

# When tree generation is correlated with state change

April 11, 2008

## 1 The paper

### 1.1 Probability model

The likelihood of the full tree is constructed by peeling from the tree tips backwards in time to root of the tree. Nevertheless, the probability model is a “forward” model - that is, it assumes that the process begins at some point and continues for some length of time  $T$  and then arrives at the present. This model assumes

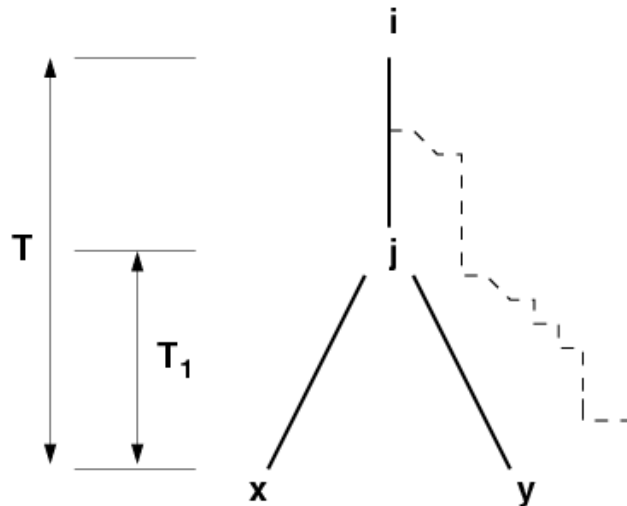


Figure 1: Diagram for evolution of a character that affects speciation rates. The character starts in state  $i$ , and after a fixed time  $T$  has 2 descendants in states  $x$  and  $y$ . All other descendants (dotted line) go extinct before time  $T$  has fully elapsed. The speciation occurs at time  $T_1$  before present.

that all surviving leaf nodes have been sampled.

We note that, while speciation times  $T_1, T_2, \dots$  are random variables, as are the character states at the internal nodes and the tips, the total duration  $T$  of the evolution is a parameter. This may cause some confusion about what it means to calculate the likelihood of an observed tree with “fixed” speciation times (and topology). That is, the total time  $T$  is “fixed” in the sense of being conditioned on, whereas the other speciation times  $T_i$  are “fixed” in a different sense. The times  $T_i$  enter the likelihood calculation with a specific value  $\{T_i = \hat{T}_i\}$  just like the states at the leaf nodes enter the likelihood calculation. The fact that this method does not estimate or integrate out the speciation times does not mean that they are not random in the model.

Thus, the total time is a parameter, but the speciation times are random variables with observed values. We note that the total time  $T$  does not get a  $dt$  because it is not random. Thus the probability (density) for the above figure is  $P_{T,\mu,\lambda}(x,y,T_1)dT_1$ .

## 1.2 Peeling on the tree

We define  $f_{ij}(t, T)$  to be the probability that, after elapsed time  $t$ , a species in state  $i$  has at least one descendant in state  $j$ , and that all other descendants go extinct before time  $T$ . To make this more explicit, we define  $N(t)$  to be the number of descendants at time  $t$ , and we define  $g_{ijk}(t)$  to be the probability that a species in state  $i$  has  $k$  descendants after elapsed time  $t$  and that the first one is on state  $j$ . We define the extinction probability  $\mathcal{E}(t)$  as the probability that after time  $t$  a species has no surviving descendants. Finally,  $P_i(\cdot)$  refers to the probability given that the process that starts in state  $i$ .

Given these definitions, we can derive the following formula for  $f_{ij}(t, T)$ :

$$\begin{aligned}
f_{ij}(t, T) &= P_i(N(t) > 0 \text{ and } \{ \text{a descendent has state } j \} \text{ and } \{ \text{all others go extinct by } T \}) \\
&= \sum_{k=1}^{\infty} P_i(N(t) = k \text{ and } \{ \text{a descendant has state } j \} \text{ and } \{ \text{all others go extinct by } T \}) \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^k P_i(N(t) = k \text{ and } \{ \text{descendant } l \text{ has state } j \} \text{ and } \{ \text{all others go extinct by } T \}) \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^k P_i(N(t) = k \text{ and } \{ \text{descendant } l \text{ has state } j \}) \times P_i(\text{all others go extinct by } T | \text{foregoing}) \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^k g_{ijk}(t) \times \mathcal{E}(T-t)^{k-1} \\
&= \sum_{k=1}^{\infty} g_{ijk}(t) \times k \times \mathcal{E}(T-t)^{k-1}.
\end{aligned}$$

