

Modèles épidémiologiques, données multi-sources et identifiabilité

Lieu : INRAE, MaIAGE

Encadrants : Elisabeta Vergu (INRAE) & Bernard Cazelles (Sorbonne Université)

L'émergence et la réémergence de maladies infectieuses, dont la Covid-19 en est la plus récente illustration, ont mis en évidence les nombreux défis auxquels est confrontée la communauté scientifique pour aider à anticiper, évaluer, gérer et contrôler les épidémies. Dans ce contexte, la modélisation mathématique fournit des outils puissants pour la description et la prédiction des dynamiques épidémiques. En particulier, la confrontation des modèles aux données, permet non seulement d'estimer des paramètres clés de ces dynamiques mais également d'affiner les prédictions.

Le travail proposé concerne l'exploration de l'utilisation simultanée de différents types de données pour estimer les paramètres des modèles de dynamiques épidémiques de type SIR, notamment du point de vue de l'identifiabilité de ces modèles. L'identifiabilité est une propriété importante des modèles, consistant en la possibilité d'estimer de manière unique les valeurs des paramètres à partir des observations disponibles. En pratique, cette question, difficile, n'est pas toujours abordée.

Lorsque l'on s'intéresse à la modélisation de la propagation d'un pathogène dans une population, on dispose en général du nombre de personnes infectées en fonction du temps (prévalence = nombre de personnes infectées à un temps donné ou incidence = nombre de nouveaux infectés par unité de temps). Cependant dans de plus en plus de situations, sont disponibles des enquêtes épidémiologiques mesurant la séroprévalence dans une population par rapport au pathogène étudié, c'est-à-dire, le nombre de personnes ayant été infectées au cours de l'épidémie. On se propose ici de quantifier l'information et donc le gain en terme d'identifiabilité qu'apportent les données de séroprévalence, utilisées conjointement avec les données de prévalence ou d'incidence. Pour cela, on commencera par explorer un modèle-jouet, un modèle SIRS par exemple, puis ce modèle pourra être complexifié. Une application à un cas réel est également envisagée. Pour l'instant nous disposons de données concomitantes d'incidence et de séroprévalence, pour l'épidémie de Zika en Polynésie Française, pour l'épidémie de Chikungunya à la Martinique, et pour l'épidémie de Covid-19 en Irlande.

Ce travail nécessitera l'utilisation de techniques d'inférence Bayésiennes (MCMC et/ou SMC)^{1,2}. Différents logiciels sont utilisables en fonction des affinités du candidat : les packages SSM³ ou LibBI⁴ avec une interface R mais aussi pomp⁵ ou STAN⁶ sous R. En ce qui concerne l'identifiabilité, elle sera étudiée par des approches numériques basées sur des profils de vraisemblance^{7,8} complétées par des techniques d'analyse de sensibilité du modèle par rapport aux paramètres^{9,10}.

Références

1. Andrieu, C., Doucet, A., Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72, 269–342.
2. Bretó, C., He, D., Ionides, E.L., King, A.A. (2009). Time Series Analysis via Mechanistic Models. *The Annals of Applied Statistics*, 3, 319-348.
3. Dureau, J., Ballesteros, S., Bogich, T. (2013). SSM: Inference for time series analysis with State Space Models. <https://github.com/JDureau/ssm/blob/master/doc/doc.pdf>.
4. Funk, S., & King, A. A. (2020). Choices and trade-offs in inference with infectious disease models. *Epidemics*, 30, 100383.
5. King, A. A., Nguyen, D., & Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69, 12, 10.18637/jss.v069.i12
6. Chatzilena, A., van Leeuwen, E., Ratmann, O., Baguelin, M., & Demiris, N. (2019). Contemporary statistical inference for infectious disease models using Stan. *Epidemics*, 29, 100367.
7. Evans, N. D., White, L. J., Chapman, M. J., Godfrey, K. R., & Chappell, M. J. (2005). The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical Biosciences*, 194(2), 175-197.
8. Eisenberg, M. C., & Hayashi, M. A. (2014). Determining identifiable parameter combinations using subset profiling. *Mathematical Biosciences*, 256, 116-126.
9. Tuncer, N., & Le, T. T. (2018). Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*, 299, 1-18.
10. Tuncer, N., Gulbudak, H., Cannataro, V. L., & Martcheva, M. (2016). Structural and practical identifiability issues of immunological vector–host models with application to rift valley fever. *Bulletin of Mathematical Biology*, 78(9), 1796-1827.

Compétences recherchées

Master 2 ou stage de fin d'études d'ingénieur en mathématiques appliquées, statistiques appliquées, physique / systèmes complexes ou biologie computationnelle.

Compétences attendues : connaissances en inférence statistique ; programmation bas niveau (C/C++, Python) ; bonne connaissance du langage R ou matlab ; appétence pour la modélisation appliquée à la biologie ; maîtrise de l'anglais scientifique (lecture d'articles).

Déroulement du stage

Le stage se déroulera dans l'unité MaIAGE (Mathématiques et Informatique Appliquées du Génome à l'Environnement) du centre INRAE de Jouy-en-Josas (78350). La durée prévue est de 4 à 6 mois. Début possible en début d'année 2021. Gratification : 550€/mois environ (taux légal).

Candidature

Envoyer CV (inclure la liste des cours suivis en M1/M2 et notes disponibles) et lettre de motivation à cazelles@biologie.ens.fr et elisabeta.vergu@inrae.fr