Simulation and inference of phylodynamic individual level models

Justin Angevaare[1], Zeny Feng[1], Rob Deardon[2]

[1]University of Guelph
[2]University of Calgary

Epidemics6 - November 30, 2017

Overview

Phylodynamics

Individual level models

Transmission pathway ILMs

Phylodynamic ILMs

Pathogen.jl

Phylodynamics

- The joint or conditional investigation of the dynamics of disease spread and evolution
- Appropriate when epidemiological and evolutionary processes occur at similar time scales
- Requires dense genetic sampling of a pathogen during an epidemic

Phylodynamics software I

- Typically a phylogenetic tree is inferred, and from that a transmission network generated
- MCMC; Metropolis-Hasting with specialized proposal methods for exploring phylogenetic tree space used
- Some challenges persist with low acceptance rates

Phylodynamics software II

- We leverage parameters from Individual Level Models (ILMs) (Deardon et al., 2010)
- Generate phylogenetic tree from transmission network proposals
  - "high acceptance" proposals

ILMs

- Time inhomogeneous Poisson process of disease spread through a heterogeneous population
- Inference typically conducted in Bayesian framework

Exposure in ILMs

- Individuals are in a single disease state during any period t
- Transition of i from susceptible to exposed occurs with rate

$$\lambda_{SE}(i,t) = \left[\Omega_S(i) \sum_{k \in I_{(t)}} \Omega_T(k)\kappa(i,k)\right] + \epsilon(i,t) \text{ for } i \in S_{(t)} \tag{1}$$

where,

- $I_{(t)}$ is the set of infectious individuals during time period t,
- $S_{(t)}$ is the set of susceptible individuals during time period t,
- $\Omega_S(i)$ is an susceptibility function,
- $\Omega_T(k)$ is a transmissability function,
- $\kappa(i,k)$ is an infection kernel, and,
- $\epsilon(i,t)$ is a sparks function.

ILM likelihood

$$L(D|\theta) = \prod_{t=1}^{T-1} \psi(t)\upsilon(t) \exp\left\{-\upsilon(t)\Delta_t\right\}, \tag{2}$$

where,

$$\psi(t) = \begin{cases} \frac{\lambda_{SE}(i,t)}{\upsilon(t)} & \text{if } i \in (S_{(t)} \cap E_{(t+1)}) \\ \frac{\lambda_{EI}(j,t)}{\upsilon(t)} & \text{if } j \in (E_{(t)} \cap I_{(t+1)}) \\ \frac{\lambda_{IR}(k,t)}{\upsilon(t)} & \text{if } k \in (I_{(t)} \cap R_{(t+1)}) \end{cases} \tag{3}$$

$$\upsilon(t) = \sum_{i \in S_t} \lambda_{SE}(i,t) + \sum_{j \in E_t} \lambda_{EI}(j,t) + \sum_{k \in I_t} \lambda_{IR}(k,t). \tag{4}$$

where Δ_t is the length of t^{th} inter-event period.

Transmission pathway ILM extension

- Extension allowing simulation and inference of infection sources
- Necessary for phylodynamic extension

Exposure in transmission pathway ILM

- Separate exposure rates are defined for each susceptible-infectious combination as

$$\lambda^*_{SE}(i,k,t) = \Omega_S(i)\Omega_T(j)\kappa(i,j) \text{ if } i \in S_{(t)}, k \in I_{(t)}, \tag{5}$$

and for exposures from an exogenous source as

$$\lambda^*_{SE}(i,t) = \epsilon(i,t) \text{ if } i \in S_{(t)}. \tag{6}$$

Transmission pathway ILM likelihood

$$L(D|\theta) = \prod_{t=1}^{T-1} \zeta(t)\upsilon(t) \exp\left\{-\upsilon(t)\Delta_t\right\}, \tag{7}$$

