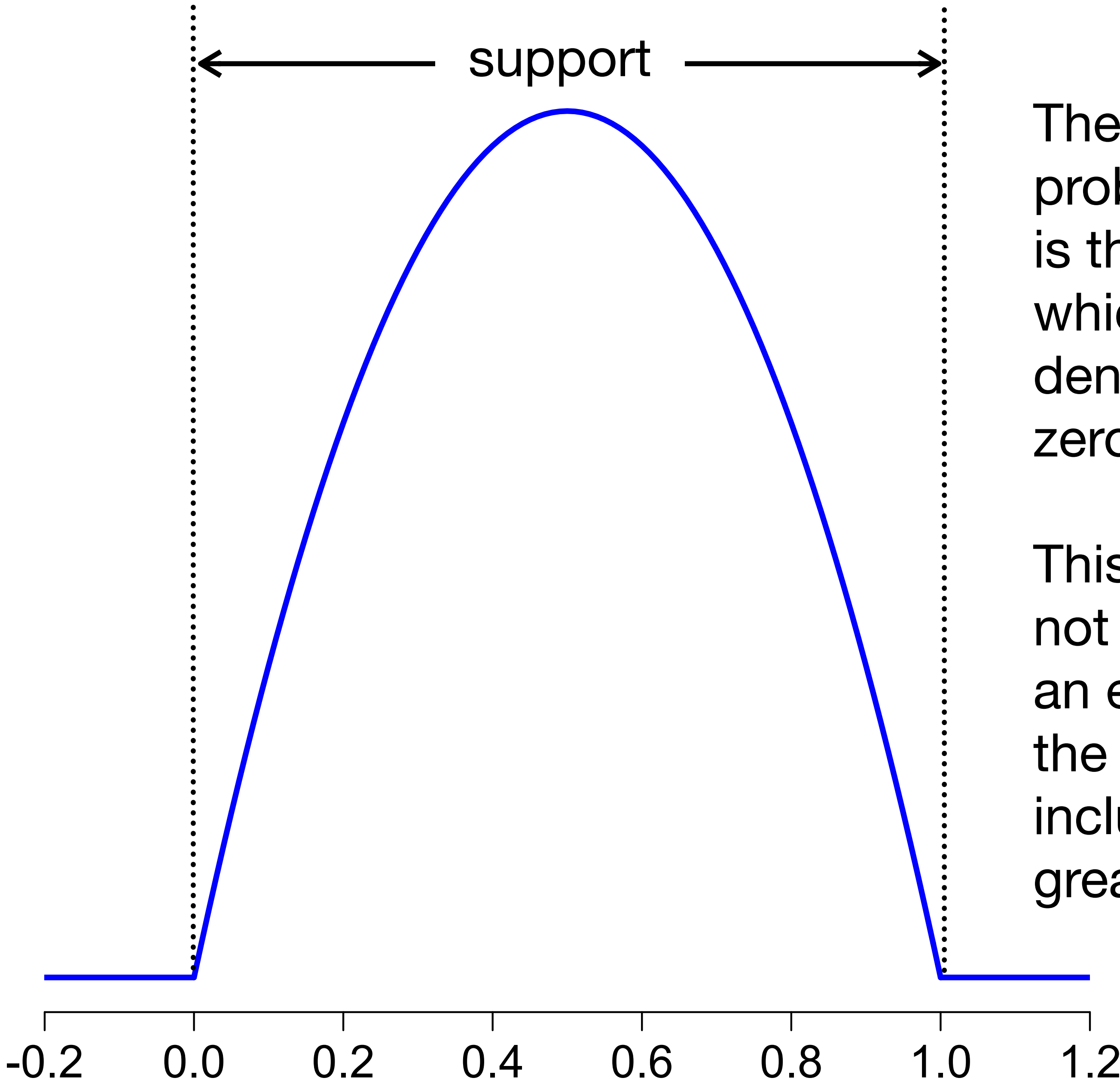
Interval in which density > 0 is the **support**



The **support** of a probability distribution is the interval over which its probability density is greater than zero.

This distribution would not be an ideal prior for an edge length because the support does not include any value greater than 1.

