

Bayesian Phylogenetics

Mark Holder (with **big** thanks to Paul Lewis)

Outline

- Intro
 - **What** is Bayesian Analysis?
 - **Why** be a Bayesian?
- **What is required** to do a Bayesian Analysis? (Priors)
- **How** can the required calculations be done? (MCMC)
- Prospects and Warnings

Simple Example:

Vesicoureteral Reflux (VUR) - valves between the ureters and bladder do not shut fully.

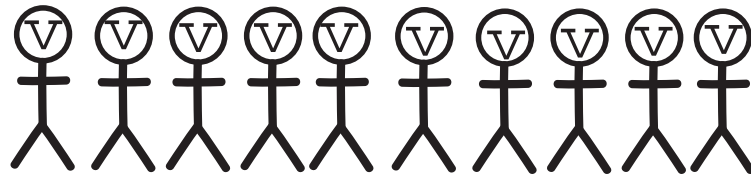
- leads to urinary tract infections
- if not corrected, can cause serious kidney damage
- effective diagnostic tests are available, but they are expensive and invasive


- $\approx 1\%$ of children will have VUR
- $\approx 80\%$ of children with VUR will see a doctor about an infection
- $\approx 2\%$ of *all* children will see doctor about an infection

Should a child with 1 infection be screened for VUR?

1% of the population has VUR

$$\Pr(V) = 0.01$$



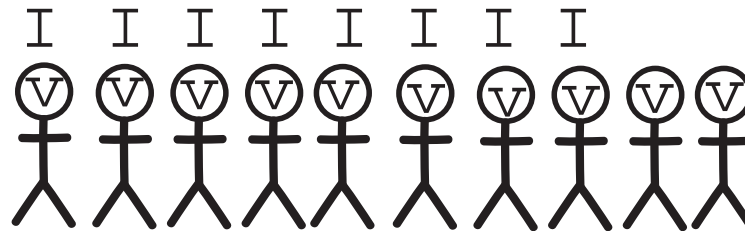
 = 0.1% of the population

80% of kids with VUR
get an infection
 $\Pr(I|V) = 0.8$

$\Pr(I|V)$ is a ***conditional*** probability

So, 0.8% of the population has
VUR and will get an infection

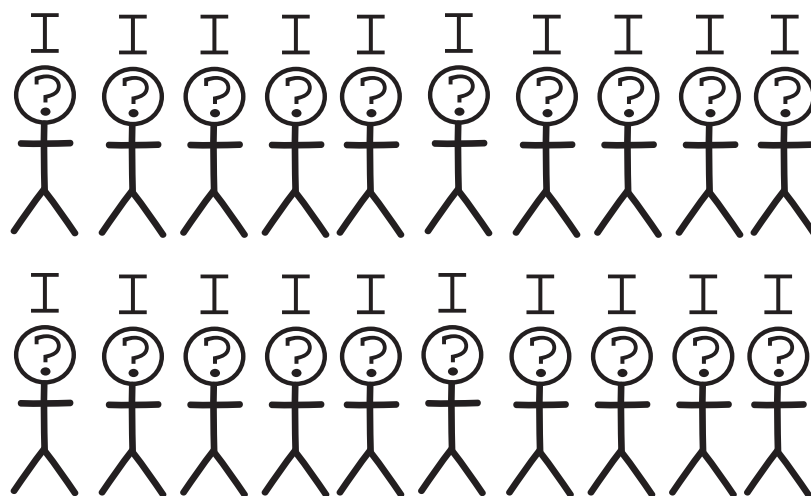
$$\Pr(V) \Pr(I|V) = 0.01 \times 0.8 = 0.008$$
$$\Pr(I, V) = 0.008$$



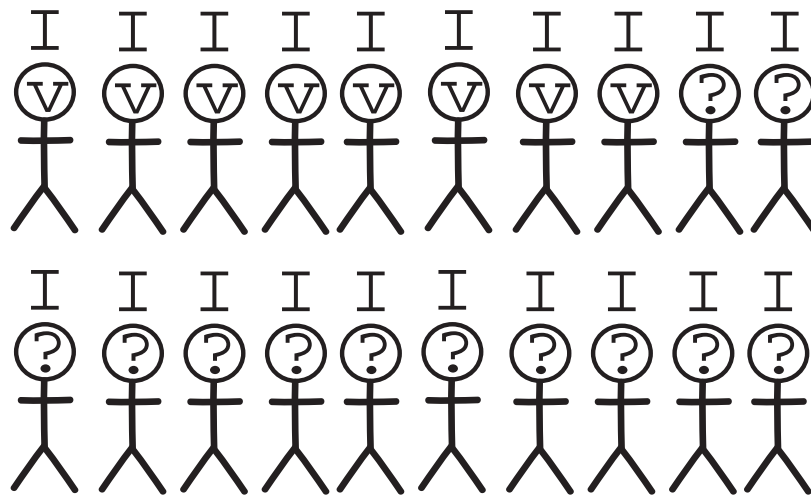
$\Pr(I, V)$ is a **joint** probability

2% of the population gets
an infection

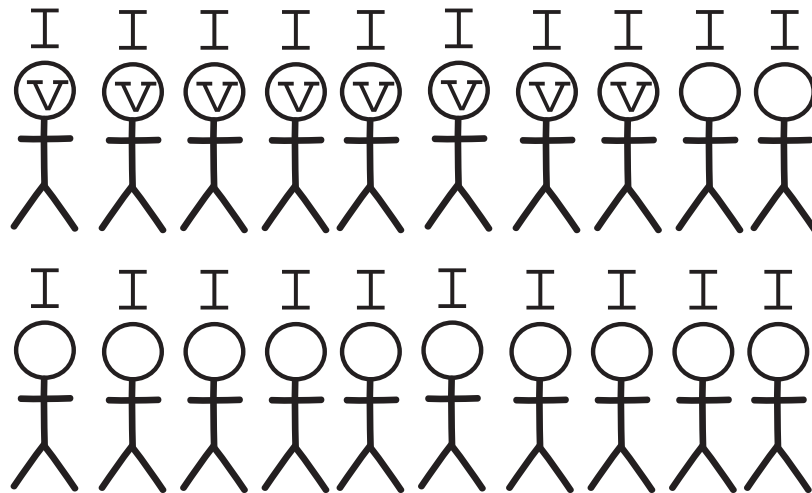
$$\Pr(I) = 0.02$$



We just calculated that
0.8% of kids have VUR
and get an infection



The other 0.12% must
not have VUR



So, 40% of kids with
infections have VUR

$$\Pr(V|I) = 0.4$$

$$Pr(V|I) = \frac{Pr(V)Pr(I|V)}{Pr(I)}$$

$$Pr(V|I) = \frac{0.01 \times 0.8}{0.02}$$

$$= 0.40$$

Pr(I) is higher for females.

$$Pr(I|\text{♀}) = 0.03 \quad Pr(I|\text{♂}) = 0.01$$

$$Pr(V|I, \text{♀}) = \frac{0.01 \times 0.8}{0.03} \quad \left| \quad Pr(V|I, \text{♂}) = \frac{0.01 \times 0.8}{0.01}$$

$$Pr(V|I, \text{♀}) = 0.267 \quad \left| \quad Pr(V|I, \text{♂}) = 0.8$$

Bayes' Rule

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

$$Pr(\text{Hypothesis}|\text{Data}) = \frac{Pr(\text{Hypothesis})Pr(\text{Data}|\text{Hypothesis})}{Pr(\text{Data})}$$

$$Pr(\text{Tree}|\text{Data}) = \frac{Pr(\text{Tree})Pr(\text{Data}|\text{Tree})}{\mathbf{Pr}(\mathbf{Data})}$$

We can ignore $\mathbf{Pr}(\mathbf{Data})$

(2nd half of this lecture)

$$Pr(\text{Tree}|\text{Data}) \propto \mathbf{Pr}(\mathbf{Tree})Pr(\text{Data}|\text{Tree})$$

$Pr(\text{Tree})$ is the *prior* probability of the tree.

$$Pr(\text{Tree}|\text{Data}) \propto Pr(\text{Tree})\mathbf{Pr}(\mathbf{Data}|\mathbf{Tree})$$

$Pr(\text{Tree})$ is the *prior* probability of the tree.

$\mathbf{Pr}(\mathbf{Data}|\mathbf{Tree})$ is the likelihood of the tree.

$$Pr(\text{Tree}|\text{Data}) \propto Pr(\text{Tree})L(\mathbf{Tree})$$

$$\mathbf{Pr}(\mathbf{Tree}|\mathbf{Data}) \propto Pr(\mathbf{Tree})L(\mathbf{Tree})$$

$Pr(\mathbf{Tree})$ is the *prior* probability of the tree.

$L(\mathbf{Tree})$ is the likelihood of the tree.

