A Brief Introduction to the Likelihood Function and Bayesian Inference

Brian R. Moore & Sebastian Höhna
Department of Evolution and Ecology
University of California, Bervis

Bodega Workshop, 2017

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a <u>single</u> site:

- Using Monte Carlo simulation
- Using brute force
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an <u>entire</u> alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

Statistical Estimation of Phylogeny: An Outline

Data

sequence
alignment

Model

tree with
branch lengths

substitution
model

Inference Method

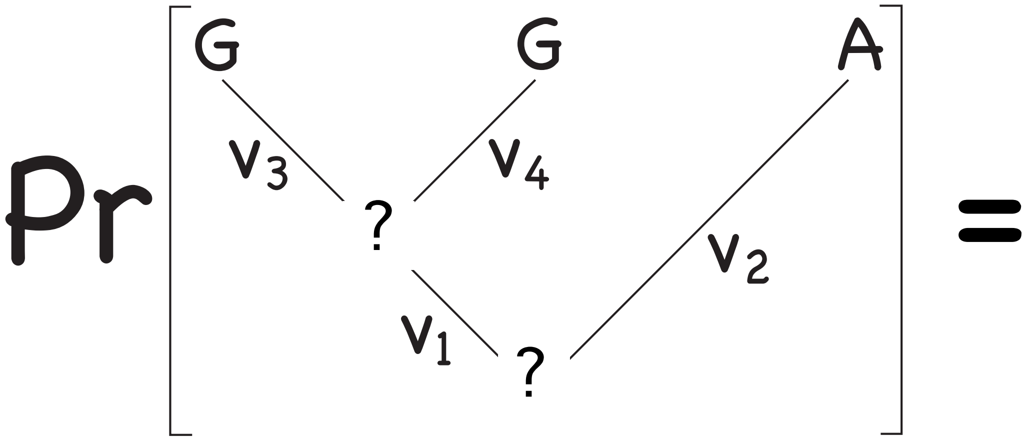
maximum
likelihood

Bayesian
inference

Parameter Estimates

phylogeny &
other model
parameters

An Introduction to the Likelihood Function



$$\pi_i \times p_{ij}(v_1) \times p_{iA}(v_2) \times p_{jG}(v_3) \times p_{jG}(v_4)$$

π_i	Stationary frequencies
$p_{ij}(v)$	Transition probabilities

An Introduction to the Likelihood Function

What is likelihood?

The likelihood is a quantity that is proportional to the probability of observing/realizing the data under a fully specified model/hypothesis.

$$\mathrm{Likelihood}(\mathrm{Parameter}) = \mathrm{Constant} \times \mathrm{Probability}(\mathrm{Data}|\mathrm{Parameter})$$

An Introduction to the Likelihood Function

What is likelihood?

The likelihood is a quantity that is proportional to the probability of observing/realizing the data under a fully specified model/hypothesis.

$$\text{Likelihood}(\text{Parameter}) \propto \text{Probability}(\text{Data}|\text{Parameter})$$

An Introduction to the Likelihood Function

What is likelihood?

The likelihood is a quantity that is proportional to the probability of observing/realizing the data, $\mathbf{X}$, under a fully specified model/hypothesis.

$$\text{Likelihood}(\text{Parameter}) \propto \text{Probability}(\mathbf{X}|\text{Parameter})$$

An Introduction to the Likelihood Function

What is likelihood?

The likelihood is a quantity that is proportional to the probability of observing/realizing the data, $\mathbf{X}$, under a fully specified model/hypothesis, Θ.

$$\text{Likelihood}(\Theta) \propto \text{Probability}(\mathbf{X}|\Theta)$$

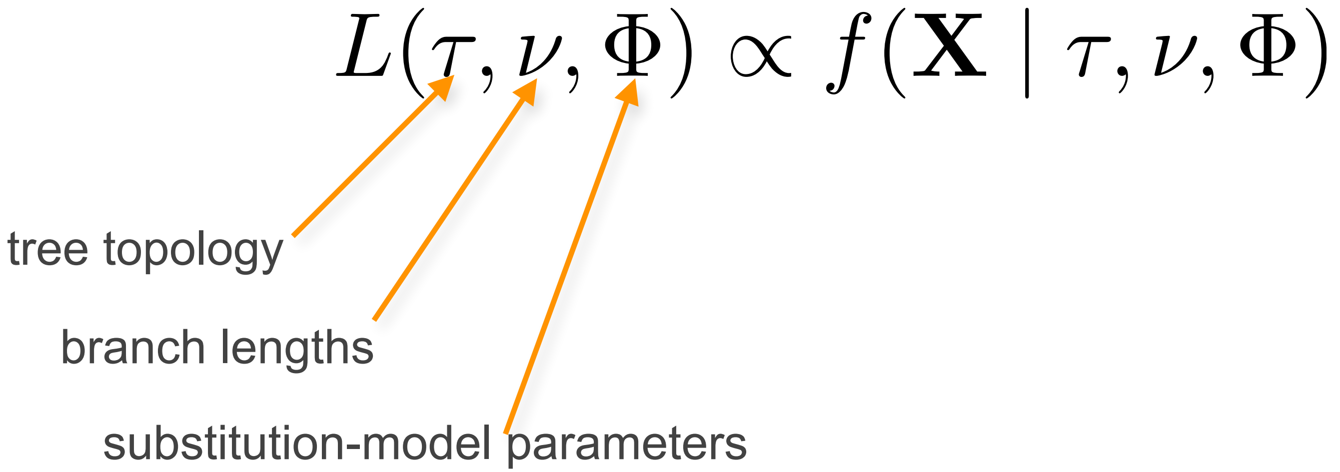
An Introduction to the Likelihood Function

What is likelihood?

The likelihood is a quantity that is proportional to the probability of observing/realizing the data, $\mathbf{X}$, under a fully specified model/hypothesis, Θ.

$$L(\Theta) \propto f(\mathbf{X}|\Theta)$$

For phylogenetic problems, the likelihood is proportional to the probability of observing the sequence alignment, $\mathbf{X}$, under a fully specified phylogenetic model.



$$L(\tau, \nu, \Phi) \propto f(\mathbf{X} \mid \tau, \nu, \Phi)$$

tree topology

branch lengths

substitution-model parameters

The likelihood is a score that measures the fit of the model to the data, providing a basis for comparing different hypotheses/parameter values on the same data

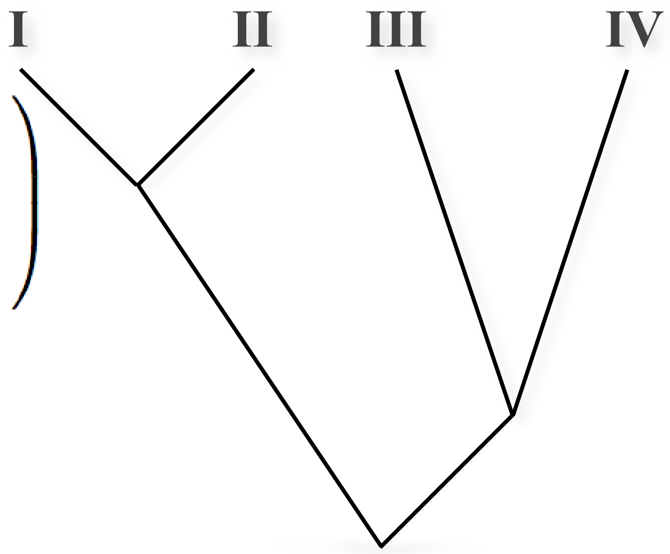
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

Consider a simple simulation experiment to develop our intuition

Imagine we have a fully specified phylogenetic model for four species:

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$

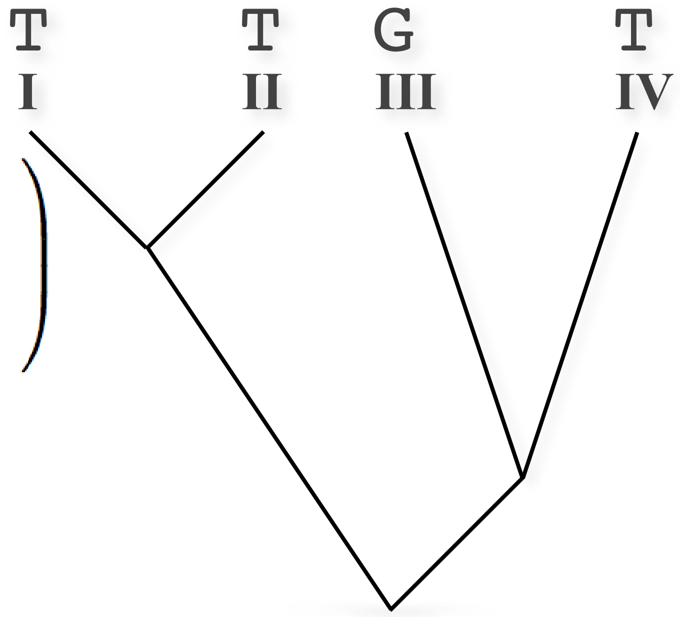


An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?



$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$

The stationary frequencies have been estimated from the rate matrix:

$$\mathbf{P}(100.0) = \begin{pmatrix} 0.138 & 0.188 & 0.495 & 0.179 \\ 0.138 & 0.188 & 0.495 & 0.179 \\ 0.138 & 0.188 & 0.495 & 0.179 \\ 0.138 & 0.188 & 0.495 & 0.179 \end{pmatrix}$$

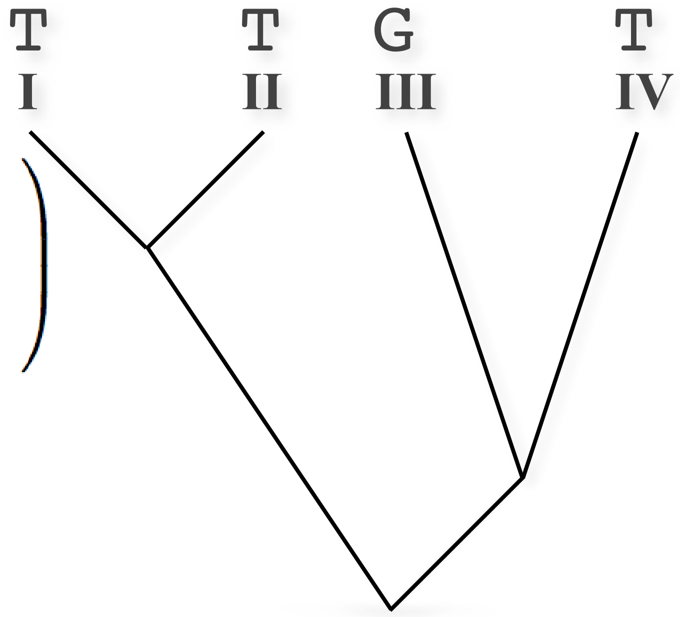
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

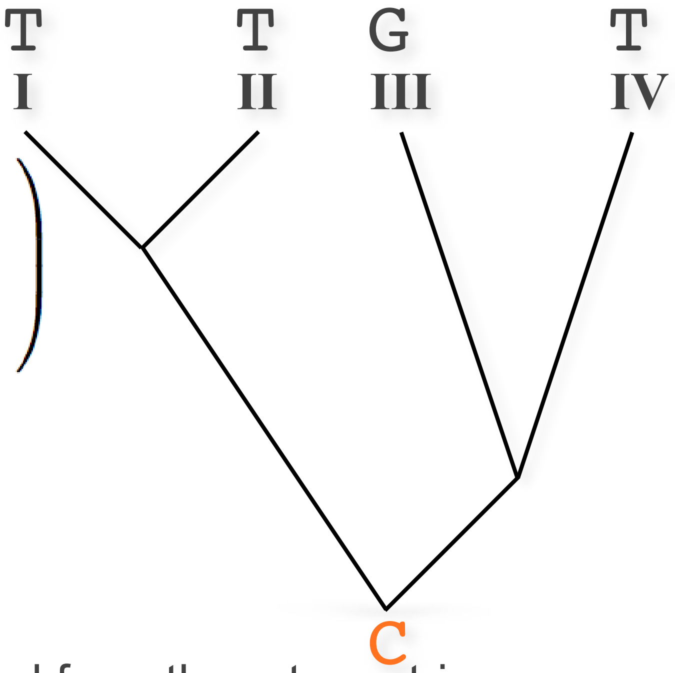
A = 0.138 0-0.138 = A
C = 0.188 0.138-0.326 = C
G = 0.495 0.326-0.821 = G
T = 0.179 0.821-1 = T

An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?



