Bayesian Phylogenetic Inference

Outline
1. Introduction, history
2. Advantages over ML
3. Bayes’ Rule
4. The Priors
5. Marginal vs Joint estimation
6. MCMC
7. Posteriors vs Bootstrap values

Posterior Probabilities and Bootstrap Values

Priors - noninformative vs informative

- example: 3 urns
  - urn 1 - full of coins, 80% are biased coins
  - urn 2 - full of coins, 10% are biased coins
  - urn 3 - empty and always has been

You are given a coin and asked to come up with a prior probability for it being from urn 1, urn 2, or urn 3

Obviously the prior for urn 3 is zero!

You might assign a P = 50% for urn 1 and urn 2, then flip your coin (gather data) & use Bayes’ Rule to obtain a posterior probability - see how the data change your prior (modify your beliefs in light of new evidence)

Posterior Probabilities and Bootstrap Values

Recent study (Pickett & Randle 2005) which indicates a disturbing but potentially very interesting problem

Clade priors appear to be correlated with their size - smallest and largest clades have the highest prior probability - midsized clades have the lowest

Posterior Probabilities and Bootstrap Values

Example - 5 taxa (A-E) yield 105 bifurcating rooted trees

A clade of 3 taxa (A,B,C) is in only 9 of the 105 trees

A clade of 2 taxa (A,B) is in 15 of the 105 trees

Prior to seeing any data we would expect a clade of (A,B) to be 1.7 times as likely as a clade of (A,B,C)

Although the trees are all equiprobable a priori, all the clades aren’t

Posterior Probabilities and Bootstrap Values

Low BS values for mid-sized clades - an artifact? Saturation?
Posterior Probabilities and Bootstrap Values

Bayesian Inference has a number of advantages over ML:
- more complex & realistic models
- multi-partition/model analysis (meta-models)
- faster (compared to ML bootstrapping)
- arguably superior branch support than BS values

But, of course, these advantages come with some disadvantages:
- PP more likely to accept false hypotheses
- much more sensitive to model violation
- What about Long Branch Attraction?

Posterior Probabilities and Bootstrap Values

Long Branch Attraction - Bayesian vs Maximum Likelihood

Bayesian methods use the same models of evolution as ML (correcting the data for multiple hits & use of branch lengths during tree searching).

Thus, it was thought by many (?) that Bayesian methods would be as successful with Long Branch Attraction problems as was ML.

- Cummings et al. (2003) addressed this question...

Posterior Probabilities and Bootstrap Values

Long Branch Attraction - Bayesian vs Maximum Likelihood

Performance difference probably due to search strategy differences
- ML uses branch & bound and always found the most likely topology
- Bayesian uses MCMC and spread its search efforts across the most likely topology and the other less-likely topologies

Comment: The simulated data were simulated in perfect accordance with the ML model (parameter values fixed) but not (Huelsenbeck & Rannala 2004) in perfect accordance with the Bayesian model (which assumes parameters are not fixed)

Would the Bayesian analysis still do worse than ML in the Felsenstein zone if its assumptions were 100% satisfied?

Did not quantify the performance difference between MP & Bayesian - how intermediate is Bayesian?

Ancestral Character State Reconstruction

aka Character Evolution

Once you have a reliable tree...

You may or may not be able to reliably estimate the evolution of a particular character

Outline
1. Leave the character of interest in the dataset?
2. Different parsimony reconstructions (ACCTRAN / DELTRAN)
3. What about missing or inapplicable data?
4. Examples - frog songs
5. Maximum Likelihood character state reconstruction
6. Software - PAUP*, MacClade, Mesquite, BayesMultiState, BayesTraits
Ancestral Character State Reconstruction

Various types of evolutionary change
- Reconstruction of hypothetical ancestral states
- Morphological, cytogenetic, or developmental processes
- Biogeographic events
- Behavioral or coevolutionary strategies
- Testing for correlated character change