where,

$$\zeta(t) = \begin{cases} \frac{\lambda^*_{SE}(i,k,t)}{\upsilon(t)} \text{ if } i \in (S_{(t)} \cap E_{(t+1)}) \text{ by endogenous exposure by } k \\ \frac{\lambda^*_{SE}(i,t)}{\upsilon(t)} \text{ if } i \in (S_{(t)} \cap E_{(t+1)}) \text{ by exogenous exposure} \\ \frac{\lambda_{EI}(j,t)}{\upsilon(t)} \text{ if } j \in (E_{(t)} \cap I_{(t+1)}) \\ \frac{\lambda_{IR}(k,t)}{\upsilon(t)} \text{ if } k \in (I_{(t)} \cap R_{(t+1)}) \end{cases} \tag{8}$$

Phylodynamic ILM extension

- Joint model of disease spread and evolution through a heterogeneous population
- Combines transmission pathway ILM with a phylogenetic tree consistent with transmission network
- Exposure times are assumed to be pathogen divergence dates

Phylodynamic ILM simulation

- Gillespie (1977) stochastic simulation method can be utilized:
- Event time is generated from an exponential distribution based on sum of rates
- Event type is generated from discrete distribution
- Event rates are updated

Phylodynamic ILM inference: likelihood

- Full likelihood is the product of the transmission pathway ILM likelihood (Eq. 7) with corresponding phylogenetic tree likelihood
- Phylogenetic tree likelihood can be calculated using the pruning algorithm of Felsenstein (1973, 1981)

Phylodynamic ILM inference: MCMC I

- Specialized MCMC algorithms necessary for phylodynamic ILM
- Algorithm must efficiently explore full posterior distribution of:
  - Augmented data (transmission network & event timings)
  - ILM and substitution model parameters

Phylodynamic ILM inference: MCMC II

- We implement an MCMC algorithm with several different step types:
  1. Propose ILM or substitution model parameters sample
  2. Propose new set of event times
  3. Propose new set of exposure sources

MCMC III

- Proposals to transmission network are generated with a discrete probability distribution. An individual i will be proposed to have been exposed by k with probability

$$\frac{\lambda^*_{SE}(i,k,t)}{\lambda^*_{SE}(i,t) + \sum_k \lambda^*_{SE}(i,k,t)}, \tag{9}$$