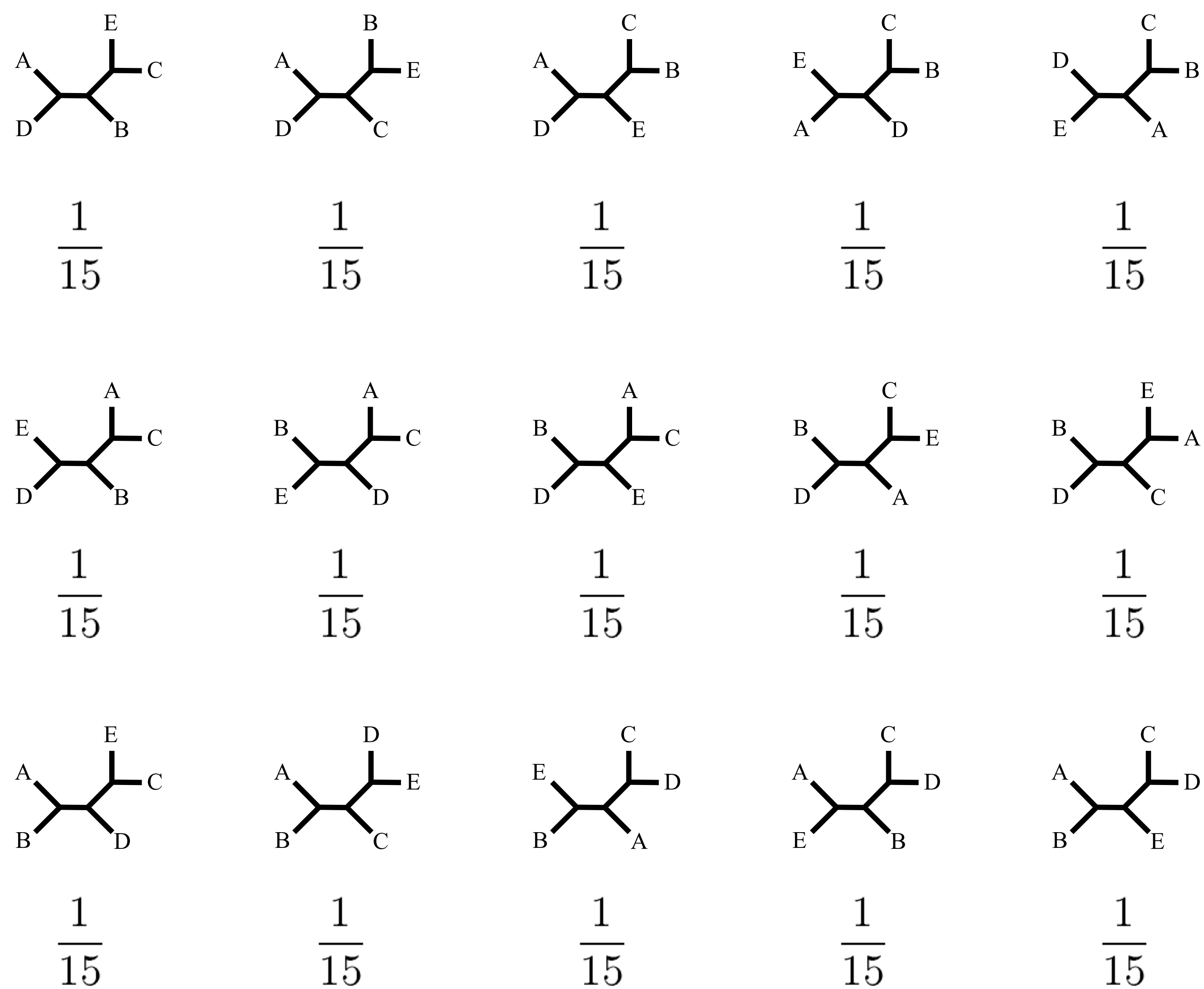
Tour of common priors

Prior distributions

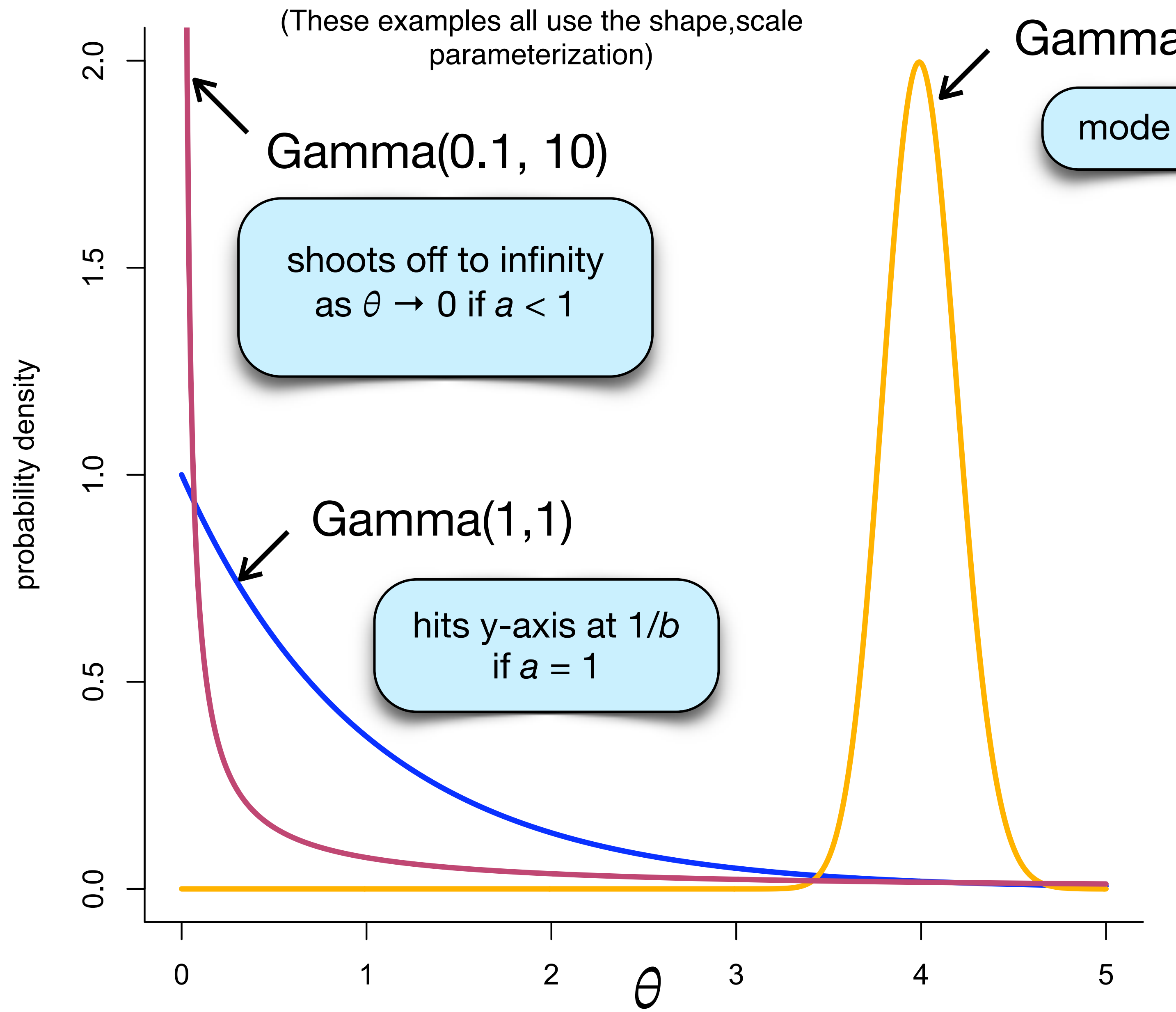
- **Discrete uniform** for topologies
- **Gamma** for edge lengths, tree length, rate ratios, and other parameters
- **Exponential** is a special case of Gamma
- **Lognormal** is an alternative to Gamma
- **Dirichlet** for state frequencies, GTR relative rates, edge length proportions
- **Beta** is a special case of Dirichlet
- **Uniform** is a special case of Beta

Commonly-used prior distributions

For **topologies**, a **discrete Uniform distribution** is common



Gamma distribution



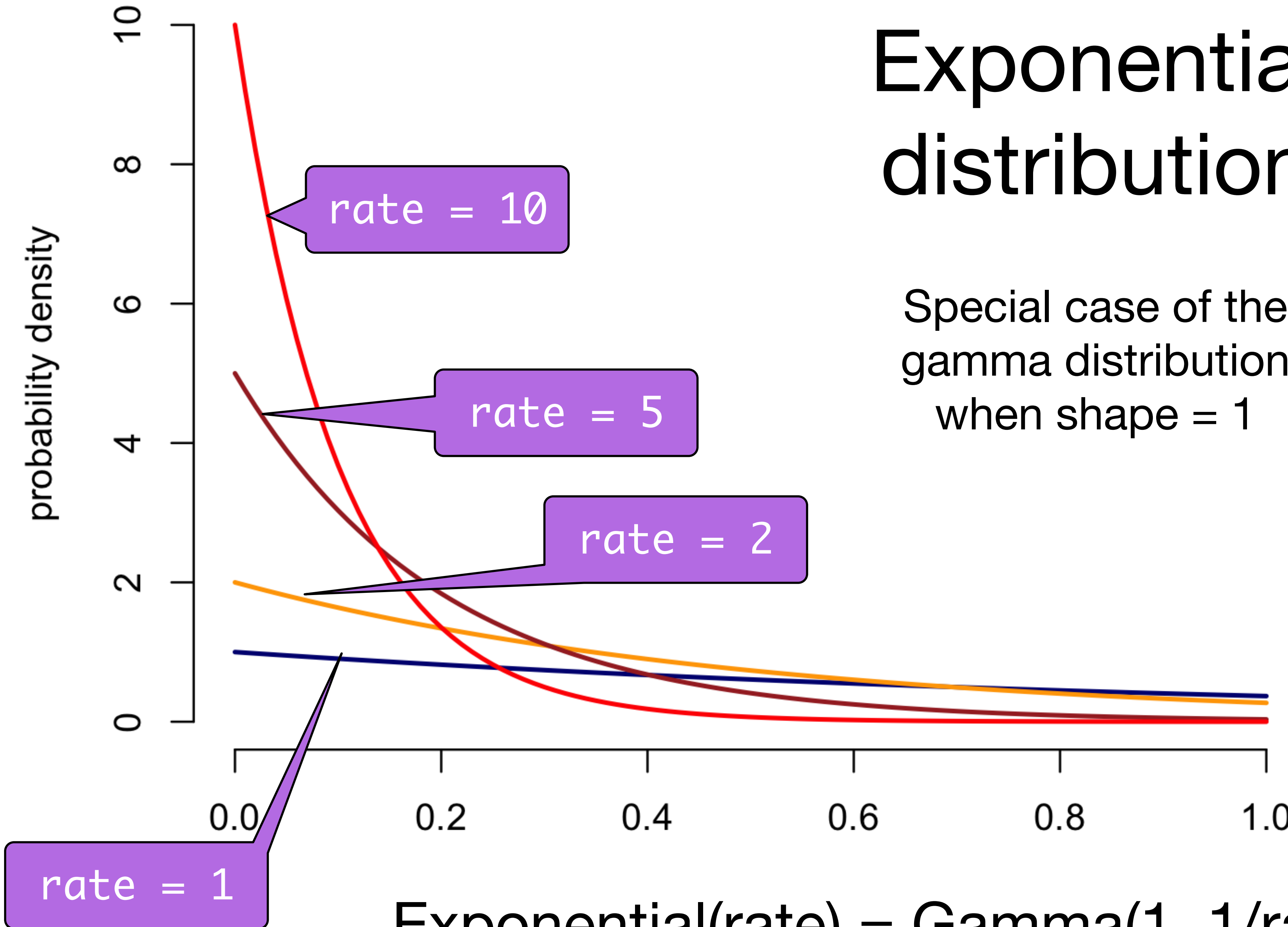
Gamma distributions are ideal for parameters that range from 0 to infinity (e.g. branch lengths)

