Density for trees based on the Yule process

Graham Jones

2011-11-06

1 Introduction

In order to design a prior for a network including allopolyploidization events, it is useful to be able to relate the probabilities of trees with different numbers of tips. Specifically, the Yule process implies a prior in which all labelled histories on n tips are equally likely, and also provides a density on node heights. Usually in Bayesian phylogenetic analysis it is not necessary to calculate the normalization factors, since they it is the same for all trees on n tips. If n varies, the normalization factors are necessary, and that is what this note is about.

2 The density

Suppose there are n tips, and that the birth rate is λ. The heights of the nodes are $t_1, t_2, \dots, t_{n-1}$ and they are ordered so that $t_1 \geq t_2, \geq \cdots \geq t_{n-1} \geq 0$. The root is at t_1. If an improper uniform prior on $[0, \infty)$ for the origin time t_0 of the tree is assumed, the density for the ordered heights is given in [1] as

$$\Pr(t_1, t_2, \dots, t_{n-1}) = n! \lambda^{n-1} \exp[-\lambda(2t_1 + t_2 + \cdots + t_{n-1})] \tag{1}$$

Each set of ordered heights corresponds to a number of labelled histories. Given the heights, one can work backwards in time, choosing a pair of nodes at each step for the heights $t_{n-1}, t_{n-2}, \dots, t_1$, and each set of choices gives a different labelled history. There are

$$\binom{n}{2}\binom{n-1}{2}\cdots\binom{2}{2} = \frac{n!(n-1)!}{2^{n-1}} \tag{2}$$

choices, and all are equally likely. The density for a particular labelled history is threfore

$$\begin{aligned} & \frac{2^{n-1}}{n!(n-1)!} \times n! \lambda^{n-1} \exp[-\lambda(2t_1 + t_2 + \cdots + t_{n-1})] \\ = \; & \frac{2^{n-1}}{(n-1)!} \lambda^{n-1} \exp[-\lambda(2t_1 + t_2 + \cdots + t_{n-1})] \end{aligned} \tag{3}$$

These values are comparable for different n.

It is easier to picture the space of labelled histories if we transform the heights by $y_i = \exp(-\lambda t_i)$ for $1 \le i \le n-1$. The y_i satisfy $0 \le y_1 \le y_2 \le \cdots \le y_{n-1} \le 1$, a region S_{n-1} inside the unit hypercube of dimension $n-1$. In S_{n-1} the density frrom equation (1) becomes

$$\Pr(y_1, y_2, \ldots, y_{n-1}) = n!\, y_1 \tag{4}$$

Each point in S_{n-1} corresponds to $2^{-(n-1)} n!(n-1)!$ labelled histories. Alternatively one can imagine $2^{-(n-1)} n!(n-1)!$ copies of each with density $(2^{n-1}/(n-1)!)\, y_1$.

2.1 Example for 4 tips

Figure 1 shows the region S_3 in x, y, z space, that is the region $0 \le z \le y \le x \le 1$, which occupies one sixth of the unit cube. The density is $24z$. Each point in S_3 corresponds to 18 labelled histories on 4 tips.

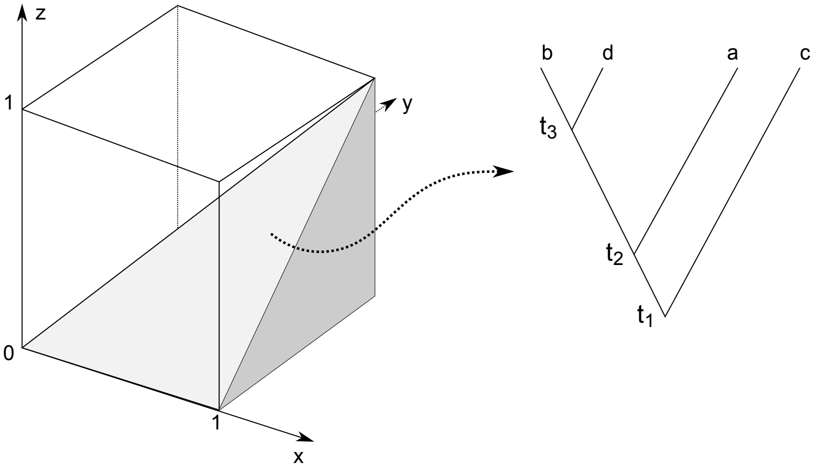

Figure 1: Tree space

In order to take a point (x, y, z) to a labelled history with 4 tips a, b, c, d, first transform to node heights: $t_1 = -\lambda^{-1}\log(z),\ t_2 = -\lambda^{-1}\log(y),\ t_3 = -\lambda^{-1}\log(x)$. Then choose a pair from a, b, c, d, such as $\{b, d\}$, which join at t_3. Then choose a pair from the remaining three lineages $a, c, \{b, d\}$, which join at t_2, and finally the remaining pair of lineages join at the root at t_1.

References

[1] Tanja Gernhard, “The conditioned reconstructed process”, Journal of Theoretical Biology Volume 253, Issue 4, 21 August 2008, Pages 769-778, doi:10.1016/j.jtbi.2008.04.005 (http://dx.doi.org/10.1016/j.jtbi.2008.04.005)