$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$

The stationary frequencies have been estimated from the rate matrix:

A = 0.138 0-0.138 = A
C = 0.188 0.138-0.326 = C
G = 0.495 0.326-0.821 = G
T = 0.179 0.821-1 = T

We randomly selected state `C`

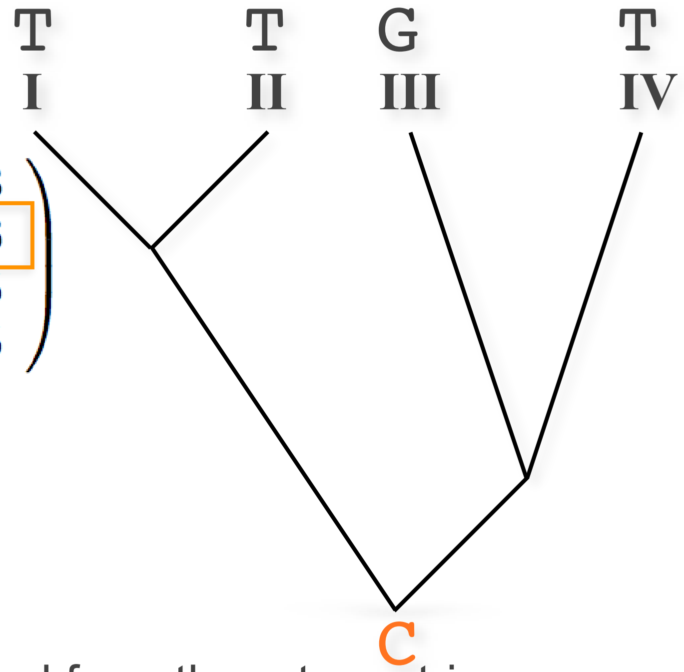
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

`A = 0.138      0-0.138       = A`
`C = 0.188      0.138-0.326 = C`
`G = 0.495      0.326-0.821 = G`
`T = 0.179      0.821-1       = T`

Simulate histories along each branch in a pre-order traversal

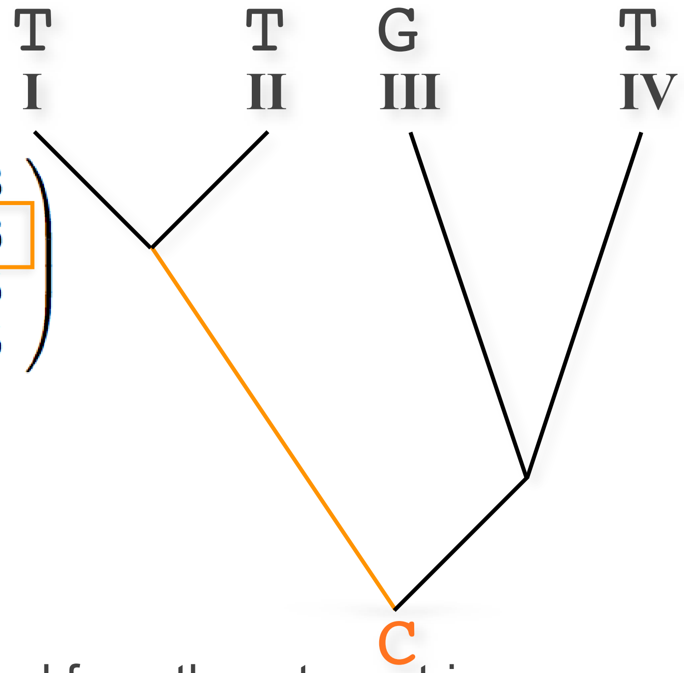
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ \boxed{0.148 & -1.069 & 0.415 & 0.506} \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

```
A = 0.138       0-0.138       = A
C = 0.188       0.138-0.326 = C
G = 0.495       0.326-0.821 = G
T = 0.179       0.821-1       = T
```

Simulate histories along each branch in a pre-order traversal

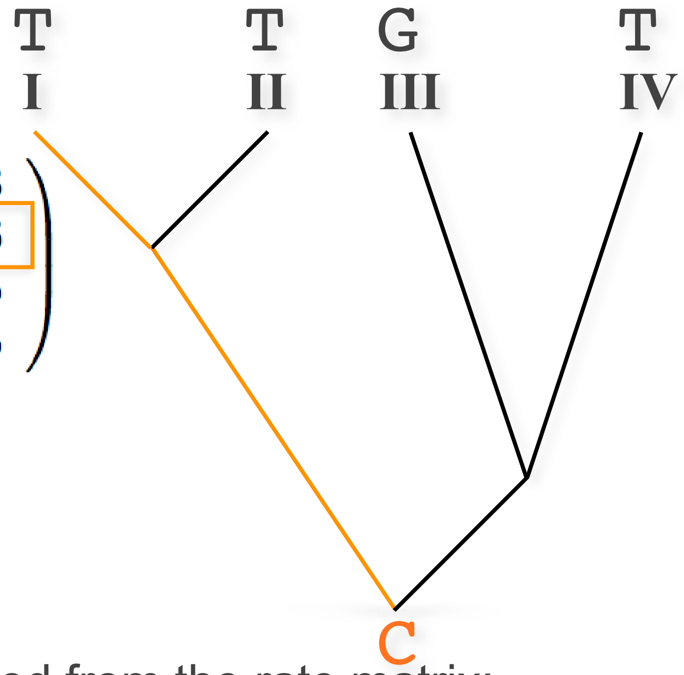
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

A = 0.138　　0-0.138 = A
C = 0.188　　0.138-0.326 = C
G = 0.495　　0.326-0.821 = G
T = 0.179　　0.821-1 = T

Simulate histories along each branch in a pre-order traversal

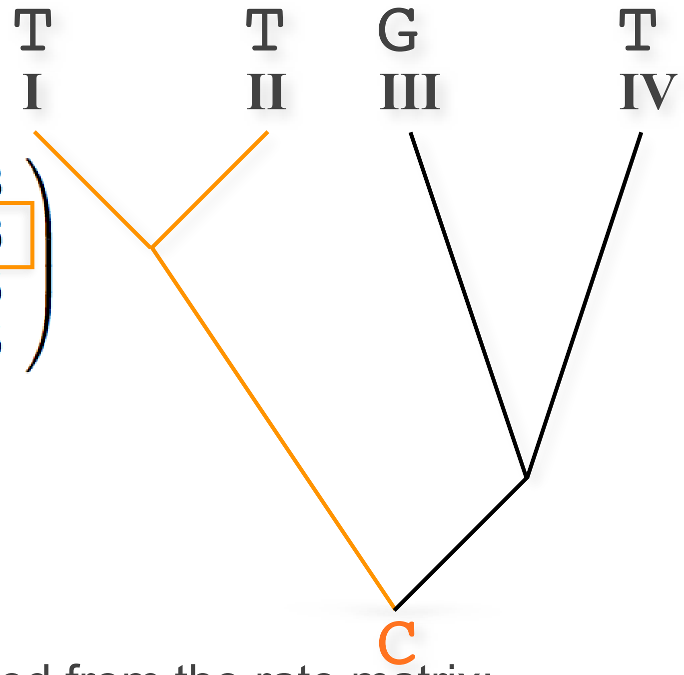
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

A = 0.138   0-0.138   = A
C = 0.188   0.138-0.326 = C
G = 0.495   0.326-0.821 = G
T = 0.179   0.821-1   = T

Simulate histories along each branch in a pre-order traversal

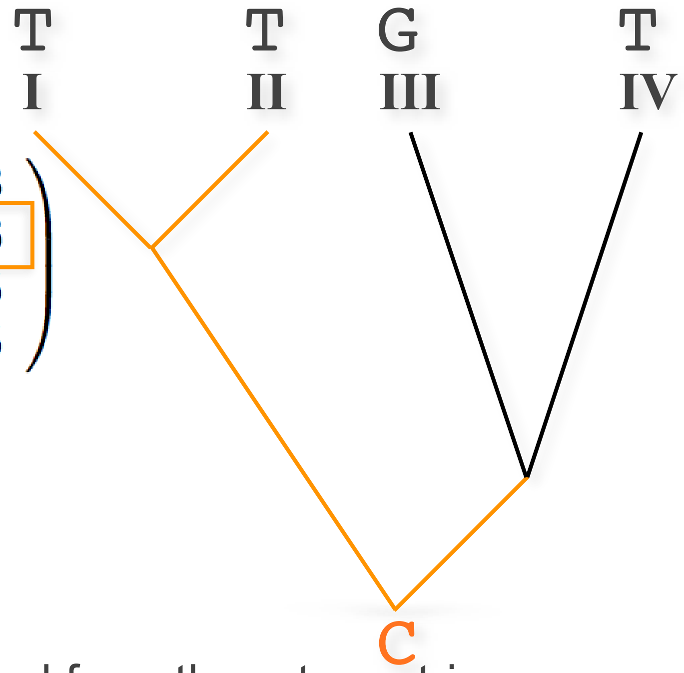
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

A = 0.138   0–0.138 = A
C = 0.188   0.138–0.326 = C
G = 0.495   0.326–0.821 = G
T = 0.179   0.821–1 = T

Simulate histories along each branch in a pre-order traversal

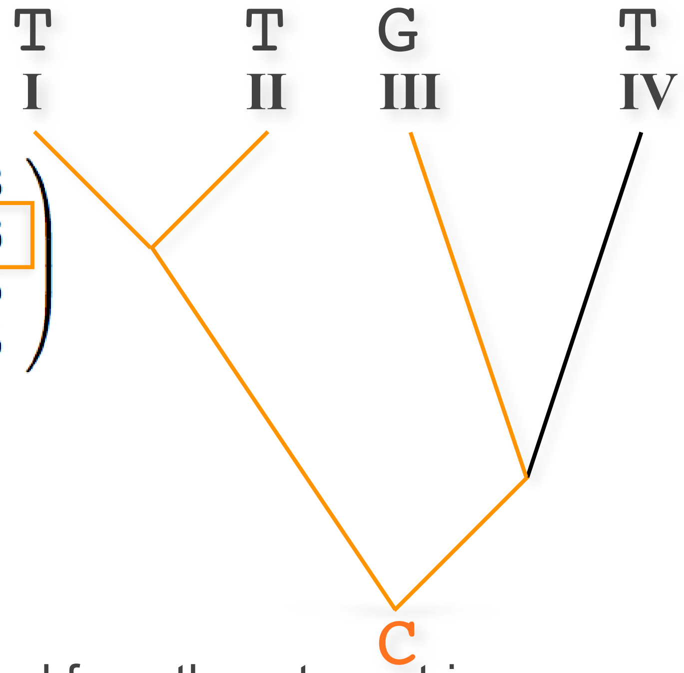
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



The stationary frequencies have been estimated from the rate matrix:

```
A = 0.138      0-0.138       = A
C = 0.188      0.138-0.326   = C
G = 0.495      0.326-0.821   = G
T = 0.179      0.821-1       = T
```

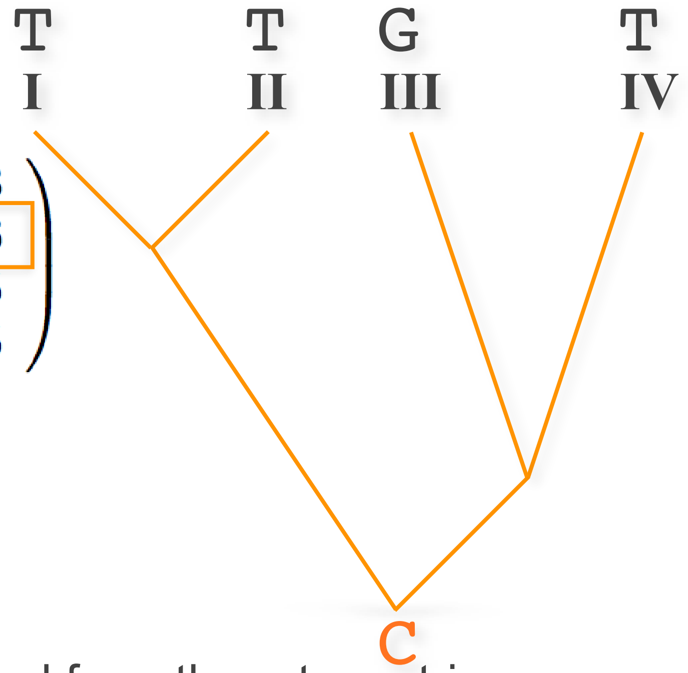
Simulate histories along each branch in a pre-order traversal

An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?



