Phylogenetics

Introduction to Bayesian inference

Rachel Warnock, Laura Mulvey
rachel.warnock@fau.de

May 19, 2022

Today

Discuss the Paul Lewis lecture series

Exercise: tree building using Bayesian inference in RevBayes

Public holdiays

Thursday 26.05 and Thursday 16.06 are public holidays.

Can we reschedule next week's class for Friday 27.05 16:00?

Can we reschedule the class on 16.06 to Tuesday 14.05 16:00?

Part 1 - Trees and likelihood

Note that the answers provided are just *my* interpretation of these concepts and your (correct) answers may vary!

Trees and likelihood

1. *In your own words, how would you describe the terms conditional probability and likelihood?*

Trees and likelihood

1. In your own words, how would you describe the terms conditional probability and likelihood?

$\rightarrow$ conditional probability - the probability of an event, depending on the value of some other event.

Trees and likelihood

1. In your own words, how would you describe the terms <u>*conditional probability*</u> *and* <u>*likelihood*</u>*?*

$\rightarrow$ conditional probability - the probability of an event, depending on the value of some other event.

$\rightarrow$ likelihood - the probability of observing something, given a set of parameter values and assumptions.

Trees and likelihood

2. *Why do we calculate likelihoods on a log scale?*

Trees and likelihood

2. *Why do we calculate likelihoods on a log scale?*

$\rightarrow$ because likelihoods can get incredibly small - see for yourself using R!

Trees and likelihood

3. How does the probability of transitioning from one character state to another (e.g. from A to T) change over time?

use the Transition probability tool to explore this further

Trees and likelihood

3. How does the probability of transitioning from one character state to another (e.g. from A to T) change over time?

use the Transition probability tool to explore this further

$\rightarrow$ the probability of change increases with time.

Trees and likelihood

Felsenstein's pruning algorithm

4a. What do we need this for?

Trees and likelihood

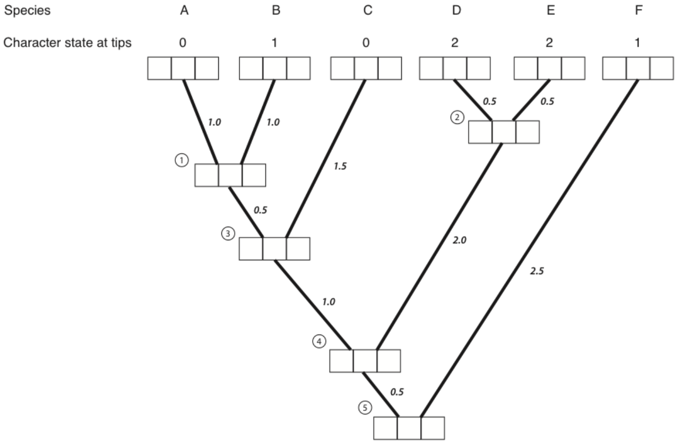
Felsenstein's pruning algorithm

4a. What do we need this for?

$\rightarrow$ to calculate the likelihood of a tree (given an alignment and a substitution model, taking into account all possible ancestral states at every node).

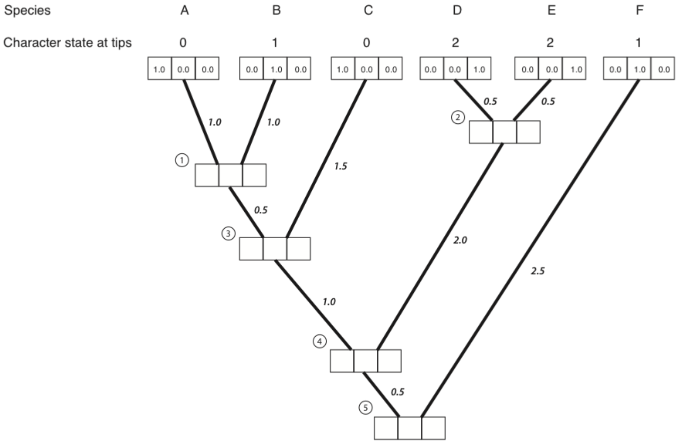
Trees and likelihood

4b. Can you describe the gist of Felsenstein's pruning algorithm?

