type permettant de simuler une telle propagation d'un virus ;
- De générer des données simulées, avec bruit, à partir de ce modèle et de certains paramètres choisis ;
- De tester des calibrations de ce modèle basées sur certaines des données simulées, afin de vérifier si les paramètres choisis peuvent être retrouvés avec une incertitude acceptable ;
- D'utiliser une méthode de Monte Carlo par chaîne de Markov (MCMC) type Metropolis ;
- D'étudier les calibrations possibles à partir de sous-ensembles de données parmi les suivantes : hospitalisations journalières, décès journaliers, test sérologique ponctuel, testing ponctuel sur une partie aléatoire de la population, testing ponctuel sur une partie aléatoire de la population non-symptomatique.

Vu la nature du mémoire, un suivi du cours Modélisation mathématique des maladies infectieuses (SMATM224) est important.

Références provisoires

- [1] Franco N, COVID-19 Belgium : Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts, *Epidemics*, Volume 37, 2021, 100490, <https://doi.org/10.1016/j.epidem.2021.100490>.
- [2] Held L, Hens N, O'Neill P, Jacco W. *Handbook of Infectious Disease Data Analysis*. Boca Raton : Chapman and Hall/CRC ; 2020.

2 Étude de la vitesse de convergence de la méthode de nouvelle génération

Contexte

Dans le cadre de l'étude mathématique d'une épidémie, des données de contacts sociaux, permettant de connaître le nombre moyen de contacts journaliers entre des individus de différents âges ou lors de différentes activités, sont fort importantes. Ces données sont obtenues via des enquêtes, éventuellement régulières, effectuées auprès d'un sous-échantillon de la population. Elles permettent de construire une « matrice de nouvelle génération » qui permet de modéliser l'incidence (nouvelles infections sur une période donnée) suivant différentes catégories d'âge. Cette méthode est basée sur le théorème de Perron-Frobenius et la méthode de la puissance itérée, ainsi que sur l'hypothèse qu'une distribution stationnaire du nombre d'infections est rapidement atteinte. Or cette hypothèse dépend de l'analyse spectrale des matrices de nouvelle génération (différence entre la plus grande et la seconde plus grande valeurs propres) qui n'est en pratique pas nécessairement vérifiée.

Objectifs

L'objectif principal de ce mémoire est de vérifier la vitesse de convergence de cette méthode sur les données réelles collectées en Belgique durant l'épidémie de COVID-19. En particulier, il sera demandé :

- De se familiariser avec la méthode de nouvelle génération (next generation matrix), également à large domaine, et de pouvoir comprendre et expliquer son utilisation ;
- De se familiariser et d'analyser les données réelles collectées en Belgique dans le cadre de l'enquête CoMix ;
- De pouvoir utiliser le package R Socrates/Socialmixr permettant d'extraire et formater ces données ;
- D'étudier la vitesse de convergence de la méthode par analyse spectrale à différentes époques de l'épidémie en Belgique, et éventuellement dans d'autres pays ;
- D'effectuer une analyse de sensibilité de la méthode en utilisant une technique de bootstrap (rééchantillonnage).

Vu la nature du mémoire, un suivi du cours Modélisation mathématique des maladies infectieuses (SMATM224) est important. La partie principale de ce mémoire devant être effectuée à l'aide d'un package R, une certaine connaissance du langage R est également importante.

Références provisoires

- [1] Willem L et al, SOCRATES : an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for COVID-19. BMC Research Notes. 2020 ; 13(1) :293. <https://doi.org/10.1186/s13104-020-05136-9>
- [2] Franco N et al., Inferring age specific differences in susceptibility to and infectiousness upon SARS-CoV-2 infection based on Belgian social contact data. PLoS Comput Biol 18(3) : e1009965 (2022). <https://doi.org/10.1371/journal.pcbi.1009965>
- [3] Coletti P et al. CoMix : comparing mixing patterns in the Belgian population during and after lockdown. Scientific Reports. 2020 ; 10(1) :21885. <https://doi.org/10.1038/s41598-020-78540-7>
- [4] Verelst F et al., SOCRATES-CoMix : a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries. BMC Medicine. 2021 ; 19(1) :254. <https://doi.org/10.1186/s12916-021-02133-y>
- [5] <http://www.socialcontactdata.org>

3 Étude de la non-convexité de la calibration d'un modèle épidémiologique

Contexte

Lors de l'apparition de l'épidémie de COVID-19 en Europe en mars-avril 2020, de nombreux modèles mathématiques ont dû tant bien que mal être calibrés. Cependant, en l'absence de connaissances initiales de donc de bons « priors », la calibration initiale des modèles fut extrêmement fastidieuse. Cela est dû à la présence à la fois de nombreux minima locaux situés très loin des minima globaux, mais aussi à la présence potentielle de différentes zones de minima dont il est difficile de savoir si elles sont séparées ou reliées.

Objectifs

L'objectif de ce mémoire est de faire une exploration à l'aide d'un modèle relativement simple pour obtenir une sorte de cartographie des minima locaux et de comprendre davantage les difficultés de calibration. En particulier, il sera demandé :

- De construire un modèle de type SIR(QD), initialement avec un très petit nombre de paramètres puis de progressivement le complexifier ;
- De construire une méthode de calibration de type recherche aléatoire basée sur les données réelles de l'épidémie de COVID-19 en Belgique au début de l'épidémie ;
- D'étudier la structure non-convexe de l'espace des paramètres et son évolution au fur et à mesure de la complexification du modèle ;
- D'élaborer une méthode d'analyse et de présentation des résultats pertinente.

