Some of these slides have been borrowed from Dr. Paul Lewis, Dr. Joe Felsenstein. Thanks!

Paul has many great tools for teaching phylogenetics at his web site:

http://hydrodictyon.eeb.uconn.edu/people/plewis

Markov **c**hain **M**onte **C**arlo

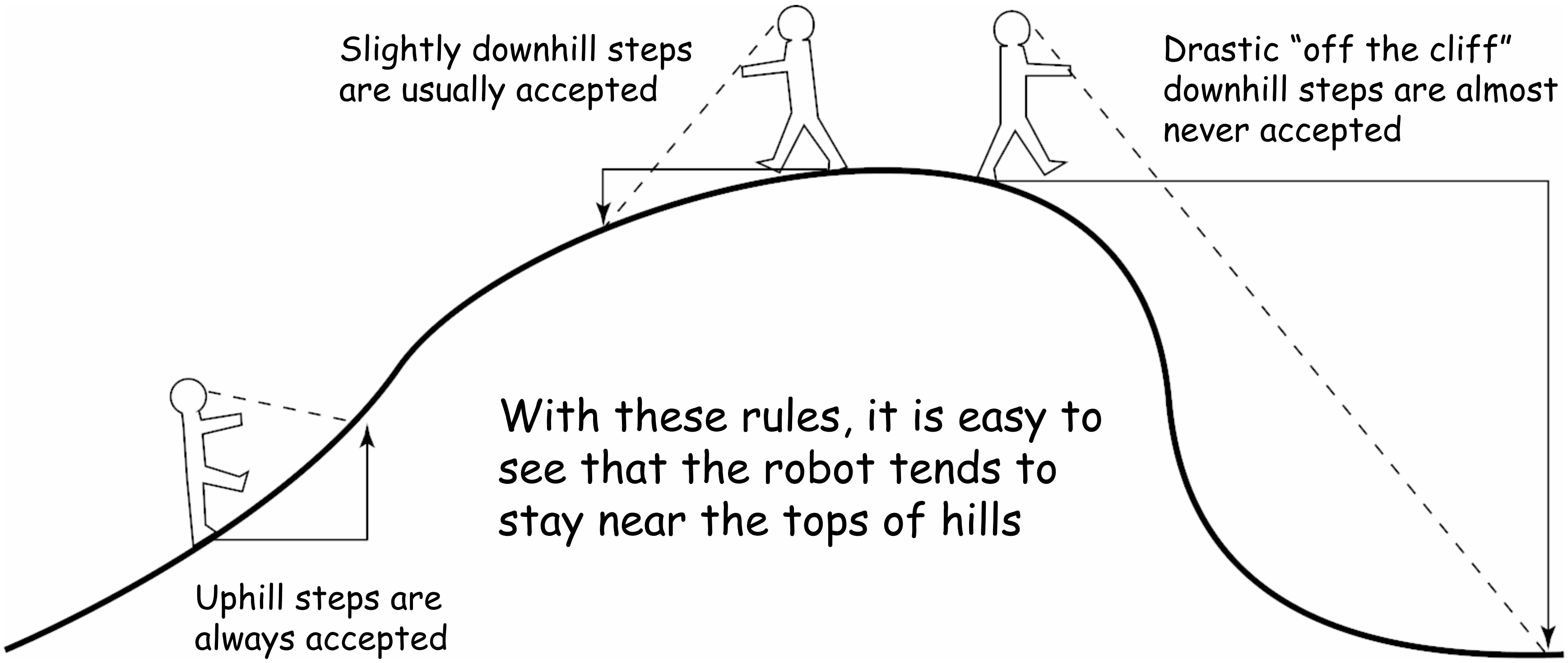
- Simulates a walk through parameter/tree space.
- Lets us estimate posterior probabilities for any aspect of the model
- Relies on the ***ratio*** of posterior densities between two points

$$R = \frac{Pr(\mathsf{Point}_2|\mathsf{Data})}{Pr(\mathsf{Point}_1|\mathsf{Data})}$$

$$R = \frac{\frac{Pr(\mathsf{Point}_2)L(\mathsf{Point}_2)}{Pr(\mathsf{Data})}}{\frac{Pr(\mathsf{Point}_1)L(\mathsf{Point}_1)}{Pr(\mathsf{Data})}}$$

$$R = \frac{Pr\left(\mathsf{Point}_2\right)L\left(\mathsf{Point}_2\right)}{Pr\left(\mathsf{Point}_1\right)L\left(\mathsf{Point}_1\right)}$$

MCMC robot's rules



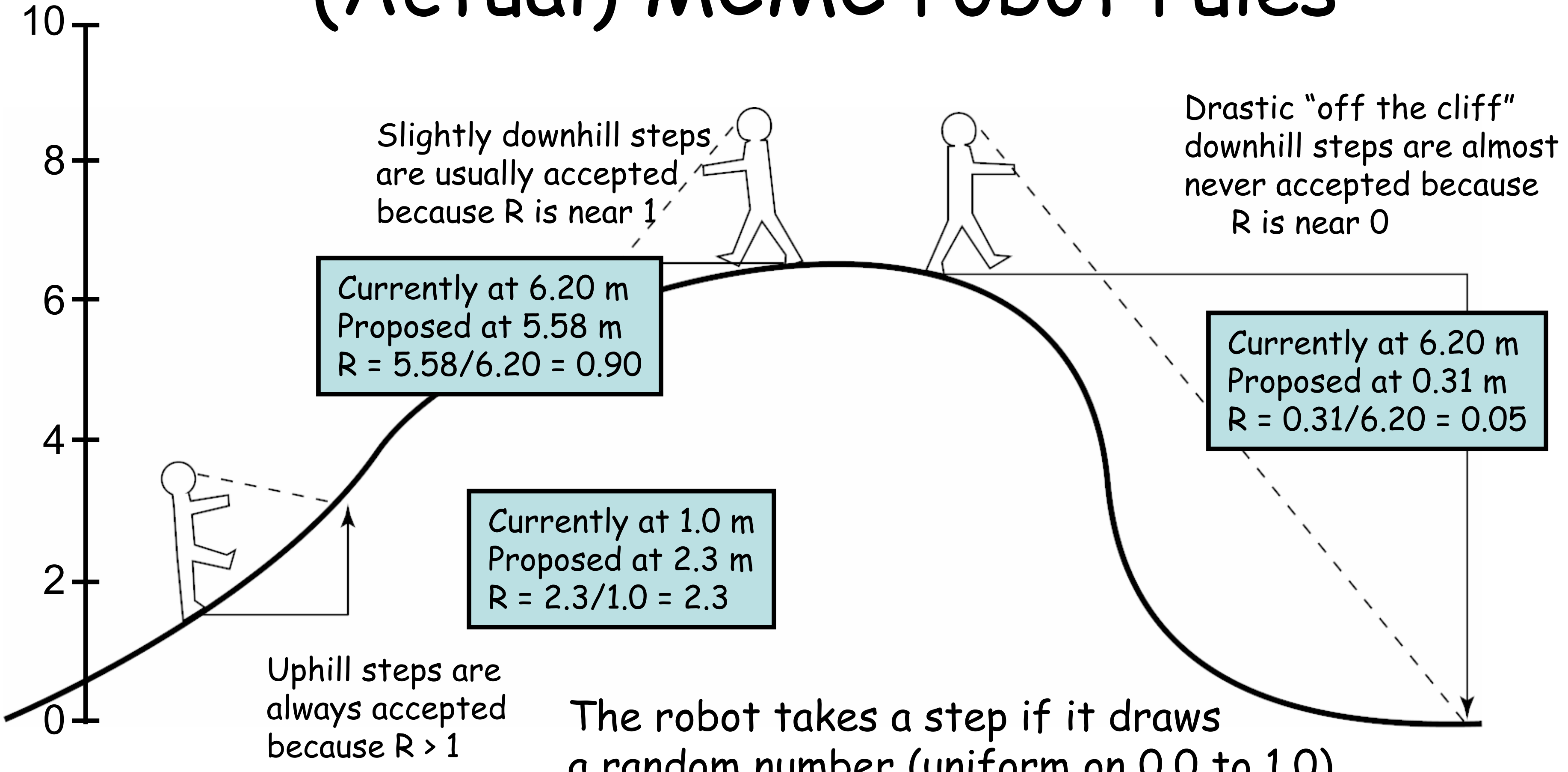
Slightly downhill steps are usually accepted

Drastic "off the cliff" downhill steps are almost never accepted

With these rules, it is easy to see that the robot tends to stay near the tops of hills

Uphill steps are always accepted

(Actual) MCMC robot rules



Slightly downhill steps are usually accepted because R is near 1

Currently at 6.20 m
Proposed at 5.58 m
R = 5.58/6.20 = 0.90

Drastic "off the cliff" downhill steps are almost never accepted because R is near 0

Currently at 6.20 m
Proposed at 0.31 m
R = 0.31/6.20 = 0.05

Currently at 1.0 m
Proposed at 2.3 m
R = 2.3/1.0 = 2.3

Uphill steps are always accepted because R > 1

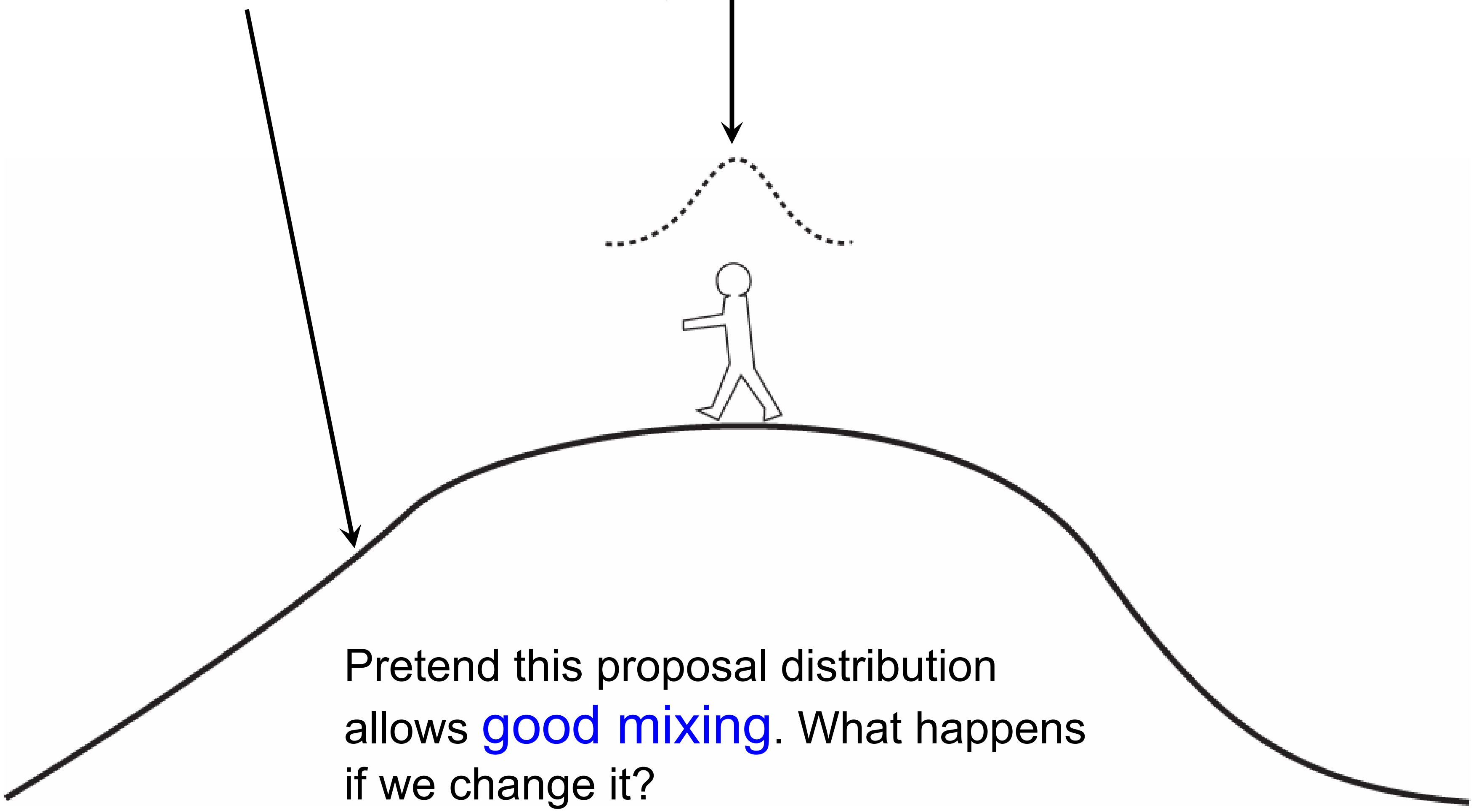
The robot takes a step if it draws a random number (uniform on 0.0 to 1.0), and that number is less than or equal to R

Target vs. proposal distributions

- The **target distribution** is the posterior distribution of interest
- The **proposal distribution** is used to decide which point to try next
  - you have much flexibility here, and the choice affects only the **efficiency** of the MCMC algorithm
  - MCMC using a **symmetric** proposal distribution is the Metropolis algorithm (Metropolis et al. 1953)
  - Use of an **asymmetric** proposal distribution requires a modification proposed by Hastings (1970), and is known as the Metropolis-Hastings algorithm

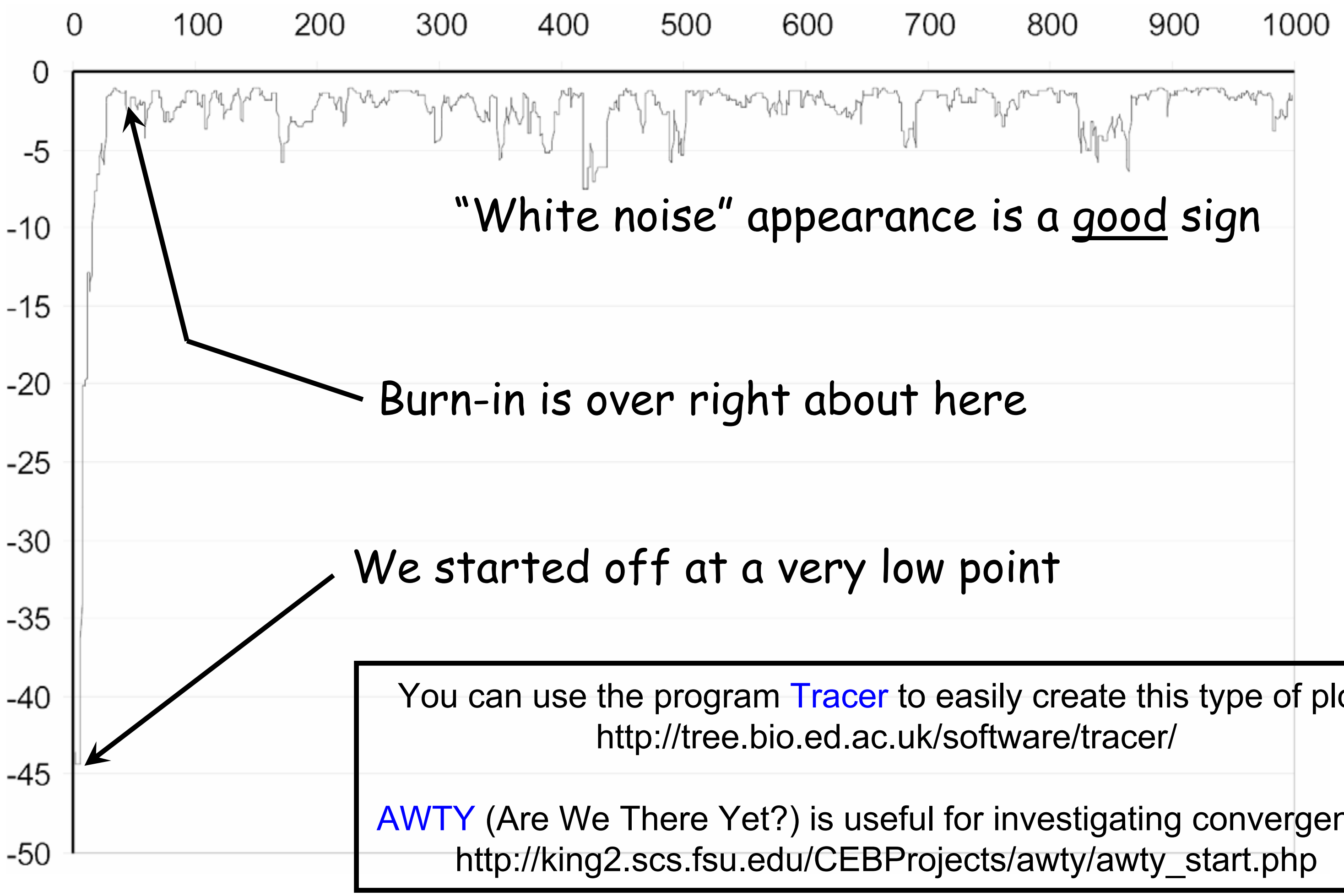
Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087-1092.

Target vs. Proposal Distributions



Pretend this proposal distribution allows good mixing. What happens if we change it?

Trace plots



"White noise" appearance is a good sign

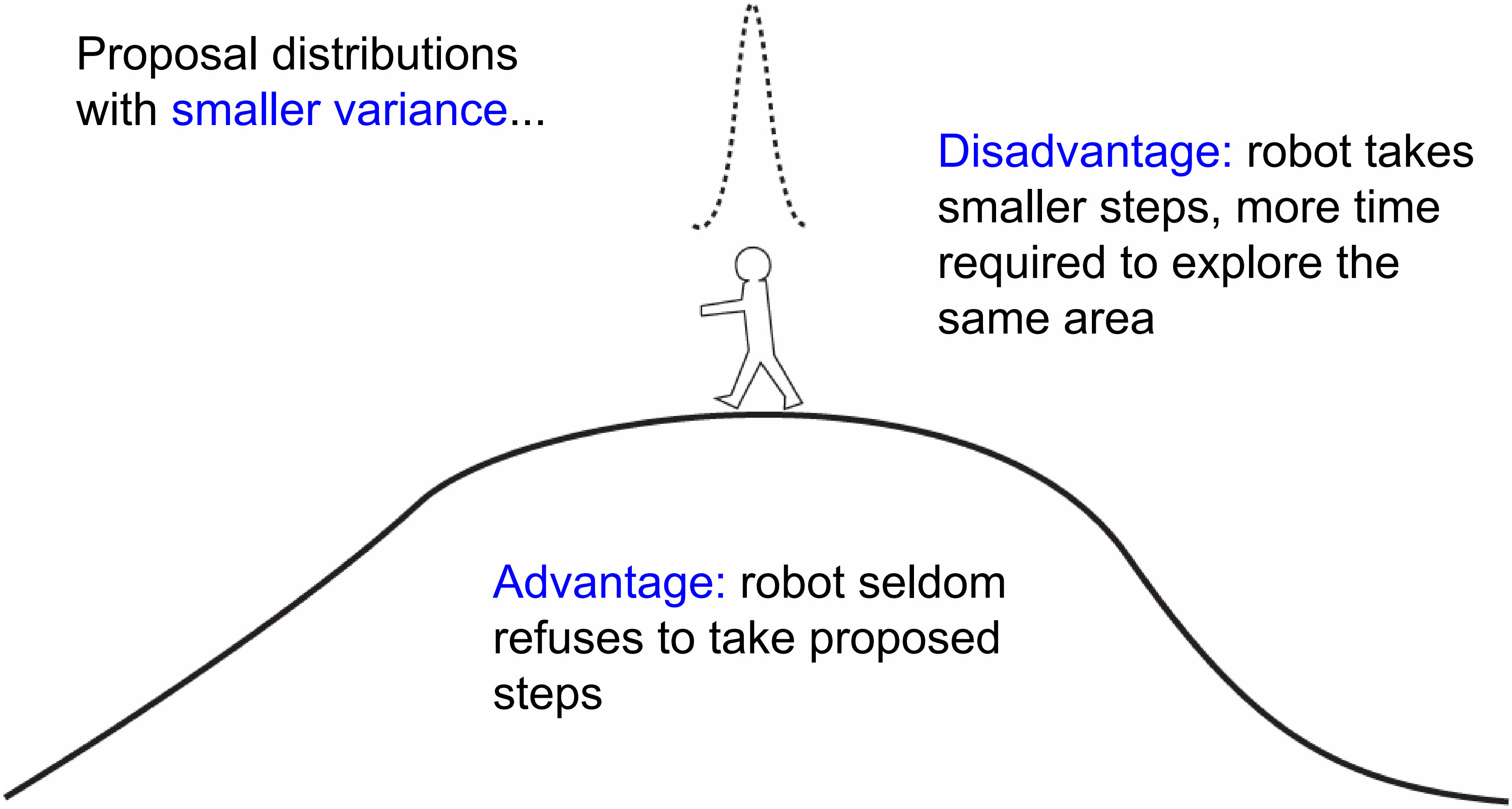
Burn-in is over right about here

We started off at a very low point

You can use the program Tracer to easily create this type of plot:
http://tree.bio.ed.ac.uk/software/tracer/