$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$

The stationary frequencies have been estimated from the rate matrix:

`A = 0.138` `0-0.138 = A`
`C = 0.188` `0.138-0.326 = C`
`G = 0.495` `0.326-0.821 = G`
`T = 0.179` `0.821-1 = T`

Simulate histories along each branch in a pre-order traversal

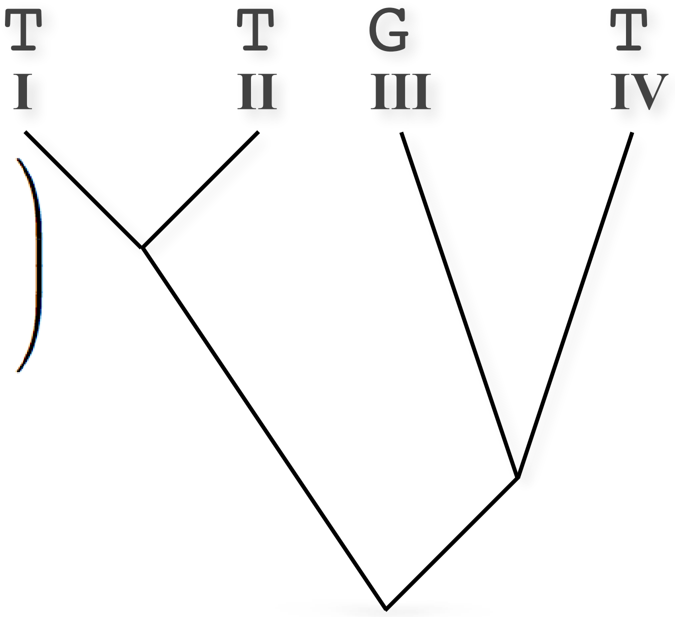
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



We repeat the simulation $100{,}000{,}000$ million times and record the frequency of outcomes that match the observed tip states.

The observed site pattern, `TTGT`, is one of $4^4 = 256$ possible site patterns.

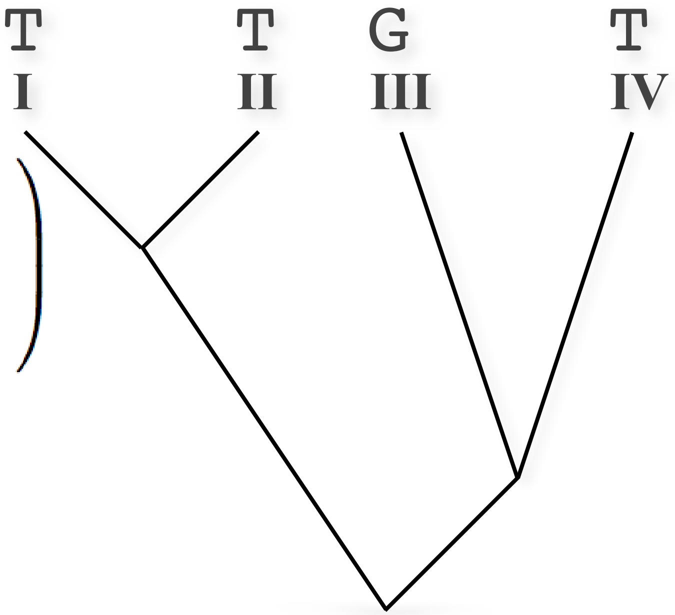
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



Of the 100,000,000 simulations, 850,358 realize the observed data (`TTGT`); the estimated probability of the observed data is therefore 0.00850358.

The probability includes all of the possible histories that can give rise to the data.

	Number of Changes							
Pattern	0	1	2	3	4	5	6	7
TTGT	0.000000	0.881408	0.075358	0.037817	0.004725	0.000622	0.000062	0.000007

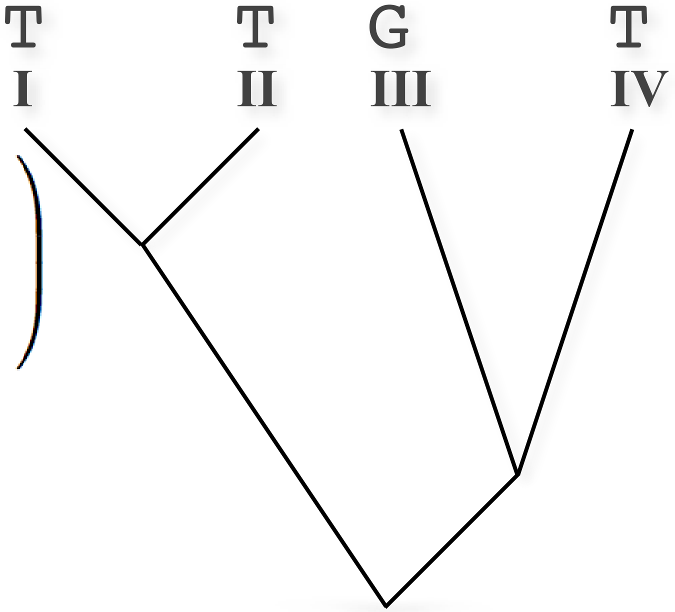
An Introduction to the Likelihood Function

OK, so what does likelihood ***really*** mean?

The four species have the following site pattern: `TTGT`

What is the likelihood of observing the site pattern under this model?

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1.916 & 0.541 & 0.787 & 0.588 \\ 0.148 & -1.069 & 0.415 & 0.506 \\ 0.286 & 0.170 & -0.591 & 0.135 \\ 0.525 & 0.236 & 0.594 & -1.355 \end{pmatrix}$$



Using Monte Carlo simulation to estimate site likelihood is pedagogical, but too inefficient for the analysis of real data.

e.g., there are $>1{,}000{,}000$ possible site patterns for a tree with 10 species.

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a single site:

- ~~Using Monte Carlo simulation~~
- Using brute force
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an entire alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a <u>single</u> site:

- ~~Using Monte Carlo simulation~~
- Using brute force
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an <u>entire</u> alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

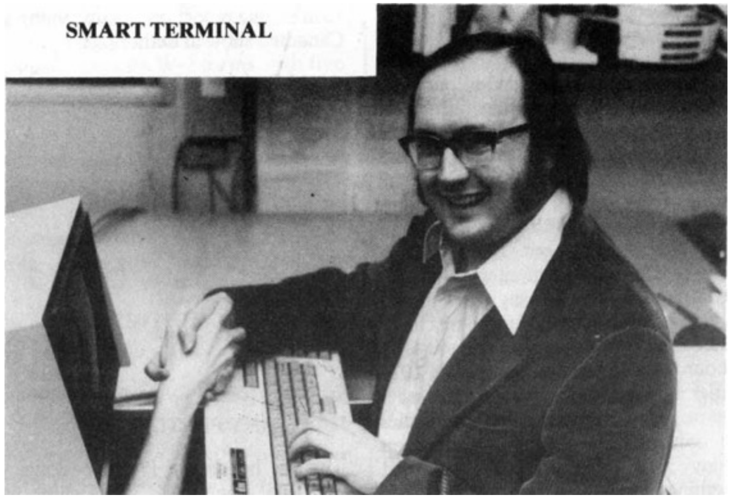
How Do We Calculate the Likelihood?

Species	Sequence data
Species I	`TCG--CACCGGCGCAGTCA....`
Species II	`TCGTTCA--GGCG--GTCA....`
Species III	`GCGTTCACCGGCGCAGTCA....`
Species IV	`TCGTTCACCGGCGCAGTCA....`

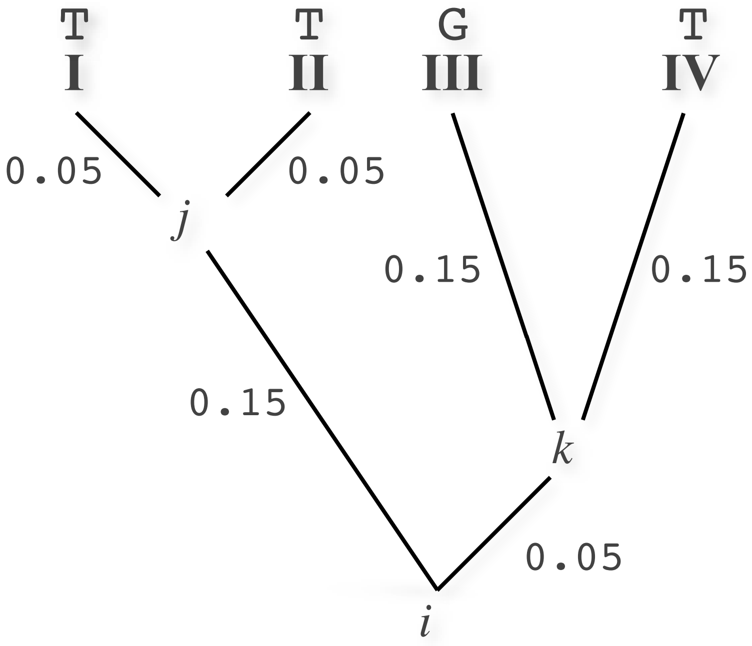
How Do We Calculate Site Likelihoods?

Brute-force solution

We could calculate the probability of observing the site pattern TTGT by summing over all possible ancestral-state configurations for internal nodes (j, k, l) that could give rise to the observed states



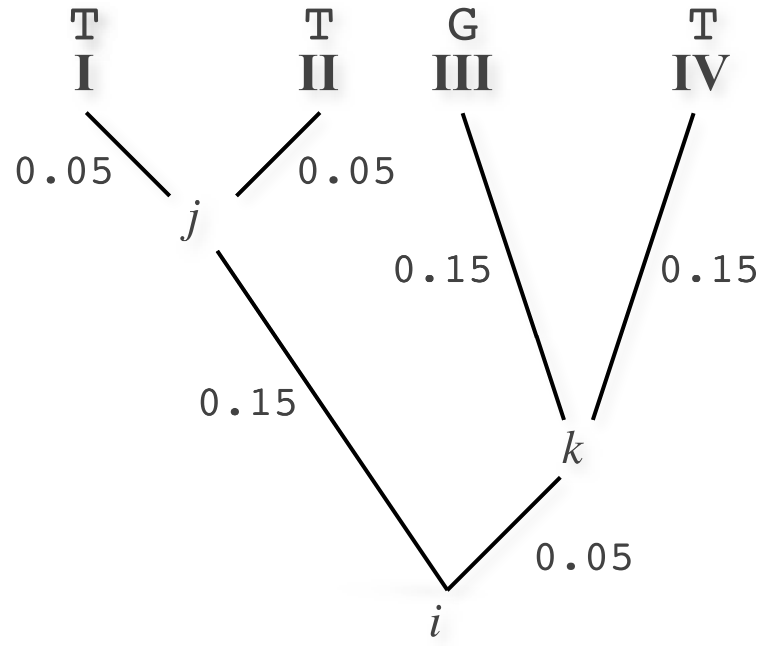
Joe Felsenstein (c.1973)



How Do We Calculate Site Likelihoods?

Brute-force solution

We could calculate the probability of observing the site pattern TTGT by summing over all possible ancestral-state configurations for internal nodes (j, k, l) that could give rise to the observed states



$$\sum_{i \in (A,C,G,T)} \sum_{j \in (A,C,G,T)} \sum_{k \in (A,C,G,T)} \pi_i \times p_{ij}(0.15) \times p_{jT}(0.05) \times p_{jT}(0.05) \times p_{ik}(0.05) \times p_{kG}(0.15) \times p_{kT}(0.15)$$

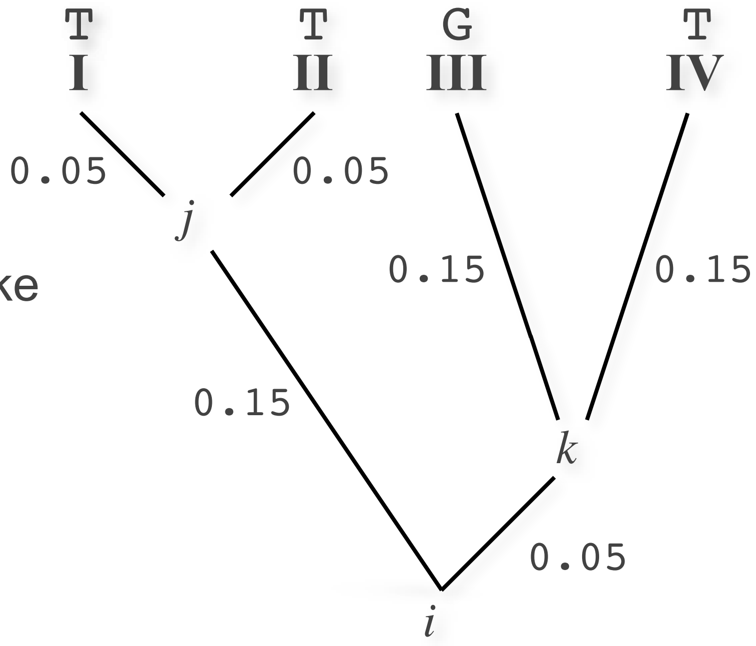
How Do We Calculate Site Likelihoods?