a = shape
b = scale
mean = ab
variance = ab^2

a = shape
r = rate
mean = a/r
variance = a/r^2

Exponential distribution

Special case of the gamma distribution when shape = 1

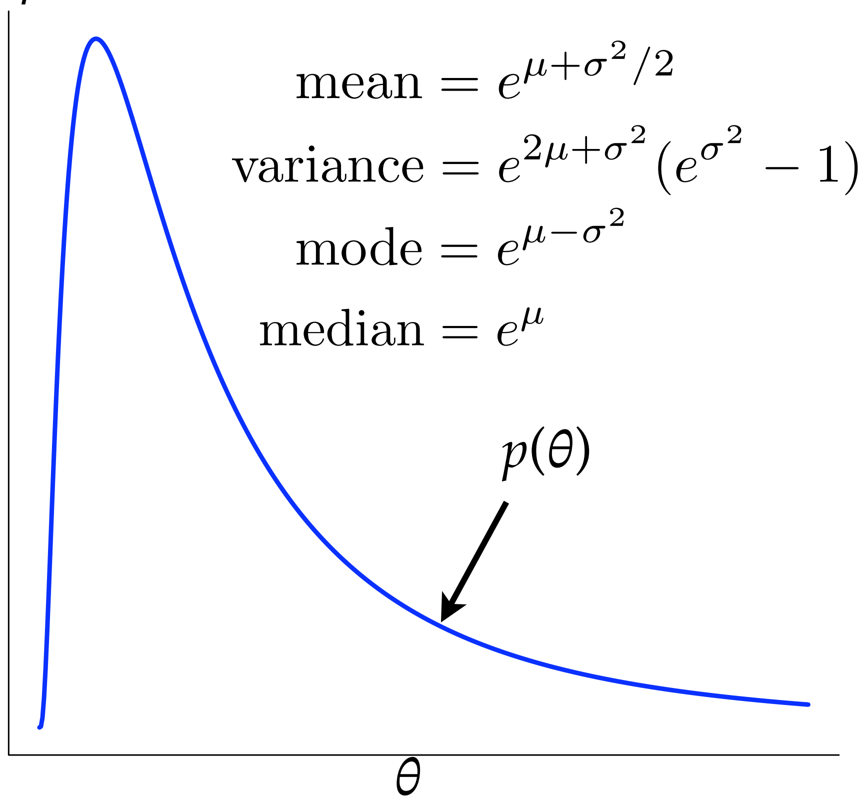


Exponential(rate) = Gamma(1, 1/rate)

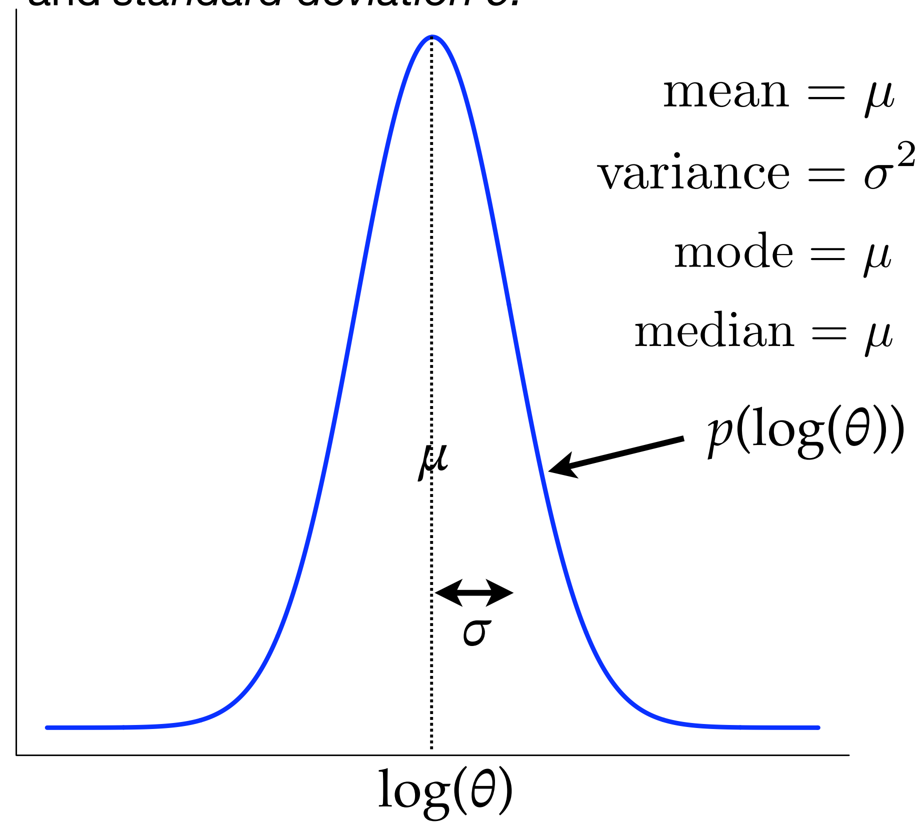
shape scale

Log-normal distribution

If θ is **log-normal** with *parameters* μ and σ...



...then **log(θ)** is **normal** with *mean μ* and *standard deviation σ.*

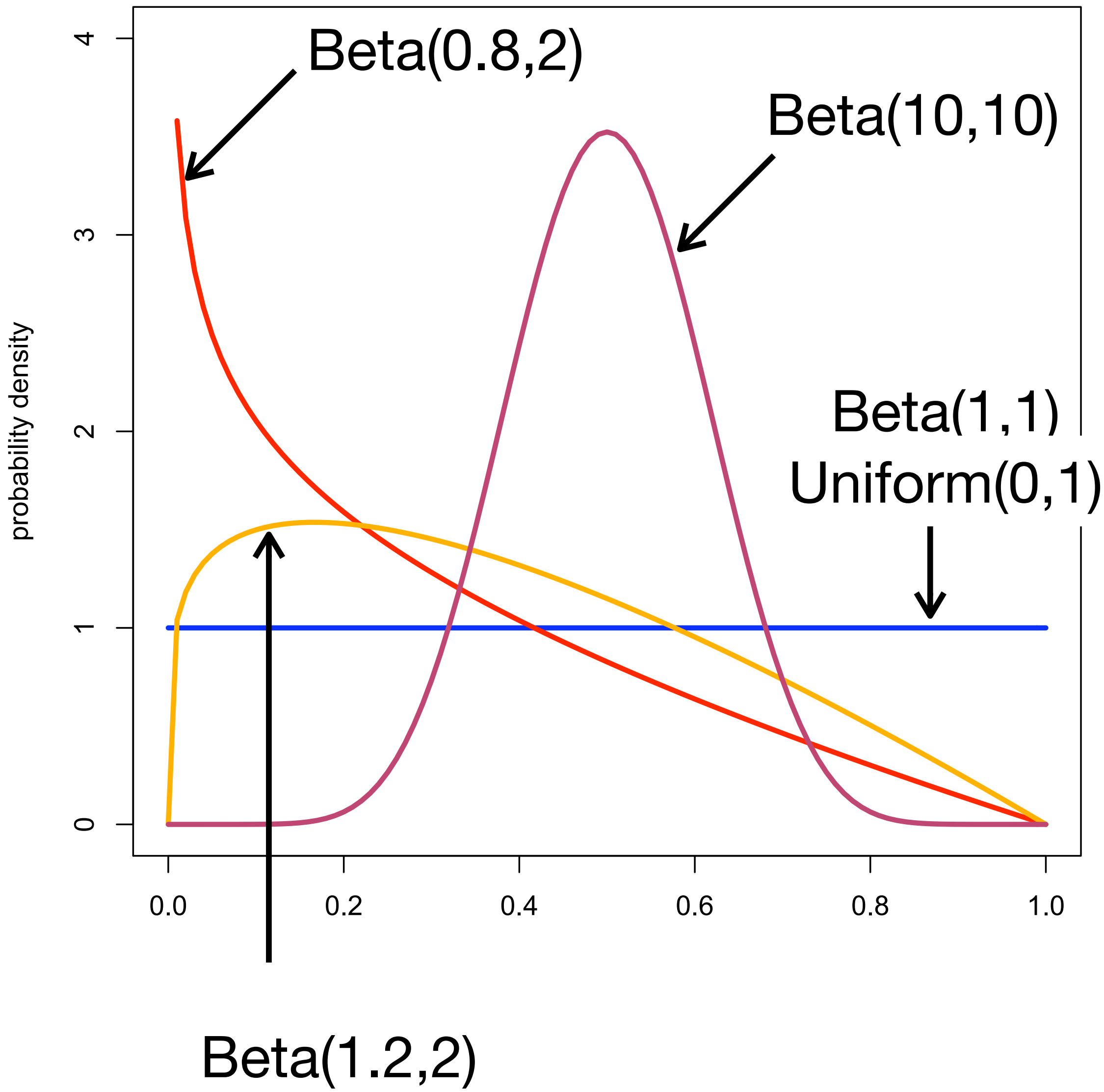


Important: *μ and σ do **not** represent the mean and variance of θ: they are the mean and variance of log(θ)!*