## 1.3 Likelihood Calculation

We therefore consider a node  $n$  with parent  $p$  and left and right children  $l$  and  $r$ . If the node  $n$  is not a leaf node, then likelihood calculation is the following:

$$D_{n,i}(t_p) = P_{i,t_p}(\text{data}_n) = \sum_j f_{ij}(t_p - t_n, t_p) \times \lambda_j dt \times D_{l,j}(t_n) \times D_{r,j}(t_n)$$

where  $t_n$  and  $t_p$  are the times of the node and its parent, and where  $\text{data}_n$  is understood to include the data under node  $n$ , including  $t_n$  and other random speciation times, but conditional on the starting time  $t_p$ . If the node  $n$  is a leaf node, then we would have

$$D_{n,i}(t_p) = f_{ij}(t_p, t_p)$$

where time 0 is understood to be the present, where data is collected.

## 1.4 Likelihood and the exponential

In the specific case where speciation and extinction are independent, then  $f_{ij}(t)$  should separate into a probability for the state and a probability for the tree. This is because  $g_{ijk}(t)$  should separate into the probability  $\{e^{Qt}\}_{ij}$  for the state transition and the probability  $g_k(t)$  that there will be  $k$  surviving descendants. Thus, if independence holds, we have

$$\begin{aligned}
f_{ij}(t, T) &= \sum_{k=1}^{\infty} \{e^{Qt}\}_{ij} g_k(t) \times k \times \mathcal{E}(T-t)^{k-1}. \\
&= \{e^{Qt}\}_{ij} \times \sum_{k=1}^{\infty} g_k(t) \times k \times \mathcal{E}(T-t)^{k-1}.
\end{aligned}$$

We know that the matrix exponential displays the special property that  $e^{Qt_1}e^{Qt_2} = e^{Q(t_1+t_2)}$ . This in some sense follows from conditioning on the state  $X(t_1)$  of a Markov chain at time  $t_1$ , and invoking the Markov property to infer that the chain states before and after  $t_1$  are conditionally independent if  $X(t_1)$  is known. This suggests that the function  $f_{ij}(t, T)$  might display a similar factorization.

In fact it does (details not demonstrated), because it also obeys a similar Markov property. That is, if we take a time interval  $t_1 + t_2$ , then the events that happen before and after  $t_1$  and  $t_2$  are independent conditional on what happens at  $t_1$ . Thus

$$f_{ij}(t_1 + t_2, T) = \sum_k f_{ik}(t_1, T) f_{kj}(t_2, T - t_1).$$

As a result, we also have

$$\begin{aligned} D_{n,i}(t_p + \Delta) &= \sum_j f_{ij}(t_p - t_n + \Delta, t_p) \times \lambda_j dt \times D_{l,j}(t_n) \times D_{r,j}(t_n) \\ &= \sum_k f_{ik}(\Delta, t_p + \Delta) \sum_j f_{kj}(t_p - t_n, t_p) \times \lambda_j dt \times D_{l,j}(t_n) \times D_{r,j}(t_n) \\ &= \sum_k f_{ik}(\Delta, t_p + \Delta) D_{n,k}(t_p) \end{aligned}$$

## 1.5 Differential Calculation

This relationship allows us to peel back in time along a branch, as follows:

$$f_{ij}(dt + t, T) = \sum_k f_{ik}(dt, T) f_{kj}(t, T).$$

Now,  $f_{ik}(dt, T)$  has the following form:

$$\begin{aligned} f_{ik}(dt, T) &= \sum_{k=1}^{\infty} g_{ijk}(dt) \times k \times \mathcal{E}(T - dt)^{k-1} \\ &= \sum_{k=1}^{\infty} g_{ijk}(dt) \times k \times \mathcal{E}(T)^{k-1}. \end{aligned}$$