AWTY (Are We There Yet?) is useful for investigating convergence:
http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php

Target vs. Proposal Distributions



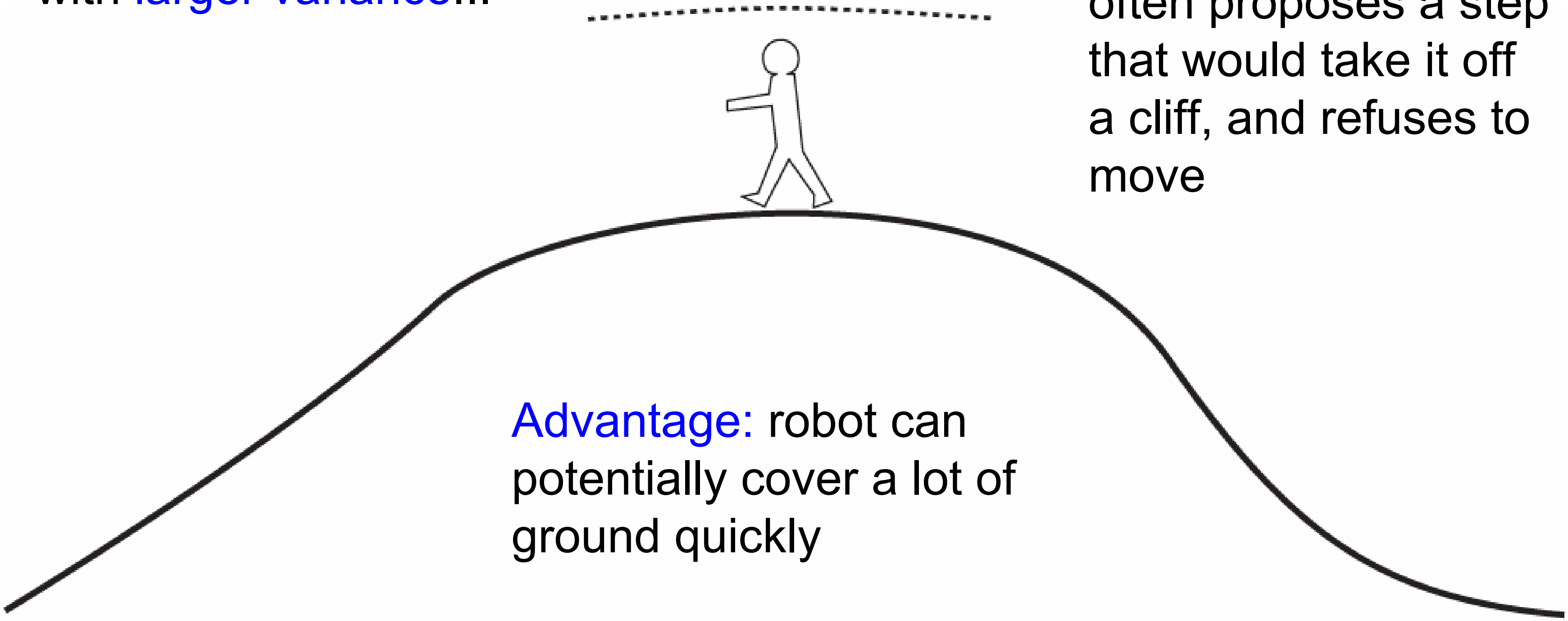
Proposal distributions with smaller variance...

Disadvantage: robot takes smaller steps, more time required to explore the same area

Advantage: robot seldom refuses to take proposed steps

Target vs. Proposal Distributions

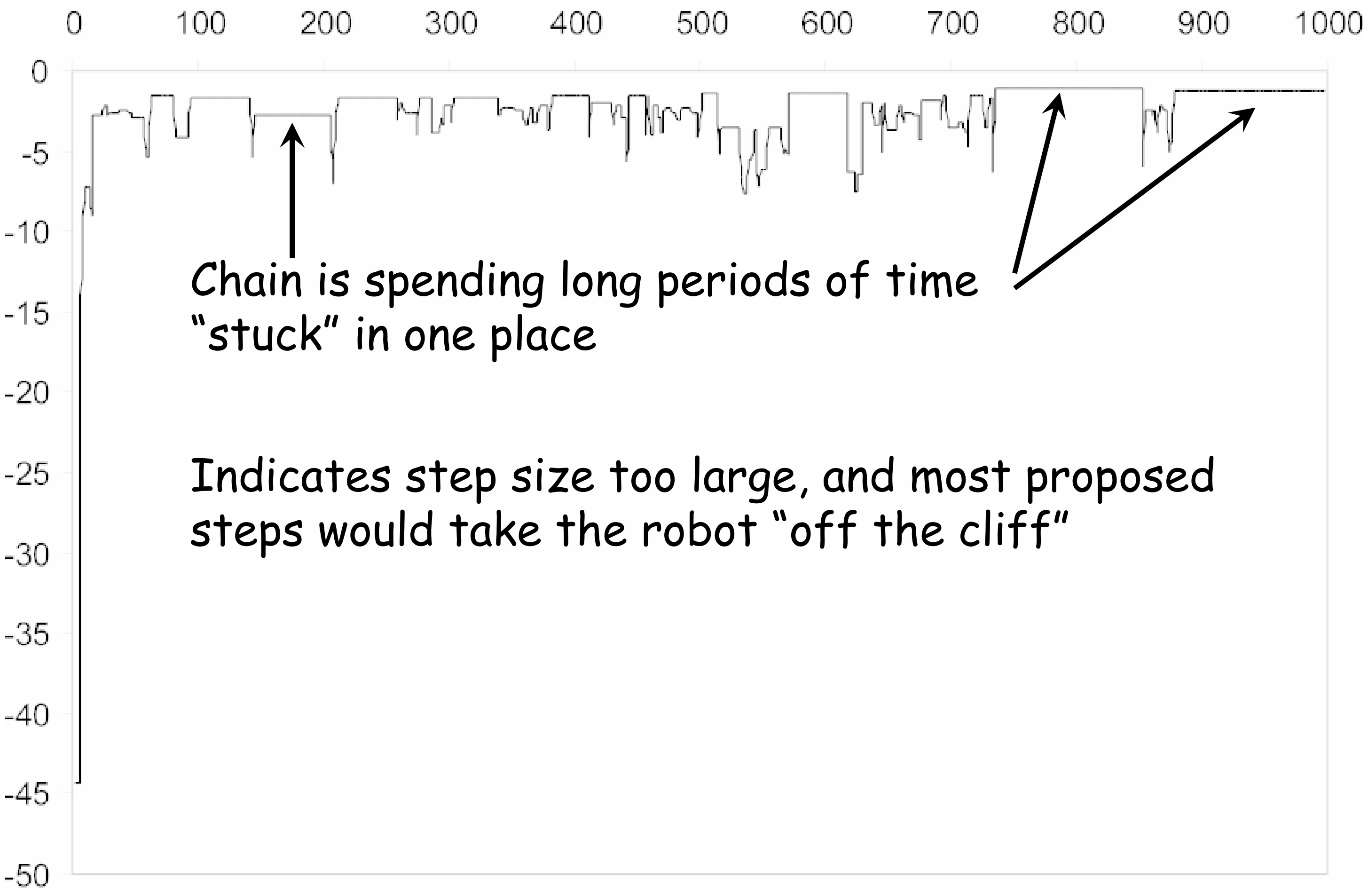
Proposal distributions with larger variance...



Disadvantage: robot often proposes a step that would take it off a cliff, and refuses to move

Advantage: robot can potentially cover a lot of ground quickly

Poor mixing

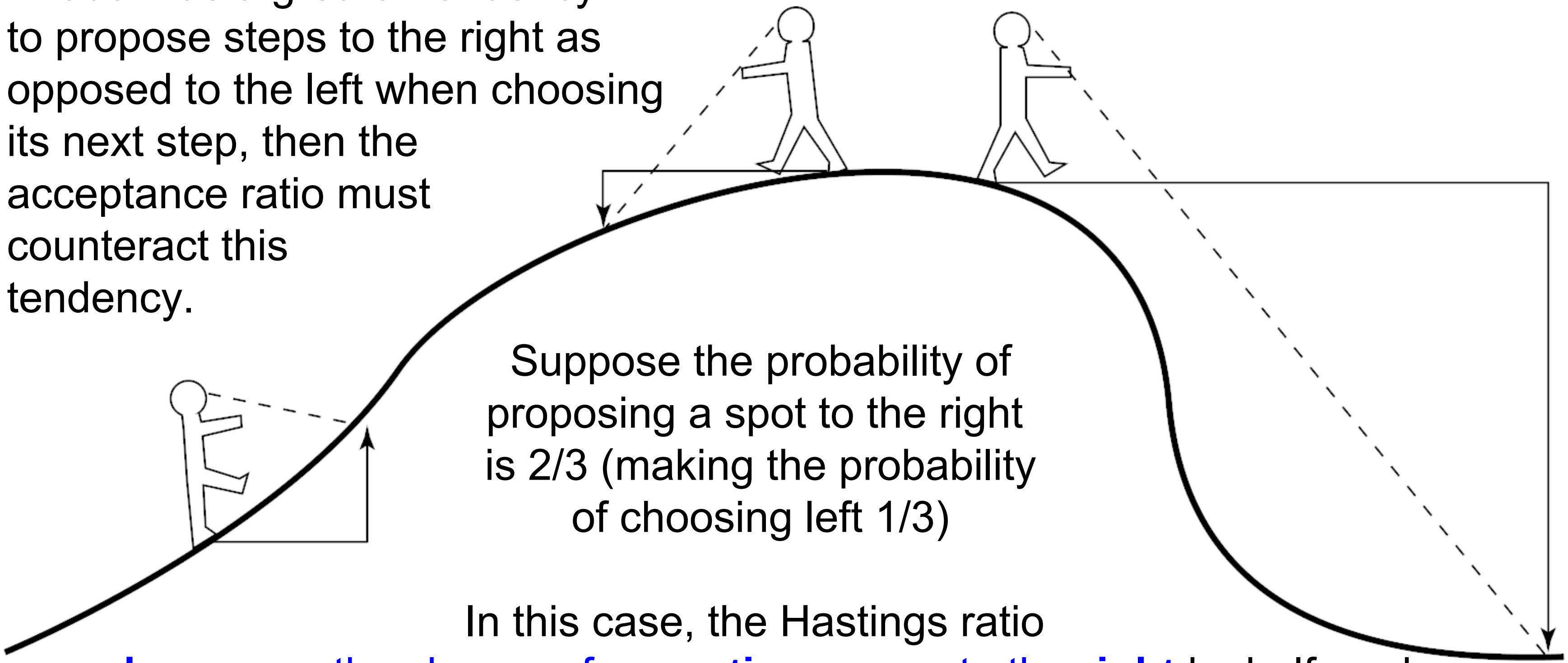


Chain is spending long periods of time "stuck" in one place

Indicates step size too large, and most proposed steps would take the robot "off the cliff"

The Hastings ratio

If robot has a greater tendency to propose steps to the right as opposed to the left when choosing its next step, then the acceptance ratio must counteract this tendency.



Suppose the probability of proposing a spot to the right is 2/3 (making the probability of choosing left 1/3)

In this case, the Hastings ratio **decreases** the chance of **accepting** moves to the **right** by half, and **increases** the chance of **accepting** moves to the **left** (by a factor of 2), thus **exactly compensating** for the asymmetry in the proposal distribution.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97-109.

MCRobot

Windows program download from:
http://www.eeb.uconn.edu/people/plewis/software.php

Metropolis-coupled Markov chain Monte Carlo (MCMCMC, or MC^3)

- MC^3 involves running **several chains simultaneously**
- The **cold chain** is the one that counts, the rest are **heated chains**
- Chain is heated by raising densities to a power less than 1.0 (values closer to 0.0 are warmer)

Geyer, C. J. 1991. Markov chain Monte Carlo maximum likelihood for dependent data. Pages 156-163 *in* Computing Science and Statistics (E. Keramidas, ed.).

What is a heated chain?

R is the ratio of posterior probability densities.

Instead of using R in the acceptance/rejection decisions, a heated chain uses $R^{\frac{1}{1+H}}$

Heating a chain makes the surface it explores **flatter**.

In MrBayes: $H =$ “Temperature” $*$(The Chain’s index)
The cold chain has index 0, and the default temperature is 0.2

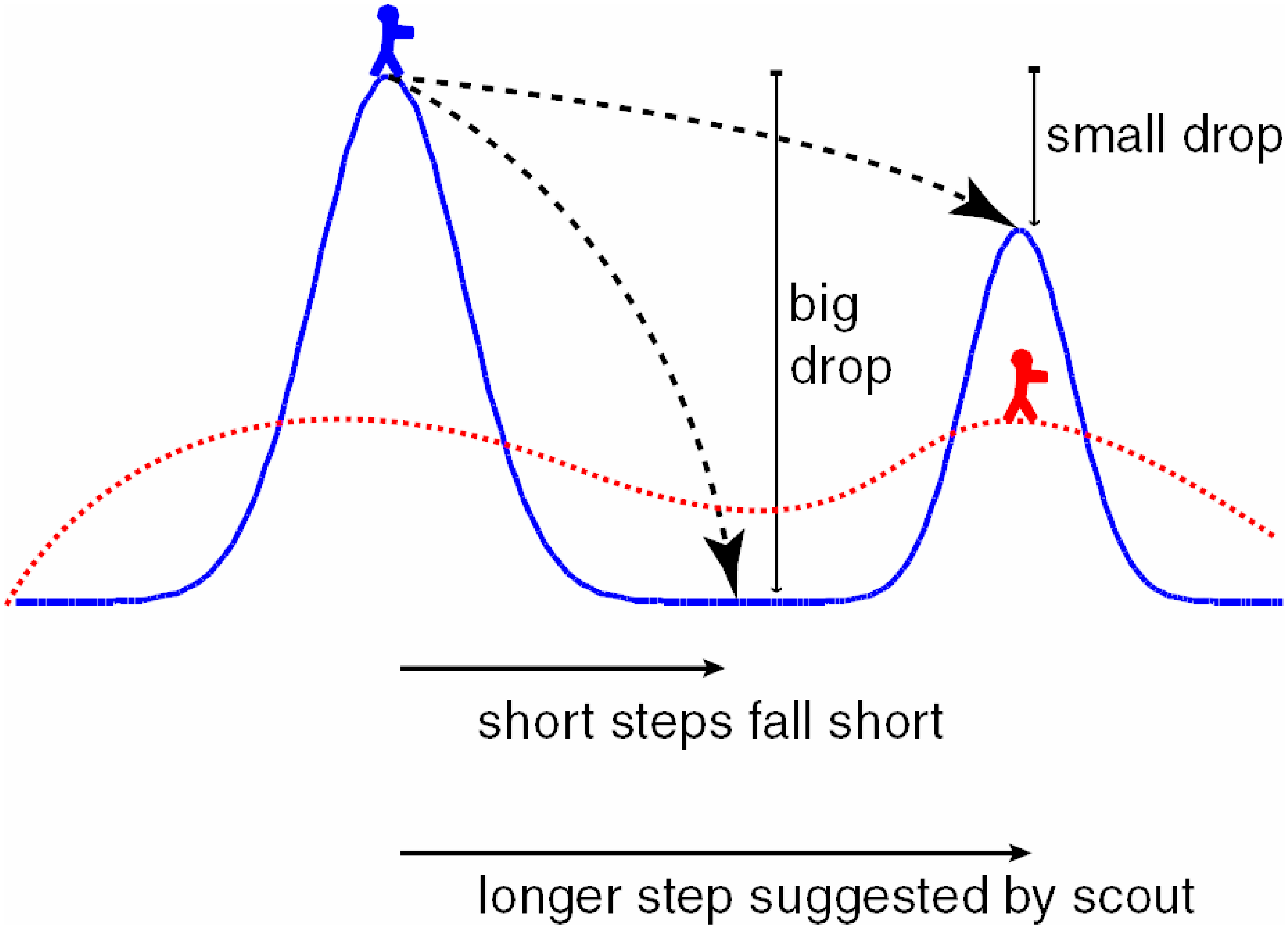
Acceptance Probability for chains with Temp = 0.2

	Chain			
R	1	2	3	4
1.2	1.0000	1.0000	1.0000	1.0000
0.8	0.8000	0.8303	0.8527	0.8600
0.4	0.4000	0.4660	0.5197	0.5640
0.01	0.0100	0.0215	0.0373	0.0562

Acceptance Probability for chains with Temp = 0.5

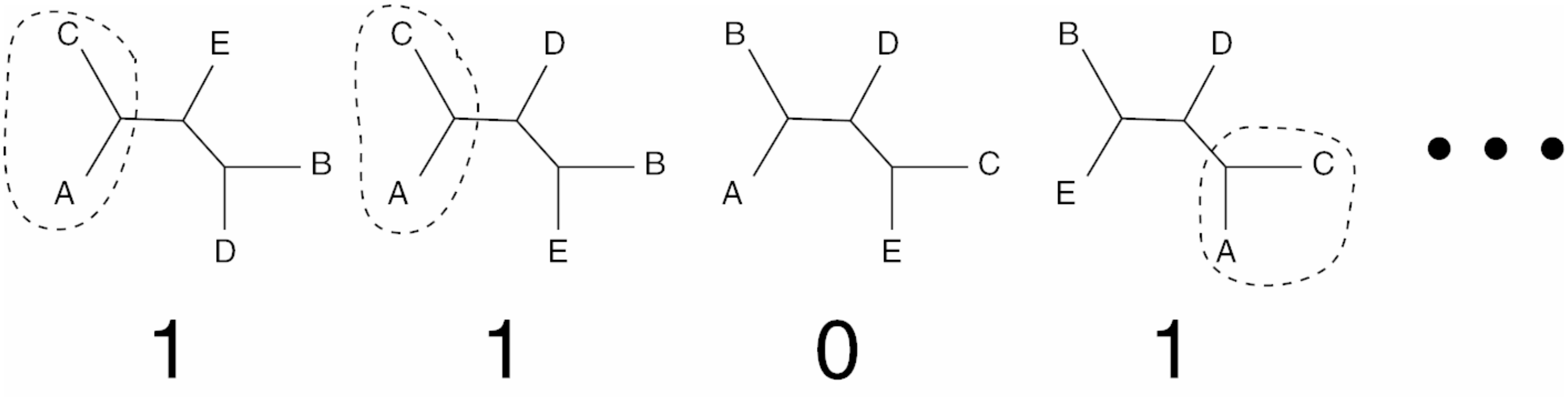
	Chain			
R	1	2	3	4
0.8	0.8000	0.8618	0.8944	0.9146
0.4	0.4000	0.5429	0.6325	0.6931
0.01	0.0100	0.0464	0.1000	0.1585

Heated chains act as scouts for the cold chain



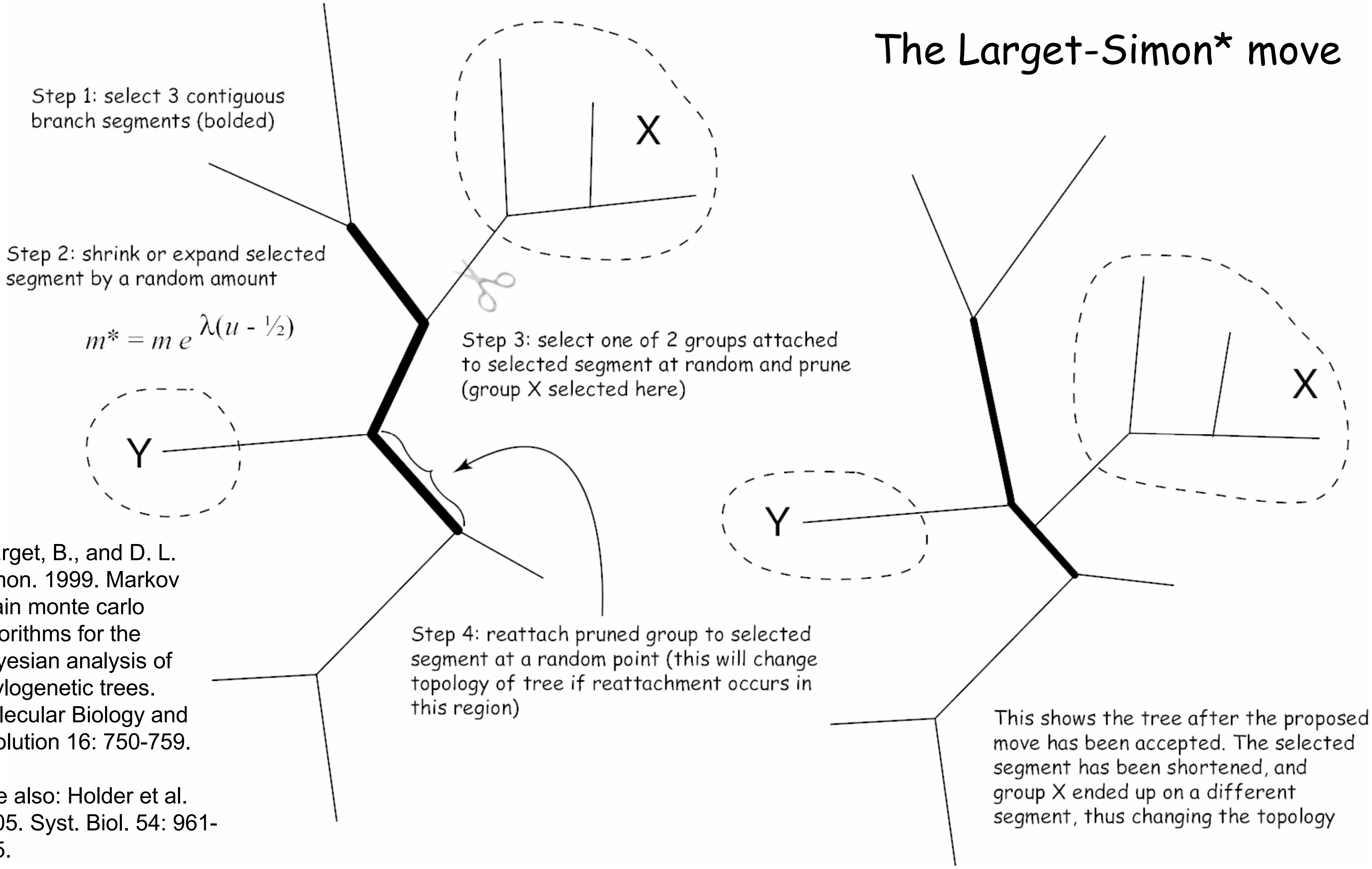
(the following slides come directly from Paul Lewis' lecture at the Woods Hole Workshop on Molecular Evolution – thanks, Paul).