To choose μ and σ to yield a particular mean (m) and variance (v) for θ, use these formulas:

$$\sigma^2 = \log\left(1 + \frac{v}{m^2}\right) \qquad \mu = \log(m) - \sigma^2/2$$

Beta distribution



mean

$$\frac{a}{a+b}$$

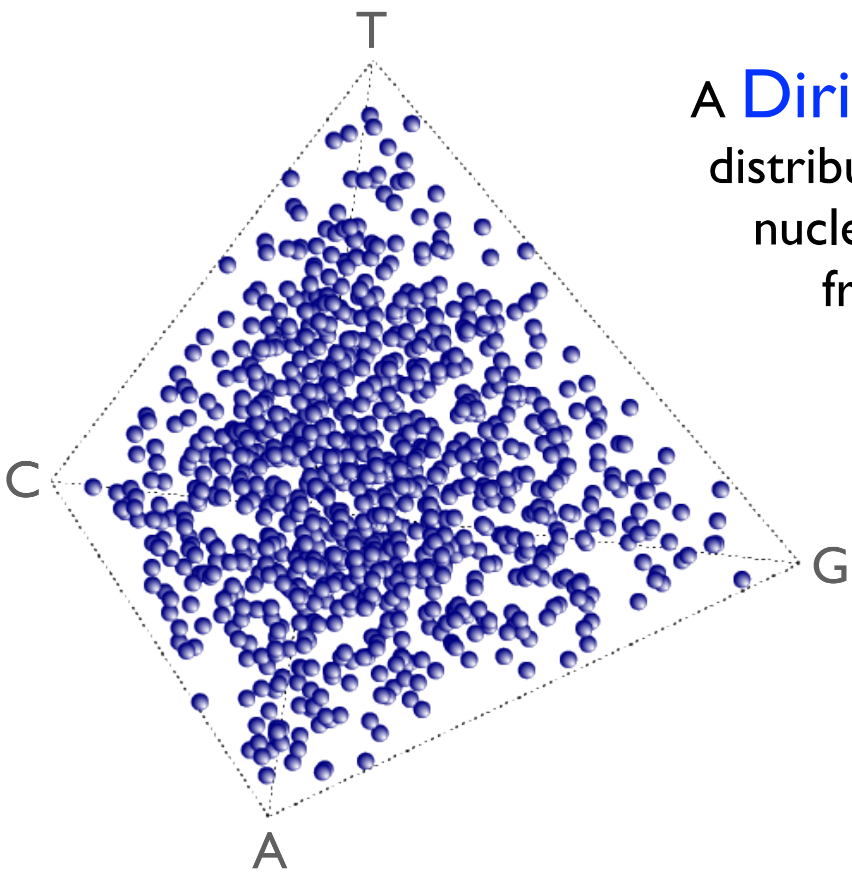
variance

$$\frac{(\text{mean})(1-\text{mean})}{a+b+1}$$

Dirichlet(*a*,*b*,*c*,*d*) distribution

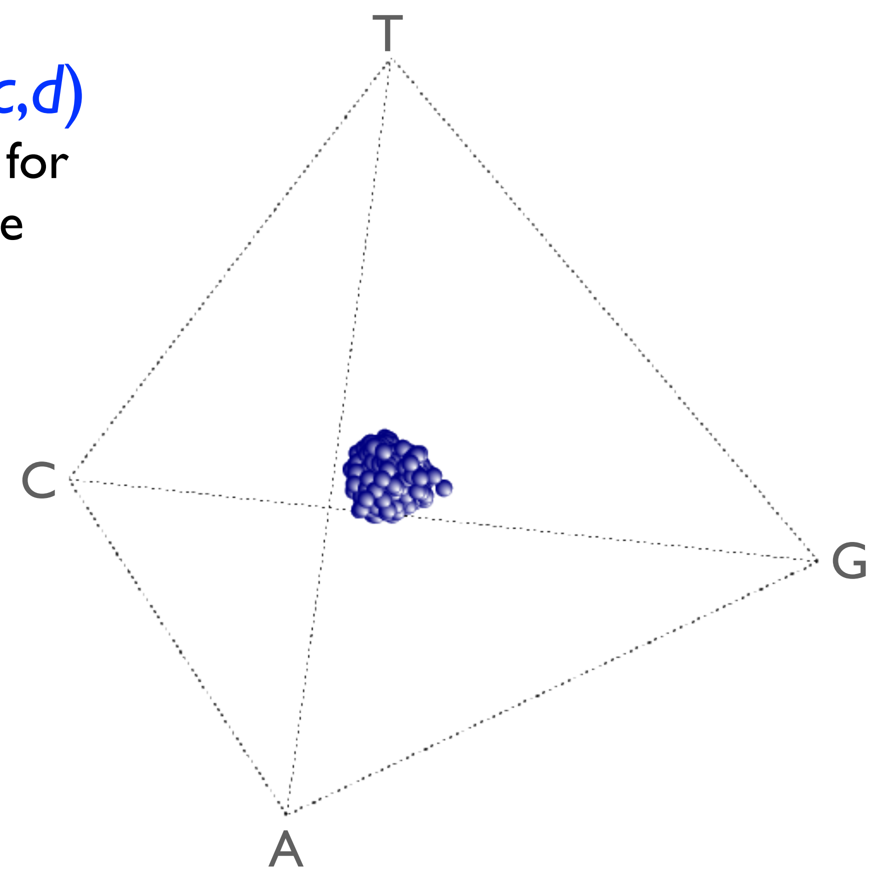
Flat: $a = b = c = d = 1$

(every combination equally probable)



Informative: $a = b = c = d = 100$

(frequencies tend to be nearly equal)



A Dirichlet(a,b,c,d) distribution is ideal for nucleotide relative frequencies.

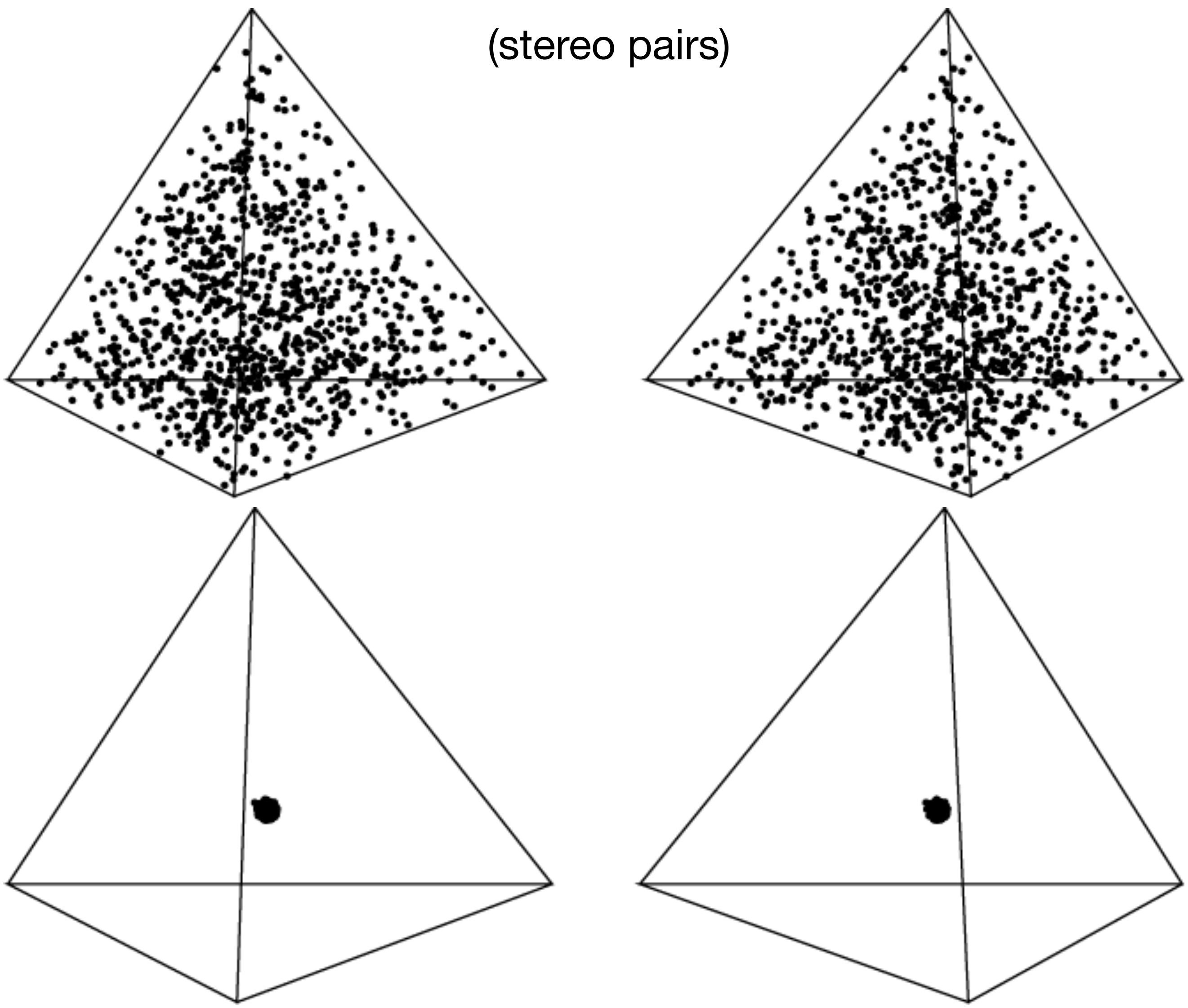
https://plewis.github.io/applets/dirichlet-prior/