Brute-force solution

Although more efficient than Monte Carlo simulation, this approach is too expensive to be practical.

For example, a tree with 100 species has $(S-1) = 99$ internal nodes, and so entails $4^{99} = 4.02 \times 10^{59}$ possible ancestral-state configurations

Even if a computer could evaluate 1 billion configurations/second, the calculation would take *way* too long …



How Do We Calculate Site Likelihoods?

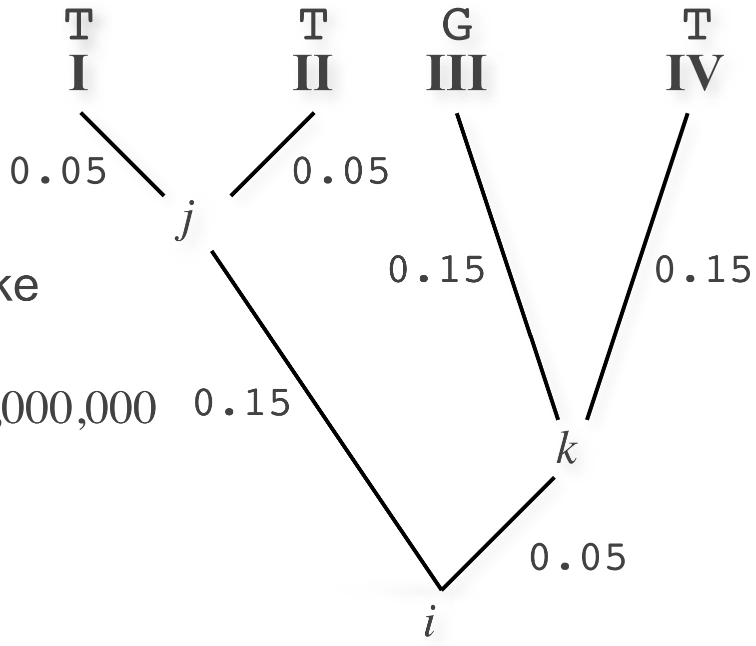
Brute-force solution

Although more efficient than Monte Carlo simulation, this approach is too expensive to be practical.

For example, a tree with 100 species has $(S-1) = 99$ internal nodes, and so entails $4^{99} = 4.02 \times 10^{59}$ possible ancestral-state configurations

Even if a computer could evaluate 1 billion configurations/second, the calculation would take *way* too long …

127,000,000,000,000,000,000,000,000,000,000,000,000,000,000 YEARS!!!



Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a single site:

- ~~Using Monte Carlo simulation~~
- ~~Using brute force~~
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an entire alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

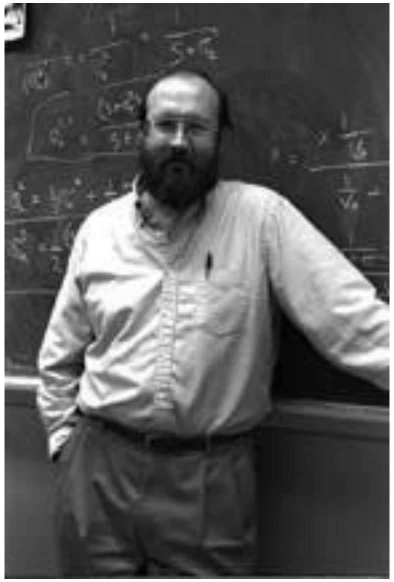
IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

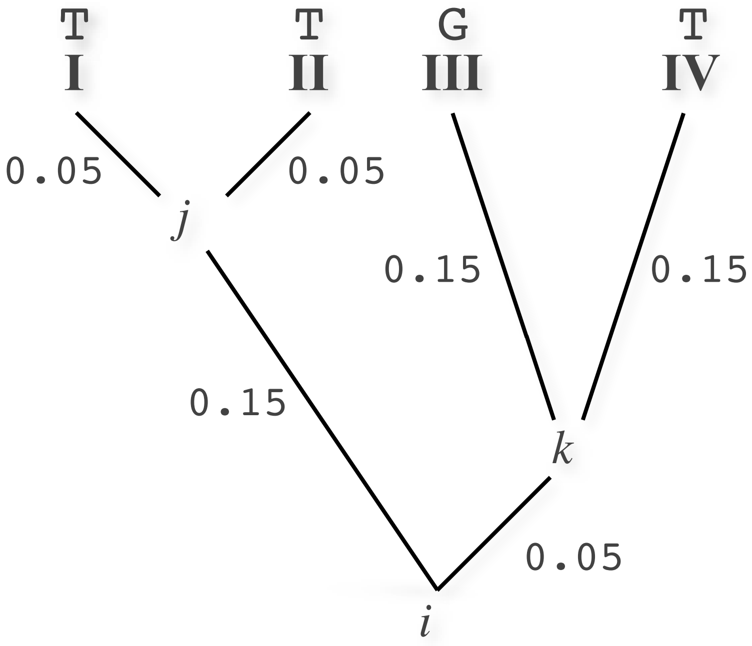
How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Make the math mirror the tree to avoid redundant calculations.



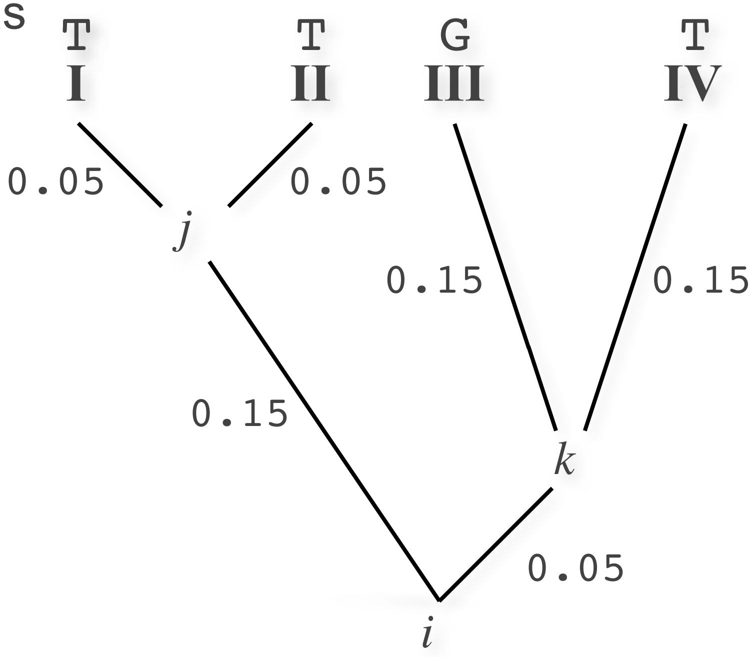
Joe Felsenstein (c.1981)



How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

- Make the math mirror the tree to avoid redundant calculations.
- Multiple summations over ancestral-state configurations at internal nodes are moved as far to the right as possible.

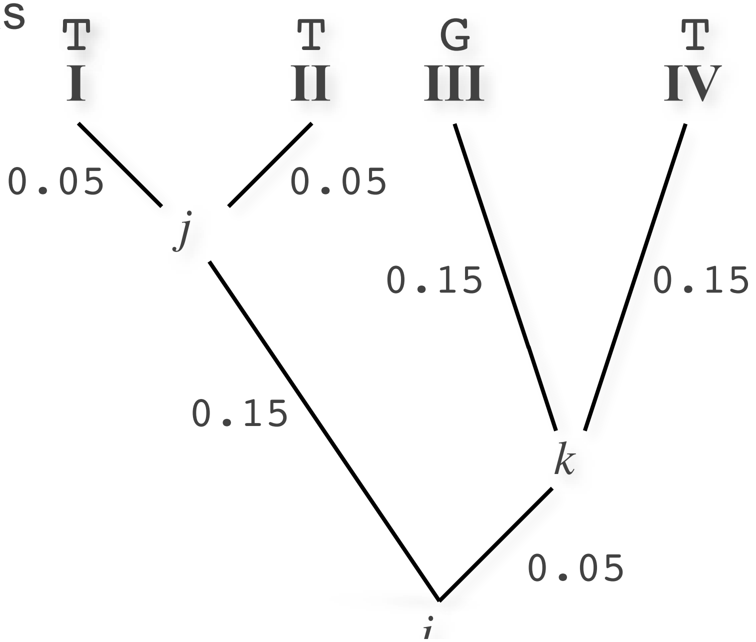


How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Make the math mirror the tree to avoid redundant calculations.

Multiple summations over ancestral-state configurations at internal nodes are moved as far to the right as possible.



The pruning algorithm reduces this (64 evaluations):

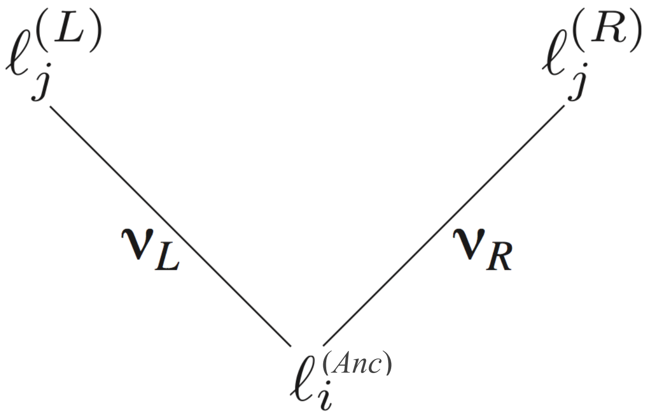
$$\sum_{i\in(A,C,G,T)}\sum_{j\in(A,C,G,T)}\sum_{k\in(A,C,G,T)} \pi_i \times p_{ij}(0.15) \times p_{jT}(0.05) \times p_{jT}(0.05) \times p_{ik}(0.05) \times p_{kG}(0.15) \times p_{kT}(0.15)$$

to this (12 evaluations):

$$\sum_{i\in(A,C,G,T)} \pi_i \left(\sum_{j\in(A,C,G,T)} p_{ij}(0.15) \times p_{jT}(0.05) \times p_{jT}(0.05)\right)\left(\sum_{k\in(A,C,G,T)} p_{ik}(0.05) \times p_{kG}(0.15) \times p_{kT}(0.15)\right)$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm



Evaluating the likelihood of a site involves the recursive calculation of *conditional likelihoods* from the tips of the tree to the root.

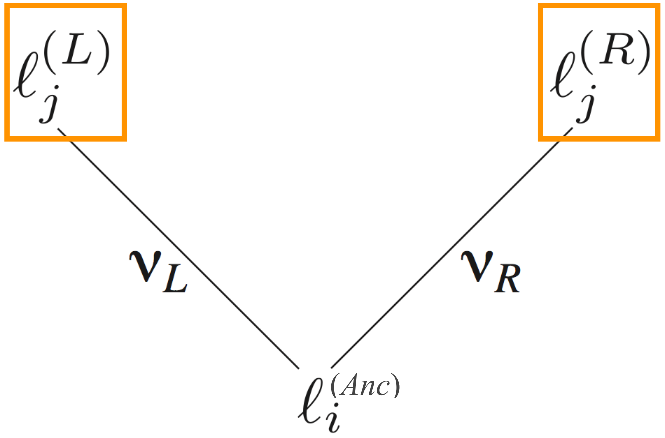
The conditional likelihoods, ℓ_i , are the probabilities of the observations above point i in the tree, conditional on state i at that node.

$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)} \right)$$

where $i = \{\mathrm{A, C, G, T}\}$.

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

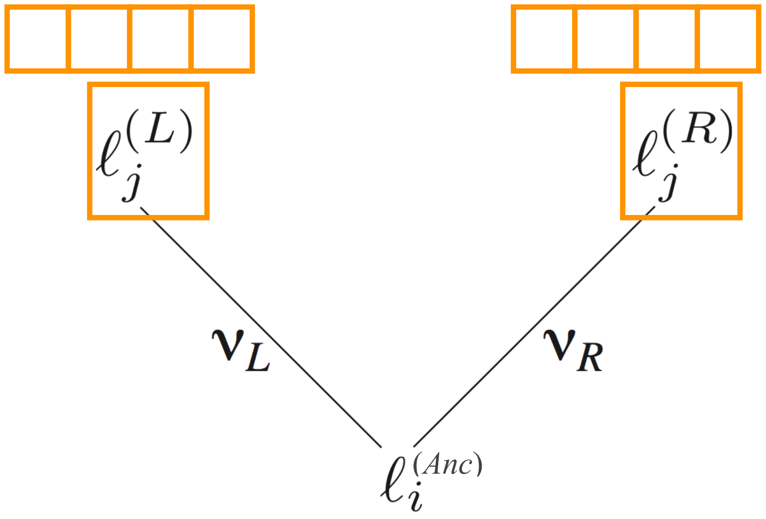


$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L) \boxed{\ell_j^{(L)}} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R) \boxed{\ell_j^{(R)}} \right)$$

These conditional likelihoods are the likelihoods of observing i at or above the end of each of the branch...