So, what's all this got to do with phylogenetics?



Imagine drawing tree topologies randomly from a bin in which the number of copies of any given topology is proportional to the (marginal) posterior probability of that topology. Approximating the posterior of any particular attribute of tree topologies (e.g. existence of group AC in this case) is simply a matter of counting.

Moving through treespace



The Larget-Simon* move

Step 1: select 3 contiguous branch segments (bolded)

Step 2: shrink or expand selected segment by a random amount

$$m^* = m\, e^{\lambda(u - ½)}$$

Step 3: select one of 2 groups attached to selected segment at random and prune (group X selected here)

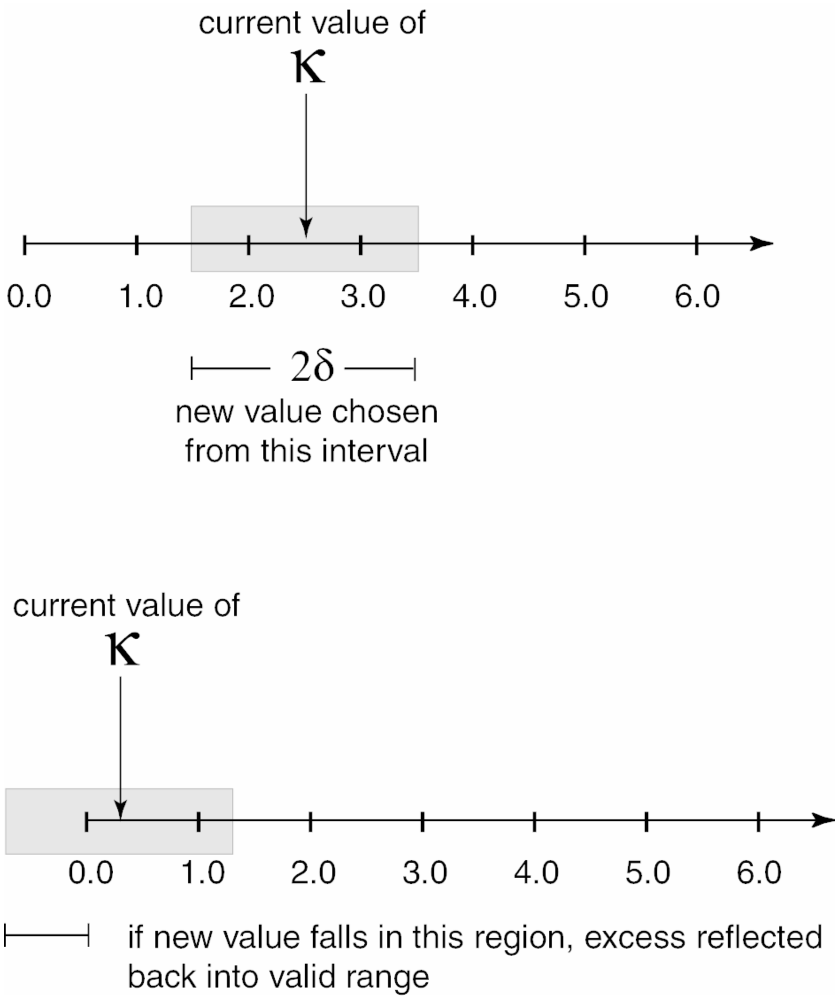
Step 4: reattach pruned group to selected segment at a random point (this will change topology of tree if reattachment occurs in this region)

This shows the tree after the proposed move has been accepted. The selected segment has been shortened, and group X ended up on a different segment, thus changing the topology

*Larget, B., and D. L. Simon. 1999. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular Biology and Evolution 16: 750-759.

See also: Holder et al. 2005. Syst. Biol. 54: 961-965.

Moving through parameter space



Using κ (ratio of the transition rate to the transversion rate) as an example of a model parameter.

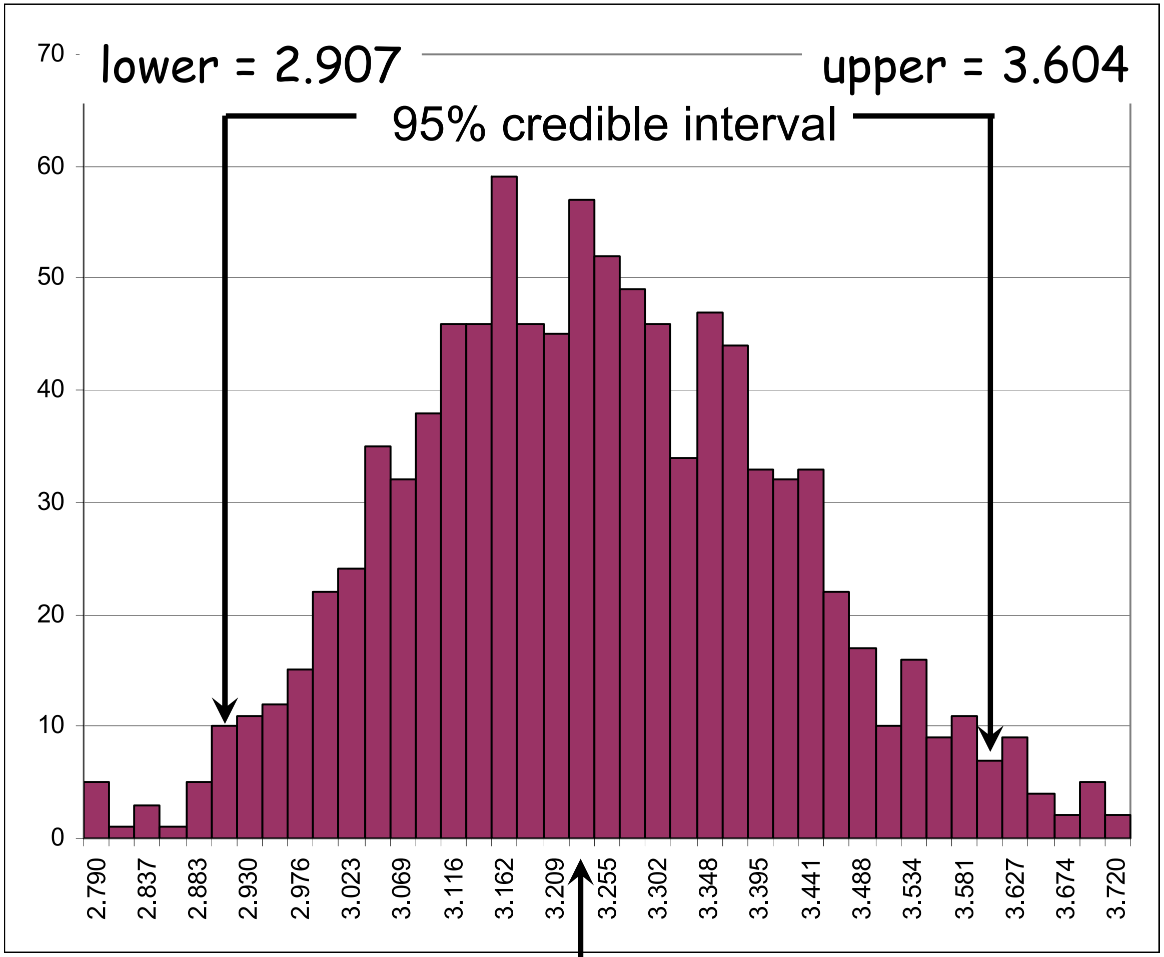
Proposal distribution is uniform from κ-δ to κ+δ

The "step size" of the MCMC robot is defined by δ: a larger δ means that the robot will attempt to make larger jumps on average.

Putting it all together

- **Start with** random tree and arbitrary initial values for branch lengths and model parameters
- **Each generation** consists of one of these (chosen at random):
  - Propose a **new tree** (e.g. Larget-Simon move) and either accept or reject the move
  - Propose (and either accept or reject) a **new model parameter value**
- Every *k* generations, save tree topology, branch lengths and all model parameters (i.e. **sample the chain**)
- After *n* generations, **summarize sample** using histograms, means, credible intervals, etc.

Marginal posterior distributions



lower = 2.907

upper = 3.604

95% credible interval

mean = 3.234

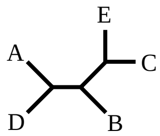
Histogram created from a sample of 1000 κ values.

From: Lewis, L., and Flechtner, V. 2002. Taxon 51: 443-451.

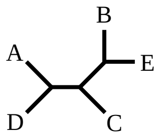
IV. Prior distributions

Commonly-used Prior Distributions

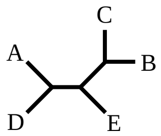
- For **topologies**: discrete Uniform distribution



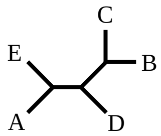
$\frac{1}{15}$



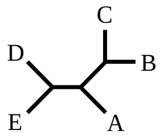
$\frac{1}{15}$



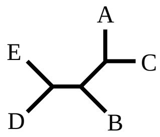
$\frac{1}{15}$



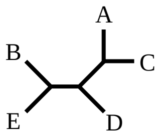
$\frac{1}{15}$



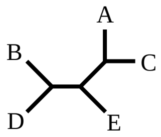
$\frac{1}{15}$



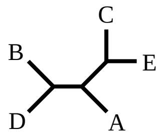
$\frac{1}{15}$



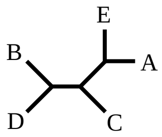
$\frac{1}{15}$



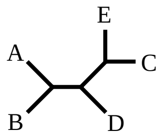
$\frac{1}{15}$



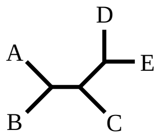
$\frac{1}{15}$



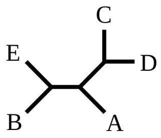
$\frac{1}{15}$



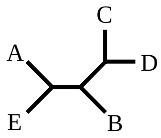
$\frac{1}{15}$



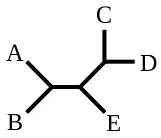
$\frac{1}{15}$



$\frac{1}{15}$



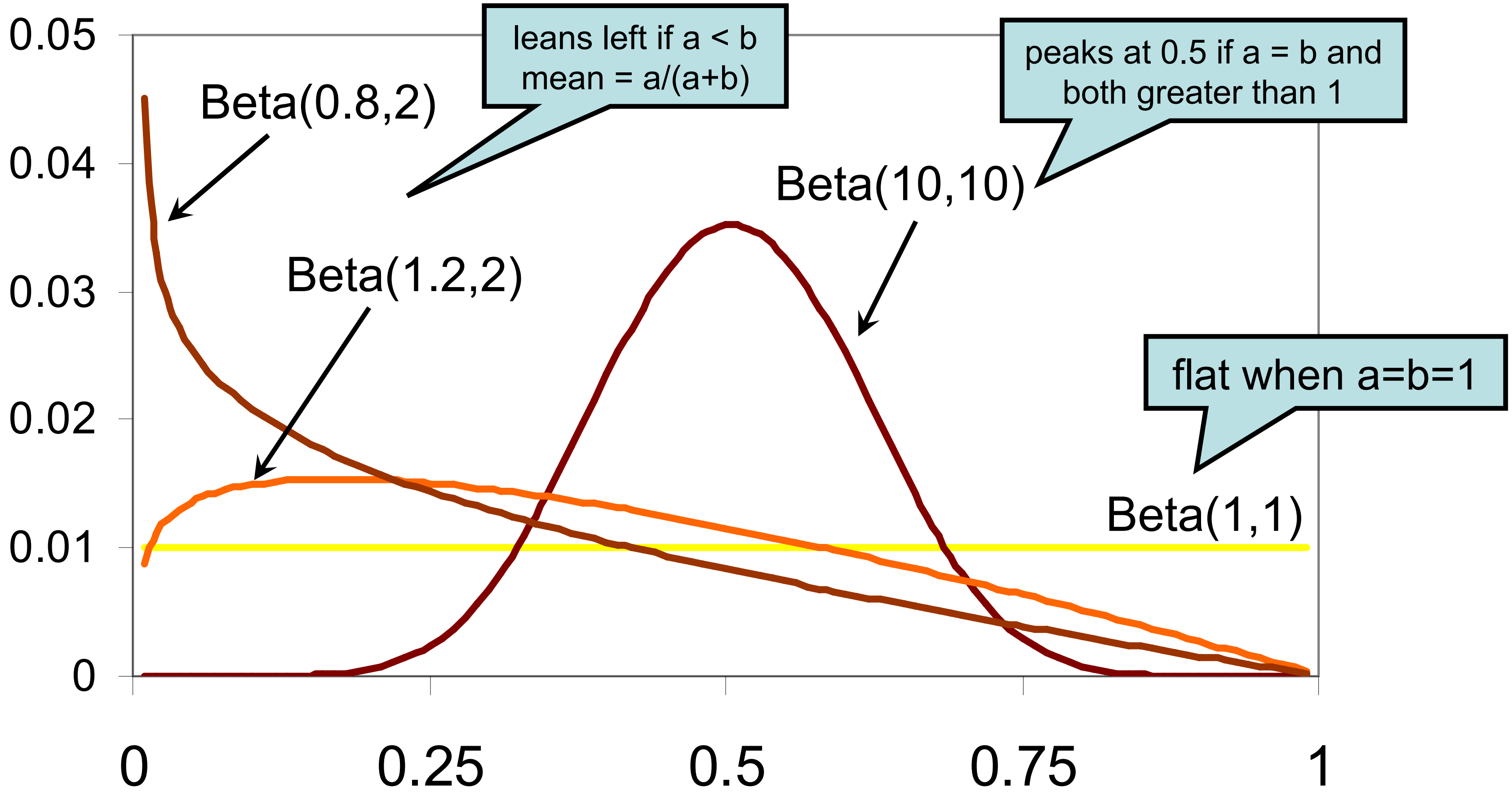
$\frac{1}{15}$



$\frac{1}{15}$

Commonly-used Prior Distributions

- For **proportions (e.g. pinvar)**: Beta(a,b) distribution



leans left if a < b
mean = a/(a+b)

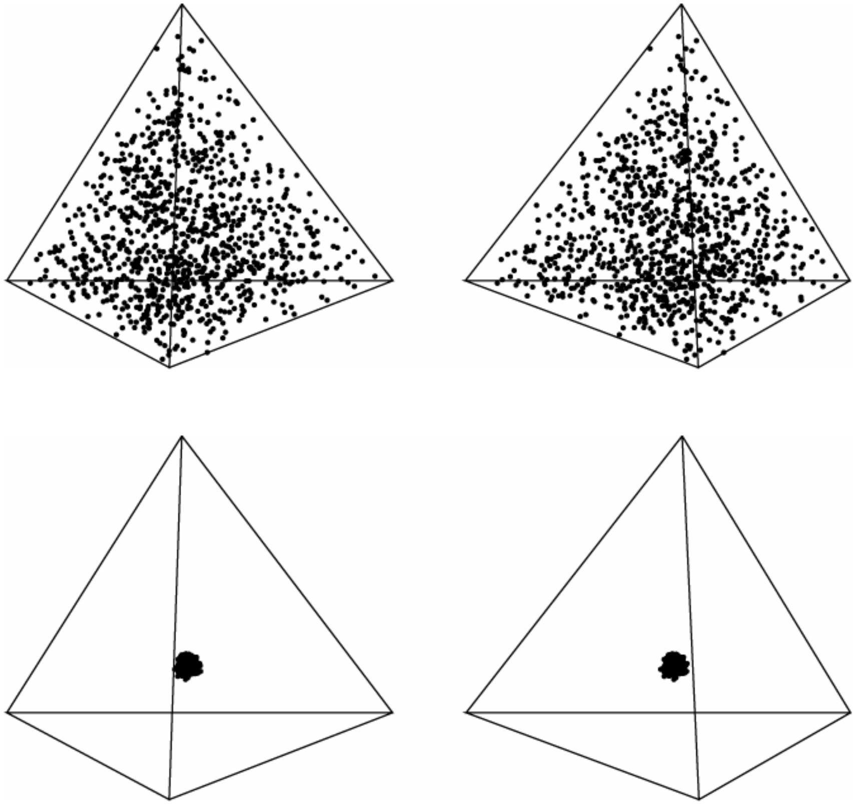
peaks at 0.5 if a = b and both greater than 1

flat when a=b=1

Commonly-used Prior Distributions

- For **base frequencies**: Dirichlet(a,b,c,d) distribution

$a\rightarrow\pi_A$, $b\rightarrow\pi_C$, $c\rightarrow\pi_G$, $d\rightarrow\pi_T$



Flat prior:

$a = b = c = d = 1$

Informative prior:

$a = b = c = d = 300$

(stereo pairs)

(Thanks to Mark Holder for pointing out to me that a tetrahedron could be used for plotting a 4-dimensional Dirichlet)

Commonly-used Prior Distributions

- For **GTR model relative rates**: Dirichlet(a,b,c,d,e,f) distribution
  - ➢ a→r_{AC}, b→r_{AG}, c→r_{AT}, d→r_{CG}, e→r_{CT}, f→r_{GT}
  - ➢ flat when a=b=c=d=e=f=1
  - ➢ all relative rates nearly equal to each other if a=b=c=d=e=f and large (e.g. 300)
  - ➢ to create a vague prior that makes the rate of transitions slightly higher than the rate of transversions, could choose a=c=d=f=1 and b=e=2
  - ➢ mean for r_{AC} is a/s where s=a+b+c+d+e+f
  - ➢ variance for r_{AC} is $a(s-a)/[s^2(s+1)]$
  - ➢ Beta(a,b) equals Dirichlet(a,b)

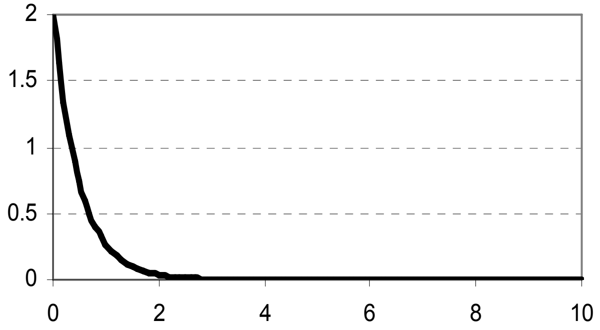
Common Priors (cont.)