Vu la nature du mémoire, un suivi du cours Modélisation mathématique des maladies infectieuses (SMATM224) est important.

Références provisoires

- [1] Franco N, COVID-19 Belgium : Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts, *Epidemics*, Volume 37, 2021, 100490, <https://doi.org/10.1016/j.epidem.2021.100490>.
- [2] Held L, Hens N, O'Neill P, Jacco W. *Handbook of Infectious Disease Data Analysis*. Boca Raton : Chapman and Hall/CRC ; 2020.

4 Étude de la méthode d'optimisation Direct Search pour la calibration d'un modèle épidémiologique

Contexte

Lors de l'apparition de l'épidémie de COVID-19 en Europe en mars-avril 2020, de nombreux modèles mathématiques ont dû tant bien que mal être calibrés. Cependant, en l'absence de connaissances initiales de donc de bons « priors », la calibration initiale des modèles fut extrêmement fastidieuse. Les techniques utilisées étaient des techniques usuelles de base type recherche aléatoire et méthode de Monte Carlo par chaîne de Markov (MCMC), mais des techniques plus évoluées n'ont pu être envisagées dû au manque de temps disponible pour les étudier et les implémenter. Or de telles méthodes pourraient potentiellement faire gagner énormément de temps de calcul.

Objectifs

L'objectif de ce mémoire est d'étudier l'une de ces méthodes connue sous le nom de Direct Search ou Pattern Search, en essayant en pratique une implémentation de l'algorithme en libre accès. En particulier, il sera demandé :

- De construire un modèle de type SIR(QD) en C++, initialement avec un très petit nombre de paramètres puis de progressivement le complexifier ;
- De construire une méthode de calibration de type recherche aléatoire basée sur les données réelles de l'épidémie de COVID-19 en Belgique au début de l'épidémie ;
- D'étudier la méthode Direct Search afin de pouvoir comprendre et expliquer son fonctionnement ;
- De tester cette méthode de calibration en utilisant le logiciel d'optimisation NOMAD
- D'effectuer des comparaisons de la vitesse de calibration suivant les deux méthodes.

Vu la nature du mémoire, un suivi du cours Modélisation mathématique des maladies infectieuses (SMATM224) est important. Vu que le logiciel NOMAD est implémenté en C++, il est préférable de travailler avec ce langage, qui est cependant assez proche du C, mais permet d'utiliser en plus de la programmation orientée objet.

Références provisoires

- [1] <https://www.gerad.ca/fr/software/nomad/>
- [2] Franco N, COVID-19 Belgium : Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts, *Epidemics*, Volume 37, 2021, 100490, <https://doi.org/10.1016/j.epidem.2021.100490>.
- [3] Held L, Hens N, O'Neill P, Jacco W. *Handbook of Infectious Disease Data Analysis*. Boca Raton : Chapman and Hall/CRC ; 2020.

5 Comparaison d'indicateurs d'incertitude pour un modèle SIR/SEIR

Contexte

Les modèles mathématiques complexes utilisés lors de l'épidémie de COVID-19 présentent généralement une étude de l'incertitude à l'aide de techniques statistiques du type Monte Carlo par chaîne de Markov (MCMC), très adaptées sur des systèmes complexes et difficilement analysables analytiquement. Cependant, les systèmes dynamiques plus simples bénéficient d'outils plus robustes permettant de tester la stabilité, basés par exemple sur la théorie de Lyapunov.

Objectifs

L'objectif principal de ce mémoire est de comparer les deux techniques d'estimation de l'incertitude-stabilité par les méthodes MCMC et la théorie de Lyapunov. En particulier, il sera demandé :

- De se familiariser avec la théorie de Lyapunov ;
- De construire un modèle de type SIR/SEIR relativement simple pour permettre d'appliquer les deux méthodes ;
- De générer des données simulées, avec bruit, à partir de ce modèle et de certains paramètres choisis ;
- D'étudier la stabilité de ce modèle suivant des variations des paramètres en utilisant des techniques de type exposant de Lyapunov ;
- D'utiliser une méthode de Monte Carlo par chaîne de Markov (MCMC) type Metropolis pour calibrer ce modèle à partir de certaines données simulées, avec incertitude ;
- De comparer les résultats donnés par les deux méthodes.

Vu la nature du mémoire, un suivi du cours Modélisation mathématique des maladies infectieuses (SMATM224) est important ainsi que le cours Théorie qualitative des systèmes dynamiques (SMATM104).

Références provisoires

- [1] A.M. Lyapunov, The General Problem of the Stability of Motion, Taylor & Francis, London, (1992).
- [2] Yi N, Zhang Q, Mao K, Yang D, Li Q. Analysis and control of an SEIR epidemic system with nonlinear transmission rate. Math Comput Model. 2009 Nov;50(9) :1498-1513. <https://doi.org/10.1016/j.mcm.2009.07.014>.
- [3] Franco N, COVID-19 Belgium : Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts, Epidemics, Volume 37, 2021, 100490, <https://doi.org/10.1016/j.epidem.2021.100490>.
- [4] Held L, Hens N, O'Neill P, Jacco W. Handbook of Infectious Disease Data Analysis. Boca Raton : Chapman and Hall/CRC ; 2020.