Ancestral Character State Reconstruction

Characters are data to help estimate the phylogeny
Characters can also be "mapped" onto a phylogeny that is obtained from other data
Such mapped characters cannot have influenced the estimate of phylogeny - removes issue of circularity
NOTE: Strength of argument depends on reconstruction:
- If homologous with character in, remove character to strengthen argument
- If homoplasious with character in, leave character in to strengthen argument

Ancestral Character State Reconstruction

If interested in evolution of particular characters
Conduct analysis with and without the characters in the matrix
- This determines if the characters have an influence on the analysis (results differ with and without)
- Be aware of and avoid issue / criticism of circularity

Ancestral Character State Reconstruction

Outline
1. Leave the character of interest in the dataset?
2. Different parsimony reconstructions (ACCTRAN / DELTRAN)
3. What about missing or inapplicable data?
4. Examples - frog songs
5. Maximum Likelihood character state reconstruction
6. Software - PAUP*, MacClade, Mesquite

Ancestral Character State Reconstruction

Character evolution:
Melanism - homology or homoplasy?
Leave character in to strengthen argument of homoplasy

Ancestral Character State Reconstruction

Parsimony Branch Lengths
Parsimony will help find the shortest tree but even on this shortest tree there are often many different ways the character evolution can be mapped on the tree
2 common, but arbitrary, ways to map characters using parsimony:

ACCTRAN - accelerated transformation - when alternative mappings are possible this causes characters to change closer to the root and later reverse (prefers reversals)

DELTRAN - delayed transformation - when alternative mappings are possible this causes characters to change further from the root (prefers parallelisms)
Parsimony Branch Lengths

Character mapping is directly related to the estimated branch length.

Note: ACCTRAN and DELTRAN have nothing to do with finding the shortest tree - they only relate to how the character evolution is mapped on the tree.

Also, these are extremes across a continuum of options.

Unlike ML there is no objective means of choosing which of many character mappings to prefer (since branch lengths are ignored by parsimony).

ML & Bayesian MCMC - branch lengths

Branch lengths are estimated during tree search.

ML & Bayesian MCMC

Parsimonious Character Optimization

OR parallelism
2 separate origins
0 => 1 (DELTRAN)

Homoplastic characters often have alternative equally parsimonious optimizations.

Commonly used varieties are:
- ACCTRAN - accelerated transformation
- DELTRAN - delayed transformation

Consequently, branch lengths are not always fully determined.

PAUP reports minimum and maximum branch lengths.

Ancestral Character State Reconstruction

Outline:
1. Leave the character of interest in the dataset?
2. Different parsimony reconstructions (ACCTRAN / DELTRAN)
3. What about missing or inapplicable data?
4. Examples - frog songs
5. Maximum Likelihood character state reconstruction
6. Software - PAUP*, MacClade, Mesquite
Slight Tangent - Missing Data
Data can be missing for a number of reasons

- specimens / structures not seen (eg females, incomplete fossils)
- gaps in DNA data
- structure is inapplicable to some species (eg wings)
- actually, gaps and inapplicables are the same reason!

Missing & Inapplicable Data

Coding in matrix: (all treated the same in analysis!)

? For missing data that may become “found” someday
eg when females are seen etc.

- For inapplicable data that will never be “found”
eg color of wings on wingless species OR nucleotide site 375 on species
that lacks that site (due to an indel)

N For DNA data that is present but the state is unknown (could be an A, T, G, or C)

Missing & Inapplicable Data


Argued that coding inapplicable data as equivalent missing can lead to undesirable results:

OTUs 1-4 and 11-14 have tails and the others don’t
Clade on the left is unresolved

Gaps in Molecular Alignments

Some regions are unalignable - these are often excluded from analysis (prefer a lack of data to misleading data)

Missing & Inapplicable Data

Missing data are ignored in tree building and can lead to alternative equally parsimonious optimizations in the absence of homoplasy

