

# The Yule model

ALEXEI J. DRUMMOND

22. May, 2007

## 1 Probability density for Yule model

The Yule model of branching is a pure birth process in which each branch has associated with it a birth rate ( $\lambda$ ) determining the instantaneous rate at which the branch gives birth to a new branch (bifurcates into two branches). Starting at the root of a tree (the first bifurcation), there are two descendant branches, each with a birth rate of  $\lambda$ . This gives rise to the following probability density function for the time to the second bifurcation ( $t_2$ ):

$$p(t_2|\lambda) = 2\lambda e^{-2\lambda t_2} \quad (1)$$

This is simply an exponential distribution with a mean of  $\frac{1}{2\lambda}$ . Likewise the probability density for the time from the  $(i-1)$ th to the  $i$ th bifurcation ( $t_i$ ) is:

$$p(t_i|\lambda) = i\lambda e^{-i\lambda t_i} \quad (2)$$

So the probability density for a tree that has *just reached* the  $n$ th bifurcation is:

$$q(\mathbf{t}|\lambda, n) = \prod_{i=2}^{n-1} i\lambda e^{-i\lambda t_i} \quad (3)$$

However most trees are not sampled exactly at the moment of the final bifurcation event. This means there is a final time  $t_n$  over which *none* of the  $n$  branches bifurcated. The probability of waiting  $t_n$  time without seeing any of the  $n$  branches bifurcate is  $e^{-n\lambda t_n}$  giving rise to a total probability of a tree of  $n$  tips:

$$p(\mathbf{t}|\lambda, n) = (n-1)!\lambda^{n-2} \prod_{i=2}^n e^{-i\lambda t_i} \quad (4)$$

Remembering that  $e^a e^b = e^{a+b}$  and defining the total tree length  $s = \sum_{i=2}^n i t_i$  we have:

$$p(\mathbf{t}|\lambda, n) = (n-1)!\lambda^{n-2} e^{-\lambda s} \quad (5)$$

This probability density is the same as that of equation (3) in Nee *et al* (2001), despite Nee making the confusing assertion that his equation (3) is conditional on the root height. From the construction above, it is clear that no such conditioning exists.

## 2 Expectation of the tree height and total tree length under the Yule model

Defining the total tree height,  $t_{MRC A} = \sum_{i=2}^n t_i$ , it is easy to show that the expected tree height is:

$$E(t_{MRC A}) = \sum_{i=2}^n \frac{1}{i\lambda} \quad (6)$$

So for a 4 taxon tree and  $\lambda = 2$  the expected height is:

$$E(t_{MRC A}) = \frac{1}{4} + \frac{1}{6} + \frac{1}{8} = 0.5416666 \quad (7)$$

The total expected tree length ( $s$ ) of a Yule branching process is:

$$E(s) = \sum_{i=2}^n \frac{1}{\lambda} = \frac{n-1}{\lambda} \quad (8)$$

So for a 4 taxon tree and  $\lambda = 2$  the expected height is:

$$E(s) = \frac{3}{2.0} = 1.5 \quad (9)$$