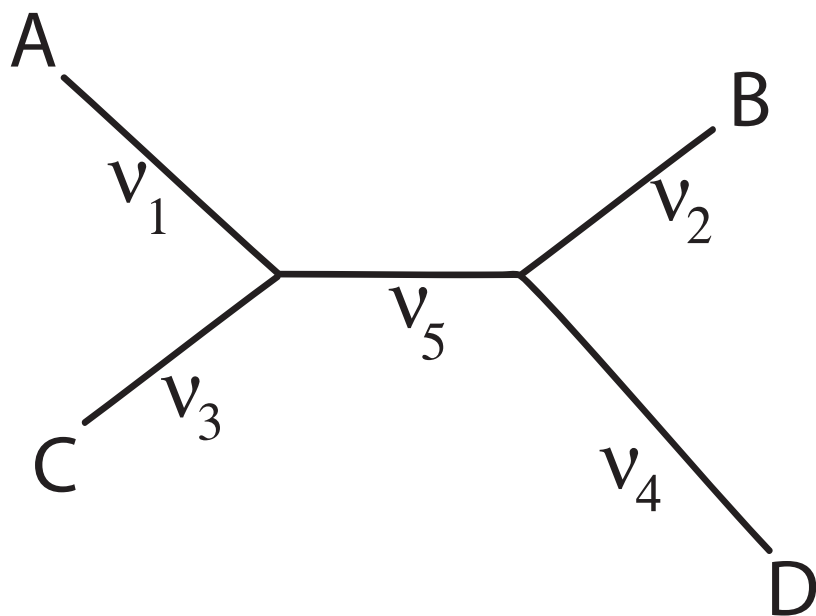
$\mathbf{Pr}(\mathbf{Tree}|\mathbf{Data})$ is the *posterior* probability of the tree.

The posterior probability is a great way to evaluate trees:

- Ranks trees
- Intuitive measure of confidence
- Is the ideal “weight” for a tree in secondary analyses
- Closely tied to the likelihood

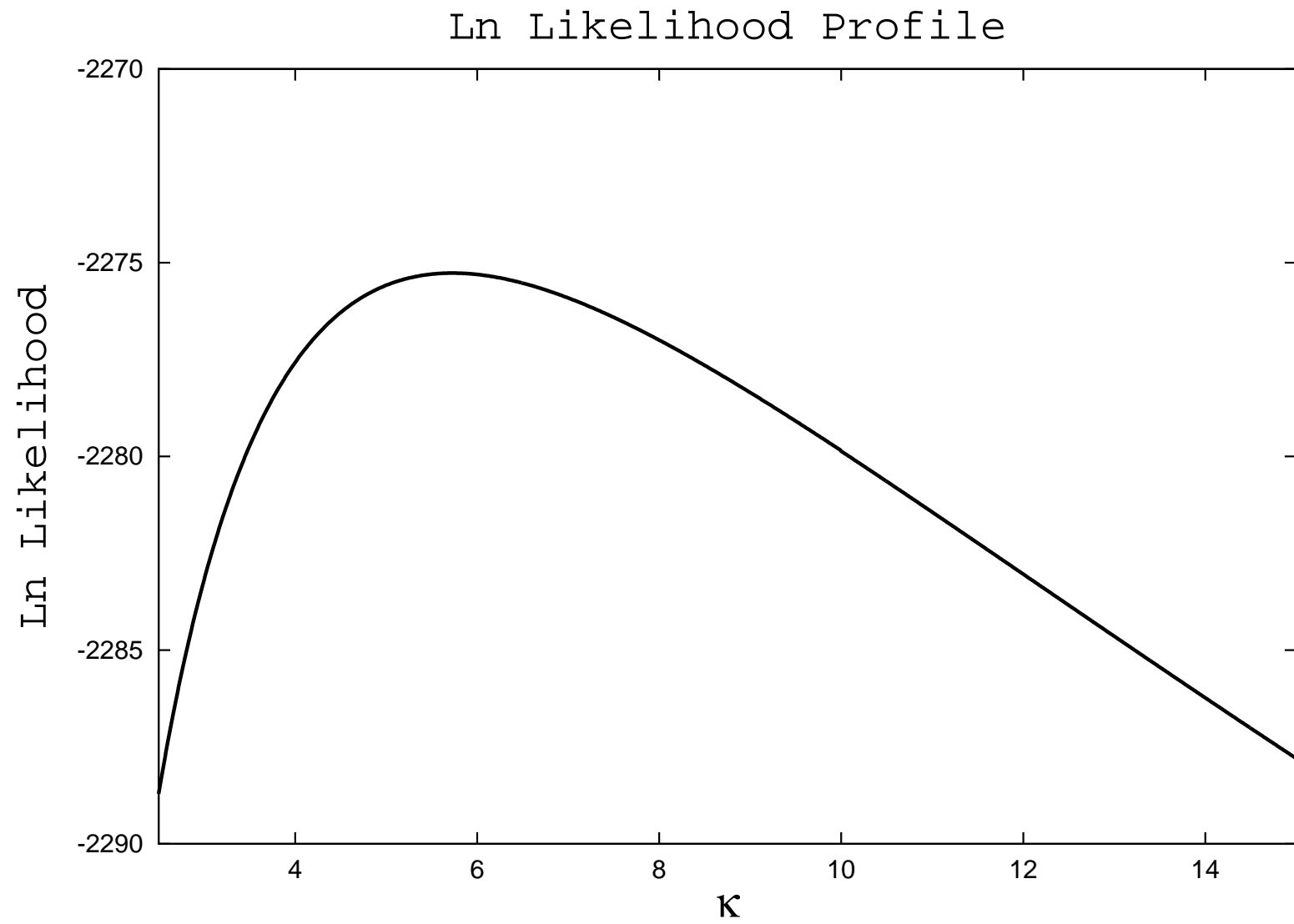
Our models don't give us $L(\text{Tree})$

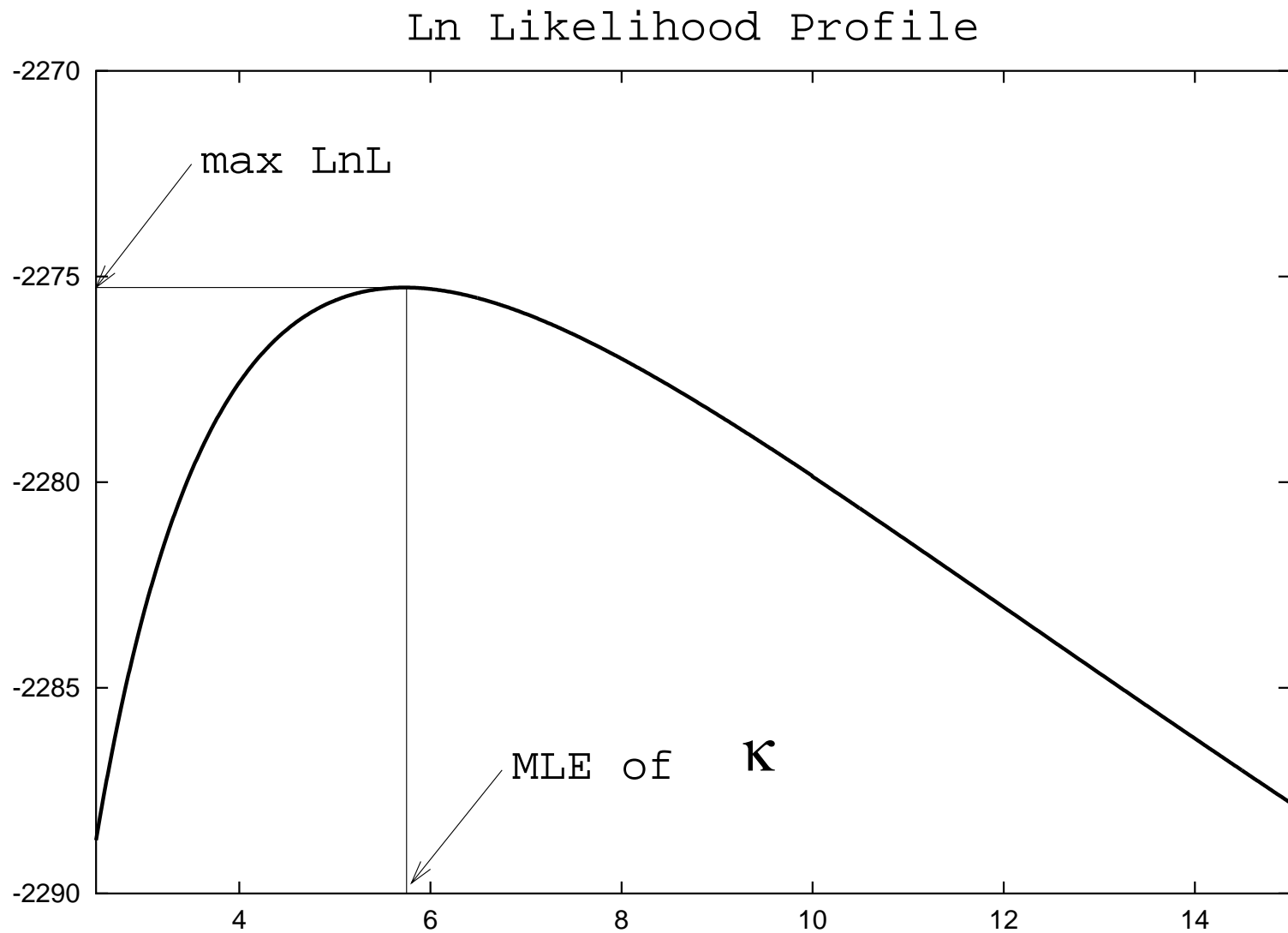
They give us things like $L(\text{Tree}, \kappa, \alpha, \nu_1, \nu_2, \nu_3, \nu_4, \nu_5)$



“Nuisance Parameters”

Aspects of the evolutionary model that we don't care about, but are in the likelihood equation.