- For other **model parameters** and **branch lengths**: Gamma(a,b) distribution
  - Exponential(λ) equals Gamma(1, λ^{-1})
  - Mean of Gamma(a,b) is a×b
    - mean of an Exponential(10) distribution is 0.1
  - Variance of a Gamma(a,b) distribution is $a \times b^2$
    - variance of an Exponential(10) distribution is 0.01

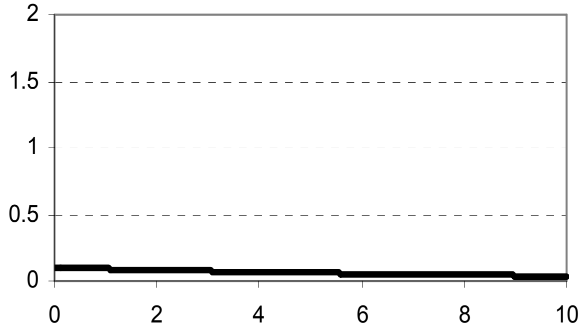
Note: be aware that in many papers the Gamma distribution is defined such that the second (scale) parameter is the *inverse* of the value b used in this slide! In this case, the mean and variance would be a/b and a/b^2, respectively.

Priors for model parameters with no upper bound

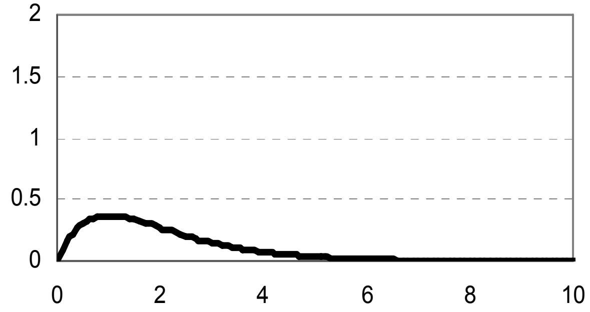
Exponential(2) = Gamma(1,½)



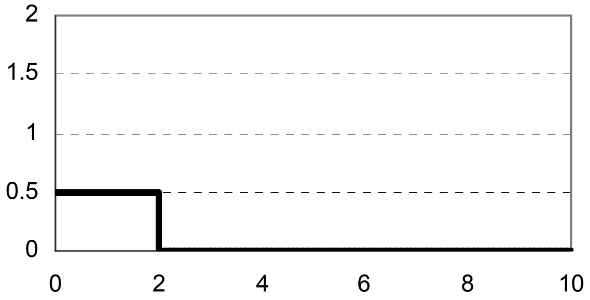
Exponential(0.1) = Gamma(1,10)



Gamma(2,1)



Uniform(0,2)

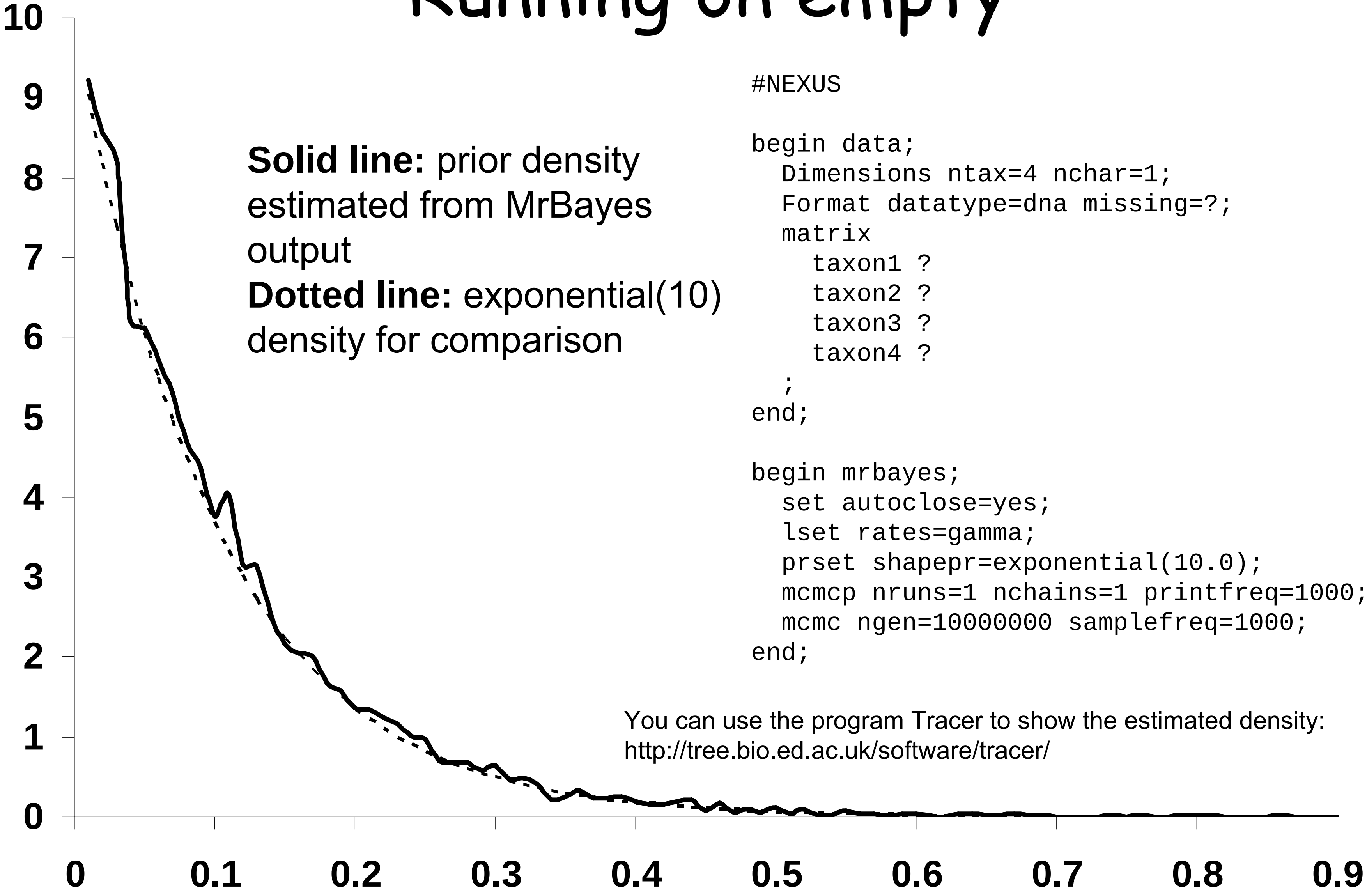


See chapter 18 in Felsenstein, J. (2004. Inferring Phylogenies.Sinauer) before using.

More About Priors

- Running on empty
- Prior as enemy
- Prior as friend
- Flat vs. informative priors
- Proper vs. improper priors
- Hierarchical models
- Empirical Bayes

Running on empty



Solid line: prior density estimated from MrBayes output
Dotted line: exponential(10) density for comparison

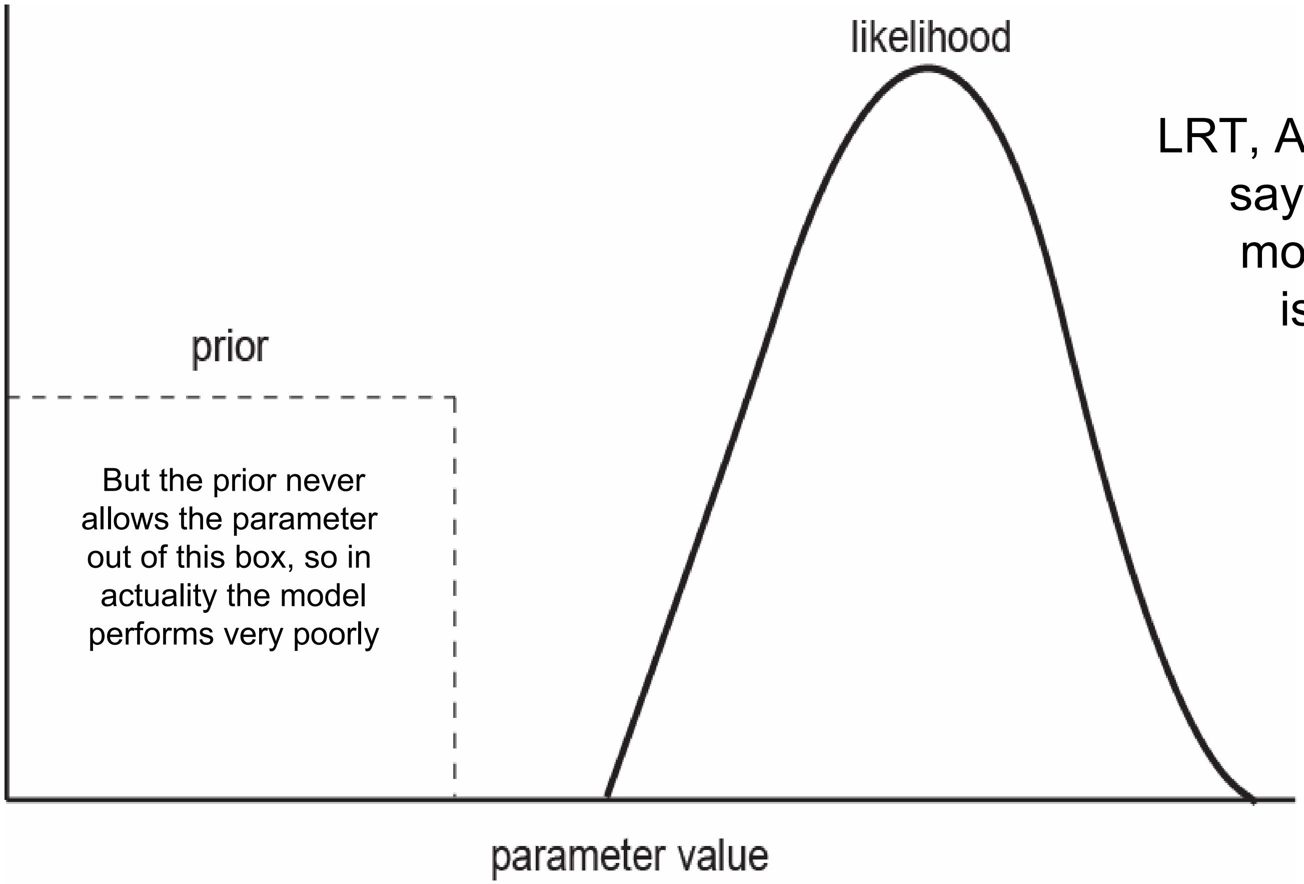
```
#NEXUS

begin data;
  Dimensions ntax=4 nchar=1;
  Format datatype=dna missing=?;
  matrix
    taxon1 ?
    taxon2 ?
    taxon3 ?
    taxon4 ?
  ;
end;

begin mrbayes;
  set autoclose=yes;
  lset rates=gamma;
  prset shapepr=exponential(10.0);
  mcmcp nruns=1 nchains=1 printfreq=1000;
  mcmc ngen=10000000 samplefreq=1000;
end;
```

You can use the program Tracer to show the estimated density:
http://tree.bio.ed.ac.uk/software/tracer/

More About Priors

- Running on empty
- Prior as enemy
- Prior as friend
- Flat vs. informative priors
- Proper vs. improper priors
- Hierarchical models
- Empirical Bayes

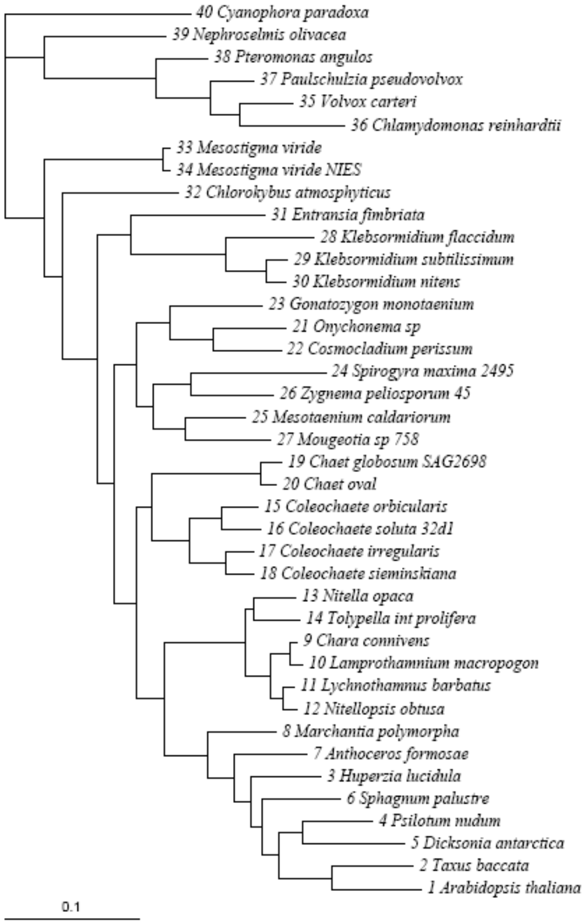
The choice of prior distributions can potentially turn a good model bad!



But the prior never allows the parameter out of this box, so in actuality the model performs very poorly

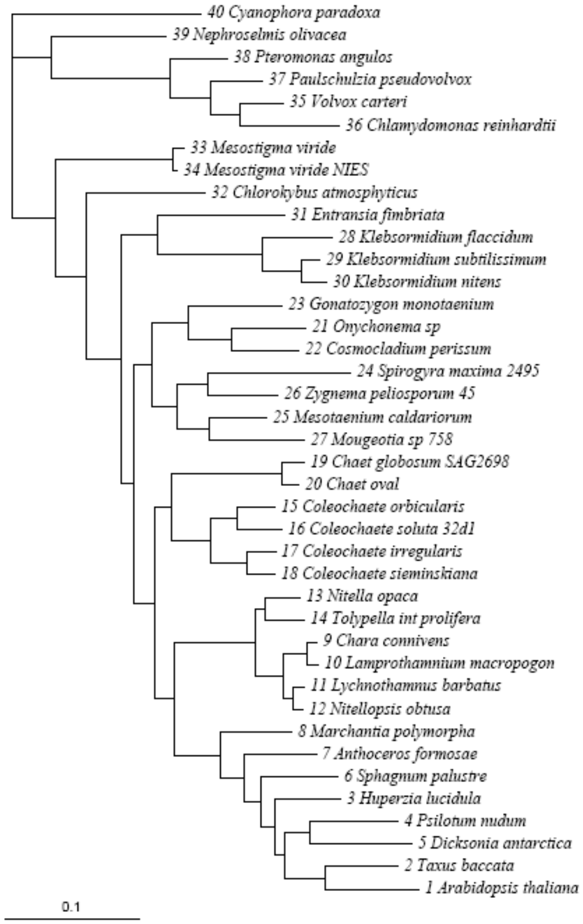
LRT, AIC and BIC all say this is a great model because it is able to attain such a high maximum likelihood score

Internal branch length prior mean 0.1



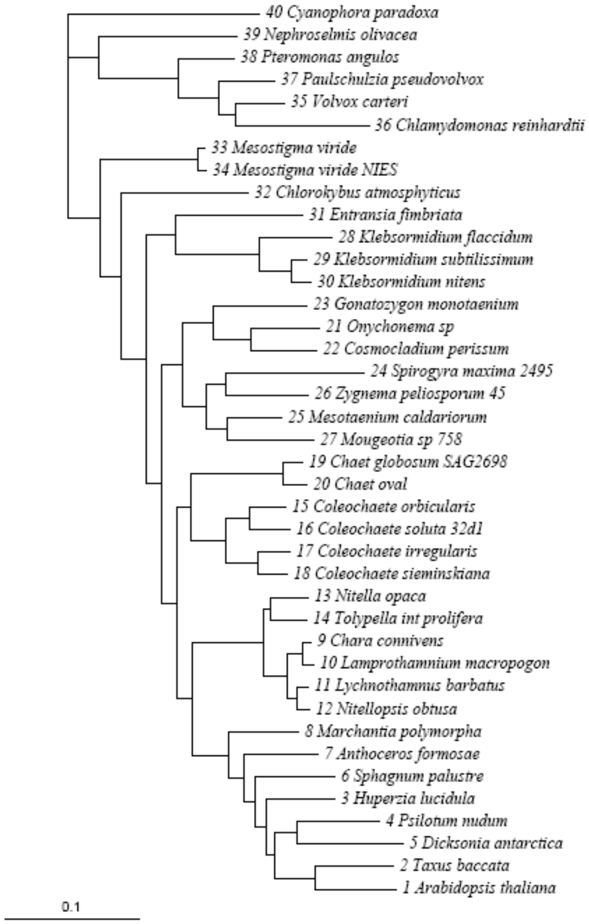
This is a reasonably vague internal branch length prior

Internal branch length prior mean 0.01



Not much effect yet...

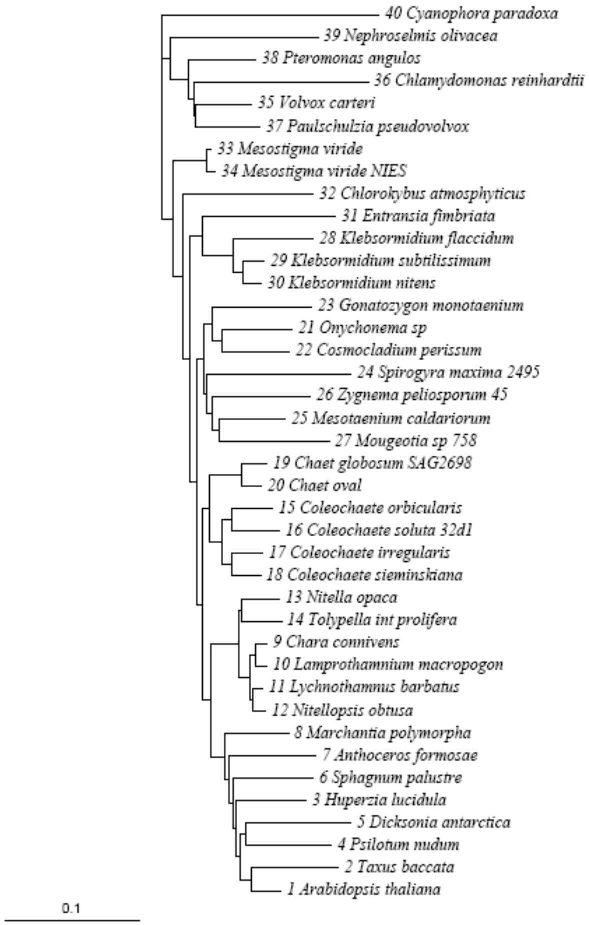
Internal branch length prior mean 0.001



Notice how the internal branch lengths are shrinking...