Is there information in data about nucleotide frequencies?

(stereo pairs)

Prior

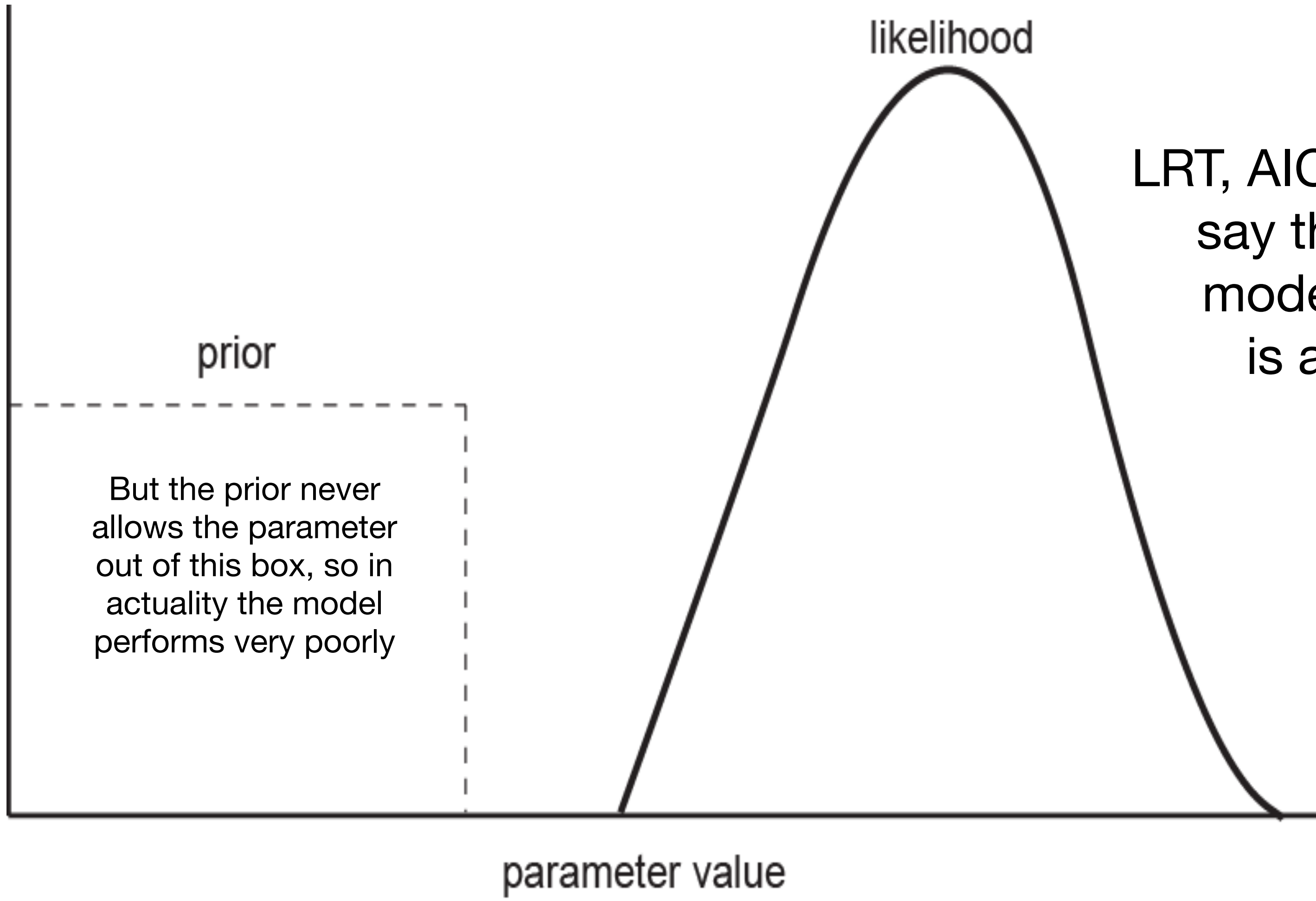
Posterior



1688 aligned nucleotide sites for 34 taxa

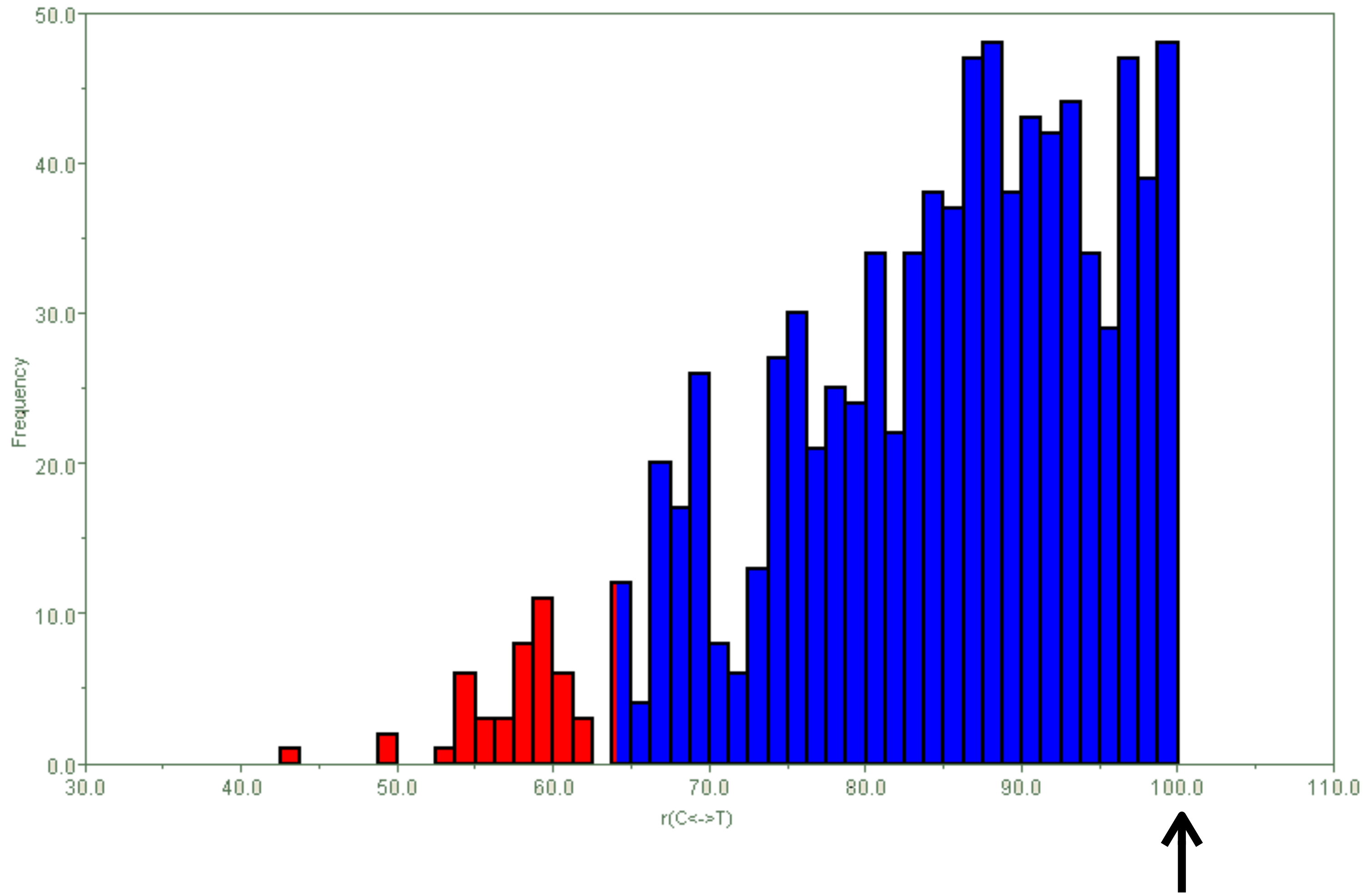
Mending (prior) fences

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



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

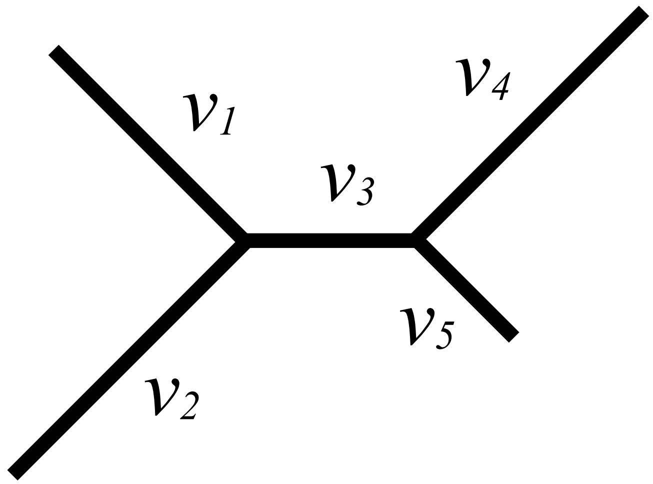
Marginal posterior distribution of r_{CT}



Zwickl and Holder (2004)

Beware of induced priors

Induced tree length prior



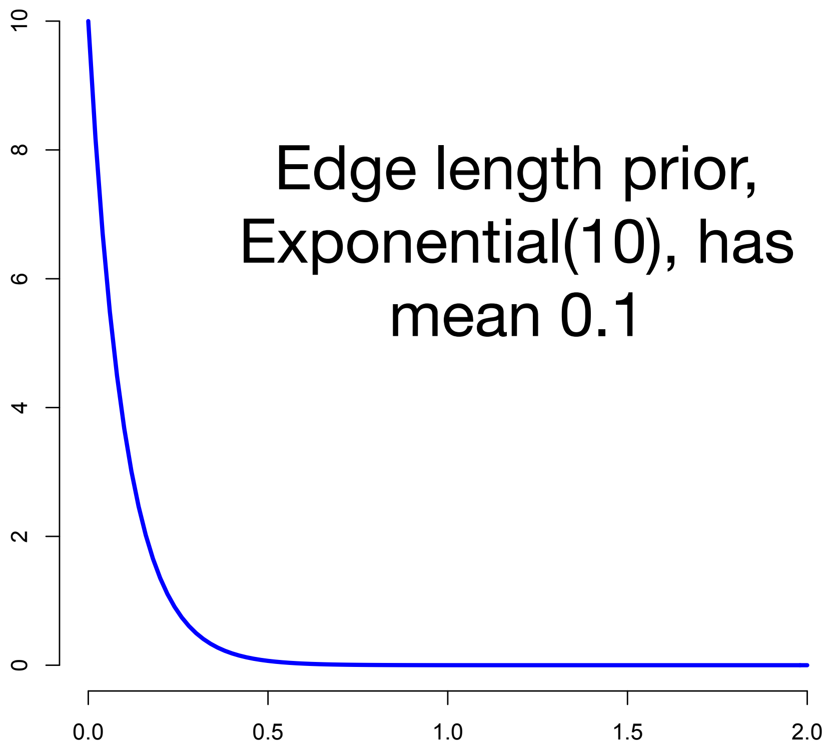
$$T = v_1 + v_2 + v_3 + v_4 + v_5$$

Prior placed on edge lengths **induces** a prior on tree length *T*