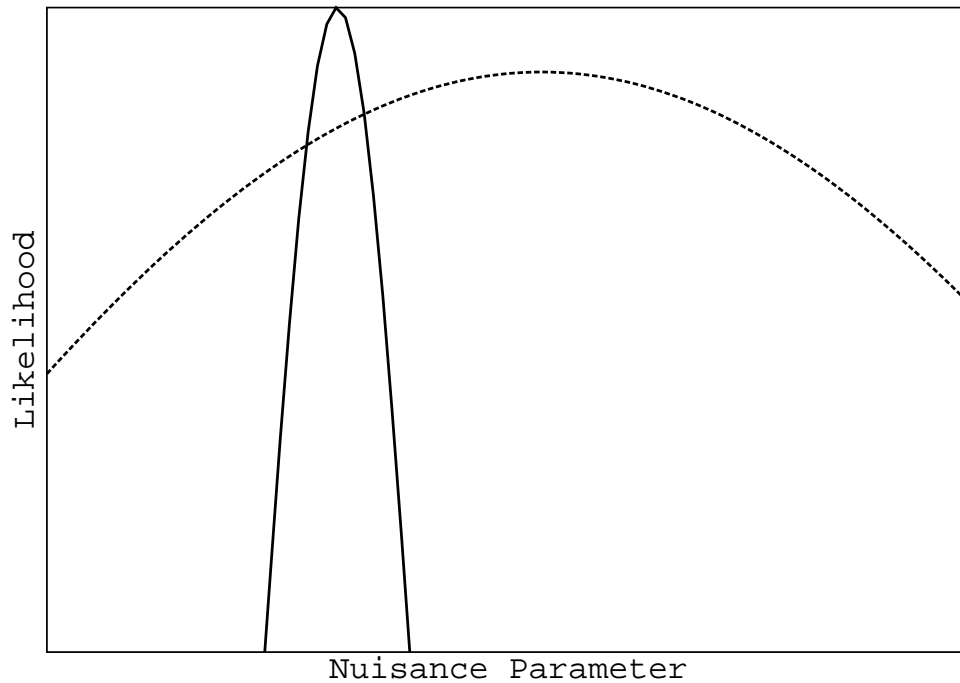
Marginalizing over (integrating out) nuisance parameters

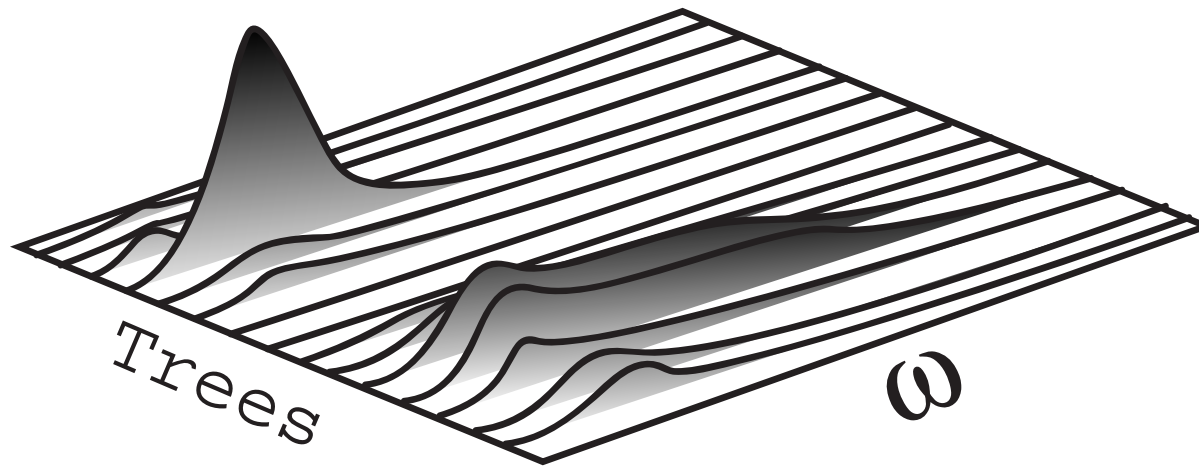
$$L(\text{Tree}) = \int L(\text{Tree}, \kappa) Pr(\kappa) d\kappa$$

- Removes the nuisance parameter
- Takes the entire likelihood function into account

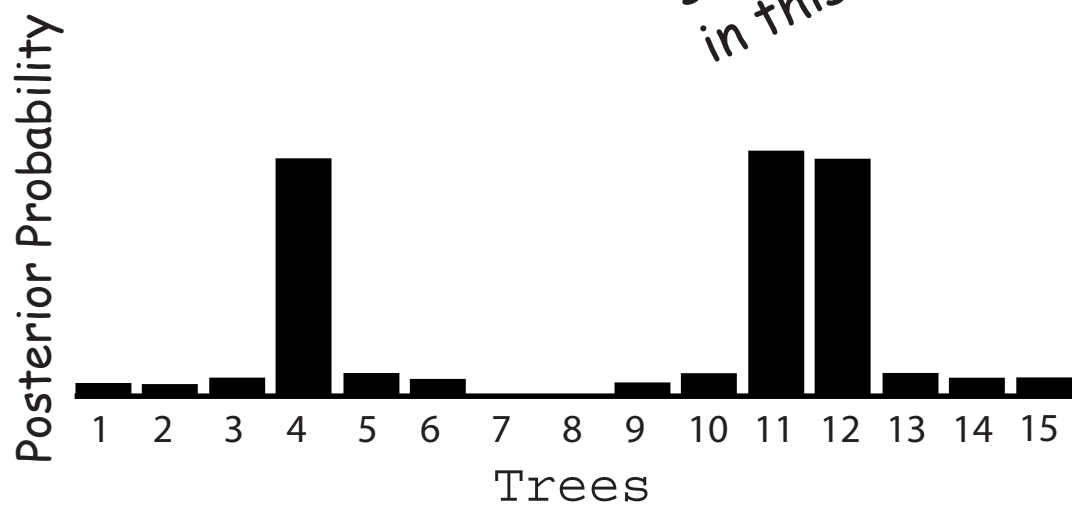
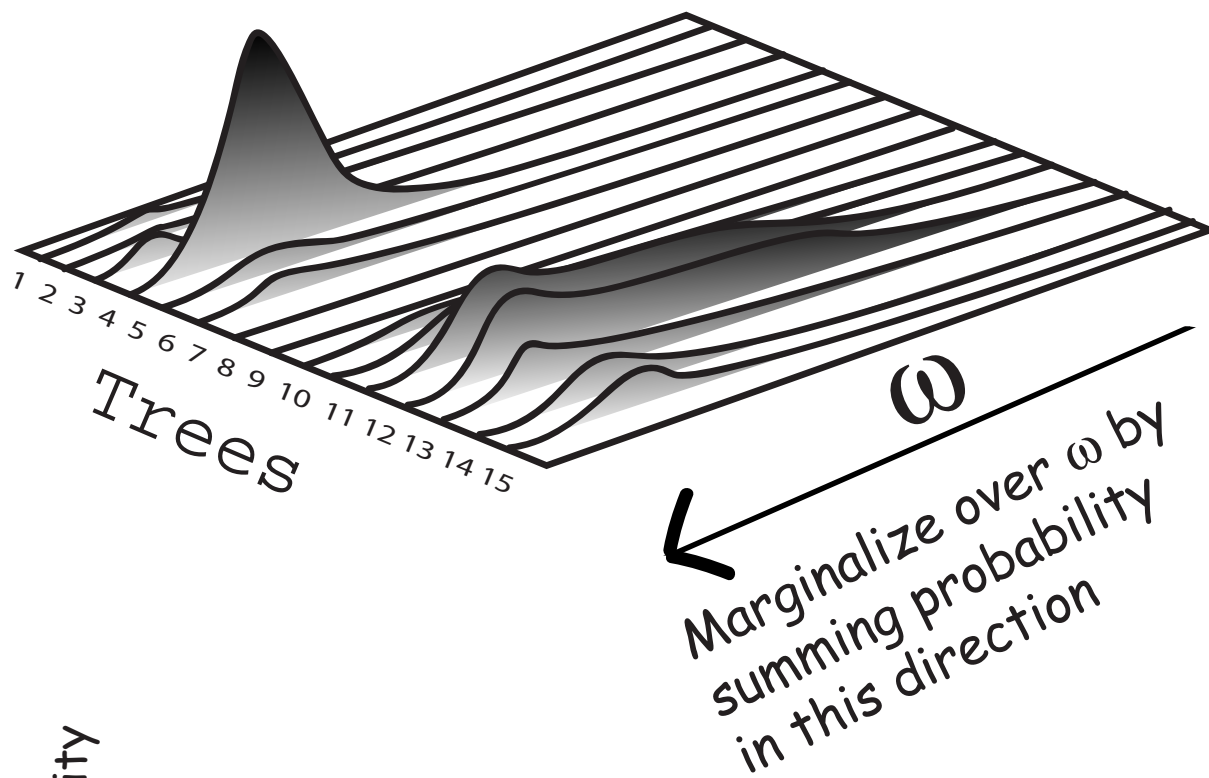
- Avoids estimation errors
- Requires a prior for the parameter

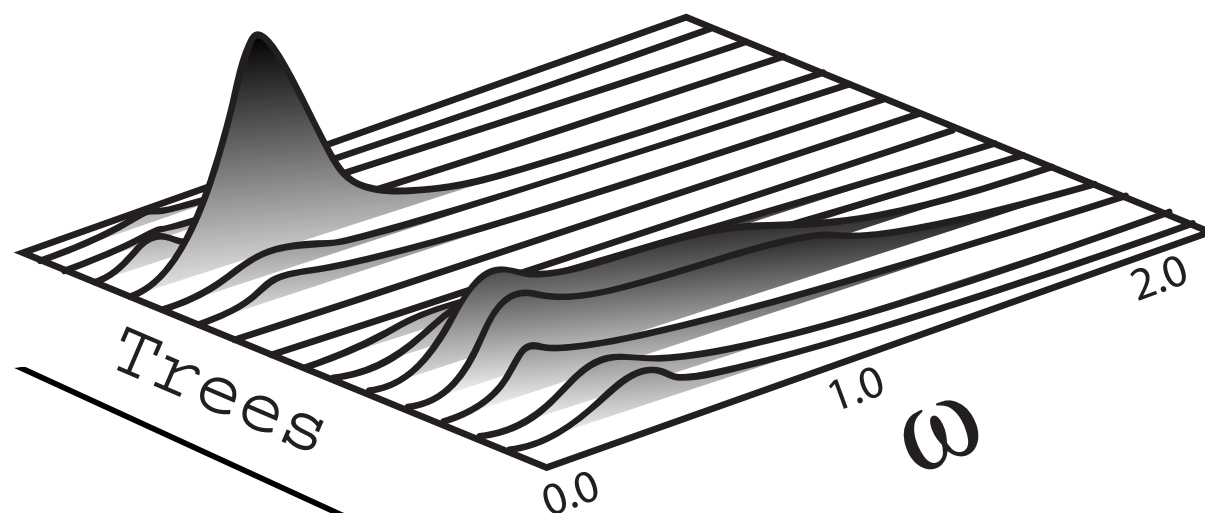
When there is substantial uncertainty in a parameter's value, marginalizing can give qualitatively different answers than using the MLE.



