MASCOT: Parameter and state inference under the marginal structured coalescent approximation: Supplementary Material

Nicola F. Müller[1,2,*], David Rasmussen[1,2,3,4], and Tanja Stadler [1,2,*]

[1]ETH Zurich, Department of Biosystems Science and Engineering, 4058 Basel, Switzerland
[2]Swiss Institute of Bioinformatics (SIB), Switzerland
[3]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA
[4]Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA

[1]Corresponding authors

1 Derivation of the derivative of the conditional lineage state probabilities

$$
\begin{aligned}
P_t(L_i = l_i|T) &= \frac{P_t(L_i = l_i, T)}{\sum_{a=1} P_t(L_i = a, T)} = \frac{P_t(L_i = l_i, T)}{P_t(T)} \\
\frac{d}{dt} P_t(L_i = l_i|T) &= \frac{d}{dt} \frac{P_t(L_i = l_i, T)}{P_t(T)} = \frac{P_t(T) \frac{d}{dt} P_t(L_i = l_i, T) - P_t(L_i = l_i, T) \frac{d}{dt} P_t(T)}{P_t(T)^2} \\
&= \frac{\frac{d}{dt} P_t(L_i = l_i, T)}{P_t(T)} - \frac{P_t(L_i = l_i|T) \frac{d}{dt} P_t(T)}{P_t(T)}
\end{aligned}
$$

$\frac{\frac{d}{dt} P_t(L_i=l_i,T)}{P_t(T)}$ can be received by dividing equation 3 in Müller et al. (2017) with $P_t(T)$:

$$
\begin{aligned}
\frac{\frac{d}{dt} P_t(L_i = l_i, T)}{P_t(T)} &= \sum_{a=1}^{m} \Big(\mu_{a l_i} P_t(L_i = a|T) - \mu_{l_i a} P_t(L_i = l_i|T) \Big) \\
&\quad - P_t(L_i = l_i|T) \sum_{a=1}^{m} \frac{\lambda_a}{2} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{k=1 \\ k \neq j,i}}^{n} P_t(L_j = a|T) P_t(L_k = a|T) \\
&\quad - P_t(L_i = l_i|T) \lambda_{l_i} \sum_{\substack{k=1 \\ k \neq i}}^{n} P_t(L_k = l_i|T) \qquad (1)
\end{aligned}
$$

The same way, we can derive $\frac{P_t(L_i=l_i|T) \frac{d}{dt} \sum_{a=1}^{n} P_t(L_i=a,T)}{P_t(T)}$ from equation 3 in Müller et al. (2017):

$$
\begin{aligned}
\frac{P_t(L_i = l_i|T) \frac{d}{dt} \sum_{a=1}^{n} P_t(L_i = a, T)}{P_t(T)} &= \frac{P_t(L_i = l_i|T)}{P_t(T)} \Big[\sum_{a=1}^{m} \sum_{b=1}^{m} \Big(\mu_{ba} P_t(L_i = b, T) - \mu_{ab} P_t(L_i = a, T) \Big) (= 0) \\
&\quad - \sum_{a=1}^{m} P_t(L_i = a, T) \sum_{b=1}^{m} \frac{\lambda_b}{2} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{k=1 \\ k \neq j,i}}^{n} P_t(L_j = b|T) P_t(L_k = b|T) \\
&\quad - \sum_{a=1}^{m} P_t(L_i = a, T) \lambda_a \sum_{\substack{k=1 \\ k \neq i}}^{n} P_t(L_k = a|T) \Big] \\
&= -\frac{P_t(L_i = l_i|T)}{P_t(T)} \Big[P_t(T) \sum_{b=1}^{m} \frac{\lambda_b}{2} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{k=1 \\ k \neq j,i}}^{n} P_t(L_j = b|T) P_t(L_k = b|T) \\
&\quad + P_t(T) \sum_{a=1}^{m} P_t(L_i = a|T) \lambda_a \sum_{\substack{k=1 \\ k \neq i}}^{n} P_t(L_k = a|T) \Big] \\
&= -P_t(L_i = l_i|T) \frac{P_t(T)}{P_t(T)} \sum_{a=1}^{m} \frac{\lambda_a}{2} \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq j}}^{n} P_t(L_j = a|T) P_t(L_k = a|T) \qquad (2)
\end{aligned}
$$

