Phylogenetic Inference using RevBayes

The Birth-Death Skyline Model & Temporal Evolution in Viruses

1 Exercise: Estimating Temporal Evolution in the Hepatitis C Virus

1.1 Introduction

A better understanding of the evolutionary processes driving changes in the temporal dynamics of infectious diseases are essential to improving treatment and public health measures. Bayesian inference methods are powerful tools for investigating rates of diversification, transmission, and removal of pathogens in an epidemic.

In this exercise, we will estimate temporally shifting diversification rates in a dataset of contemporaneously sampled Hepatitis C (HCV) sequences from an epidemic in Egypt (sampled in 1999). The analyses are intended to replicate those performed in the study by Stadler et al. (2013).

1.2 Getting Started

This tutorial provides the **`Rev`** scripts and data files to estimate dynamic diversification patterns in a Hepatitis C epidemic (Stadler et al., 2013).

- Download data and output files from: http://bit.ly/1tZ9Xeu

1.3 Launch RevBayes

Execute the RevBayes binary. If this program is in your path, then you can simply type in your Unix terminal:

- **`$ rb`**

When you execute the program, you will see the program information, including the current version number and functions that will provide information about the program — **`contributors()`** and **`license()`**.

1.4 Specifying the Birth-Death Skyline Model

Stadler (2011) described the *birth-death shift model* for inferring temporal changes in rates of speciation and extinction. This work was extended by Stadler et al. (2013) to estimate changing rates of transmission and other diversification patterns in rapidly evolving infectious diseases, where they described a 'piecewise' model of speciation and extinction termed the *birth-death skyline model*.

Figure 1 depicts the conditional dependence structure of the various parameters in the model using a probabilistic graphical model. This model describes the distribution of trees and branching times (Ψ). The process is conditioned on the stochastic node representing the origin time T. This parameter is the time the first lineage originated, or the introduction of the virus. For any given time-tree the T is greater than the first transmission event, i.e., the root age of the tree. In their paper, Stadler et al. (2013), assumed a lognormal prior on the origin of the epidemic. Under this model, a fixed number m of parameter intervals

are assumed, leading to $m-1$ rate-change events over the time of the process. Thus, for any state of T the vector of deterministic nodes $\mathcal{M}$ has values such that the $m-1$ event times are evenly spaced between the origin and present day.

The rate parameters are assumed at each of the $m-1$ intervals. Therefore, the process is conditioned on a vector of m transmission rates (λ) and m rates of becoming uninfectious (δ). Often, epidemiologists have more information on the effective reproductive number of a pathogen R, making it more intuitive to devise priors for this parameter instead of the transmission rate. Since λ is a function of R and δ, thus for any interval i in the time series, λ_i is a deterministic node given by the function: $\lambda_i = R_i \delta_i$. The stochastic nodes, R_i and δ_i are then given lognormal prior densities for each interval i. A third rate parameter governs this process and is the rate of sampling through time ψ. This parameter, however, is given the constant value $\psi = 0$ and not depicted in this model because our focus is on situations when all sequences are sampled at the same time. Since the tips are contemporaneous, we must then have a single parameter ρ that represents the probability of sampling lineages at time 0. Here, we will fix ρ to a very small value: $1e-6$.