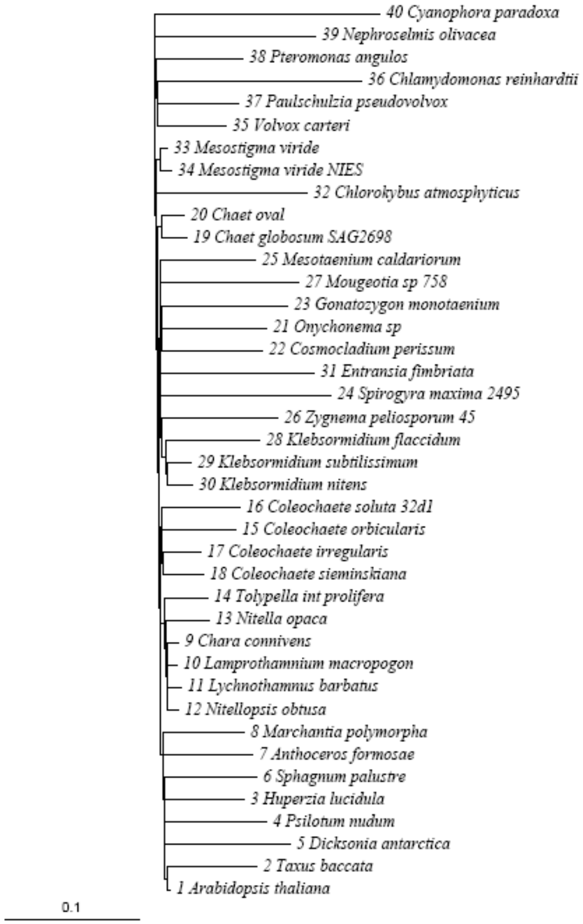
(Trees in this series are drawn to same scale)

Internal branch length prior mean 0.0001



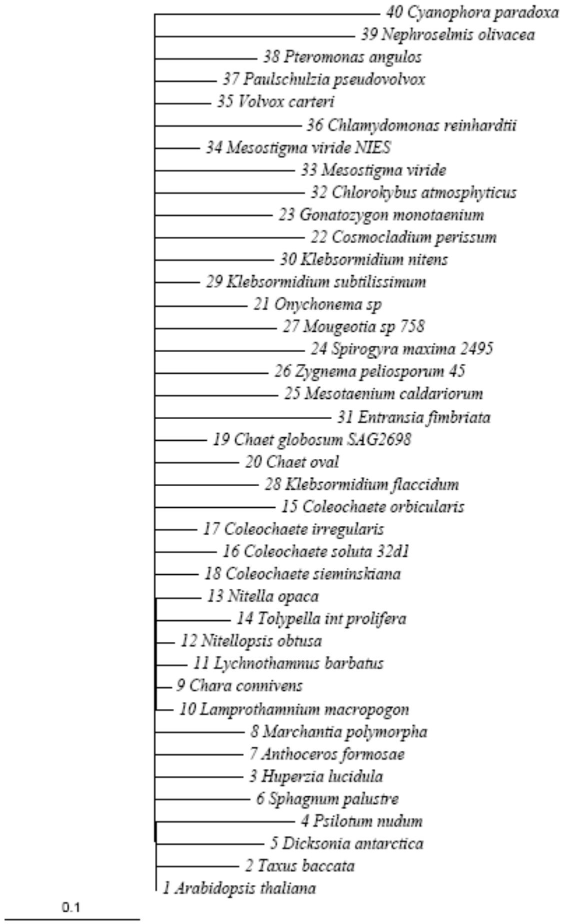
Model compensating for small internal branch lengths by increasing the external branch lengths...

Internal branch length prior mean 0.00001



Internal branch length prior now so informative that it is beginning to noticeably override the likelihood...

Internal branch length prior mean 0.000001



The internal branch length prior is calling the shots now.

More About Priors

- Running on empty
- Prior as enemy
- Prior as friend
- Flat vs. informative priors
- Proper vs. improper priors
- Hierarchical models
- Empirical Bayes

Too many parameters, too little information

H H T T T T H

7 coins flipped once

3/7 = 0.43

1 parameter model behaves well

1.0 1.0 0.0 0.0 0.0 0.0 1.0

7 parameter model behaves badly

Under maximum likelihood, parameter values tend to go to extremes if there is too little information.

Priors *add information* and can keep models in check

More About Priors

- Running on empty
- Prior as enemy
- Prior as friend
- Flat vs. informative priors
- Proper vs. improper priors
- Hierarchical models
- Empirical Bayes

Flat prior: posterior proportional to likelihood

$$
\begin{aligned}
\text{posterior} \longrightarrow f(\theta|D) &= \frac{f(D|\theta)f(\theta)}{f(D) \longleftarrow \text{constant}} \\
&\propto f(D|\theta)\enclose{circle}{f(\theta)} \longleftarrow \text{constant} \\
&\propto f(D|\theta) \longleftarrow \text{likelihood}
\end{aligned}
$$

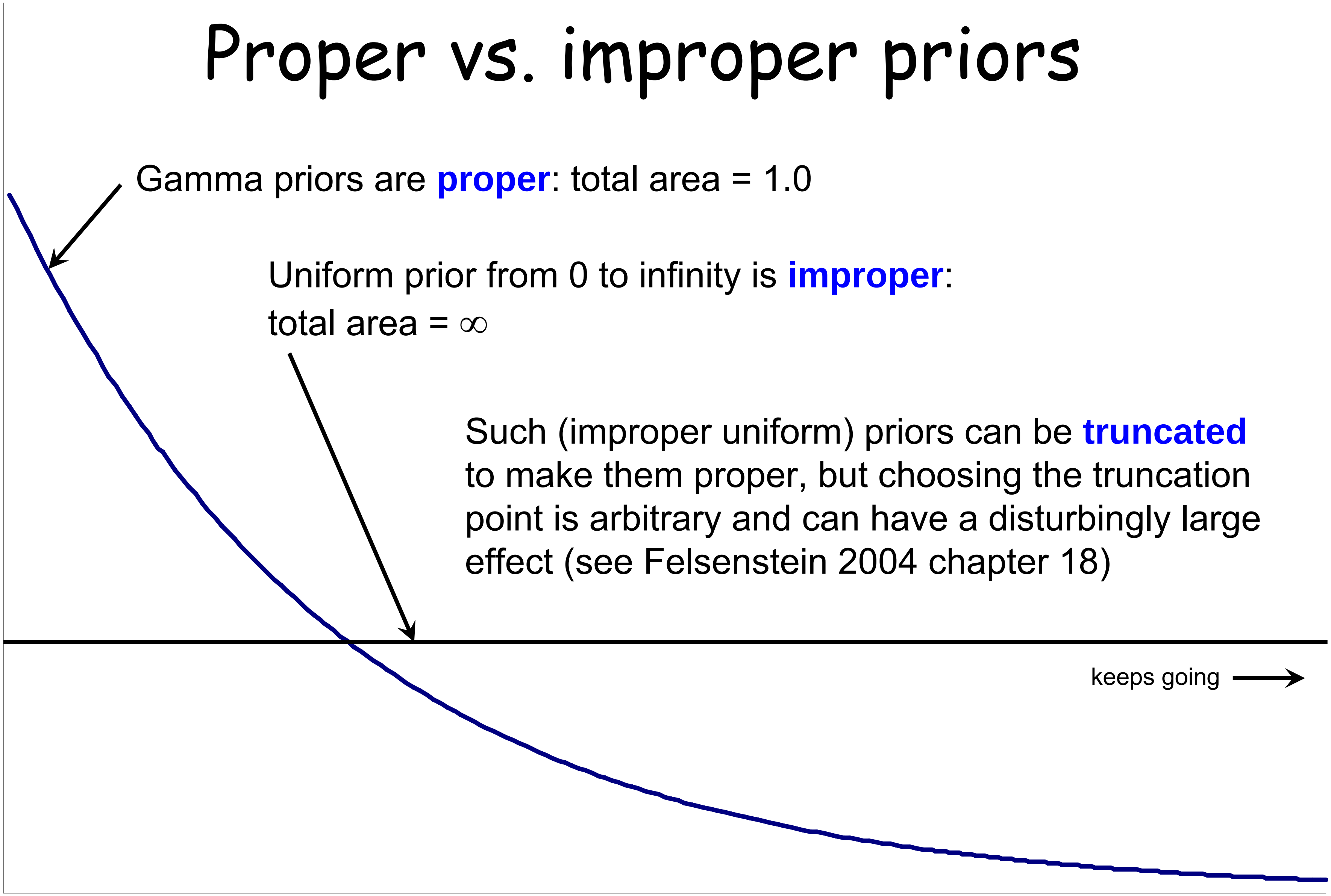
Under a flat prior, the posterior distribution *peaks* at the same place as the likelihood function, but:

- posterior mean usually differs from the maximum likelihood estimate
- flat priors are not possible for most parameters

More About Priors

- Running on empty
- Prior as enemy
- Prior as friend
- Flat vs. informative priors
- Proper vs. improper priors
- Hierarchical models
- Empirical Bayes

Proper vs. improper priors



Gamma priors are **proper**: total area = 1.0

Uniform prior from 0 to infinity is **improper**:
total area = ∞

Such (improper uniform) priors can be **truncated** to make them proper, but choosing the truncation point is arbitrary and can have a disturbingly large effect (see Felsenstein 2004 chapter 18)

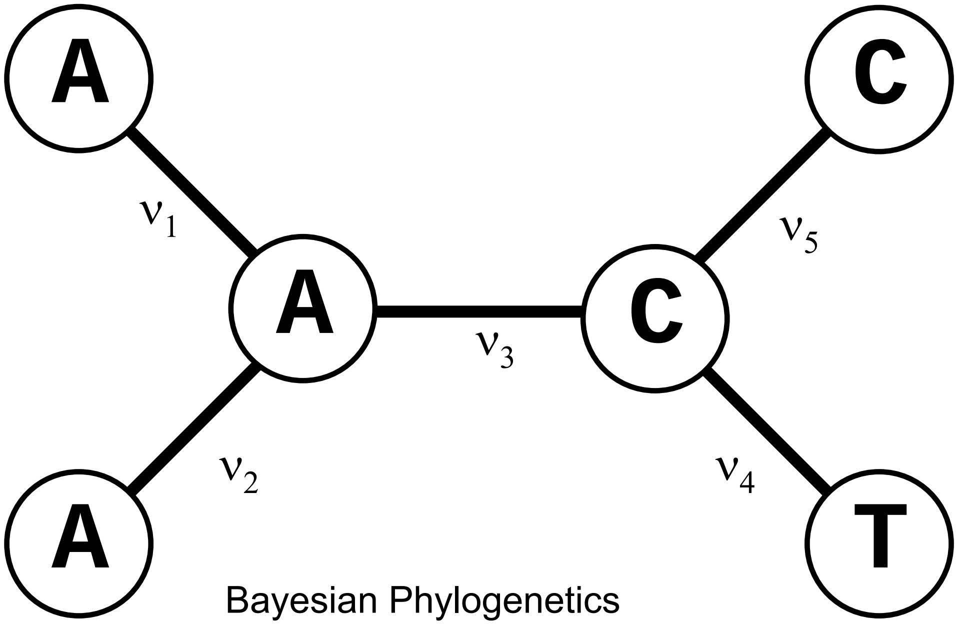
More About Priors

- Running on empty
- Prior as enemy
- Prior as friend
- Flat vs. informative priors
- Proper vs. improper priors
- Hierarchical models
- Empirical Bayes

In a non-hierarchical model, all parameters are present in the likelihood function

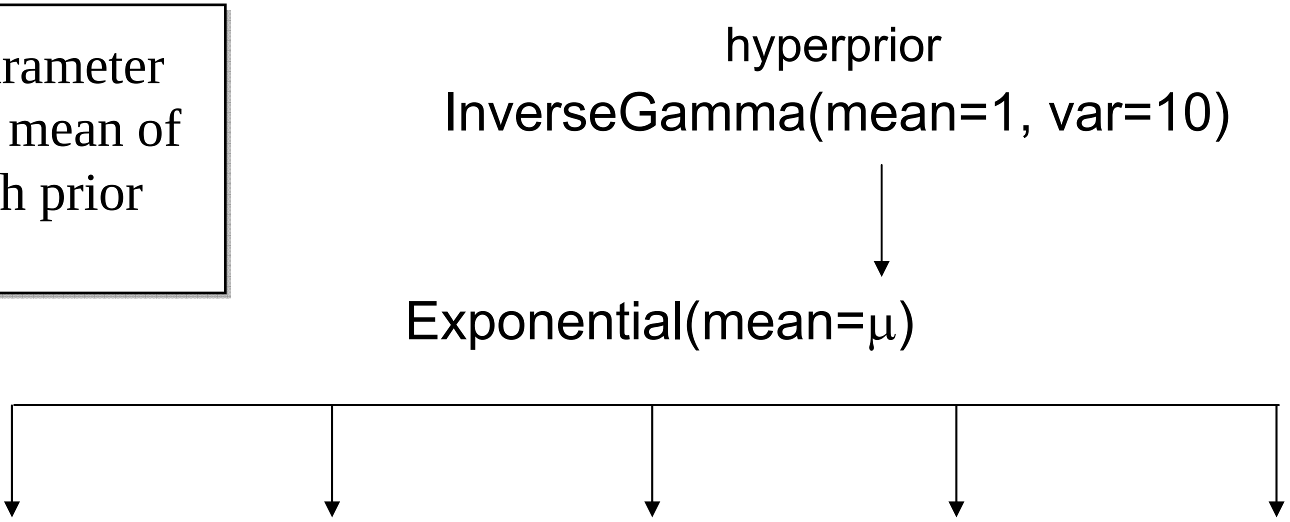
Exponential(mean=0.1)

$$L_k = \tfrac{1}{4}\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_1/3}\right]\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_2/3}\right]\left[\tfrac{1}{4}-\tfrac{1}{4}e^{-4\nu_3/3}\right]\left[\tfrac{1}{4}-\tfrac{1}{4}e^{-4\nu_4/3}\right]\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_5/3}\right]$$



Hierarchical models add **hyper**parameters not present in the likelihood function

μ is a hyperparameter governing the mean of the edge length prior



hyperprior

InverseGamma(mean=1, var=10)

Exponential(mean=μ)

$$L_k = \tfrac{1}{4}\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_1/3}\right]\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_2/3}\right]\left[\tfrac{1}{4}-\tfrac{1}{4}e^{-4\nu_3/3}\right]\left[\tfrac{1}{4}-\tfrac{1}{4}e^{-4\nu_4/3}\right]\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_5/3}\right]$$

For example, see Suchard, Weiss and Sinsheimer. 2001. MBE 18(6): 1001-1013.

Empirical Bayes

This uses some aspects of the data to determine some aspects of the prior, which is not acceptable to purists, who prefer using the hierarchical approach.

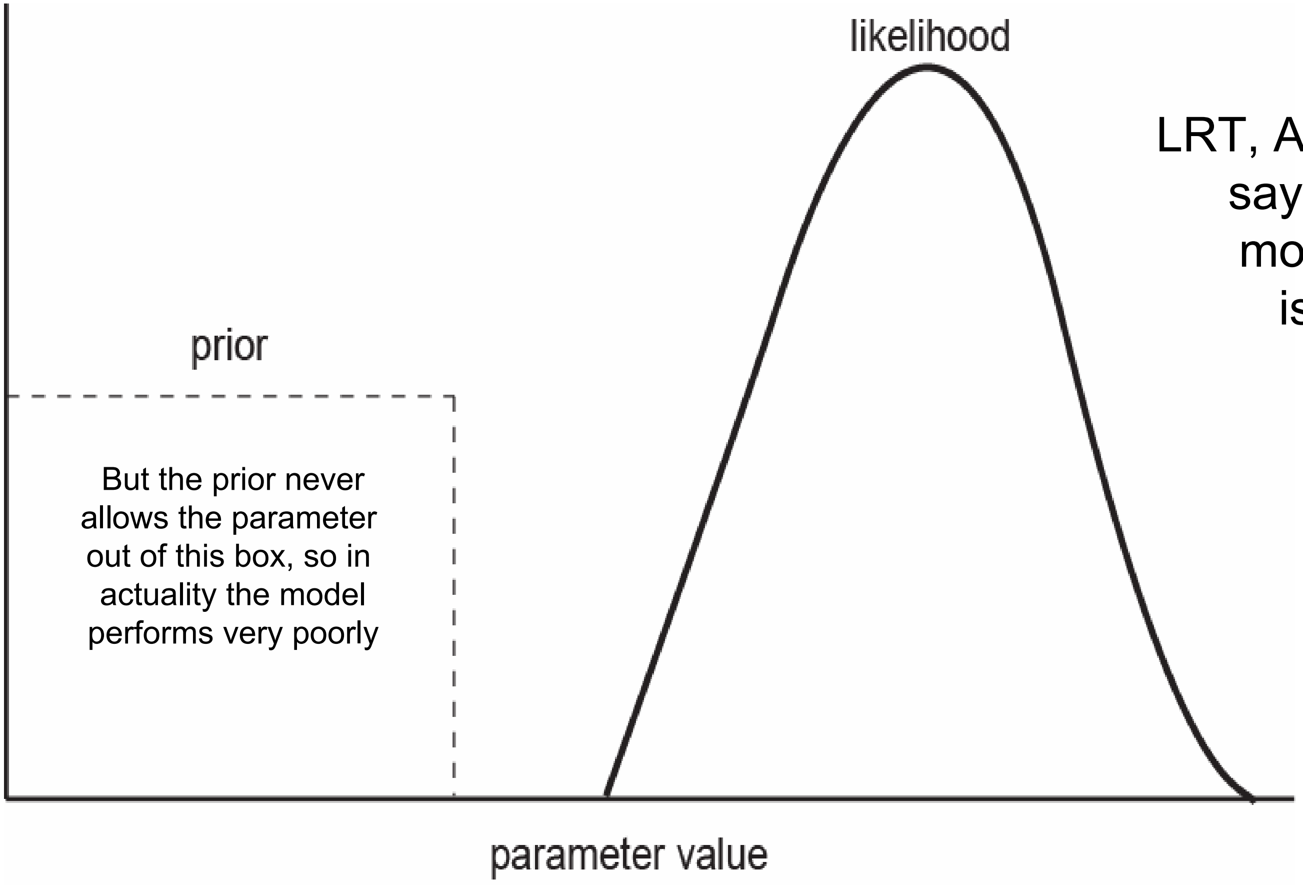
An empirical Bayesian would use the maximum likelihood estimate (MLE) of the length of an average branch here

Exponential(mean=MLE)

$$L_k = \tfrac{1}{4}\left[\tfrac{1}{4} + \tfrac{3}{4} e^{-4\nu_1/3}\right]\left[\tfrac{1}{4} + \tfrac{3}{4} e^{-4\nu_2/3}\right]\left[\tfrac{1}{4} - \tfrac{1}{4} e^{-4\nu_3/3}\right]\left[\tfrac{1}{4} - \tfrac{1}{4} e^{-4\nu_4/3}\right]\left[\tfrac{1}{4} + \tfrac{3}{4} e^{-4\nu_5/3}\right]$$

V. Bayesian model selection

LRT, AIC and BIC only evaluate *part* of a Bayesian model (i.e. the likelihood)



But the prior never allows the parameter out of this box, so in actuality the model performs very poorly

LRT, AIC and BIC all say this is a great model because it is able to attain such a high maximum likelihood score

Marginal probabilities of models

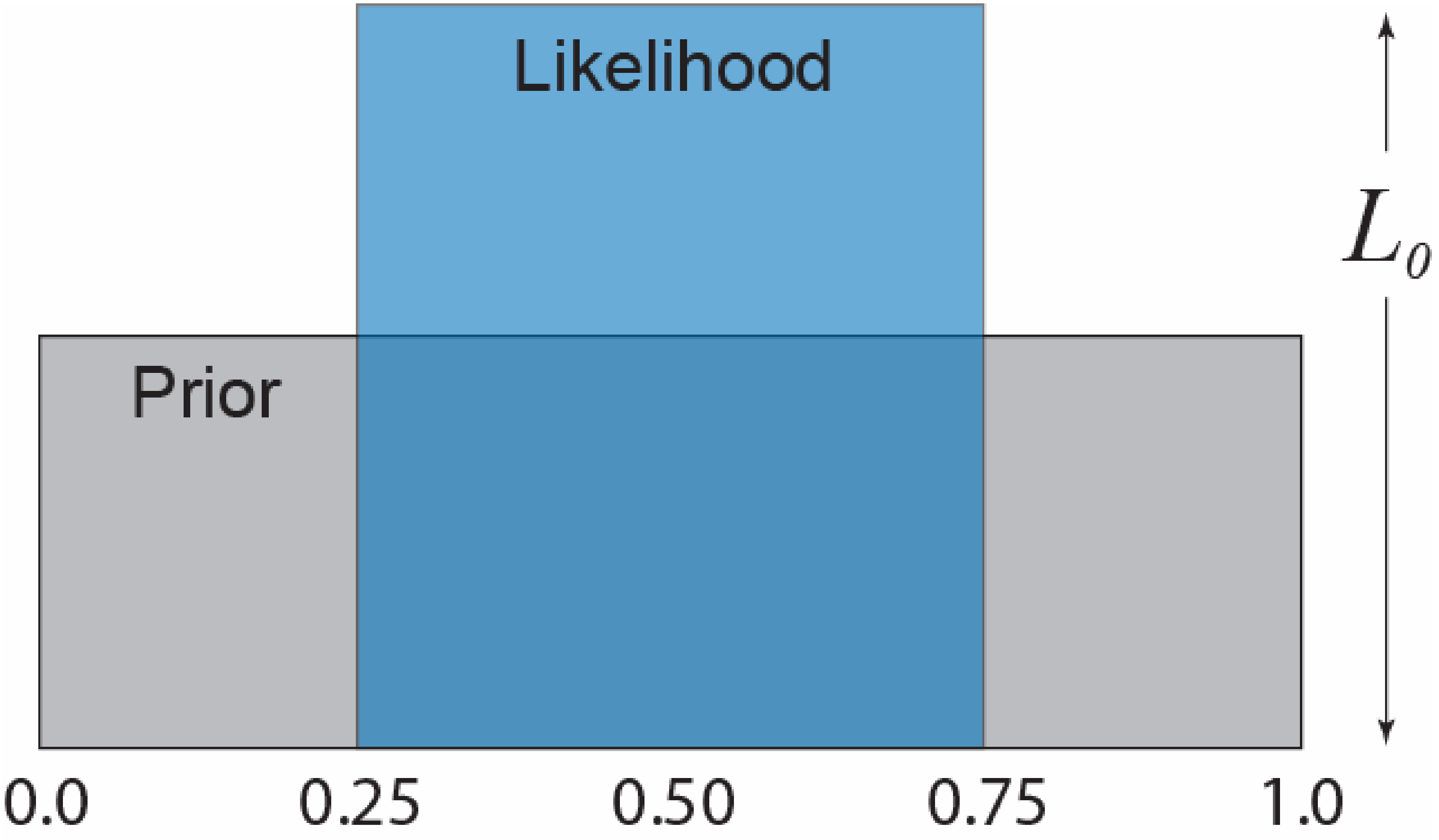
$$\Pr(D) = \int_\theta f(D|\theta)\ f(\theta)\ d\theta$$