  and by an exogenous source with probability

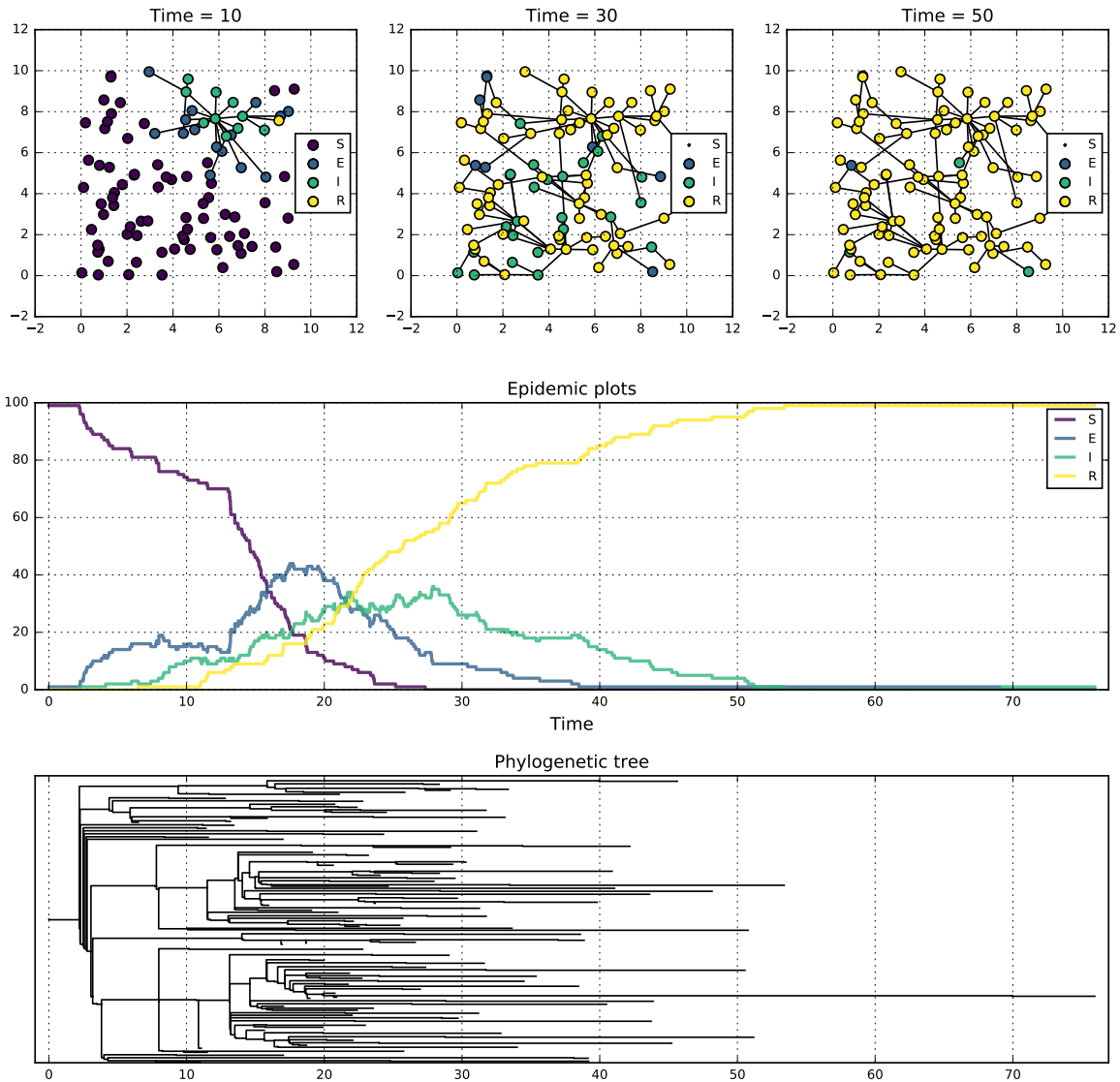
$$\frac{\lambda^*_{SE}(i,t)}{\lambda^*_{SE}(i,t) + \sum_k \lambda^*_{SE}(i,k,t)}. \tag{10}$$