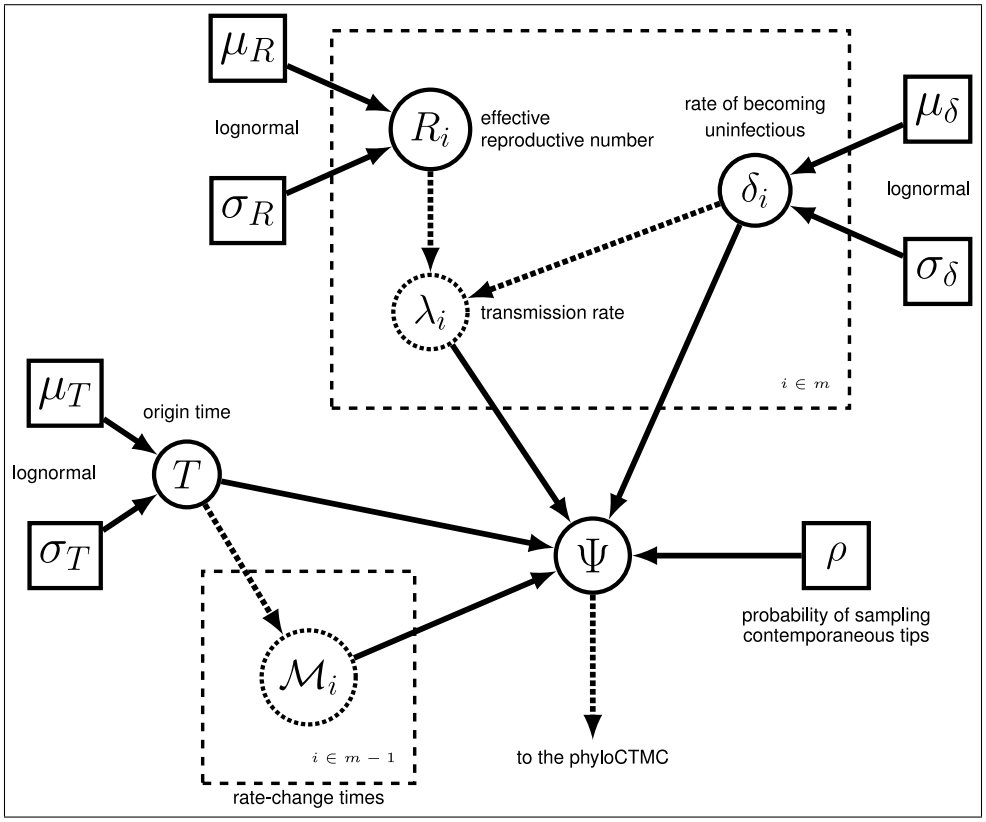

Figure 1: The graphical model of the *birth-death skyline* model for contemporaneously smapled sequences.

Load the Data and Starting Tree

```
RevBayes > D <- readCharacterData(file="data/HCV_data.nex")
```

```
RevBayes > T <- readTrees("data/HCV_data_start.tre")[1]
```

```
RevBayes > mi <- 0
```

1.4.1 The Origin Time

```
RevBayes > origin ~ dnUnif(200.0, 400.0)
```

```
RevBayes > origin.setValue(325.0)
```

```
moves[mi++] <- mvOriginTimeSlide(origin, timetree, delta=3.75, tune=true,
   weight=3.0)
```

1.4.2 Rate-Change Events

```
RevBayes > n_intervals <- 10
```

```
RevBayes > change_times <- n_intervals - 1
```

```
for(i in 1:change_times){
        relative_event_times[i] <- (i * (1.0 / n_intervals))
}
```

```
absolute_times := origin * relative_event_times
```

1.4.3 Sampling Parameters

```
s[1] <- 0.0
```

```
rho[1] <- 1e-6
```

1.4.4 The Rate of Becoming Uninfectious

```
for(i in 1:n_intervals){
        delta[i] ~ dnLnorm(1.0, 1.25)
        delta[i].setValue(1.0)
}
```

1.4.5 The Effective Reproductive Number

```
for(i in 1:n_intervals){
        Re[i] ~ dnLnorm(1.0, 1.25)
        Re[i].setValue(2.0)
}
```

1.4.6 The Birth Rate

```
lambda := Re * delta
```

1.4.7 The Time Tree under the Birth-Death Skylind Model

```
timetree ~ dnSkySerialBDP(origin=origin, lambda=lambda, lambdaTimes=
   absolute_times, mu=delta, muTimes=absolute_times, psi=s, rho=rho,
   timeSinceLastSample=0.0, condition="nTaxa", names=names)
```

```
timetree.setValue(T)
timetree.lnProb
```

```
th := treeHeight(timetree)
```

```
mean_d := mean(delta)
mean_r := mean(Re)
mean_l := mean(lambda)
```