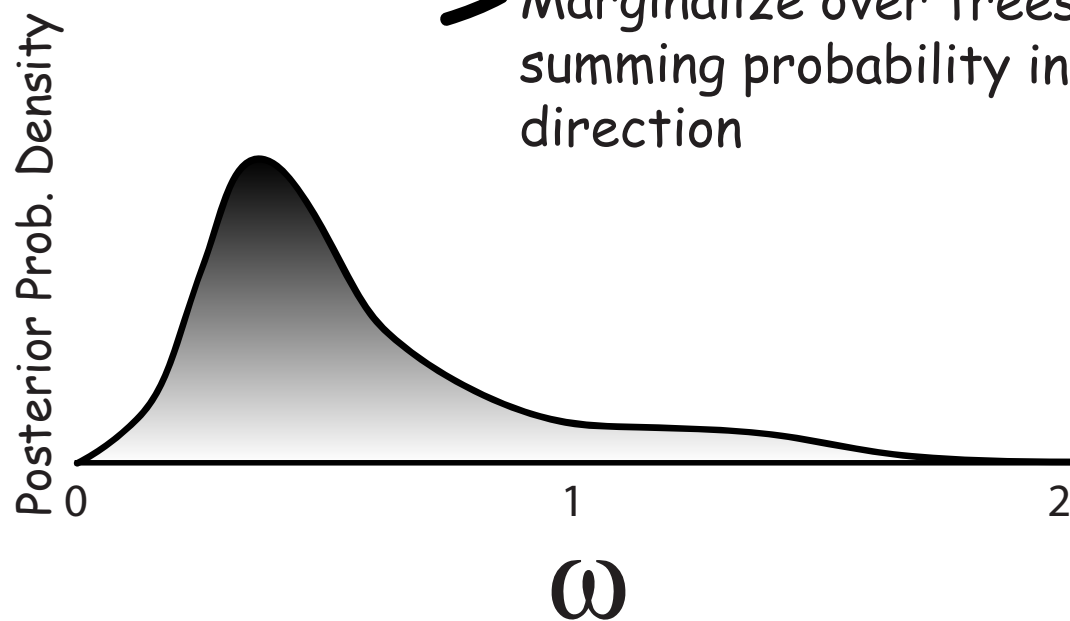


Joint posterior probability
density for trees and ω





→ Marginalize over trees by summing probability in this direction



The Bayesian Perspective

Pros	Cons
Posterior probability is the ideal measure of support	Is it robust?
Focus of inference is flexible	
Marginalizes over nuisance parameters	Requires a prior

Priors

- Probability distributions
- Specified *before* analyzing the data
- Needed for
 - Hypotheses (trees)
 - Parameters

Probability Distributions

Reflect the action of random forces

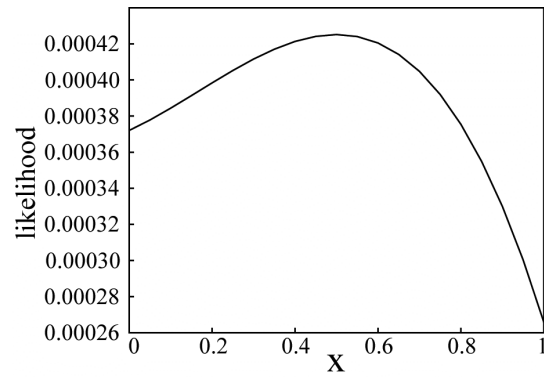
Probability Distributions

Reflect the action of random forces

OR

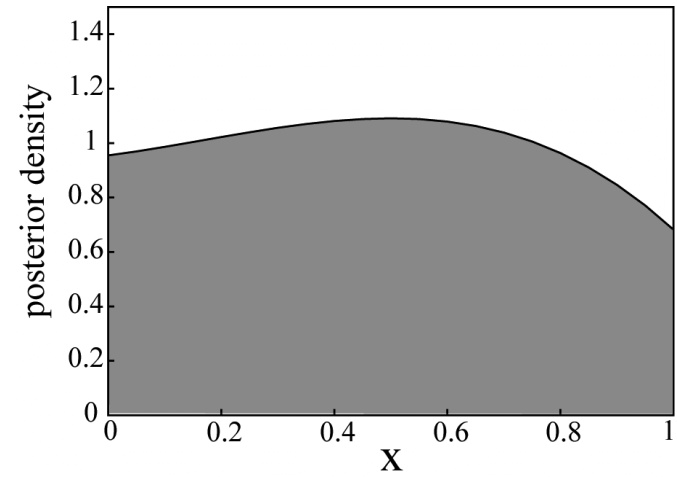
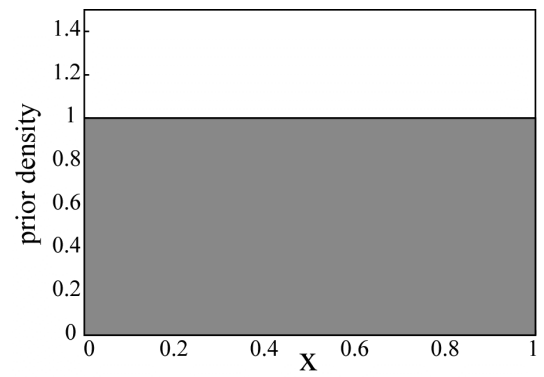
(if you're a Bayesian)

Reflect your uncertainty

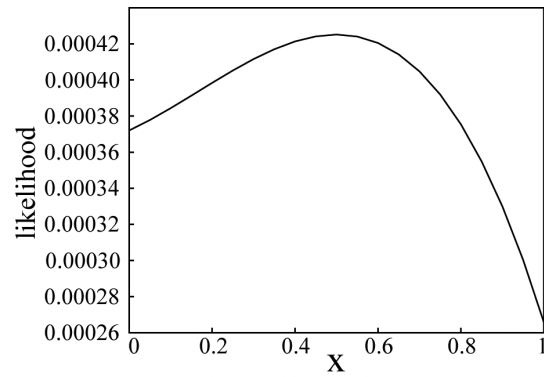


X

\propto

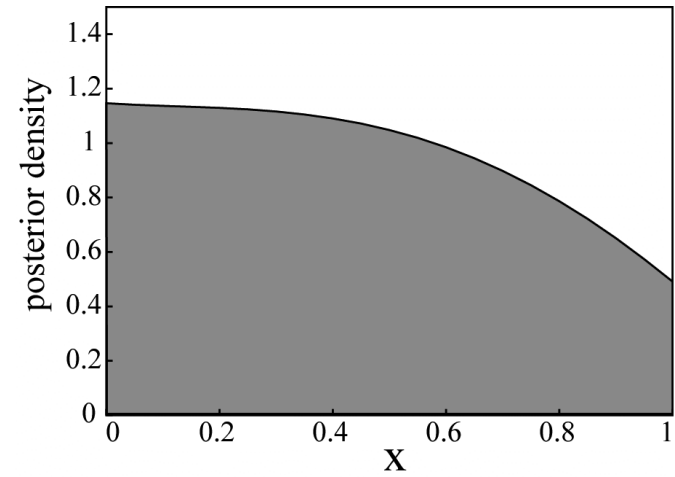
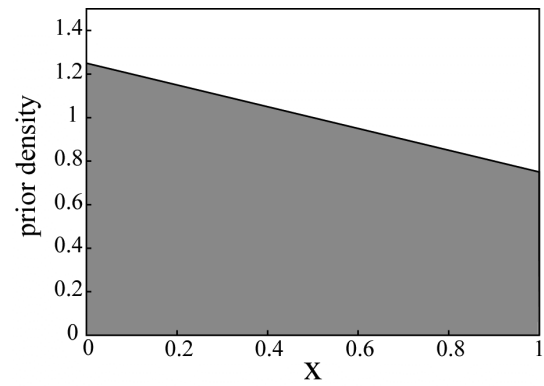