Marginal probability of the data (denominator in Bayes' rule). This is a weighted average of the likelihood, where the weights are provided by the prior distribution.

$$\Pr(D|M) = \int_\theta f(D|\theta, M)\ f(\theta|M)\ d\theta$$

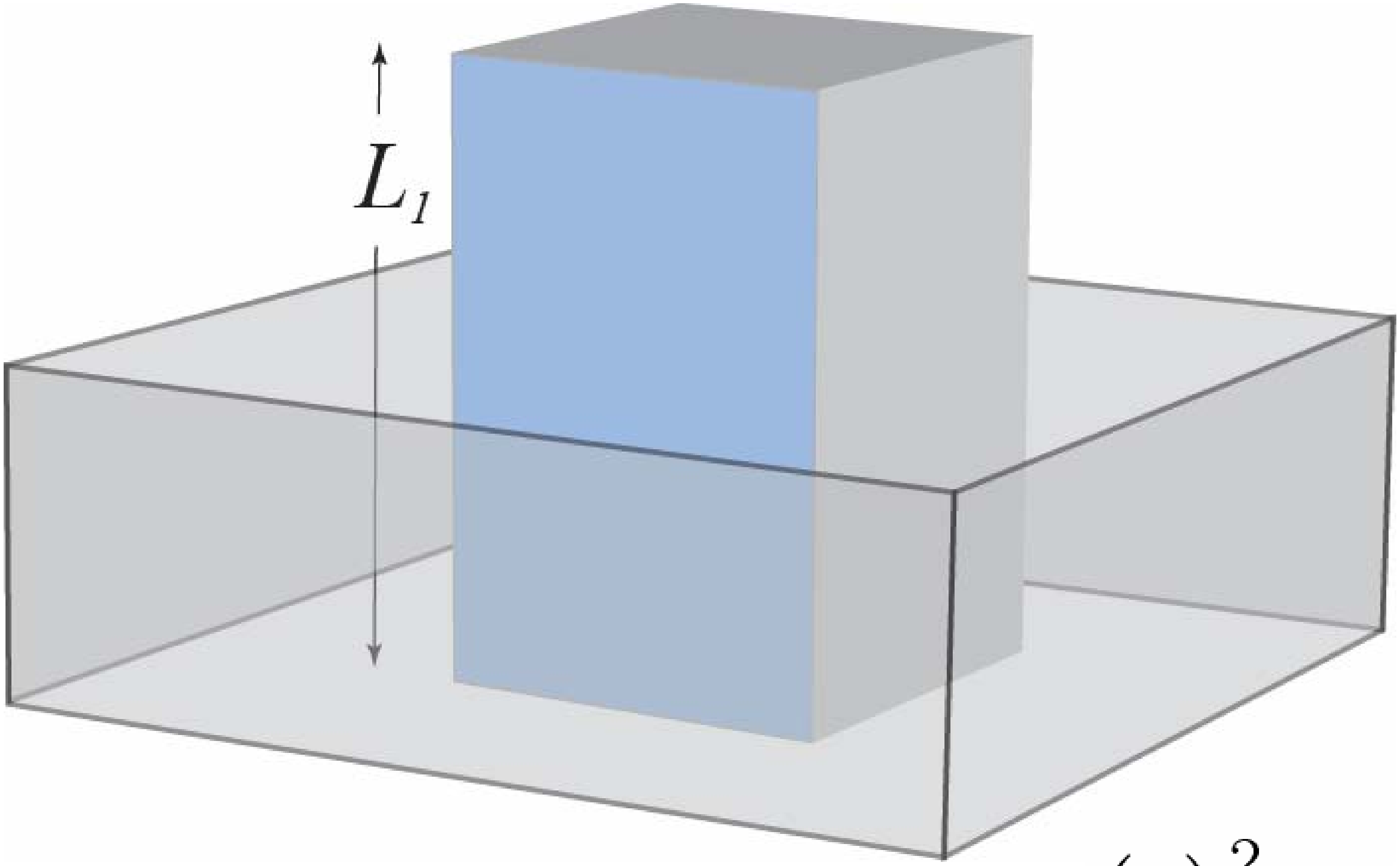
Often left out is the fact that we are also conditioning on M, the model used. $\Pr(D|M_1)$ is comparable to $\Pr(D|M_2)$ and thus the marginal probability of the data can be used to compare the average fit of different models as long as the data D is the same.

Bayes Factor: 1-param. model



Average likelihood = $\left(\frac{1}{2}\right) L_0$

Bayes Factor: 2-param. model



Average likelihood = $\left(\frac{1}{2}\right)^2 L_1$

Bayes Factor is ratio of marginal model likelihoods

1-parameter model M_0: (½) L_0

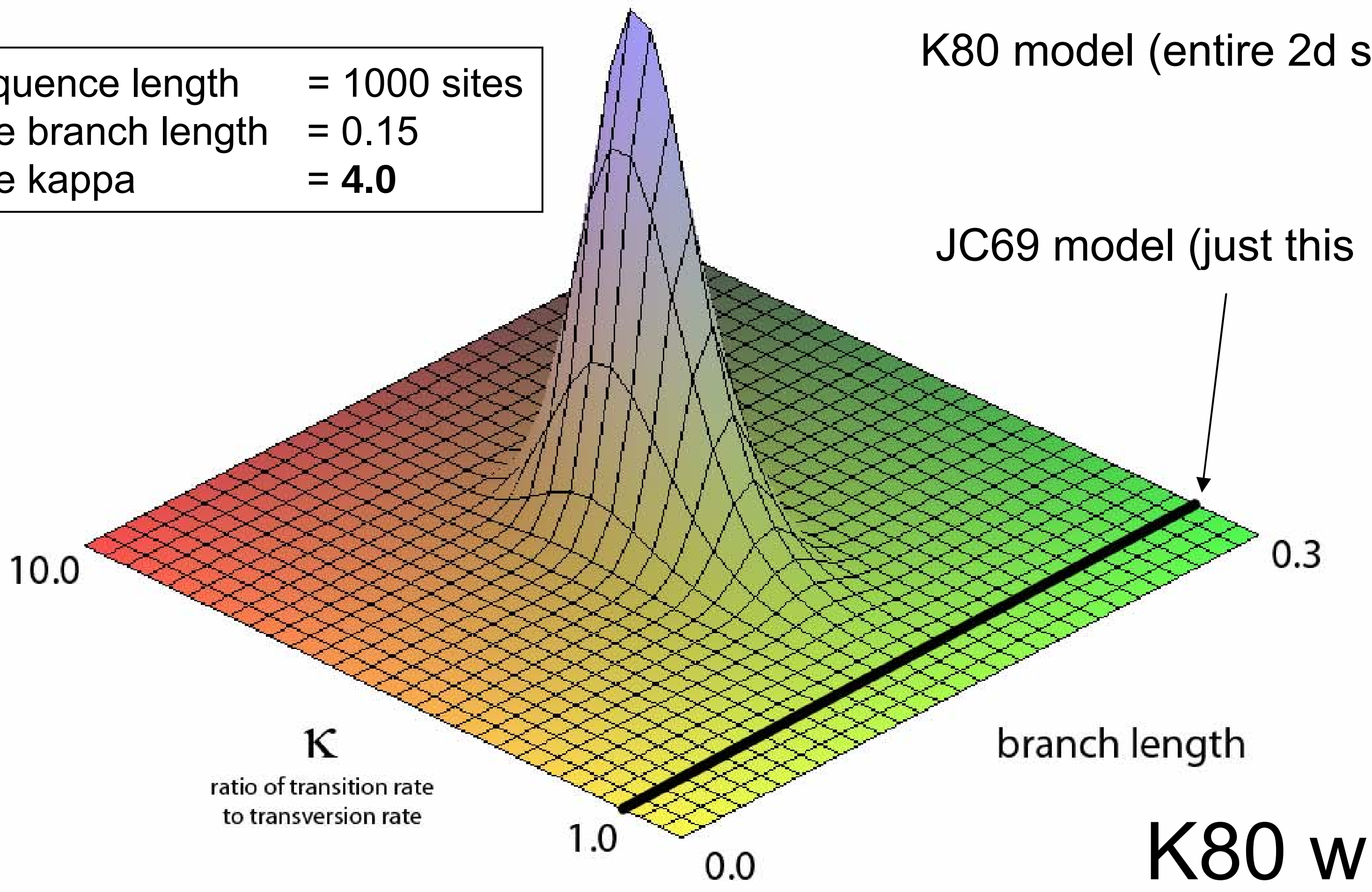
2-parameter model M_1: (¼) L_1

Bayes Factor favors M_0 unless L_1 is at least *twice* as large as L_0

All other things equal, more complex models are penalized by their extra dimensions

Marginal Likelihood of a Model

sequence length	= 1000 sites
true branch length	= 0.15
true kappa	= **4.0**



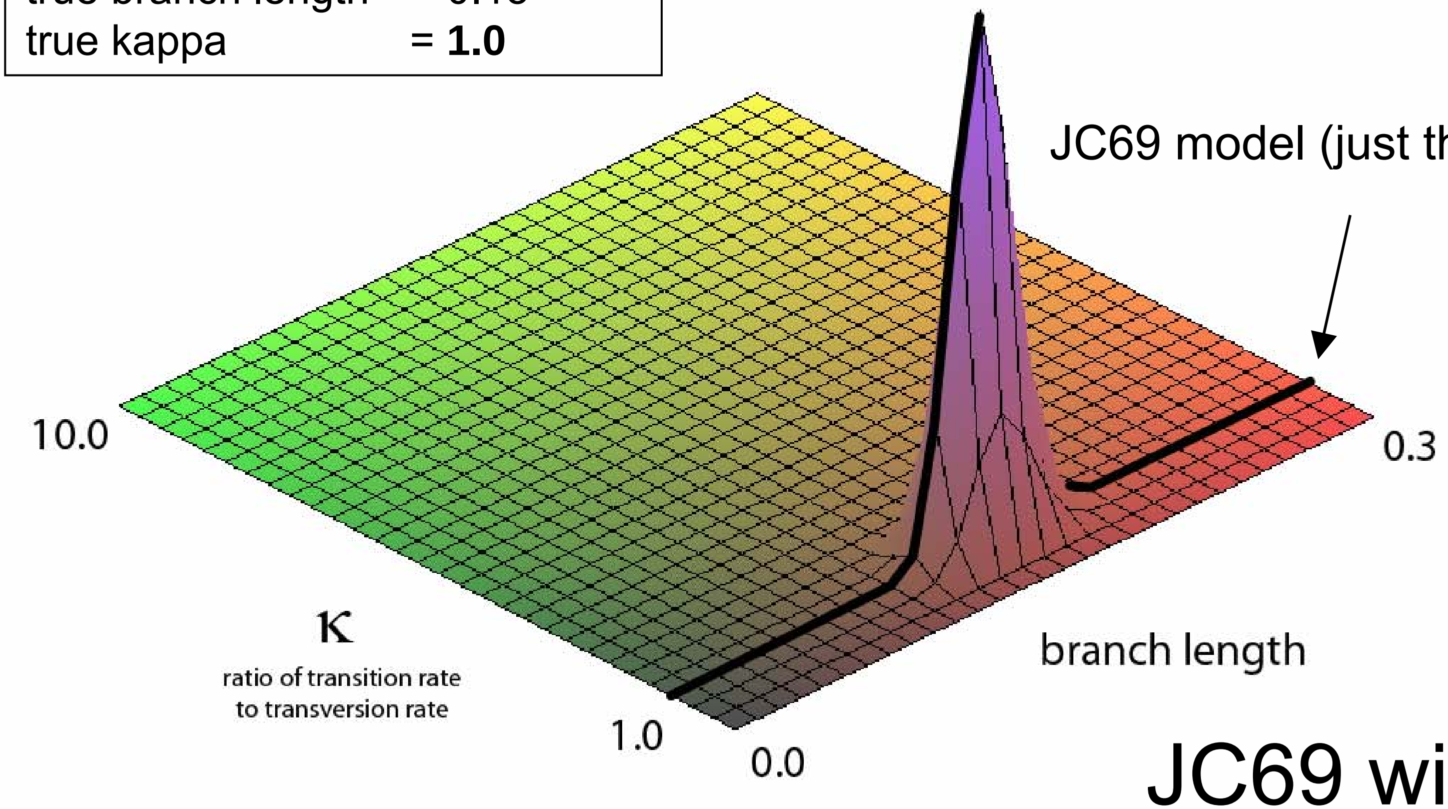
K80 model (entire 2d space)

JC69 model (just this 1d line)

K80 wins

Marginal Likelihood of a Model

sequence length = 1000 sites
true branch length = 0.15
true kappa = **1.0**

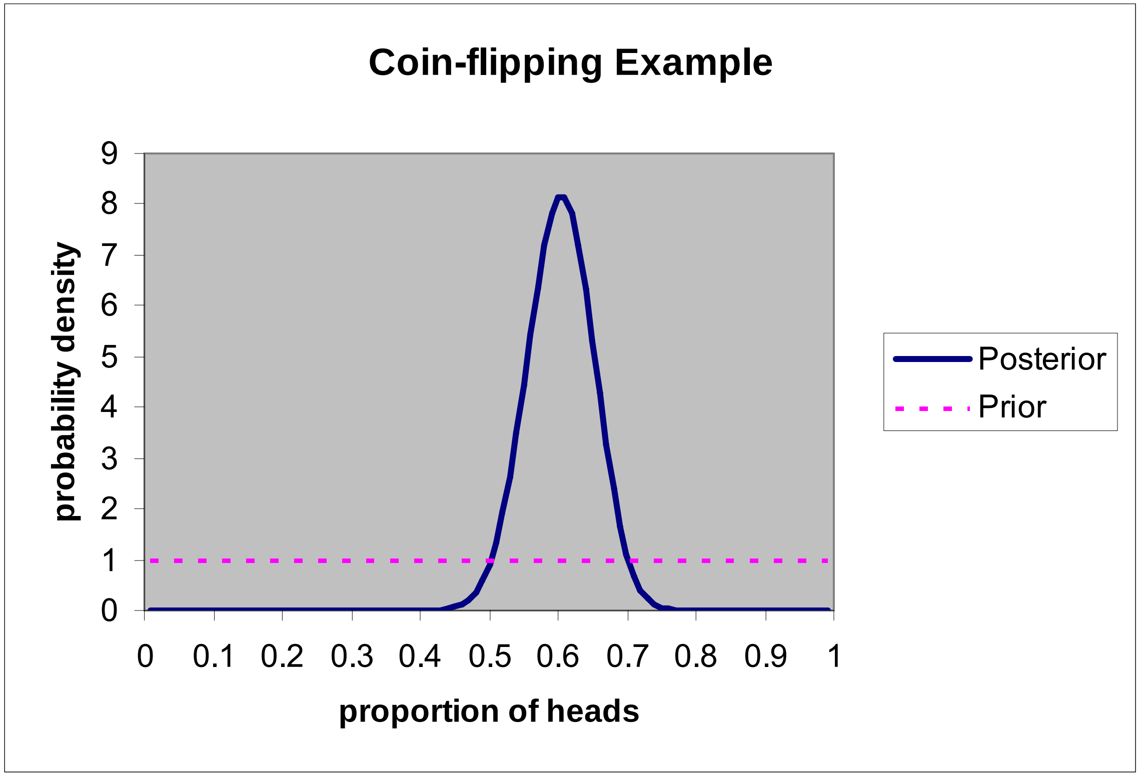


K80 model (entire 2d space)

JC69 model (just this 1d line)

JC69 wins

Direct Method



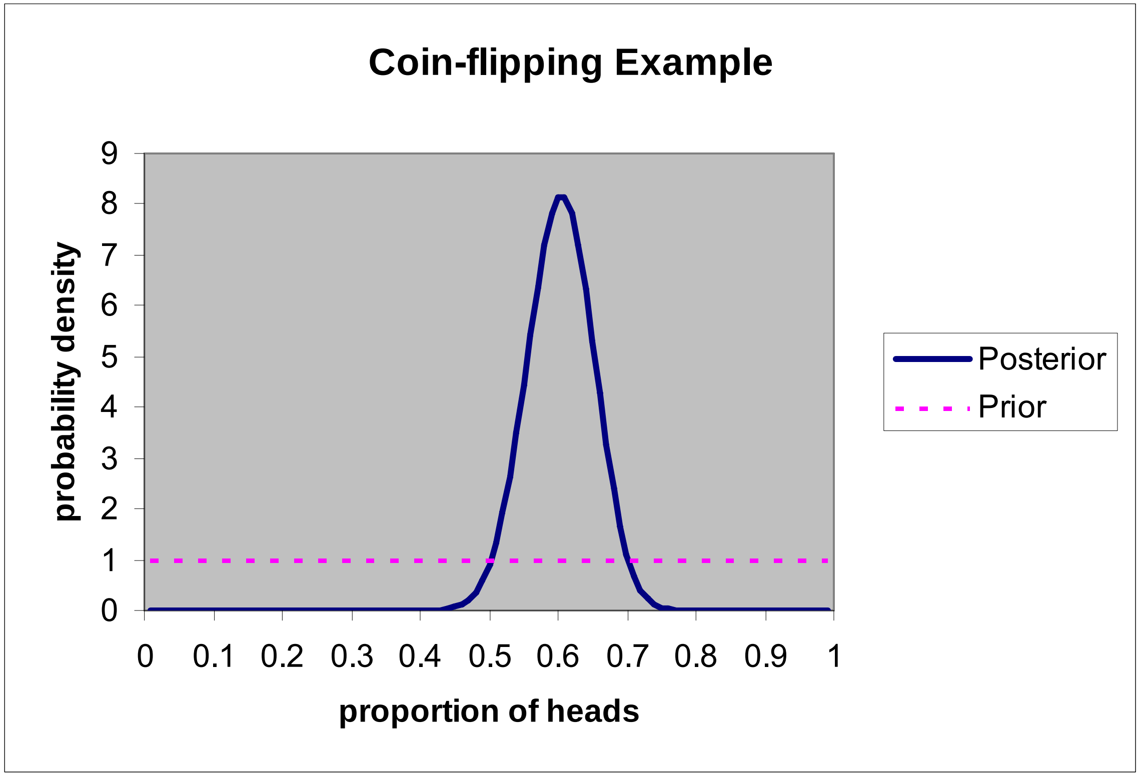
Sample values from the ***prior***. In this case, draw proportion of heads from Uniform(0,1)

Compute likelihood for each point drawn from the prior

Marginal likelihood is the ***arithmetic*** mean of the sampled likelihoods

Problem: tends to **underestimate** marginal likelihood because **few draws** from the prior will be in the **highest** part of the likelihood

Harmonic Mean Method



Sample values from the ***posterior***.

