

Notes

David Helekal

June 16, 2020

1 Simulation

1.1 Coalescent Preliminaries

The coalescent is a CTMC defined on the set $\{1...n\}$, parametrised via the coalescent rate, in our case $g/Ne(t)$, where g is a scale parameter and $Ne(t)$ the population size at time t . The transition rates of the coalescent process are given by

$$\rho(j, j-1) = \binom{j}{2} \cdot \frac{g}{Ne(t)}$$

The waiting times in the homogenous case are exponentially distributed

$$W_j \sim 1 - \exp\left\{-\frac{g\binom{j}{2}}{Ne(t)}\right\}$$

In the inhomogenous case, the waiting times can be derived as follows: For an inhomogenous CTMC, let $E_j(t)$ be the total exit rate from state j at time t . By the markov property individual exit events from a given state only depend on the state and given time, i.e. they form a time-inhomogenous poisson process. As such the probability of no events in an interval $[t, t+s]$ $s \in \mathbb{R}^+$ is

$$\exp\left\{-\int_t^{t+s} E_j(\tau) d\tau\right\} = \exp\left\{-\int_0^s E_j(t+\tau) d\tau\right\} \quad (1)$$

The waiting times are defined as

$$W_j(t) = \inf\{s : X(t+s) \neq j \mid X(t) = j\} \quad (2)$$

As such

$$W_j(t) > s \Rightarrow \forall \tau \in [t, t+s] X(\tau) = j \quad (3)$$

Furthermore the above relation doesn't hold iff an exit event has occurred in the time interval $[t, t+s]$. As such:

$$\begin{aligned} P[W_j(t) > s] &= P[\text{no exit events in } [t, t+s]] = \exp\left\{-\int_0^s E_j(t+\tau) d\tau\right\} \\ P[W_j(t) < s] &= 1 - \exp\left\{-\int_0^s E_j(t+\tau) d\tau\right\} \end{aligned}$$

In the case of phylodynamic coalescent this becomes

$$W_j(t) \sim 1 - \exp\left\{-\int_0^s \frac{g\binom{j}{2}}{Ne(t+\tau)} d\tau\right\} \quad (4)$$

Note, the waiting times are still memoryless:

$$P[W_j(t) > s + u \mid W_j(t) > s] = P[W_j(t) > s + u \mid X(s) = j] \quad (5)$$

By markov property

$$P[W_j(t) > s + u \mid X(s) = j] = P[W_j(t + s) > u] \quad (6)$$

1.2 Homogenous case

The sampling process conditioned on sampling times follows a slightly modified gillespie scheme.

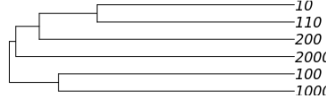


Figure 1: An example simulated coalescent tree

```
f <- (sampling_times, Ne): //Sampling times in descending order
extant_lineages <- 1
future_lineages <- length(sampling_times)-1
t <- sampling_times[1]
idx <- 1

while extant_lineages > 1 or future_lineages > 0:
  if extant_lineages < 2:
    idx++
    t <- sampling_times[idx]
    extant_lineages++
    future_lineages--
  else:
    delta_t <- t-sampling_times[idx+1]
    rate <- binom(extant_lineages,2)/Ne
    w_t ~ exp(rate)

    if w_t < delta_t:
      coalesce_lineages
      extant_lineages--
      t <- t+w_t
    else:
      idx++

```

```

t <- sampling_times[idx]
extant_lineages++
future_lineages—

```

1.3 Inhomogenous Case

In the inhomogenous case, the scheme is similar, with the key difference that the sampling times now follow a much more complex distribution. As such a sampling scheme such as rejection sampling will be required (?)

1.4 Multistrain+Inhomogenous Case

In this case, coalescent nodes have an added colour property, and each colour coalesces according to a colour specific, time dependent case. Nodes of non-identical colour can coalesce iff at least one of them is the last remaining node of a given colour.

Given M colours, M population size functions $\{Ne_j(t)\}_{1 \leq j \leq M}$, and initial population size N , Let $Y(t)$ be a CTMC with the state space:

$$S = \{\mathbf{s} \in \mathbb{Z}_+^N : |\mathbf{s}| \leq N, |\mathbf{s}| \geq 1\} \quad (7)$$

and the transition rates

$$\mathbf{s} \rightarrow \mathbf{s} - \mathbf{e}_j \quad \binom{s_j}{2} Ne_j(t) + \delta_{s_j,1} Ne_j(t) \sum_{i \neq j} s_i \quad (8)$$