- This is accounted for in Metropolis-Hastings acceptance ratio

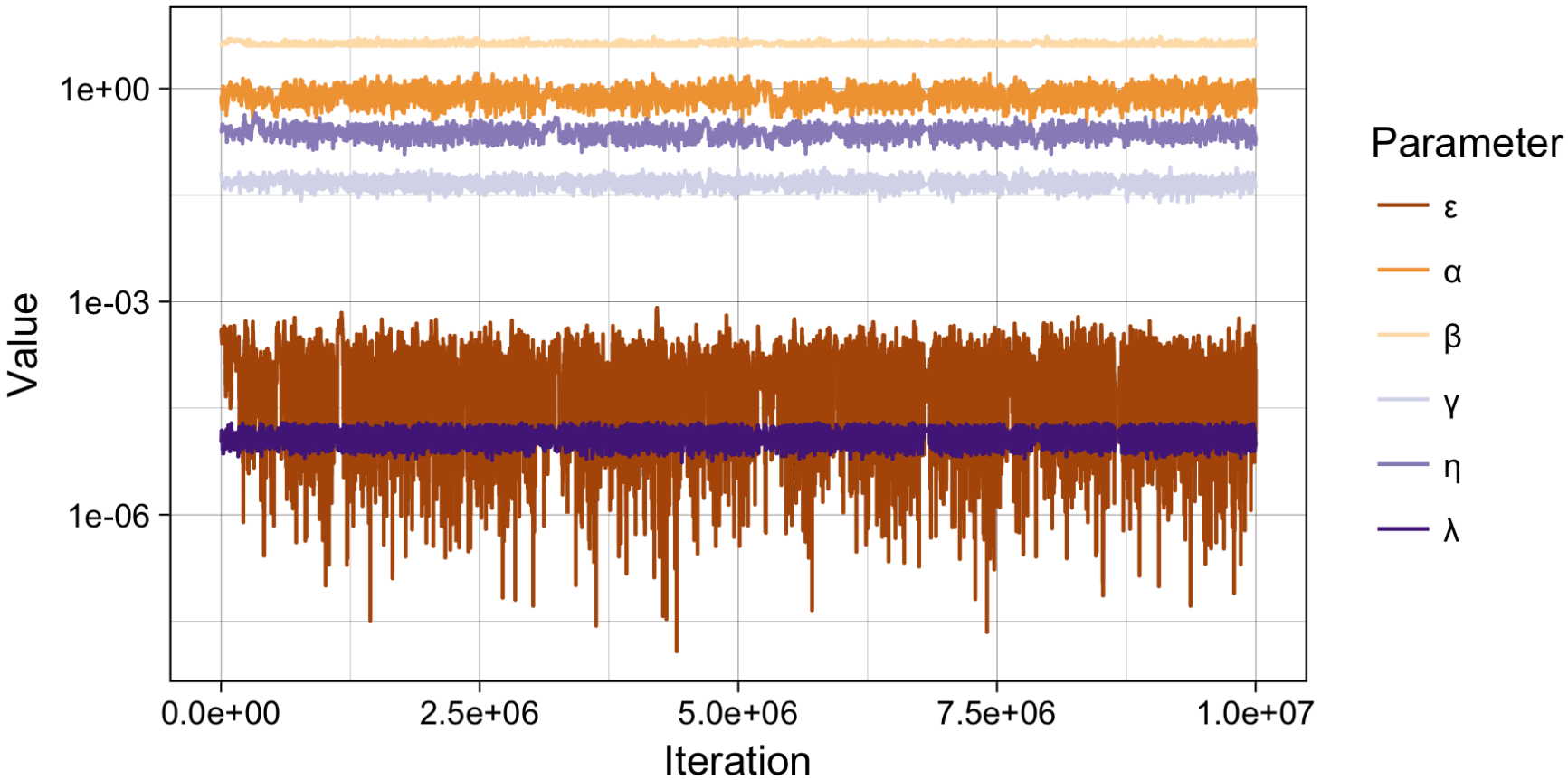
Pathogen.jl

- jangevaare/Pathogen.jl is a flexible full featured package which is in development for Phylodynamic modelling in Julia
- Julia is a fast, high level language designed for scientific computing applications that is approaching it's 1.0 release
- SEIR, SEI, SIR, and SI ILMs with fully customizable $\Omega_S, \Omega_T, \kappa, \epsilon, \Omega_L$, and Ω_R. Many substitution models.
- Initial work has included some simplifying assumptions regarding pathogen diversity.

Phylodynamic simulations in Pathogen.jl



Phylodynamic inference in Pathogen.jl



Initial findings and ongoing work

- Inference of $\lambda^*_{SE}(i, t)$, and exogenous exposures sensitive to external diversity
- Incorporation of more realistic external pathogen source diversity
- Incorporation of within host diversity
- Development of guidelines for MCMC tuning
- Further speed and usability optimizations in Pathogen.jl

References

R. Deardon, S. P. Brooks, B. T. Grenfell, M. J. Keeling, M. J. Tildesley, N. J. Savill, D. J. Shaw, and M. E. Woolhouse. Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica*, 20(1): 239–261, 2010.

J. Felsenstein. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, 25(5):471, 1973.

J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17(6):368–376, 1981.

D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

Thank you!

jangevaa@uoguelph.ca | jangevaare

Appendices

Latency in ILMs

- Transition of j from exposed to infectious occurs with rate

$$\lambda_{EI}(j,t) = \Omega_L(j) \text{ for } j \in E_{(t)} \tag{11}$$

where,

- $E_{(t)}$ is the set of exposed individuals during time period t, and,
- $\Omega_L(j)$ is a latency function.