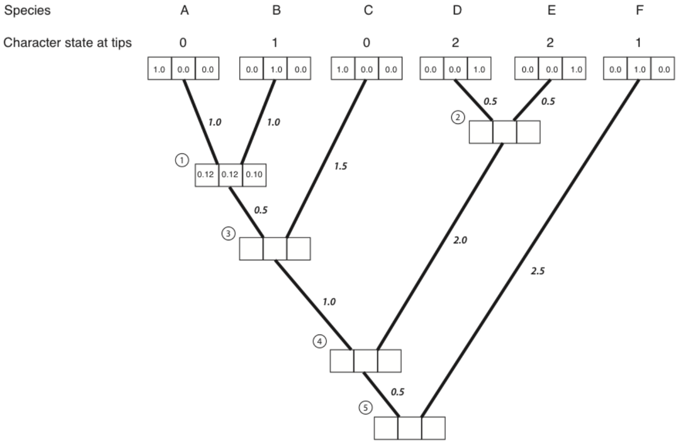


Trees and likelihood

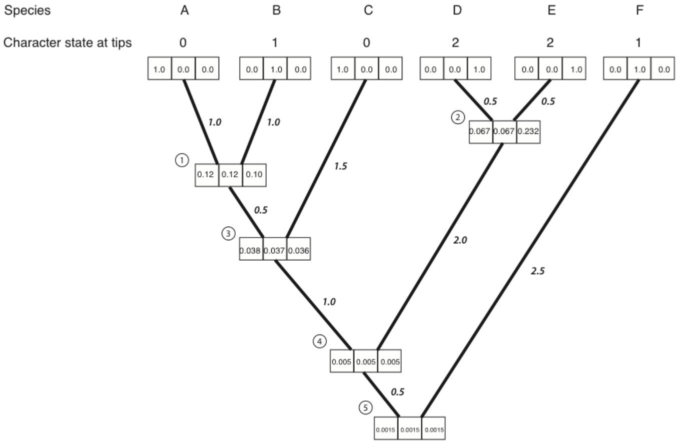
4b. Can you describe the gist of Felsenstein's pruning algorithm?



4b. Can you describe the gist of Felsenstein's pruning algorithm?



4b. Can you describe the gist of Felsenstein's pruning algorithm?



Part 2 - trees, likelihood and rate heterogeneity (substitution models)

Substitution models

1. What are the assumptions of the following substitution models? Consider the rate of change between character states and the state frequencies.

JC $\rightarrow$

HKY $\rightarrow$

GTR $\rightarrow$

Substitution models

1. What are the assumptions of the following substitution models? Consider the rate of change between character states and the state frequencies.

JC $\rightarrow$	equal frequencies, equal rates
HKY $\rightarrow$	unequal frequencies, unequal rates between transversions and transitions
GTR $\rightarrow$	unequal frequencies, unequal rates

Substitution models

Can you briefly describe the following three approaches that account for rate variation among characters?

2a. Site specific rates

Substitution models

Can you briefly describe the following three approaches that account for rate variation among characters?

2a. Site specific rates

$\rightarrow$ assign sites to separate partitions and allow each partition to have its own set of parameters.

Substitution models

Can you briefly describe the following three approaches that account for rate variation among characters?

2a. Site specific rates

$\rightarrow$ assign sites to separate partitions and allow each partition to have its own set of parameters.

2b. Invariant sites model

Substitution models

Can you briefly describe the following three approaches that account for rate variation among characters?

2a. Site specific rates

$\rightarrow$ assign sites to separate partitions and allow each partition to have its own set of parameters.

2b. Invariant sites model

$\rightarrow$ assign a subset of sites to a "constant" (i.e. non-variable) category.

Substitution models

2c. Discrete Gamma model

Substitution models

2c. Discrete Gamma model

$\rightarrow$ calculate the likelihood assuming there are discrete
(e.g. 4) rate categories. Variation in rate categories is represented by a gamma distribution and the distribution parameters are calculated from the data.

Substitution models

3. For a given substitution model, what do the values in the Q matrix and the P matrix represent?

Substitution models

3. For a given substitution model, what do the values in the <u>Q matrix</u> and the <u>P matrix</u> represent?

$\rightarrow$ the Q matrix is the instantaneous rate matrix, i.e. the instantaneous rates of change between 2 character states.

Substitution models

3. For a given substitution model, what do the values in the *<u>Q matrix</u> and the <u>P matrix</u> represent?*

→ the Q matrix is the instantaneous rate matrix, i.e. the instantaneous rates of change between 2 character states.

→ the P matrix is the transition probability matrix, i.e. the probabilities of change between 2 character states after relative time t, which is represented by the branch lengths.