Compute likelihood for each point drawn from posterior

Marginal likelihood is the ***harmonic*** mean of the sampled likelihoods

Problem: tends to **overestimate** marginal likelihood because **few draws** from the posterior will be in the **lowest** part of the likelihood

Thermodynamic Integration[1]

- Special MCMC analysis performed in which the distribution explored slowly changes from posterior to prior
- Produces much more accurate[2] marginal likelihood estimates:

log(marg. like.)	MSE	Method
-167.316	29.62	Harmonic mean
-172.783	0.01	Thermodynamic Integration
-172.743	0.00	True value

- More computation needed than for typical Bayesian MCMC analysis

[1] Lartillot & Phillippe. 2005. Computing Bayes factors using thermodynamic integration. Syst. Biol. 55:195-207.

[2] Work in collaboration with Wangang Xie, Meng Hui Chen, Lynn Kuo and Yu Fan. In this case, model and tree were simple enough that the marginal likelihood could be determined analytically (i.e. the true value is known).

How would we like our phylogenetic inference methods to behave?

Ideally, the methods would return the true tree with strong support for every grouping in the tree.

Why is this perfect performance not possible?

- systematic errors
- sampling errors

What properties are important when choosing between methods? Assessments of support for different aspects of the tree should be:

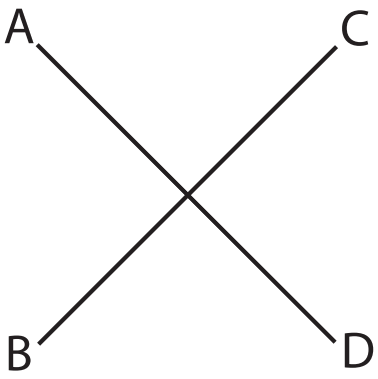
- interpretable
- reliable
- if we do not feel that the statements of support are always believable, then we may prefer to be conservative

Ways to make Bayesian statements of support more conservative:

- Polytomy prior
- data-size dependent priors
- majority-rule consensus trees
- more complex models, robust Bayesian techniques

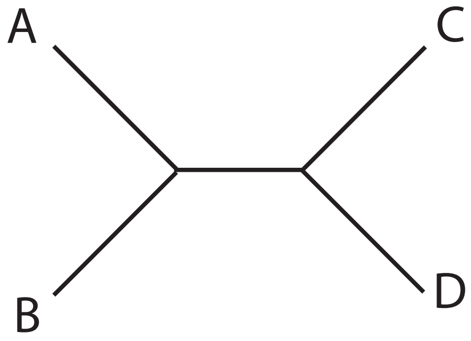
Simulating from stars

model tree

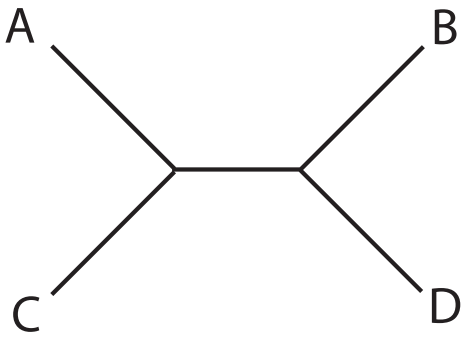


inferred trees

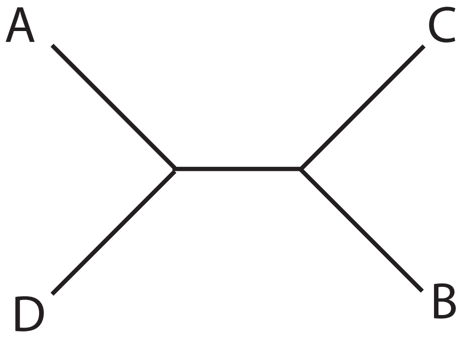
“expected” support



$$\frac{1}{3}$$



$$\frac{1}{3}$$



$$\frac{1}{3}$$

Results of star tree simulations

100,000 sites simulated

Tree 1	Tree 2	Tree 3	Tree 1	Tree 2	Tree 3
0.3029	0.2922	**0.4049**	0.2990	0.3288	**0.3722**
0.4607	0.1362	0.4031	0.3172	0.0464	**0.6364**
0.6704	0.0975	0.2321	0.1584	**0.7969**	0.0447
0.6120	0.1852	0.2028	**0.4625**	0.3600	0.1775
0.3605	0.3570	0.2825	**0.7077**	0.0881	0.2042
0.5455	0.2505	0.2040	0.0884	0.0262	**0.8854**
0.4253	**0.4254**	0.1493	**0.9551**	0.0422	0.0027
0.1595	**0.7465**	0.0940	0.1826	**0.5511**	0.2663
0.4436	0.1697	0.3867	0.3043	**0.4224**	0.2733
0.3994	0.3904	0.2102	**0.6559**	0.0707	0.2734
0.1151	**0.5912**	0.2937	0.0073	**0.9892**	0.0035
0.8333	0.0951	0.0716	0.2703	**0.4112**	0.3185
0.8317	0.0736	0.0947			

Tree geometry

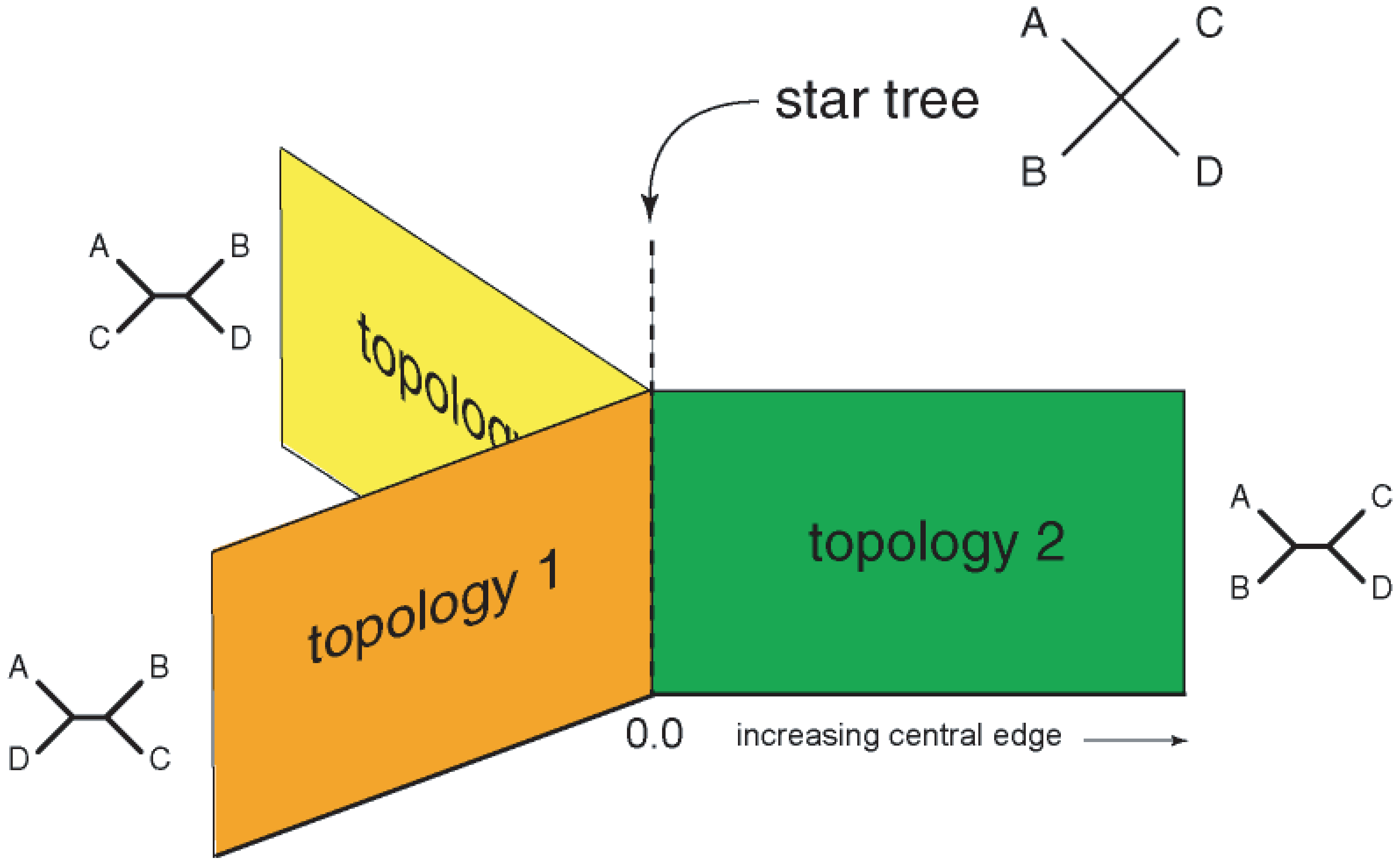


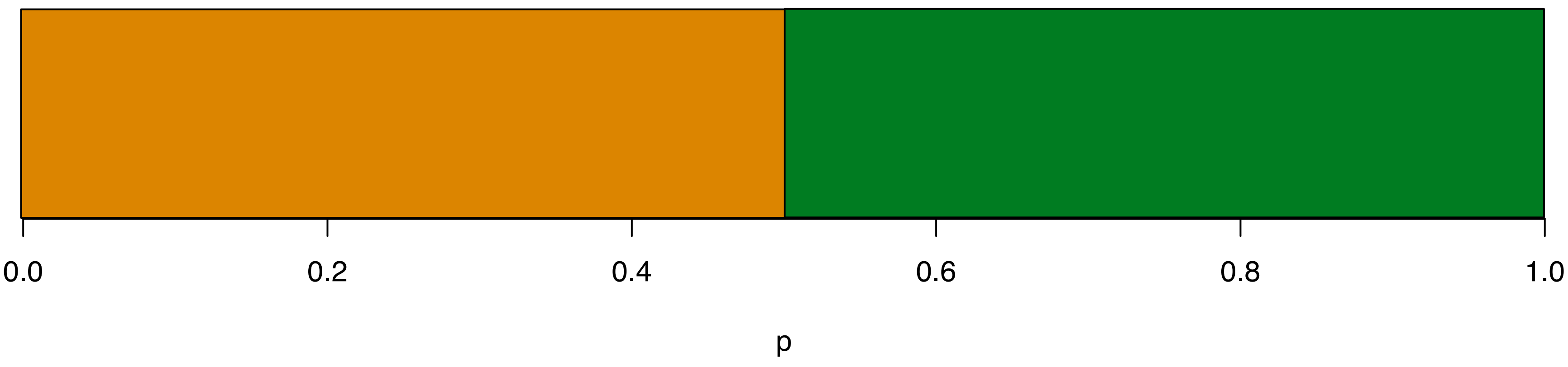
figure courtesy of Paul Lewis

Coin-flipping analogy

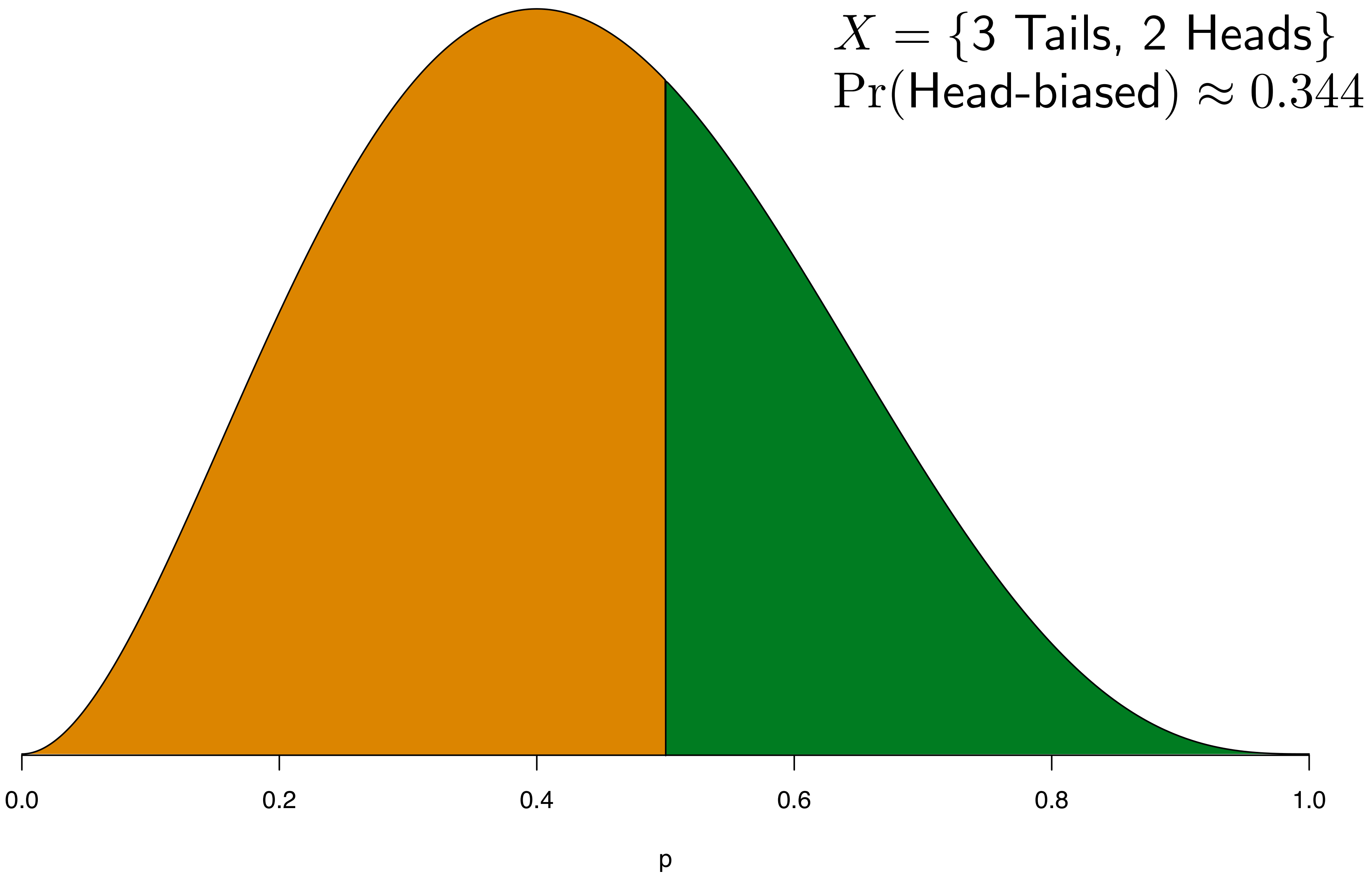
p is the probability of Heads

Uniform prior, no data.

$\mathrm{Pr}(\text{Head-biased}) = 0.5$



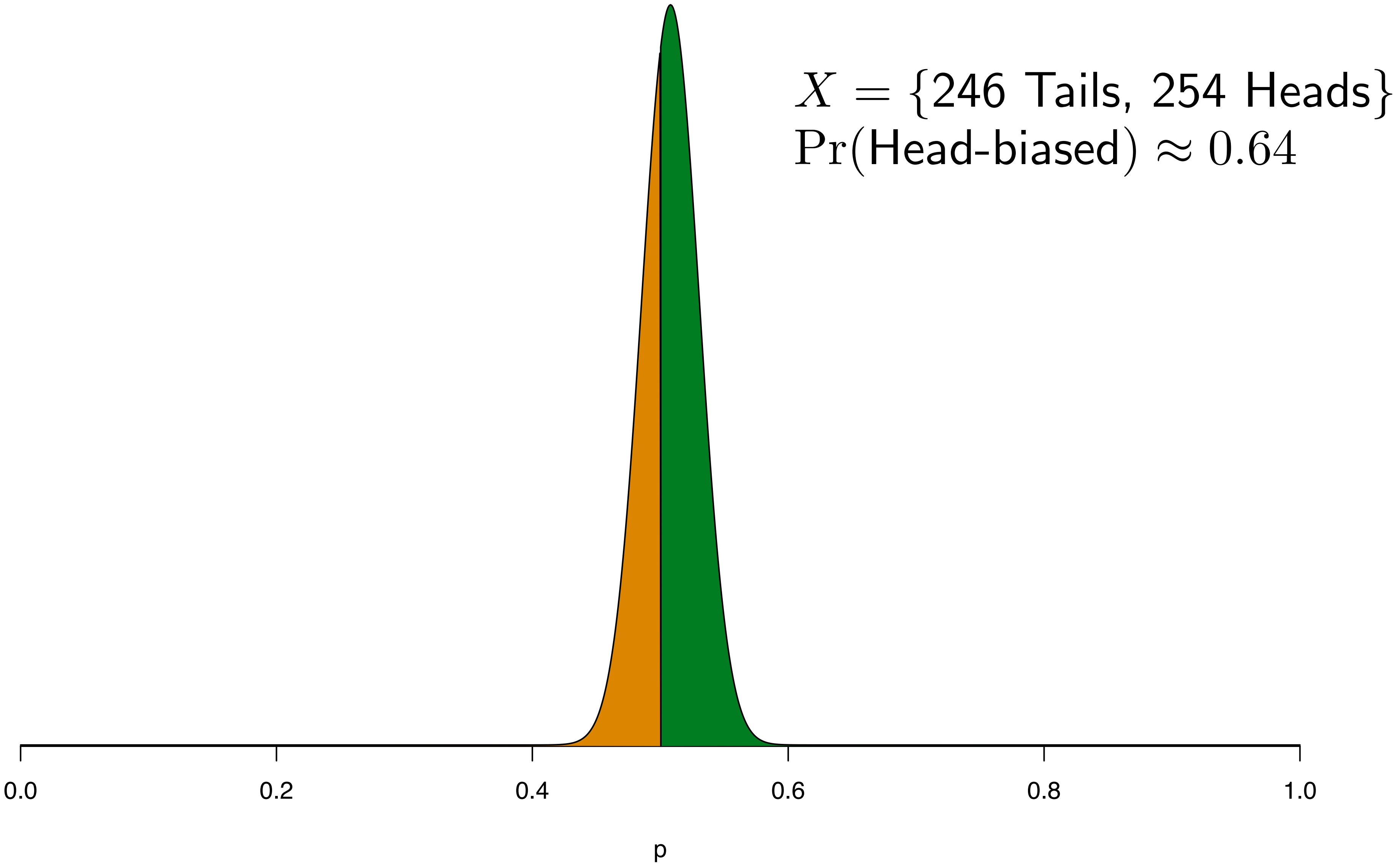
Coin-flipping analogy



$X = \{3 \text{ Tails, } 2 \text{ Heads}\}$

$\Pr(\text{Head-biased}) \approx 0.344$

Coin-flipping analogy



$X = \{\text{246 Tails, 254 Heads}\}$
$\Pr(\text{Head-biased}) \approx 0.64$

Despite the fact that $p = 0.5$:

$$\Pr(\textsf{Head-biased}|\textsf{Data}) \sim \textsf{Uniform}(0, 1)$$

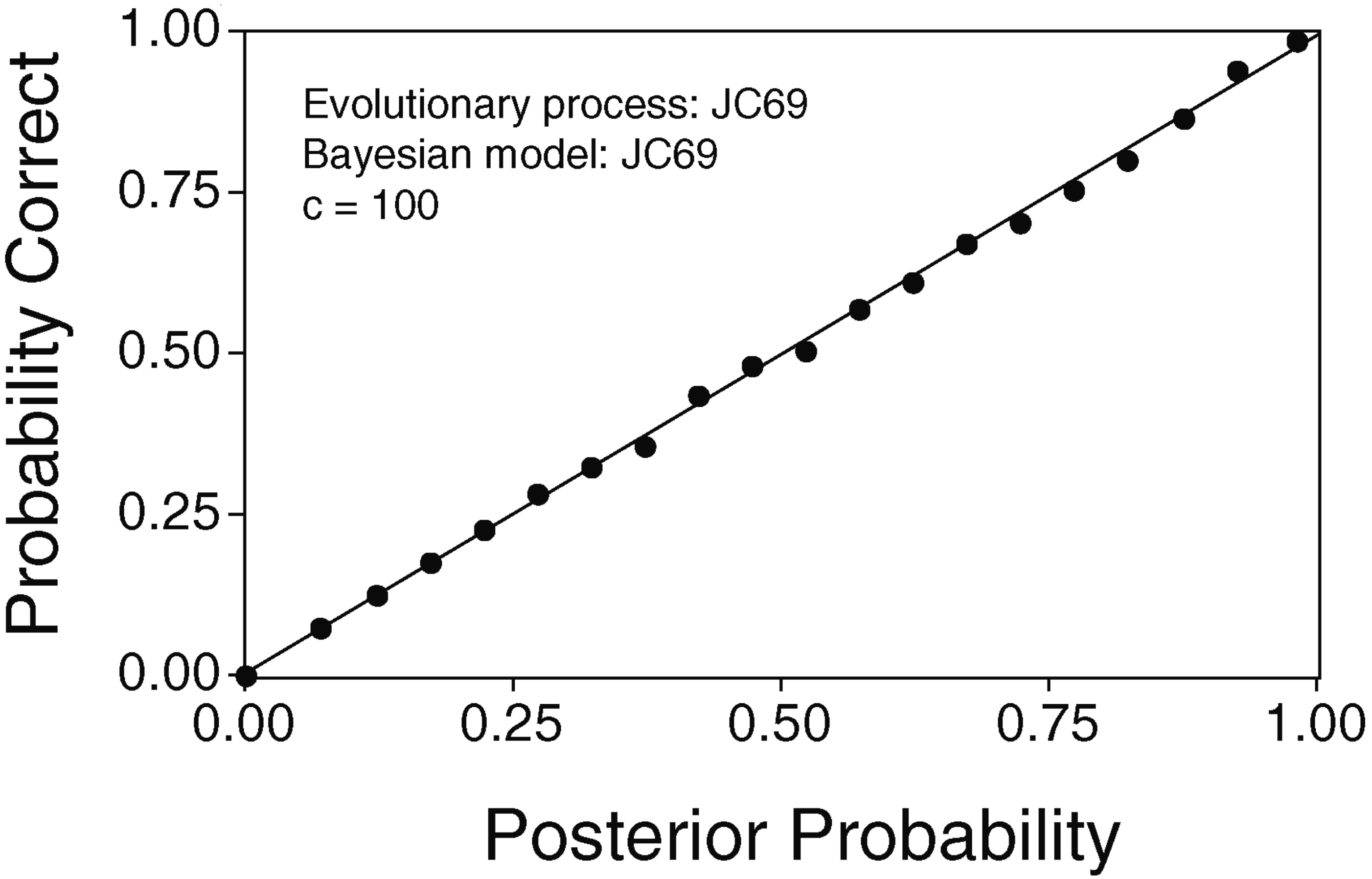
even as the sample size $\rightarrow \infty$.