Combining equations 1 and 2 then yields:

$$\begin{aligned}
\frac{\frac{d}{dt}P_t(L_i = l_i, T)}{P_t(T)} - \frac{P_t(L_i = l_i|T)\frac{d}{dt}P_t(T)}{P_t(T)} \quad &= \quad \sum_{a=1}^{m}\Big(\mu_{al_i}P_t(L_i = a|T) - \mu_{l_i a}P_t(L_i = l_i|T)\Big) \\
&\quad -P_t(L_i = l_i|T)\sum_{a=1}^{m}\frac{\lambda_a}{2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{k=1\\k\neq j,i}}^{n}P_t(L_j = a|T)P_t(L_k = a|T) \\
&\quad -P_t(L_i = l_i|T)\lambda_{l_i}\sum_{\substack{k=1\\k\neq i}}^{n}P_t(L_k = l_i|T) \\
&\quad +P_t(L_i = l_i|T)\frac{P_t(T)}{P_t(T)}\sum_{a=1}^{m}\frac{\lambda_a}{2}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}P_t(L_j = a|T)P_t(L_k = a|T) \\
&= \quad \sum_{a=1}^{m}\Big(\mu_{al_i}P_t(L_i = a|T) - \mu_{l_i a}P_t(L_i = l_i|T)\Big) \\
&\quad +P(L_i = l_i|T)\sum_{a=1}^{m}\lambda_a P_t(L_i = a|T)\sum_{\substack{k=1\\k\neq i}}^{n}P_t(L_k = a|T) \\
&\quad -P_t(L_i = l_i|T)\lambda_{l_i}\sum_{\substack{k=1\\k\neq i}}^{n}P_t(L_k = l_i|T)
\end{aligned} \tag{3}$$

2 Derivation of the derivative of $P_t(T)$

The expression for $\frac{d}{dt}P_t(T)$ can be easily derived from $\frac{d}{dt}P_t(L_i = a, T)$, by summing over every possible state a:

$$\frac{d}{dt}P_t(T) = \sum_{a=1}^{m}\frac{d}{dt}P_t(L_i = a, T) = \frac{d}{dt}\sum_{a=1}^{m}P_t(L_i = a, T)$$

This summation was done for deriving equation 2, and thus, we showed that:

$$\frac{d}{dt}P_t(T) = -P_t(T)\sum_{b=1}^{m}\frac{\lambda_b}{2}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}P_t(L_j = b|T)P_t(L_k = b|T)$$

3 Derivation of the second derivative

The second derivative can be easily received by taking the derivative of equation 3 in this supplement with respect to time:

$$
\begin{aligned}
\frac{d^2 P_t(L_i = l_i|T)}{dt^2} &= \frac{d}{dt}\Bigg(\sum_{a=1}^{m}\Big(\mu_{al_i} P_t(L_i = a|T) - \mu_{l_i a} P_t(L_i = l_i|T)\Big)\Bigg) \\
&+\frac{d}{dt}\Bigg(P_t(L_i = l_i|T)\sum_{a=1}^{m}\lambda_a P_t(L_i = a|T)\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T)\Bigg) \\
&-\frac{d}{dt}\Bigg(P_t(L_i = l_i|T)\lambda_{l_i}\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T)\Bigg) \\
&= \sum_{a=1}^{m}\Bigg(\mu_{al_i}\frac{d}{dt}P_t(L_i = a|T) - \mu_{l_i a}\frac{d}{dt}P_t(L_i = l_i|T)\Bigg) \\
&+\frac{d}{dt}P_t(L_i = l_i|T)\Bigg(\sum_{a=1}^{m}\lambda_a P_t(L_i = a|T)\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T)\Bigg) \\
&+P_t(L_i = l_i|T)\Bigg(\sum_{a=1}^{m}\lambda_a \frac{d}{dt}P_t(L_i = a|T)\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T)\Bigg) \\
&+P_t(L_i = l_i|T)\Bigg(\sum_{a=1}^{m}\lambda_a P_t(L_i = a|T)\sum_{\substack{k=1\\k\neq i}}^{n} \frac{d}{dt}P_t(L_k = a|T)\Bigg) \\
&-P_t(L_i = l_i|T)\lambda_{l_i}\frac{d}{dt}\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T) \\
&-\frac{d}{dt}P(L_i = l_i|T)\lambda_{l_i}\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T)
\end{aligned} \tag{4}
$$

