

Bayesian inference of reproduction number from genomic data

Background

Genomic data are increasingly being used to understand infectious disease epidemiology. Isolates from a given outbreak are sequenced, and the patterns of shared variation are used to infer which isolates within the outbreak are most closely related to each other. Unfortunately, the phylogenetic trees typically used to represent this variation are not directly informative about epidemiological processes, such as the temporal variations in the reproduction number $R(t)$ which represents the average number of secondary infections caused by each infected individual. Accurate estimation of this $R(t)$ function is vitally important to understand the effect of control measures on infectious disease outbreaks.

Objectives

The aim of this project is to develop and apply a stochastic model of infectious disease outbreaks in which the $R(t)$ function has an effect on the genomic relationships between sampled individuals.

Proposed research plan

1. Develop a simple stochastic model of outbreaks, for example using a branching process
2. Implement simple simulations under this model to study the effect of variations in $R(t)$ on genomic relationships
3. Derive a mathematical formulation for the likelihood which may require the use of data augmentation techniques, numerical approximation and/or Monte-Carlo approximation (such as a particle filter)
4. Implement a Bayesian approach to infer the parameters of the model based initially on an observed dated phylogeny, for example using a Markov Chain Monte Carlo
5. Test the methodology using simulated datasets where answers are available and/or to real datasets on outbreaks of a complex disease such as tuberculosis or covid

Starter References

1. Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am J Epidemiol.* 2004;160: 509–516.
2. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol.* 2013;178: 1505–12.
3. Didelot, X., Gardy, J., and Colijn, C. Bayesian inference of infectious disease transmission from whole genome sequence data. *Mol Biol Evol.* 2014;31:1869–1879.
4. Didelot X, Fraser C, Gardy J, Colijn C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Mol Biol Evol.* 2017;34: 997–1007.
5. Ignatieva A, Hein J, Jenkins PA. A characterisation of the reconstructed birth–death process through time rescaling. *Theor Popul Biol.* 2020;134:61-76
6. Ju N, Heng J, Jacob PE. Sequential Monte Carlo algorithms for agent-based models of disease transmission. arXiv:2101.12156 preprint.