The nature of the phenomenon

- Polytomies are given 0 prior probability.
- We are asking methods to choose between several *incorrect* answers.
- *Not* a damning flaw in Bayesian analyses (or an indication of a bug in the software).

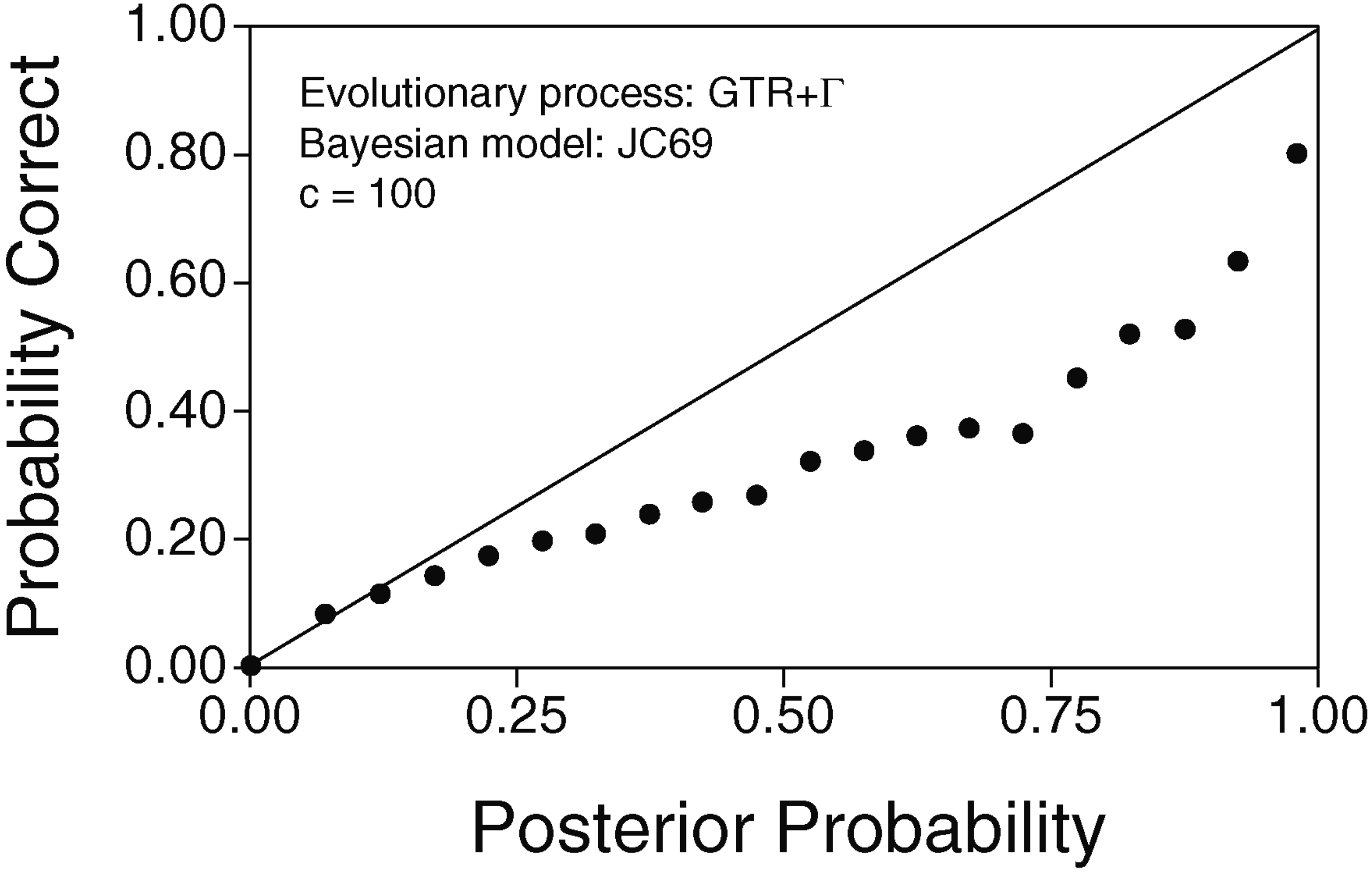
Behavior of Bayesian inference on trees drawn from the prior

From **?**:



Behavior of Bayesian inference when the inference model is too simple

From **?**:

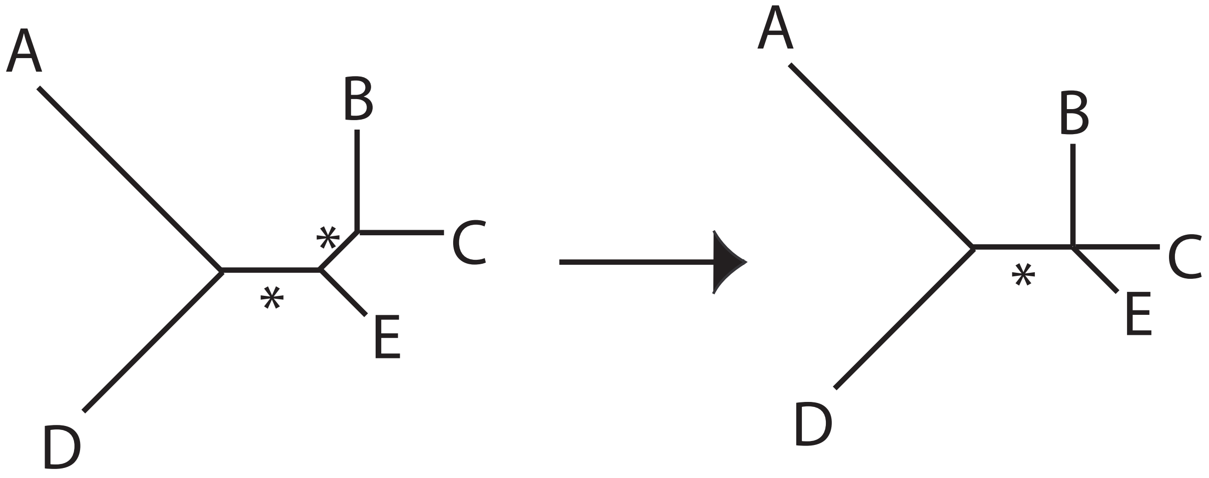


Creating a more conservative analysis

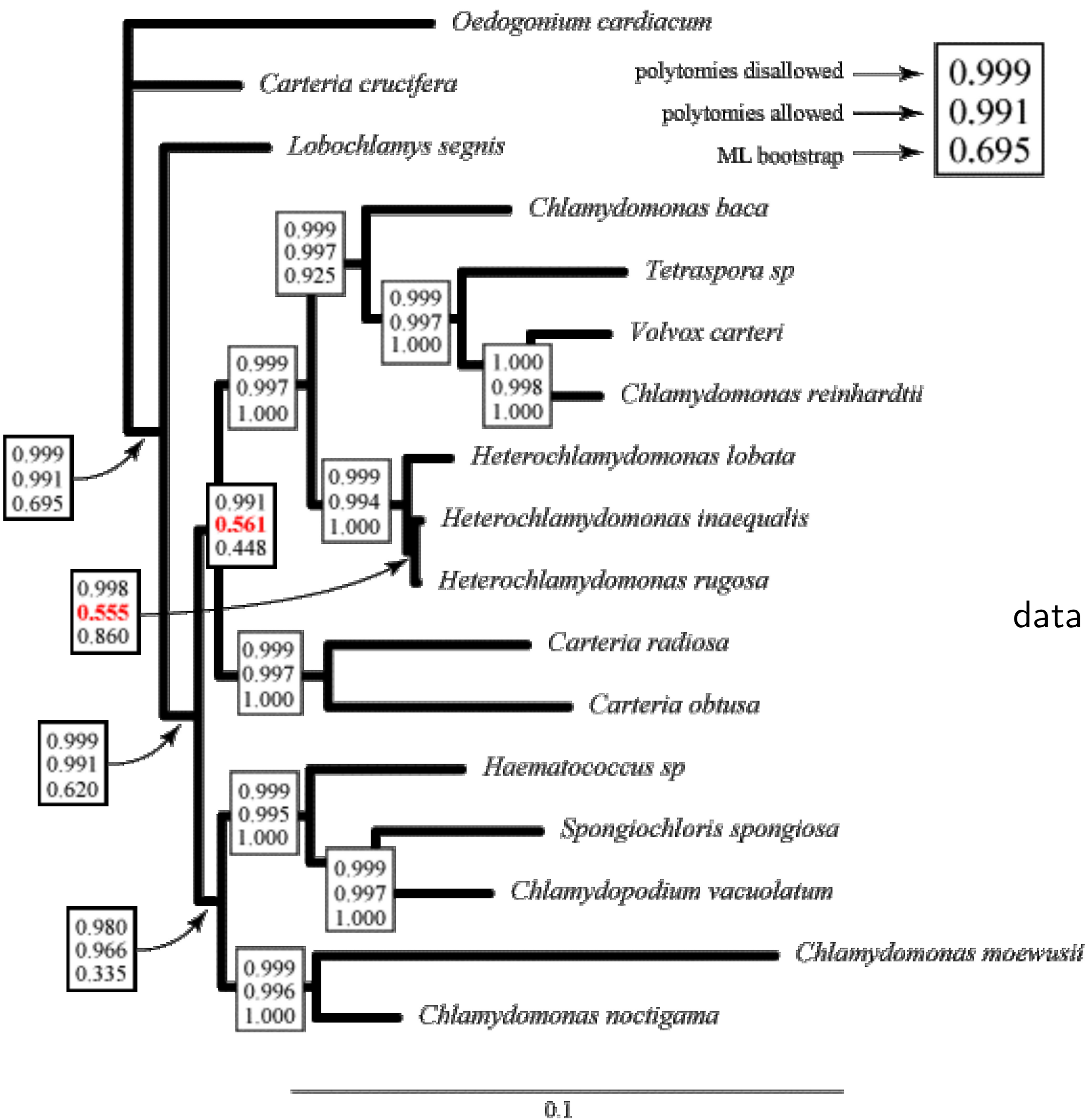
Allowing for polytomies in Bayesian analyses:

- polytomies express uncertainty
- must place a prior probability on unresolved trees
- new MCMC proposals must be invented (see **?**, for details)

Delete Edge Move



Effects of allowing for polytomies



data from **?**

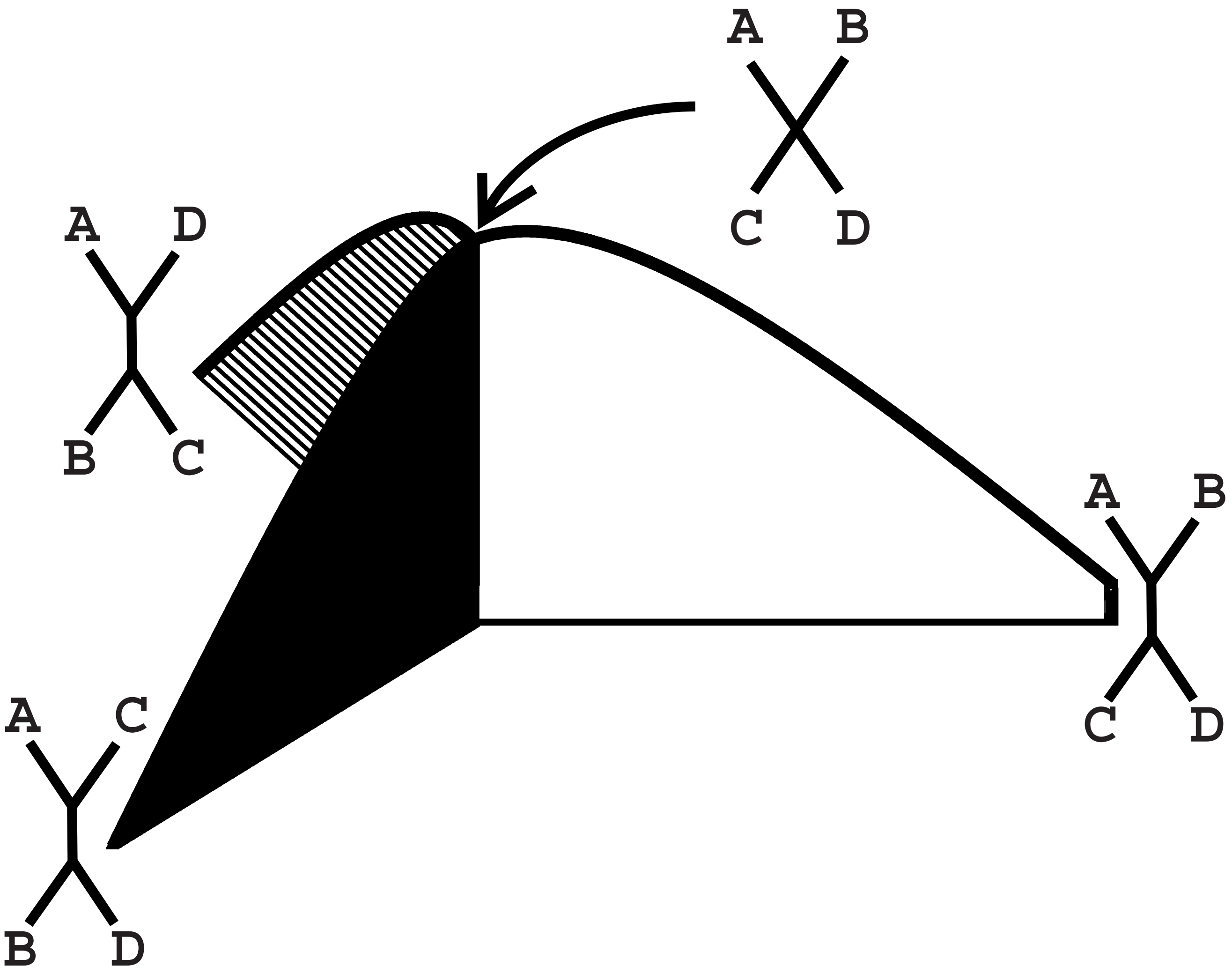
Polytomy MCMC Wrap up

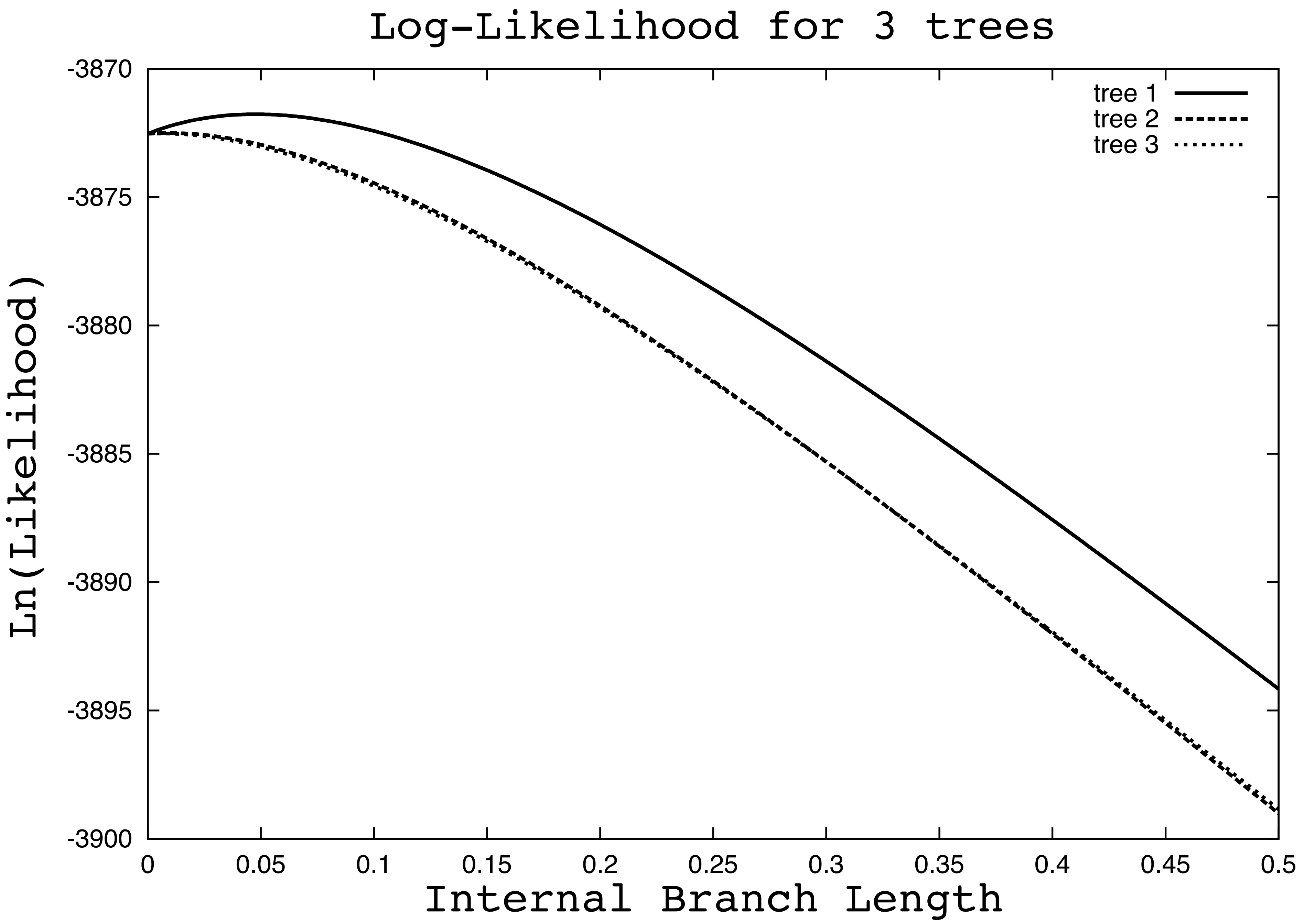
- Allowing unresolved trees is one way to make the Bayesian tree inference more conservative

- Even strong priors in favor of polytomies do not give up too much power

Different priors on the internal and external branches

? suggested using strong priors that favor short lengths for the internal branches of the tree. This can lower the support for potentially spurious groupings.


A
B
C
D
A
D
B
C
A
B
C
D
A
C
B
D


Log-Likelihood for 3 trees
Ln(Likelihood)
-3870
-3875
-3880
-3885
-3890
-3895
-3900
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Internal Branch Length
tree 1
tree 2
tree 3