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm



$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L) \boxed{\ell_j^{(L)}} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R) \boxed{\ell_j^{(R)}} \right)$$

These conditional likelihoods are either known (if tip) or have already been computed (in a previous step).

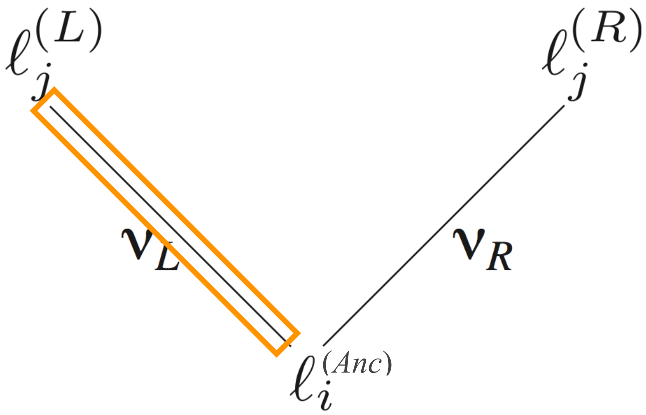
if the branch is a tip, the ℓ_i of the observed state is 1, otherwise it is 0 (*e.g.*, depicted for states A and C in L and R).

if the branch is not a tip, the ℓ_i of each possible state has previously been evaluated in a prior step.

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

So, *how* do we calculate the conditional likelihoods?



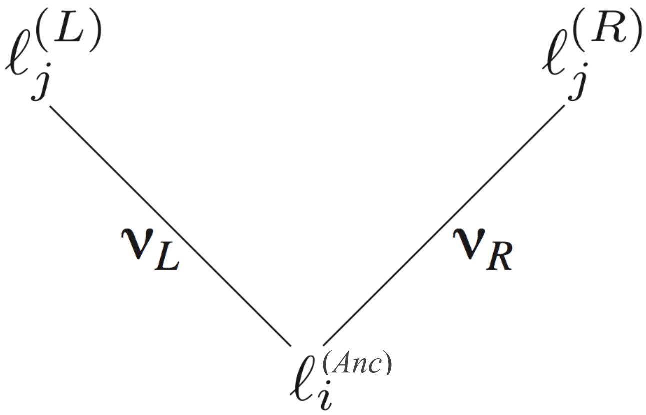
$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right)$$

Let’s first focus on the left descendant branch.

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

So, *how* do we calculate the conditional likelihoods?



$$\ell_i^{(Anc)} = \left(\boxed{\sum_{j\in(A,C,G,T)}} \boxed{p_{ij}(\nu_L)} \boxed{\ell_j^{(L)}} \right)$$

for each possible end state sum…

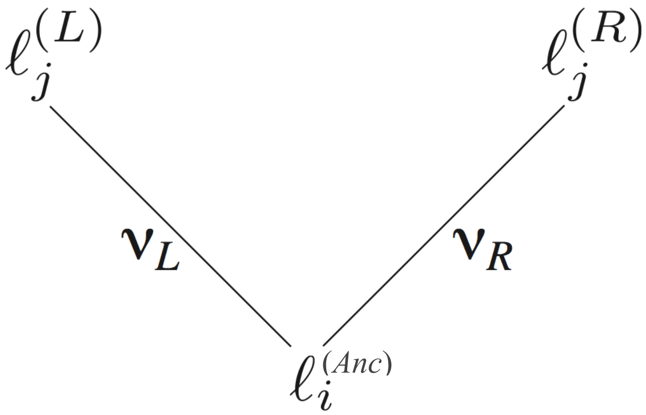
…the product of the transition probabilities of changes from i to j over branch L…

…and the conditional likelihood of each end state, j.

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

So, *how* do we calculate the conditional likelihoods?



$$\ell_i^{(Anc)} = \left(\sum_{j\in(A,C,G,T)} \boxed{p_{ij}(\nu_L)} \ell_j^{(L)} \right)$$

$$\mathbf{Q} = q_{ij} = \begin{pmatrix} - & \mu a \pi_C & \mu b \pi_G & \mu c \pi_T \\ \mu a \pi_A & - & \mu d \pi_G & \mu e \pi_T \\ \mu b \pi_A & \mu d \pi_C & - & \mu f \pi_T \\ \mu c \pi_A & \mu e \pi_C & \mu f \pi_G & - \end{pmatrix}$$

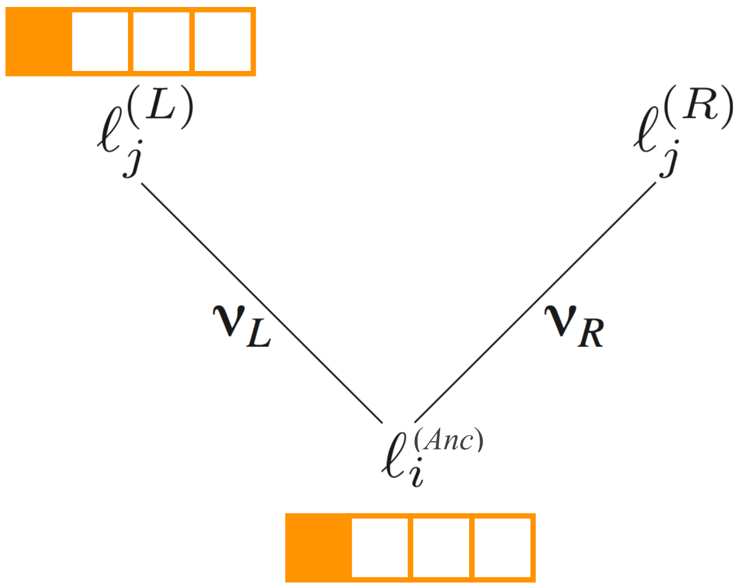
$$\mathbf{P}(v) = e^{\mathbf{Q}v}$$

Recall that the transition probabilities, $\mathbf{P}_{ij}(v_L)$, of histories ending in state j that were initiated from state i and run over branch L are either approximated (by Monte Carlo simulation) or solved by exponentiating the product of the instantaneous-rate matrix, $\mathbf{Q}$, and branch length, v_L.

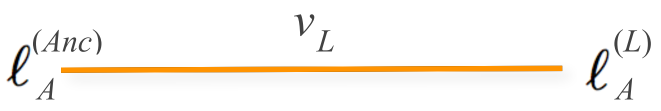
How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

First, set the start state, i, to $i = \mathrm{A}$
and set the end state, j, to $j = \mathrm{A}$



$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right)$$

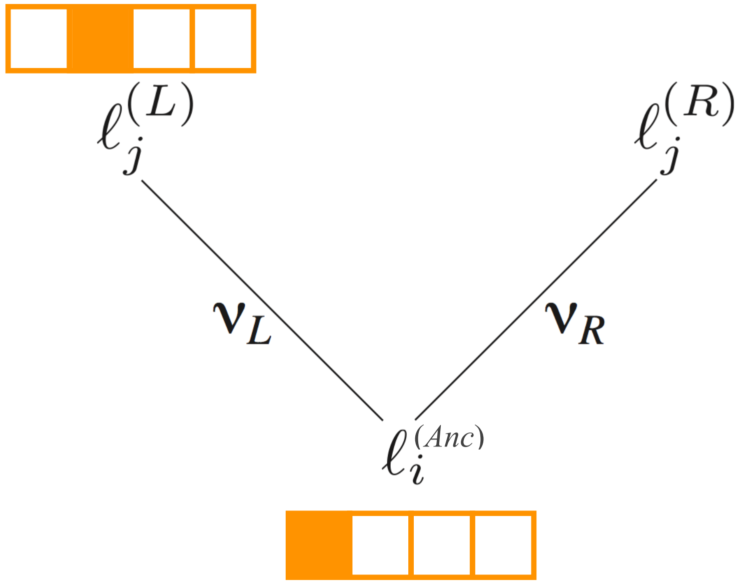


The transition probabilities $p_{AA}(v_L)$ over branch v_L are calculated by exponentiating the product of the instantaneous rate matrix, $\mathbf{Q}$, and branch length v_L.

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

First, set the start state, i, to $i = \mathrm{A}$
and set the end state, j, to $j = \mathrm{A}$
next set the end state, j, to $j = \mathrm{C}$
then set the end state, j, to $j = \mathrm{G}\ldots$

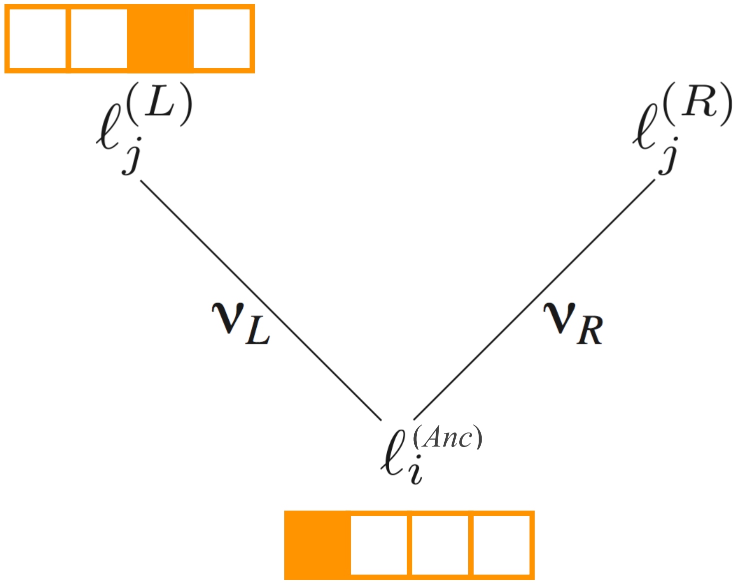


$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right)$$

Then we do the same thing for end state $j = \mathrm{G}\ldots$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm



First, set the start state, i, to $i = \mathrm{A}$
and set the end state, j, to $j = \mathrm{A}$
next set the end state, j, to $j = \mathrm{C}$
then set the end state, j, to $j = \mathrm{G}$
then set the end state, j, to $j = \mathrm{T}$

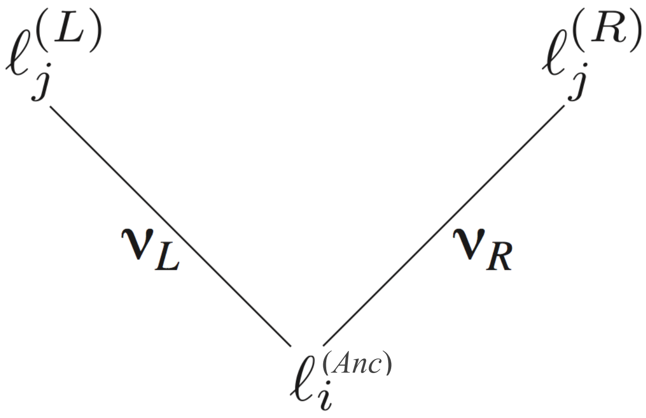
$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right)$$

Finally, we do the same thing for end state $j = \mathrm{T}$

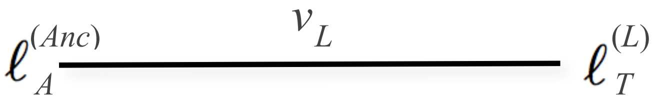
How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Next, we sum the fractional likelihoods for each of the four end states $j = \{A, C, G, T\}$…



$$\ell_i^{(Anc)} = \left(\boxed{\sum_{j\in(A,C,G,T)}} p_{ij}(\nu_L)\ell_j^{(L)} \right)$$

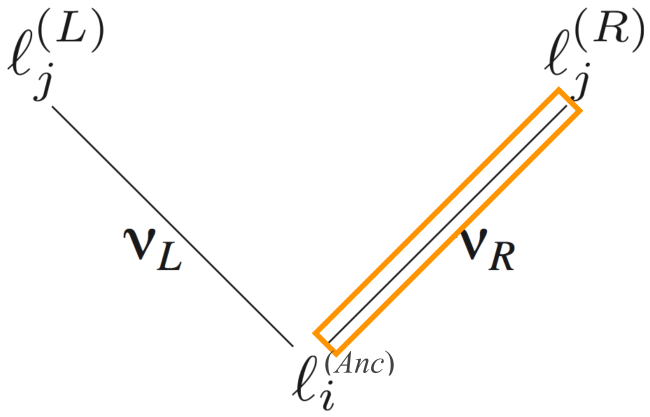


Then we sum the fractional likelihoods for each of the four end states $j = \{A, C, G, T\}$…

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Next, we repeat the process for the right branch, v_R.