$$v_i \sim \text{Exponential}(\lambda)$$ ⟵ Mean = 1/λ

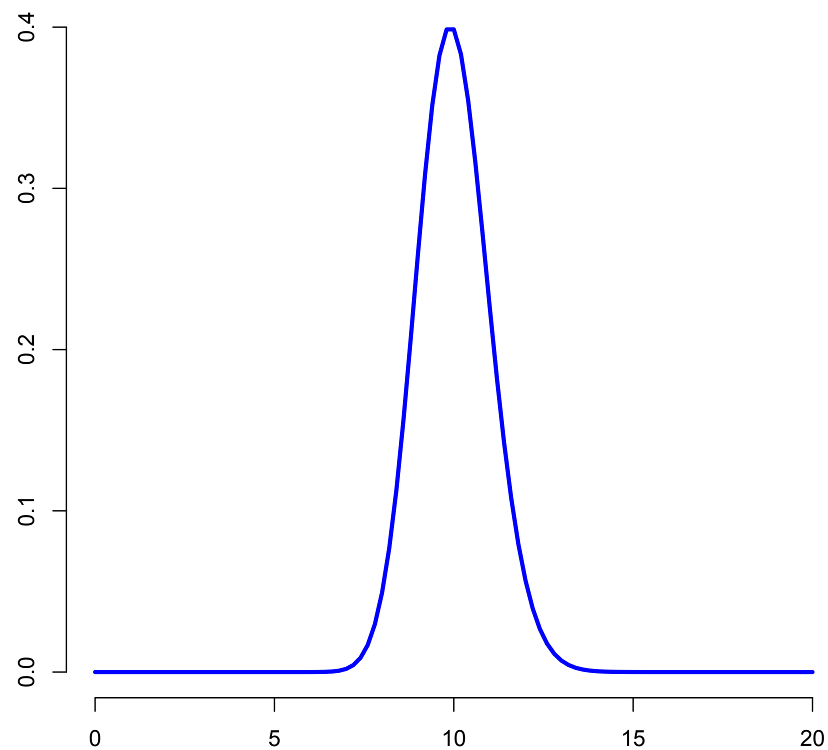
$$T \sim \text{Gamma}\left(5, \frac{1}{\lambda}\right)$$ ⟵ Mean = 5/λ

Edge length vs. **tree** length



Edge length prior, Exponential(10), has mean 0.1

For tree with 100 edges, the induced tree length prior, Gamma(100, 0.1), has mean 10



Gamma-Dirichlet prior

Better to place a prior on *T* directly and place a separate prior on edge length *proportions*:

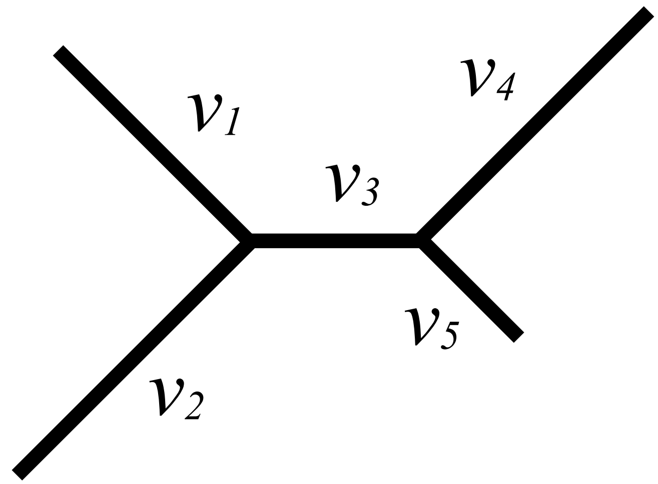
$$p(T)$$

Gamma(a,λ)

$$p\left(\frac{v_1}{T}, \frac{v_2}{T}, \frac{v_3}{T}, \frac{v_4}{T}, \frac{v_5}{T}\right)$$