slide courtesy of Derrick Zwickl

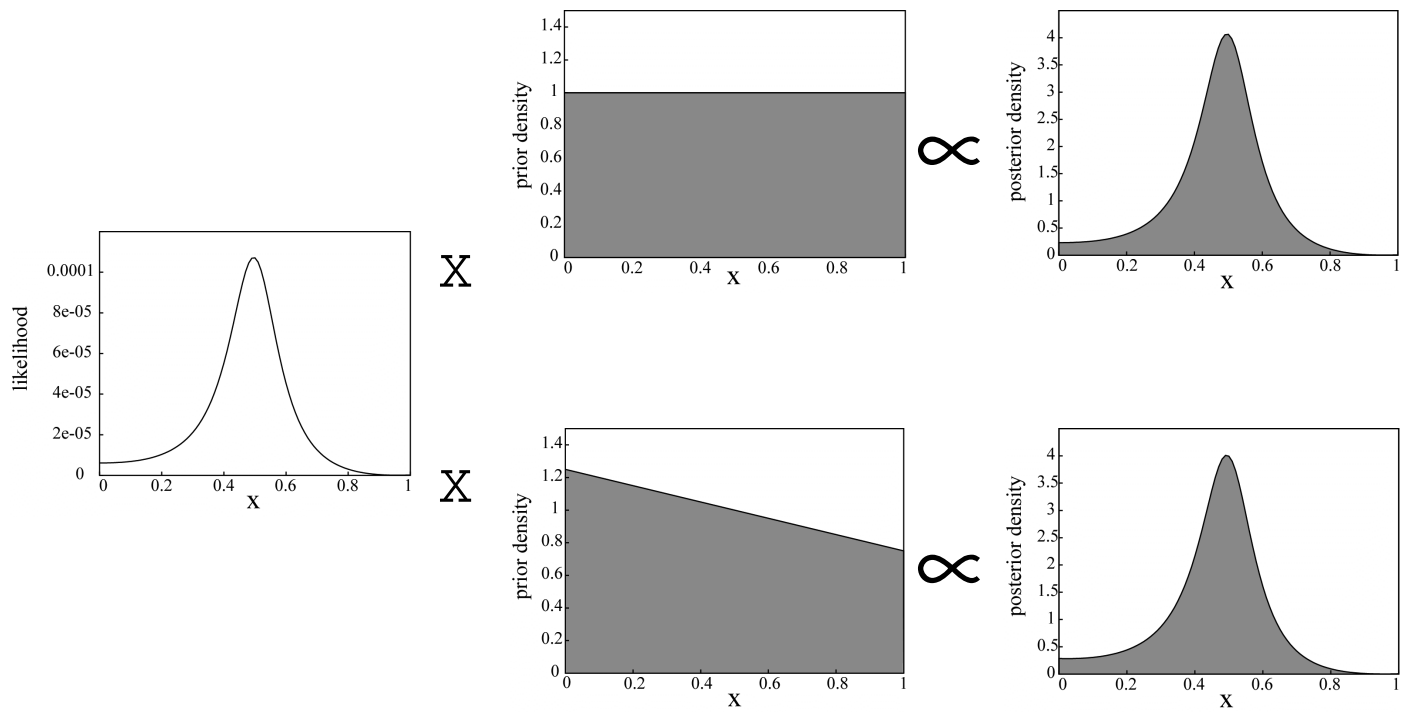


X

\propto



slide courtesy of Derrick Zwickl

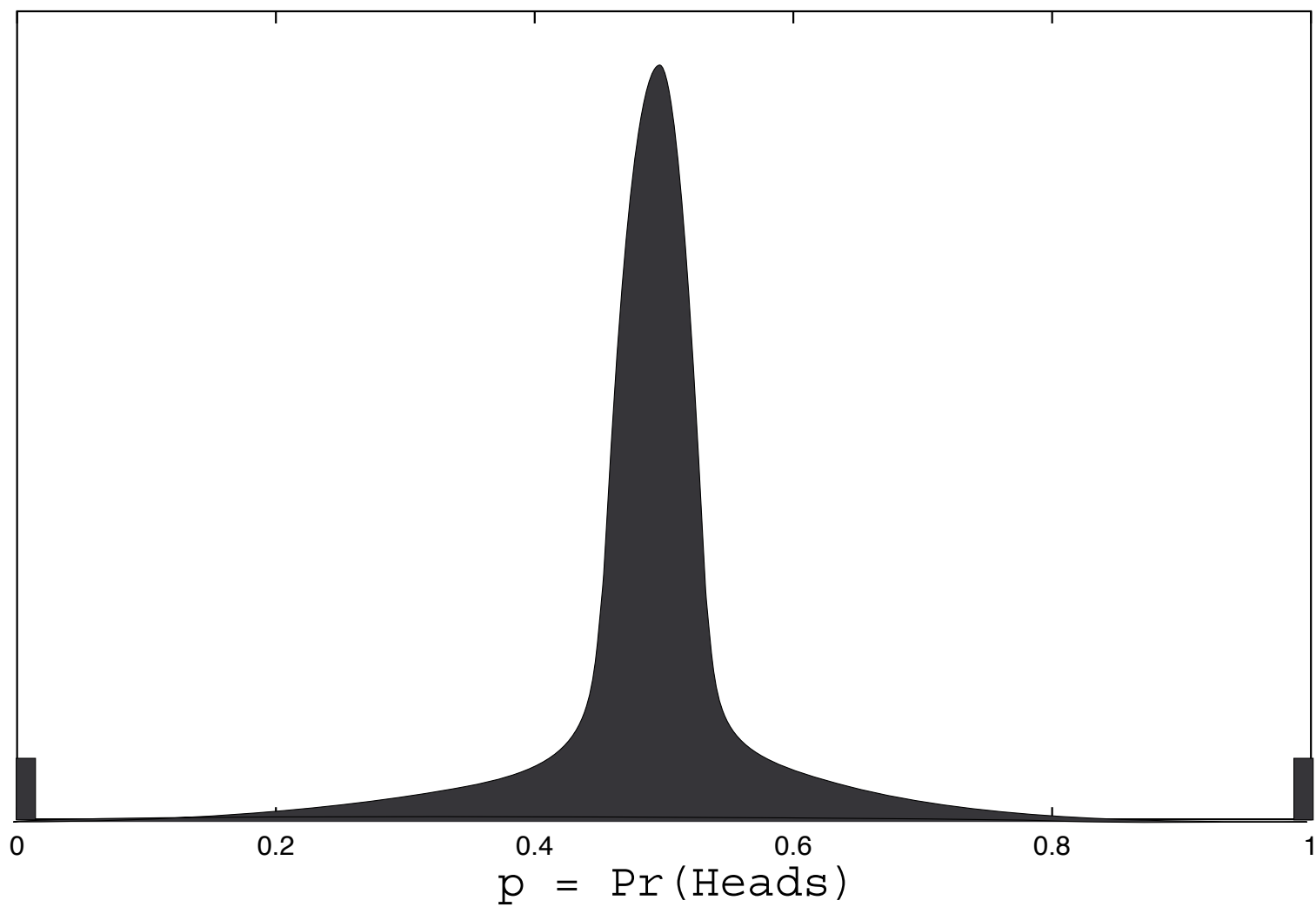


slide courtesy of Derrick Zwickl

Considerations when choosing a prior for a parameter

- What values are most likely?

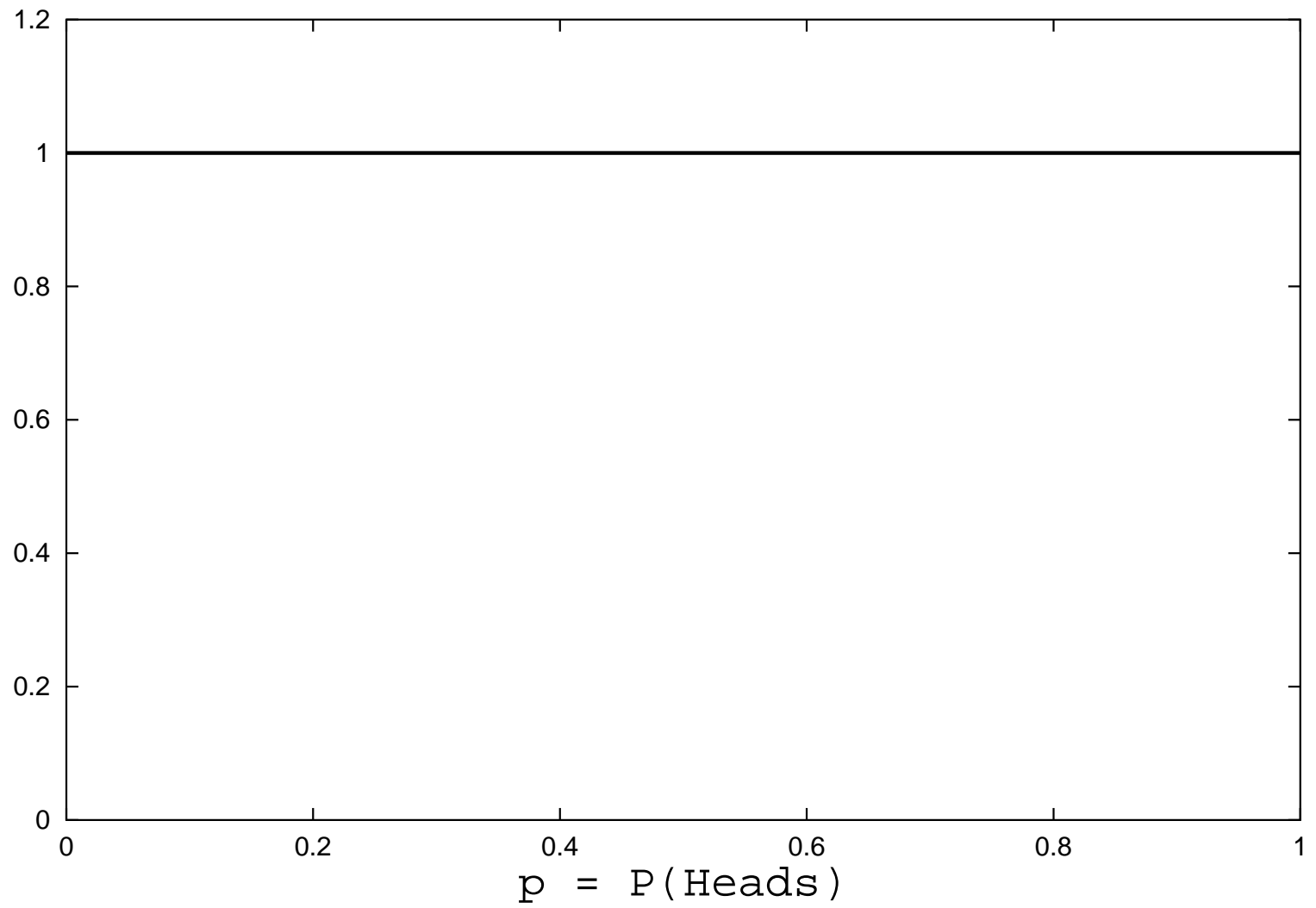
Subjective Prior on $\text{Pr}(\text{Heads})$



