

# The serially-sampled coalescent

ALEXEI J. DRUMMOND

July 12, 2007

## 1 A simple example

Consider the situation in which there are 4 individuals sampled, two in the present (A, B) and two sampled  $\tau$  time units in the past. Going back in time, the probability that there is no coalescent between A and B before time  $\tau$  is:

$$p_{nc} = e^{-\tau/\theta} \quad (1)$$

And consequently the probability of coalescence is:

$$p_c = 1 - p_{nc} \quad (2)$$

This is also the probability that the coalescent tree is one of the following topologies: ((A,B),(C,D)), (((A,B),C),D), (((A,B),D),C).

Now consider the topology ((A,B),(C,D)). Conditional on coalescence of (A,B) before time  $\tau$  it has a probability of  $\frac{1}{3}$ . However if there is no coalescence before time  $\tau$  it has its normal coalescent probability of  $\frac{1}{9}$  (being a symmetrical tree shape). This gives a total probability for this tree shape of:

$$p_{((A,B),(C,D))} = \frac{p_c}{3} + \frac{p_{nc}}{9} \quad (3)$$

Likewise the probability of topologies (((A,B),C),D) and (((A,B),D),C) can be calculated as:

$$p_{(((A,B),C),D)} = \frac{p_c}{3} + \frac{p_{nc}}{18} \quad (4)$$

$$p_{(((A,B),D),C)} = \frac{p_c}{3} + \frac{p_{nc}}{18} \quad (5)$$

The probability of the two remaining symmetrical trees are:

$$p_{((A,C),(B,D))} = \frac{p_{nc}}{9} \quad (6)$$

$$p_{((A,D),(B,C))} = \frac{p_{nc}}{9} \quad (7)$$

The probability of each of the remaining asymmetric trees is:

$$\frac{p_{nc}}{18} \quad (8)$$

Taking  $\tau/\theta = 0.5$  then  $p_{nc} = 0.607$  and  $p_c = 0.393$  giving a probability of ((A,B),(C,D)) of:

$$P((A,B),(C,D)) = 0.199 \quad (9)$$

the probability of  $((A,B),C),D$  is:

$$P(((A,B),C),D) = 0.165 \quad (10)$$

the probability of  $((A,C),(B,D))$  is:

$$P((A,C),(B,D)) = 0.0674 \quad (11)$$

and the probability of  $((C,D),B),A$  is:

$$P(((C,D),B),A) = 0.0337 \quad (12)$$

Work out the rest :-). Check out `examples/testCoalescent.xml` to see these results from an MCMC run.