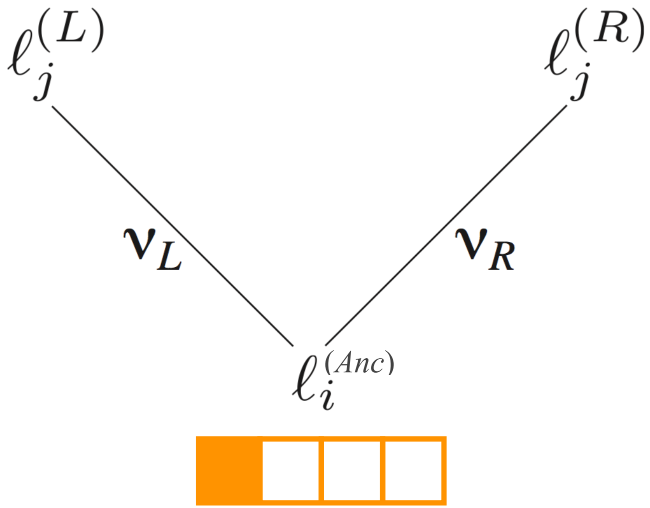


$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right) \times \boxed{\left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)} \right)}$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Finally, we calculate $\ell_A^{(Anc)}$ as the product of the conditional likelihoods of the two descendant branches, which makes explicit the assumption that substitutions along these two lineages are independent.

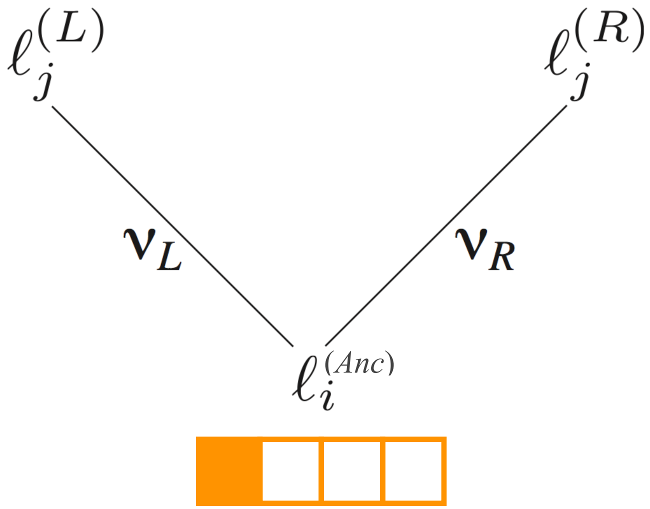


$$\ell_i^{(Anc)} = \boxed{\left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right)} \times \boxed{\left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)} \right)}$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Now we have computed $\ell_A^{(Anc)}$…

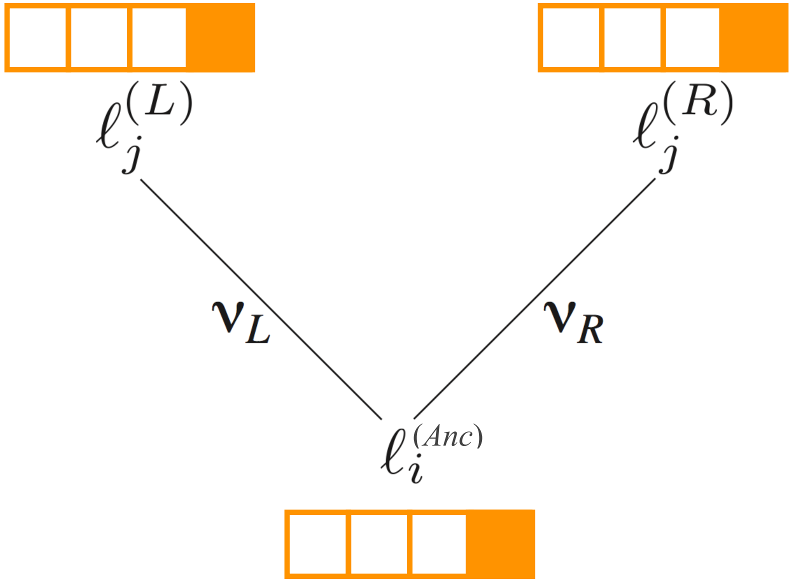


$$\ell_i^{(Anc)} = \left(\sum_{j\in(A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)}\right) \times \left(\sum_{j\in(A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)}\right)$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

So, now we need to repeat the entire process
for $i = \mathrm{C}$,
for $i = \mathrm{G}$,
and for $i = \mathrm{T}$.

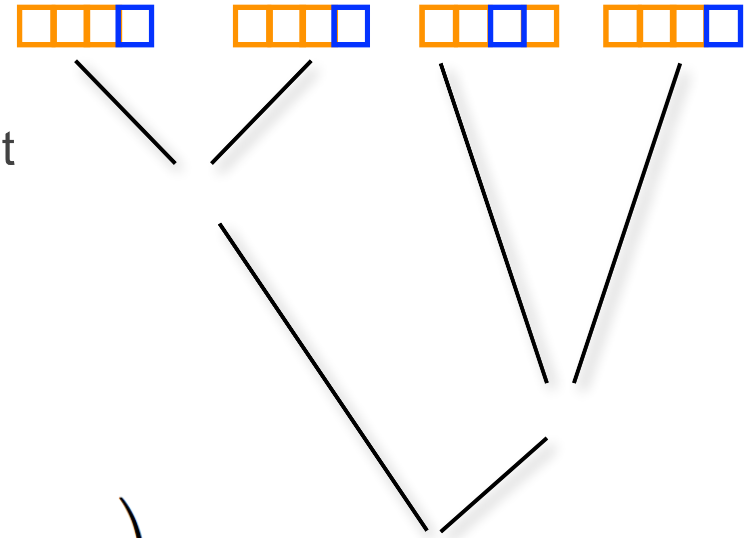


$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)} \right)$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Then we need to repeat the entire process for each of the more inclusive nodes toward the root, where the conditional likelihoods of the tips are first recorded...

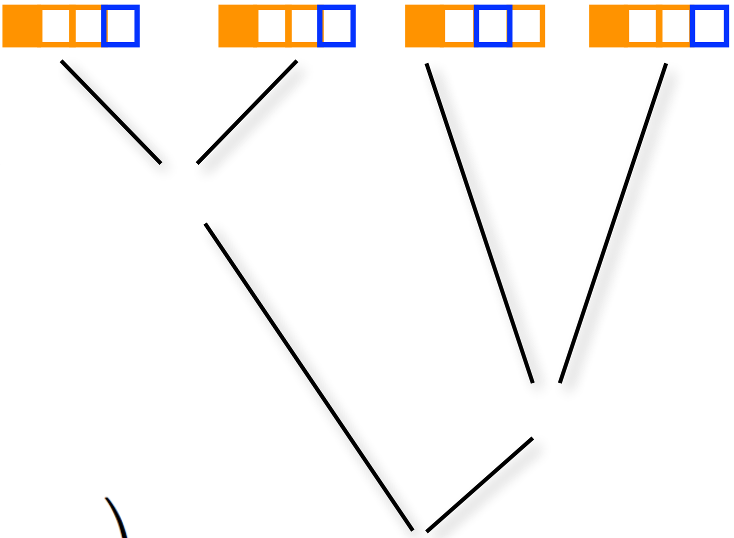


$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L) \ell_j^{(L)} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R) \ell_j^{(R)} \right)$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Then we need to repeat the entire process for each of the more inclusive nodes toward the root, where the conditional likelihoods of the tips are first recorded...

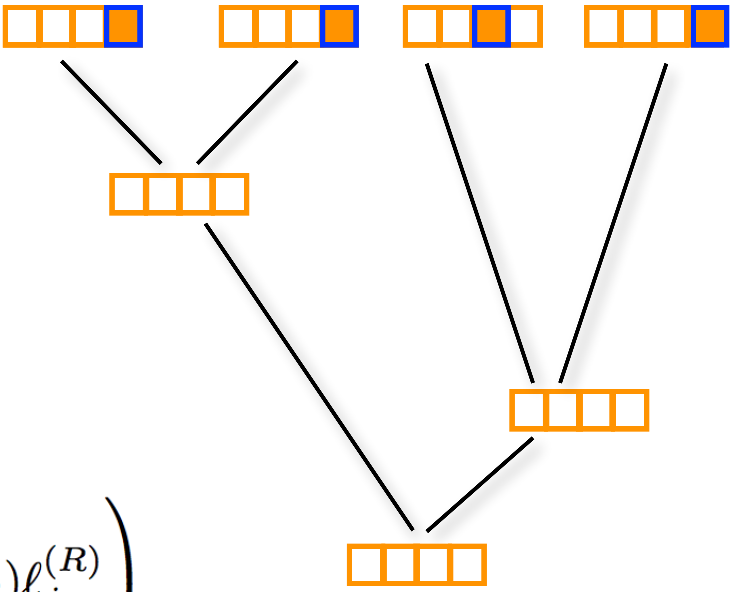


$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)} \right)$$

How Do We Calculate Site Likelihoods?

The Felsenstein Pruning Algorithm

Upon reaching the root of the tree, the conditional likelihood of each state is ‘weighted’ by the prior probability (stationary frequency) of the corresponding state, π_i, to give the unconditional probability of the data, x_i.



$$\ell_i^{(Anc)} = \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_L)\ell_j^{(L)} \right) \times \left(\sum_{j \in (A,C,G,T)} p_{ij}(\nu_R)\ell_j^{(R)} \right)$$

$$\mathbb{P}(TTGT) = \pi_A \ell_A^{(\mathrm{Root})} + \pi_C \ell_C^{(\mathrm{Root})} + \pi_G \ell_G^{(\mathrm{Root})} + \pi_T \ell_T^{(\mathrm{Root})}$$

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a single site:

- ~~Using Monte Carlo simulation~~
- ~~Using brute force~~
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an entire alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

Calculating the Likelihood of an Alignment

Calculating the likelihood of an alignment assumes independent of sites

We generally assume that the substitution process is independent across sites, which allows us to combine their probabilities as the product of the N site likelihoods.

$$L(\tau, \nu, \Theta) \propto f(\mathbf{X} \mid \tau, \nu, \Theta) = \prod_{i=1}^{N} f(x_i \mid \tau, \nu, \Theta)$$

That is, we compute the likelihood for each site in the alignment (one site at a time) and then combine them as the product

Although this is a convenient assumption, it is not very biologically realistic, but it can be relaxed.

Calculating the Likelihood of an Alignment

Compressing the alignment using unique site patterns

Because we assume that sites are independent, sites with the same site pattern (*e.g.*, TTGT) will have the same likelihood

Therefore, we can avoid redundant calculations (and increase computational efficiency) by identifying the set of N' unique site patterns.

```
species I      TCGCACCGGCGTA...
species II     TCGTTCAGGCGTT...
species III    GCGTTCACCGGGG...
species IV     TCGTTCACCCGTT...
```

$$L(\tau, \nu, \Theta) \propto f(\mathbf{X} \mid \tau, \nu, \Theta) = \prod_{i=1}^{N'} f(x_i \mid \tau, \nu, \Theta)^{c_i}$$

where c_i is the number of instances of the unique site pattern i.

Calculating the Likelihood of an Alignment

Why do we calculate log likelihoods?

Conditional likelihoods for a given site are real numbers with values < 1.

Because an alignment typically has thousands of sites, the product of these numbers can quickly become too small to be held correctly in computer memory (a problem called '*underflow*').

To avoid underflow, we sum the logs of the site likelihoods:

$$L(\tau, \nu, \Theta) \propto f(\mathbf{X} \mid \tau, \nu, \Theta) = \prod_{i=1}^{N} f(x_i \mid \tau, \nu, \Theta)$$

$$L(\tau, \nu, \Theta) \propto \ln L(\mathbf{X} \mid \tau, \nu, \Theta) = \sum_{i=1}^{N} \ln f(x_i \mid \tau, \nu, \Theta)$$

Using the log likelihood does not alter the estimates of parameters.

Statistical Estimation of Phylogeny: An Outline

Data	**Model**	**Inference Method**	**Parameter Estimates**
sequence alignment	tree with branch lengths	maximum likelihood	phylogeny & other model parameters
	substitution model	Bayesian inference	

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a single site:

- ~~Using Monte Carlo simulation~~
- ~~Using brute force~~
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an entire alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

Bayesian Inference

Conditional Probability

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}$$

Joint Probability

The probability of observing both A and B, $\Pr(A,B)$, is therefore:

$$\Pr(A,B) = \Pr(B)\Pr(A \mid B)$$

and by the same reasoning:

$$\Pr(A,B) = \Pr(A)\Pr(B \mid A)$$

which is the probability of observing A times the probability of observing B given that A has occurred.