Considerations when choosing a prior for a parameter

- What values are most likely?
- How do you express ignorance?
 - *vague* distributions

Flat Prior on p



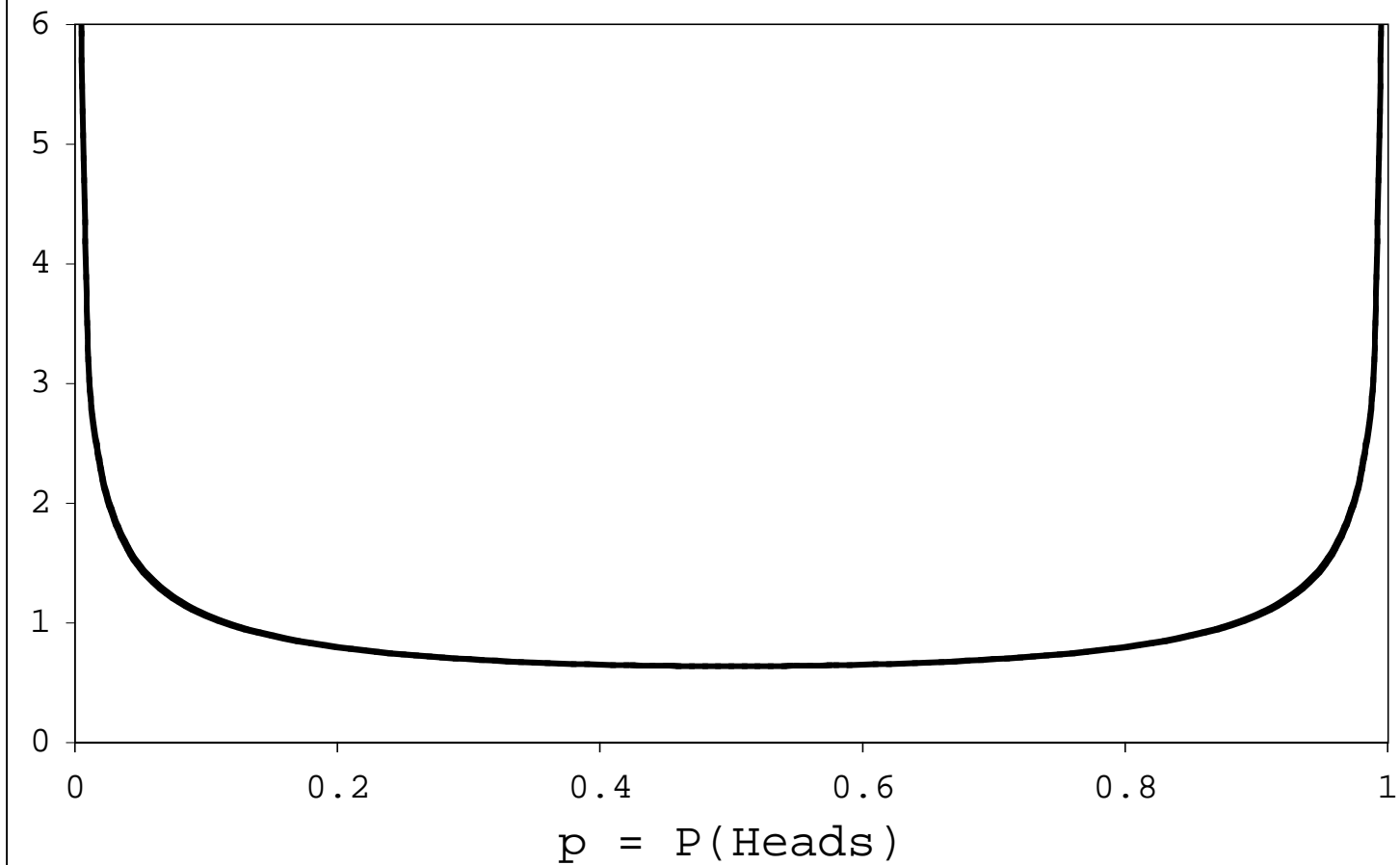
“Non-informative” priors

- Misleading term
- Used by many Bayesians to mean “prior that is expected to have the smallest effect on the posterior”
- **Not** always a uniform prior

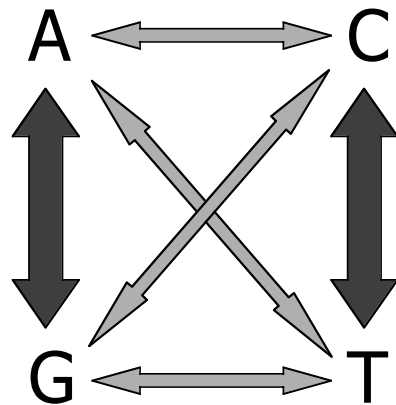
Considerations when choosing a prior for a parameter

- What values are most likely?
- How do you express ignorance?
 - *vague* distributions
 - **How easily can the likelihood discriminate between parameter values?**