Dirichlet(c,c,d,c,c)

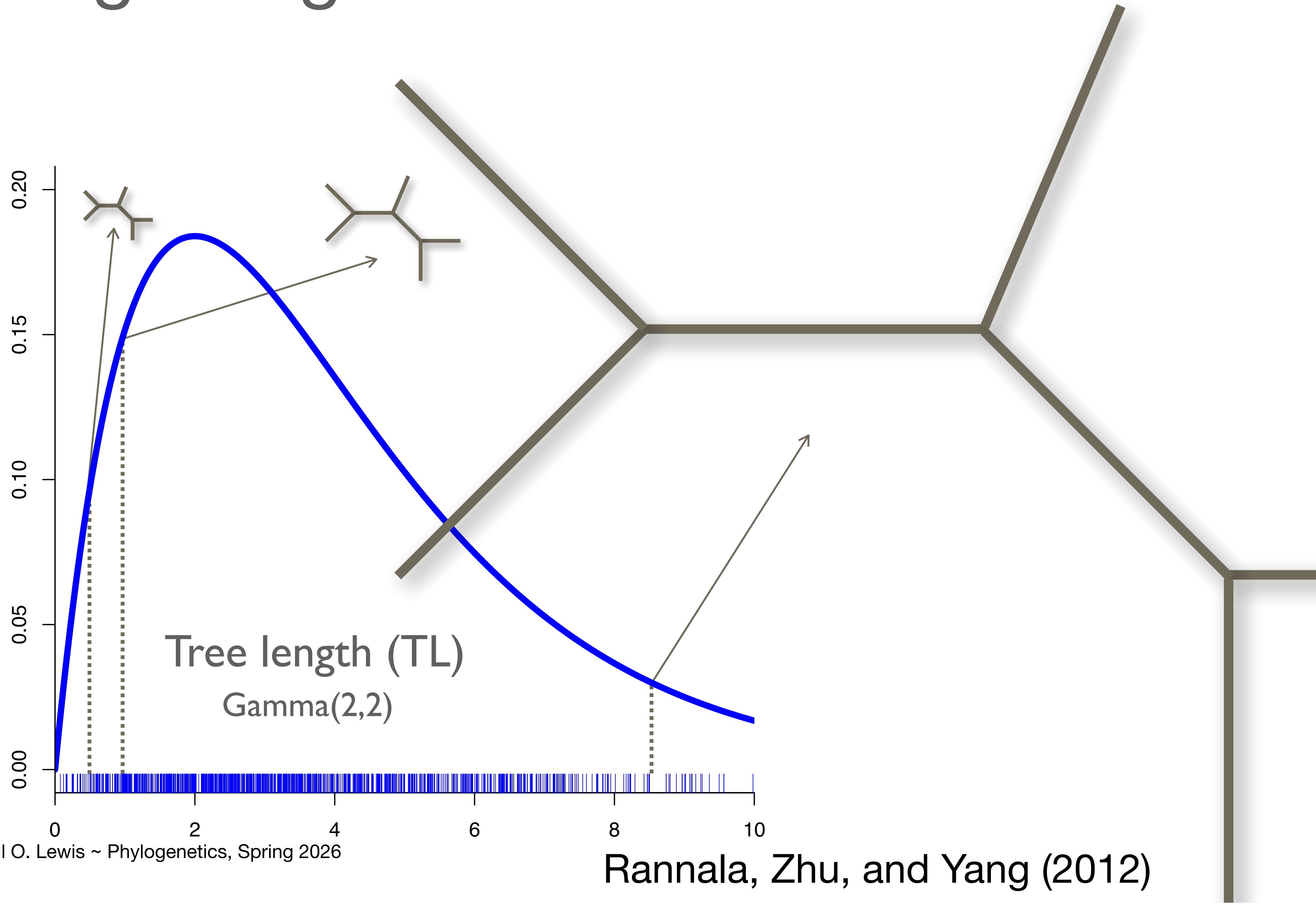
Note: *rate* parameterization: mean T = a/λ



c = d = 1 corresponds to a **flat prior** on edge length proportions; all edge length proportions have the same probability density.

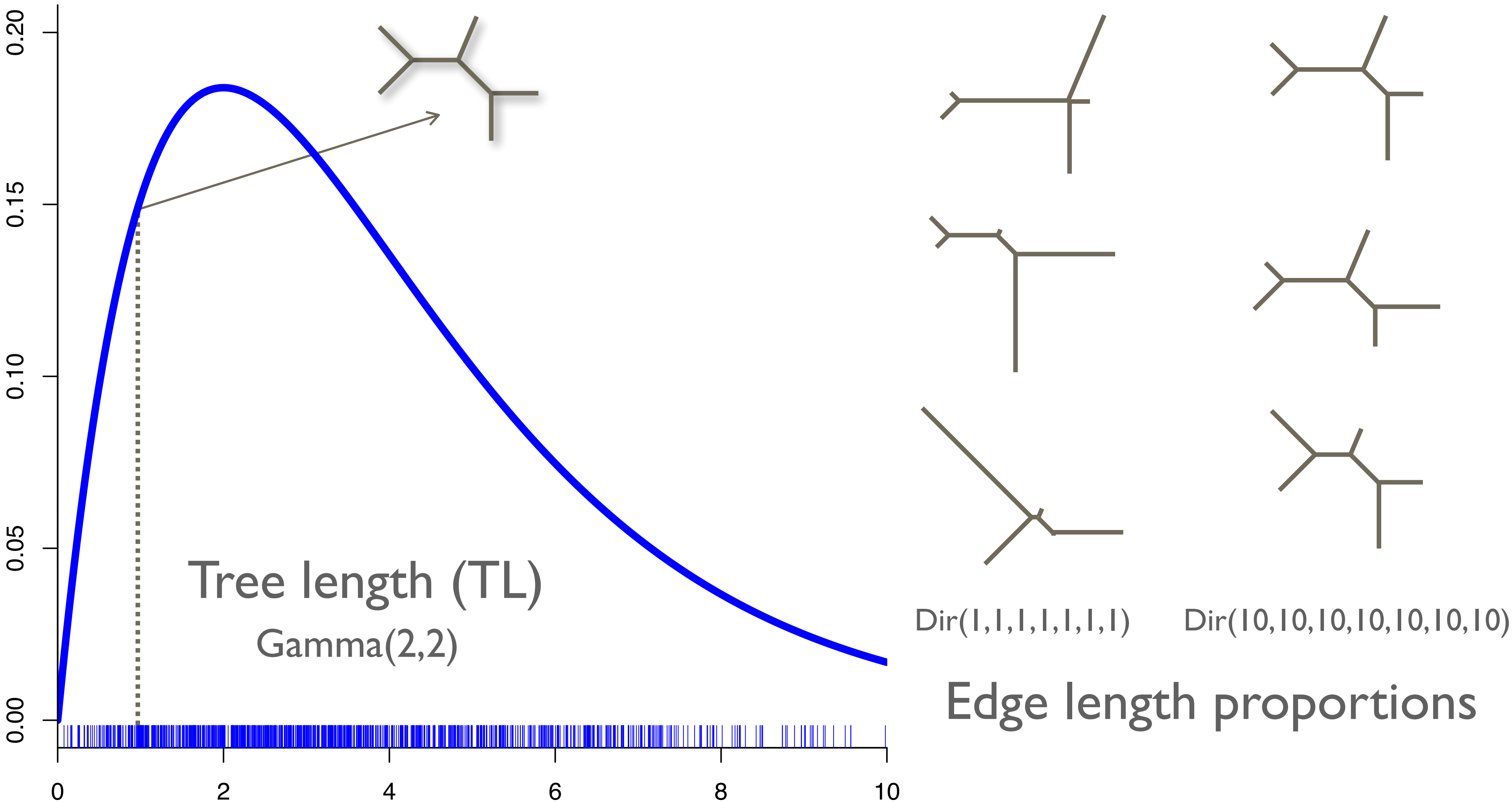
Rannala, Zhu, and Yang (2012)