Bayesian Inference

~~Conditional Probability~~ Bayes Theorem

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(A)\Pr(B \mid A)}{\Pr(B)}$$

Joint Probability

The probability of observing both A and B, $\Pr(A,B)$, is therefore:

$$\Pr(A,B) = \Pr(B)\Pr(A \mid B)$$

and by the same reasoning:

$$\Pr(A,B) = \Pr(A)\Pr(B \mid A)$$

which is the probability of observing A times the probability of observing B given that A has occurred.

Bayesian Inference

Bayes Theorem

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\,\Pr(A)}{\Pr(B)}$$

Bayesian Inference

Bayes Theorem

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(\boxed{A} \mid \boxed{B}) = \frac{\Pr(B \mid A)\,\Pr(A)}{\Pr(B)}$$

unobserved outcome

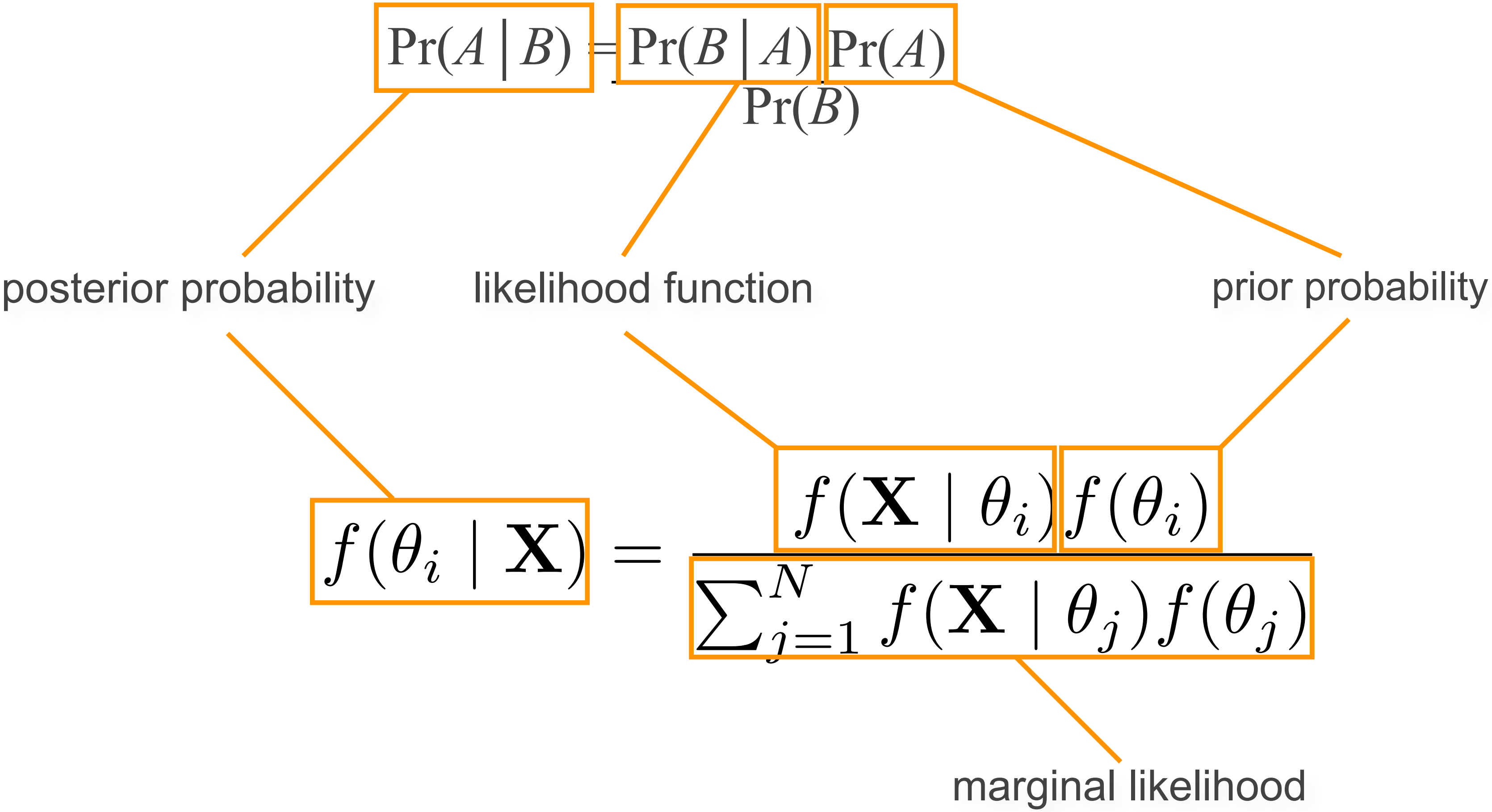
observed outcome

$$f(\boxed{\theta_i} \mid \boxed{\mathbf{X}}) = \frac{f(\mathbf{X} \mid \theta_i) f(\theta_i)}{\sum_{i=1}^{N} f(\mathbf{X} \mid \theta_j) f(\theta_j)}$$

Bayesian Inference

Bayes Theorem

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.



$$\boxed{\Pr(A \mid B)} = \frac{\boxed{\Pr(B \mid A)}\,\boxed{\Pr(A)}}{\Pr(B)}$$

posterior probability

likelihood function

prior probability

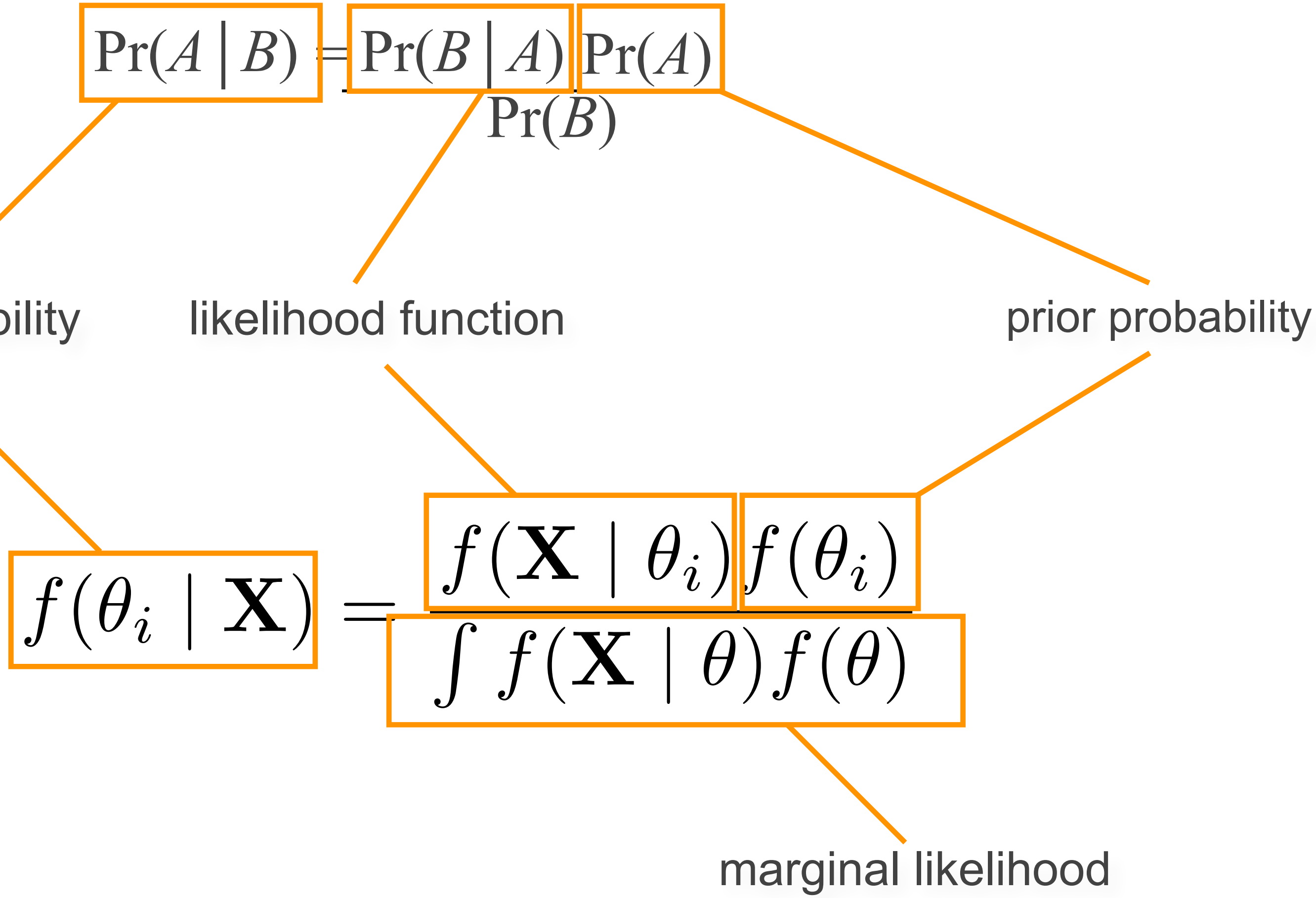
$$\boxed{f(\theta_i \mid \mathbf{X})} = \frac{\boxed{f(\mathbf{X} \mid \theta_i)}\,\boxed{f(\theta_i)}}{\boxed{\sum_{j=1}^{N} f(\mathbf{X} \mid \theta_j) f(\theta_j)}}$$

marginal likelihood

Bayesian Inference

Bayes Theorem

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.



$$\boxed{\Pr(A \mid B)} = \frac{\boxed{\Pr(B \mid A)}\,\boxed{\Pr(A)}}{\Pr(B)}$$

posterior probability

likelihood function

prior probability

$$\boxed{f(\theta_i \mid \mathbf{X})} = \frac{\boxed{f(\mathbf{X} \mid \theta_i)}\,\boxed{f(\theta_i)}}{\boxed{\int f(\mathbf{X} \mid \theta) f(\theta)}}$$

marginal likelihood

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a single site:

- ~~Using Monte Carlo simulation~~
- ~~Using brute force~~
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an entire alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

Bayesian Inference

Generic statistical paradigm	Coin tossing
pose a substantive question	Is this a fair coin? Or, what is the probability of observing heads in a single toss?
develop a stochastic model with parameters that, if know, would answer the question	Binomial probability distribution with parameter θ (probability of observing heads)
collect observations that are informative about model parameters	toss the coin n times and record the number of heads, x.
find the best estimate of model parameters (by some means) conditioned on (*i.e.*, given) the data at hand	find the best estimate of the θ parameter using Bayesian inference

Bayesian Inference

Example: Coin tossing

$$\underbrace{\boxed{f(\theta \mid x)}}_{\text{posterior probability}} = \frac{\overbrace{\boxed{f(x \mid \theta)}}^{\text{likelihood function}}\overbrace{\boxed{f(\theta)}}^{\text{prior probability}}}{\underbrace{\boxed{\int_0^1 f(x \mid \theta) f(\theta)}}_{\text{marginal likelihood}}}$$

Bayesian Inference

Example: Coin tossing

likelihood function

$$f(\theta \mid x) = \frac{\boxed{f(x \mid \theta)}\, f(\theta)}{\int_0^1 f(x \mid \theta) f(\theta)}$$

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$f(x|\theta) = \binom{n}{x}\underset{\substack{\uparrow\\ \text{heads}}}{\theta^{x}}(1-\theta)^{n-x}$$

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$f(x|\theta) = \binom{n}{x}\theta^{x}(1 \underset{\substack{\uparrow \\ \text{tails}}}{-} \theta)^{n-x}$$

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$f(x|\theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x}$$

possible orderings
of x heads in n tosses

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:

discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

possible orderings
of x heads in n tosses

This is called the Binomial coefficient, and is read 'n choose x':