We therefore consider  $g_{ijk}(dt)$ , which is the probability that a particle of type  $i$  gives rise to  $k$  particles in time  $dt$ , and one of the particles has type  $j$ . For  $k = 1, 2, \dots$  we have<sup>1</sup>

$$\begin{aligned} g_{ij1}(dt) &= (1 - \mu_i dt) \times (1 - \lambda_i dt) \times (1_{i=j} + q_{ij} dt) \\ g_{ij2}(dt) &= (1 - \mu_i dt) \times (\lambda_i dt) \times (1_{i=j} + q_{ij} dt), \end{aligned}$$

where  $q_{ii} = -\sum_{j \neq i} q_{ij}$ . We can ignore  $k > 2$  because this requires that 2 events happen in an interval of time  $dt$ . Multiplying these expressions through and discarding higher powers of  $dt$ , we get

$$\begin{aligned} g_{ij1}(dt) &= 1_{i=j}(1 - dt(\mu_i + \lambda_i)) + q_{ij} dt \\ g_{ij2}(dt) &= 1_{i=j} \lambda_i dt. \end{aligned}$$

Therefore, we have

---

<sup>1</sup>Note that I wrote  $1_{i=j}$  instead of the more standard Kronecker delta  $\delta_{ij}$  to express an identity matrix/identity condition. This seemed clearer here, but is actually useful notation in other areas.

$$\begin{aligned}
f_{ik}(dt, T) &= g_{ij1}(dt) + 2\mathcal{E}(T) \times g_{ij2}(dt) \\
&= [1_{i=j}(1 - dt(\mu_i + \lambda_i)) + q_{ij}dt] + 2\mathcal{E}(T) \times [1_{i=j}\lambda_i dt] \\
&= 1_{i=j} + dt [q_{ij} + 1_{i=j}(2\mathcal{E}(T)\lambda_i - \mu_i - \lambda_i)].
\end{aligned}$$

Or, alternately,

$$\begin{aligned}
f_{ii}(dt, T) &= 1 + dt [2\mathcal{E}(T)\lambda_i - \mu_i - \lambda_i + q_{ii}] \\
&= 1 + dt \left[ 2\mathcal{E}(T)\lambda_i - \mu_i - \lambda_i - \sum_{j \neq i} q_{ij} \right] \\
f_{ij}(dt, T) &= dt \times q_{ij} \text{ (if } j \neq i)
\end{aligned}$$

## 1.6 Differential Calculation for likelihood

Applying the above differential calculation to the likelihood, we have:

$$D_{n,i}(t_p + dt) = \sum_k f_{ik}(dt, t_p) D_{n,k}(t_p).$$

Therefore, we have

$$\begin{aligned}
D_{n,i}(t_p + dt) - D_{n,i}(t_p) &= (f_{ii}(dt, t_p) - 1) D_{n,i}(t_p) + \sum_{k \neq i} f_{ik}(dt, t_p) D_{n,k}(t_p) \\
&= dt \left[ 2\mathcal{E}(T)\lambda_i - \mu_i - \lambda_i - \sum_{k \neq i} q_{ik} \right] D_{n,i}(t_p) + dt \sum_{k \neq i} q_{ik} D_{n,k}(t_p).
\end{aligned}$$

This yields the resulting differential equation:

$$\frac{d}{dt_p} D_{n,i}(t_p) = \left[ 2\mathcal{E}(T)\lambda_i - \mu_i - \lambda_i - \sum_{k \neq i} q_{ik} \right] D_{n,i}(t_p) + \sum_{k \neq i} q_{ik} D_{n,k}(t_p)$$

## 1.7 Forwards differential equation

Would it be possible to set up a forwards differential equation? What I mean by this is to use the relationship

$$f_{ij}(t + dt, T) = \sum_k f_{ik}(t, T) f_{kj}(dt, T - t)$$

instead of

$$f_{ij}(dt + t, T) = \sum_k f_{ik}(dt, T) f_{kj}(t, T),$$

thereby conditioning in what happens right before the end of an interval, instead of what happens right after the start of an interval. (Unlike the Kolmogorov forwards and backwards equations, which are actually taking derivatives of different variables, these are derivatives of the same variable at different points.)

## 2 Conditioning on the sample

How would one take a forwards model like this and turn it into a backwards model?

I presume that this involves running the model for an infinite length of time, and then looking into the past. But it is not clear to me how one would do this when/if the population size is infinite, which I presume it would be, conditional on not being 0, if  $\lambda > \mu$ .