Jeffrey's (Default) Prior



Example: The Kimura model



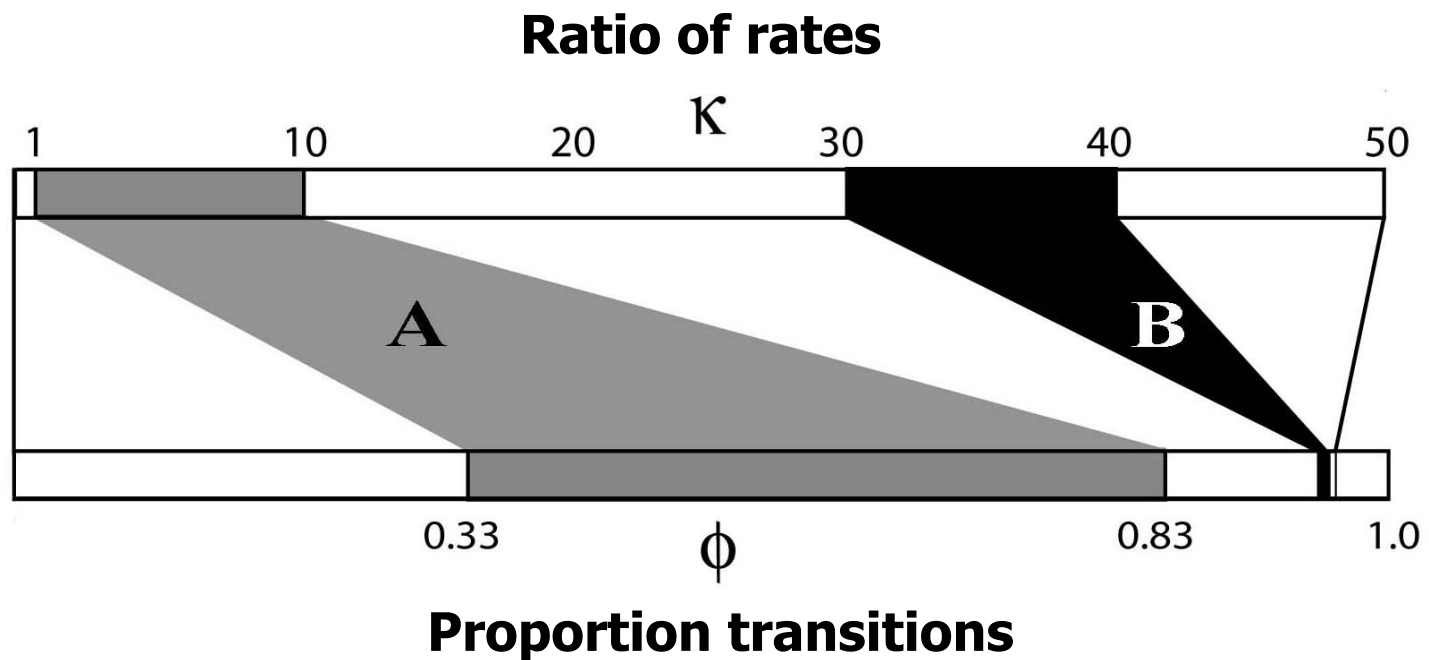
Ratio of rates $(0, \infty)$

$$\kappa = \frac{r_{ti}}{r_{tv}}$$

Proportion transitions $(0, 1)$

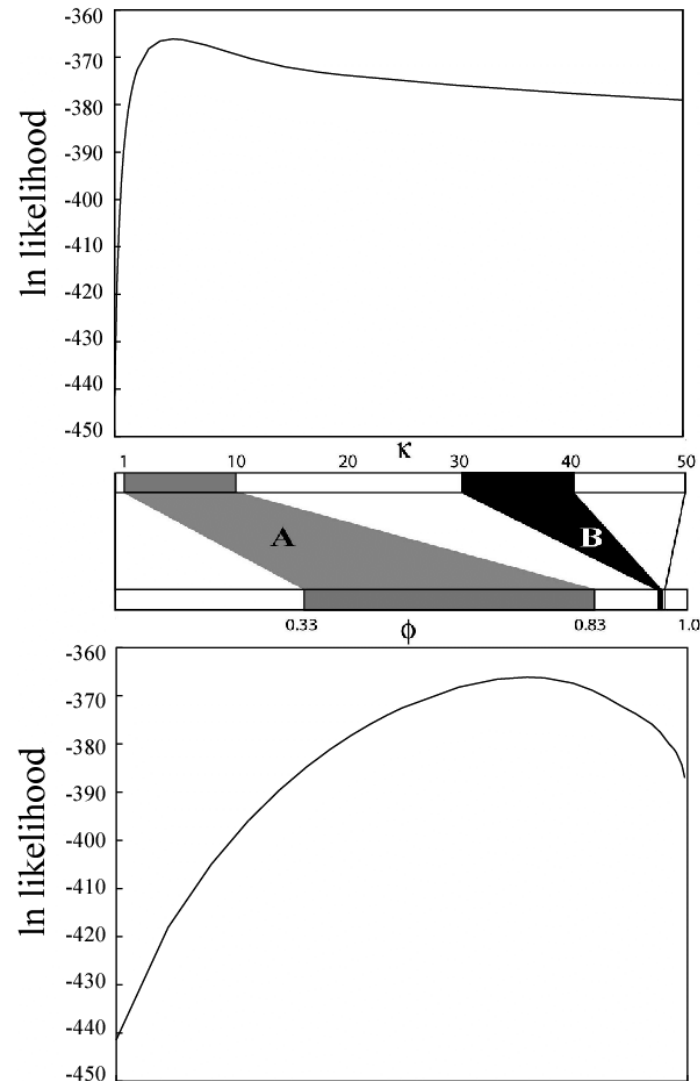
$$\phi = \frac{r_{ti}}{r_{ti} + 2r_{tv}}$$

- κ and ϕ map onto the predictions of K80 very differently

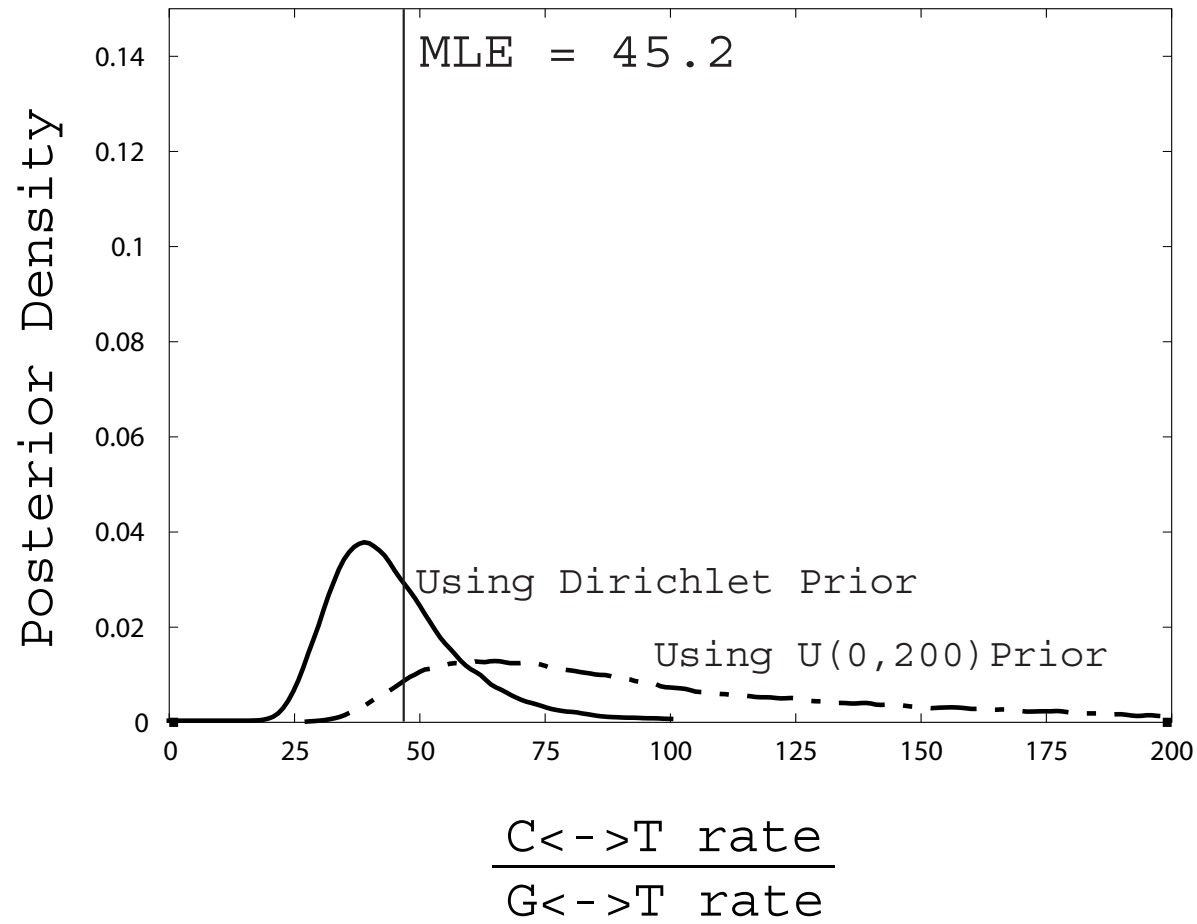


K80 : κ and ϕ

- The likelihood surface is tied to the model predictions
- The ML estimates are equivalent
- The curve shapes (and integrals) are quite different