1 ? ? 0 0

A B C D E

Abundant missing data can lead to multiple equally parsimonious trees or multiple equally parsimonious character mappings on one tree.
This can be a serious problem with morphological data (especially fossil data) but is less of a problem with molecular data unless analyses are of incomplete data

Missing & Inapplicable Data

When OTUs 5-10 are coded as missing for tail color (since they lack tails) Parsimony analysis will treat them as absent and infer this character evolution:

The clade on the left becomes resolved

“In this example the two tailed clades have influenced each other, even though they are widely separated on the tree, because the intervening taxa with missing data allowed the influence to leak through.” p.577” - Maddison
This effect has been used to argue that OTUs that are inapplicable for a certain character should be coded not as equivalent to missing (?) but instead should be given a separate state to ‘block’ the undesirable effects.

So

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTU1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>OTU2</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OTU3</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OTU4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Becomes

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTU1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>OTU2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>OTU3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>OTU4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

What undesirable effect might result from this new coding?

If they are homologous then they should influence each other! The state blue is constructed as plesiomorphic for both clades - if they are not homologous they should be coded as separate characters! eg tail type 1 and tail type 2.

If they are not homologous change matrix from:

tail: present/absent

tail color: red/blue

into:

Tail type 1: present/absent
Tail type 1 color: red/blue
Tail type 2: present/absent
Tail type 2 color: red/blue

### Ancestral Character State Reconstruction

**Female Responses to Ancestral Advertisement Calls in Túngara Frogs**

Michael J. Ryan and A. Stanley Rand

Phylogenetic techniques were used to estimate and reconstruct advertisement calls at ancestral nodes. These calls were used to investigate the degree of preference of female Túngara frogs (Phrynomerus punctatus) for both extant and ancestral calls. Females did not discriminate between calls of males of their own species and calls at their most recent ancestral node. They also recognized calls of some extant species and all four ancestral nodes as the signals of appropriate mates. Both shared ancestral history, and call convergence might differentially influence call preferences.

Not just estimation of ancestral states - literal reconstruction of ancestral states-


**Outline**

1. Leave the character of interest in the dataset?
2. Different parsimony reconstructions (ACCTRAN / DELTRAN)
3. What about missing or inapplicable data?
4. Examples - frog songs
5. Maximum Likelihood character state reconstruction
6. Software - PAUP*, MacClade, Mesquite
Ancestral Character State Reconstruction

Reconstruction of ancestral frog calls

4. Generally, females of species "descendant" to reconstructed ancestor responded positively to (recognized some components of) computer-generated calling song, but basically only of immediate or near-immediate ancestor (not ancestor of other clades)

5. Suggested to researchers that evolution of parameters of male calling song may well have proceeded along pathway with intermediate stages similar to hypothetical reconstructed ancestors-->simple ancestral song diversifying through gradually accumulated innovations

Outline
1. Leave the character of interest in the dataset?
2. Different parsimony reconstructions (ACCTRAN / DELTRAN)
3. What about missing or inapplicable data?
4. Examples - frog songs
5. Maximum Likelihood character state reconstruction
6. Software - PAUP*, MacClade, Mesquite, BayesMultiState, BayesTraits

Terms - from lecture & readings
ancestral character state reconstruction
 ACCTRAN
 DELTRAN
 missing data
 inapplicable data

Study questions
Picket & Randle (2005) noted an interesting and (if true) disturbing point regarding all branch support measures - what was it? They provided no correction for BS & Jackknifing but what correction did they suggest for PP?

Why would one expect Bayesian Inference to be less prone to long branch attraction than Parsimony?

Given Cummings et al. (2003) findings what best explains why Bayesian Inference seems less decisive in the Felsenstein zone than ML?

If one is interested in reconstructing the evolution of a character what is the issue of circularity if one uses the character to help infer the topology? When would you want to leave the character in versus out?

What are two reasons a cell in a data matrix might be empty?

What are two circumstances in which a parsimony reconstruction of ancestral states is likely to be wrong?