Edge lengths: Gamma-Dirichlet

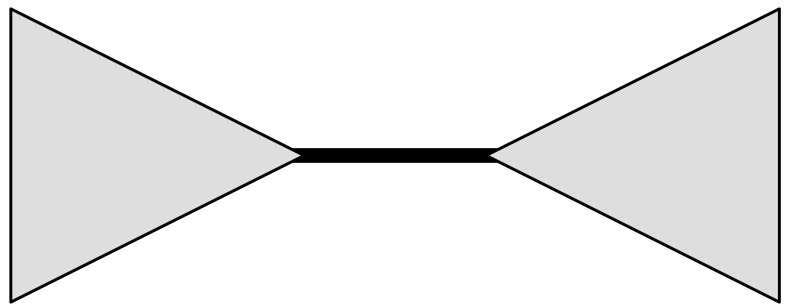


Rannala, Zhu, and Yang (2012)

Edge lengths: Gamma-Dirichlet



Induced prior on splits

m taxa  n taxa

$$\Pr(\text{split}) = \frac{\left[\begin{array}{c}\text{number of rooted}\\ \text{trees with m taxa}\end{array}\right]\left[\begin{array}{c}\text{number of rooted}\\ \text{trees with n taxa}\end{array}\right]}{\text{number of unrooted trees with n+m taxa}}$$

m=2, n=8: Pr(split) = 0.0667

m=5, n=5: Pr(split) = 0.0001

Running on empty

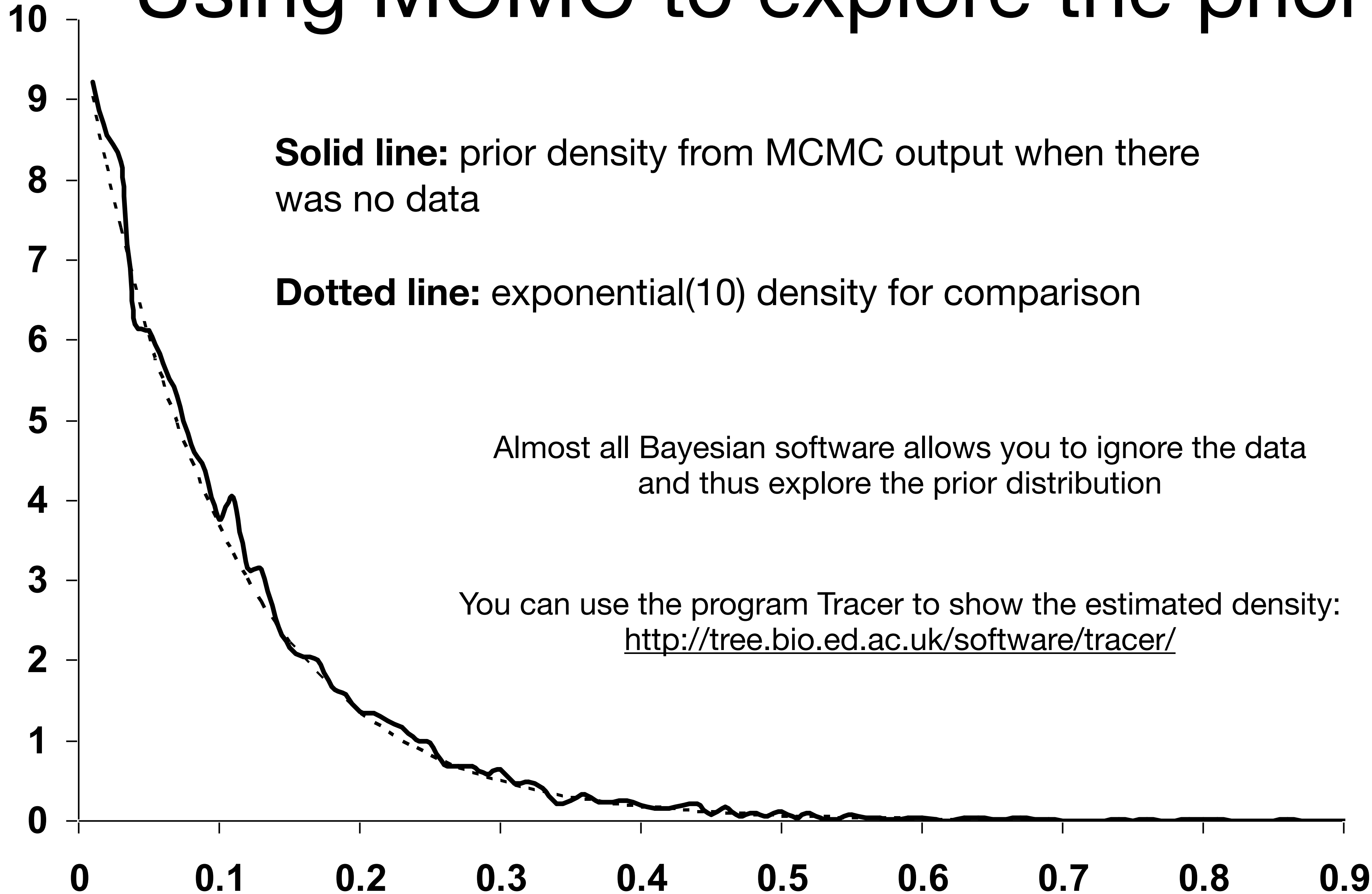
Using MCMC to explore the prior

Solid line: prior density from MCMC output when there was no data

Dotted line: exponential(10) density for comparison

Almost all Bayesian software allows you to ignore the data and thus explore the prior distribution

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

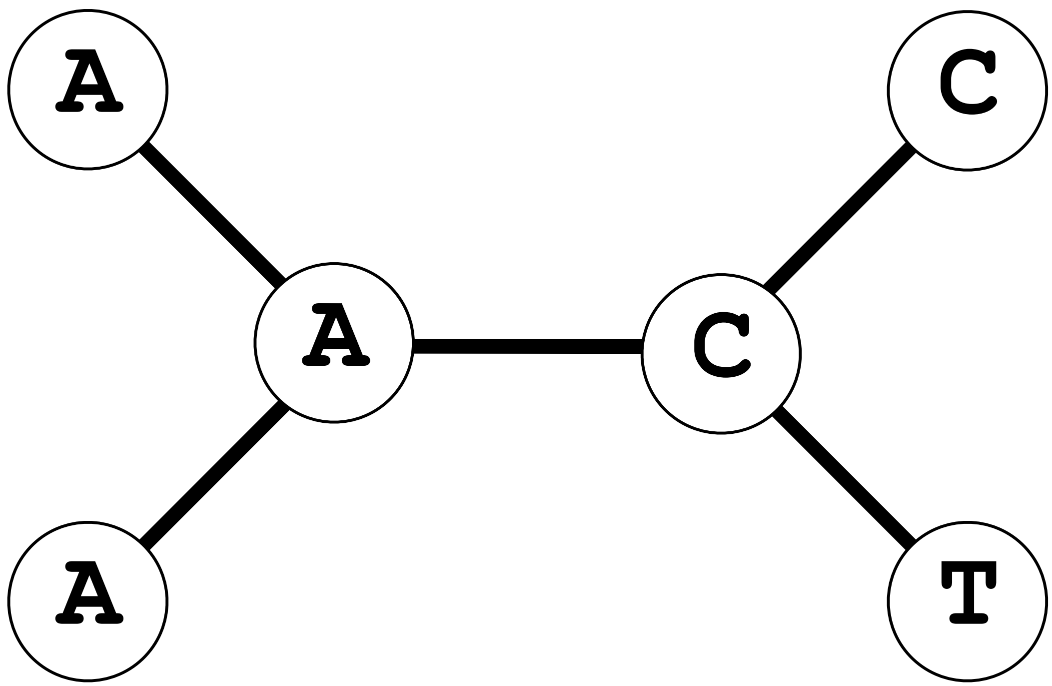


Hierarchical models
vs.
Empirical Bayes

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

Prior: Exponential, mean=0.1

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



Hierarchical models add *hyperparameters* not present in the likelihood function

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

hyperprior

Prior: Exponential, mean *μ*

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

During an MCMC analysis, *μ* will hover around a reasonable value, sparing you from having to decide what value is appropriate. You still have to specify a hyperprior, however.

Suchard, Weiss and Sinsheimer (2001)

Empirical Bayes

uses data, hence "empirical"

↓

Prior: Exponential, mean=MLE

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

Brown et al. (2010)