Effects of the Prior in the GTR model

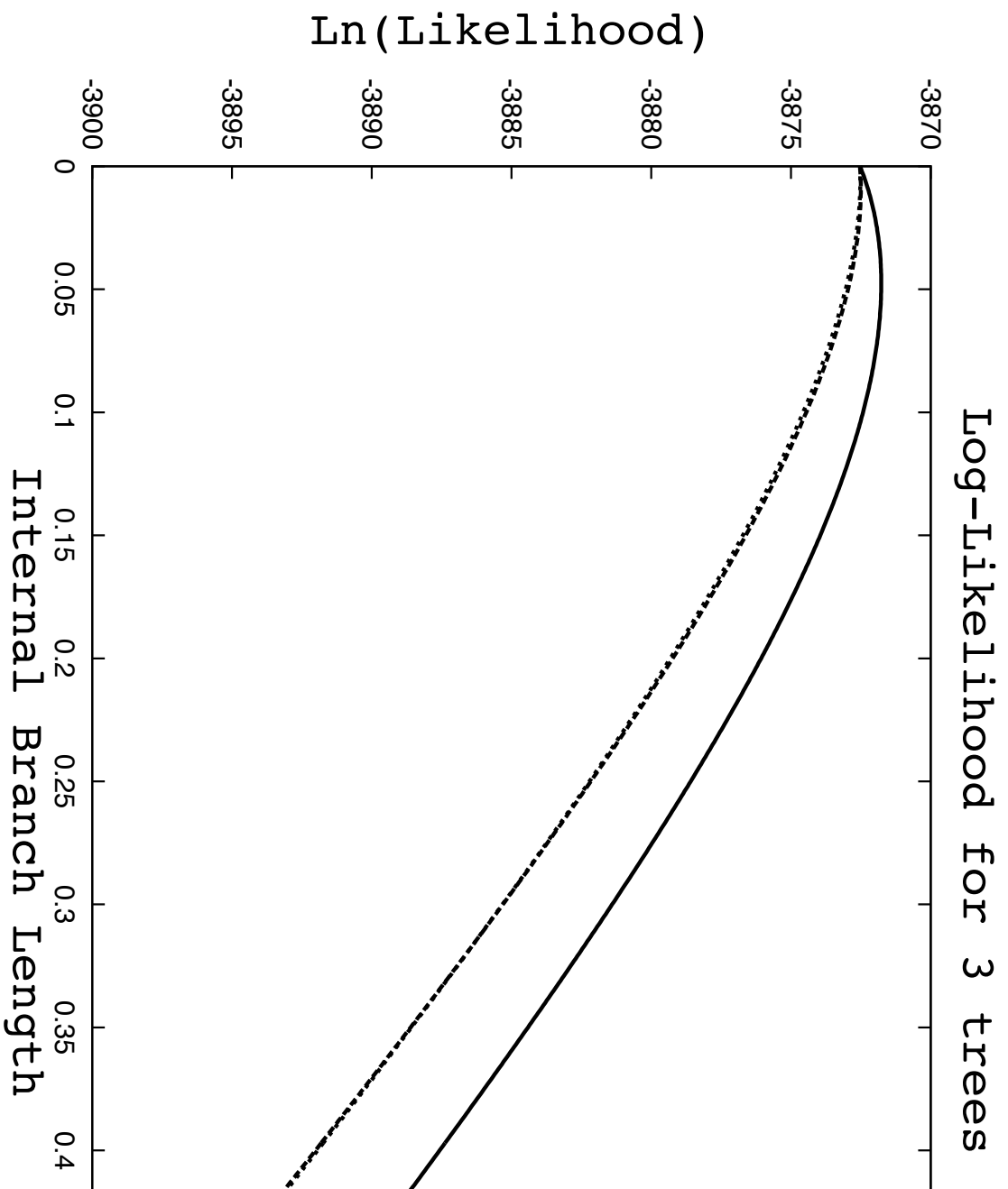


Minimizing the effect of priors

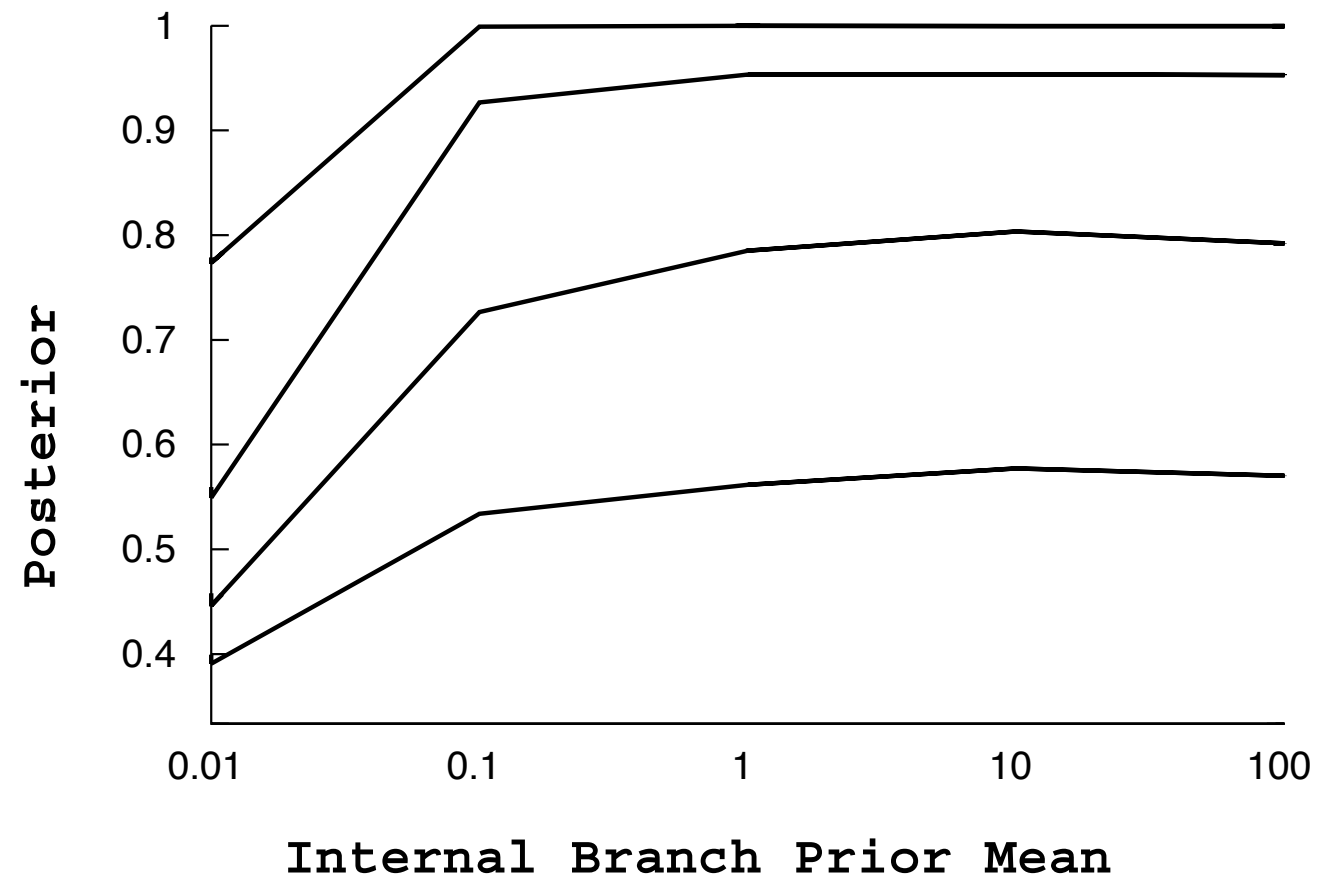
- Flat \neq non-informative
- Familiar model parameterizations may perform poorly in a Bayesian analysis with flat priors.

Considerations when choosing a prior for a parameter

- What values are most likely?
- How do you express ignorance? (minimally informative priors)
- **Are some errors better than others?**



The *Tree's* Posterior vs
the *Branch Length's* Prior Mean



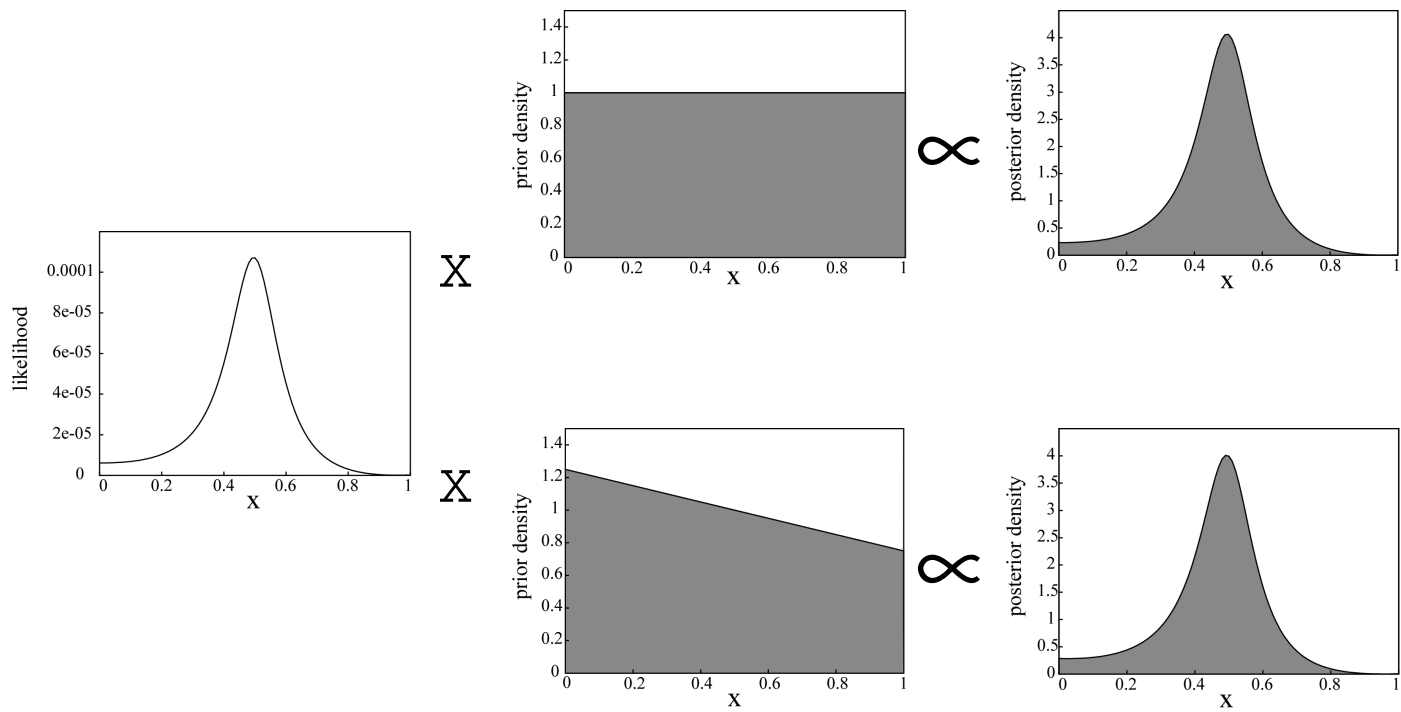
We might make analyses more conservative by

- Favoring short internal branch lengths
- Placing some prior probability on “star” trees (Lewis *et al.*)

We need to worry about sensitivity of our conclusions to all “inputs”

- Data
- Model
- Priors

Often priors will be the *least* of our concerns



slide courtesy of Derrick Zwickl

The prior can be a benefit (not just a necessity)
of Bayesian analysis

- Incorporate previous information
- Make the analysis more conservative

But...

It can be hard to say “I don’t know”

Priors can strongly affect the analysis *if...*

- The prior strongly favors some parameter values, OR
- The data (via the likelihood) are not very informative (little data or complex model)

Because Bayesian inference relies on marginalization, the priors for *all* parameters can affect the posterior probabilities of the hypotheses of interest.

How do we calculate a posterior probability?

$$Pr(\text{Tree}|\text{Data}) = \frac{Pr(\text{Tree})L(\text{Tree})}{\mathbf{Pr}(\mathbf{Data})}$$

In particular, how do we calculate **Pr(Data)**?

$Pr(\text{Data})$ is the marginal probability of the data, so

$$Pr(\text{Data}) = \sum_i Pr(\text{Tree}_i) L(\text{Tree}_i)$$

But this is a sum over all trees (there are ***lots*** of trees).

Recall that even $L(\text{Tree}_i)$ involves multiple integrals.

$$Pr(\mathbf{D}) = \sum \int \int \int \int \int \text{Posterior Probability Density}$$

$$L(\text{Tree}_i, \kappa, \alpha, \nu_1, \nu_2, \nu_3, \nu_4, \nu_5) Pr(\text{Tree}_i) Pr(\kappa) Pr(\alpha) Pr(\nu_1) Pr(\nu_2) \cdots$$

Markov chain Monte Carlo

- 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

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

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

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