1.4.8 Moves on the Time Tree and Birth-Death Model Parameters

Moves on the Rate Parameters

```
for(i in 1:n_intervals){
   moves[mi++] <- mvScale(delta[i], lambda=1.0, tune=true, weight=1.0)
   moves[mi++] <- mvScale(Re[i], lambda=0.1, tune=true, weight=1.0)
}
moves[mi++] <- mvVectorScale(delta,lambda=1.0,tune=true,weight=2.0*
   n_intervals)
moves[mi++] <- mvVectorScale(Re,lambda=1.0,tune=true,weight=2.0*n_intervals)
moves[mi++] <- mvVectorSingleElementScale(Re,lambda=30.0,tune=true,weight
   =2.0)
moves[mi++] <- mvVectorSingleElementScale(delta,lambda=30.0,tune=true,weight
   =2.0)
```

Moves on the Time Tree

```
### moves on node ages ###
moves[mi++] <- mvNodeTimeSlideUniform(timetree, weight=30.0)
moves[mi++] <- mvRootTimeSlide(timetree, delta=3.75, tune=true, weight=5.0)

### moves on tree topology ###
moves[mi++] <- mvNNI(timetree, weight=8.0)
moves[mi++] <- mvNarrow(timetree, weight=8.0)
moves[mi++] <- mvFNPR(timetree, weight=8.0)
moves[mi++] <- mvSubtreeScale(timetree, weight=5.0)
```

1.4.9 The Clock Rate

```
clock_rate <- 0.79e-3
```

1.4.10 The GTR+G Model

```
sf ~ dnDirichlet( v(1,1,1,1) )
er ~ dnDirichlet( v(1,1,1,1,1,1) )
Q := gtr(er,sf)
```

```
shape ~ dnExponential( 0.1 )
gamma_rates := discretizeGamma( shape, shape, 4 )
```

```
### moves on GTR+G parameters ###
moves[mi++] <- mvSimplexElementScale(er, alpha=10.0, tune=true, weight=3.0)
moves[mi++] <- mvSimplexElementScale(sf, alpha=10.0, tune=true, weight=3.0)
moves[mi++] <- mvScale(shape, lambda=0.75, tune=true, weight=3.0)
```

1.4.11 The Phylogenetic CTMC

```
phySeq ~ dnPhyloCTMC(tree=timetree, Q=Q, siteRates=gamma_rates, branchRates=
   clock_rate, nSites=n_sites, type="DNA")
phySeq.clamp(D)
```

```
### workspace model wrapper ###
mymodel <- model(er)
```

1.4.12 MCMC Analysis

```
monitors[1] <- filemonitor(filename="output/hcv_skybdp_mcmc.log", posterior=
   true, prior=true, likelihood=true, printgen=10, origin, th, lambda, er,
   sf, shape, absolute_times, Re, delta, mean_l, mean_r, mean_d)
monitors[2] <- filemonitor(filename="output/hcv_skybdp_mcmc.trees",printgen
   =10, timetree)
monitors[3] <- screenmonitor(printgen=10, origin, mean_l, th, mean_r, mean_d
   )
```

```
### workspace mcmc ###
mymcmc <- mcmc(mymodel, monitors, moves)

### pre-burnin to tune the proposals ###
```

```
mymcmc.burnin(generations=2000,tuningInterval=150)

### run the MCMC ###
mymcmc.run(generations=40000)

### display proposal acceptance rates and tuning ###
mymcmc.operatorSummary()
```

```
tt <- readTreeTrace("output/hcv_skybdp_mcmc.trees", "clock")
tt.summarize()
# write MAP tree to file
mapTree(tt, "output/hcv_skybdp_MAP.tre")
```

Useful Links

- RevBayes: https://github.com/revbayes/code
- TreePar: http://cran.r-project.org/web/packages/TreePar/index.html
- Tree Thinkers: http://treethinkers.org

Questions about this tutorial can be directed to:

- Tracy Heath (email: tracyh@berkeley.edu)
- Tanja Stadler (email: tanja.stadler@bsse.ethz.ch)
- Sebastian Höhna (email: sebastian.hoehna@gmail.com)

Version dated: August 28, 2014

Relevant References

Gernhard T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology. 253:769–778.

Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences, USA. 108:6187–6192.

Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences. 110:228–233.