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

The likelihood function for the Binomial distribution:

$$L(\theta)f(x \mid \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

With some algebra, we can solve for θ to find the MLE:

$$\hat{\theta} = \frac{x}{n}$$

Bayesian Inference

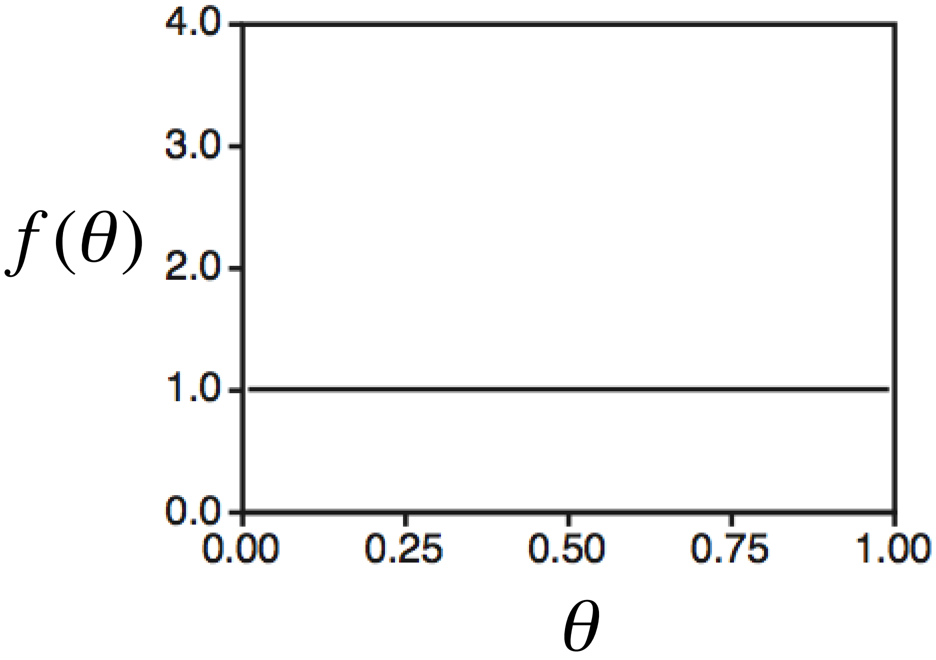
Example: Coin tossing

$$f(\theta \mid x) = \frac{f(x \mid \theta)\overset{\text{prior probability}}{\boxed{f(\theta)}}}{\int_0^1 f(x \mid \theta) f(\theta)}$$

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:



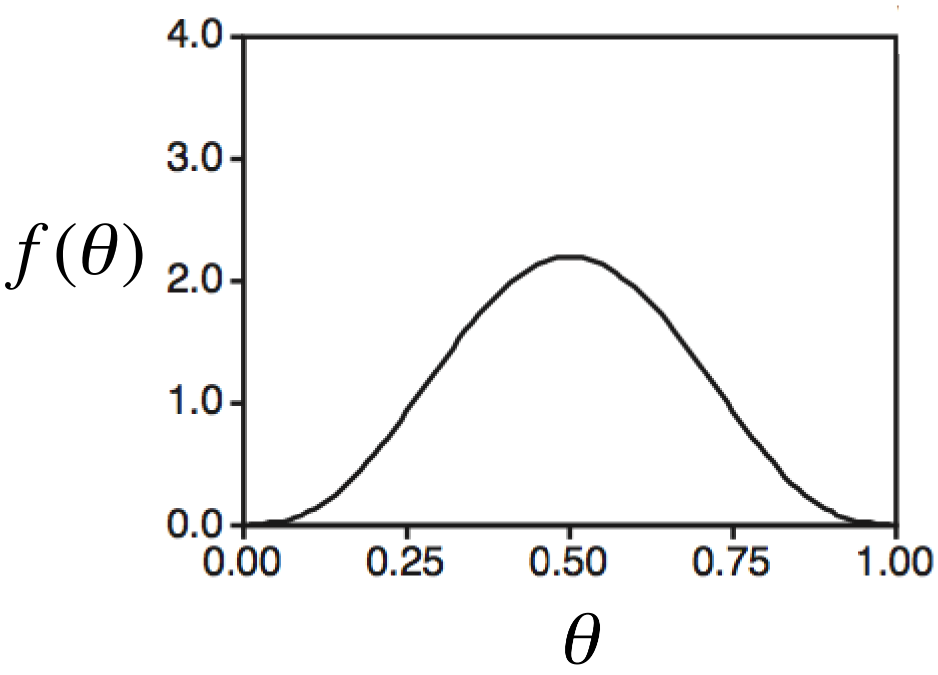
uniform prior: $alpha = beta = 1$

NOTE: uniform prior $\neq$ uninformative

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:

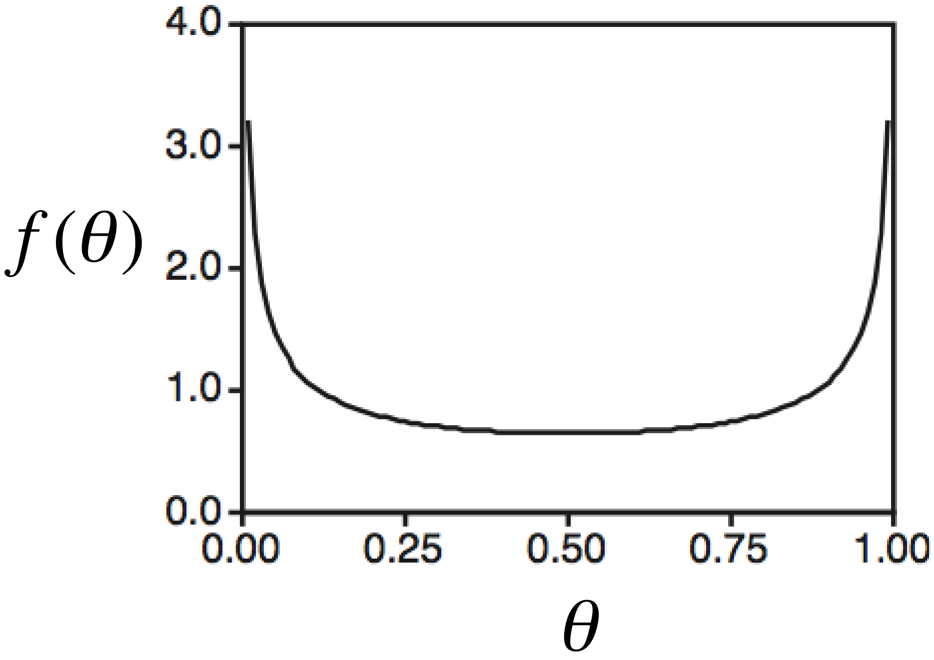


nonuniform prior: $alpha = beta = 4$

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:

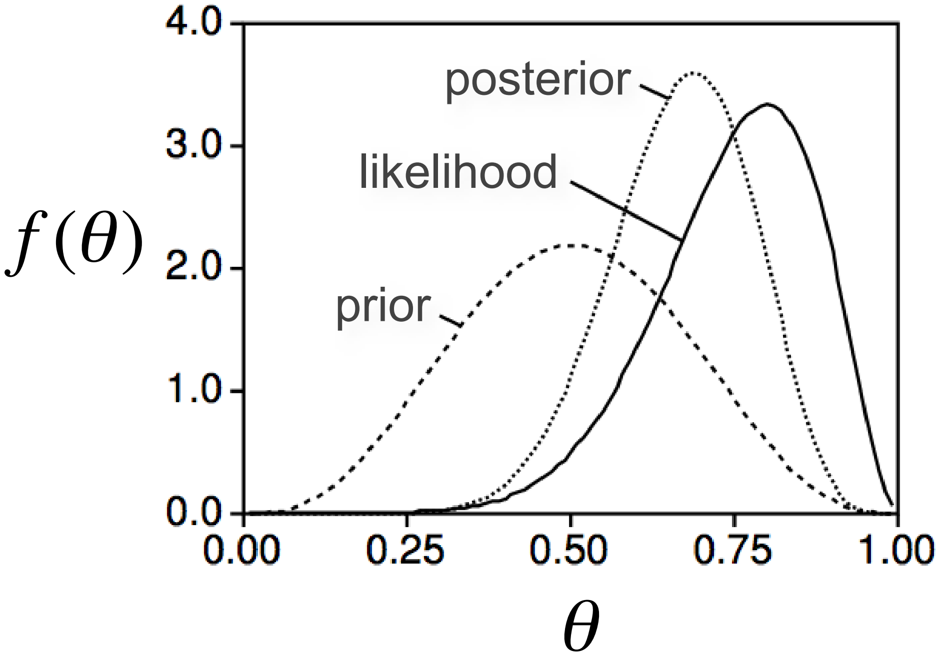


nonuniform prior: $alpha = beta = 0.5$

Bayesian Inference

Example: Coin tossing

The impact of the prior probability distribution on the estimated posterior probability:

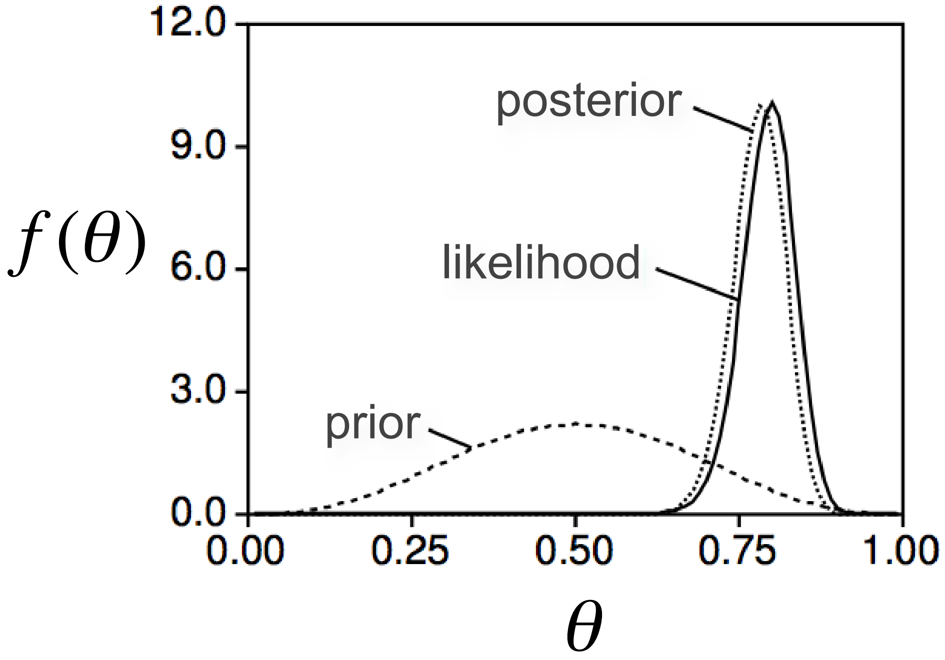


$x = 8$ heads in $n = 10$ tosses

Bayesian Inference

Example: Coin tossing

The impact of the prior probability distribution on the estimated posterior probability:



$x = 80$ heads in $n = 100$ tosses

The impact of the prior decreases as the number of observations increases.

Outline

I. A brief introduction to the likelihood function

What, exactly, is likelihood anyway?

II. Calculating the likelihood of the data

How to calculate the likelihood of a single site:

- ~~Using Monte Carlo simulation~~
- ~~Using brute force~~
- Using the Felsenstein pruning algorithm

How to calculate the likelihood of an entire alignment:

III. Introduction to Bayesian inference

Becoming fascinated with posteriors:

- Deriving Bayes theorem
- A non-phylogenetic example
- Bayesian inference of phylogeny

IV. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

Bayesian Inference of Phylogeny (on one slide)

posterior probability

likelihood function

prior probability

$$\boxed{f(\text{Parameter} \mid \text{Data})} = \frac{\boxed{f(\text{Data} \mid \text{Parameter})}\boxed{f(\text{Parameter})}}{\boxed{f(\text{Data})}}$$

marginal likelihood

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_{(2s-5)!!})$

   branch lengths $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_{(2S-3)})$

2. Model of character change $\boldsymbol{\Phi} = (\theta, \pi, \alpha, T)$

   relative substitution rates $\boldsymbol{\theta} = (\theta_{AC}, \theta_{AG}, \theta_{AT}, \theta_{CG}, \theta_{CT}, \theta_{GT})$

   stationary frequencies $\boldsymbol{\pi} = (\pi_A, \pi_C, \pi_G, \pi_T)$

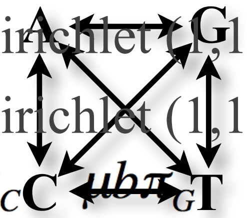
IV. Priors on parameters

$\sim$Uniform

$\sim$Dirichlet (1,...,1)

$\sim$Dirichlet (1,1,1,1,1,1)

$\sim$Dirichlet (1,1,1,1)



III. Phylogenetic likelihood function

$$L(\tau, \nu, \Theta) \propto f(\mathbf{X} \mid \tau, \nu, \Theta) = \prod_{i=1}^{N} f(x_i \mid \tau, \nu, \Theta)$$

$$Q = q_{ij} = \begin{pmatrix} - & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu a\pi_A & - & \mu d\pi_G & \mu e\pi_T \\ \mu b\pi_A & \mu d\pi_C & - & \mu f\pi_T \\ \mu c\pi_A & \mu e\pi_C & \mu f\pi_G & - \end{pmatrix}$$

V. Posterior Probability

$$f(\tau, \nu, \Phi \mid \mathbf{X}) = \frac{f(\mathbf{X} \mid \tau, \nu, \Phi) f(\tau, \nu, \Phi)}{f(\mathbf{X})}$$