Part 3a - Introduction to Bayesian statistics

1. In your own words can you describe each component of Bayes' rule? Which parts are difficult to understand?

Bayes' theorem

$$P(\text{parameters} \mid \text{data, model}) = \frac{P(\text{data} \mid \text{parameters, model})\, P(\text{parameters} \mid \text{model})}{P(\text{data} \mid \text{model})}$$

Bayes' theorem

$$P(\text{parameters} \mid \text{data, model}) = \frac{P(\text{data} \mid \text{parameters, model})\, P(\text{parameters} \mid \text{model})}{P(\text{data} \mid \text{model})}$$

posterior

likelihood

priors

marginal probability of the data

Bayes' theorem

P (data | parameters, model) $\leftarrow$ the model used to calculate the **likelihood**.

P (parameters | model) $\leftarrow$ this represents our **prior knowledge** of the model parameters.

P (parameters | data, model) $\leftarrow$ the **posterior** reflects our combined knowledge based on the likelihood and the priors.

Bayes' theorem

P (data | parameters, model) $\leftarrow$ the model used to calculate the **likelihood**.

P (parameters | model) $\leftarrow$ this represents our **prior knowledge** of the model parameters.

P (parameters | data, model) $\leftarrow$ the **posterior** reflects our combined knowledge based on the likelihood and the priors.

P (data | model) $\leftarrow$ the probability of the data integrated over all possible parameter values. Can be thought of as a normalising constant (i.e., ensuring the posterior sums to one).

Bayes' theorem

$$\text{Posterior} \propto \text{Likelihood} \times \text{Priors}$$

The posterior probability is proportional to the numerator, i.e. the likelihood times the prior.

2. Can you describe the difference between <u>discrete</u> and <u>continuous</u> variables?

2. Can you describe the difference between discrete and continuous variables?

$\rightarrow$ discrete variables have a set of predefined values (e.g. an integer, or having a tail vs. not).

$\rightarrow$ continuous variables can take on any real number value within a range (e.g. length, body mass).

2. Can you describe the difference between discrete and continuous variables?

→ discrete variables have a set of predefined values (e.g. an integer, or having a tail vs. not).

→ continuous variables can take on any real number value within a range (e.g. length, body mass).

And between probabilities and probability densities?

2. Can you describe the difference between discrete and continuous variables?

$\rightarrow$ discrete variables have a set of predefined values (e.g. an integer, or having a tail vs. not).

$\rightarrow$ continuous variables can take on any real number value within a range (e.g. length, body mass).

And between probabilities and probability densities?

$\rightarrow$ a probability assumes a singular value (e.g. $P = 0.5$).

$\rightarrow$ a probability density $-$ a range of values represented by a distribution.

3. What is the difference between vague vs. informative priors?

3. What is the difference between vague vs. informative priors?

$\rightarrow$ a vague prior is used for parameters where we have little clue what the true value is, e.g. it could be anything between 0 and infinity.

$\rightarrow$ an informative prior is used for parameters where we have some good existing knowledge about what the parameter value could be, e.g. maybe we already know the rate of evolution among a well studied group, so we could use a prior distribution with a mean equal to the known value and add a small variance.

4. *What is the aim of MCMC in Bayesian inference?*

4. What is the aim of MCMC in Bayesian inference?

$\rightarrow$ the aim to approximate the posterior distribution.
The posterior distribution is hard to calculate analytically (i.e. exactly), so we use MCMC to traverse the parameter space and at each step calculate the likelihood *x* the prior, spending time in different regions of the parameter in proportion to the posterior probability - that is, we spend most time in areas with the highest posterior probability.

Part 3b - Introduction to Bayesian statistics part 2

1. *How are steps chosen in an MCMC analysis?*

1. *How are steps chosen in an MCMC analysis?*

$\rightarrow$ these depend on the type of parameter and landscape of the parameters space.

2. Give an example of a parameter you would estimate under each of the following prior distributions and try to state why?

Gamma distribution

Lognormal distribution

Beta distribution

Dirichlet distribution

3. Why do we sometimes need to calculate the marginal likelihood?

3. Why do we sometimes need to calculate the marginal likelihood?

$\rightarrow$ this is required for model testing within a Bayesian framework.

4. What is the difference between a <u>hierarchical</u> model and a <u>non-hierarchical</u> model?

4. What is the difference between a <u>hierarchical</u> model and a <u>non-hierarchical</u> model?

$\rightarrow$ in a hierarchical model different components of the model are nested, e.g. different models can joined together to model different processes that apply to the data.

Exercise 4