Fitting notes

Trying to find values for a,b and c the transmission modifiers for each level. Ideally without using too many epidemiological measures and using tree measures instead.

Priors:

All uniformly distributed

A = 0.5-1

B = 0-0.3

C = 0-0.5

Fitted using pyabc, a module undertaking ABCSMC, and using a Euclidean distance between normalised vectors of summary statistics.

Observed and simulated data

MCC of BEAST run of 371 sequences until the 30th September 2014 in Sierra Leone. This is about the exponential growth phase of this part of the epidemic. In the simulation, this translates to 148 days, as we start on the 5th May which we think is roughly when the superspreading funeral was.

This included a phylogeographic reconstruction of within and between district jumps.

Summary statistics

For a: within chiefdom, outside of household

Tried the three sets of summary statistics from Saulnier et al 2017: lineages through time, branch lengths and topology. Each was used separately, and had total number of tips as a statistic included.

LTT: I haven’t tried the LTT summary stats yet, have instead just used the number of lineages through time forced into the same bins as the observed data, 20 bins as suggested in Saulnier. If the simulated tree doesn’t run for long enough, 0 lineages are present in that bin.

Branch lengths: Used the 14 SS based on branch length. If the result for a statistic was none, for example if there were no internal branches, it returns None and is removed from consideration for that run. Therefore even when some statistics can’t be calculated, the others from the vector will be compared between observed and simulated.

Topology: The 8 SS based on tree topology. Has same None conditions as BL set, but it doesn’t look like it ever comes up with a None.

Preliminary results suggest that the branch length set is the only one that is informative for inferring a, as this is the only one that is leading to a contraction of the posterior after successive generations of ABCSMC.

B and C: between chiefdom and between district

The number of jumps at each level, just one statistic for each, compared to the jump counts from the phylogeographic analysis.

Results

LTT:

a mean:

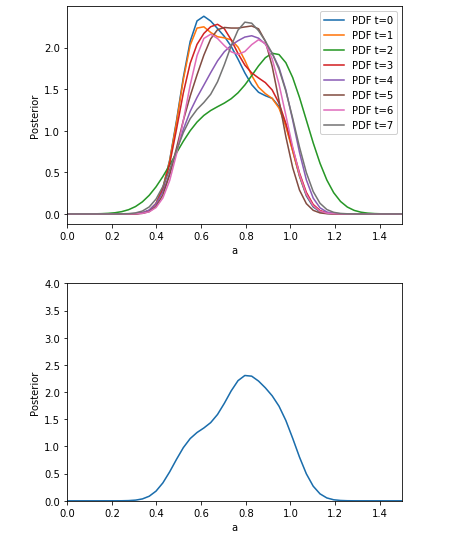
0.664879502770309

b mean:

0.099326358497549

c mean:

0.3111247427368455



Branch lengths:

a mean:

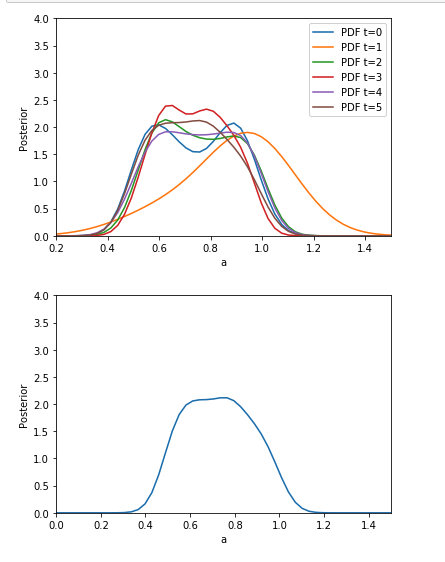
0.6519397146055267

b mean:

0.11007115337899634

c mean:

0.31662586380381375



Topology:

a mean:

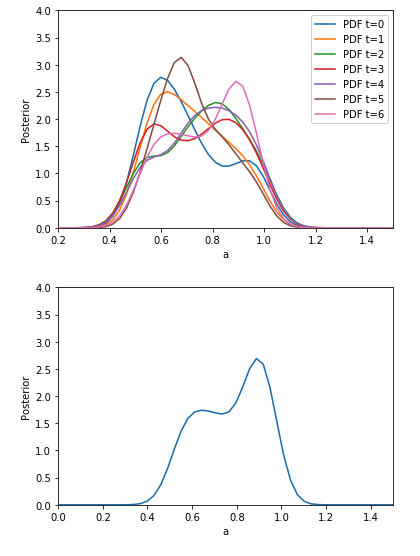
0.6749953760442663

b mean:

0.10902110905354796

c mean:

0.29937993982580796



Going to use means from branch lengths because it has the nicest posterior density for a-value. However, in the future may do a sensitivity analysis for branch lengths vs LTT. The bimodal plot for topology I think means that I won’t use that value. All the values for all parameters are quite similar across the three summary statistic sets which is reassuring.