4 Approximation of the third derivative

In order to get a fast estimate of the step size, we make two assumptions: We assume that the sum of probability mass in a state over all lineages but lineage i does not change and that the sum of the derivatives of lineage i coalescing in any state does not change, meaning:

$$\begin{aligned} \frac{d}{dt}\sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T) &= 0 \\ \frac{d}{dt}\sum_{a=1}^{m} \lambda_a P_t(L_i = a|T) \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T) &= 0 \end{aligned} \tag{5}$$

This assumption only affects the step size used for integration and allows us to simplify equation 4 to:

$$\begin{aligned} \frac{d^2 P_t(L_i = l_i|T)}{dt^2} \approx\; & \sum_{a=1}^{m} \left(\mu_{al_i} \frac{d}{dt} P_t(L_i = a|T) - \mu_{l_i a} \frac{d}{dt} P_t(L_i = l_i|T) \right) \\ & + \frac{d}{dt} P_t(L_i = l_i|T) \left(\sum_{a=1}^{m} \lambda_a P_t(L_i = a|T) \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T) \right) \\ & - \lambda_{l_i} \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T) \frac{d}{dt} P(L_i = l_i|T) \end{aligned}$$

Again taking the derivative with respect to time then yields:

$$\begin{aligned} \frac{d^3 P(L_i = l_i|T)}{dt^3} \approx\; & \frac{d}{dt} \sum_{a=1}^{m} \left(\mu_{al_i} \frac{d}{dt} P_t(L_i = a|T) - \mu_{l_i a} \frac{d}{dt} P_t(L_i = l_i|T) \right) \\ & + \frac{d}{dt} \left(\frac{d}{dt} P_t(L_i = l_i|T) \left(\sum_{a=1}^{m} \lambda_a P_t(L_i = a|T) \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T) \right) \right) \\ & - \frac{d}{dt} \left(\frac{d}{dt} P(L_i = l_i|T) \lambda_{l_i} \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T) \right) \\ =\; & \sum_{a=1}^{m} \left(\mu_{al_i} \frac{d^2}{dt^2} P_t(L_i = a|T) - \mu_{l_i a} \frac{d^2}{dt^2} P_t(L_i = l_i|T) \right) \\ & + \frac{d^2}{dt^2} P_t(L_i = l_i|T) \left(\sum_{a=1}^{m} \lambda_a P_t(L_i = a|T) \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T) \right) \\ & + \frac{d}{dt} P_t(L_i = l_i|T) \left(\frac{d}{dt} \sum_{a=1}^{m} \lambda_a P_t(L_i = a|T) \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = a|T) \right) \\ & - \frac{d^2}{dt^2} P(L_i = l_i|T) \lambda_{l_i} \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T) \\ & - \frac{d}{dt} P_t(L_i = l_i|T) \lambda_{l_i} \frac{d}{dt} \sum_{\substack{k=1\\k\neq i}}^{n} P_t(L_k = l_i|T) \end{aligned}$$

Using again the assumptions of 5, we can write:

$$
\begin{aligned}
\frac{d^3 P(L_i = l_i|T)}{dt^3} \approx & \sum_{a=1}^{m} \left(\mu_{a l_i} \frac{d^2}{dt^2} P_t(L_i = a|T) - \mu_{l_i a} \frac{d^2}{dt^2} P_t(L_i = l_i|T) \right) \\
& + \frac{d^2}{dt^2} P_t(L_i = l_i|T) \left(\sum_{a=1}^{m} \lambda_a P_t(L_i = a|T) \sum_{\substack{k=1 \\ k \neq i}}^{n} P_t(L_k = a|T) \right) \\
& - \frac{d^2}{dt^2} P(L_i = l_i|T) \lambda_{l_i} \sum_{\substack{k=1 \\ k \neq i}}^{n} P_t(L_k = l_i|T)
\end{aligned}
$$

References

Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017). The structured coalescent and its approximations. Molecular Biology and Evolution, page msx186.