Removal in ILMs

- Transition of k from infectious to removed occurs with rate

$$\lambda_{IR}(k,t) = \Omega_R(k) \text{ for } k \in I_{(t)} \tag{12}$$

where,

- $\Omega_R(k)$ is a removal function.

Phylogenetic tree likelihood I

$$
\begin{aligned}
L(D|\theta) &= \pi \times S_N, \\
S_N &= [P_{N_\mathrm{L}} \times S_{N_\mathrm{L}}] \odot [P_{N_\mathrm{R}} \times S_{N_\mathrm{R}}] \\
S_{N-1} &= [P_{N-1_\mathrm{L}} \times S_{N-1_\mathrm{L}}] \odot [P_{N-1_\mathrm{R}} \times S_{N-1_\mathrm{R}}], \\
&\dots \\
S_{\frac{N+1}{2}+1} &= \left[P_{\frac{N+1}{2}+1_\mathrm{L}} \times S_{\frac{N+1}{2}+1_\mathrm{L}}\right] \odot \left[P_{\frac{N+1}{2}+1_\mathrm{R}} \times S_{\frac{N+1}{2}+1_\mathrm{R}}\right], \\
P_j &= \exp(Q \times d_j),
\end{aligned}
$$

Phylogenetic tree likelihood II

where,

- π is a vector of nucleotide frequencies according to a nucleotide substitution model, Q
- S_j is a vector of nucleotide likelihoods at node j
- $\odot$ is the component-wise product
- N is the total number of nodes
- $S_1, \ldots, S_{\frac{N+1}{2}}$ are observed
- j_L, j_R, represent the left and right children nodes of node j, and,
- d_j is the length of the branch connecting node j to its parent node.

Felsenstein's pruning algorithm requires that the nucleotide likelihoods at internal nodes are calculated following a postorder traversal of the phylogenetic tree. As each site is assumed to be independent, the full likelihood of any phylogenetic tree is the product of site-specific likelihoods.

Alternative continuous time ILM likelihood I

$$L(D|\theta) = \prod_{t=1}^{T-1} \left[\prod_{i \in (S_{(t)} \cap E_{(t+1)})} P_{SE}(i,t) \prod_{j \in (E_{(t)} \cap I_{(t+1)})} P_{EI}(j,t) \right.$$
$$\prod_{k \in (I_{(t)} \cap R_{(t+1)})} P_{IR}(k,t) \prod_{i \in (S_{(t)} \cap S_{(t+1)})} P_{SS}(i,t)$$
$$\left. \prod_{j \in (E_{(t)} \cap E_{(t+1)})} P_{EE}(j,t) \prod_{k \in (I_{(t)} \cap I_{(t+1)})} P_{II}(k,t) \right], \quad (13)$$

where,

- D are the observed data, which includes observations of infectiousness and removal times for each individual and any risk factor information required by Ω_S, Ω_T, κ, Ω_L, and Ω_R,
- θ are the model parameters and augmented event timings for all observed individuals, and...

Alternative continuous time ILM likelihood II

$$
\begin{aligned}
P_{SE}(i,t) &= \lambda_{SE}(i,t)\exp\left\{-\lambda_{SE}(i,t)\Delta_t\right\}, && (14)\\
P_{EI}(j,t) &= \lambda_{EI}(j,t)\exp\left\{-\lambda_{EI}(j,t)\Delta_t\right\}, && (15)\\
P_{IR}(k,t) &= \lambda_{IR}(k,t)\exp\left\{-\lambda_{IR}(k,t)\Delta_t\right\}, && (16)\\
P_{SS}(i,t) &= \exp\left\{-\lambda_{SE}(i,t)\Delta_t\right\}, && (17)\\
P_{EE}(j,t) &= \exp\left\{-\lambda_{EI}(j,t)\Delta_t\right\}, && (18)\\
P_{II}(k,t) &= \exp\left\{-\lambda_{IR}(k,t)\Delta_t\right\}. && (19)
\end{aligned}
$$